***Oncorhynchus mykiss* GTseq Run-Timing Markers**

Rationale

The SFGL *O. mykiss* GTseq panel includes 15 markers with run-timing annotations. Annotations for these markers were derived from 4 published sources (Hasselman BPA Report – 1 marker, Hess, 2016 – 2 markers, Micheletti 2018 BMC Evol Bio – 3 markers, Micheletti 2018 Mol Ecol – 6 markers) and two unpublished sources (noted as Shawn Narum – 1 marker, Steven Micheletti – 2 marker, in the panel metadata).

I attempted to collate information across studies for these markers using their names or the mapping position against the genome (NCBI Accession GCA\_002163495.1, ENSEBML release 100, Omyk\_1.0) to understand how marker associations vary across both lineages and phenotypes used for association studies. Mapping studies used reads from Columbia River populations (CRITFC available in (Ford *et al.* 2020) and (Pearse 2019)) and from Rogue River populations (Dayan - <https://github.com/david-dayan/Omy_GTseq_markers>). In addition to the 4 published sources of marker annotations, I also examined 5 other studies that that consider variation at these markers in steelhead (Collins *et al.* 2020; Kannry *et al.* In Press; Pearse 2019; Prince *et al.* 2017; Willis *et al.* 2020). These results are summarized in the table below, more detailed results available in attached excel spreadsheet, and a summary of disagreements between studies is below. I also include a brief discussion.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Paper** | **Marker/ Sample Subgroup** | Chr28\_11607954 | Omy\_RAD52458-17 | **Omy\_GREB1\_05** | **Chr28\_11625241** | Chr28\_11632591 | Omy\_GREB1\_09 | **Chr28\_11658853** | **Chr28\_11667578** | **Omy\_RAD47080-54** | **Omy\_RAD15709-53** | **Chr28\_11671116** | **Chr28\_11676622** | **Chr28\_11683204** | **Chr28\_11702210** | Chr28\_11773194 |
| **Dayan** | informative | 0 | NA | 1 | 1 | 0 | 0 | 1 | 1 | NA | 1 | NA | 1 | 1 | 1 | 0 |
| **Dayan** | diagnostic | 0 | NA | 0 | 1 | 0 | 0 | 1 | 1 | NA | 1 | NA | 1 | 1 | 1 | 0 |
| Hess 2016 |  | NA | 1 | NA | NA | NA | NA | NA | NA | 1 | NA | NA | NA | NA | NA | NA |
| Michelleti BMC |  | 1 | NA | NA | 1 | 1 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Michelleti Mol Ecol |  | NA | NA | NA | NA | NA | NA | 1 | 1 | NA | NA | 1 | 1 | 1 | 1 | NA |
| Hasselman 2017 |  | NA | NA | NA | NA | NA | NA | NA | NA | NA | 1 | NA | NA | NA | NA | NA |
| **Prince 2017** |  | NA | 1 | NA | NA | NA | NA | NA | NA | 1 | 1 | NA | NA | NA | NA | NA |
| **Pearse 2019** |  | NA | NA | NA | NA | NA | NA | NA | NA | 0 | 0 | NA | NA | NA | NA | NA |
| **Kannry 2020** |  | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Willis 2020 |  | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | NA | 1 | 1 | 1 | NA | 1 |
| Collins 2020 | Coastal | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | NA | 1 | 1 | 1 | NA | 0 |
| Collins 2020 | Interior | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | NA | 1 | 1 | 1 | NA | 0 |

**Table 1:** Marker run-timing association summary across different studies. Red values: no association/linkage in study; Blue: positive associations/strong linkage; NA: marker not included in study or could not be confirmed from available results/data.

**Note 1**: these values are very rough notes and there are differences in the metric used across studies. Bold paper title indicates population other than Columbia River used. See details in text below or in full table (attached file).

**Note 2:** Omy\_RAD47080-54 and Omy\_RAD15709-53 are different GTseq panel markers that tag the same SNP (Chr28: 11667915 on the Omyk\_1.0 genome and Scaffold79929e: 649428 in the genome used for the Prince *et al* 2017 study). The SFGL will keep both markers in the panel but after evaluating genotyping success rate, will only retain data from one of these markers for downstream analyses.

**Note 3:** NA markers: Chr28\_11671116 was filtered out of final analysis because of high missingness in summer run samples, however it was diagnostic for the samples with calls. Omy\_RAD52458-17 was filtered out because of paralog issues, but was highly informative (fixed in summer sample, mixed in winter sample). Omy\_RAD47080-54 was a duplicate of Omy\_RAD15709-53 and was >99% concordant, but had more missing data.

**Note 4:** Matching of markers across different studies that use different genotyping methods or alignment to different genomes comes from combining results from CRITFC mapping results with Dayan mapping results. Check the [pdfs here](https://github.com/david-dayan/Omy_GTseq_markers/tree/master/shared_info). I don’t have the methods for these but I assume they are correct.

I examine these discrepancies below

*Chr28\_11607954*

According to panel metadata, this marker was added to the panel after identification from Pool-seq data in (Micheletti *et al.* 2018a). I did not find evidence that it is included any additional studies other the recent GTseq studies on the Columbia River (Collins 2020; Willis 2020). Therefore we only have data for Columbia River steelhead.

Interestingly, a recent GTseq study (Collins 2020) demonstrated that this marker, which is positioned on the far 5’ flank of the greb1/rock1 region, is not strongly linked to other greb1/rock1 region SNPs in the interior Columbia lineage, but is linked in the coastal lineage. Willis *et al* (2020) find a similar linkage pattern for this marker and go on to identify weaker association for this marker than others in the greb1 and intergenic regions.

*Omy\_GREB1\_05*

This marker was added to the panel from unpublished CRITFC results (Micheletti). It is not diagnostic for run timing in the Rogue, but it is highly informative (fixed in summer sample, allele frequency ~0.8 for alternative allele in winter sample). It is strongly associated with run timing in both Columbia River coastal and interior populations. Surveys of genetic variation among Columbia river steelhead populations identify two distinct linkage blocks among interior steelehead, with one block (5’) encompassing genic and immediately proximate intergenic greb1 SNPs, and a second encompassing only intergenic SNPs (Willis 2020; Collins 2020). The 5’ block showed stronger association in the interior lineage, and this SNP showed strong differences in effect size and association strength across lineages.

*Chr28\_11632591*

According to panel metadata, this marker was added to the panel after identification from Pool-seq data in (Micheletti *et al.* 2018a). I did not find evidence that it is included any additional studies other the recent GTseq studies on the Columbia River (Collins 2020; Willis 2020). Therefore we only have data for Columbia River steelhead. Similar to Omy\_GREB1\_05, this SNP is part of the 5’ linkage group and demonstrates strong differences in association strength and effect size across coastal vs interior lineages.

*Omy\_GREB1\_09*

Same as *Chr28\_11632591* in panel origin and association to the largely genic 5’ linkage group in interior Columbia River fish. However, as an individual marker, no variance in association strength across lineages.

*Chr28\_11773194*

This marker is on the far 3’ end of the region and the only marker that directly tags rock1. It is also very far from the other markers. It is unpublished from the CRITFC lab and used only in later Columbia River GTseq surveys. While it demonstrates significant association with run timing, it has the weakest association of any of the GTseq markers. It is in weaker linkage to other markers across both coastal and interior Columbia River Steelhead populations.

*Omy\_RAD47080-54 (also duplicate)*

This marker was diagnostic in the Rogue and strongly associated in Columbia River, Eel and Umpqua populations. However, unpublished surveys of steelhead variation at this SNP (available as public comment in response Northern California steelhead DPS petition – Pearse 2019) show summer and winter run with high frequency for the alternative allele in multiple populations (Trinity, Washougal, Sacramento).

*Omy\_RAD52458-17*

This marker tags one of the three snps that overlap between those Prince et al 2017 and those commonly used in GTseq studies. It was filtered out because of paralog issues, but because it is the focus of intensive studies (and litigation) sharing here that it was highly informative but not diagnostic. It was fixed in the summer sample, but demonstrated both heterozygotes and both homozygotes in the winter run sample.

Markers for Classification / Run timing marker filtering log

We use different sets of markers to classify half-pounders and late-summer run adults in (a) the half-pounder manuscript and (b) discussions with ODFW about our findings regarding diagnostic markers for *O. mykiss* on the Rogue. For (a) we build a classifier from all loci and then make assignments based on 95% credible intervals. In this approach much of this classification power comes from 7 diagnostic markers (fixed for alternative alleles) plus one additional highly informative markers (not fixed but strongly segregating). For (b) we only call individuals as winter-like or summer like if all known run-timing markers are homozygous for the winter or summer alleles.

The set of known run-timing markers underwent the following filtering:

* Raw unfiltered dataset: 15 markers
* Initial GTseq quality control: 14 markers (removed 1 marker Omy\_RAD52458-17, due to evidence of differing levels of reads from a paralogous region across individuals entering the pipeline)
* Second GTseq quality control: 12 markers. Removed 2 markers: (1) in a later draft of the ms, discovered that a redundant marker (Omy\_RAD47080-54) was mistakenly maintained in the analysis, (2) noticed a population specific pattern of missingness in a run timing marker (Chr28\_11671116 ~30% missing in early-summer run samples, but only ~2% missing in winter run).
* Of the remaining 12 markers, 5 were not strictly diagnostic (fixed for alternative alleles in the winter and early-summer run samples), leaving only 7 markers for classification in (b) above.

Discussion

Since their initial identification, the 15 run-timing markers in our GTseq panel have been used to conduct extensive genetic surveys among Columbia River steelhead (Collins *et al.* 2020; Willis *et al.* 2020), and studies in other river systems have catalogued genetic variation at a small number of same SNPs tagged by our markers using different genotyping approaches (Pearse 2019; Prince *et al.* 2017). These studies point to varying strengths of association with run-timing phenotypes across the *greb1/rock1* region and lineage specific recombination events that have produced different numbers of haplotype blocks within the genomic region across populations (Collins *et al.* 2020; Willis *et al.* 2020). Indeed, in our study, the two markers on the far 5’ and 3’ flanks of the *greb1/rock1* region do not segregate among run-timing groups, suggesting that recombination in Rogue River steelhead has potentially freed these flanking SNPs from association with causal SNPs in the interior of the region. While these flanking markers are still associated with run-timing in the Columbia River, they demonstrate weaker linkage and strength of association than others (Collins *et al.* 2020; Willis *et al.* 2020). In contrast, the remaining two uninformative markers on the Rogue River are centered within a haplotype block identified in the Columbia River steelhead that is strongly associated with run timing in both Columbia River steelhead (Collins *et al.* 2020) and steelhead on the Eel and Umpqua rivers (Prince *et al.* 2017). Taken together, the observation of uninformative markers in the Rogue despite strong associations in other lineages highlight the importance of lineage specific marker validation before application of genetic markers for management purposes.

Collins EE, Hargrove JS, Delomas TA, Narum SR (2020) Distribution of genetic variation underlying adult migration timing in steelhead of the columbia river basin. *Ecol Evol* **10**, 9486-9502.

Ford MN, Krista; Waples, Robin; Anderson, Eric C; Kardos, Marty;, Koch IM, Garrett; Miller, Michael R; Myers, Jim; Naish, Kerry; Narum, Shawn;, O'Malley KGP, Devon; Seamons, Todd; Spidle, Adrian; Swanson, Penny;, Thompson TQW, Ken; Willis, Stuart (2020) Reviewing and synthesizing the state of the science regarding associations between adult run timing and specific genotypes in chinook salmon and steelhead: Report of a workshop held in seattle, washington, 27–28 february 2020.

Hasselman DJH, Stephanie A; Matala, Amanda R; Matala, Andrew P; Micheletti, Steven J; Narum, Shawn R (2017) Genetic assessment of columbia river stocks.

Hess JE, Zendt JS, Matala AR, Narum SR (2016) Genetic basis of adult migration timing in anadromous steelhead discovered through multivariate association testing. *Proc Biol Sci* **283**.

Kannry SH, O'Rourke SM, Kelson SJ, Miller MR (In Press) On the ecology and distribution of steelhead (oncorhynchus mykiss) in california’s eel river. *Heredity*.

Micheletti SJ, Hess JE, Zendt JS, Narum SR (2018a) Selection at a genomic region of major effect is responsible for evolution of complex life histories in anadromous steelhead. *BMC Evol Biol* **18**, 140.

Micheletti SJ, Matala AR, Matala AP, Narum SR (2018b) Landscape features along migratory routes influence adaptive genomic variation in anadromous steelhead (oncorhynchus mykiss). *Mol Ecol* **27**, 128-145.

Pearse DE (2019) Northern california steelhead dps-configuration review-panel report

Prince DJ, O'Rourke SM, Thompson TQ*, et al.* (2017) The evolutionary basis of premature migration in pacific salmon highlights the utility of genomics for informing conservation. *Sci Adv* **3**, e1603198.

Willis SC, Hess JE, Fryer JK*, et al.* (2020) Steelhead (oncorhynchus mykiss) lineages and sexes show variable patterns of association of adult migration timing and age‐at‐maturity traits with two genomic regions. *Evolutionary Applications*.