Project Aims

1. What are the primary goals for this project?
   1. Analyze structure at neutral markers
      1. What neutral structure exists among basins within OC Chinook?
      2. Do spring-run OC Chinook salmon form a distinct lineage separate from OC fall-run Chinook salmon? Do spring and fall runs within a system fail to share a single MRCA using neutral markers?
      3. *Remember to synthesize what is known about hatchery supplementation of stocks in these rivers with the results of neutral structure – a lot of supplemental hatchery individuals from other systems over the past century will complicate interpretation of population structure in OC Chinook*.
   2. Examine “structure” at run timing and other adaptive markers. Are there adaptive differences? (*is this confounded by biased sampling though?*)
   3. Catalog genetic variation within OC Chinook ESU, particularly at run-timing associated markers
      1. E.g. Are there early-run associated alleles in population not thought have a spring run? Are there late-run alleles in habitats thought to be exclusively early-run? To what extent do we observe heterozygotes, and can we assess if these alleles are in Hardy-Weinberg proportions? Synthesize these results with what is known about passage barriers on each system?
2. Some additional interesting questions we could (**maybe**) address re:GREB1L region:
   1. For samples where we have meaningful phenotypes/inference of run timing (*are there any? Seems no. How comfortable are we inferring migration history from data far away in space and time from freshwater entry?*)
      1. Estimate effect size/conduct association analysis
      2. Estimate additivity/dominance
      3. Does the pattern of dominance change across populations?
   2. Is there variation across populations within OC?
      1. Different LD blocks among the populations?
   3. Spatial/ecological associations?
      1. *Do we have sufficient metadata to conduct isolation-by-resistance/spatial analyses to address ecological/evolutionary question about migration timing alleles? If so this could be worth while*

Proposed Next Steps

We’ve already pulled out the run-timing marker data and polarized alleles using data from the coastal lineage of Columbia River, Rogue River and a very rough association study using the empirical data (with ODFW assigned run used as the phenotype). What are some next steps?

1. Exploratory Data Analysis
   1. PCA of neutral and full dataset to characterize any major population structure. A basic understanding of structure is needed before we can plan downstream analyses (for example if there are two major clusters, then we should approach other analyses hierarchically, e.g. LD plots of run-timing markers should be conducted not only dataset-wide but also separately for each cluster to determine if haplotype structure varies across populations).
   2. Explore the polarized run-timing marker spreadsheet to develop questions and familiarize ourselves with the patterns.
2. Review metadata/phenotypes and sample sizes.
   1. Can we confidently assign freshwater entry timing or a proxy for this phenotype to any samples?
   2. Can we confidently assign any other run timing phenotype?
   3. What samples have meaningful sampling locations? For live samples sampling location is a lower bound of how far the individual will travel upriver. What about for carcass samples? Are we willing to examine the spatial distribution of alleles within a river system using this information? Seems very case-dependent.
3. Summarize state of each basin.
   1. What runs are thought to be extant?
   2. What runs are thought to have been present historically?
   3. What is the stocking history for each basin?
4. Umpqua
   1. What questions can we reasonably address about structure within Umpqua with our data?