

## New notes/approaches:

Several challenges have presented themselves in using the pedigree approach to detect selection through time - I'll list these here, and also discuss present some possibilities for how to approach each challenge.

### Challenge One: the issue of having certain inbreds differentially/non-randomly present in the pedigree in calling regions of IBD

- Beagle is not calling certain regions IBD (B73 problem) - this problem is not a problem with IBS/kinship as carried out in Tassel.
- The solution is to build a reference panel population first with HapMap3 BKN and CML lines.
- Will need to adjust mapping positions accordingly so that the ref. panel pop and the GBS pop match in terms of position - there is a way of doing this in Beagle
- Test out with initially small subset of individuals.

### Challenge Two: genome wide IBS vs. IBS at a given SNP (binary matrix)

- This may be a viable solution but need to learn NUMPY or pandas to produce matrices (working on it) that won't take up as much space.
- Need to think more about stats here and what is being computed here too.

### Challenge Three: Transmission disequilibrium test - this is what Justin called it

- This is more along the lines of what Graham Coop suggested in tracking reproductive success down the pedigree. In my view, the challenge would be to map a matrix of traits that are genotypically correlated to each other to particular SNPs or IBD haplotype segments - in my head it looks similar to the multivariate breeder's equation, but I haven't thought it through enough yet. **Addendum** This could also be a single locus scan - much more direct, but again not polygenic.

$$\Delta z = G\beta$$

Where  $\Delta z$  is the change in any trait, including the change in the trait of fitness  $\omega$  from one generation to the next.

$$\Delta\omega = G\beta$$

Since we can estimate  $\Delta\omega$  down the pedigree from one generation to the next, and we can combine different traits together in a G matrix (a genotypic correlation matrix between any two traits and we can do this for many traits), we can estimate the selection gradient  $\beta$  on any given set of traits. True, this does not get at polygenic selection. It gets at something different.

$$\Delta\omega * G^T = \beta$$

Although one interesting prediction to test here is that the directional selection gradient ought to get shallower through time - as  $N_e$  drops (due to alleles being chucked), and the genetic variance getting narrower through time, if selection is not relaxed.

## More quant. gen or pop gen (both really)

Looking at  $V_A$  down the pedigree on a per generational basis, where  $V_A$  is a function of allele frequencies across loci  $L$  (each of which drift stochastically in  $L$ -dimensional space but some are more correlated with each other due to LD - stolen from Coop's NOTES):

$$V_A = \sum_{l=1}^L \alpha_l^2 2p_l(1 - p_l)$$

Here,  $\alpha$  are the allelic effects which could be estimated from GWAS for a series of loci  $L$  that affect any number of traits. We expect decay in  $V_A$  but I bet you won't see it. The idea is to explain why you still get a considerable response to selection (almost certainly going to show up as  $V_A$ ) in most traits in spite of alleles being regularly chucked (we think) from the population (except alleles probably aren't being chucked very efficiently at all, due to the brilliance that is the design of Reciprocal recurrent selection experiments - see Agrawal 2006 for BOX 1 for simple explanation on how variance is maintained in the face of decreasing heterozygosity, with an example of linked loci).

**NOTE**  $V_A$  is likely not really all  $V_A$  were you to look down the pedigree (it's probably closer conceptually to " $V_I$ "), as you're likely to get different SNPs and GWAS hits through time (changing genetic landscape per each reshuffling cycle of inbreeding in an RRS or RRS propagation - Tim does not like it when I say architecture). Nonetheless, at any given time step what we want to show is that  $V_A$  actually stays overall quite high or consistent through time, and likely so does  $N_e$  (or one type of  $N_e$ ).  $N_e$  could be estimated within a line, within a heterotic group, and among heterotic groups through time in a hierarchical manner (Whitlock 1999 -  $Q_{st}$   $F_{st}$  relationship).

The other cool thing is that we could actually show that either the  $\alpha$  s for a given trait or the  $L$  s underlying a given trait change through time - but,  $V_A$  and  $N_e$  is still quite high overall, and or even add a mutational input term above to show that  $V_A$  is increasing (that's kind of going a bit crazy - but we do have population mutation rates of TEs).

===== ## Challenge Four: Mini-pedigrees

We can do allele dropping/TDT on lots of minipedigrees instead of one big one. The problem arises when we have a minipedigree of size  $n=2$ . Based on relatedness we can likely identify things that are either parent/offspring or full-sib, and based on distribution of size of IBD blocks we may be able to distinguish between those two.

Let's say we have  $n=2$  and know they are a parent-offspring pair. We have no additional information (year etc.) to help us add direction. Is this useful? Basically we have two outcomes at a locus:

1. They share an allele. I think this is uninteresting unless the unknown parent has  $> 1$  copy of the alternate allele (otherwise we can't rule out that both parents are homozygous for the favored allele, in which case we expect with probability 1 that the offspring would get the allele).
2. They don't share an allele. This is also uninteresting for the opposite reason. If favored allele is A, and the pair is AA and aa, then this could be an example of the A allele being successfully passed down (if AA is offspring and unknown parent is AA or Aa) or of the allele not being passed down (if aa is offspring and unknown parent is aa or Aa).

I think maybe we have to drop these?