

RESEARCH ARTICLE

Open Access



# Selection at a genomic region of major effect is responsible for evolution of complex life histories in anadromous steelhead

Steven J. Micheletti<sup>1\*</sup> , Jon E. Hess<sup>2</sup>, Joseph S. Zandt<sup>3</sup> and Shawn R. Narum<sup>1</sup>

## Abstract

**Background:** Disparity in the timing of biological events occurs across a variety of systems, yet the understanding of genetic basis underlying diverse phenologies remains limited. Variation in maturation timing occurs in steelhead trout, which has been associated with *greb1L*, an oestrogen target gene. Previous techniques that identified this gene only accounted for about 0.5–2.0% of the genome and solely investigated coastal populations, leaving uncertainty on the genetic basis of this trait and its prevalence across a larger geographic scale.

**Results:** We used a three-tiered approach to interrogate the genomic basis of complex phenology in anadromous steelhead. First, fine scale mapping with 5.3 million SNPs from resequencing data covering 68% of the genome confirmed a 309-kb region consisting of four genes on chromosome 28, including *greb1L*, to be the genomic region of major effect for maturation timing. Second, broad-scale characterization of candidate *greb1L* genotypes across 59 populations revealed unexpected patterns in maturation phenology for inland fish migrating long distances relative to those in coastal streams. Finally, genotypes from 890 PIT-tag tracked steelhead determined associations with early versus late arrival to spawning grounds that were previously unknown.

**Conclusions:** This study clarifies the genetic bases for disparity in phenology observed in steelhead, determining an unanticipated trait association with premature versus mature arrival to spawning grounds and identifying multiple candidate genes potentially contributing to this variation from a single genomic region of major effect. This illustrates how dense genome mapping and detailed phenotypic characterization can clarify genotype to phenotype associations across geographic ranges of species.

**Keywords:** *Oncorhynchus*, Adaptation, Genome evolution, Pooled sequencing, Sexual maturation, *greb1L*, Migration

## Background

Variation in the temporal occurrence of life history events, or phenology, occurs among a vast number of plant and animal systems [1–3]. The timing of processes such as migration, hibernation, flowering, and breeding can directly influence survival because essential resources vary over both time and space [4]. Therefore, maintaining interspecific phenological variation is often essential for a species' persistence, as the timing of biological events may be

beneficial or detrimental depending on cyclical variation of biotic and abiotic factors [5]. Thus, balancing selection can act to preserve variation in phenology [6, 7] which is often reflected by high genomic differentiation within multiple species [8, 9]. Consequently, discovery of phenology-related genomic variation is important both for understanding the genomic evolution of an organism and managing populations with phenological variation in the wild [10].

Variation in the timing of migration occurs across a variety of taxa but is particularly consistent and predictable in multiple anadromous salmonid species, which migrate from the ocean to freshwater tributaries to

\* Correspondence: [mics@critfc.org](mailto:mics@critfc.org)

<sup>1</sup>Columbia River Inter-Tribal Fish Commission, 3059F National Fish Hatchery Road, Hagerman, ID 83332, USA

Full list of author information is available at the end of the article



spawn [11]. Specifically, steelhead trout (*Oncorhynchus mykiss*) show distinct bimodal variation in the timing of entry into freshwater tributaries (freshwater entry maturation) [12, 13]. Steelhead that enter freshwater early are sexually premature (stream-maturing) and undergo maturation while in freshwater, whereas late migrating fish typically become sexually mature in the ocean prior to freshwater entry (ocean-maturing) [11, 14]. Despite these two distinct maturation-timing strategies, both stream and ocean-maturing fish spawn at similar times, and admixture occurs in coastal streams where both strategies are present [13, 15]. These alternate phenotypes do not affect population structure as fish with distinct maturation differences from the same geographic regions tend to be closely related across genome-wide markers [15–17]. In contrast to coastal populations, inland populations of steelhead are comprised exclusively of the stream-maturing type and enter freshwater sexually premature several months in advance of spawning as they migrate long distances to spawning tributaries [14, 16]. Even with similarity in maturation states at freshwater entry of inland populations of steelhead, uncertainty remains in whether variation in maturation exists near freshwater spawning grounds. Other salmonid species demonstrate temporal variation in maturation and arrival timing to inland spawning tributaries [18, 19] suggesting that variation in spawning site maturation may occur in steelhead.

The genetic basis for freshwater entry maturation in steelhead has recently been attributed to a single locus, *Growth regulation by estrogen in breast cancer-like (greb1L)*, an oestrogen target-gene. Previous genomic reduction techniques (e.g., RAD-seq) identified multiple single-nucleotide polymorphisms (SNPs) within *greb1L* that successfully differentiated ocean and stream-maturing fish [13, 15]. However, these techniques were relatively coarse and effectively evaluated 0.5–2.0% of the steelhead genome, leaving uncertainty to the genetic basis of freshwater entry maturation. The recent availability of a high-quality *O. mykiss* genome assembly and full-genome resequencing techniques enables a more precise investigation of the genomic basis of maturation in steelhead. Further, access to passive integrated transponder (PIT) tagging data offers a proxy for measuring spawning site maturation that has not been explored in previous studies.

In this study we investigated three primary questions related to the evolution of complex phenology in anadromous steelhead. First, we evaluated whether dense genome scans could reveal additional candidate genes associated with stream versus ocean maturation phenotypes in steelhead from replicated streams. Second, we tested whether a previously identified candidate gene, *greb1L*, was consistently associated with freshwater entry maturation phenotypes in steelhead across a broad geographic range that included far-migrating inland populations in

addition to previously studied coastal populations. Third, we examined spawning site arrival phenotypes from individual fish to test for genotype to phenotype associations of *greb1L* with spawning site maturation. Together, we synthesized results to explore models of selection that likely maintain genomic variation of complex maturation phenotypes across a broad distribution of steelhead.

## Methods

### Genomic resequencing for fine scale mapping of phenology traits

To acquire high-density coverage of the *O. mykiss* genome, we utilized a pooled-sequencing (Pool-seq) approach. Pool-seq involves combining individual DNA samples together and sequencing the homogenized DNA mixture together [20]. This method provides a representation of dense population allele frequencies across a reference genome [21]. However, dense representation of the genome comes at the loss of individual genotypes which are generally used to estimate metrics such as linkage disequilibrium and heterozygosity. We implemented Pool-seq in two independent spawning tributaries in the Columbia River Basin, the Klickitat and Kalama rivers, which both contain stream and ocean-maturing fish. We targeted fish for each library by peak run-times for the two phenologies: stream-maturing in July and ocean-maturing in March for total of 193 individuals collected between 2003 and 2005 (Kalama) and 2014–2017 (Klickitat). Using non-invasive samples of fin clips from fish trapped at weirs, we pooled samples accordingly: Kalama ocean-maturing samples ( $n = 50$ ), Kalama stream-maturing samples ( $n = 46$ ), Klickitat ocean-maturing samples ( $n = 47$ ), Klickitat stream-maturing samples ( $n = 50$ ) (Additional file 1: Table S1).

All four libraries were prepared using a modified NEBNext Ultra enzymatic fragmentation protocol (Additional file 1) [22, 23]. In summary, individual DNA was quantified with pico-green fluorescence on a Tecan M200 (Tecan, Männedorf, Switzerland) and normalized within two-standard deviations of the mean concentration to avoid over-representation of any given individual [23]. Pooled samples were fragmented using NEBNext Ultra dsDNA fragmentase and cleaned using a Qiagen MinElute. After ligation of Illumina adaptors, we targeted sequences with a mean size of 500 bp, performed PCR amplification, then cleaned PCR product using AMPure XP beads. All libraries were sequenced on an Illumina NextSeq 500 with a targeted 500–800 million paired-end reads per library.

### Pool-seq bioinformatics

Sequenced libraries were prepared with the *PoolParty* pipeline [23]. As part of this pipeline, raw 150 bp paired-end reads were filtered by trimming reads (to a minimum of 50 bp) with a base quality score less than

20 using the *trim-fastq.pl* script part of Popoolation [24]. Trimmed reads were then aligned to the *O. mykiss* reference assembly (Omyk\_1.0; GCA\_002163495) using bwa mem [25] with default parameters. PCR duplicates were removed using samblaster [26] and unpaired and unmapped reads were removed using the SAMtools *view* module [27]. Filtered BAM files were then combined using the SAMtools *mpileup* module, which extracts SNP and coverage information for each pool. To remove any false positive SNPs that often occur around insertion-deletions (indels) we used the *identify-genomic-indel-regions.pl* and *filter-sync-by-gtf.pl* scripts from Popoolation2 to remove any SNPs within 5 bp of indel regions [21]. Only variant positions with a minimum of 15 X depth of coverage and a maximum of 250 X depth of coverage were retained; this eliminated regions that may be paralogs (high coverage) or regions that were likely overrepresented by a small number of individuals (low coverage). Alignment and coverage statistics for all libraries were calculated using the *PPstats* module of *PoolParty*.

To determine selective sweeps or genomic regions with significant differentiation, we implemented sliding-window fixation index ( $F_{ST}$ ), a local score technique to test for statistical association, and a Cochran–Mantel–Haenszel (CMH) test [24, 28]. Sliding window  $F_{ST}$  between stream and ocean-maturing pools, was calculated using Popoolation2, using a sliding window of 5-kb with a step size of 50 bp [28]. Local score is an alternative to Fisher’s exact test (FET) for allele frequency difference that reduces false positives by incorporating linkage disequilibrium. Specifically, local score uses FET *p*-values to determine differentiated genomic regions while simultaneously considering linkage disequilibrium. Local score uses a score function related to  $-\log(10)$  *p* which will vary based on window size. Opposed to combining *p*-values within a fixed window size, local score considers the proximity of statistically significant *p*-values to determine window size iteratively. Finally, the CMH test identifies consistent differences in allele frequencies across biological replicates and computes significance between groups of interest [29]. Thus, in our libraries, the CMH test identified SNPs with allele frequency changes that occurred between stream and ocean-maturing fish from both the Kalama and Klickitat. We considered genomic regions to be significant if they showed statistical support for differentiation between stream and ocean-maturing fish in both local score analyses (analogous to a Bonferroni corrected  $\alpha = 0.05$ ), and the CMH test (Bonferroni corrected  $\alpha = 0.05$ ). Significant regions were then investigated for variant annotations using *Snpeff* [22] which predicts non-synonymous SNPs (nsSNPs) from a general feature format (GFF). Variants identified as nsSNPs are anticipated to be under

selection and more likely to be causal SNPs for a given trait [30].

Population structure analyses were performed using *PPanalyze* from the *PoolParty* pipeline. We used *PPanalyze* to remove SNPs with minor allele frequency (MAF) < 0.05 and create neighbor joining trees based on Nei’s genetic distance and 10,000 bootstraps, using both all genomic SNPs, and SNPs within a 309-kb region of chromosome 28.

### Broad scale characterization of candidate genotype frequencies

To determine association between maturation-timing and genomic regions of major effect across a large geographic scale, we isolated a single informative SNP within *greb1L* (*greb1L*-SNP) in an additional 59 collection localities across the Columbia River Basin [23] ( $n = 2915$ ; Additional file 1: Table S3). In previous studies, *greb1L*-SNP explains a large proportion of trait variation in relation to maturation and consistently differentiates stream and ocean maturing fish [13]. Steelhead populations in North America are generally divided into coastal and inland genetic lineages [11, 31]. For example, the Columbia River Basin, which primarily encompasses the states of Oregon, Washington, and Idaho in the United States, consists of a coastal lineage west of the Cascade mountain range that is genetically distinct from an inland lineage east of the Cascades [32]. These two lineages also differ in respect to maturation-timing. The coastal lineage, which generally has shorter migration distances to spawning sites (50–380 km), consists of both stream and ocean-maturing fish. The inland lineage, which requires longer travel to spawning sites (370–1500 km), only supports steam-maturing fish [18]. Due to the apparent lack of variation in maturation-timing in inland populations, previous studies have only investigated the genetic basis of steelhead migration and maturation in coastal populations [13, 15, 17]. However, we plotted genotype frequencies of *greb1L*-SNP across 59 collections including inland and coastal populations of steelhead using ArcGIS 10.5 to represent a broad geographic range.

### Individual phenotypes to refine genotype to phenotype associations

To determine an association between *greb1L*-SNP and spawning tributary arrival, we downloaded array ping dates from the Columbia Basin PIT Tag Information System (PTAGIS; [ptasgis.org](http://ptasgis.org)) for wild tagged fish between 2012 and 2016 in spawning tributaries across the Columbia River Basin. We retained data from sub-basins with known spawning tributaries which we also had sufficient individual DNA tissues ( $N > 50$ ), leading to 6 distinct sub-basins, primarily in the inland lineage, with 890 fish in total (Additional file 1: Table S4). DNA from each individual was genotyped with a panel of markers

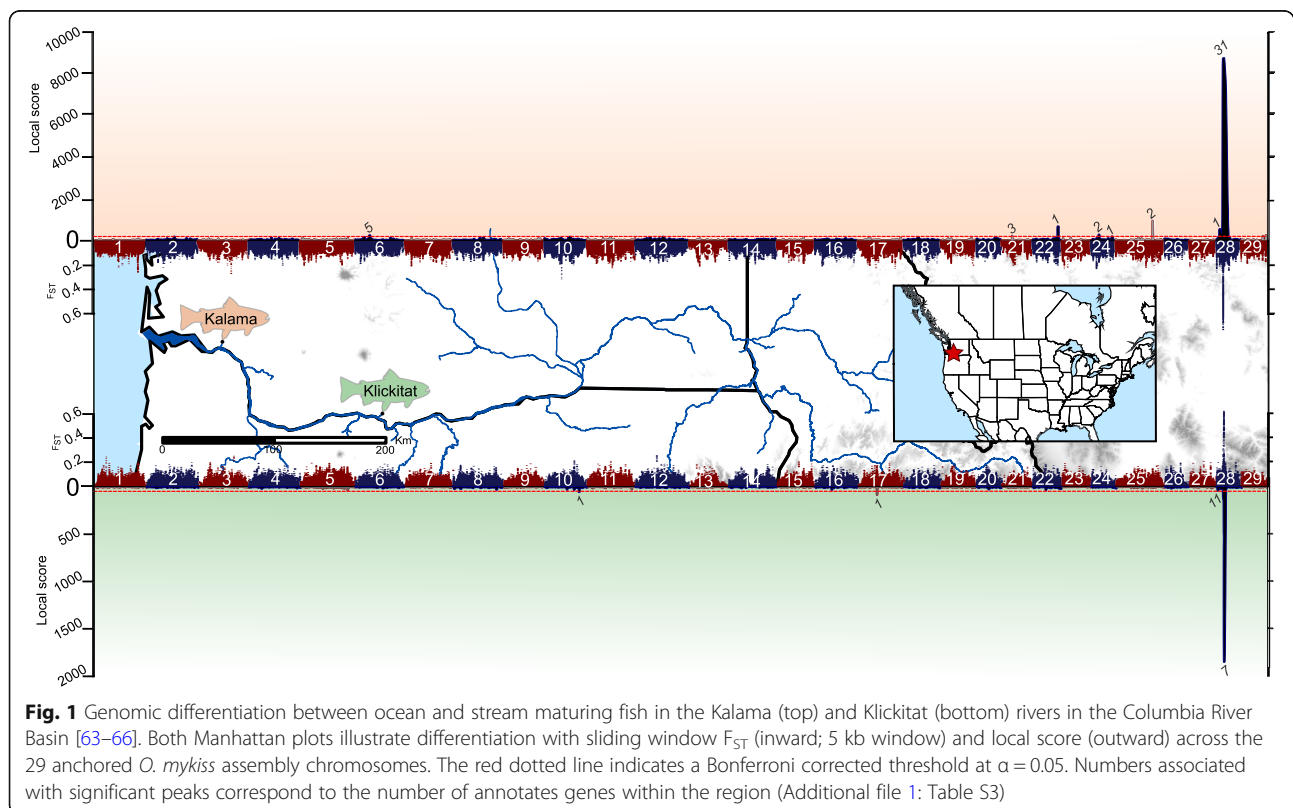
using Genotyping-in-Thousands by sequencing (GT-seq) [33] to isolate and genotype a single *greb1L*-SNP (identified as 47080\_54 in [13]). For each of the six sub-basins we determined significant associations between individual genotype (either premature [AA], heterozygote [AG], or mature [GG]) and spawning tributary arrival week using a one-way analysis of variance (ANOVA) paired with a Tukey's range test [34]. We additionally determined the variance explained by the genotype in each location using ANOVA sum of squares.

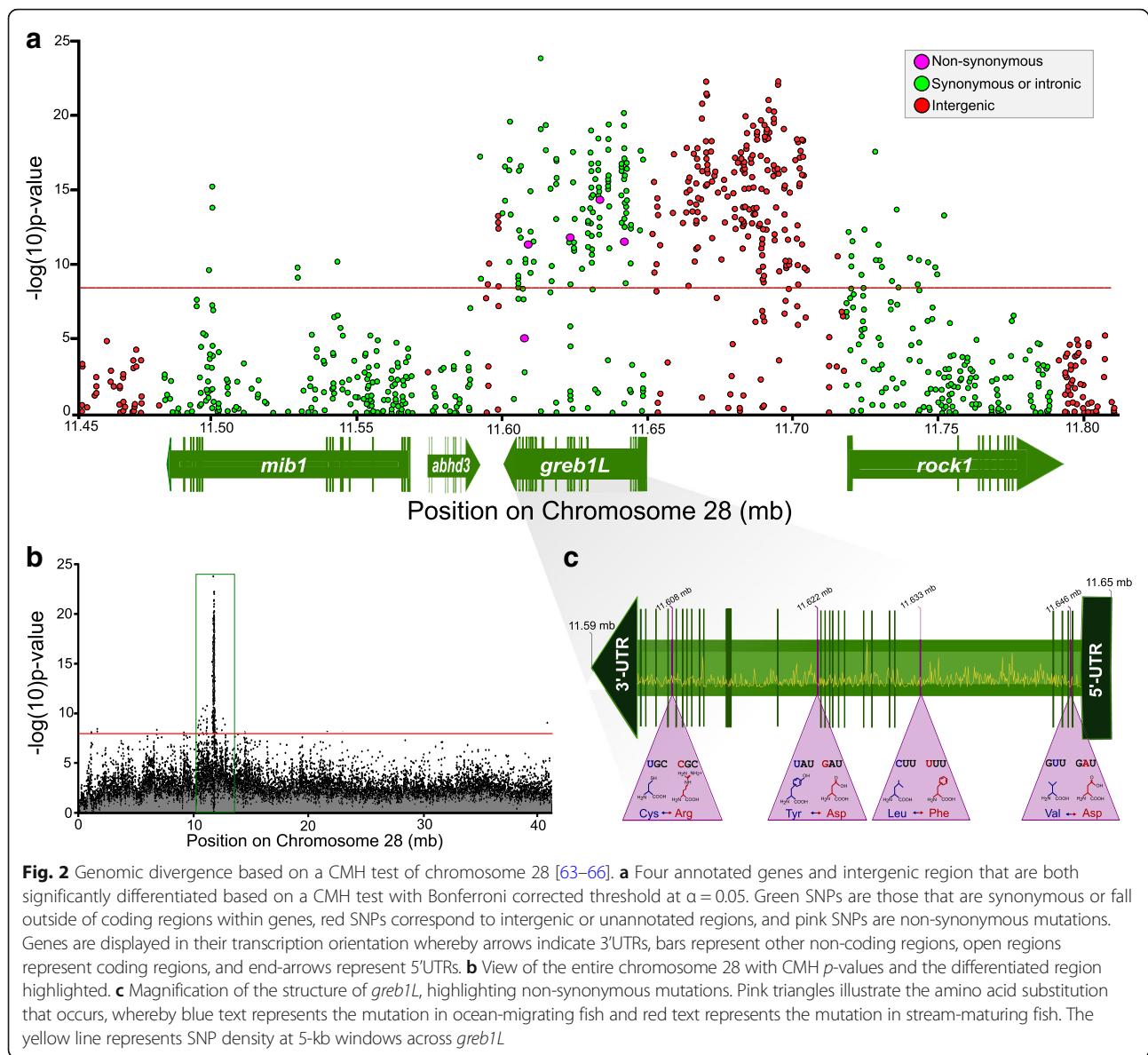
## Results

### Genomic resequencing for fine scale mapping of phenology traits

Pooled-sequencing of four libraries consisting of ocean and stream-maturing fish from the Klickitat and Kalama rivers covered 68.2% of the *O. mykiss* assembly genome between 15 X – 250 X with an average depth of coverage of 33 X (Additional file 1: Figure S1). We identified a total of 21,957,491 SNPs between all four libraries that met coverage threshold criteria, among which 5,321,204 had a minor allele frequency (MAF) > 0.05. A comparison of ocean and stream-maturing fish between both populations using local score [35] and sliding-window  $F_{ST}$  [28] revealed that multiple genomic regions, consisting of 68 annotated genes in total, were significantly differentiated (Fig. 1; Additional file 1: Table S2). However,

a single large genomic region on chromosome 28 was the only consistently significantly differentiated region between both comparisons. A Cochran–Mantel–Haenszel test (CMH test [29]) for consistent changes in allele frequencies between stream and ocean-maturing fish across both populations identified a highly differentiated 309-kb region consisting of 5,294 SNPs, 361 of which were statistically significant after a Bonferroni corrected threshold of  $\alpha = 0.05$  (Fig. 2a, b). When all genomic SNPs were considered, a neighbor-joining tree using Nei's genetic distance [36] successfully separated libraries by the expected neutral geographic structure (Klickitat libraries vs. Kalama libraries; Additional file 1: Figure S2a). However, when only SNPs within the 309-kb region were considered, the neighbor-joining tree separated libraries by stream-maturing and ocean-maturing phenologies (Additional file 1: Figure S2b). Within the 309-kb significant region were four annotated genes: Mindbomb E3 Ubiquitin Protein Ligase 1 (*mib1mib1*), *Abhydrolase Domain Containing 3* (*abhd3*), *Growth regulation by estrogen in breast cancer-like 1*, (*greb1L*), and *Rho Associated Coiled-Coil Containing Protein Kinase 1*, (*rock1*), as well as two intergenic regions (Fig. 2a). While all genes contained significantly differentiated SNPs, *greb1L* and the immediate upstream intergenic region contributed the most significantly differentiated SNPs (36% and 51% of significant SNPs, respectively).



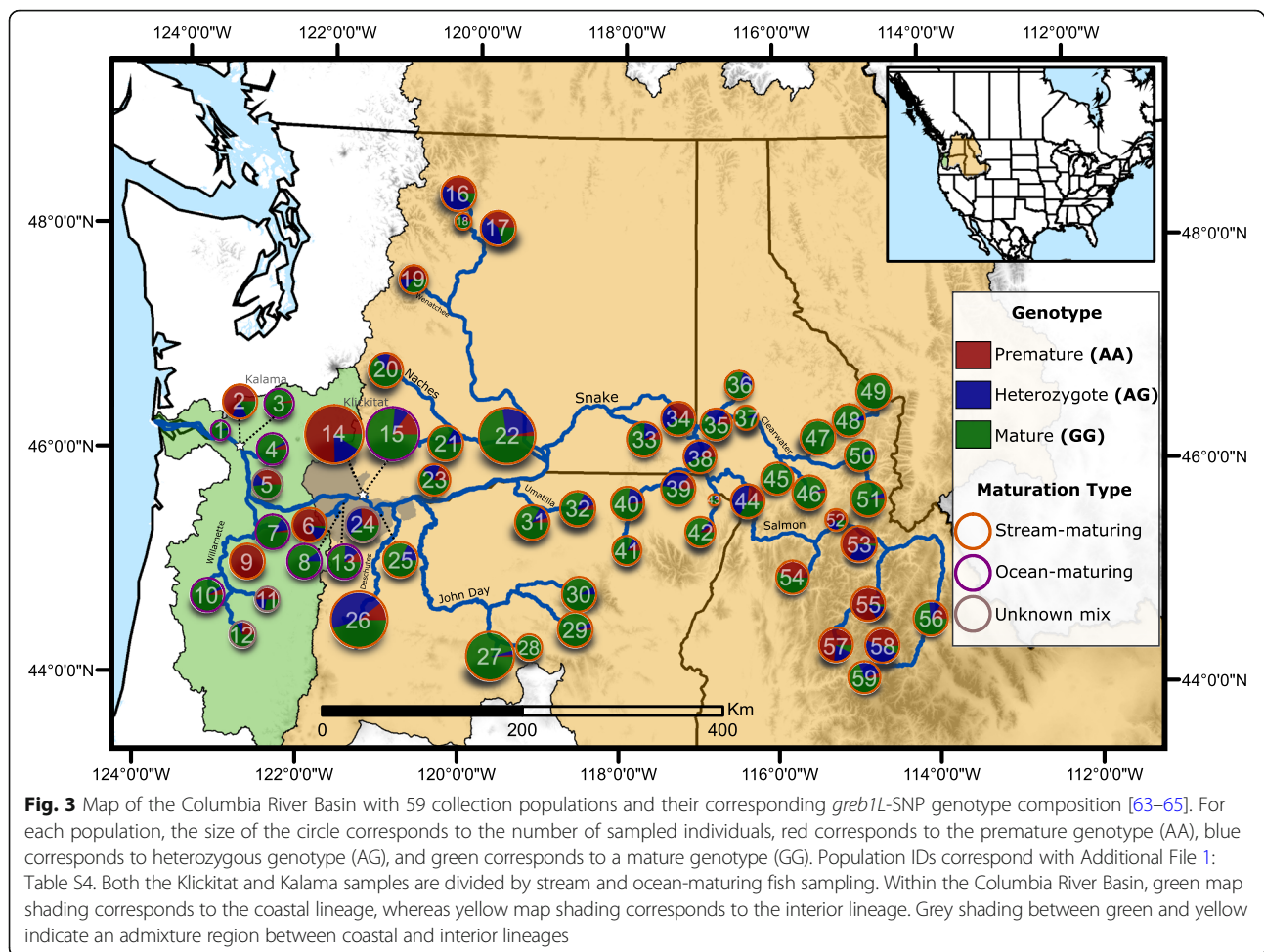


Furthermore, a genomic variant annotation analysis using SnpEff [37] revealed that *greb1L* was the only gene to contain five non-conservative, non-synonymous SNP mutations (nsSNPs), four of which surpassed significance thresholds (Fig. 2c, Additional file 1: Table S3).

#### Broad scale characterization of candidate genotype frequencies

To determine the association between the differentiated chromosome 28 region with phenology on a broad geographic scale, we examined genotype frequencies of a diagnostic *greb1L* marker (*greb1L*-SNP; Chr.28, position 11613335) across both coastal and inland populations. This *greb1L*-SNP has been demonstrated to be strongly associated with freshwater entry maturation in previous studies of coastal steelhead populations [13, 15, 17].

Specifically, a “mature” *greb1L*-SNP genotype (GG) was strongly associated with ocean-maturation and a “premature” genotype (AA) was strongly associated with stream-maturation. Genotypes for the *greb1L*-SNP were isolated from a recent study using RAD-seq in 59 populations of steelhead across the Columbia River Basin [21] (Additional file 1: Table S4). Of the 59 populations, 15 were located in coastal streams, which support both ocean and stream-maturing fish, and 44 were from inland populations that consist solely of stream-maturing fish. In coastal populations, “premature” (AA) and “mature” genotypes (GG) were associated with stream-maturation and ocean-maturation, respectively ( $\chi^2 = 195$ ,  $p < 0.001$ ; Fig. 3). It was not possible to test for genotype-phenotype association of freshwater entry maturation in inland populations since they were all stream-maturing ( $p = 1$ ); however,



genotype frequencies were variable and the majority genotype in this region was the “mature” type (GG;  $x = 65\%$ ). Prevalence of the “mature” genotype for inland populations was contrary to expectations for stream-maturing steelhead and suggested that further refinement of phenology phenotypes was necessary beyond a simplistic categorization of these populations into ocean and stream-maturing types.

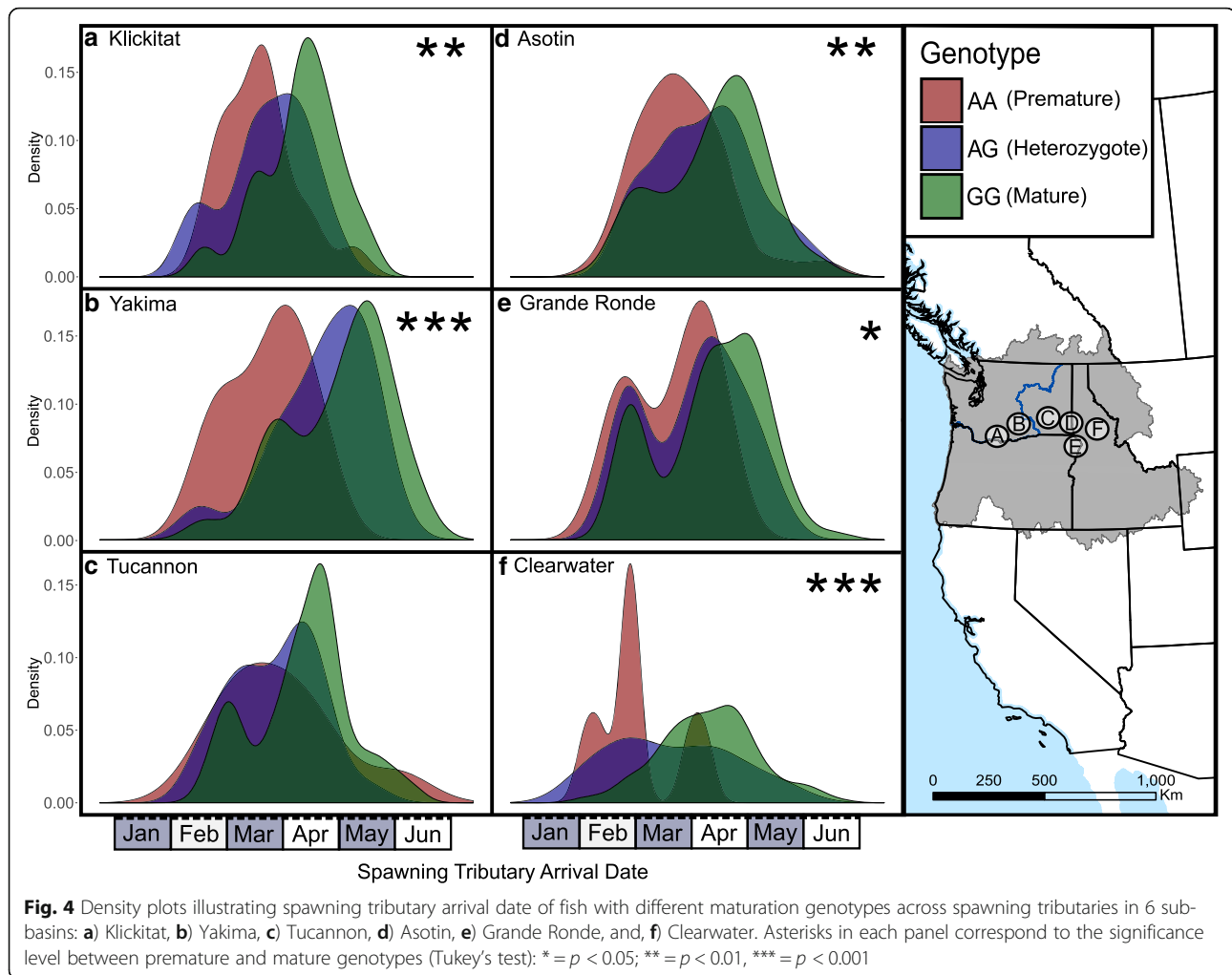
#### Individual phenotypes to refine genotype-by-phenotype associations

We genotyped *greb1L*-SNP from 890 individual steelhead that were tracked throughout their migration to spawning tributaries in six sub-basins with PIT-tag arrays ([ptagis.org](http://ptagis.org)) to test whether the chromosome 28 candidate region was more strongly associated with stream-maturing than ocean-maturing phenotype (Additional file 1: Table S5). Although the state of maturation of an individual fish is a challenging trait to measure, arrival timing to spawning sites at the end of freshwater migration can serve as a sufficient proxy for maturation [38, 39]. We used a one-way ANOVA paired with a Tukey post-hoc test to determine

significant associations between *greb1L*-SNP genotypes and spawning tributary arrival date. All six sub-basins showed a trend between spawning tributary arrival date and genotype class, with five out of the six having statistically different arrival week means between “premature” and “mature” genotypes (Fig. 4, Additional file 1: Figure S3 and Table S5). On average, fish with “premature” genotypes arrived to spawning tributaries  $3.04 \pm 1.4$  weeks prior to those with “mature” genotypes, fish with heterozygote genotypes arrived  $1.64 \pm 1.2$  weeks later than those with “premature” genotypes and  $1.51 \pm 0.74$  weeks earlier than those with “mature” genotypes. Across all six sub-basins genotypes from this single marker (*greb1L*-SNP) explained  $10\% \pm 4\%$  of spawning tributary arrival date variation (Additional file 1: Table S5).

#### Discussion

Our study confirms a genomic region of major effect underlying phenological variation in anadromous steelhead through dense genome resequencing. Previous studies have mapped SNPs generated from restriction-site associated DNA sequencing (RAD-seq) to *greb1L*, yet



**Fig. 4** Density plots illustrating spawning tributary arrival date of fish with different maturation genotypes across spawning tributaries in 6 sub-basins: **a**) Klickitat, **b**) Yakima, **c**) Tucannon, **d**) Asotin, **e**) Grande Ronde, and, **f**) Clearwater. Asterisks in each panel correspond to the significance level between premature and mature genotypes (Tukey's test): \* =  $p < 0.05$ ; \*\* =  $p < 0.01$ , \*\*\* =  $p < 0.001$

have not explicitly explored additional genes of smaller effect upstream and downstream of *greb1L* [13, 15, 17]. Using dense mapping data, we illustrated that *greb1L* is part of a larger genomic region under selection consisting genes and divergent inter-genic regions. This discovery was made possible by both the advancement in quality of the *O. mykiss* genome assembly, and through resequencing techniques (Pooled-sequencing) that provide adequate read coverage across the majority of the reference genome [20]. While previous genome scans with RAD-seq yielded significant association of markers from the *greb1L* region, fine scale mapping provided a broader understanding of the genomic basis of this phenological trait in steelhead due to higher marker density covering a large portion of the genome [40].

The additional genes of smaller effect characterized in this study indicate that the genomic basis for maturation phenology in anadromous steelhead may encompass a larger genomic region on chromosome 28 than previously understood. *greb1L* has consistently shown the most compelling associations to maturation phenotypes

[13, 15, 17]. Our results additionally highlight extreme differentiation in the upstream intergenic region of *greb1L* which likely contains many regulatory components such as transcription factors, promoters, and enhancers [41]. *greb1L* has obvious connections to sexual maturation since it mediates the interaction of oestrogen with other target proteins [13, 42]. Migrating fish, either mature, or nearing maturity, have elevated levels of oestrogen in their bloodstream which relates to multiple sexual characteristics such as egg formation and testicular development [43]. Additionally, we showed that *greb1L* contains multiple non-conservative and non-synonymous mutations. These changes in protein structure are compelling candidates that are possibility under selection, which additionally provide evidence that *greb1L* is a key gene under selection for maturation phenology [30, 44]. *rock1*, the gene directly upstream of *greb1L*, also has obvious ties to maturation as it has been connected to embryo development in zebrafish [45] and testicular development in humans [46]. Furthermore, it is a key regulator of actin-myosin contraction [47] which may be connected to

the long migration distances anadromous steelhead need to swim to reach spawning grounds [21]. *abhd3*, the gene immediately downstream from *greb1L*, was the least differentiated, yet is a physiological regulator of medium-chain phospholipids [48]. The necessity for efficient fat disposition is essential in fish during long-distance migration [19, 49]. *mib1* has less-apparent connections to migration and maturity as it primarily relates to cell apoptosis [50]; however, *mib1* does influence ventricle formation and can be connected to cardiac function and swimming performance [51, 52]. Finally, a large intergenic region is highly differentiated between *greb1L* and *rock1* which may simply be a gene that currently lacks annotation, or possibly a region that consists of enhancers or promoters for the nearby genes or non-coding RNAs. If the latter, these intergenic SNPs may play a large regulatory role in expression level [53]. Overall, this highly divergent region on chromosome 28 contains compelling candidate genes which justifies further validation and marker development to investigate maturation phenotypes of steelhead in more detail. For example, informative SNPs from these candidate genes can be included into high-throughput amplicon sequencing panels to screen large numbers of individuals [33].

Our results indicated that the candidate region on chromosome 28 is most consistently associated with arrival timing on spawning grounds rather than arrival timing at freshwater entry across populations included in this study. Traditionally, arrival timing at freshwater entry has been a phenotypic proxy for maturation timing that has been used to characterize fish entering freshwater as either sexually mature or premature [11], which partly owes to the relative ease of recording this proxy trait for returning steelhead at many lower river collection sites including dams, weirs, and hatcheries [54, 55]. Coastal tributaries support both ocean and stream-maturing fish, whereas steelhead returning to inland tributaries are all stream-maturing fish [18]. Previous studies have solely investigated freshwater entry maturation of steelhead in coastal and lowland rivers [13, 15, 17], and their findings beget expectations that all inland fish may be fixed for “premature” *greb1L* genotypes. To the contrary, using spawning tributary arrival time, we show that variation in maturation phenology does occur in inland steelhead, which are all stream-maturing fish. Specifically, about 10% of variance of time to arrival is explained by *greb1L*-SNP. While the strength of this pattern is not overwhelming, it is still compelling given that PIT-tag arrival times were used as a proxy and may be prone to some error, and only a single informative SNP was used. In addition, it currently serves as the only explanation for variation in *greb1L* genotypes in stream-maturing fish. Given this, fish with “premature” *greb1L*-SNP genotypes generally arrive to spawning tributaries early, fish with “mature” genotypes arrive later,

and heterozygous fish arrive in an intermediate timeframe. This association suggests that inland fish tend to hold in larger freshwater tributaries for several months as premature fish, and then migrate to spawning grounds in headwater tributaries over a continuum of maturation states before all fish become sexually mature and spawn together. This admixture of fish with varying phenotypes shows no Wahlund effect heterozygote deficit [21], providing additional evidence that population structure is not directly influenced by this phenology [15]. Inland steelhead with “premature” genotypes ascend to spawning grounds early (premature arrival) and continue to mature there, whereas fish with “mature” genotypes become sexually mature in freshwater downstream of spawning grounds, then move upstream to spawning grounds once they are mature (mature arrival). However, in coastal populations, mature arrival and premature arrival are likely analogous to ocean-maturing and stream-maturing phenotypes (respectively) that are commonly observed as steelhead enter freshwater systems near the ocean. Coastal and inland steelhead populations contain *greb1L* “mature” and “premature” genotypes; however, the inland *greb1L* “mature” and “premature” genotypes both exhibit an early and narrow range of freshwater entry timings, but later become diverged in their spawning tributary entry timing. In contrast, the *greb1L* “mature” and “premature” genotypes of the coastal steelhead populations instead show large disparity in freshwater entry timing from the ocean. This difference in freshwater entry timings of the *greb1L* mature genotypes of the coastal versus inland lineages may be due to the long migration distance that inland fish must swim to reach their (300–1500 km) spawning sites [21]. Regardless of their *greb1L* genotypes, all fish from inland populations must migrate early to approach spawning grounds before environmental conditions become unfavorable [56, 57] including passage through unfavorable migratory corridors [21]. Then, when near inland spawning sites, balancing selection may preserve the variation in *greb1L* genotypes which manifests as variation in spawning tributary arrival (i.e., premature or mature spawning tributary arrival). In some years, due to resources such as spawning habitat availability, and environmental conditions, it may be beneficial to arrive to a spawning tributary early (*greb1L* premature genotype), whereas in others it may be more beneficial to arrive late (*greb1L* mature genotype) [14, 58]. In addition, the benefits of tributary arrival may vary based on geographic localities. Thus, balancing selection likely maintains the genomic variation in chromosome 28, a phenomenon that may be confirmed with further studies that investigate variation in steelhead phenology in specific locations and throughout the species range.

As commonly illustrated, maintaining adaptive genetic variation is essential for protecting species at risk to



extinction [59, 60]. We provide evidence that genomic variation linked to sexual maturation is present in inland steelhead populations, which were previously assumed to be fixed for *greb1L* premature genotypes similar to fish exhibiting stream-maturing phenotypes in coastal populations. This discovery illustrates that understanding complex phenology patterns is a difficult process due to challenges of monitoring migrating fish through sexual development stages across large watersheds. However, this challenge can be overcome by monitoring efforts that apply tags and collect non-lethal tissue from migrating fish as they enter freshwater [13]. A profound understanding of complex life history traits can enable managers to maintain the necessary levels of neutral and adaptive genetic variation in wild populations to mitigate impacts of climate change on complex phenology traits [15, 61, 62].

## Conclusions

Anadromous salmonids, a recreational and culturally significant family of fishes, have complex life histories related to migration and maturation. Previous investigations into the genetic basis of maturation-timing suggested that a single gene, *greb1L*, was related to sexually premature and mature entry of steelhead into freshwater systems. Using genome resequencing for fine scale mapping of maturation traits during migration, we identified a 309-kb genomic region of major effect for freshwater entry maturation that included four candidate genes, *greb1L*, *rock1*, *mib1*, and *abhd3*. This region also includes a highly significant intergenic region between *greb1L* and *rock1* which may play a regulatory role in expression of these genes. Additionally, broad-scale SNP genotypes from *greb1L* in populations and individuals refined genotype to phenotype associations, revealing that candidate genotypes were more consistently associated with timing of arrival on spawning grounds rather than freshwater entry maturation. These results suggest that there are fitness benefits to arriving to spawning tributaries prematurely or maturely, and variation in this trait is likely to be maintained by balancing selection. Together this study illustrates the importance of high precision genomic scans and detailed phenotypes to identify targets of selection.

## Additional file

**Additional file 1:** Supporting figures, Supporting tables, Pooled-sequencing protocol. (DOCX 820 kb)

## Abbreviations

*abhd3*: Abhydrolase Domain Containing 3; ANOVA: Analysis of Variance; CMH: Cochran–Mantel–Haenszel; FET: Fisher's Exact Test; FST: Fixation Index; *greb1L*: Growth Regulation By Estrogen In Breast Cancer 1 Like; GT-seq: Genotyping-In-Thousands by sequencing; MAF: Minor Allele Frequency; *mib1*: Mindbomb E3 Ubiquitin Protein Ligase 1; PCR: Polymerase chain

reaction; PIT: Passive Integrated Transponder; RAD-seq: Restriction Site Associated DNA Sequencing; *rock1*: Rho Associated Coiled-Coil Containing Protein Kinase 1; SNP: Single-Nucleotide Polymorphism

## Acknowledgements

We thank Stephanie Harmon, Amanda Matala, and Janae Cole for laboratory support. Samples were contributed by biologists from Nez Perce Tribe, Yakama Nation, Warm Springs, and Umatilla, and agencies such as NOAA Fisheries, U.S Fish and Wildlife Service, Idaho Dept. of Fish and Game, Oregon Dept. of Fish and Wildlife, and Washington Dept. of Fish and Wildlife.

## Availability of data and materials

Raw sequencing fastq files for the four pooled-sequencing libraries are provided in the NCBI sequence read archive (SRA; <https://www.ncbi.nlm.nih.gov/sra/SRP151789>) under project SRP151789.

## Authors' contributions

SJM developed scripts, performed analyses, and wrote the manuscript. JSZ contributed to sample collection and experimental design. JEH contributed to SNP development and preliminary analyses. SRN assisted with analyses, experimental design. All authors participated in manuscript editing and revision. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

Sampling activities were conducted in accordance with the terms and conditions of the US Endangered Species Act (ESA) Section 10a1A Permit 1379-6R to Columbia River Inter-Tribal Fish Commission and annual NOAA ESA Section 4d permits to the Yakima Indian Nation. Permits issued under the ESA are reviewed by expert committees to ensure ethical treatment of animals. This included collection of non-lethal samples and precautions to reduce stress and ensure adequate recovery after release.

## Consent for publication

Not applicable

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>Columbia River Inter-Tribal Fish Commission, 3059F National Fish Hatchery Road, Hagerman, ID 83332, USA. <sup>2</sup>Columbia River Inter-Tribal Fish Commission, Portland, OR 97232, USA. <sup>3</sup>Yakama Nation Fisheries Program, Toppenish, WA 98948, USA.

Received: 21 April 2018 Accepted: 24 August 2018

Published online: 15 September 2018

## References

- Ollerton J, Lack A. Relationships between flowering phenology, plant size and reproductive success in shape *Lotus corniculatus* (Fabaceae). *Plant Ecol*. 1998;139(1):35–47.
- Edwards M, Richardson AJ. Impact of climate change on marine pelagic phenology and trophic mismatch. *Nature*. 2004;430(7002):881.
- Gordo O. Why are bird migration dates shifting? A review of weather and climate effects on avian migratory phenology. *Clim Res*. 2007;35(1/2):37–58.
- Schwartz MD. Phenology: an integrative environmental science. Dordrecht, The Netherlands: Kluwer Academic; 2003.
- Forrest J, Miller-Rushing AJ. Toward a synthetic understanding of the role of phenology in ecology and evolution. (2010).
- Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, et al. Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genet*. 2005;1(5):e60.
- Delph LF, Kelly JK. On the importance of balancing selection in plants. *New Phytol*. 2014;201(1):45–56.
- Liedvogel M, Åkesson S, Bensch S. The genetics of migration on the move. *Trends Ecol Evol*. 2011;26(11):561–9.

9. Derks MF, Smit S, Salis L, Schijlen E, Bossers A, Mateman C, et al. The genome of winter moth (*Operophtera brumata*) provides a genomic perspective on sexual dimorphism and phenology. *Genome Biol Evol.* 2015; 7(8):2321–32.
10. Allendorf FW, Hohenlohe PA, Luikart G. Genomics and the future of conservation genetics. *Nat Rev Genet.* 2010;11(10):697.
11. Quinn TP. The behavior and ecology of Pacific salmon and trout. UBC press, (2011).
12. Leider SA, Chilcote MW, Loch JJ. Comparative life history characteristics of hatchery and wild steelhead trout (*Salmo gairdneri*) of summer and winter races in the Kalama River, Washington. *Can J Fish Aquat Sci.* 1986;43(7): 1398–409.
13. Hess JE, Zandt JS, Matala AR, Narum SR. Genetic basis of adult migration timing in anadromous steelhead discovered through multivariate association testing. *Proc R Soc B.* 2016;283(1830):20153064. The Royal Society
14. Quinn TP, McGinnity P, Reed TE. The paradox of “premature migration” by adult anadromous salmonid fishes: patterns and hypotheses. *Can J Fish Aquat Sci.* 2015;73(7):1015–30.
15. Prince DJ, O'Rourke SM, Thompson TQ, Ali OA, Lyman HS, Saglam IK, et al. The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Sci Adv.* 2017;3(8):e1603198.
16. Arciniega M, Clemente AJ, Miller MR, Peterson M, Garza JC, Pearse DE. Parallel evolution of the summer steelhead ecotype in multiple populations from Oregon and northern California. *Conserv Genet.* 2016;17(1):165–75.
17. Thompson TQ, Bellinger RM, O'Rourke SM, Prince DJ, Stevenson AE, Rodrigues AT, Banks MA. Anthropogenic habitat alteration leads to rapid loss of adaptive variation and restoration potential in wild salmon populations. *bioRxiv.* (2018);310714.
18. Robards MD, Quinn TP. The migratory timing of adult summer-run steelhead in the Columbia River over six decades of environmental change. *Trans Am Fish Soc.* 2002;131(3):523–36.
19. Keefer ML, Caudill CC. Homing and straying by anadromous salmonids: a review of mechanisms and rates. *Rev Fish Biol Fish.* 2014;24(1):333–68.
20. Schlötterer C, Tobler R, Kofler R, Nolte V. Sequencing pools of individuals—mining genome-wide polymorphism data without big funding. *Nat Rev Genet.* 2014;15(11):749.
21. Micheletti SJ, Matala AR, Matala AP, Narum SR. Landscape features along migratory routes influence adaptive genomic variation in anadromous steelhead (*Oncorhynchus mykiss*). *Mol Ecol.* 2018;27(1):128–45.
22. Sweet D, Lorente M, Valenzuela A, Lorente J, Alvarez JC. Increasing DNA extraction yield from saliva stains with a modified Chelex method. *Forensic Sci Int.* 1996;83(3):167–77.
23. Micheletti SJ, Narum SR. Utility of pooled sequencing for association mapping in nonmodel organisms. *Mol Ecol Resour.* 2018;18(4):825–35.
24. Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A, Schlötterer C. PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One.* 2011;6(1): e15925.
25. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics.* 2009;25(14):1754–60.
26. Faust GG, Hall IM. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics.* 2014;30(17):2503–5.
27. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics.* 2011;27(21):2987–93.
28. Kofler R, Pandey RV, Schlötterer C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (pool-Seq). *Bioinformatics.* 2011b;27(24):3435–6.
29. Mantel N. Chi-square tests with one degree of freedom; extensions of the mantel-Haenszel procedure. *J Am Stat Assoc.* 1963;58(303):690–700.
30. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 2009;4(7):1073.
31. Utter FM, Campton D, Grant S, Milner G, Seeb J, Wishard L. Population structures in indigenous salmonid species of the Pacific northwest. In: Neil WJ, Himswonh DC, editors. *Salmonid ecosystems of the North Pacific.* Oregon state. Corvallis: University Press; 1980. p. 285–304.
32. Brannon EL, Powell MS, Quinn TP, Talbot A. Population structure of Columbia River basin Chinook salmon and steelhead trout. *Rev Fish Sci.* 2004;12(2–3):99–232.
33. Campbell NR, Harmon SA, Narum SR. Genotyping-in-thousands by sequencing (GT-seq): a cost effective SNP genotyping method based on custom amplicon sequencing. *Mol Ecol Resour.* 2015;15(4):855–67.
34. Tukey JW. Comparing individual means in the analysis of variance. *Biometrics.* 1949;5(2):99–114.
35. Fariello MI, Boitard S, Mercier S, Robelin D, Faraut T, Arnould C, Gourichon D. Accounting for linkage disequilibrium in genome scans for selection without individual genotypes: the local score approach. *Mol Ecol.* 2017; 26(14):3700–14.
36. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics.* 2004;20(2):289–90.
37. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly.* 2012;6(2):80–92.
38. Berglund I. Growth and early sexual maturation in Baltic salmon (*Salmo salar*) parr. *Can J Zool.* 1992;70(2):205–11.
39. Achord S, Matthews GM, Johnson OW, Marsh DM. Use of passive integrated transponder (PIT) tags to monitor migration timing of Snake River Chinook salmon smolts. *N Am J Fish Manag.* 1996;16(2):302–13.
40. Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, et al. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *American Naturalist.* 2016;188(4):379–97.
41. van Helden J, André B, Collado-Vides J. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J Mol Biol.* 1998;281(5):827–42.
42. Rae JM, Johnson MD, Scheys JO, Cordero KE, Larios JM, Lippman ME. GREB1 is a critical regulator of hormone dependent breast cancer growth. *Breast Cancer Res Treat.* 2005;92(2):141–9.
43. Choi YJ, Kim NN, Shin HS, Choi CY. The expression of leptin, estrogen receptors, and vitellogenin mRNAs in migrating female chum Salmon, *Oncorhynchus keta*: the effects of hypo-osmotic environmental changes. *Asian Australas J Anim Sci.* 2014;27(4):479.
44. Hartl DL, Clark AG, Clark AG. Principles of population genetics, vol. 116. Sunderland: Sinauer associates; 1997.
45. Mi H, et al. Thomas. PANTHER version 11: expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements. *Nucl. Acids Res.* 2016;45:D183–9.
46. Mizuno K, Kojima Y, Kamisawa H, Moritoki Y, Nishio H, Kohri K, Hayashi Y. Gene expression profile during testicular development in patients with SRY-negative 46, XX testicular disorder of sex development. *Urology.* 2013;82(6): 1453–e1.
47. Chun KH, Araki K, Jee Y, Lee DH, Oh BC, Huang H, et al. Regulation of glucose transport by *rock1* differs from that of *ROCK2* and is controlled by actin polymerization. *Endocrinology.* 2012;153(4):1649–62.
48. Long JZ, Cisar JS, Milliken D, Niessen S, Wang C, Trauger SA, Cravatt BF. Metabolomics annotates *abhd3* as a physiologic regulator of medium-chain phospholipids. *Nat Chem Biol.* 2011;7(11):763–5.
49. Idler DR, Bitners I. Biochemical studies on sockeye salmon during spawning migration: II. Cholesterol, fat, protein, and water in the flesh of standard fish. *Can J Biochem Physiol.* 1958;36(8):793–8.
50. Ralte AM, Sharma MC, Karak AK, Mehta VS, Sarkar C. Clinicopathological features, MIB-1 labeling index and apoptotic index in recurrent astrocytic tumors. *Pathol Oncol Res.* 2001;7(4):267.
51. Farrell AP, Johansen JA, Steffensen JF, Moyes CD, West TG, Suarez RK. Effects of exercise training and coronary ablation on swimming performance, heart size, and cardiac enzymes in rainbow trout, *Oncorhynchus mykiss*. *Can J Zool.* 1990;68(6):1174–9.
52. Luxán G, Casanova JC, Martínez-Poveda B, Prados B, D'amato G, MacGrogan D, Izquierdo-García JL. Mutations in the NOTCH pathway regulator *mib1* cause left ventricular noncompaction cardiomyopathy. *Nat Med.* 2013;19(2):193.
53. Kilpinen H, Waszak SM, Gschwind AR, Raghav SK, Witwicki RM, Orioli A, et al. Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science.* 2013;342(6159):744–7.
54. Downing SL, Prentice EF, Frazier RW, Simonson JE, Nunnallee EP. Technology developed for diverting passive integrated transponder (PIT) tagged fish at hydroelectric dams in the Columbia River basin. *Aquac Eng.* 2001;25(3):149–64.
55. High B, Peery CA, Bennett DH. Temporary staging of Columbia River summer steelhead in coolwater areas and its effect on migration rates. *Trans Am Fish Soc.* 2006;135(2):519–28.

56. Zaugg WS, Adams BL, McLain LR. Steelhead migration: potential temperature effects as indicated by gill adenosine triphosphatase activities. *Science*. 1972;176(4033):415–6.
57. Richter A, Kolmes SA. Maximum temperature limits for Chinook, coho, and chum salmon, and steelhead trout in the Pacific northwest. *Rev Fish Sci*. 2005;13(1):23–49.
58. Poulin B, Lefebvre G, McNeil R. Tropical avian phenology in relation to abundance and exploitation of food resources. *Ecology*. 1992;73(6):2295–309.
59. Berkes F, Colding J, Folke C. Rediscovery of traditional ecological knowledge as adaptive management. *Ecol Appl*. 2000;10(5):1251–62.
60. Garcia de Leaniz C, Fleming IA, Einum S, Verspoor E, Jordan WC, Consuegra S, Webb JH. A critical review of adaptive genetic variation in Atlantic salmon: implications for conservation. *Biol Rev*. 2007;82(2):173–211.
61. Kovach RP, Joyce JE, Echave JD, Lindberg MS, Tallmon DA. Earlier migration timing, decreasing phenotypic variation, and biocomplexity in multiple salmonid species. *PLoS One*. 2013;8(1):e53807.
62. Langin. Salmon spawn fierce debate over protecting endangered species, thanks to a single gene. *Science: Biology Plants & Animals*. (2018). <https://doi.org/10.1126/science.aau0709>.
63. United States Geological Survey (2007–2014). National Hydrography Dataset available on the World Wide Web (<https://nhd.usgs.gov>). Accessed 14 Feb 2018.
64. United States Geological Survey and United States Department of Agriculture, Natural Resources Conservation Service. (2013) . Federal standards and procedures for the National Watershed Boundary Dataset (WBD). 4th ed. 63 p. <http://pubs.usgs.gov/tm/11/a3/>.
65. United States Census Bureau. TIGER/line shapefiles (machinereadable data files) cartographic boundary shapefiles - nation. W,(2015).
66. Fick SE, Hijmans RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol*. 2017;37(12):4302–15.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

