Editor's comments

The reviewers make a number of important points that should be considered in revision. Of particular note, Reviewer 1 highlights the importance of considering the possible effects of purging during the experiment, and this intersects with reviewer 2's comments about lack of details of the experiment and what was being selected, and what the outcome of the experiment was. Could purging, for example, have occurred over the course of the experiment, and did yield improve within lines in addition to increased yield in hybrids? Reviewer 2 also notes a lack of details on the simulation approach; I agree with this, and also thought this was perhaps confusing wording on page 14: "Selections were executed at random".

— *We thank the editor and reviewers for their time and helpful comments. We have addressed these and other comments specifically in the text below. Text in the manuscript modified in response to the reviewers has been highlighted blue.*

On the statistical side, candidate selected regions are picked out and effectively 1-sided tests are reported for the rare allele that has increased in frequency. While the authors are reasonably cautious about these cases anyway, at least some attempt to report an experiment-wide p value or FDR for these regions seems useful.

— *In the last paragraph of the results we report 1-sided tests for a few large haplotypes. We agree with the reviewer that an experiment-wide p-value might be useful if we were performing a formal genome scan, but because our purpose was simply to show that although the overall power to distinguish selection from drift may be low, there are nonetheless individual genomic regions that show patterns of fixation suggestive of selection.* Do we want to do an FDR for the p-values in Figure 3?

Do the results provide sufficient power to reject the overdominance model? How strong would the signal be likely to be?

— *While we are not making a formal hypothesis test of the overdominance model (which seems difficult to do as it would depend on the number of loci proposed to be overdominant), we believe the preponderance of the evidence supports a simple dominance model. While we can not rule out any overdominance – and have now added text citing one such example – the genome-wide pattern does not appear to us to support such a model as the major driver of heterosis for yield.*

One additional comment: the authors note that heterozygosity increases early in the experiment, and discuss various possible explanations for this. But couldn't directional selection on rare alleles be the driver of this, i.e. early on there will be a rise in allele frequencies, increasing diversity, and then subsequently a loss due to fixations?

— *This is a good point and we have added this to the text.*

Reviews

Reviewer 1: I really enjoyed reading this manuscript. I think this is a well done study that provides novel insight in to basic processes in breeding of maize. Many of these results have implications for other species as well. I have a number of minor comments below that I think should be addressed but these do not diminish my enthusiasm for the manuscript.

The abstract could use some editing. There are a number of references in this abstract and the abstract is not well focused on the specific findings of this study.

— *The abstract has been extensivley edited to better reflect our findings, and references in the abstract have been removed.*

In the introduction the authors describe the concept of heterotic groups and state that these lines maximize performance and heterosis when crossed. I would caution that this should be clarified to heterosis for yield rather than for all traits. Research by other groups (ie Flint-Garcia et al 2009) suggests that heterosis can vary for different traits and that the definition of heterotic groups is primarily based on yield heterosis, not heterosis for other traits.

— *This is a good point and we have added mention of this in the introduction.*

The introduction references several studies that looked at genome-wide population structure in breeding programs from the Ross-Ibarra group. I hesitate to even make this suggestion (do to potential flaws) but it might be appropriate to also include Jiao et al nat Genetics 2012 here as this study also looked at a large dataset of SNPs in modern maize.

— *While we agree with the reviewer's hesitation (see Lorenz and Hoegemeyer (2013)), we have added a citation to Jiao et al. (2012).*

I was quite intrigued by the difference in distance between cycles 4 and then 8-16. The authors suggest that this could be due to some differences in genetic maintenance of the populations. I am curious if there were any particular environmental influences during that time period that may have led to stronger selection or if there were any differences in agronomic practices that have occurred during the long-term experiment (ie irrigation, fertilization, plant density, etc). it would be helpful to know if any of these factors may have contributed.

— *The major difference is the change to 20 (instead of 10) lines at cycle 8. Machine harvest was introduced at cycle 6, but we have no reason to believe this would have dramatically changed diversity patterns. From our reading of the literature and consultation with breeders who have run the experiment we are not aware of any other significant changes.*

It was a bit surprising to me that there was no differentiation among the founder inbreds of the two populations. If PCA was run only on the founders (without any other lines) do you get separation of the two groups or not?

— *PCA of just the founder lines does not show differentiation corresponding to the two populations.* Justin, can you do this quick and dirty to confirm?

Figure 3 is quite useful and helps to show large physical regions that have lost heterozygosity. The authors discuss the physical-genetic difference. It would be useful to have a supplemental figure that shows these same values (cycle 16 heterozygosity on y-axis) versus genetic position on x-axis. I am curious if there are ever regions of multiple cM on chromosome arms that become fixed. This would be a nice complementary figure. These are provided for some chromosomes in 4B and 5B but would be useful to see genome wide.

— *Heterozygosity plotted on a genetic map was shown in the original supplemental figure 2 (which the reviewer pointed out we failed to reference). We have switched this figure for the original figure 2. As shown in the new figure 3, there are a few regions on chromosome arms that have been fixed and span multiple cM.* do we need to note this in the main text? is it an intesting observation, maybe implying rapid selection?

The large regions of fixation are quite intriguing. Is there any evidence for major flowering time or disease resistance QTL in these regions?

— Justin, any ideas? Do we do some formal correspondence with NAM results? Or say "geez this is hard because QTL are big and we're not sure we'd believe the results?"

I am somewhat baffled as to why the Fig S1 was put in supplemental materials. A central aspect of this study is this issue of genetic drift relative to selection and there are two pages of results text for this section. However, this useful figure is only in supplemental materials. I would strongly encourage moving this figure to the actual manuscript. If space is an issue then only use figu 4 or fig 5 as these are quite redundant with each other in terms of conceptual views. The legend to fig S1 could use additional detail. I am guessing that the black lines are 95% CI but this is not clear. Please explain the black lines.

— *Following the reviewer's suggestion we have moved this figure to the main text; the figure legend now explains the black lines. The old figures 4 and 5 have been combined into a single multi-panel plot.*

Figure S2 is not ever referenced in the text as far as I could find. The data table of the same results is referenced but I dont think the figure is.

— *Figure S2 and Figure 3 have been switched as described above; the supplmental figure is now referenced in the caption to the (new) Figure 3.*

The last paragraph of the results section assesses the fixation of particular haplotypes over time. I think it could be very useful to visualize this data for each of the large blocks. This could be done by using a graph of allele frequency over time

for each region. Right now these results are presented in text form for only two regions but it would be very interesting to see these results for all eight regions shown in table 1. What is the fixation at each cycle in the observed data and how often was this region fixed in simulations. I appreciate the authors point that these fixations do occur at low rates in simulations and since they are studying a genome-wide pattern a number of low frequency events are expected to occur by chance.

— *Some of these regions are visible in Figure 4. We have now included plots similar to figure 4 for all chromosomes in the supplement.* Justin is this doable?

The authors interpret the rate of heterozygosity in terms of selection and drift. I am assuming that they are interpreting selection to mean the selection for advantageous alleles in hybrid test combinations. However, I wonder about the inadvertent selection to maintain heterozygosity? In work from McMullen and Buckler it was argued that the gain in vigour due to heterozygosity of pericentromeric regions was larger than expected and that this may be due to increased ability to cross and maintain these plants. Was this concept included in the modeling of the data? In other words, while drift will tend to increase homozygosity in a population there may be counteracting forces of heterosis effects that will maintain heterozygosity in the same population. Just curious about this concept.

— *Yes, this is indeed a possibilitiy and one we mentioned in the text as a possible explanation for the fact that the observed data show higher heterozygosity than the simulations in early generations. We have now added citations to McMullen et al. (2009) and Gore et al. (2009) in this sentence to help highlight the point.*

The authors spend some time in the discussion regarding models of heterosis and extrapolation of these results to discuss selection. While this manuscript describes materials that are quite relevant for maize breeding programs I think it is important to note that these types of materials are not universal. In this case the synthetic populations are derived from inbred lines. This is important in terms of the nature of deleterious alleles that remain. Strongly deleterious alleles have likely been purged from the founders as each founder was an inbred line that had been selected for some level of vigor. This means that the remaining single-locus deleterious alleles will all be of quite minor effect and therefore selection is likely to act on alleles of minor effect rather than major effect deleterious alleles. I think that is would be useful to remind readers of this attribute of the population. My guess is that if this experiment was done with a maize (or teosinte) population that had never been passed through an inbred "bottleneck" that there was be some alleles of major effect that would be purged quickly and that the role of drift and selection might be slightly different.

— *This is an excellent point. We currently have experiments underway to test precisely this point using selfed progeny of both landrace and teosinte populations. We have added discussion of this point to the text.*

Reviewer 2: Girke et al. in "The genomic impacts of drift and selection for hybrid performance in maize" report a retrospective examination of genomic patterns of diversity in a maize experimental population. The data set consists of genotyping of 38,000 SNPs genotyped in individuals from various phases of the selection experiment. While the manuscript consists primarily of population genetic analysis, the authors attempted to better integrate a popular maize genetic map with the current reference genome.

The strength of the manuscript is that the authors design simulations that closely parallel the experimental population design to determine exceptions for change in heterozygsosity and allele frequency differentiation. The simulations are very useful in that clearly establish a major role of genetic drift, owing in large part to small effective population size, in the differentiation of experimental populations. There are a couple of points of clarity needed with regard to the simulations. First, it isn't clear how recombination was handled within the simulation, other than an indication that "all recombinations events in the simulations were carried out in R with the hyped software package" (p 11). Presumably some number (0 - 2?) recombination events per arm generation was included in the simulation, with probability of recombination at a location determine by genetic map length. Something like this needs to be stated.

— *We have added three sentences to the methods describing additional details of the recombination model. More information about the specific functions as well as detailed code is available at the url link to the hypred R package.*

Also related to the simulations, there are multiple references in the text to empirical results as "deviation from simulation" (e.g., p 15) when I suspect what the authors wish to say is that the results differ from neutral expectations based on simulations.

— *We have removed the phrase "deviation from simulation" throughout, replacing it with phrases mentioning differences from neutral simulations.*

The other major weakness of the paper is that relatively little background on the experimental populations is provided. There is a long list of citations of prior work on this experimental population, but little exposition.

– Justin? Jode?

A better presentation of the specific target of selection is needed (I infer that selection is on improved yield through increased hybrid performance). Ultimately in the Discussion, there is additional discussion of the phenotypes, but the presentation is minimal (again pointing to multiple citations) and provides little insight into the nature of the selection observed. Better discussion of phenotypes potentially targeted by selection is essential if the manuscript is to end, as it does, with a call for further dissection of targets of selection in this population.

– Jode? Justin?

A third major issue that manuscript seems to be not at all informed by multiple experimental evolution studies conducted on Drosophila and other model systems. This contributes to a couple of awkward citations of personal communication (p 17 and 19), where ideas about loss of diversity and need for replication of experimental populations have been reported in the primary literature. See Drosophila papers by Tom Turner and Anthony Long and colleagues.

– *This was an oversight on our part; we have now incorporated citation to some of this literature and hopefully clarified the discussion.*

Finally, the Discussion seems a bit overly long and could potentially be more tightly written.

– *We have endeavored to streamline the discussion where possible while incorporating additional citations (as above) and discussion of load as mentioend by reviewer 1.*

Minor points: - Careful copy editing is needed. Heterozygsoity in the figures and figure captions is referred to variously as "H" with and without italics and also as heterozygosity spelled out, sometimes within the same figure or figure caption. Simiarily, Fst needs to be presented in a consistent manner.

– *We have changed text and labels to improve consistency in using heterozygosity vs "H".*

Also edit citations for consistent capitalization and in=text citations for consistent spacing.

– *We have moved the ms from word to latex; we believe this has fixed many of these issues. We have also corrected bibtex entries in several places to improve consistency of capitalization.*

Also: On page 7, line 10: A sentence should not begin with a number.

– *Changed.*

Page 10, line 14: Citation format is incorrect.

– *Fixed.*

Page 11, line 16: Number of SNPs passing filters should be reported in Materials and Methods

– *This is mentioned in the very beginning of the results, where we felt it would be easier to find for readers that may skim the methods rapidly.*

Figure 2 caption: What is H combined and how was it calculated?

– *This is simply total heterozygosity pooling across all plants. Text explaining this has been added to the figure legend.*

Figure 4 and 5: Please include the simulated data (mean + variance) so the pattern is more apparent.

— *We were unable to devise a means of adding simulated data to the plot in the new Figure 4 without substantially obfuscating the main points of the figure. We have however marked individual regions which have lower heterozygosity than seen in our simulations.*

Figures could be better labeled, for example, simply label panels A & B of Figure 4 as "Genetic Map and "Physical Map" and the cycles as "Cycle."

— *We have tried to improve figure labelling throughout. In places instead of adding labels (e.g. Fig 4 mentioned by the reviewer) we have added additional explanatory text to the caption in hopes of clarifying.*

When discussing the plant material that was used in the analysis (page 7, line 13), two individuals that appeared to be mislabeled had their labels switched. However, later in the paragraph, another labeling issue was discovered, and these individuals were removed from the analyses. What is the rationale behind the differential treatment for these two labeling issues?

— *Four plants were found that appeared noticeably out of place on the PCA. For two of these we have insufficient evidence to know whether they were switched, mislabelled, contamination, etc. These were removed from the analysis. The remaining two were clearly a label switch in which "plant 31" from cycle 4 was labelled as cycle 8 and "plant 31" from cycle 8 was labelled as cycle 4. Because the labels corresponded to the position swap seen in the PCA, we were confident in correcting the labels for these plants and kept them in the analysis.*

For the integration of the physical and genetic maps, there are two concerns: 1) Do the SNPs on the Maize 50K SNP chip already have reference positions? How were the SNP contextual sequences mapped back to the reference genome? If it was a BLAST-type tool, was there an e-value threshold that was used to remove low-confidence positions?

— *Details of the chip and validation of the markers are published in Ganal et al. (2011) and do not seem appropriate to repeat here. All SNPs have an assigned position on the reference genome, but our inspection of the genetic map led us to move some of these markers as detailed in the text.*

2) When integrating the genetic and physical maps, how many SNPs occurred in mis-mapped blocks?

— Justin? *We have added this number to the text.*

When the cycles of selection shift from 10 individuals to 20 individuals, does this really constitute a doubling in effective population size?

— *Yes. While a multi-generation Ne (e.g. harmonic mean of N) will not immediately double, the single-generation Ne has effectively doubled.*

A subheading on heterosis would help better organize the Discussion.

— *Done*

The Hudson et al. 1992 measure of Fst that the authors report is a sequence-based measure based on average pairwise diversity within and between loci. How were the "loci" defined here?

— *Hudson et al. 1992 Fst is based on pairwise differences between sequences. We can thus calculate this from SNPs in a window (or individual SNPs) without knowing anything about the number of bp.*

On page 6, what is the difference in a "found inbred" and a "founder line?" I suspect this distinction will be lost on many readers.

— *We have deleted "line" here in order to clarify. No distinction was intended.*

On page 15."consistent with strong genetic drift imposed by selection practices," needs to be reworded.

— *"selection practices" has been replaced with "experimental design" for clarity*

In summary, this is an interesting manuscript that would benefit from improved clarity of presentation.

Literature Cited

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.

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.

Lorenz, A., and T. Hoegemeyer, 2013 The phylogenetic relationships of us maize germplasm. Nature genetics 45: 844–845.

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.