Several questions need to be resolved before embarking on the final analysis for the McKenzie River Genetic Pedigree Project.

Flagged some (in purple) that would benefit from discussion with Kathleen.

Report Plan

* How do we want to collaborate to write the report?
  + Something similar to NSNT: start from previous report and update with new values
  + DID provides methods, results, outlines some of the discussion. KGO provides drafts intro and discussion
  + DID writes new draft from scratch, using the proposal as framework.
  + Go step by step, write methods, results first. Then talk
  + Compare formatting to NSNT, make some choices in advance, talk to KGO about it. It does matter to the corps.
    - No TOC
    - For example title page
    - Get a template running and revise it as a way to work on it together
    - Fonts consistent across report and figures

Analysis Choices

1. What changes to the assignment approach are we comfortable with for the 2016-2020 offspring / 2011 – 2015 parents?
   1. Unlike the previous NSNT assignment approach, the previous MCKR assignment approach is more reproducible. I was able to closely reproduce the assignments made by Nick in one offspring year (97.5% agreement).
   2. I also caught some errors in the code. The approach described in the reports and manuscripts does not match what was actually done, and the pedigree could probably be improved by bringing the code in line with the stated methods.
   3. In addition to the relatively minor code errors, there are several shortcomings in the previous MCKR approach. Most likely bias CRR/TLF downward a small amount. Some result in impossible parent offspring trios and have unknown net effects on CRR/TLF. Others simply favor less likely parents over more likely parents.
2. How to present 2010 parent year results?
   1. 2010 did not have the full age structure included in previous reports (no year 6 returns yet)? Should we re-evaluate the whole year (i.e. assign 3, 4, 5 and 6 year offspring with the new approach), or just attempt to assign 6 year old offspring and revise previous TLF/CRR upward.
   2. Just look to see, play by ear about including it into the report. E.g if only a few fish assign, then don’t bother
3. Is there any interest in extending new assignment (if taken) approach beyond the 2016-2020 offspring / 2011 – 2015 parents?
   1. If we need to make changes to pedigree approach for the 2011-2015 parent years to fix previous errors or improve the pedigree, should we apply the approach to ALL previous data (2007 – 2015 parent years). This would provide the benefit of being able to identify longer term trends and increase statistical power in the modeling by increasing sample size (also see points 5 and 6 below). Software version changes mean that results aren’t comparable across the reports, regardless of what approach we take to assign parentage in 2011 – 2015. Is this sufficient justification to re-analyze data from cohorts that are already reported on in previous reports/manuscripts without casting doubt on the previous findings?
   2. Assign 2007-2020, only present results 2011-2020
4. LSDR/”Recycling”
   1. I’d need to look at the corrected metadata to be totally sure of the details, but in general USACE no longer does LSDR. NOR fish that arrive at cougar are tagged and released downstream or back in the tailrace, regardless of date. They are only reintroduced above the dam if they manage to enter the trap a second time. This is a lot of handling stress and they require they fish to double a portion of their upstream migration. Is this worth it, or should a LSDR approach be used again? Should we think about the effects of the fire on downstream habitat when we frame the discussion?
   2. The stated objective in the proposal is a Chi squared test comparing assignment rate before or after September 1st? Can we go a little further and plot the assignment rate though time to estimate cutoff dates for LSDR to produce different rates of sampling immigrants and excluding F1s?
5. Grandparentage / Adfluvials / Precocial Males
   1. There is no plan in the proposal to address unsampled parents above the dam (adfluvials and precocial males) through grandparentage assignment as in previous reports. Is this right?
   2. What to do with 2014 precocial males? I have only one sample of precocial males from above the dam. Should we include these as parents in final results? If so, how? There are no other precocial male samples in other years, but there are some jacks from the Cougar Trap that were outplanted.
   3. If we choose to examine grandparentage to identify unsampled parents, should we infer a new pedigree using all the data (2007 – 2020)
   4. Total flexibility to add objectives
   5. Time and what’s manageable to make a decision here. Only add if there is sufficient time
   6. 2014: include them in the pedigree, only include in the report they contribute
   7. Email Ryan about 2014 precocials, ask if he knows why the LSDR approach was abandoned (it’s Kathleen who wants to know).
6. HOR/NOR Fitness comparison
   1. Methods for comparing HOR/NOR fitness are not described in proposal. Should I take the same approach in 2016 report
      1. Include origin and origin\*sex interaction in individual year GLMs, present estimated effect if significant.
   2. Should we also include origin as predictor in a multi-year GLMM?
   3. What about comparing fitness of F1s (NOR offspring of HORxHOR) to descendants of NORxHOR and NORxNOR for the subset of years where we can conduct grandparentage assignment? Does this motivate inferring a pedigree across all years (2007 – 2020) using a consistent approach? Hold off until we know we have time and will do grandparentage, do it talk about it later.

Summary of Previous Assignment Issues

* Discrepancies between methods and actual approach
  + The methods state best parentage in CERVUS is determined by LOD. In the code, parentage is selected one parent at a time on the basis of parent offspring duo mismatches, with LOD used only to break ties (but also see below).
  + When two candidate parents have the same number of mismatches, Sard’s results sometimes choose the less likely parent (lower LOD score), this is different from the stated methods. The offending piece of code settles ties using the LOD for the parent offspring trio instead of the parent in question’s LOD score, despite the fact the code chooses each parent separately. This is a minor problem and results in about 1% of assignments different from what would be produced if the code followed the methods.
  + The methods state that disagreements between CERVUS and COLONY assignments (other than for non-assignments) are resolved based on the number of mismatches. However, in the case where both parents disagree, the code above defaults to COLONY so long as both pedigrees made full trio assignments, even if there were more mismatches.
  + The ms states that offspring that assigned to both parents in a given year took precedence over assignments to a single parent in a different year. This doesn’t actually happen in the code. If an impossible parent pair (different years) is selected in both CERVUS and COLONY. The offspring is not assigned a parent in the final consensus pedigree. At no point do we go back and accept the next best parentage.
* Other shortcomings of previous assignment approach in SFMK
  + MAJOR: The CERVUS pedigree is built on the number of mismatches one parent at a time, (i.e. searches for the best dam by mismatches, then the best sire, not the best parent offspring trio).
    - Introduces the problem of creating impossible pedigrees (e.g. parents in different years), which are subsequently filtered, leading to unassigned offspring and biasing the central result of the study (CRR/TLF) downward.
    - Allows for non-Mendelian inheritance by not taking into account the full parent offspring trio. For example if the offspring genotype is AB, a one parent at a time approach would allow for a parent pair of AC AD, when we know this trio is impossible.
    - Throws out nearly all the power that CERVUS has to make assignments over older exclusion based methods, but this isn’t that important (see below).
  + MINOR: CERVUS is run separately for each offspring year (3 times), the results are then concatenated and the best parentage is selected from the concatenated results table. This renders all statistics other than the number of mismatches meaningless (likelihoods can’t be compared across different datasets). Fortunately, the code as-is rarely relies on LOD except to break ties, so it has little effect in the final pedigree.
  + MINOR: The algorithm for choosing which of several likely POPs in CERVUS relies first on number of mismatches, disregarding the more informative LOD score unless there is a tie. This is also how the assignment approaches work in all systems, and I don’t propose changing it.

Holiday Farm Fire

* How did the Holiday Farm fire (started 9/7/2020, last known date of Cougar Trap operations 9/3) influence operations in 2020?
* Fire extended from Near Leaburg to South Fork MCKR. What is the extent of damage to spawning habitat? What effects on water quality/temp?
  + Oregon ETART suggests turbidity and connectivity effects are not severe, but temperature and contaminant effects are severe and will negatively impact population productivity.
* Should USACE still recycle NOR fish if habitat below Cougar is severely degraded?

Sample Metadata

* Need to build a script to fix the “recycled” problem in the metadata.
  + The exact dates and approach used varies between years, but generally, NOR fish that enter the Cougar trap after a certain date are fin clipped and assigned a sample ID, floy tagged, and released downstream. They are only reintroduced above the dam if they come back a second time. However, since they are assigned sample number on their first capture, not the second, they all have their downstream release site saved in progeny, not their final release site. Fixing this is not simple because of the way the metadata files are structured. It will take some heavy lifting to get this done.
  + Also note that the number of parents in the “sample\_summary” document reflects the error above, and not the correction. Therefore there are more candidate parents in each year than described in that document. Fortunately, all were genotyped anyway as potential offspring so this mistake doesn’t require any additional lab work.
* Cougar Dates
  + Very few samples from the Cougar trap have date information in progeny, but we have date info for almost all of these individuals. However, it is linked to either the floy tag number (new data), or Nicks old sample id (old data), not the current sample id (progeny metadata). Will have to do a lot of work here too to get all this data in a single place.
* 2020 NOR
  + The 2020 Cougar trap samples have some issues. These have been resolved over email, but let’s confirm the metadata in my repository and on progeny incorporates the resolution.
* Cougar trap arithmetic
  + Sometimes it’s unclear how the values on the summary tab of the cougar trap metadata files shared from USACE match up with the individual level metadata (floy tag data) on the other tab. Let’s work through this and reach out to Chad if there’s anything we can’t resolve.
* Individual Sample issues:
  + Missing metadata samples
    - MRA14\_0120 / OtsAC14SFMK\_0120: This individual has no metadata but was previously included as an offspring by Sard. This made sense, but what to do now that 2014 is a parent year? My inclination is to include it, given that the massive variance in TLF among individuals, every parent we miss can potentially have a big effect on the final estimates of TLF/CRR etc… The problem is sorting this TLF into reintros or outplants of we can’t figure out if is NOR or HOR
    - 2015 no metadata samples (n = 3). These have no release info. Similar to the 2014 individual, previously not an issue because 2015 wasn’t a parent year, but should we include them or not.
    - 2016 no metadata sample (n = 1)
    - Include them in the pedigree, maybe include an appendix, let’s discuss later, maybe have a disclaimer, we have some flexibility
  + 2014 Precocial Males:
    - There are 12 precocial males (already genotyped) in 2014, but there is limited metadata about them. All are considered as offspring previously (so NOR), and I have a note that they are from above the dam, but no idea how they were sampled.
    - Separately, there are 6? (mix NOR and HOR) jacks in 2014 from the Cougar trap, with one outplanted above the dam. he numbers here are different in every source (Cougar metadata tab 1 and tab2, progeny (where they are not listed as jacks at all), and Nicks datasets). Let’s make a final determination on all of these.