Review

October 15th, 2014

General Comments

Pan *et al.* constructed 12 segregating populations, analyzed the genetic recombination landscape of these populations and correlated the recombination hotspots with various genomic features. The authors present nice data sets that could be useful for genetic studies. However, I am not convinced by the interpretation of the data, especially the methods used for correlation studies. and more work need to be done to either clarify their methods and distinguish their results from those of Bauer *et al.*[1]. Below I include comments on specific aspects of the paper; I hope they may be useful in revising the manuscript. There are a number of places with typo (e.g. `non-colineraity` on page 9 should be `non-colinearity`) or grammatical error (e.g.), careful revision of the text with this in mind would improve legibility. Please do not hesitate to contact me if you need help to clarify questions or comments.

[1] Bauer E, Falque M, Walter H, Bauland C, Camisan C, Campo L, Meyer N, Ranc N, Rincent R, Schipprack W: Intraspecific variation of recombination rate in maize. Genome Biol 2013, 14:R103.

Background

- Bulleted list of concerns
- Only concerns that are sufficiently important that they should prevent publication of the paper

Results

- The number of recombination bins and the length of genetic maps were calculated using the same set of SNPs to measure recombination events in different ways, it is obvious to me that the two values should be highly correlated. I do not think it is necessary to present this correlation value unless you have a point to make.
- I have concerns about the claim that you have improved the reference genome. Because B73 is not a founder line for most of the 12 populations, gaven the large number of genomic variations of maize [ˆc2], is it possible that the non-colinearity regions you observed are population specific inversions but not reference genome errors? Or caused by genotyping errors or mis-placement on the genetic map. Your have to rule out these possibilities for the purpose of reference genome improvement.
- For a better comparison of recombination events in 12 popualtions, x-axis and band width of the histograms in **Figure S2** should keep

the same. And you should backup this statement `As expected, the longer the chomosome is the more recombination events occur` or drop it.

- It may be not statistically legitimate to compare genomic features of hotspot regions with randomly selected genomic regions given SNPs on the SNP50 array may enriched in genic regions. And most of the *P* values lower than 2.2e-16 seem worrisome. Instead of choosing random genomic regions, recombinant coldspot regions of same size and SNP density should be selected for testing as described in {Myers *et al.*}[^c3]. In addition, SNP array suffers from some degree of ascertainment bias, I do not know how much it will affect this analysis, but address this concern here or in discussion seems warranted.
- You proposed a hypothesis about the function of genes with intragenic recombination and conducted GO term enrichment test. However, it seems like you neither rejected nor validated your hypothesis. Therefore, the intrepretation may belong to discussion.
- Again, the relationship of hotspot and gene expression should also using regions of same size and SNP density as control rather than random selected region.
- A citation should be given for the "genome-wide significance level" for determining the threshold of GWAS. Or why not just use well accepted `FDR` or `bonferroni` method to control multiple test problem?
- In the text, you should check whether **Fig. 4A** pointed to the right figure.
- In the part of result section, the header said `intragenic recombination is significantly associated with gene expression and phenotypic variation in maize`, however, I could not find any evidence in the text, except some case studies, to backup this conclusion if I did not miss anything.

Discussion

- You may want to talk more about how to use the 65 recombination hotspots in marker assisted selection.
- I am confused about this [2], are the beginning and end part of the genomic region per se.

[2] Recombination is more likely to occur at the beginning and end of the genomic elements and no the genomic element regions `per se`, ...

Methods

- Bulleted list of concerns
- Even if the authors didn't fix these the papers would be ok, but they might improve the paper.

Tables and Figures

- Figure 3A: legend of # *Genes with >= 1 SNP* should be # *Genes with <= 1 SNP*. And check the label of y-axis of Figure 3C.

  Pick one of: * Major revisions