**Re-Run Planning**

**Other Work**

What should/needs to be done before starting re-runs?

* Score McKenzie GTs (12 days)
* Chum Report Revisions (< 2 days)
* Halfpounder first revisions (10 days?)
* Periodic review of faculty prep and meet (1 day)
* UWR USACE meeting (3 days)
* Vacation days (2-3 days?) (I plan to take a couple long weekends in July/ August)

This puts me at ~ 6 work weeks from today (or right around August 1st), without doing any substantial OC Chinook work.

From here, I plan to spend the first 2-3 weeks of August prepping McKenzie analysis - get all code working and first draft results (computational notebook, not text for report) from the non-rerun dataset. Then go do re-runs after August 20th (estimated 3 - 5 days , but plan for 10 days see below). Return to Portland, score new GTs and update genotype table, then re-render the notebook with updated data and start drafting the report (mid September).

Important note to self: Getting the analysis drafted and running will allow me to identify any issues with metadata that are best resolved in the lab. Be sure I’m confident that I don’t need any more information before setting off to Newport.

**Re-run timing estimates using only plates 5-8:**

These estimates assume that individuals need to be re-run at ALL loci. Planning to re-run individuals only at panels where they are missing genotypes will potentially reduce the lab work time, but increase the planning/prep time before getting to the lab.

*Number of re-runs* ***­****for >7 loci*

If we use my scoring from plates 5-8 as a benchmark, and decide to only conduct re-runs for individuals that would other be filtered from the analysis due to missingness (scored at 7 or fewer loci – the approach used by Dave previously), we will need to conduct re-runs at about 1% of all samples. This is about 28 individuals. Alternatively, using Dave Jacobsen’s more conservative scoring for markers 212, 215 and 311, we would need to re-run about 57 individuals.

*Number of re-runs for >=10 loci*

Using a much higher standard, if we use my scoring from plates 5-8 as a benchmark, and decide to conduct re-runs for any individual with at least 2 missing loci, we will need to conduct re-runs at about 9% of all samples, or ~260 individuals. Going by Dave’s more conservative scoring, 15% or 405 samples.

*Carcass Samples*

Using plates 5-8 (of 28) is conservative because I got better in the lab over time, but it is anti-conservative because plates 5-8 contain no carcass samples which are expected to perform worse. In the full dataset there are ~116 carcass samples. I haven’t done anything quantitative, but glancing at the genotypes of carcass samples suggest they did not perform substantially worse than other samples, about 6% of individual genotypes are unscoreable.

*Timing (for re-running 14% of samples)*

In both cases above (either DPJ and DID scoring standard, and either >1 missing locus, or >3 missing loci), we have less than or close to the number of samples in a 4 quadrant ABI run (sometimes less than even a single plate). Let’s be conservative and assume will use 1 full 384 quadrant ABI run (14% of samples need to be re-run). Let’s also assume we have all information ready to go in advance and can hit the ground running. Being very conservative, in the past, prepping a 4 quadrant ABI run took about 3 days: 1 day to aliquot DNA and run PCRs, 1 day to coload amplified DNA and prep the ABI run, 1 day to wait for results. This 3 day estimates fits with Sandra’s prediction of re-runs for this project requiring <4 days of work.

However, if we want to start from fresh DNA, that will add about 2 days for a 4 quadrant ABI run (1 LONG day to cherry pick, plate and digest + a few hours to extract once the overnight digest is complete, let’s be conservative and call this 2 full days). All individuals flagged as needing re-runs only failed at a subset, not all, loci, so I suspect fresh extractions won’t improve genotype quality, but let’s plan for this anyway in case this isn’t the case with other samples (ABI plates 5-8 correspond to DNA extraction plates 5-8, so we’re basically working with a sample size of 4 here, other DNA extractions may have been worse)

So the total conservative estimate is 5 workdays, with one (cherry picking and plating DNA) running long.

In the past I’ve been able to work 10 days straight pretty well. Let’s assume this is the case and plan for 10 work days dedicated to re-runs. This allows a full 2 fold more time than estimated in the assumed worst case for re-runs.

I also need to budget a day to archive DNA and clean up.

**Estimates with all samples scored**

Let’s re-do the timing estimate once all samples are scored and we finalize the decision on which samples need to be re-run (e.g. >1 or >3 missing loci)