

# The genomic impacts of drift and selection for hybrid performance in maize

Justin P. Gerke<sup>\*1</sup>, Jode W. Edwards<sup>†</sup>, Katherine E. Guill<sup>‡</sup>, Jeffrey Ross-Ibarra<sup>§2</sup>, and Michael D. McMullen<sup>\*‡</sup>

<sup>\*</sup>Division of Plant Sciences, University of Missouri, Columbia MO 65211

<sup>†</sup>Corn Insects and Crop Genetics Research Unit, USDA-Agricultural Research Service, Ames, IA, 50011

<sup>‡</sup>Plant Genetics Research Unit, USDA-Agricultural Research Service, Columbia MO 65211

<sup>§</sup>Department of Plant Sciences, Center for Population Biology, and Genome Center, University of California, Davis, CA 95616

Revised manuscript for *Genetics*, September 15, 2015

**ABSTRACT** Although maize is naturally an outcrossing organism, modern breeding utilizes highly inbred lines in controlled crosses to produce hybrids. The United States Department of Agriculture reciprocal recurrent selection experiment between the Iowa Stiff Stalk Synthetic (BSSS) and the Iowa Corn Borer Synthetic No. 1 (BSCB1) populations represents one of the longest running experiments to understand the response to selection for hybrid performance. To investigate the genomic impact of this selection program, we genotyped the progenitor lines and over 600 individuals across multiple cycles of selection using a genome-wide panel of ~ 40,000 SNPs. We confirmed previous results showing a steady temporal decrease in genetic diversity within populations and a corresponding increase in differentiation between populations. Thanks to detailed historical information on experimental design, we were able to perform extensive simulations using founder haplotypes to replicate the experiment in the absence of selection. These simulations demonstrate that while most of the observed reduction in genetic diversity can be attributed to genetic drift, heterozygosity in each population has fallen more than expected. We then took advantage of our high-density genotype data to identify extensive regions of haplotype fixation and trace haplotype ancestry to single founder inbred lines. The vast majority of regions showing such evidence of selection differ between the two populations, providing evidence for the dominance model of heterosis. We discuss how this pattern is likely to occur during selection for hybrid performance, and how it poses challenges for dissecting the impacts of modern breeding and selection on the maize genome.

## Introduction

Hybrid maize, first developed in the early 20th century (Crow 1998), rapidly and completely replaced mass-selected open pollinated varieties in the United States (Crabb and Hughes 1947). The shift towards development of inbred lines based on their ability to generate good hybrids – referred to as combining ability – constituted an abrupt change from the open-pollinated mass selection that breeders practiced for millennia (Anderson 1944; Troyer 1999). Maize inbred lines are now partitioned into separate heterotic groups that maximize performance and hybrid vigor (heterosis) for yield when inbreds from different heterotic groups are crossed with each other (Tracy and Chandler 2006).

While the founders of these heterotic groups were not initially differentiated, multiple studies with molecular markers

have indicated that these heterotic groups have diverged genetically over time to become highly structured and isolated populations, resulting in a dramatic restructuring of population genetic variation (Duvick *et al.* 2004; Ho *et al.* 2005; Feng *et al.* 2006). Advances in high throughput genotyping and the development of a maize reference genome now enable the observation of maize population structure at high marker density across the whole genome (Ganal *et al.* 2011; Chia *et al.* 2012). These studies have examined a broad spectrum of germplasm at various points in the history of maize to search for the signals of population structure and artificial selection (Hufford *et al.* 2012; van Heerwaarden *et al.* 2012; Jiao *et al.* 2012). Although selective sweeps from the initial domestication of maize are clearly visible, localized genomic signals of selection during modern breeding are difficult to observe (Hufford *et al.* 2012; van Heerwaarden *et al.* 2012) despite steady, heritable improvement in phenotype (Duvick 2005). The lack of distinct selection signals in broad germplasm collections may be due to population-specific selection within germplasm sub-

<sup>1</sup>Corresponding author. Current Address: DuPont Pioneer, Johnston, Iowa 50131. E-mail: justin.gerke@gmail.com

<sup>2</sup>Corresponding author. Department of Plant Sciences, University of California, Davis, California 95616. E-mail: rossibarra@ucdavis.edu

groups; in this case selection should be easier to detect in populations maintained in an individual program. This possibility is supported by the measured success in identifying targets of selection in individual experimental populations under directional selection for specific phenotypes such as seed size and ear number (Hirsch *et al.* 2014; Beissinger *et al.* 2014).

In this study, we examine a different experiment in which the method and target of selection — reciprocal recurrent selection for hybrid yield — closely mirror those used in the generation of modern maize hybrids (Comstock *et al.* 1949; Duvick *et al.* 2004). Reciprocal recurrent selection is a method, initially proposed by Comstock *et al.* (1949), in which lines from two populations are evaluated based on the phenotype of the hybrids each line produces when crossed with lines from the opposing population. The selected lines are then intermated within each population to generate lines for the next cycle of recurrent selection. This procedure results in two closed and genetically isolated populations that simultaneously evolve improved combining ability with one another. The USDA-ARS at Ames, Iowa has conducted a reciprocal recurrent selection program (hereafter referred to as the Iowa RRS) with the Iowa Stiff Stalk Synthetic (BSSS) and Iowa Corn Borer Synthetic No. 1 (BSCB1) populations for 18 cycles of selection for hybrid yield (Penny and Eberhart 1971; Edwards 2011). This program represents one of the best-documented public experiments on selection for combining ability and hybrid performance. Extensive records on population sizes, breeding methods, selection differentials, and seed are available for all cycles of selection. This set of resources makes the Iowa RRS an ideal test case for the study of the genomic impact of hybrid breeding in maize. The Iowa RRS experiment provides additional relevance because lines derived from the BSSS population have had a major impact on the development of commercial hybrids (Duvick *et al.* 2004; Darrah and Zuber 1986), the formation of modern heterotic groups (Troyer 1999; Senior *et al.* 1998), and the choice of a maize reference genome (Schnable *et al.* 2009).

## Methods

### **The BSSS and BSCB1 recurrent selection program**

The Iowa RRS program was initiated in 1949 (Penny and Eberhart 1971). The BSSS population was formed from 16 inbred lines selected for stalk strength in 1933 and 1934 (Sprague, 1946) and the BSCB1 population was formed from a set of 12 inbred lines (Hallauer *et al.* 1974). The founder inbreds (Table S1) were randomly mated to create each ‘cycle 0’ base population. Testcross progenies were formed by self-pollinating 100 individuals in each population and simultaneously crossing the same 100 individuals onto 10 plants (as females) in the reciprocal population (e.g., a BSSS plant was self-pollinated and crossed to 10 females in BSCB1). Seed from the 10 females pollinated by a single male plant was bulked, and the 200 testcross progenies were grown in replicated multi-

environment yield trials. Ten testcross families were selected from among the BSSS males (crossed to BSCB1 as a tester) and BSCB1 males (crossed to BSSS as a tester) based primarily on grain yield. Self-pollinated seed from the 10 selected males in each population were planted in the following season and plant-to-plant crosses made between each of the 45 possible pairs; seed from these crosses formed the cycle 1 populations.

Additional changes to this procedure occurred in later cycles (Penny and Eberhart 1971; Keeratinijakal and Lamkey 1993). In cycles six through eight, individual S1 plants (progeny of one generation of self-pollination) were used as males instead of noninbred plants. Beginning in cycle eight, the number of testcross progenies selected to form the next cycle of selection was doubled to 20. Beginning in cycle 10, the method of producing testcross progenies for evaluation was changed to generate a single set of reciprocal full sib families between the populations rather than two sets of half-sib families. This procedure has been continued to the present, with reciprocal hybrid testcross progeny derived from the cycle 18 population in 2014 and evaluated in 2015 for formation of cycle 19. The phenotypes evaluated for selection have been consistent across all 19 cycles of selection: grain yield (dry matter yield of maize grain per acre) has been the most important, but selection was also applied for reduced moisture content at harvest and reduced root and stalk lodging. Plant densities have increased consistently across cycles of selection as the populations have become consistently better adapted to high plant density with continued selection (Brekke *et al.* 2011a).

### **Plants and inbred lines used**

The plants and inbred lines used in this experiment are listed in Tables S1 and S2. We genotyped 34–36 plants from each of the BSSS and BSCB1 populations at selection cycles 0, 4, 8, 12, and 16 (Table S3). These plants represent descendants of the original populations which have been randomly mated to maintain seed. We also genotyped the founder inbreds for each population, with the exception of F1B1, CI.617, WD456, and K230 for which seed was not available. The data for founder line CI.540 was not used because the genotyped material was heterozygous. A number of derived lines were also genotyped for calibrating phasing and imputation procedures (see below).

### **Genotype data**

Plants from the cycles of selection, founders, and derived lines were grown in a greenhouse and tissue was collected at the 3 leaf stage. Tissue was lyophilized, ground, and DNA extracted by a CTAB procedure (Saghai-Marof *et al.* 1984). Samples were genotyped using the 24-sample Illumina MaizeSNP50 array (Ganal *et al.* 2011) according to the Illumina Infinium protocol, and imaged on an Illumina BeadStation at the University of Missouri DNA core facility. Genotypes were determined with the GenomeStudio v2010.2 software using the manufac-

turer's MaizeSNP50\_B.egt cluster file. The design of the maize SNP50 Chip included a relatively small ascertainment panel of inbred lines, introducing a bias in the frequencies of SNPs included on the chip (Ganal *et al.* 2011). However, because our simulations are based not on theoretical expectations but instead on sampling from the observed data at cycle 0, we expect ascertainment bias to have a minimal impact on our results.

We called 48,919 SNPs on the Illumina platform from the MaizeSNP50\_B.egt cluster file. Genotypes with quality scores of 50 or less were recoded as missing data. Three plants were removed from the data due to an excess of missing data (the derived line B10, a cycle 0 plant from BSSS, and a cycle 8 plant from BSSS). In addition, BSCB1 plant 31 from cycle 4 appeared switched with plant 31 from cycle 8 based on our principal component analysis (PCA), so we switched the labels for these two genotypes to correct the mistake. To avoid structure among the missing data, we removed any SNP that was coded as missing in more than 3 plants in either group of founders or any group of plants from a particular cycle and population. Preliminary analysis by PCA and heatmap plots of distance matrices revealed two additional likely mix-ups. Plant 23 from BSSS cycle 8 was a clear outlier from the BSSS population as a whole and plant 2 from BSSS cycle 0 is likely a mislabeled plant from cycle 16. Since there was no evidence suggesting when mislabelings occurred, each of these plants was removed from the analysis. The final genotyping data set contained 39,261 SNPs and is available at [http://figshare.com/articles/Gerke\\_et\\_al\\_Iowa\\_RRS/1515061](http://figshare.com/articles/Gerke_et_al_Iowa_RRS/1515061).

### **Integrating the genetic and physical map**

Our simulation approach requires co-linear genetic and physical maps. We therefore took steps to improve the positions of the SNP markers on the genetic and physical maps relative to version 5A.59 of the maize genome assembly. The probable physical position of each SNP was obtained by comparing SNP context sequences to the genome sequence. For this purpose, SNP context sequences were defined as the sequence 25 bp upstream of the SNP, the bp representing the SNP itself, followed by 25 bp downstream of the SNP, making a total sequence length of 51bp. When a single genomic location was queried by two separate probes on the array, we chose the probe with higher quality calls and dropped the other marker from the dataset. To assign a genetic position for each SNP, we used a map derived from the B73xMo17 (IBM) mapping population similar to the IBM framework map in Ganal *et al.* (2011). This genetic map contains 4,217 framework SNP markers, which provides a much higher density than the map used to order the 5A.59 release of the maize genome sequence. As a result, we identified several places in the genome where the physical positions were incorrect according to our genetic map. These cases included both simple reversals of the physical map relative to the genetic map, and also the assignment of blocks of markers to the wrong linkage group, which we refer to as mis-mapped

blocks. To maintain collinearity between the genetic and physical map, the physical positions of these SNPs were reassigned as follows. Individual mis-mapped markers small reversals and mis-mapped blocks (< 10 kb) were removed from the data. Small rearrangements of this sort are more likely to represent mis-mapped paralogous sequence than true errors in the physical map. When larger reversals were identified, we transposed the physical positions of the SNPs from one end of the segment to the other. Mis-mapped blocks were often larger than the physical gap into which they were moved. We therefore assigned the first SNP of the block to a position 10 kb downstream from the previous SNP on the correct linkage group. We then recalculated genomic coordinates for the rest of the chromosome based on the marker distances within the translocated segment. The last SNP of the block was also given a 10 kb cushion between itself and the next SNP on the correct linkage group. Non-framework SNPs which had a physical position but no genetic coordinates were moved along with their framework neighbors if the nearest flanking framework markers were also moved. However, it is unclear whether non-framework SNPs just outside of these anchors should be kept in place or moved along with the adjacent SNPs. Since most inversions were small relative to the genetic map (and would therefore still fall in the same window of a sliding window analysis), these SNPs were left in place. However, markers bordering translocations were removed to ensure there were no markers mapped to the incorrect linkage group. Among the SNPs used for analyses, 15 were mis-mapped to a different linkage group and 1,585 were moved within a linkage group.

Approximate genetic positions for non-framework SNPs were interpolated based on their physical positions with the approx() function in R (R Core Team 2015) with the 4,217 framework SNPs used as a reference. The IBM genetic map distances were then converted to single-meiosis map distances using the formulae of Winkler *et al.* (2003). Finally, SNPs located at physical positions outside of those bounded by the genetic map (such as the telomeres) were assigned the genetic position of their nearest mapped neighbor. Since moved segments were arbitrarily joined 10 kb from their nearest genetic neighbor, we acknowledge that the physical positions of these markers are only estimates. However, the estimated junctions are small relative to the genetic windows used for our analysis. The final map used is provided as supplemental material (Supplementary File S1).

### **Haplotype phasing**

Although the genotypes of the plants from each population are unphased, the homozygous genotypes of the founders and derived inbreds provide excellent prior information for a probabilistic estimation of genotype phase in the populations. We therefore used fastPHASE (Scheet and Stephens 2006) to estimate the genotype phase of each plant. To estimate the error in phasing, we created test cases by combining the genotypes of

two derived inbreds into a hypothetical F1 hybrid of unknown phase. This F1 was presented to fastPHASE with the rest of the data, except that its parent inbreds were removed. Analyses of several hypothetical F1s from different cycles of selection revealed very low phasing error rates (Table S4). Therefore the phased genotypes of cycle 0 plants were used as the starting data for simulations (see below).

### Diversity and principal component analysis

Heterozygosity ( $H$ ) was measured as  $H = 2p(1 - p)$ , where  $p$  and  $(1 - p)$  are the frequencies of the two SNP alleles.  $F_{ST}$  (Hudson *et al.* 1992) was calculated using the HBKpermute program in the analysis package (<https://github.com/molpopgen/analysis>) of the software library libsequence (Thornton 2003). All results were plotted using the R package ggplot2 (Wickham 2009). We conducted PCA by singular value decomposition, as described in McVean (2009).

### Simulations

Our simulation sought to model the effects of genetic drift in the Iowa RRS experiment independent of any selection, and our model thus closely followed the published methods of the Iowa RRS (Penny and Eberhart 1971; Keeratinijakal and Lamkey 1993). Starting individuals in each population were constructed by randomly sampling two distinct haplotypes with replacement from the phased haplotypes of cycle 0. In the actual random mating scheme used in the Iowa RRS experiment, a single pairing could only contribute four gametes to the next generation (two kernels each from two ears), and our simulation reflects this. Advanced cycles were simulated by randomly mating gametes from self-fertilized plants of the previous cycle until 10 new individuals were created. The first cycle involved two rounds of random mating, whereas all subsequent cycles used one round. After cycle 5, the process employed two rounds of selfing instead of one. After cycle 7, the population size was increased from 10 to 20. At cycles 4, 8, 12, and 16, the plants were randomly mated to match the sample size of the observed data. The genotypes of these simulated random matings are the final results of each simulation and were analyzed in the same way as the observed data. Simulated recombination was carried out in R with the hypred software package ([cran.r-project.org/web/packages/hypred/](http://cran.r-project.org/web/packages/hypred/)). The number of crossovers between two parental gametes is drawn from a Poisson distribution with  $\lambda = L$ , where  $L$  is the length of the chromosome in Morgans. Crossover breakpoints are drawn from a uniform distribution over the interval  $(0, L)$ .

Simulations were executed in parallel on a computing cluster, with unique random number seeds drawn for each simulation. Statistics were calculated for each simulation using the same formula as the experimental data. We used non-overlapping sliding windows of equal genetic distance

to account for the non-independence of markers in low-recombination regions when calculating measures of significance. For the haplotype-based, single-locus simulations, recombination was simply replaced with binomial sampling of two alleles.

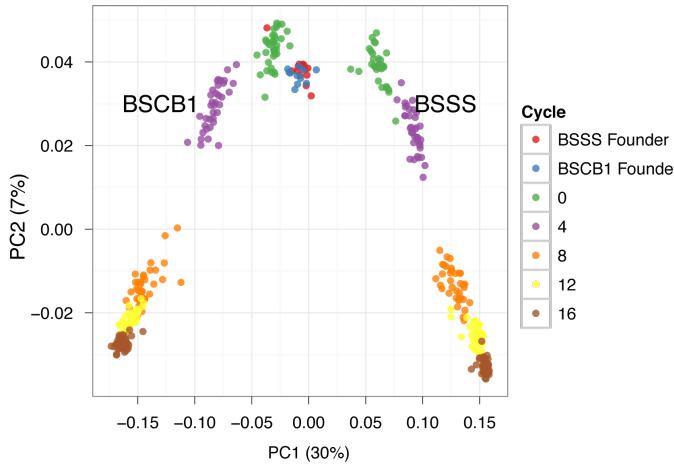
## Results

### Population structure and genetic diversity

Founder inbreds and samples from cycles 0, 4, 8, 12, and 16 were genotyped at 39,261 SNPs that passed a set of quality filters and could be assigned collinear genetic and physical map positions (See Materials and Methods for details). Change in population structure throughout the Iowa RRS experiment can be observed visually by a principal component analysis (PCA). Analysis of individuals from all the selection cycles (Figure 1) clearly separates the BSSS and BSCB1 populations along the first axis of variation, with increasing separation as the experiment progressed. The second axis of variation primarily separates the cycles from one another within each population. There is no separation between the founders of the two populations, and projection of later cycles onto a PCA of the founders shows no distinction between BSSS and BSCB1 (Figure S1). At cycle 0, however, the BSSS population shows more divergence from the founders than does BSCB1, likely due to drift during either the populations construction or subsequent maintenance. Structure continued to develop within each population over the course of the experiment. There is an especially wide gap between cycles 4 and 8, which correlates with the addition of an extra generation of self-pollination prior to selection at each cycle. The distance between cycles then decreases dramatically after cycle 8, and corresponding to the increased effective population size (See Materials and Methods).

No new genetic material was intentionally introduced into either population after the experiment's inception, so the substantial increase in genetic distance could only arise from the loss of genetic diversity within each population. Consistent with previous studies of the Iowa RRS (Messmer *et al.* 1991; Labate *et al.* 1997; Haggdorn *et al.* 2003; Hinze *et al.* 2005), genome-wide genetic diversity (expected heterozygosity,  $H$ ) decreases steadily across cycles of selection in both populations (Figure 2). The loss of heterozygosity is smaller when the two populations are considered together, indicating the loss of different alleles within BSCB1 and BSSS. This genetic differentiation is reflected by the tenfold increase in  $F_{ST}$  between the founder lines and the populations at cycle 16.

We noticed an irregular increase in the number of polymorphic markers between BSSS cycles 4 and 8 (Table S5). All of these newly polymorphic markers were present at extremely low frequency and were spread among various individuals. This may represent a series of minor alleles that were not captured in our sample of cycle 4 individuals and thus appeared to resurface at cycle 8. Alternatively, the pattern may be the

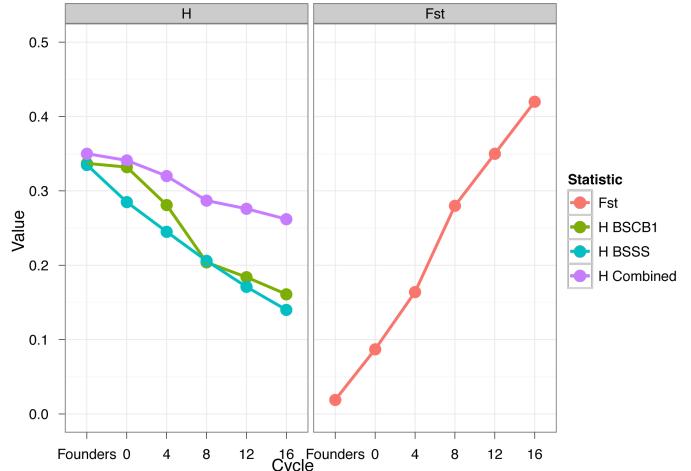


**Figure 1** Principal component analysis of the SNP data from Iowa RRS. The axes represent the first two eigenvectors from an analysis of cycles 0-16, with projection of the founder lines onto the vector space. The variation explained by each eigenvector is given in parentheses on the axes. The populations steadily diverge at increasing cycles, with less distinction visible between the founder groups. The comparatively large distance between cycles 4 and 8 corresponds to a switch from one to two generations of selfing at each cycle. The smaller separation between cycles 8-16 corresponds to an increase in effective population size from ten to twenty. The BSSS cycle 0 population has drifted away from the BSSS founders, despite the absence of intentional selection during the creation and maintenance of cycle 0.

result of minor contamination at some point in the populations history. It was observed that an allele of the sugary gene associated with sweet corn appeared in the population at this time (O.S. Smith, personal communication), suggesting contamination may be the cause. However, the low frequency of the new alleles (typically only one or two alleles out of 72 possible in 36 diploid samples) means their effect on population diversity is minimal. We did not attempt to incorporate this contamination into our simulation approaches, as it only makes our tests for low heterozygosity slightly more conservative.

### Fixation of large genomic regions

Figure 3 shows heterozygosity varying along the genome at cycle 16 of each population. Of particular note are extremely large pericentromeric regions of zero or near-zero heterozygosity spanning tens of megabases. These regions experience low rates of meiotic recombination, which creates an expanded physical map relative to their genetic length (Ganal *et al.* 2011). In general, the majority of fixed haplotype segments are small (< 2 cM) in genetic space regardless of their physical size; one exception is an 8 cM region on chromosome 1 in the BSCB1 population.

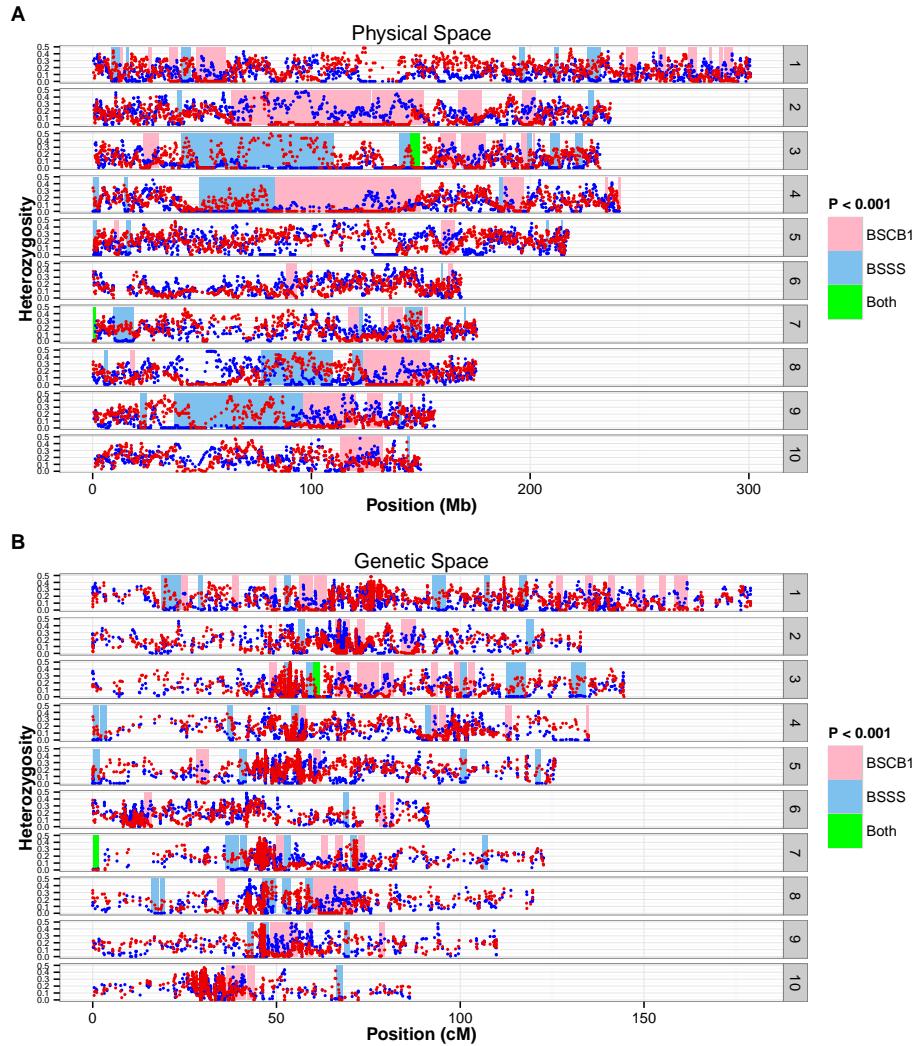


**Figure 2** Heterozygosity ( $H$ , left panel) and  $F_{ST}$  (right panel) plotted as a function of selection cycle in each population. Heterozygosity is shown for each population separately and then for the total population pooling all plants.

The sheer physical size of the pericentromeric regions yields extremely high marker density on the genetic map, allowing for clear resolution of haplotype phasing and recombination breakpoints. To further examine the fixation in these regions, we computationally imputed haplotype phase in the BSSS and BSCB1 populations and used the phased data to track haplotype frequencies and founder of origin. In most cases, these fixed haplotypes can be traced back to single founder inbreds. For example, in BSSS, a 60 Mb (2 cM) region of chromosome 9 became fixed by cycle 12 and traces back to the founder Os420, and in BSCB1, a 60 Mb (3 cM) region became fixed on chromosome 4 and traces back to A340 (Figure 4). Table 1 gives a summary of the large genomic regions that have become fixed or nearly fixed by cycle 16. These regions represent blocks of linked loci that show no evidence of recombination since at least the development of the founding inbred lines in the 1920s and 1930s.

### The role of genetic drift

There is clear evidence of phenotypic improvement in response to selection in the Iowa RRS populations (Smith, 1983; Keeratinijakal and Lamkey, 1993; Schnicker and Lamkey, 1993; Holthaus and Lamkey, 1995; Brekke *et al.*, 2011a; Brekke *et al.*, 2011b; Edwards, 2011) and large changes in genetic structure indicated by molecular markers. A central issue for these maize populations and others like them is whether the changes observed at the molecular level are caused directly by selection on phenotype, or indirectly due to the genetic drift that selection imposes through inbreeding and small effective population sizes. To gauge the roles of selection and drift, we conducted simulations of the crossing and selection schemes used in the RRS experiment. Selection was executed at random in each



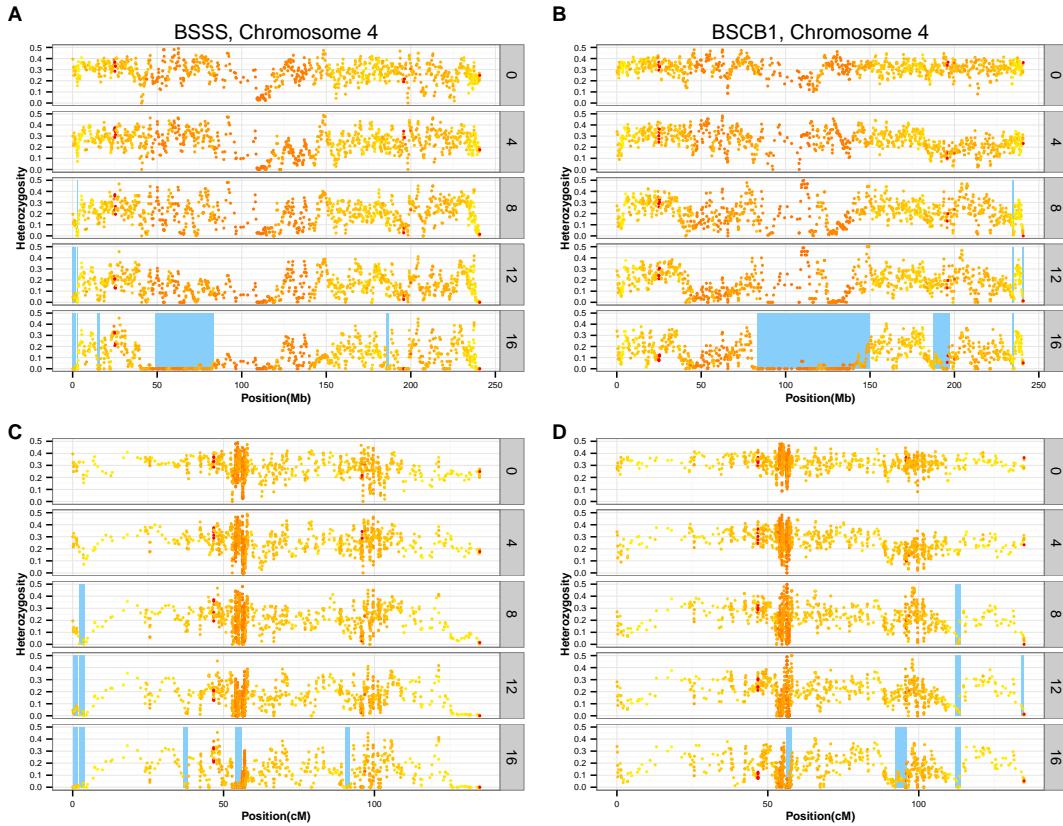
**Figure 3** Heterozygosity at cycle 16 across all ten chromosomes in each population, calculated on 15-marker sliding windows with 5 marker steps. Heterozygosity values in BSSS (blue dots) and BSCB1 (red dots) are superimposed in one panel. 2 cM windows of heterozygosity lower than 10 of 10,000 simulations ( $P < 0.001$ ) are shaded in light blue (BSSS) or pink (BSCB1), and correspond to a genome-wide false discovery rate of  $< 3\%$  in both populations. Two regions genome-wide show significantly low heterozygosity in both populations and are shaded green. (A) Physical map. (B) Genetic map. The same data are plotted separately for each population in Figure S2.

simulation, so the patterns observed across simulations represent the expected distribution of effects caused only by recombination and genetic drift. We conducted 10,000 simulations, modeling recombination using the IBM genetic map (Lee *et al.* 2002), which is based on a cross between a BSSS-derived inbred line used to construct the reference genome (B73) and an inbred with co-ancestry from both the BSSS and BSCB1 founder germplasm (Mo17).

Averaged across the genome, the vast majority of the reduction in diversity observed in both populations can be attributed to genetic drift. Nonetheless, we do observe differences from

values generated under our neutral simulations (Figure 5). The observed data show higher than expected heterozygosity at cycle 4, which could be explained by several factors, including differences between simulated and actual breeding practices, under-sampling of diversity in cycle 0, selection acting to increase the frequency of initially rare haplotypes, or even selection acting to maintain heterozygosity (e.g. McMullen *et al.* 2009; Gore *et al.* 2009).

As the cycles progress, heterozygosity falls more rapidly than expected in both populations, and the observed values at cycle 16 are significantly lower than the simulated data. Simu-



**Figure 4** Heterozygosity in each cycle across chromosome 4 of the BSSS (left) and BSCB1 (right) plotted on the physical (top) and genetic (bottom) map. Each panel consists of five plots representing cycles 0,4,8,12, and 16 of the experiment. Heterozygosity is calculated on 15-marker sliding windows, with 5 marker steps between each calculation. Each data point is color-coded based on a linear transformation of recombination rate (red = low recombination rate). Shaded regions represent 2 cM windows with heterozygosity values significantly lower than expected by simulation at a given cycle at  $P < 0.001$ . Plots for other chromosomes are shown in Figure S3.

**Table 1 Ancestry of haplotypes fixed in the cycle 16 population.**

Population	Chr.	Interval (cM)	Interval (Mb)	Founder	Derived Lines
BSSS	3	53.6-55.3	67.7-123	CI187-2	B94
BSSS	3	57.3-64.8	129.2-157.1	ND <sup>a</sup>	B89, B94
BSSS	4	52.8-55.6	39.9-82.7	CI187-2	B89, B94, B67, B72, B39, B43
BSSS	9	41.1-44.8	20.8-26.6	Oh <sup>b</sup>	B89, B94, B43, B17, B72, B84, B67
BSSS	9	45.7-47.5 <sup>c</sup>	30.8-90.4	Os420	B89, B94
BSCB1	2	67-67.5	80.6-114.5	CC5	B90, B91, B95, B97, B99
BSCB1	4	55.6-57	82.7-140	ND <sup>a</sup>	B90, B95, B97
BSCB1	8	61.5-67.7	125.1-145.6	P8	B90, B97, B91, B99, B54

<sup>a</sup> ND = Not Determined (Either a recombinant haplotype or originates from an un-genotyped founder)

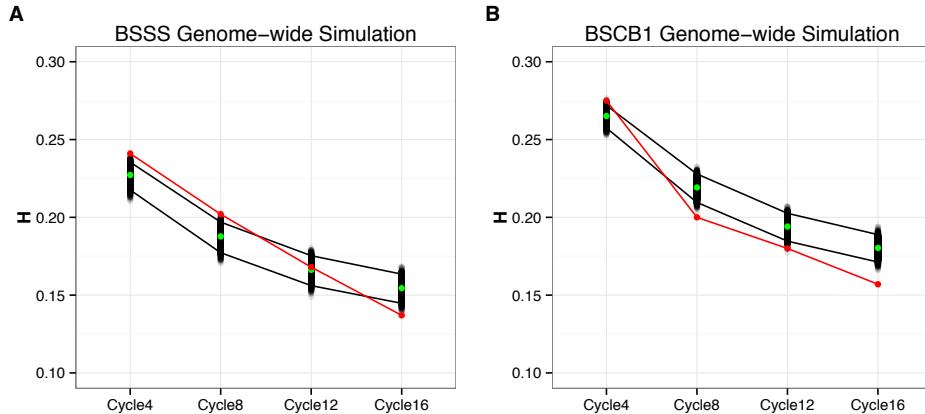
<sup>b</sup> Although Oh7 is a BSCB1 founder, it is a descendant of CI.540, an un-genotyped BSSS founder. BSSS segments matching Oh7 presumably derive from CI.540

<sup>c</sup> Founders Ind\_B2 (BSSS), CI187-2 (BSSS), R4 (BSCB1), and I205 (BSCB1) are all IBD at this region of chromosome 9

lations across a number of different marker densities were consistent with this result (data not shown).

To examine the behavior of specific regions, we compared observed and simulated results for each 2 cM segment of the

genome. The dynamics observed across most of the genome are largely insensitive to window size (we tested from 2-4 cM, data not shown), and are consistent with strong genetic drift imposed by the experimental design. A subset of loci were



**Figure 5** Heterozygosity in each population, observed vs. simulated data. Heterozygosity was calculated as the average across all markers genome-wide in the BSSS (A) and BSCB1 (B) populations. The observed data is marked by the red line, the simulations by gray dots, and the median of the simulations by a green dot. Black lines represent the 99% and 1% quantiles of the simulated data.

flagged as significant (Figure 3), and these loci almost always overlapped regions of fixation or near-zero heterozygosity in one population. Simulated values in these regions are often quite low as well (Supplementary File S2), however, indicating that drift alone can explain most of the drop in diversity.

Since the population size of the Iowa RRS is small (10–20), many biallelic SNPs should fix by chance regardless of their starting minor allele frequencies. Observed differences from the simulated neutral expectation thus do not arise from changes in allele frequencies per se, but rather from the fixation of linked markers across larger than expected genetic distances. The validity of significance cutoffs therefore depend on the accuracy of our genetic map. While the maize genetic map is known to vary among genetic backgrounds across short distances (McMullen *et al.* 2009), broad-scale patterns of recombination appear relatively stable across diverse germplasm (Rodgers-Melnick *et al.* 2015, but see Bauer *et al.* (2013)). Differences between observed and simulated results could be due to selection, variation or inaccuracy in the genetic map, or a combination of these factors.

To explore the roles of selection and drift independent of the genetic map, we returned to the large regions of fixation in the centromeres which showed no recombination across the full RRS experiment. Given the lack of recombination, each of these regions can be analyzed by the simulation of a single locus, and the high density of markers allows the clear resolution of the individual haplotypes. We used the computationally phased data to measure the frequency of the fixed haplotype at each cycle, and assessed the probability of observing the fixation event given the initial frequency. The BSSS chromosome 9 haplotype fixed at cycle 16 was at low frequency at cycle 0 (7 out of 68 haplotypes), but increased rapidly in frequency by cycle 8 (66 out of 70 haplotypes). Simulation of the haplotype as a single locus in the RRS experiment produces this

increase in frequency in only 3.9% of 1000 independent simulations, whereas the haplotype was lost in over 80% of the simulations. In BSCB1, a 30 Mb, a < 1cM region of chromosome 2 became nearly fixed by cycle 8 (67/70) despite a prevalence of 4/72 at cycle 0, which occurred 1.5% of the time by simulation. Although these results suggest that selection may have pushed these haplotypes to fixation, the fact that fixation of such a rare haplotype still occurred in some simulations speaks to the strong genetic drift imposed upon the BSSS and BSCB1 populations. Interestingly, each of these two genomic regions harbored a different cycle 0 haplotype at higher frequency, but these higher-frequency haplotypes were subsequently lost within the RRS population. In other cases, the haplotypes which eventually fixed were at moderate frequency in the cycle 0 populations and drift to fixation in the majority of simulations. Several key inbreds in the stiff-stalk heterotic group — B73, B37 and B14 — were derived from the BSSS population (Darrah and Zuber 1986; Troyer 1999). B37 and B14 were derived from cycle 0, and B73 was derived from a half-sib recurrent selection program also started with the BSSS population. We examined these three inbreds at the pericentromeric regions listed in Table 1, and found that in most cases they carry different haplotypes from those that rose to high frequency in the RRS experiment.

## Discussion

Our analysis of the Iowa RRS experiment reveals a steady loss of diversity in the BSSS and BSCB1 populations as they became increasingly differentiated from one another over time. Principal component analysis shows that as the effective population size and the rates of inbreeding were altered, the rates of change in population structure were altered as well. These patterns of population structure, diversity, and differentiation

between BSSS and BSCB1 can be largely reproduced by simulation without any selection, supporting the hypothesis that the majority of the genetic structure observed can be attributed to genetic drift alone, despite effective selection for phenotypic improvement. Similar observations have been recently made in other reciprocal recurrent selection programs, even in comparisons of multiple replicated populations (Romay *et al.* 2012; Lamkey and Lorenz 2014). Reciprocal recurrent selection serves as a model for the method of hybrid maize improvement (Duvick *et al.* 2004), and similar patterns of diversity and population structure can be seen broadly across North American maize germplasm (van Heerwaarden *et al.* 2012). Genetic drift has thus most likely played a large role in the current genetic structure of modern maize. These patterns differ markedly, however, from experimental evolution in systems such as *Drosophila* (Turner *et al.* 2011; Burke *et al.* 2010) and *E. coli* (Tenaillon *et al.* 2012) in which the effects of selection on diversity are readily discernible. A key difference between these studies and those in maize are the effective population sizes, which were kept much lower in maize to produce a short-term phenotypic response to selection within available field testing resources.

Although drift can explain most of the genetic structure genome-wide, phenotypic data provide clear evidence that selection has altered the frequencies of favorable alleles in the BSSS and BSCB1 populations. Numerous experiments have shown that the selected populations and the hybrids formed from them exhibit genetic gain for hybrid yield as well plant architecture and tolerance to high-density planting (Smith 1983; Keeratinijakal and Lamkey 1993; Schnicker and Lamkey 1993; Holthaus and Lamkey 1995; Brekke *et al.* 2011a,b; Edwards 2011; Lauer *et al.* 2012). We find that heterozygosity falls more than expected across the genome as a whole, and, though drift imposes limitations on the power to detect selection at individual loci, genomic regions of extremely low diversity evident at cycle 16 are unlikely to be produced by drift alone. We further show that an identity-by-descent, haplotype-based approach provides additional power to identify selected regions, as it can distinguish between the fixation of rare and common haplotypes. These analyses show that the most likely targets of selection occur at different loci in the two populations, a result consistent with analyses in commercial breeding programs (Feng *et al.* 2006) and which may help explain the lack of selection seen in previous analyses across numerous breeding programs (van Heerwaarden *et al.* 2012).

The observation that different targets of selection are observed in opposing heterotic populations bears implications for the genetic mechanisms responsible for heterosis and the success of maize hybrids. Classic overdominance models of heterosis predict that at a single locus two distinct alleles confer heterozygote advantage when combined. Alternatively, the dominance model predicts that heterosis is driven by dominance effects and the complementation of linked alleles in low-recombination regions (dominance or pseudo-overdominance).

In the case of true overdominance, selection should thus lead to decreased heterozygosity at the same locus in both populations as complementary haplotypes are fixed in each group (e.g. Guo *et al.* 2014). We find little evidence to support this genetic phenomenon, finding only two 2 cM windows genome-wide in which both populations show significantly reduced heterozygosity. Although we cannot rule out soft sweeps of complementary overdominant alleles, the observed pattern more parsimoniously favors a dominance model, in which fixation of a haplotype in one population simply selects against that same haplotype in the other population. Although strongly deleterious variants were likely purged during the inbreeding process leading to the founder lines, many weakly deleterious alleles can be found segregating at low frequencies among inbreds (Mezmouk and Ross-Ibarra 2014). Because deleterious alleles will be rare in both populations, most haplotypes in the second population will have a different suite of deleterious variants and will complement the fixed haplotype reasonably well. We expect that selection against homozygosity of the fixed haplotype will thus have little impact on diversity in the second population. Although our results better fit the simpler dominance model, the ability to distinguish between models will depend strongly on allele frequencies as well as the effects of selection, and drift. This is especially true because in a model of hybrid complementation, genetic drift in one population can alter the selective value of alleles in the other population. Given these complexities, empirical evaluation of the effects of putatively selected haplotypes will play a key role in distinguishing opposing genetic models.

## Acknowledgements

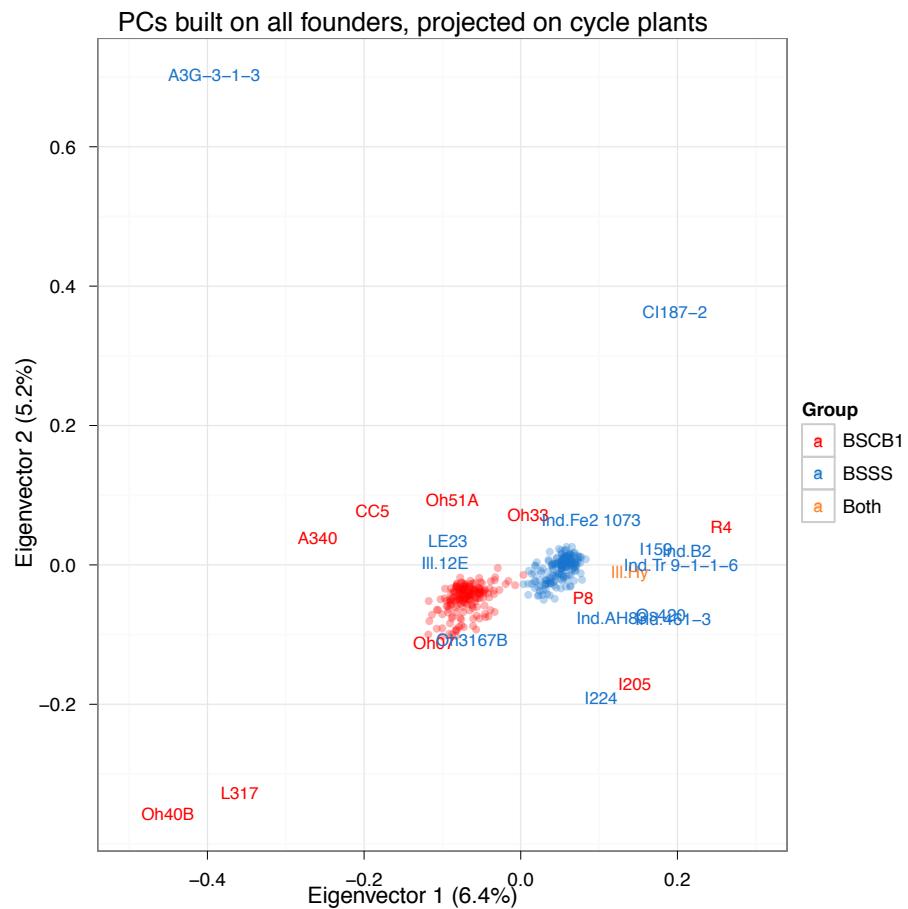
We thank Oscar “Howie” Smith, members of the Ross-Ibarra lab, and two anonymous reviewers for comments on earlier versions of the manuscript. We also thank the editor Stephen Wright for his patience. J.P.G received support for this research as a Merck Fellow of the Life Sciences Research Foundation. This research was supported by the National Science Foundation (IOS-0820619) and funds provided to USDA-ARS (MDM). Names of products are necessary to report factually on available data: however, neither the USDA, nor any other participating institution guarantees or warrants the standard of the product and the use of the name does not imply approval of the product to the exclusion of others that may also be suitable.

## Literature Cited

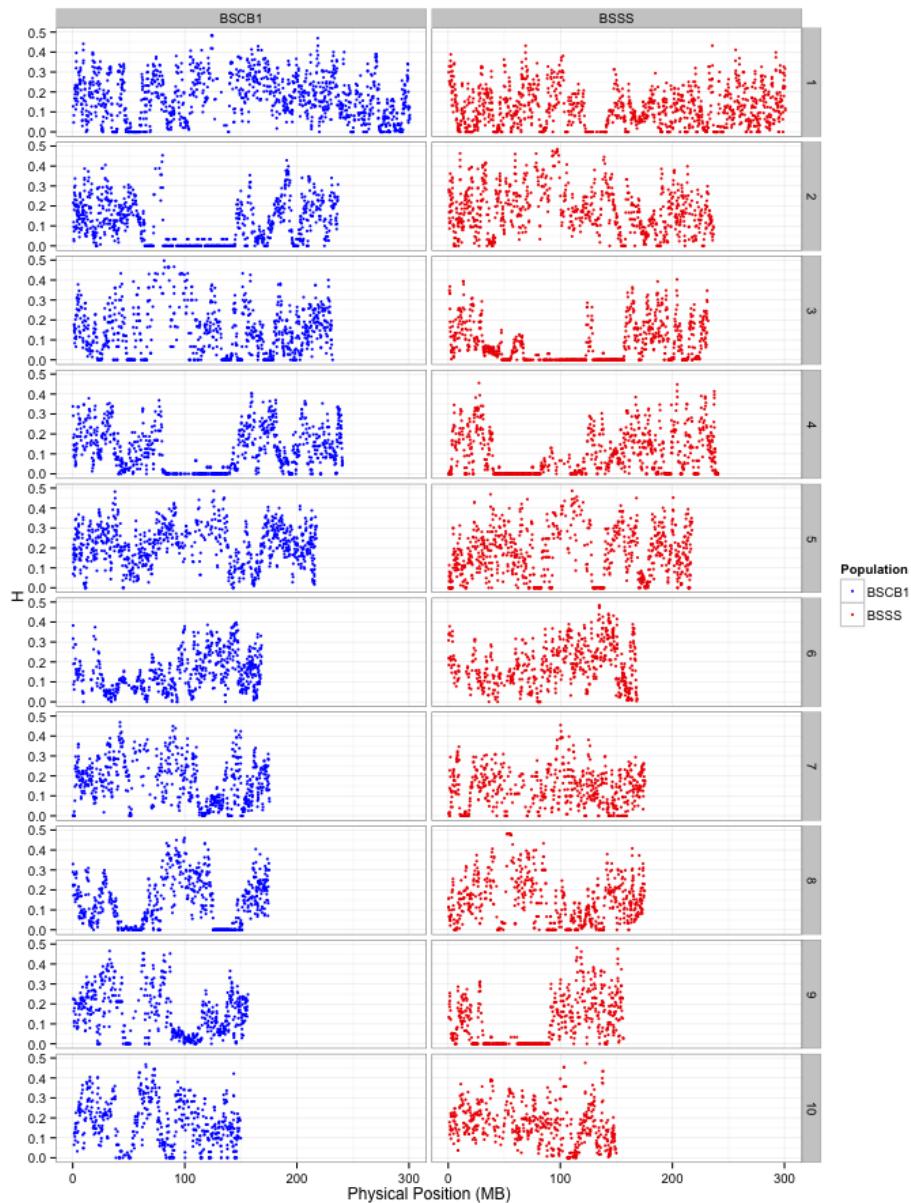
- Anderson, E., 1944 The sources of effective germ-plasm in hybrid maize. Annals of the Missouri Botanical Garden pp. 355–361.
- Bauer, E., M. Falque, H. Walter, C. Bauland, C. Camisan, *et al.*, 2013 Intraspecific variation of recombination rate in maize. Genome Biology 14.

- Beissinger, T. M., C. N. Hirsch, B. Vaillancourt, S. Deshpande, K. Barry, *et al.*, 2014 A genome-wide scan for evidence of selection in a maize population under long-term artificial selection for ear number. *Genetics* 196: 829–840.
- Brekke, B., J. Edwards and A. Knapp, 2011a Selection and adaptation to high plant density in the iowa stiff stalk synthetic maize (l.) population. *Crop Science* 51: 1965–1972.
- Brekke, B., J. Edwards and A. Knapp, 2011b Selection and adaptation to high plant density in the iowa stiff stalk synthetic maize (l.) population: II. plant morphology. *Crop Science* 51: 2344–2351.
- Burke, M. K., J. P. Dunham, P. Shahrestani, K. R. Thornton, M. R. Rose, *et al.*, 2010 Genome-wide analysis of a long-term evolution experiment with drosophila. *Nature* 467: 587–590.
- Chia, J. M., C. Song, P. J. Bradbury, D. Costich, N. de Leon, *et al.*, 2012 Maize hapmap2 identifies extant variation from a genome in flux. *Nat. Genet.* 44: 803–807.
- Comstock, R. E., H. Robinson and P. Harvey, 1949 Breeding procedure designed to make maximum use of both general and specific combining ability. *Agronomy Journal*.
- Crabb, A. R., and H. Hughes, 1947 *The hybrid-corn makers. Prophets of plenty.* Rutgers University Press, New Brunswick.
- Crow, J. F., 1998 90 years ago: the beginning of hybrid maize. *Genetics* 148: 923–928.
- Darrah, L., and M. Zuber, 1986 1985 united states farm maize germplasm base and commercial breeding strategies. *Crop Science* 26: 1109–1113.
- Duvick, D. N., 2005 The contribution of breeding to yield advances in maize (*zea mays l.*). *Advances in Agronomy* 86: 83–145.
- Duvick, D. N., J. Smith, M. Cooper *et al.*, 2004 Long-term selection in a commercial hybrid maize breeding program. *Plant Breeding Reviews* 24: 109–152.
- Edwards, J., 2011 Changes in plant morphology in response to recurrent selection in the iowa stiff stalk synthetic maize population. *Crop Science* 51: 2352–2361.
- Feng, L., S. Sebastian, S. Smith and M. Cooper, 2006 Temporal trends in ssr allele frequencies associated with long-term selection for yield of maize. *Maydica* 51: 293.
- Ganal, M. W., G. Durstewitz, A. Polley, A. Bérard, E. S. Buckler, *et al.*, 2011 A large maize (*Zea mays L.*) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLoS One* 6: e28334.
- Gore, M., J. Chia, R. Elshire, Q. Sun, E. Ersoz, *et al.*, 2009 A first-generation haplotype map of maize. *Science* 326: 1115–1117.
- Guo, M., M. A. Rupe, J. Wei, C. Winkler, M. Goncalves-Butruille, *et al.*, 2014 Maize argos1 (zar1) transgenic alleles increase hybrid maize yield. *Journal of experimental botany* 65: 249–260.
- Hagdorn, S., K. R. Lamkey, M. Frisch, P. E. Guimaraes and A. E. Melchinger, 2003 Molecular genetic diversity among progenitors and derived elite lines of bsss and bscb1 maize populations. *Crop Science* 43: 474–482.
- Hallauer, A. R., S. Eberhart and W. Russell, 1974 Registration of maize germplasm1 (reg. no. gp 26 to gp 34). *Crop Science* 14: 341–342.
- Hinze, L. L., S. Kresovich, J. D. Nason and K. R. Lamkey, 2005 Population genetic diversity in a maize reciprocal recurrent selection program. *Crop Science* 45: 2435–2442.
- Hirsch, C. N., S. A. Flint-Garcia, T. M. Beissinger, S. R. Eichten, S. Deshpande, *et al.*, 2014 Insights into the effects of long-term artificial selection on seed size in maize. *Genetics* 198: 409–421.
- Ho, J., S. Kresovich and K. Lamkey, 2005 Extent and distribution of genetic variation in us maize: historically important lines and their open-pollinated dent and flint progenitors. *Crop Science* 45: 1891–1900.
- Holthaus, J. F., and K. R. Lamkey, 1995 Population means and genetic variances in selected and unselected iowa stiff stalk synthetic maize populations. *Crop Science* 35: 1581–1589.
- Hudson, R., D. D. Boos and N. Kaplan, 1992 A statistical test for detecting geographic subdivision. *Molecular biology and evolution* 9: 138–151.
- Hufford, M., X. Xu, J. van Heerwaarden, T. Pyhajarvi, J. Chia, *et al.*, 2012 Comparative population genomics of maize domestication and improvement.\*. *Nature Genetics* 44: 808–811.
- Jiao, Y., H. Zhao, L. Ren, W. Song, B. Zeng, *et al.*, 2012 Genome-wide genetic changes during modern breeding of maize. *Nature genetics* 44: 812–815.
- Keeratinijakal, V., and K. R. Lamkey, 1993 Responses to reciprocal recurrent selection in bsss and bscb1 maize populations. *Crop Science* 33: 73–77.
- Labate, J. A., K. R. Lamkey, M. Lee and W. L. Woodman, 1997 Molecular genetic diversity after reciprocal recurrent selection in bsss and bscb1 maize populations. *Crop Science* 37: 416–423.

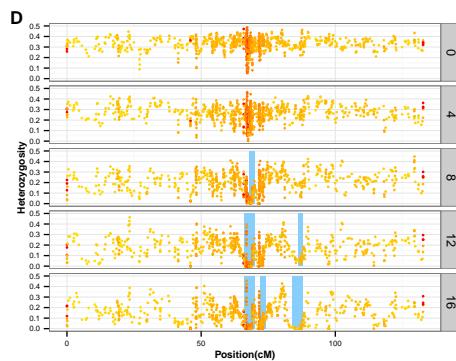
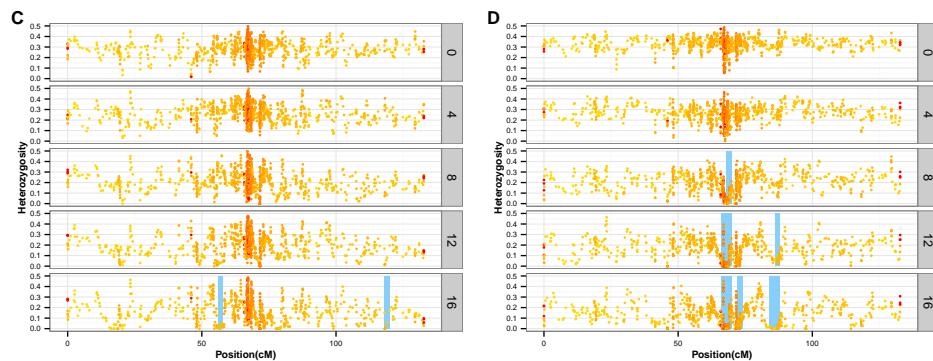
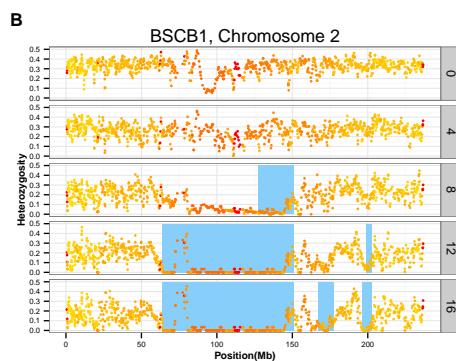
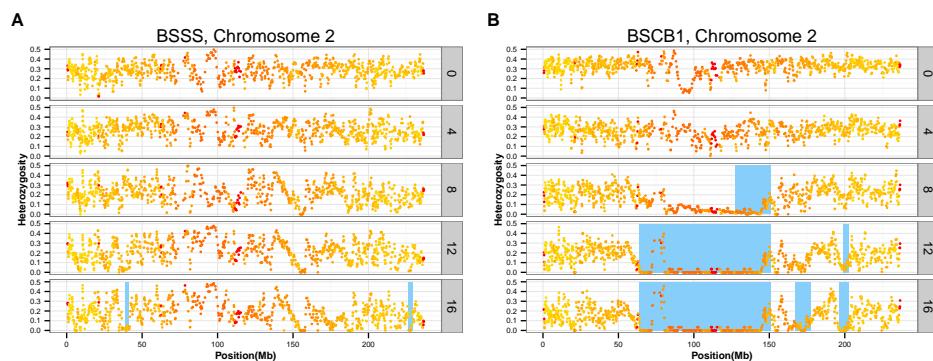
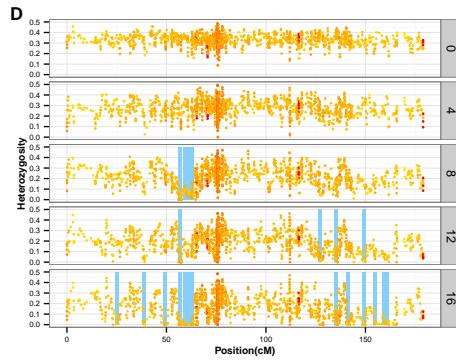
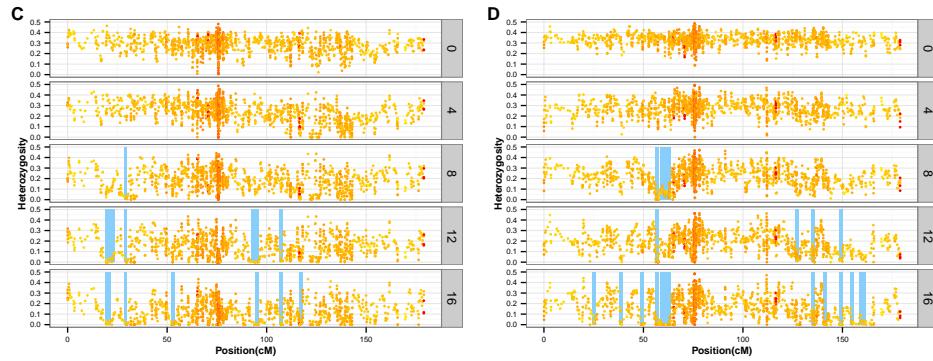
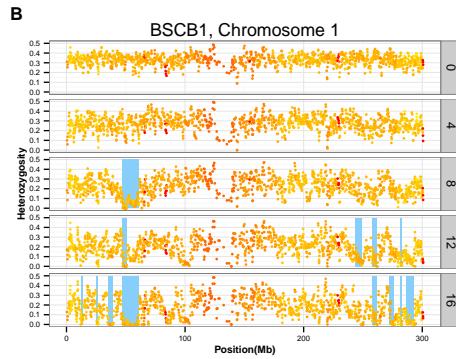
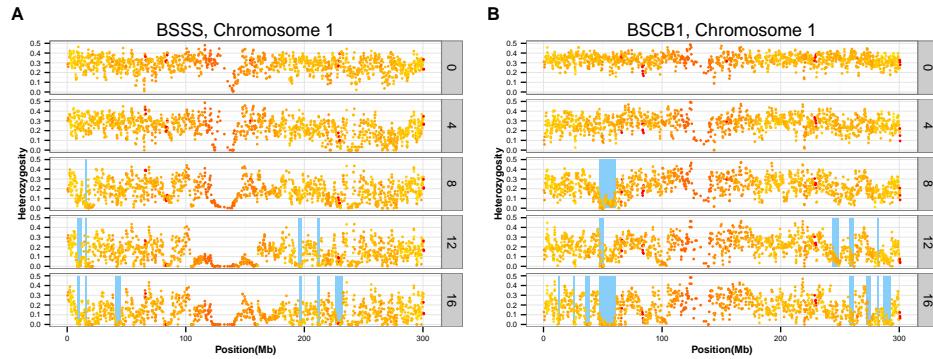
- Lamkey, C., and A. Lorenz, 2014 Relative effect of drift and selection in diverging populations within a reciprocal recurrent selection program. *Crop Science* 54: 576–585.
- Lauer, S., B. D. Hall, E. Mulaosmanovic, S. R. Anderson, B. Nelson, *et al.*, 2012 Morphological changes in parental lines of pioneer brand maize hybrids in the us central corn belt. *Crop Science* 52: 1033–1043.
- Lee, M., N. Sharopova, W. D. Beavis, D. Grant, M. Katt, *et al.*, 2002 Expanding the genetic map of maize with the intermated b73 × m017 (ibm) population. *Plant Molecular Biology* 48: 453–461.
- McMullen, M., S. Kresovich, H. Villeda, P. Bradbury, H. Li, *et al.*, 2009 Genetic properties of the maize nested association mapping population.\*. *Science* 325: 737–740.
- McVean, G., 2009 A genealogical interpretation of principal components analysis. *PLoS Genetics* 5: e1000686.
- Messmer, M., A. Melchinger, M. Lee, W. Woodman, E. Lee, *et al.*, 1991 Genetic diversity among progenitors and elite lines from the iowa stiff stalk synthetic (bsss) maize population: comparison of allozyme and rflp data. *Theoretical and Applied Genetics* 83: 97–107.
- Mezmouk, S., and J. Ross-Ibarra, 2014 The pattern and distribution of deleterious mutations in maize. *G3 (Bethesda)* 4: 163–171.
- Penny, L. t., and S. Eberhart, 1971 Twenty years of reciprocal recurrent selection with two synthetic varieties of maize (*zea mays* l.). *Crop Science* 11: 900–903.
- R Core Team, 2015 *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing Vienna, Austria.
- Rodgers-Melnick, E., P. J. Bradbury, R. J. Elshire, J. C. Glaubitz, C. B. Acharya, *et al.*, 2015 Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proceedings of the National Academy of Sciences* 112: 3823–3828.
- Romay, M. C., A. Butrón, A. Ordás, P. Revilla and B. Ordás, 2012 Effect of recurrent selection on the genetic structure of two broad-based spanish maize populations. *Crop Science* 52: 1493–1502.
- Saghai-Marof, M. A., K. M. Soliman, R. A. Jorgensen and R. W. Allard, 1984 Ribosomal DNA spacer-length polymorphisms in barley - Mendelian inheritance, chromosomal location, and population-dynamics. *Proc. Natl. Acad. Sci. USA* 81: 8014–8018.
- Scheet, P., and M. Stephens, 2006 A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *The American Journal of Human Genetics* 78: 629–644.
- Schnable, P. S., D. Ware, R. S. Fulton, J. C. Stein, F. Wei, *et al.*, 2009 The B73 maize genome: complexity, diversity, and dynamics. *Science* 326: 1112–1115.
- Schnicker, B. J., and K. R. Lamkey, 1993 Interpopulation genetic variance after reciprocal recurrent selection in bsss and bscb1 maize populations. *Crop Science* 33: 90–95.
- Senior, M., J. Murphy, M. Goodman and C. Stuber, 1998 Utility of ssrs for determining genetic similarities and relationships in maize using an agarose gel system. *Crop Science* 38: 1088–1098.
- Smith, O., 1983 Evaluation of recurrent selection in bsss, bscb1, and bs13 maize populations. *Crop Science* 23: 35–40.
- Tenaillon, O., A. Rodríguez-Verdugo, R. L. Gaut, P. McDonald, A. F. Bennett, *et al.*, 2012 The molecular diversity of adaptive convergence. *Science* 335: 457–461.
- Thornton, K., 2003 Libsequence: a c++ class library for evolutionary genetic analysis. *Bioinformatics* 19: 2325–2327.
- Tracy, W., and M. Chandler, 2006 The historical and biological basis of the concept of heterotic patterns in corn belt dent maize pp. 219–233 in *Plant Breeding: The Arnel R. Hallauer International Symposium*. Wiley Online Library.
- Troyer, A. F., 1999 Background of us hybrid corn. *Crop Science* 39: 601–626.
- Turner, T. L., A. D. Stewart, A. T. Fields, W. R. Rice and A. M. Tarone, 2011 Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *drosophila melanogaster*.
- van Heerwaarden, J., M. B. Hufford and J. Ross-Ibarra, 2012 Historical genomics of north american maize. *Proceedings of the National Academy of Sciences* 109: 12420–12425.
- Wickham, H., 2009 *ggplot2: elegant graphics for data analysis*. Springer Science & Business Media.
- Winkler, C. R., N. M. Jensen, M. Cooper, D. W. Podlich and O. S. Smith, 2003 On the determination of recombination rates in intermated recombinant inbred populations. *Genetics* 164: 741–745.

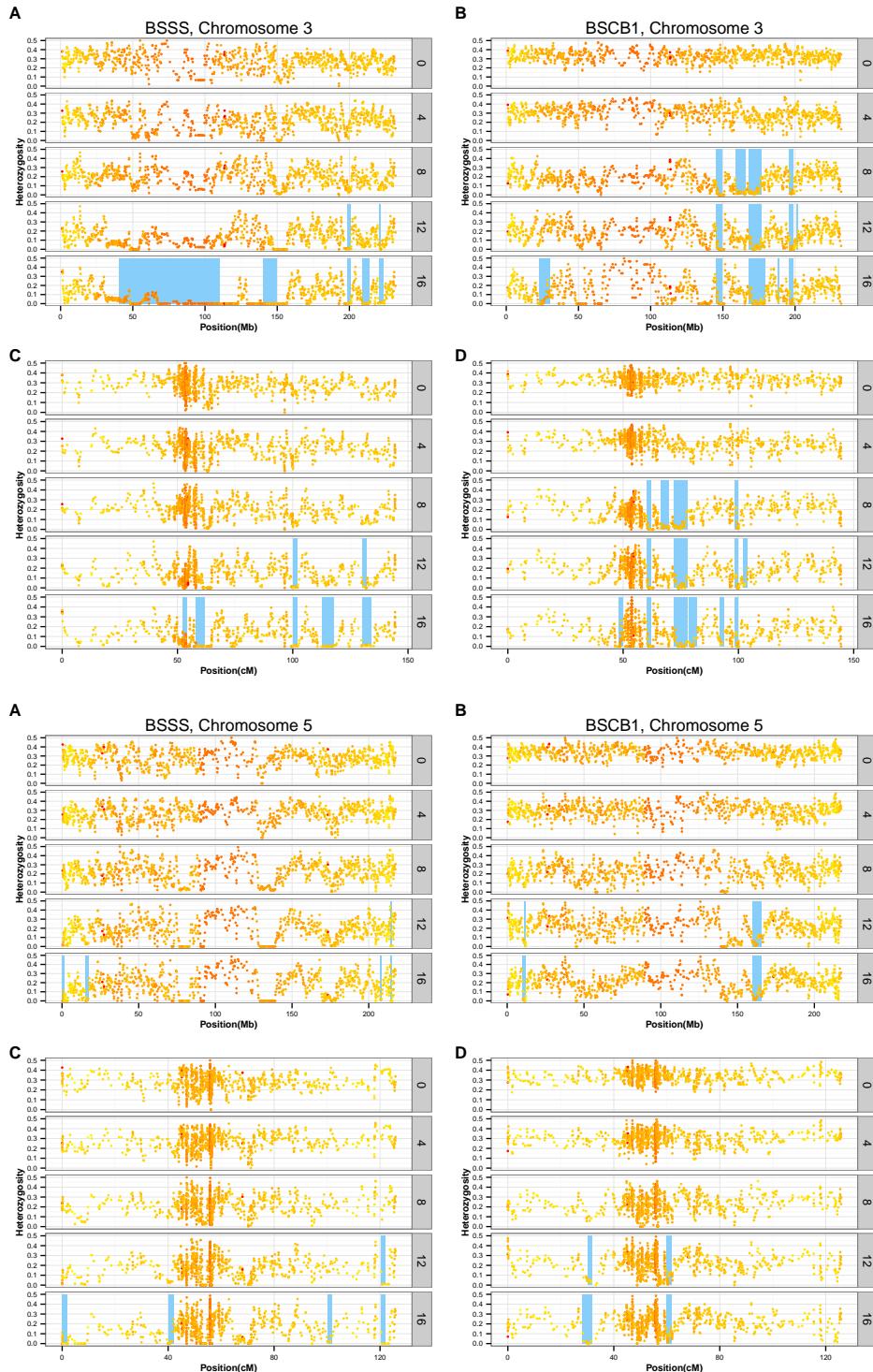


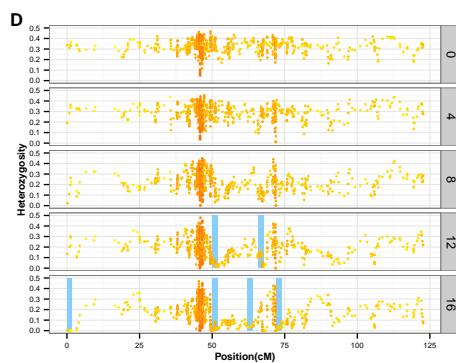
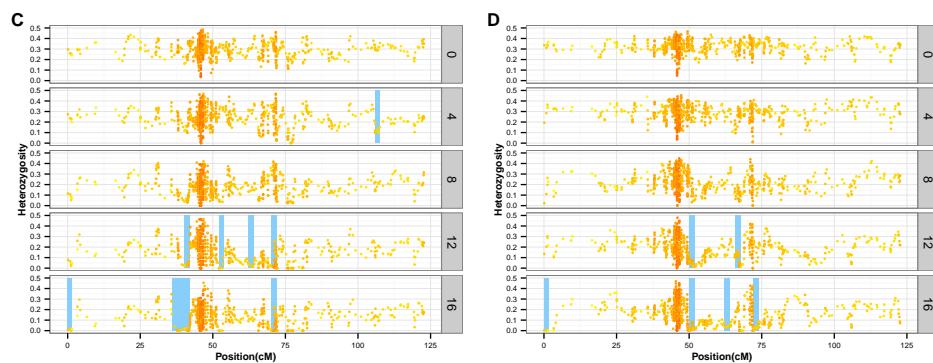
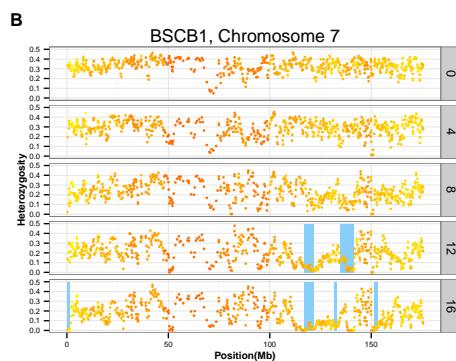
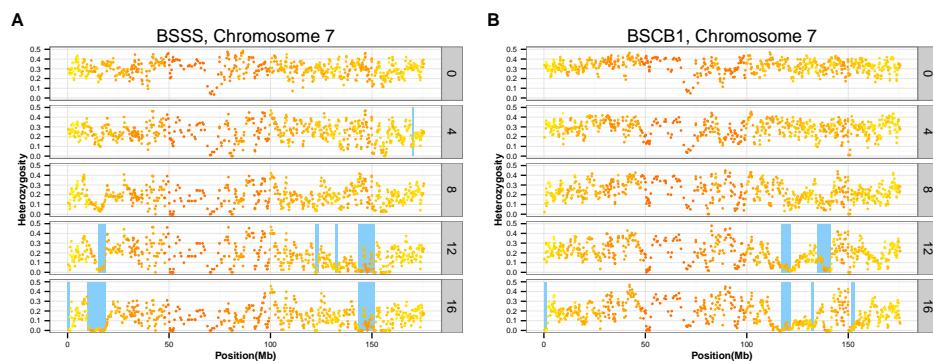
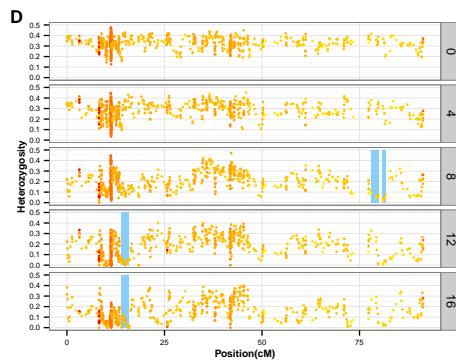
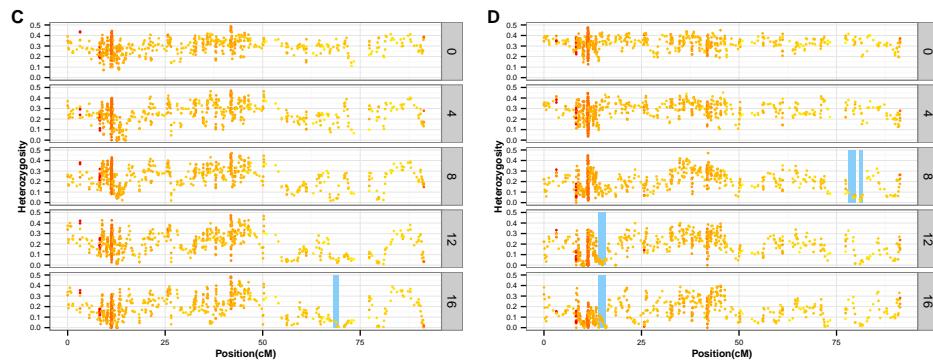
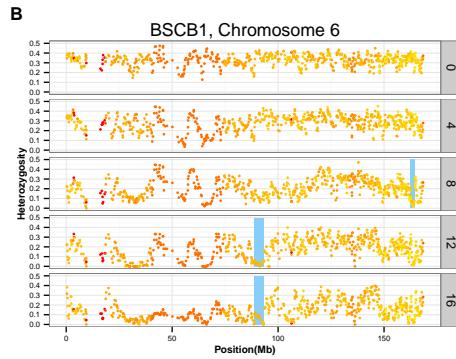
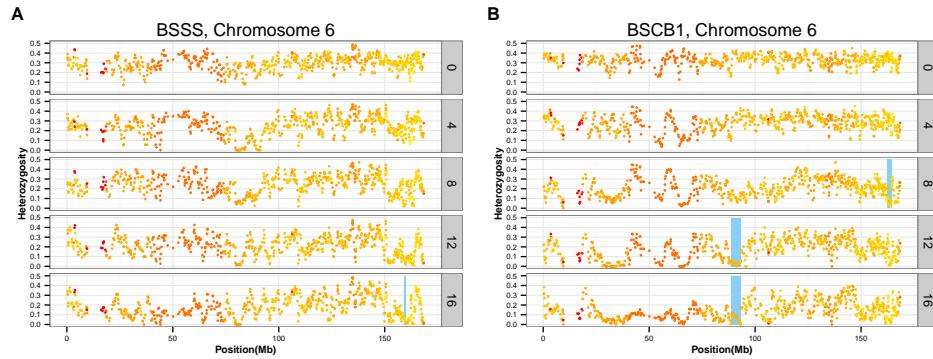
**FIGURE S1** Principle component analysis of founder inbred lines. The names of founder inbreds are shown on the graph; all other points represent BSSS (blue) and BSCB1 (red) individuals projected onto the PCA of the founders.

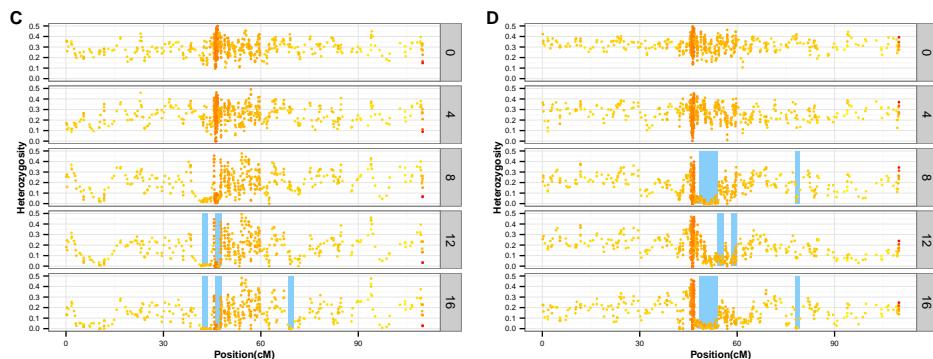
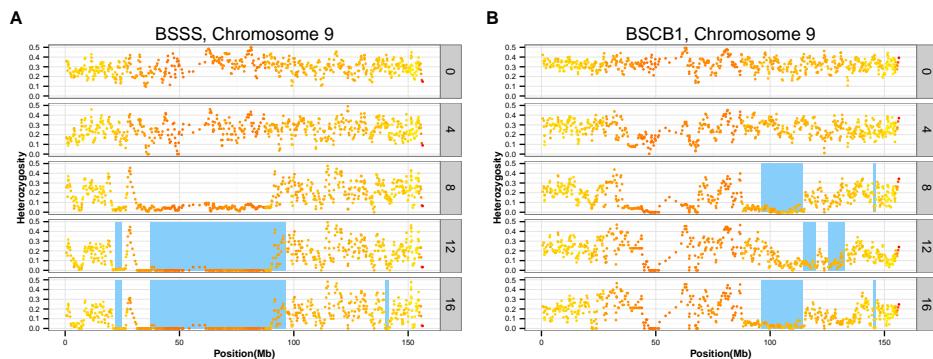
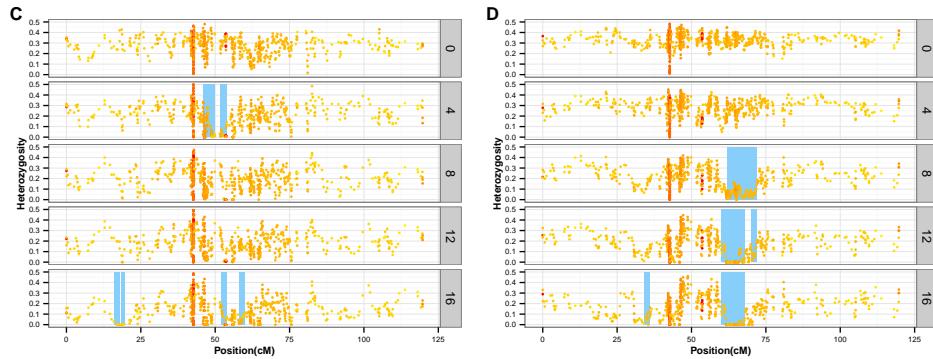
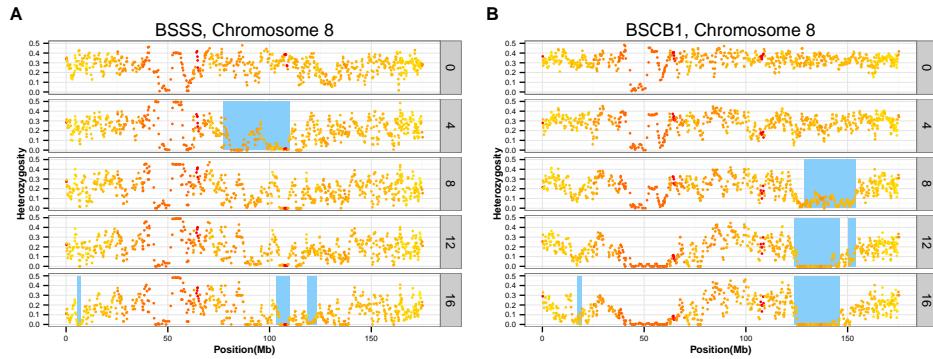


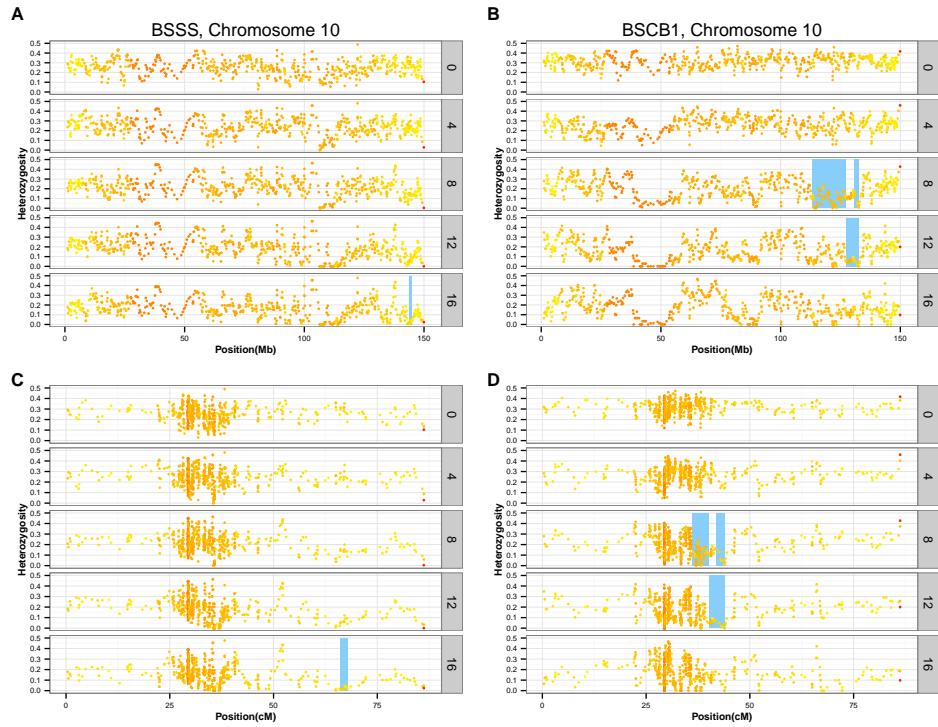
**FIGURE S2** Heterozygosity ( $H$ ) at cycle16 across all ten chromosomes in each population.  $H$  is calculated on 15-marker sliding windows with 5 marker steps. Each point is plotted at the midpoint of the 15-marker window.











**FIGURE S3** Heterozygosity in each cycle across chromosomes of the BSSS (left) and BSCB1 (right) plotted on the physical (top) and genetic (bottom) map. Details are as in Fig. 4.

**TABLE S1 Founder lines genotyped.**  
BSSS Founders Genotyped

Inbred	Background / Pedigree	Notes
Ind_Tr9_1_1_6	Reid Early Dent (Troyer Strain)	
Oh3167B	Echelberger Clarage	
I224	Iodent	
Ind467(744)	Reid Medium	
CI.187-2	Krug-Nebraska Reid x IA Gold Mine	
Os420	Osterland Yellow Dent	
I159	Iodent	
A3G-3-1-3	BL345BxIAI129	
Ind_Fe2_1073	Troyer Reid (Early)	Parent of unavailable line F1B1
III_Hy	IL High Yield	
III_12E	unknown	
Ind_AH83	Funk 176A	
Ind_B2	Troyer Reid (Late Butler)	Parent of unavailable line F1B1
LE23	IL Low Ear	

BSCB1 Founders Genotyped

Inbred	Background / Pedigree
I205	Iodent
Oh51A	[(OH56xWf9)Oh56] (Wooster Clarage x ?)
A340	4-29 x 64 (Silver King x Northwestern Dent)
III_Hy	IL High Yield
Oh33	Clarage
Oh07	C.I.540xIII.L
R4	Funk Yellow Dent
Oh40B	eight line LSC composite
P8	Palin Reid
L317	LSC
CC5	Golden Glow (W23)

Founders not analyzed

Inbred	Group	Reason
CI.540	BSSS	heterozygous genotype
F1B1	BSSS	unavailable
CI.617	BSSS	unavailable
WD456	BSSS	unavailable
K230	BSCB1	source segregates phenotypically

**TABLE S2** Derived lines genotyped.

Inbred	Group	Cycle	Notes
B10	BSSS	0	thrown out; poor data quality
B42	BSCB1	0	
B14A	BSSS	0	Cuzco x B14
B43	BSSS	0	
B10	BSSS	0	
B37	BSSS	0	
B44	BSSS	0	
B17	BSSS	0	
B69	BSSS	0	
B39	BSSS	0	
B90	BSCB1	7	
B40	BSSS	0	
B54	BSCB1	0	
B78	BSSS	8	from half-sib recurrent selection program
B72	BSSS	3	from half-sib recurrent selection program
B84	BSSS	7	from half-sib recurrent selection program
B94	BSSS	8	
B99	BSCB1	10	
B11	BSSS	0	
B89	BSSS	7	
B95	BSCB1	7	
B91	BSCB1	8	
B67	BSSS	0	
B73	BSSS	5	from half-sib recurrent selection program
B97	BSCB1	9	

**TABLE S3 Number of individuals genotyped at each cycle.**

Population	Cycle	# plants
BSSS	0	34
BSSS	4	36
BSSS	8	35
BSSS	12	36
BSSS	16	36
BSCB1	0	36
BSCB1	4	36
BSCB1	8	35
BSCB1	12	36
BSCB1	16	36

**TABLE S4 Switch error rates from computationally phasing hybrids simulated from derived inbred lines.**

Simulated 'Hybrid'	Derived Lines Used as Priors	Population	Possible Switches	Switch errors	Rate
B11xB67	all	BSSS	12195	41	0.003
B17xB44	all	BSSS	12230	58	0.005
B39xB37	all	BSSS	11485	129	0.011
B43xB69	all	BSSS	12266	26	0.002
B73xB72	all	BSSS	12043	47	0.004
B78xB94	all	BSSS	11658	51	0.004
B89xB84	all	BSSS	11517	53	0.005
B11xB67	cycle 0	BSSS	12195	45	0.004
B17xB44	cycle 0	BSSS	12230	54	0.004
B39xB37	cycle 0	BSSS	11485	126	0.011
BB43xB69	cycle 0	BSSS	12266	32	0.003
B42xB54	all	BSCB1	11260	38	0.003
B90xB97	all	BSCB1	6499	52	0.008
B91xB95	all	BSCB1	7830	56	0.007
B99xB97	all	BSCB1	6779	52	0.008
BB42xB54	cycle 0	BSCB1	11260	57	0.005

**TABLE S5 Evidence of minor contamination between cycles 4 and 8 in BSSS.**  
Polymorphic SNPs

SNPs polymorphic in cycle	But not cycle	BSSS	BSCB1
4	0	290	225
8	0	2242	319
12	0	1227	181
16	0	344	128
8	4	4499	1081
12	4	2328	669
16	4	551	477
12	8	1821	1405
16	8	178	997
16	12	542	666
0	Founders	1202	1707
4	Founders	822	885
8	Founders	1201	798
12	Founders	816	550
16	Founders	445	460

Heterozygosity of SNPs in cycle 8 but not cycle 4 of the BSSS.

	BSSS	BSCB1
Mean	0.052	0.052
Median	0.028	0.028