Hi Jeff,

I added a few comments below to your questions.

Hi Andy,

Many thanks for going through the paper and contributing your thoughts.  Jinxing and I met today to go over them, and our thoughts are as follows:

GERP-Score:  I think the circular shuffling has essentially the same effect of comparing GERP score haplotypes to haplotypes alone, doesn't it?  We previously also tried the SNPs and SNPs w/ GERP comparison, but fitting all the SNPs does a darn good job and actually adding GERP doesn't help much.  Perhaps this is because when doing all the SNPs (or even all the SNPs with GERP ≤0) you have enough stuff in LD with GERP SNPs that you gain little by adding GERP values?

**I think you are correct. I was directly thinking about the SNPs w/ SNPs and GERP step you mention above when I made this comment. I think the work that Jinliang did in his GCA/SCA step backs up your hypothesis as well. No big or significant change in likelihood between the models with just the SCA term, just the GERP term, and the SCA + GERP term indicates that you are indeed capturing the same dominance variation with both scores.**

Circular Shuffling: Good call, we'll up this to 50Mb

-Jinliang: I actually did the analysis with circular re-shuffling of 50k SNPs (> 100Mb on average).

BayesC: This is a good idea, but we may wait to play with it.  The other part of the paper that isn't in the current manuscript is a GWAS that my former postdoc Sofiane did.  The idea was to look for enrichment in GWAS hits for SNPs with high GERP scores. I think (correct me if I'm wrong) that this is getting at the same thing your BayesC suggestion would do. Isn't it? In retrospect I like the BayesC approach better, but if Sofiane has finished his part I think we could just include that instead.  Thoughts?

**The GWAS approach will give you much better results for what you are looking for and I think will answer some of the questions concerning SNP vs. GERP vs. SNP + GERP above. The GWAS approach would be much better and would give you better results since you can control more for population structure. One problem with BayesABC methods is there is no control for population structure. A SNP can be found in several iteration for many reason, many of them due to population structure. This will be really important for this diallele since several of the inbreds have very different genetic backgrounds, even within heterotic groups, as well as very different flowering.**

GCA/SCA: Some of this gets away I think from the main focus of the paper, on deleterious enrichment. We have already shown in simple models that once we include SCA, GERP scores don't add much.  I suspect this is because SCA is essentially telling you about complementation already.  So what I was thinking was to do a simple model to ask if GERP scores help explain SCA more than relatedness alone (SCA ~ GERP + relatedness). Is this useful?  I'm a bit afraid it won't show an effect of GERP, but maybe genome-wide relatedness isn't saying much about pericentromere relatedness, which is what I think the GERP score (summed absolute difference in GERP score between two parents) is telling us.

**I totally agree that no change including the SCA + GERP (or GERP + SCA) is explaining the same variability. I don’t think we would need to do much more than this to prove the point than Jinliang has done in the ANOVA/Mixed model steps.  It’s a pretty good way to infer that they are explaining the same thing.**

**After thinking through my comments, I felt a bit bad about the comments concerning including SCA and GCA in the phenotypic analysis since 1) you had already run it and all other analyses depend on this data and 2) in my experience, the results will not change too much. I think the main goal of all of the SCA and GCA work is to just show that the GERP score and the SCA are explaining the same non-additive variance in the model.**

**I hope this helps! Let me know if you need anything from me!**

**Andy**

We'll keep working on the draft and send you an update.  Please let me know if you have other thoughts on the above, or if I'm way off base (as is often the case).  Thanks again for your thoughts on this!

Cheers,

Jeff