

# Predicted Genetic Gains from Targeted Recombination in Elite Biparental Maize Populations

Sofía P. Brandariz and Rex Bernardo\*

Dep. of Agronomy and Plant Genetics, Univ. of Minnesota, 411 Borlaug Hall, 1991 Upper Buford Circle, Saint Paul, MN 55108.

**ABSTRACT** Targeted recombination is the ability to induce or select for specific recombination points on chromosomes. A first study with the intermated B73 × Mo17 maize (*Zea mays* L.) population showed that targeted recombination doubles the predicted gains for yield and other agronomic traits. Our objective was to assess the predicted gains from targeted recombination for quantitative traits in multiple, elite maize populations. A total of 969 biparental maize populations were phenotyped at four to 12 environments in the United States from 2000 to 2008. Positions of one and two targeted recombinations per chromosome were determined from genomewide marker effects for 2911 single nucleotide polymorphism (SNP) loci. Relative efficiency ( $RE_{\text{targeted}}$ ) was calculated as the predicted response to targeted recombination divided by the predicted response to nontargeted recombination. On average, targeted recombination doubled the predicted genetic gains for yield, moisture, and test weight. For each trait,  $RE_{\text{targeted}}$  ranged from around 60 to 400% among the populations, and targeted recombination did not increase gains in around 4% of the populations. The  $RE_{\text{targeted}}$  tended to decrease as the similarity between the parents increased. Having targeted recombination on three chromosomes (for yield and test weight) to seven chromosomes (for moisture) led to the same or greater predicted gain than nontargeted recombination. Marker intervals for targeted recombination varied across populations and traits. Overall, our results for multiple, elite maize populations indicated that targeted recombination is a most promising breeding approach.

**Abbreviations:** CRISPR, clustered regularly interspaced short palindromic repeats; GCA, general combining ability;  $RE_{\text{targeted}}$ , relative efficiency; RR-BLUP, ridge regression–best linear unbiased prediction; SNP, single nucleotide polymorphism.

## CORE IDEAS

- Targeted recombination is the ability to have specific recombination points.
- In hundreds of maize populations, targeted recombination doubled the predicted gains.
- Targeted recombination is a potentially powerful breeding approach.

**M**AIZE breeding usually involves crossing two inbreds (A and B) to form a breeding population, developing selfed lines or doubled haploids from the A/B population, and evaluating the lines for their test-cross performance (Hallauer, 1990). Breeders therefore choose which crosses to make and which progeny to select. However, breeders have not attempted to control recombination among loci for quantitative traits and have simply relied on the results of random meiosis and fertilization during the development of breeding lines (Bernardo, 2017).

Targeted recombination is the ability to induce or select for specific recombination points so that genetic gains can be maximized. The marker intervals where targeted recombination should occur can be determined from genomewide marker effects (Bernardo, 2017; Ru and Bernardo, 2018). The procedure assumes that for a given trait, at least two quantitative trait loci (between

Citation: Brandariz, S.P., and R. Bernardo. 2019. Predicted genetic gains from targeted recombination in elite biparental maize populations. *Plant Genome* 12:180062. doi: 10.3835/plantgenome2018.08.0062

Received 30 Aug. 2018. Accepted 5 Dec. 2018. \*Corresponding author (bernardo@umn.edu).

This is an open access article distributed under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). Copyright © Crop Science Society of America 5585 Guilford Rd., Madison, WI 53711 USA

which targeted recombination is to occur) are on a given chromosome, and this assumption is likely met for a complex trait such as grain yield. Suppose a chromosome has three SNP markers and the effects of marker alleles (in order on the chromosome) carried by two parental inbreds are as follows: [0.2, 0.4, -0.5] in parent A, and [-0.2, -0.4, 0.5] in parent B. If recombination occurs between the second and third markers, a [0.2, 0.4, 0.5] homologue that maximizes the genetic gain can be recovered and converted into a doubled haploid. The same process can be done with many SNP loci on a chromosome and across multiple chromosomes.

An initial study with the intermated B73 × Mo17 maize population showed that in this classic maize population, predicted gains with one targeted recombination per chromosome ( $x = 1$ ) were twice the predicted gains with nontargeted recombination for yield and other agronomic traits (Bernardo, 2017). Predicted gains were higher with two targeted recombinations per chromosome ( $x = 2$ ). While this initial Bernardo (2017) study indicated that targeted recombination is a promising breeding approach, we do not know if the predicted gains will also be doubled in newer, elite maize germplasm. We also do not know the extent of variation in the usefulness of targeted recombination among multiple breeding populations. Lastly, information is lacking on factors that influence the usefulness of targeted recombination for quantitative traits.

Our main objective of this study was to assess the predicted gains from targeted recombination for yield, moisture, and test weight in multiple maize populations. Our specific goals were to (i) determine the extent of variation in predicted gains from targeted recombination across elite maize populations, (ii) identify factors that cause targeted recombination to be ineffective, and (iii) determine the number of chromosomes on which targeted recombination needs to be performed to achieve equal or greater predicted gains as nontargeted recombination.

## MATERIALS AND METHODS

### Phenotypic and Marker Data

The data have been previously described (Jacobson et al., 2014, 2015a,b; Lian et al., 2014; Brandariz and Bernardo, 2018) but are also briefly described here for the readers' convenience. Monsanto provided us with testcross phenotypic and SNP marker data for 969 biparental maize populations. The populations were evaluated for yield ( $\text{Mg ha}^{-1}$  at  $155 \text{ g H}_2\text{O kg}^{-1}$ ), moisture ( $\text{g H}_2\text{O kg}^{-1}$ ), and test weight ( $\text{kg hL}^{-1}$ ) at four to 12 environments in the United States from 2000 to 2008 (Jacobson et al., 2014). Only the  $F_2$  populations with heritability ( $h^2$ ) significantly different from zero ( $P = 0.05$ ) were used in this study. The  $h^2$  had a mean and range (in parentheses) of 0.46 (0.18, 0.92) for yield, 0.67 (0.25, 0.91) for moisture, and 0.60 (0.20, 0.92) for test weight.

The parents of the populations were genotyped with 2911 SNP markers, whereas the progeny were genotyped with 25 to 123 markers. The genotypes at each locus were

coded as 1 if the line was homozygous for the SNP allele from parent A, -1 if the line was homozygous for the SNP allele from parent B, and 0 if the line was heterozygous. Marker loci that were monomorphic between the two parental inbreds or that had a minor allele frequency  $< 0.10$  were excluded within each population (Lian et al., 2014; Jacobson et al., 2015a). The SNP data for the progeny were then imputed from the parental SNP data, based on the conditional probability of a nonobserved marker genotype given the two flanking-marker genotypes (Jacobson et al., 2015a). Monsanto provided us with a consensus map for all populations. The linkage map comprised 1668 cM, and the chromosome sizes ranged from 103 cM for chromosome 6 to 245 cM for chromosome 1.

### Genomewide Marker Effects

Marker effects were obtained by ridge regression–best linear unbiased prediction (RR-BLUP) as implemented in the rrBLUP package (Endelman, 2011) in R (R Development Core Team, 2018). Two training population models were used for estimating the marker effects: the A/B model, and the general combining ability (GCA) model. In the A/B model, marker effects were estimated from the A/B population itself. For each trait, the performance of an individual was predicted from information on the remaining  $N - 1$  individuals as  $y_p = \mu + \mathbf{Xm}$ , where  $y_p$  was the predicted performance of the individual;  $\mu$  was the estimated overall mean from RR-BLUP analysis of the  $N - 1$  individuals used in the training population;  $\mathbf{X}$  was an  $1 \times N_M$  incidence matrix with elements of 1, -1, and 0; and  $\mathbf{m}$  was an  $N_M \times 1$  vector of marker effects estimated from the remaining  $N - 1$  individuals (Jacobson et al., 2014). The shrinkage factor for marker effects was a function of restricted maximum likelihood estimates, obtained in the rrBLUP package, of the residual variance and a common (across SNP loci) marker variance. The marker effects and predictive ability ( $r_{MP}$ ), the latter defined as the correlation between marker-predicted values and phenotypic values, were estimated by deleting one individual at a time as described above and with cross-validation across environments. Cross-validation across environments was conducted by iterating through all possible partitions of the environments into a training set and validation set (Jacobson et al., 2014).

In the GCA model, the training population was obtained by pooling all prior A/\* populations (\* being a parent from the same heterotic group as A and B) and all prior \*/B populations to predict the performance of progeny in the A/B cross. For each trait, marker effects were estimated separately within each A/\* and \*/B population (Jacobson et al., 2014). Such within-population analysis accounted for differences in  $\mu$  among populations, which in turn were due to population structure as well as differences in the sets of environments used to evaluate each A/\* and \*/B population. Marker effects were averaged across the A/\* and \*/B populations (Jacobson et al., 2014; 2015a). The performance of all  $N$  individuals in the A/B

population was then predicted as  $\mathbf{y} = \mu\mathbf{1} + \mathbf{X}\mathbf{m}$ , where  $\mathbf{y}$  was an  $N \times 1$  vector of predicted performance;  $\mathbf{1}$  was an  $N \times 1$  vector with elements equal to 1;  $\mathbf{X}$  was an  $N \times N_M$  incidence matrix with elements of 1, -1, and 0; and  $\mathbf{m}$  was an  $N_M \times 1$  vector of marker effects averaged across the A/\* and \*/B populations (Jacobson et al., 2014).

For the GCA model, 27  $F_2$  populations were selected as the A/B populations on the basis of having at least four A/\* and \*/B crosses, a minimum population size of 50 lines, and an entry-mean heritability ( $h^2$ ) significantly greater than zero for each trait as described by Jacobson et al. (2014, 2015a). Both the A/B model and GCA model were used for these 27 A/B populations, and the model that led to the higher  $r_{MP}$  was selected for each trait and A/B population. For the rest of the populations, only the A/B model was used. Once predictions were obtained, populations were filtered according to having (i)  $r_{MP} > 0.40$  or (ii) a correlation between marker-predicted values and true genotypic values ( $r_{MG}$ )  $> 0.65$ . The  $r_{MG}$  values were estimated as  $r_{MP}/h$  (DeKkers, 2007; Lee et al., 2008).

The number of  $F_2$  populations with  $h^2 > 0$  was 706 for yield, 707 for moisture, and 698 for test weight. The number of  $F_2$  populations with  $r_{MP} > 0.40$  was 30 for yield, 314 for moisture, and 187 for test weight. The number of  $F_2$  populations with  $r_{MG} > 0.65$  was six for yield, 51 for moisture, and 17 for test weight.

### Predicted Gains from Targeted Recombination

Marker effects, calculated with either the A/B model or GCA model as described above, were used to determine the marker intervals where one and two targeted recombinations should occur on each chromosome (Bernardo, 2017). First, one targeted recombination was considered at each marker interval on each chromosome. The performance of two doubled haploids, induced from each of the two resulting homologues, was calculated as the sum of the effects of the alleles carried by each homologue. The homologue with superior performance (higher yield and test weight and lower moisture) was identified. The procedure was repeated for all 10 chromosomes. The predicted performance of the doubled haploid resulting from targeted recombination was calculated as the sum of  $\mu$  from RR-BLUP analysis plus the sum of the gains from targeted recombination within each chromosome. Second, two targeted recombinations were considered between all pairs of marker intervals within each chromosome assuming independent recombinations on each chromosome. Subsequent procedures were the same as those with one targeted recombination per chromosome.

For each  $F_2$  population, the response to selection with targeted recombination ( $R_{Targeted}$ ) was calculated as the predicted performance of the best doubled haploid from targeted recombination minus the estimate of  $\mu$  from the RR-BLUP analysis. The response with nontargeted recombination ( $R_{Nontargeted}$ ) was estimated as the marker-predicted performance of the best observed line minus the estimate of  $\mu$  from RR-BLUP analysis. Because

of the shrinkage of RR-BLUP marker effects toward the mean, marker-predicted values for each trait are closer to the mean in comparison with the corresponding phenotypic values. The use of marker-predicted values instead of phenotypic values to calculate  $R_{Nontargeted}$  therefore accounted for the difference in scale because of shrinkage in RR-BLUP (Bernardo, 2017). The relative efficiency of selection with targeted recombination vs. nontargeted recombination was calculated for each population as  $RE_{Targeted} = (R_{Targeted}/R_{Nontargeted}) \times 100$  (Bernardo, 2017). Confidence intervals ( $P = 0.05$ ) for the difference in the genetic gain with targeted recombination vs. nontargeted recombination were conducted by obtaining the difference in the genetic gain for each chromosome and performing 1000 bootstrap samples within each chromosome (Ru and Bernardo, 2018; Bernardo, 2017).

### Factors Associated with the Relative Efficiency of Targeted Recombination

The correlation was calculated between  $RE_{Targeted}$  and  $r_{MP}$ ,  $r_{MG}$ ,  $h^2$ , and the marker similarity ( $S_{AB}$ ) between the parents of the biparental cross. The  $S_{AB}$  between parents was estimated as the simple matching coefficient across the SNP loci (Sokal and Michener, 1958; Jacobson et al., 2015b; Brandariz and Bernardo, 2018). First, we calculated the within-locus simple matching coefficients by considering the possible combinations of marker genotypes (MM, Mm, and mm) between the parents. The simple matching coefficient was 1 between MM and MM or between mm and mm, 0 between MM and mm, and 0.50 between Mm and any other genotype (MM, Mm, or mm). Second, we calculated the mean of the within-locus simple matching coefficients across the SNP loci.

## RESULTS

### Mean and Variability of the Relative Efficiency of Targeted versus Nontargeted Recombination

On average, having one ( $x = 1$ ) or two ( $x = 2$ ) targeted recombinations per chromosome doubled the predicted genetic gain for yield, moisture, and test weight. For yield, the mean  $RE_{Targeted}$  was 217% with  $x = 1$  and 233% with  $x = 2$  for populations with  $h^2 > 0$ ; 230% with  $x = 1$  and 258% with  $x = 2$  for populations with  $r_{MP} > 0.40$ ; and 243% with  $x = 1$  and 292% with  $x = 2$  for populations with  $r_{MG} > 0.65$  (Fig. 1). The mean  $RE_{Targeted}$  for moisture and test weight was also  $\sim 200\%$  for all subsets of populations meeting the  $h^2$ ,  $r_{MP}$ , and  $r_{MG}$  criteria (Fig. 1).

While the mean  $RE_{Targeted}$  was around 200%, the individual  $RE_{Targeted}$  values differed among the populations. For the populations with  $h^2 > 0$ ,  $RE_{Targeted}$  for yield ranged from 60 to 454% with  $x = 1$  and from 72 to 452% with  $x = 2$ . However, for populations with  $r_{MP} > 0.40$  or with  $r_{MG} > 0.65$ ,  $RE_{Targeted}$  for yield was always  $> 100\%$  (Fig. 1). For moisture,  $RE_{Targeted}$  ranged from 63 to 422% with  $x = 1$ , and from 80 to 449% with  $x = 2$  for populations with  $h^2 > 0$ . Similar variation in  $RE_{Targeted}$  was found among populations with  $r_{MP} > 0.40$  or with  $r_{MG} > 0.65$

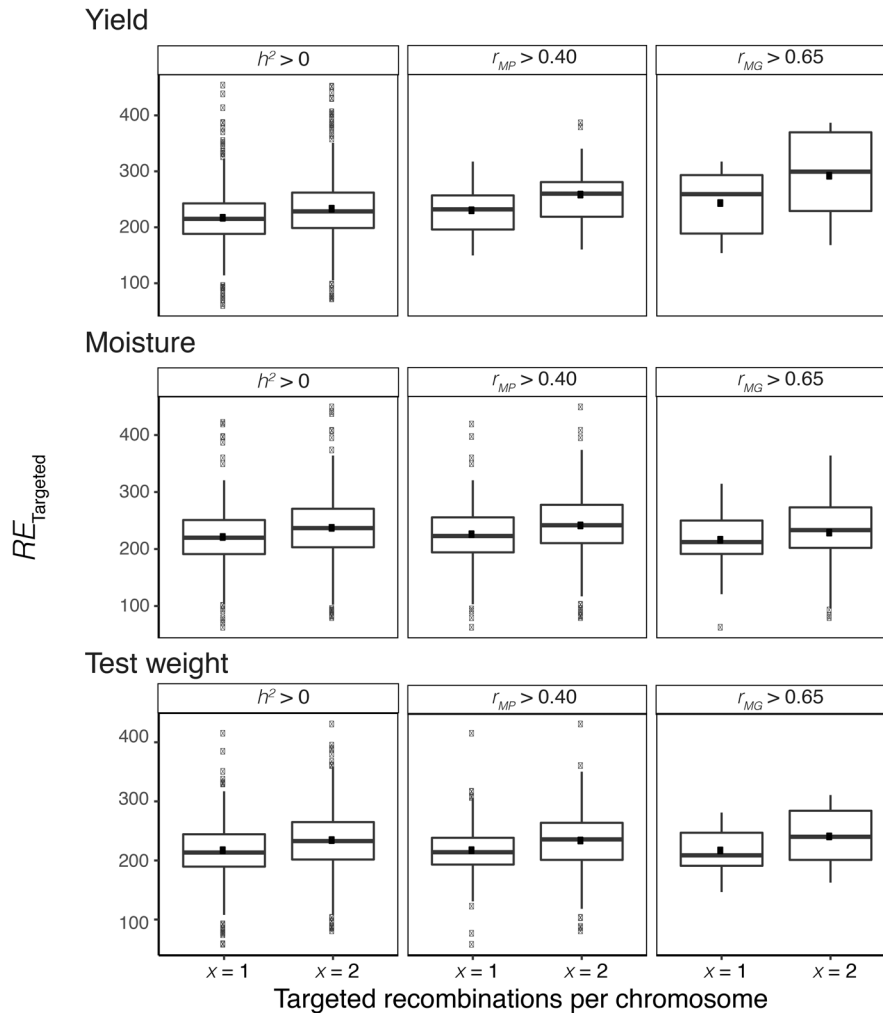


Fig. 1. Box-plot of relative efficiency ( $RE_{\text{Targeted}}$ , %) of selection with one ( $x = 1$ ) and two ( $x = 2$ ) targeted recombinations per chromosome vs. nontargeted recombination, for maize  $F_2$  populations with: heritability on an entry-mean basis ( $h^2 > 0$ ); correlation between marker-predicted values and phenotypic values ( $r_{MP} > 0.40$ ); and correlation between marker-predicted values and true genotypic values ( $r_{MG} > 0.65$ ).

(Fig. 1). The  $RE_{\text{Targeted}}$  for test weight ranged from 60 to 415% with  $x = 1$  and from 81 to 431% with  $x = 2$  with  $h^2 > 0$ . Some populations with  $r_{MP} > 0.40$  had  $RE_{\text{Targeted}} < 100\%$ , but none of the populations with  $r_{MG} > 0.65$  had  $RE_{\text{Targeted}} < 100\%$  (Fig. 1).

The variation in  $RE_{\text{Targeted}}$  (i.e., ratio between predicted gains) was accompanied by variation in the predicted gain itself (i.e.,  $R_{\text{Targeted}}$ ), which was the numerator of  $RE_{\text{Targeted}}$ . For the populations with  $h^2 > 0$ ,  $R_{\text{Targeted}}$  for yield ranged from  $<0.1$  to  $3.0 \text{ Mg ha}^{-1}$  with  $x = 1$  and from  $<0.1$  to  $3.7 \text{ Mg ha}^{-1}$  with  $x = 2$ . These  $R_{\text{Targeted}}$  values resulted in predicted yields ranging from  $9$  to  $17 \text{ Mg ha}^{-1}$  with  $x = 1$  and  $x = 2$  (Fig. 2). Approximately 97% of predicted yields were lower than  $15.7 \text{ Mg ha}^{-1}$  ( $250 \text{ bushels acre}^{-1}$ ) with either  $x = 1$  and  $x = 2$  with  $h^2 > 0$  (Fig. 2).

The  $R_{\text{Targeted}}$  for moisture ranged from less than  $-0.1$  to  $-64 \text{ g kg}^{-1}$  with  $x = 1$  and from less than  $-0.1$  to  $-69.3 \text{ g kg}^{-1}$  with  $x = 2$ , resulting in predicted moisture values of  $123$  to  $313 \text{ g kg}^{-1}$  with  $x = 1$  and  $x = 2$  (Fig. 2). For test weight, the  $R_{\text{Targeted}}$  ranged from  $<0.1$  to  $5 \text{ kg hL}^{-1}$

with  $x = 1$  and from  $<0.1$  to  $5.7 \text{ kg hL}^{-1}$  with  $x = 2$ , resulting in predicted test weights of  $68$  to  $82 \text{ kg hL}^{-1}$  with  $x = 1$  and  $x = 2$  (Fig. 2).

#### Frequency of Populations in which Targeted Recombination is Likely Ineffective

Targeted recombination was predicted to be ineffective when (i)  $RE_{\text{Targeted}}$  was  $<100\%$  or (ii)  $RE_{\text{Targeted}}$  was numerically  $>100\%$  but was not statistically different ( $P = 0.05$ ) from  $100\%$ . The numbers of populations with  $RE_{\text{Targeted}} < 100\%$  were as follows: 13 for yield (1.8%) with  $x = 1$  and eight (1.1%) with  $x = 2$  (five in common); six for moisture (0.9%) with  $x = 1$  and eight (1.1%) with  $x = 2$  (two in common); and 10 for test weight (1.4%) with  $x = 1$  and seven (1.0%) with  $x = 2$  (three in common). In these populations, the best line had more than two nontargeted recombinations on some chromosomes (results not shown). The numbers of populations with  $RE_{\text{Targeted}}$  exceeding  $100\%$ , but not significantly greater than  $100\%$ , were as follows: 20 for yield (2.8%) with  $x = 1$  and 19 (2.7%) with  $x = 2$  (14 in common); 15 for

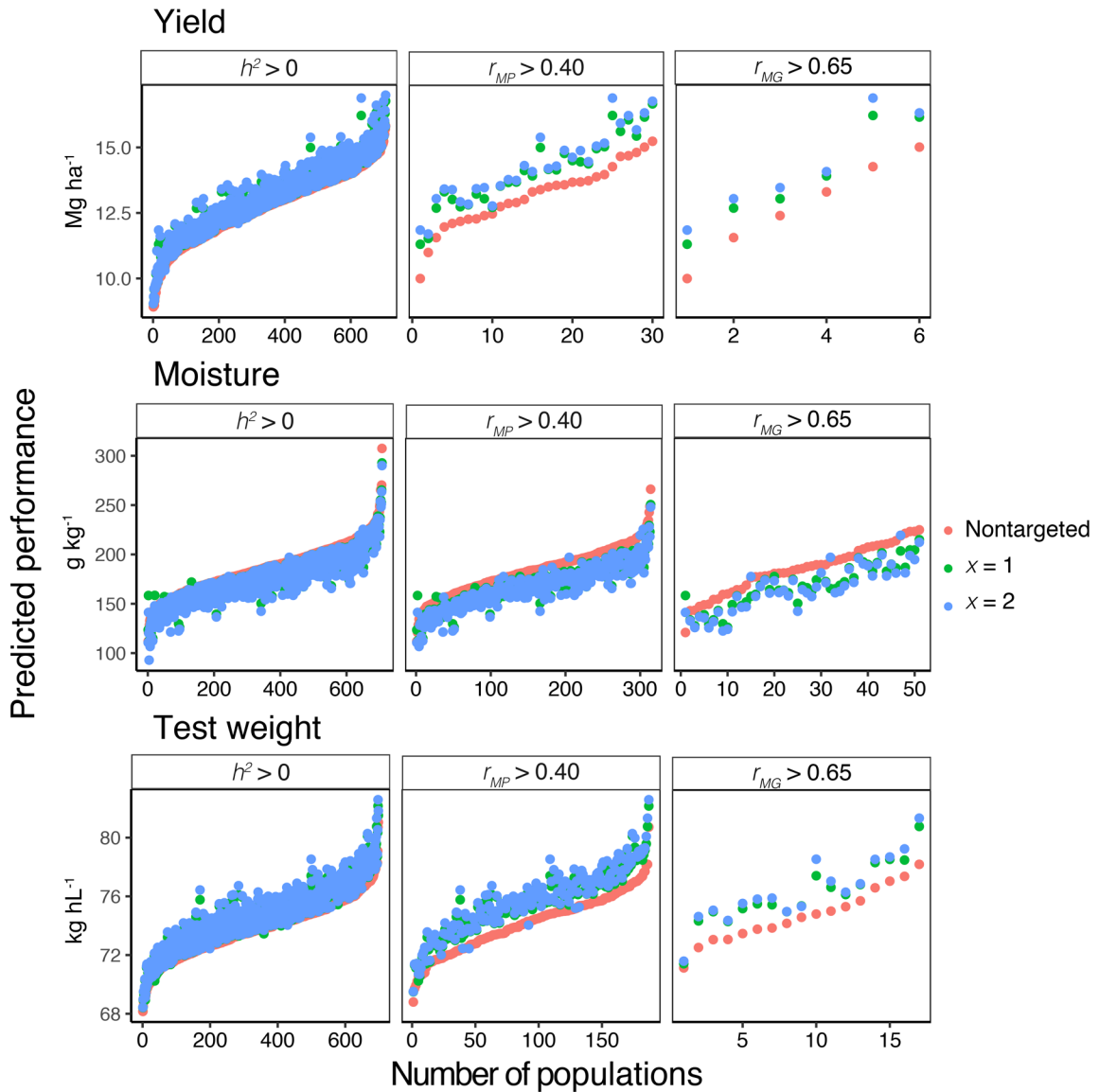


Fig. 2. Predicted performance with nontargeted recombination and one ( $x = 1$ ) and two ( $x = 2$ ) targeted recombinations per chromosome, for maize  $F_2$  populations with: heritability on an entry-mean basis ( $h^2 > 0$ ); correlation between marker-predicted values and phenotypic values ( $r_{MP} > 0.40$ ); and correlation between marker-predicted values and true genotypic values ( $r_{MG} > 0.65$ ).

moisture (2.3%) with  $x = 1$  and 21 (3.0%) with  $x = 2$  (11 in common); and 17 for test weight (2.4%) with  $x = 1$  and 19 (2.7%) with  $x = 2$  (11 in common). Adding the frequencies of  $RE_{\text{Targeted}} < 100\%$  and  $RE_{\text{Targeted}}$  not significantly greater than 100% led to an overall frequency of  $\sim 4\%$  of populations for which targeted recombination for a given trait was ineffective.

The correlations were low ( $-0.04$  to  $0.14$ ) and mostly nonsignificant between  $RE_{\text{Targeted}}$  and values of  $h^2$ ,  $r_{MP}$ , and  $r_{MG}$  for all traits. However, the correlation between  $RE_{\text{Targeted}}$  and the marker similarity ( $S_{AB}$ ) between the parents of each cross was significant for all traits. The correlations between  $RE_{\text{Targeted}}$  and  $S_{AB}$  were as follows:  $-0.23$  for yield with both  $x = 1$  and  $x = 2$ ;  $-0.25$  for moisture with  $x = 1$  and  $-0.27$  with  $x = 2$ ; and  $-0.24$  for test weight with  $x = 1$  and  $-0.25$  with  $x = 2$ .

### Chromosome Contributions and Targeted-Recombination Positions

In general, chromosomal contributions to the total predicted gain (averaged across populations with  $h^2 > 0$ ) were proportional to the sizes (in cM) of the chromosomes (Fig. 3). However, chromosomal contributions were slightly larger than expected for chromosomes 1, 5, and 2 and were slightly smaller than expected for chromosomes 8 and 9. The mean and range (in parentheses) of the minimum number of chromosomes needed for targeted recombination to reach equal or greater predicted gains compared with nontargeted recombination was 3 (1, 8) with  $x = 1$  and 3 (1, 9) with  $x = 2$  for yield; 7 (1, 9) with  $x = 1$  and 7 (3, 9) with  $x = 2$  for moisture; and 3 (1, 9) with both  $x = 1$  and  $x = 2$  for test weight.

The most frequent positions where one and two targeted recombinations should occur varied across



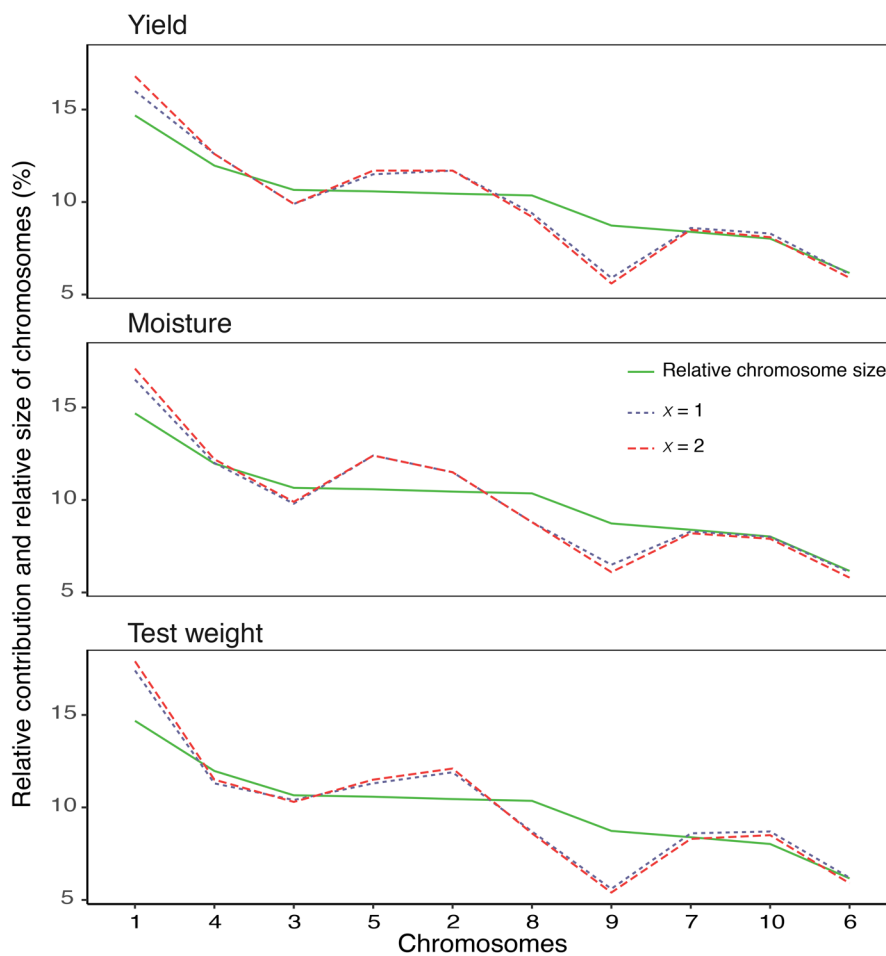


Fig. 3. Mean percentage (%) of (i) chromosome contribution relative to the total predicted gain with one ( $x = 1$ ) and two ( $x = 2$ ) targeted recombinations per chromosome and (ii) size of each chromosome relative to the total linkage map across maize populations with heritability on an entry-mean basis ( $h^2 > 0$ ).

populations. For yield, the most frequent position with  $x = 1$  on chromosome 1 had an incidence of 14.6% (Table 1). In other words,  $\sim 15\%$  of the populations had the same targeted-recombination position for  $x = 1$  on chromosome 1. Chromosomes 6 and 9 had the highest incidence of the most frequent positions for a given trait (Table 1). For these two chromosomes,  $\sim 40\%$  of the populations had the same targeted-recombination

position for  $x = 1$  and  $\sim 30\%$  of the populations shared a position for  $x = 2$ .

Targeted-recombination positions also varied across traits. For  $x = 1$  on chromosome 1, only 3% of the populations shared the same targeted-recombination position for all three traits, 24% of the populations shared the same position for two traits, and 73% had unique positions for each trait (Table 2). More than 50% of the populations did not share targeted-recombination positions across traits except for chromosomes 6 and 9 with  $x = 1$  (Table 2).

Table 1. Incidence (%) of the most frequent positions for one ( $x = 1$ ) and two ( $x = 2$ ) targeted recombinations per chromosome across maize  $F_2$  populations with heritability on an entry-mean basis  $> 0$ .

Chromosomes	Yield		Moisture		Test weight	
	$x = 1$	$x = 2$	$x = 1$	$x = 2$	$x = 1$	$x = 2$
1	14.6	4.2, 2.8	15.4	3.5, 4.2	14.3	3.9, 3.6
2	24.6	10.6, 10.1	19.7	8.6, 9.1	24.5	10.2, 10.1
3	25.5	14.0, 13.8	29.4	12.3, 12.1	27.2	14.3, 14.3
4	21.6	7.5, 8.0	24.2	9.2, 9.1	20.2	6.4, 6.5
5	22.8	8.9, 9.1	22.3	9.2, 9.2	21.8	8.6, 8.8
6	44.8	34.6, 34.7	42.3	32.2, 32.3	39.4	31.2, 31.1
7	27.8	16.3, 16.3	26.6	14.3, 14.2	26.7	15.0, 15.1
8	29.2	16.4, 16.5	27.3	14.1, 14.1	25.8	13.3, 13.3
9	41.6	33.2, 33.1	43.7	33.6, 34.1	41.4	33.1, 33.1
10	30.2	15.3, 15.2	25.6	10.6, 10.8	27.4	13.9, 14.1

## DISCUSSION

### Targeted Recombination as a Promising Breeding Approach

Our results showed that targeted recombination is predicted to substantially improve genetic gains for yield, moisture, and test weight in elite maize breeding populations. On average, targeted recombination doubled the predicted gains for all traits. The  $RE_{\text{Targeted}}$  values in this study were similar to those found for yield in the B73  $\times$  Mo17 maize population (212% with  $x = 1$  and 254% with  $x = 2$ ; Bernardo, 2017). The results from this study with hundreds of elite maize populations therefore provided

Table 2. Percentage of maize  $F_2$  populations (with heritability on an entry-mean basis  $> 0$ ) for which targeted-recombination positions were common for different numbers of traits.

No. of targeted recombinations per chromosome ( $x$ )	Traits with same targeted-recombination position	Chromosome									
		1	2	3	4	5	6	7	8	9	10
1	3	3	4	8	3	5	19	6	8	19	7
	2	24	33	36	30	33	42	38	31	46	37
	0	73	63	56	66	62	39	55	61	35	55
2	3	3, 3	5, 3	7, 6	2, 2	5, 5	17, 18	5, 7	7, 7	19, 18	6, 7
	2	26, 27	33, 32	33, 36	25, 31	34, 30	38, 39	35, 31	32, 34	41, 42	38, 38
	0	71, 70	62, 64	59, 58	73, 67	62, 65	45, 43	60, 62	61, 59	39, 40	56, 54

strong evidence of the potential value of targeted recombination. As discussed in the next section, this assumes that inducing artificial targeted recombination or pyramiding natural recombinations that already occur at the desired marker intervals becomes feasible.

While the predicted gains were doubled on average, the  $RE_{\text{Targeted}}$  for each trait ranged from around 60 to 400% among the populations, showing that predicted gains from targeted recombination may be population dependent. More importantly, targeted recombination was ineffective in only  $\sim 4\%$  of the populations per trait. This low frequency suggests that targeted recombination will usually be superior to nontargeted recombination. The infrequent values of  $RE_{\text{Targeted}} < 100\%$  were due to having more than two nontargeted recombinations on some chromosomes. A larger number of targeted recombinations ( $x > 2$  per chromosome) is expected to increase the gains from targeted recombination, but achieving more than one to two targeted recombinations per chromosome will likely be difficult. Furthermore, the gains with nontargeted recombination can be increased by growing a larger population and selecting the best individual. But because the standardized selection differential is not a linear function of the proportion of individuals selected, more-stringent selection is unable to double the gains with nontargeted recombination when the initial population is already large (Bernardo, 2017). However, for the 4% of populations for which targeted recombination is ineffective, selecting the best individual out of a larger population could be a feasible approach.

The  $RE_{\text{Targeted}}$  values tended to decrease as the marker similarity between the parents of the biparental cross increased ( $r \approx -0.25$ ). However, the effectiveness of targeted recombination was not associated with variation in  $h^2$ ,  $r_{\text{MP}}$ , and  $r_{\text{MG}}$  (assuming  $h^2 > 0$ ). That being said, breeders need to consider both the  $RE_{\text{Targeted}}$  values and the response to targeted recombination, as a high  $RE_{\text{Targeted}}$  value is not meaningful if the gains ( $R_{\text{Targeted}}$ ) for yield in megagrams per hectare or moisture in grams per kilogram are small. Higher values of  $h^2$ ,  $r_{\text{MP}}$ , and  $r_{\text{MG}}$  are known to increase the response to genomewide selection and higher values of these three parameters are therefore desirable.

Our results showed that targeted recombination does not need to occur on all chromosomes to achieve equal or greater gains compared with nontargeted recombination. This result was consistent with the results for the B73  $\times$  Mo17 maize population (Bernardo, 2017). Individual

chromosomes accounted for different proportions to the total predicted gains, and having targeted recombination on about three (for yield and test weight) to seven (for moisture) chromosomes was predicted to achieve equal or greater gains than nontargeted recombination.

The  $R_{\text{Targeted}}$  values were for a doubled haploid, whereas the  $R_{\text{Nontargeted}}$  values were for the best  $F_3$  line. We did not expect this difference in inbreeding level to affect our results because, from theory, the expected testcross genotypic value of a homozygous line is equal to the testcross genotypic value of the  $F_3$  (or other selfing generation) line from which the homozygous line was derived (Bernardo, 1991). These theoretical results are supported by maize empirical data that showed that the mean testcross performance (across five testers) of the best  $S_1$  line out of 50 (7.92 Mg ha $^{-1}$ ) was very close to the testcross performance of the best  $S_8$  line out of 50 (8.10 Mg ha $^{-1}$ ) (Lopez-Perez, 1979). Therefore, we expect a similar  $R_{\text{Nontargeted}}$  if we could have estimated it for a doubled haploid.

### Designing a Maize Breeding Program that Incorporates Targeted Recombination

The variation in  $RE_{\text{Targeted}}$ , per-chromosome contributions, and targeted-recombination positions indicated that prioritization of breeding populations, chromosomes, and traits will be important in targeted recombination. The breeding populations that are prime candidates for targeted recombination will obviously be those that combine a high  $RE_{\text{Targeted}}$  and  $R_{\text{Targeted}}$ . Crosses between closely related inbreds, which have a high marker similarity, will likely be excluded. Given that the position of a targeted recombination affects the performance for multiple traits, a selection index can be used. One or two targeted-recombination positions per chromosome can then be identified to maximize the selection index value. Targeted recombination on all chromosomes might be infeasible. If so, breeders will need to determine how much gain can be achieved if targeted recombination is done on a certain subset of chromosomes and selection on the remaining chromosomes is for the best natural recombinants (Bernardo, 2017).

The training population for obtaining the marker effects also needs to be chosen. In the A/B model (Jacobson et al., 2014), the training population is genetically identical to the population in which targeted recombination will be done. Therefore, the A/B model is likely preferable as long as the number of lines in the A/B cross

is large enough to obtain high values of  $r_{MP}$ . However, a disadvantage of the A/B model is that progeny in the A/B cross need to be phenotyped to estimate the marker effects. An alternative is to use the GCA model, in which the use of prior A/\* and \*/B crosses as the training population circumvents the need to phenotype progeny in the A/B cross (Jacobson et al., 2014).

Targeted recombination can be achieved in two ways. The most feasible way involves screening and pyramiding recombination events with a procedure similar to marker-assisted backcrossing. In concept, this approach involves performing foreground selection for chromosomes that carry a targeted recombination and background selection across the rest of the chromosomes (Bernardo, 2017). Marker-assisted backcrossing of transgenes involves the introgression of double-recombination events because the transgenes are flanked on each side by a recombination event. Approaches for marker-assisted gene pyramiding (Servin et al., 2004; De Beukelaer et al., 2015) should therefore readily apply to pyramiding double targeted recombinations ( $x = 2$ ) on multiple chromosomes, although the chromosomal segment between two targeted recombinations is likely to be longer than the transgene. The approach would be slightly different for pyramiding single targeted recombinations ( $x = 1$ ) per chromosome.

A second way of achieving targeted recombination involves CRISPR (clustered regularly interspaced short palindromic repeats) technology (Cong et al., 2013; Ran et al., 2013; Hsu et al., 2014). A CRISPR system has been used to induce mitotic targeted recombination for fine mapping a gene for manganese sensitivity in yeast (Sadhu et al., 2016). In addition, CRISPR has been used for targeted recombination in cultivated tomato (*Solanum lycopersicum* L.) and wild tomato (*S. pimpinellifolium* L.) (Filler Hayut et al., 2017). For crop plants, a protocol that involves a multiplex CRISPR system for inducing homologous recombinations, screening cells with the targeted recombinations, and regenerating cells into plants is yet to be developed. Thus, the feasibility of CRISPR technology for routinely inducing targeted recombination in plants is still unknown.

The development of protocols for targeted recombination, either by pyramiding natural combinations or by the use of CRISPR, involves many logistical and technical details that will need to be resolved. Such details are well beyond the scope of this study. Nevertheless, we are currently conducting proof-of-concept studies that involve pyramiding natural recombinations to validate the predicted gains from targeted recombination in elite maize germplasm.

### Conflict of Interest

The authors declare that there is no conflict of interest.

### ACKNOWLEDGMENTS

Sofia P. Brandariz was supported by Ph.D. fellowship funded by Monsanto (now Bayer). We thank Drs. David Butruille, Mike Lohuis, and Sam Eathington for allowing us to use the data sets in this study, and Dr. Shengqiang Zhong for his valuable input and insights on our work with these data sets since 2011.

### REFERENCES

- Bernardo, R. 1991. Correlation between testcross performance of lines at early and late selfing generations. *Theor. Appl. Genet.* 82:17–21. doi:10.1007/BF00231272
- Bernardo, R. 2017. Prospective targeted recombination and genetic gains for quantitative traits in maize. *Plant Genome* 10. doi:10.3835/plantgenome2016.11.0118
- Brandariz, S.P., and R. Bernardo. 2018. Maintaining the accuracy of genome-wide predictions when selection has occurred in the training population. *Crop Sci.* 58:1226–1231. doi:10.2135/cropsci2017.11.0682
- Cong, L., F.A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P.D. Hsu, X. Wu, W. Jiang, L.A. Marraffini, and F. Zhang. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* 339:819–823. doi:10.1126/science.1231143
- De Beukelaer, H., G. De Meyer, and V. Fack. 2015. Heuristic exploitation of genetic structure in marker-assisted gene pyramiding problems. *BMC Genet.* 16:2. doi:10.1186/s12863-014-0154-z
- Dekkers, J.C.M. 2007. Prediction of response to marker-assisted and genomic selection using selection index theory. *J. Anim. Breed. Genet.* 124:331–341. doi:10.1111/j.1439-0388.2007.00701.x
- Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4:250–255. doi:10.3835/plantgenome2011.08.0024
- Filler Hayut, S., C.M. Bessudo, and A.A. Levy. 2017. Targeted recombination between homologous chromosomes for precise breeding in tomato. *Nat. Commun.* 8:15605. doi:10.1038/ncomms15605
- Hallauer, A.R. 1990. Methods used in developing maize inbreds. *Maydica* 35:1–16.
- Hsu, P.D., E.S. Lander, and F. Zhang. 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157:1262–1278. doi:10.1016/j.cell.2014.05.010
- Jacobson, A., L. Lian, S. Zhong, and R. Bernardo. 2014. General combining ability model for genome-wide selection in a biparental cross. *Crop Sci.* 54:895–905. doi:10.2135/cropsci2013.11.0774
- Jacobson, A., L. Lian, S. Zhong, and R. Bernardo. 2015a. Marker imputation before genome-wide selection in biparental maize populations. *Plant Genome* 8. doi:10.3835/plantgenome2014.10.0078
- Jacobson, A., L. Lian, S. Zhong, and R. Bernardo. 2015b. Minimal loss of genetic diversity after genome-wide selection within biparental maize populations. *Crop Sci.* 55:783–789. doi:10.2135/cropsci2014.10.0744
- Lee, S.H., J.H.J. van der Werf, B.J. Hayes, M.E. Goddard, and P.M. Visscher. 2008. Predicting unobserved phenotypes for complex traits from whole-genome SNP data. *PLoS Genet.* 4:e1000231. doi:10.1371/journal.pgen.1000231
- Lian, L., A. Jacobson, S. Zhong, and R. Bernardo. 2014. Genome-wide prediction accuracy within 969 maize biparental populations. *Crop Sci.* 54:1514–1522. doi:10.2135/cropsci2013.12.0856
- Lopez-Perez, E. 1979. Comparisons among five different testers for the evaluation of unselected lines of maize (*Zea mays* L.). Ph.D. dissertation, Iowa State University, Ames.
- Ran, F.A., P.D. Hsu, J. Wright, V. Agarwala, D.A. Scott, and F. Zhang. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8:2281–2308. doi:10.1038/nprot.2013.143
- R Development Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/> (accessed 18 July 2018).
- Ru, S., and R. Bernardo. 2018. Targeted recombination to increase genetic gain in self-pollinated crops. *Plant and Animal Genome Conf. XXVI*, San Diego, CA. 13–17 Jan. 2018.
- Sadhu, M.J., J.S. Bloom, L. Day, and L. Kruglyak. 2016. CRISPR-directed mitotic recombination enables genetic mapping without crosses. *Science* 352:1113–1116. doi:10.1126/science.aaf5124
- Servin, B., O.C. Martin, M. Mézard, F. Hospital. 2004. Toward a theory of marker-assisted gene pyramiding. *Genetics* 168:513–523. doi:10.1534/genetics.103.023358.
- Sokal, R.R., and C.D. Michener. 1958. A statistical method for evaluating systematic relationships. *Univ. Kans. Sci. Bull.* 38:1409–1438.