Detecting Signatures of Selection using the Switchgrass GWAS Panel and Genomic Data

**Idea:** select for high and low biomass populations in switchgrass using seed from the GWAS panel plantings. Use replicated populations to determine the loci under selection.

**Evolutionary background:** Drift never changes the frequencies of alleles in replicated populations, whereas selection does. So you can use replicated populations, take the average of allele frequency, and you will get the ancestral allele frequency back out, if the locus is not under selection. Using replicated populations is thus an excellent trick for differentiating between drift and selection.

Let's use the huge size of the GWAS panel, combining material that has never been grown together, and the fact that switchgrass is an obligate outcrosser to look for new (or old) allelic combinations that confer high (and low) biomass in specific environments.

(Parentheses indicate smaller details to consider).

**Experimental Design:**

1. Minimize inbreeding
   1. Within family selection to minimize LD buildup (stand planting and collection of seed/DNA from multiple individuals)
   2. Maximize number of families (more stands is better)
2. Bidirectional selection long term - for and against trait of interest (not sure how to accomplish this… choose the smallest living plants in the field also? This doesn't deal with lethal allelic combinations… or that smaller plants might be within-season volunteers)
3. Dense genotyping or sequencing of individuals (HudsonAlpha)
4. Control demography (GWAS panel seed does this)
5. Weir Fisher Theta statistic for analyzing selected sites in the data - worked the best of any statistic in similar selection studies.

**Methods:**

1. Collect seed from the GWAS plantings at (three?) sites. Mix the seed within sites. For example, at PKLE:
2. Sequence (ten?) pooled subsets of seed (germinated seed?) to determine starting allelic frequencies that will be planted.
3. Plant twenty stand plantings of subsets of seed using broadcast seeding (or whatever).
4. (After n seasons) In ten stands, sequence the individuals with the largest biomass. In the other ten, sequence the individuals with the smallest biomass.
5. Potentially select seed from these stands (the biggest and smallest plants, but several of them/at the family level) and broadcast them in twenty new stands next year (one for each old stand). Repeat the sequencing of largest and smallest individuals to determine how allele frequencies have shifted.
6. Do this at three distinct locations where the diversity panel is being grown. Test (using Weir Fisher Theta statistic) which allele frequencies are under selection at each site, and how these allele frequencies differ between sites.

***David Lowry*** says they have infinite land at MSU, so that's one possible location to do this at. We probably have more space at Pickle, even though that's an octoploid site (but who really cares about ploidy… if those are the biggest plants, they're the biggest…)

**Sequencing Experimental Design:**

First generation:

10 seed pools (both germinable and non-germinable seed, potentially)

(10 seed pools with germinable seed)

10 season n high biomass pools

10 season n low biomass pools

(Second generation?)

(10 seed pools (both germinable and non-germinable seed, potentially))

(10 seed pools with germinable seed)

(10 season n+k high biomass pools)

(10 season n+k low biomass pools)

At three locations.

Minimum 30 pools from one location that would need high depth sequencing per generation. For one generation, ideally 120 pools. For two generations, up to 240 pools. Not super difficult to do stand establishment, growout, and tissue collection, but then sequence later if the initial results are promising. Identifying the largest/smallest plants in a stand might be difficult.

Weir Fisher Theta statistic to test for differences in allele frequency between:

1. Germinable and complete seed pools (selection for non-lethal allele combinations)
2. High biomass season n and germinable and/or complete seed pools (selection for germination and high biomass)
3. Low biomass season n and germinable and/or complete seed pools (selection for germination and low biomass)

**Product:** seed from strong biomass stands adapted to the region to sell to farmers.