Comments:

In this paper Renny-Byfield and co-workers describe an analysis of an ancient polyploid event in the lineage leading to cotton. This analysis parallels earlier work examining gene loss rates, expression, and transposon content among duplicated regions of various plant genomes and is probably closest to previous analyses of of maize and brassica. Intriguingly the authors find no evidence to support a model explaining biased gene loss based on spreading of transposon silencing from siRNA target elements inserted near the less expressed copy of differentially expressed gene pairs. I think this is an interesting paper, and while I do not think the evidence they present makes a conclusive case for their model or represents a mortal wound to the model previously put forward, I still think this manuscript is a significant contribution to the field. However, there are a few issues the authors should address prior to publication.

Most critically: it is not clear to me from reading the manuscript and methods whether the data presented in figures 2-5 represent analyses only of conserved gene pairs, of all genes within LF and MF regions with a syntenic ortholog in chocolate, or all genes within the respective regions, syntenic or not.

We examined only those genes that are syntenic paralogs within LF and MF, thus only genes that have a paralog to compare to are considered, not all the genes in each fraction. So, in short, and like the reviewer later suggests we took the most conservative approach. I’ve modified the results and methods to reflect this.

It is also not clear what the total number of genes included in each of these analyses are.

This should be clearer now.

Both of these pieces of information are critical to interpreting the results of this study. Obviously examining only conserved syntenic genes between regions represents the most controlled analysis. However if there are too few of these to be informative, restricting the analysis to only genes with a syntenic ortholog in chocolate would still be vastly superior to an "all genes" approach, as multiple lines of evidence suggest non-syntenic recently inserted genes belong in a class by themselves.

Because so many comparisons are made between the analyses conducted here and the maize (polyploidy ~10 mya) and brassica (polyploidy ~20 mya) events, I'm surprised the authors don't talk more about the fact that the event they are focused on is much more ancient than either previous event at 60 mya.

I have added a few words dotted throughout the MS to draw more attention to this aspect of the data.

Particularly when dealing with transposons, both sequence divergence and sequence deletion will likely make ancient transposon insertions much more difficult to identify in cotton than in the two species previously studied, which are already older than ideal for identifying transposon insertions which date back to the original polyploid. I don't think this is a fatal flaw with the analysis, but it's certainly a potential explanation for some of the differences between their observations and those in maize and brassica and should be addressed.

This is a difficult one, I wasn’t sure where to add this in the MS. I’ve had a go but please feel free to modify.

In the results section the authors discuss an observation that LF regions contain higher surrounding local GC content than homoeologous MF regions of the cotton genome. However I cannot tell from the methods section whether they took into account gene density. In other words, do the differences they observe reflect a difference in the proportion of sequence in a genome region which is exonic (tends to be higher GC) or different in the GC content of non-coding DNA between the homoeologous regions? The first interpretation would make this result predictable and unsurprising based upon the expected differential gene loss and transposon density between these regions. The second intepretation would make this an intriguing result and might indicate something about the fate of regulatory sequences (which also tend to have higher GC content than other noncoding DNA) across subgenomes.

We have a problem here, there are likely to be quite a few overlaps between our 5kb windows and nearby genes, this could very well explain the difference. We may have to re-think this one. I have moved this figure to the Supp Info. We need to reduce the number of figures by one anyways andthis seemed like a good one to cut from the main text.

In the discussion section the authors put forward a model where MF regions have lower levels of recombination as a result of higher transposon density and therefore are less able to purge mildly deleterious mutations than on the more recombinationally active LF regions resulting in a greater accumulation of gene deletions. This appears to be similar to the differences between pericentromeric regions and euchromatic chromosome arms. My two concerns are 1) lower recombination rates tend to be correlated with lower rates of sequence deletion through intrastrand nonhomologous recombination which would suggest there would be many fewer gene deletions in MF regions if the authors' predictions about rates of recombination are accurate

I’m not sure what the reviewer means in this case. There are quite a few examples where genome contraction and sequence removal are correlated with recombination rate (we site a few and there are more). Where as the reviewer seems to suggest it’s the other way round. Either way it’s not the rate of sequence loss we are arguing for but the rate at which they are *fixed* that’s important. Presumably they are easier to fix in regions of low recombination, at least that seems reasonable to me. I think we can argue this after the next round of review, but we need to be careful.

2) the authors do not appear to have made any attempt to test their prediction by comparing recombination rates between duplicated regions in the cotton genome. I can imagine plenty of reasons this might be difficult to perform. Insufficient availability of high resolution genetic maps being the simplest. However because it is apparently an easy hypothesis to test if the authors are unable to test it they should mention why that is.

This is worth mentioning. I have added a sentence in the MS.