bioRxiv preprint doi: https://doi.org/10.1101/056853; this version posted June 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Title

The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation

Authors and Affiliations

Daniel J. Prince[1,2], Sean M. O'Rourke[1], Tasha Q. Thompson[1], Omar A. Ali[1], Martha Arciniega[3,4], Hannah S. Lyman[1], Ismail K. Saglam[1,5], Anthony J. Clemento[3,4], Thomas J. Hotaling[6], Andrew P. Kinziger[7], Adrian P. Spidle[8], John Carlos Garza[3,4], Devon E. Pearse[3,4], Michael R. Miller[1,2]

[1]Department of Animal Science, University of California, 1 Shields Ave, Davis, CA 95616, USA
[2]Center for Watershed Sciences, University of California, 1 Shields Ave, Davis, CA 95616, USA
[3]Fisheries Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries Service, 110 Shaffer Rd, Santa Cruz, CA 95060, USA
[4]Institute of Marine Sciences, University of California, 1156 High St, Santa Cruz, CA 95060, USA
[5]Ecological Sciences Research Laboratories, Department of Biology, Hacettepe University, Beytepe, Ankara 06800, Turkey
[6]Salmon River Restoration Council, 25631 Sawyers Bar Rd, Sawyers Bar, CA 96027, USA
[7]Department of Fisheries Biology, Humboldt State University, 1 Harpst St, Arcata, CA 95521, USA
[8]Northwest Indian Fisheries Commission, 6730 Martin Way E, Olympia, WA 98516, USA

Corresponding Author

Michael R. Miller
Department of Animal Science
University of California
1 Shields Ave
Davis CA 95618 USA
micmiller@ucdavis.edu
+1-530-304-4719

Abstract

The delineation of conservation units (CUs) is a challenging issue that has profound implications for minimizing the loss of biodiversity and ecosystem services. CU delineation typically seeks to prioritize evolutionary significance and genetic methods play a pivotal role in the delineation process by quantifying overall differentiation between populations. While CUs that primarily reflect overall genetic differentiation do protect adaptive differences between distant populations, they do not necessarily protect adaptive variation within highly connected populations. Advances in genomic methodology facilitate the characterization of adaptive genetic variation, but the potential utility of this information for CU delineation is unclear. Here we use genomic methods to investigate the evolutionary basis of premature migration in Pacific salmon, a complex behavioral and physiological adaptation that exists within highly-connected populations and has experienced severe declines. Strikingly, we find that premature migration is associated with the same single locus across multiple populations in each of two different species. Patterns of variation at this locus suggest that the premature migration alleles arose from a single evolutionary event within each species and were subsequently spread to distant populations through straying and positive selection. Our results reveal that complex adaptive variation can depend on rare mutational events at a single locus, demonstrate that CUs reflecting overall genetic differentiation can fail to protect evolutionarily significant variation that has substantial ecological and societal benefits, and suggest that a supplemental framework for protecting specific adaptive variation will sometimes be necessary to prevent the loss of significant biodiversity and ecosystem services.

bioRxiv preprint doi: https://doi.org/10.1101/056853; this version posted June 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Invaluable economic, ecological, and cultural benefits are being lost worldwide as biodiversity decreases due to human actions (Millennium Ecosystem Assessment 2005; Krauss *et al.* 2010; Pimm *et al.* 2014). Legislation exists in many countries throughout the world that provides a framework to protect unique species as well as population segments below the species level (Allendorf et al. 2012). Protection is achieved by assessing the health of a defined conservation unit and, if the unit is at risk, critical habitat is preserved/restored and stressors are restricted until the risk is eliminated. Assessing risk and developing a protection strategy is not possible without first establishing unit boundaries. Since the number of units that can be effectively managed is resource-limited (Waples 1998), the delineation of units must be strategic and prioritize evolutionary significance (Ryder 1986; Waples 1991; Dizon *et al.* 1992; Moritz 1994; Crandall *et al.* 2000; Fraser and Bernatchez 2001). Several criteria, such as genetic and ecological exchangeability (Crandall *et al.* 2000), have been proposed for assessing evolutionary significance for conservation unit delineation, but directly evaluating these criteria in natural populations is difficult (Allendorf *et al.* 2012).

Genetic methods play a pivotal role in evaluating evolutionary significance for conservation unit delineation (Crandall *et al.* 2000). To this end, genetic data from different regions of the genome are combined to produce measurements of overall genetic differentiation between populations. These measurements represent typical regions of the genome and are used as a proxy for evolutionary significance in conservation unit delineation. However, because most genomic regions are primarily influenced by gene flow and genetic drift as opposed to selection, these measurements may fail to account for important adaptive differences between populations (Funk *et al.* 2012). Recent advances in genetic methodology facilitate the identification and evolutionary analysis of adaptively important loci (Colosimo *et al.* 2005; Miller *et al.* 2007; Baird *et al.* 2008; Hohenlohe *et al.* 2010; Miller *et al.* 2012; Pearse *et al.* 2014; Barson *et al.* 2015; Ali *et al.* 2016) and provide an alternative way to assess evolutionary significance, but the utility of these loci for conservation unit delineation is unclear and disputed (Allendorf *et al.* 2010; Funk *et al.* 2012; Shafer *et al.* 2015, 2016; Garner *et al.* 2016).

Pacific salmon (*Oncorhynchus* spp.) provide a unique opportunity to investigate the application of genetic tools to the conservation of biodiversity below the species level (Waples 1991, 1998; Allendorf *et al.* 1997; Pennock and Dimmick 1997; Wainwright and Waples 1998). Despite extensive conservation efforts, Pacific salmon have been extirpated from almost forty percent of their historical range in the contiguous U.S. while many remaining populations have experienced dramatic declines and face increasing challenges from climate change (Levin and Schiewe 2001; Augerot 2005; Gustafson *et al.* 2007; Moyle *et al.* 2008; Muñoz *et al.* 2015). Reintroduction attempts of extirpated populations are largely unsuccessful as precise natal homing across highly heterogeneous environments has resulted in divergent selection and abundant local adaptation (Taylor 1991; Fraser *et al.* 2006; Miller *et al.* 2012; Anderson *et al.* 2014). Thus, maintaining existing stocks is critical for preserving the species themselves as well as the communities and ecosystems which rely on their presence (Committee on Protection and Management of Pacific Northwest Anadromous Salmonids *et al.* 1996). Genetic methods have been used extensively in the process of delineating conservation units in Pacific salmon (called evolutionarily significant units [ESUs] or distinct population segments [DPSs] depending on the species) and, as a consequence of patterns of gene flow, have resulted in units that primarily reflect geography (Busby *et al.* 1996; USFWS and NOAA 1996; Myers *et al.* 1998; NOAA 2006). While current ESUs and DPSs certainly protect adaptive differences between distant populations, adaptations within highly connected populations are not necessarily protected (Crandall *et al.* 2000; Moyle *et al.* 2008). However, the evolutionary significance of these adaptations and the potential long-term consequences of not independently protecting them are poorly understood.

Perhaps the most recognized example of differential adaptation within highly connected populations of Pacific salmon is variation in adult migration timing (also called run timing) (Moyle

bioRxiv preprint doi: https://doi.org/10.1101/056853; this version posted June 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2002; Quinn 2005; Quinn *et al.* 2015). In contrast to typical adult salmon that mature sexually before freshwater migration, premature migrating individuals have a complex behavioral and physiological adaptation which allows them to access distinct habitats and spawn earlier in the season (Quinn *et al.* 2015). Premature migrating populations also distribute ocean-derived nutrients higher into watersheds and provide additional, more-coveted harvest opportunities due to their distinct migration time and high fat content (Hearsey and Kinziger 2014). Due to their extended time in freshwater and reliance on headwater habitat, premature migrating populations have suffered grossly disproportionate impacts from human actions such as dam building, mining, and logging (Busby *et al.* 1996; Myers *et al.* 1998; Kareiva *et al.* 2000; Committee on Endangered and Threatened Fishes in the Klamath River Basin *et al.* 2004; Moyle *et al.* 2008; Williams *et al.* 2013; Quinn *et al.* 2015). Genetic analyses typically find little differentiation between proximate premature and mature migrating populations (Allendorf 1975; Chilcote *et al.* 1980; Thorgaard 1983; Nielsen and Fountain 1999; Waples *et al.* 2004; Clemento III 2006; Kinziger *et al.* 2013; Brieuc *et al.* 2015; Arciniega *et al.* 2016), and as a result, they are generally grouped into the same ESU or DPS (Busby *et al.* 1996; Myers *et al.* 1998). Therefore, despite the extirpation or substantial decline of premature migrating populations, the ESUs or DPSs to which they belong usually retain relatively healthy mature migrating populations and thus have low extinction risk overall (Busby *et al.* 1996; Myers *et al.* 1998; Williams *et al.* 2013). Here we investigate the genetic and evolutionary basis of premature migration to explore potential consequences of not independently protecting this beneficial adaptation as well as the utility of genomics for informing conservation.

Results

Initial genomic analysis consistent with current steelhead DPS delineations

Dramatic examples of premature migration are observed in coastal (non-interior) populations of steelhead (anadromous rainbow trout; *O. mykiss*) and Chinook salmon (*O. tshawytscha*). In these populations, premature migrating individuals (called summer steelhead or spring Chinook) utilize receding spring flows during freshwater migration to reach upstream habitat prior to hostile summer conditions in the lower watershed, hold for several months in deep cool pools while their gametes mature, then spawn at similar times to mature migrating individuals which have just entered freshwater (Moyle 2002; Quinn *et al.* 2015). We began our investigation by compiling a set of 148 steelhead samples from five coastal locations across four DPSs in California and Oregon (Figure 1A). Four of the locations (Eel, New, Siletz, and North Umpqua) represent the few remaining watersheds with significant wild premature migrating populations. The fifth location, Scott, contains only mature migrating individuals. Our sampling focused as much as possible on individuals that could be confidently categorized as premature or mature migrating based on collection date and location (Figure 1B; Table S1).

To collect high-resolution genomic information from these samples, we prepared individually barcoded restriction-site associated DNA (RAD) libraries, sequenced them using paired-end Illumina technology, and aligned the sequence reads to a recent draft of the rainbow trout genome (Berthelot *et al.* 2014; Table S2, S1). We then used a probabilistic framework to discover SNPs and genotype them in each individual (Korneliussen *et al.* 2014). A total of 9,864,960 genomic positions were interrogated in at least 50% of individuals and 615,958 high-confidence SNPs (p-value $< 10^{-6}$) were identified. Of these, 215,345 SNPs had one genotype posterior greater than 0.8 in at least 50% of the samples. Population structure characterization and genome-wide analyses in nonmodel organisms are typically carried out with far fewer SNPs (Narum *et al.* 2013). We conclude that the sequence data obtained is appropriate for genome-wide measurements as well as high-resolution analyses of specific genomic regions.

To characterize the genetic structure of these populations, we performed a principal component analysis and estimated pairwise F_{ST} using genome-wide genotype data (Fumagalli *et al.* 2014). The first two principal components (PCs) revealed four distinct groups corresponding to the four current DPSs

bioRxiv preprint doi: https://doi.org/10.1101/056853; this version posted June 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Figure 1C). The Siletz and North Umpqua, which are two distinct locations within the Oregon Coast DPS, did not break into distinct groups until PC6 (Figure 1D), indicating relatively low genetic differentiation between distinct locations within a DPS. In all cases, individuals with different migration phenotypes from the same location were in the same group. The pairwise F_{ST} estimates also revealed strong genetic differentiation between locations but little differentiation between migration phenotypes from the same location (Figure 1E). The mean pairwise F_{ST} between migration groups from the same location was 0.032 (range: 0.018 - 0.039, n = 3) whereas the mean between groups from different locations was 0.125 (range: 0.049 - 0.205, n = 25). Thus, as found in previous analyses, the overall genetic structure among steelhead populations is predominantly influenced by geography as opposed to migration phenotype. We conclude that measurements of overall genetic differentiation from genome-wide SNP data are consistent with current steelhead DPS delineations.

Premature migrating steelhead explained by a single genetic evolutionary event at a single locus

To identify genomic loci associated with premature migration, we performed association mapping of migration category. We used a likelihood ratio test which follows a Chi-squared distribution (Kim *et al.* 2011) to compare 181,954 SNPs between migration categories in the North Umpqua and found 14 SNPs that were significant (p-value < Bonferroni-corrected α of 0.001). Strikingly, all of these SNPs were located within a 211,251 bp region (568,978 - 780,229) on a single 1.95 Mbp scaffold (Figure 2A; Table S3). Furthermore, when this analysis was repeated with Eel individuals using 170,678 SNPs, we obtained a virtually identical pattern of association (Figure 2B). The strongest associated SNPs in both sample locations were flanking two restriction sites approximately 50 kb apart and located just upstream and within a gene identified as *GREB1L* (Figure 2C). The strength of these associations was unexpected given the phenotypic complexity of premature migration and the relatively low number of samples analyzed. We conclude that the same single locus is strongly associated with migration phenotype in at least two DPSs.

To investigate the evolutionary history of this locus, we sequenced three amplicons, each of approximately 500 bp, from the *GREB1L* region in all individuals (Figure 2C; Table S4, S5, S1) and used these sequences to construct a gene tree based on parsimony (Felsenstein 1989). Strikingly, the tree contained two distinct monophyletic groups corresponding to migration phenotype (Figure 2D). Every individual separated into the appropriate migration category clade, irrespective of their location of origin. Although there was little differentiation within the mature migration clade, premature migrating individuals from the Siletz and North Umpqua formed a monophyletic group that was nested within the Eel/New clade and more divergent from the mature migration clade. Interestingly, three individuals originally classified as mature migrating were at an intermediate position in the tree (Figure 2D), indicating possible heterozygosity at the causative polymorphism. These individuals tended to be collected earlier and higher in the watershed (mean: Jan. 26; Table S1) than typical mature migrating individuals (mean: Feb. 21; Table S1), suggesting an intermediate phenotype for heterozygotes. The overall tree topology is inconsistent with premature migration alleles originating from independent evolutionary events in different locations. We conclude that there is a complete association between variation at this locus and migration category and that the premature migration alleles from all locations arose from a single evolutionary event.

To examine the evolutionary mechanisms leading to the dispersal of the premature migration allele, we summarized patterns of genetic variation using two estimators of theta ($\theta = 4N\mu$). One estimator is based on average pairwise differences (θ_π; Tajima 1983) and the other is based on number of segregating sites (θ_S; Watterson 1975). When genome-wide data were used, both estimators produced similar thetas for each migration category (Figure 2E). The *GREB1L* region of mature migrating individuals also produced thetas similar to the genome-wide analysis. However, premature migrating individuals from the North Umpqua had strikingly lower thetas (Figure 2E) and a significantly skewed site frequency spectrum (Tajima's D = -2.08, p-value = 0.001; Tajima 1989)

bioRxiv preprint doi: https://doi.org/10.1101/056853; this version posted June 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

indicative of strong, recent positive selection in the *GREB1L* region. Premature migrating individuals from the Eel also had reduced thetas in the *GREB1L* region (premature: θ_π/kb = 2.48, θ_S/kb = 2.67; mature: θ_π/kb = 3.59, θ_S/kb = 4.00), but the site frequency spectrum was not significantly skewed, consistent with an older selection event. While both demography and selection can reduce nucleotide diversity and skew the site frequency spectrum, this pattern is specific to the *GREB1L* region as opposed to genome-wide, implicating selection as the cause. Furthermore, the combination of a stronger signature of selection and a more divergent sequence pattern in the northern premature migration clade, suggests a relatively recent (perhaps since the last glacial maximum) movement of the premature migration allele to the north. We conclude that strong positive selection facilitated dispersal of the premature migration allele after an initial evolutionary event.

Premature migrating Chinook also explained by a single genetic evolutionary event in GREB1L region

To broaden our investigation into premature migration, we compiled a set of 250 Chinook samples from nine locations across five ESUs in California, Oregon, and Washington (Figure 3A). Similar to steelhead, our sampling focused as much as possible on individuals that could be confidently categorized as premature or mature migrating based on collection time and location (Table S6). We then prepared individually barcoded RAD libraries, sequenced them using paired-end Illumina technology, and aligned the sequence reads to the same rainbow trout reference assembly used above (Table S7, S6). No reference genome is available for Chinook, and rainbow trout, which diverged from Chinook approximately 10-15 mya (Crête-Lafrenière *et al.* 2012; Zhivotovsky 2015), is the closest relative with a draft genome assembly. Using the methods described above, a total of 3,910,009 genomic positions were interrogated in at least 50% of the samples and 301,562 high-confidence SNPs (p-value < 10^{-6}) were identified. Of these, 55,797 SNPs had one genotype posterior greater than 0.8 in at least 50% of the samples. Although the alignment success was lower and subsequent SNP discovery and genotyping produced fewer SNPs compared to steelhead, the large number of SNPs discovered and genotyped should still be adequate for downstream analysis.

To characterize the genetic structure of these populations, we performed a principal component analysis and estimated pairwise F_{ST} using the genotype information described above. The first two PCs revealed four groups: the largest group contained all coastal ESUs, the second contained the two Puget Sound ESU locations, and the last two groups corresponded to the two locations within the Upper Klamath-Trinity Rivers ESU and were only differentiated by the second axis (Figure 3B). In all cases, individuals from the same location but with different migration phenotypes were in the same group, and locations within groups became differentiated as additional PCs were examined (data not shown). The mean pairwise F_{ST} between migration categories from the same location was 0.037 (range: 0.009 - 0.093, n = 7) and the mean between groups from different locations was 0.097 (range: 0.021 - 0.199, n = 113; Figure 3C). Thus, similar to what we found in steelhead, the overall genetic structure is strongly influenced by geography as opposed to migration phenotype. We conclude that measurements of overall genetic differentiation from genome-wide SNP data are consistent with current Chinook ESUs.

To investigate the genetic architecture and evolutionary basis of premature migration in Chinook, we conducted association mapping using the methods described above with 113,435 SNPs. Strikingly, we again found a single significant peak of association (p-value < Bonferroni-corrected α of 0.001) that contained 4 SNPs within 57,380 bp (537,741 – 595,121) in the same *GREB1L* region identified in steelhead (Figure 3D; Table S8). We next examined allele frequencies at these four SNPs and found a strong and consistent shift between all premature and mature migrating populations independent of location (Figure 3E). Thus, despite having lower genomic resolution and fewer samples per location, these results clearly demonstrate that the *GREB1L* region is also the primary locus associated with premature migration in Chinook. Furthermore, the shift of allele frequencies in the same direction across all populations is inconsistent with the premature migration alleles in Chinook being a product of multiple independent evolutionary events. Although the genomic region was

bioRxiv preprint doi: https://doi.org/10.1101/056853; this version posted June 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

consistent between species, the SNPs identified in Chinook were distinct from those in steelhead (Table S8, S3). In other words, the premature and mature migrating Chinook haplotypes are more similar to each other than to either of the steelhead haplotypes and vice versa, suggesting independent genetic evolutionary events in each species. We conclude that the same evolutionary mechanism used in steelhead, with a single genetic evolutionary event in the *GREB1L* region that subsequently spread to different locations, also explains premature migration in Chinook.

Discussion

Our association analysis across multiple populations in each of two different species as well as an independent analysis on an additional steelhead population (Hess *et al.* 2016) suggests that either the function or regulation of *GREB1L* is modified in premature migrating individuals. *GREB1L* is highly conserved across vertebrates, encodes a nuclear hormone receptor coactivator, and is differentially regulated by feeding and fasting in AgRP neurons of the hypothalamic arcuate nucleus in mice (Mohammed *et al.* 2013; Henry *et al.* 2015). The strength of the associations as well as the known role of AgRP neurons in modulating diverse behavior and metabolic processes such as foraging and fat storage (Dietrich *et al.* 2015; Henry *et al.* 2015) provides evidence for and an explanation as to how the complex premature migration phenotype could be controlled by this single locus. An alternative explanation is that the *GREB1L* region only influences a subset of the phenotypic components of premature migration and that other important loci were not identified due to technical or biological reasons. Regardless, our results indicate that an appropriate genotype at this locus is necessary for successful premature migration.

Given that premature migration alleles at this locus are critical for premature migration, our results on the evolutionary history of these alleles provide important insights into the potential for premature migration to persist during declines and reemerge if lost. Finding that the same locus is associated with premature migration in both steelhead and Chinook indicates that genetic mechanisms capable of producing this phenotype are very limited. Furthermore, our observation of only two genetic evolutionary events that produced a premature migration allele in the 20-30 million years of evolutionary time between steelhead and Chinook (Crête-Lafrenière *et al.* 2012; Zhivotovsky 2015) suggests that mutational events creating new premature migration alleles are exceptionally rare. The combination of these results indicates that the premature migration allele and phenotype cannot be expected to readily re-evolve if the alleles are lost.

The rarity of mutational events that produce premature migration alleles at this locus highlights the importance of existing premature migration alleles. In addition to their presence in premature migrating individuals, premature migration alleles can exist in the heterozygous state. However, two lines of evidence suggest that premature migration alleles could not be maintained in the heterozygous state. First, the observed bimodal phenotypic distribution and our results in support of an intermediate heterozygous phenotype suggests some level of selection against heterozygotes (Moyle 2002; Quinn *et al.* 2015). Second, the genetic differentiation between premature migration alleles from California and Oregon steelhead indicates that, unlike mature migration alleles, premature migration alleles are not freely moving across this area. This result reveals that mature migrating populations do not act as an influential source or conduit of premature migration alleles despite being abundant and broadly distributed and provides further evidence for selection against heterozygotes. Therefore, although a broad scale survey of these alleles would be useful to better understand their current distribution, the evidence strongly suggests that premature migrating populations are ultimately necessary for the spread and maintenance of these alleles.

Previously, studies revealing that overall genetic structure among populations of steelhead and Chinook primarily reflects geography (as opposed to migration phenotype) suggested that premature migration evolved independently in many locations within each species (Thorgaard 1983; Waples *et al.* 2004). This implied that premature migration is evolutionarily replaceable over time frames relevant to

bioRxiv preprint doi: https://doi.org/10.1101/056853; this version posted June 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

conservation planning (e.g. tens to hundreds of years; Waples *et al.* 2004) and is not an important component in the evolutionary legacy of the species (Williams *et al.* 2013). Although these interpretations were logical given the data available at that time, our results provide novel insight into the evolutionary basis of premature migration and have direct conservation implications. Our results demonstrate that the evolution was not independent in each location but instead relied on pre-existing genetic variation, and thus, suggest that the widespread extirpation and decline of premature migrating populations has greatly diminished the potential restoration and expansion (e.g. into new habitats that become available with climate change) of premature migration across at least a substantial proportion of the range for both species (Miller *et al.* 2012).

The combination of three key results from this study has broad conservation implications which highlight the utility of genomics for informing conservation. First, we present a striking example of how a single allele at a single locus can have significant economic, ecological, and cultural importance. Second, we show that mutations producing an important allele can be very rare from an evolutionary perspective, suggesting the allele will not readily re-evolve if lost. Lastly, we observe that patterns of significant adaptive allelic variation can be completely opposite from patterns of overall genetic differentiation. Taken together, our results demonstrate that conservation units reflecting overall genetic differentiation can fail to protect evolutionarily significant variation that has substantial ecological and societal benefits, and suggest that a supplemental framework for protecting specific adaptive variation will sometimes be necessary to prevent the loss of significant biodiversity and ecosystem services.

Methods

Sample collection and molecular biology

Fin clips were taken from live adults or post-spawn carcasses (Table S1, S6), dried on Whatman qualitative filter paper (Grade 1), and stored at room temperature. DNA was extracted with either the DNeasy Blood & Tissue Kit (Qiagen) or a magnetic bead-based protocol (Ali *et al.* 2016) and quantified using Quant-iT PicoGreen dsDNA Reagent (Thermo Fisher Scientific) with an FLx800 Fluorescence Reader (BioTek Instruments).

SbfI RAD libraries were prepared with well and plate (when applicable) barcodes using either the traditional or new RAD protocol (Ali *et al.* 2016) and sequenced with paired-end 100 bp reads on an Illumina Hiseq 2500 (Table S2, S7). In some cases, the same sample was included in multiple libraries to improve sequencing coverage.

For amplicon sequencing, genomic DNA extractions were rearrayed into 96-well plates and diluted 1:40 with low TE pH 8.0 (10 mM Tris-HCl, 0.1 mM EDTA). 2 µl of this diluted sample was used as PCR template for each of three amplicons in the *GREB1L* region (Figure 2; Table S4). Multiple forward primers were synthesized for each amplicon. Each forward primer contained a partial Illumina adapter sequence, a unique inline plate barcode, and the amplicon specific sequence (Table S4, S5). Initial PCR reactions were performed in 96-well plates using OneTaq DNA polymerase (New England Biolabs) at the recommended conditions with an annealing temperature of 61°C and 35 cycles. These reaction plates were then combined into a single plate that preserved the well locations. The pooled PCR products were cleaned with Ampure XP beads (Beckman Coulter) and a second round of PCR with 8 cycles was performed to add the remaining Illumina adapter sequence and a unique TruSeq barcode to each well (Table S4, S5). From each final PCR reaction, 2 µl was removed, pooled, and purified with Ampure XP beads. The final amplicon library was sequenced with paired-end 300 bp reads on an Illumina MiSeq.

RAD analysis

RAD sequencing data were demultiplexed by requiring a perfect barcode and partial restriction site match (Ali *et al.* 2016). Sequences were aligned to a slightly modified version of a recent rainbow trout genome assembly (Berthelot *et al.* 2014; see scaffold79929e assembly and annotation) using the

bioRxiv preprint doi: https://doi.org/10.1101/056853; this version posted June 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

backtrack algorithm of BWA (Li and Durbin 2009) with default parameters. SAMtools (Li *et al.* 2009) was used to sort, filter for proper pairs, remove PCR duplicates, and index BAM files (Table S2, S7). In cases where the same sample was sequenced in multiple libraries, BAM files from the same sample were merged prior to indexing using SAMtools (Table S1, S2, S6, S7).

Additional BAM file sets were generated to account for technical variation among samples. To minimize variation associated with the two distinct library preparation protocols used in Chinook (Ali *et al.* 2016; Table S7), a set of single-end BAM files were generated for Chinook that contained only trimmed reads from the restriction-site end of the RAD fragments. To prepare these files, these reads were trimmed to 75 bp from the 3' end after removing 5 bp from the 5' end. Next, paired-end alignments were performed and processed as above. Lastly, reads from the variable end of RAD fragments were removed (Table S7). To remove variation associated with variable sequencing depth, a set of subsampled BAM files were generated by using SAMtools to randomly sample approximately 120,000 alignments from paired-end BAM files for steelhead and approximately 60,000 alignments from single-end BAM files for Chinook.

All RAD analyses were performed using ANGSD (Korneliussen *et al.* 2014) with a minimum mapping quality score (minMapQ) of 10, a minimum base quality score (minQ) of 20, and the SAMtools genotype likelihood model (GL 1; Li 2011). Unless otherwise noted, samples with less alignments than required for subsampling were excluded (Table S1, S6), and only sites represented in at least 50% of the included samples (minInd) were used.

Principal component analyses (PCA) and association mapping were performed by identifying polymorphic sites (SNP_pval 1e-6), inferring major and minor alleles (doMajorMinor 1; Skotte *et al.* 2012), estimating allele frequencies (doMaf 2; Kim *et al.* 2011), and retaining SNPs with a minor allele frequency of at least 0.05 (minMaf). For PCA, subsampled BAM files were used and genotype posterior probabilities were calculated with a uniform prior (doPost 2). The ngsCovar (Fumagalli *et al.* 2013) function implemented in ngsTools (Fumagalli *et al.* 2014) was used to calculate a covariance matrix from called genotypes. For association mapping, paired-end BAM files and a frequency test with known major and minor alleles (doAsso 1; Kim *et al.* 2011) were used. The number of Chinook samples per population was highly variable, so to reduce location specific effects, the analysis used no more than 10 samples from each population.

Genome-wide F_{ST} between population pairs was estimated by first estimating a site frequency spectrum (SFS) for each population (doSaf; Nielsen *et al.* 2012) using paired-end BAM files for steelhead and single-end BAM files for Chinook. Two-dimensional SFS and global F_{ST} (weighted) between each population pair were then estimated using realSFS (Korneliussen *et al.* 2014).

To calculate Watterson's theta (Watterson 1975), Tajima's theta (Tajima 1983), and Tajima's D (Tajima 1989), SFS estimated as described above were used as priors (pest) with paired-end BAM files to calculate each statistic for each site (doThetas) which were averaged to obtain a single value for each statistic (Korneliussen *et al.* 2013). The analysis was restricted to 565,000 – 785,000 bp of scaffold79929e for the *GREB1L* region analysis.

The coalescent simulation program ms (Hudson 2002) was used to determine 95% confidence intervals for the theta estimates from 10,000 simulations under a neutral demographic model. The input number of chromosomes was equal to the number of individuals used to calculate the theta statistics. For genome-wide confidence intervals, 100 independent loci and an input theta of 1, which is the approximate theta of a single RAD tag, were used. For the *GREB1L* region confidence intervals, a single locus and the empirical theta estimates were used. The significance of the empirical Tajima's D value was evaluated by generating a Tajima's D distribution from 10,000 ms simulations under a neutral demographic model. A single locus and the average between empirical values of Watterson's and Tajima's thetas in the *GREB1L* region were used. A Tajima's D distribution was also generated using the extremes of the theta confidence intervals and the empirical value remained significant.

Allele frequencies were estimated (doMaf 1; Kim *et al.* 2011) for the significant Chinook SNPs

bioRxiv preprint doi: https://doi.org/10.1101/056853; this version posted June 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

in each population that had at least 4 individuals with enough alignments for subsampling. Paired-end BAM files were used with the reference genome assembly as the prespecified major allele (doMajorMinor 4). Since some populations had low samples sizes, all samples were included regardless of alignment number.

Amplicon analysis

Amplicon sequence data were demultiplexed by requiring perfect barcode and primer matches. Sequences were aligned to the reference genome assembly described above using the BWA-SW algorithm (Li and Durbin 2010) with default parameters and SAMtools was used to sort, filter for proper pairs, and index BAM files (Table S5).

Phylogenetic analysis was performed on samples in which two or more amplicons had at least 20 alignments (Table S1, S5). Genotypes for all sites were called using ANGSD with the SAMtools genotype likelihood model, a uniform prior, and a posterior cutoff of 0.8. The genotype output file was parsed and converted into biallelic consensus sequences with IUPAC nucleotide code denoting heterozygous positions. These consensus sequences were used as an input for RRHS (Lischer *et al.* 2014) to produce 1000 output files that each contained one randomly generated haplotype sequence for each sample. Parsimony trees were constructed for each RRHS output and a consensus tree was called using PHYLIP (Felsenstein 1989).

In the initial phylogentic analysis, one sample from the Eel River that was originally classified as premature migrating clustered in the mature migration clade (Table S1). A PCA analysis specific to the Eel River placed this sample at an intermediate position between mature migrating and premature migrating sample groups. Furthermore, this was the only Eel River sample that was homozygous for a haplotype associated with residency (Pearse *et al.* 2014). Examination of the original sampling information revealed that this fish was much smaller than others and collected upstream from the main premature steelhead holding area (Clemento III 2006), indicating that it was a resident trout as opposed to an anadromous steelhead. Therefore, this sample was removed and the analysis was rerun.

Scaffold79929e assembly and annotation

Our initial RAD analysis was aligned against a published reference genome assembly (Berthelot *et al.* 2014) and identified highly associated SNPs on three independent scaffolds. Given the state of the assembly, the sizes of the scaffolds with highly associated SNPs, and the positions of the highly associated SNPs on the scaffolds, we hypothesized that these scaffolds might be physically linked despite not being connected in the current assembly. We aligned four large-insert mate-pair libraries to the published assembly to look for linkages and estimate the distance between linked scaffolds (Table S9). A perfect sequence match was required and alignments to regions with high coverage were discarded. The resulting alignments from all libraries strongly supported a linear assembly with a total size of 1,949,089 bp that included the three associated scaffolds as well as four others (Table S10, S9). This assembled scaffold was named scaffold79929e (e for extended), added to the published assembly, and the seven independent scaffolds that composed it were removed to create the modified reference assembly used in this study.

Scaffold79929e was annotated with MAKER (Cantarel *et al.* 2008) using rainbow trout and Atlantic salmon (*Salmo salar*) EST sequences from the NCBI database, the UniProt/Swiss-Prot database for protein homology, a rainbow trout repeat library (Berthelot *et al.* 2014) for masking, AUGUSTUS (human) and SNAP (mamiso) gene predictors, a maximum intron size of 20,000 bp for evidence alignments, and otherwise default parameters.

Acknowledgments

The authors thank Chong Liu, Josh Saxon, Peter Tronquet, Oregon Department of Fish and Wildlife, and California Department of Fish and Wildlife for assistance with sample acquisition as well as Fred

bioRxiv preprint doi: https://doi.org/10.1101/056853; this version posted June 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Allendorf, Jamie Ashander, and Robin Waples for valuable comments on an earlier version of the manuscript. M.R.M. thanks Chris Doe, Yniv Palti, and Gary Thorgaard for invaluable support and encouragement. This work used the Vincent J. Coates Genomics Sequencing Laboratory at the University of California at Berkeley, supported by NIH S10 Instrumentation grants S10RR029668 and S10RR027303.

References

Ali O. A., O'Rourke S. M., Amish S. J., Meek M. H., Luikart G., Jeffres C., Miller M. R., 2016 RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping. Genetics **202**: 389–400.

Allendorf F. W., 1975 Genetic Variability in a Species Possessing Extensive Gene Duplication: Genetic Interpretation of Duplicate Loci and Examination of Genetic Variation in Populations of Rainbow Trout.

Allendorf F. W., Bayles D., Bottom D. L., Currens K. P., Frissell C. A., Hankin D., Lichatowich J. A., Nehlsen W., Trotter P. C., Williams T. H., 1997 Prioritizing Pacific Salmon Stocks for Conservation. Conserv. Biol. **11**: 140–152.

Allendorf F. W., Hohenlohe P. A., Luikart G., 2010 Genomics and the future of conservation genetics. Nat. Rev. Genet. **11**: 697–709.

Allendorf F. W., Luikart G. H., Aitken S. N., 2012 *Conservation and the Genetics of Populations*. Wiley-Blackwell, Hoboken.

Anderson J. H., Pess G. R., Carmichael R. W., Ford M. J., Cooney T. D., Baldwin C. M., McClure M. M., 2014 Planning Pacific Salmon and Steelhead Reintroductions Aimed at Long-Term Viability and Recovery. North Am. J. Fish. Manag. **34**: 72–93.

Arciniega M., Clemento A. J., Miller M. R., Peterson M., Garza J. C., Pearse D. E., 2016 Parallel evolution of the summer steelhead ecotype in multiple populations from Oregon and Northern California. Conserv. Genet. **17**: 165–175.

Augerot X., 2005 Atlas of Pacific Salmon. Integr. Comp. Biol. **45**: 952–952.

Baird N. A., Etter P. D., Atwood T. S., Currey M. C., Shiver A. L., Lewis Z. A., Selker E. U., Cresko W. A., Johnson E. A., 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS One **3**: e3376.

Barson N. J., Aykanat T., Hindar K., Baranski M., Bolstad G. H., Fiske P., Jacq C., Jensen A. J., Johnston S. E., Karlsson S., Kent M., Moen T., Niemelä E., Nome T., Næsje T. F., Orell P., Romakkaniemi A., Sægrov H., Urdal K., Erkinaro J., Lien S., Primmer C. R., 2015 Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. Nature **528**: 405–408.

Berthelot C., Brunet F., Chalopin D., Juanchich A., Bernard M., Noël B., Bento P., Da Silva C., Labadie K., Alberti A., Aury J.-M., Louis A., Dehais P., Bardou P., Montfort J., Klopp C., Cabau C., Gaspin C., Thorgaard G. H., Boussaha M., Quillet E., Guyomard R., Galiana D., Bobe J., Volff J.-N., Genêt C., Wincker P., Jaillon O., Crollius H. R., Guiguen Y., 2014 The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nat. Commun. **5**: 3657.

Brieuc M. S. O., Ono K., Drinan D. P., Naish K. A., 2015 Integration of Random Forest with population-based outlier analyses provides insight on the genomic basis and evolution of run timing in Chinook salmon (Oncorhynchus tshawytscha). Mol. Ecol. **24**: 2729–2746.

Busby P. J., Wainwright T. C., Bryant G. J., Lierheimer L. J., Waples R. S., Waknitz F. W., Lagomarsino I. V., 1996 *Status review of west coast steelhead from Washington, Idaho, Oregon, and California*. US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service.

Cantarel B. L., Korf I., Robb S. M. C., Parra G., Ross E., Moore B., Holt C., Alvarado A. S., Yandell M., 2008 MAKER: An easy-to-use annotation pipeline designed for emerging model organism

bioRxiv preprint doi: https://doi.org/10.1101/056853; this version posted June 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

genomes. Genome Res. **18**: 188–196.

Chilcote M. W., Crawford B. A., Leider S. A., 1980 A Genetic Comparison of Sympatric Populations of Summer and Winter Steelheads. Trans. Am. Fish. Soc. **109**: 203–206.

Clemento III A. J., 2006 Subpopulation Structure of Steelhead Trout (Oncorhynchus mykiss) in the Middle Fork Eel River as Determined by Microsatellite DNA Polymorphisms.

Colosimo P. F., Hosemann K. E., Balabhadra S., Villarreal G., Dickson M., Grimwood J., Schmutz J., Myers R. M., Schluter D., Kingsley D. M., 2005 Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. Science **307**: 1928–1933.

Committee on Endangered and Threatened Fishes in the Klamath River Basin, Board on Environmental Studies and Toxicology, Division on Earth and Life Studies, National Research Council, 2004 *Endangered and Threatened Fishes in the Klamath River Basin:: Causes of Decline and Strategies for Recovery*. National Academies Press, Washington, D.C.

Committee on Protection and Management of Pacific Northwest Anadromous Salmonids, National Research Council, Division on Earth and Life Studies, Commission on Life Sciences, 1996 *Upstream:: Salmon and Society in the Pacific Northwest*. National Academies Press, Washington, D.C.

Crandall K. A., Bininda-Emonds O. R. P., Mace G. M., Wayne R. K., 2000 Considering evolutionary processes in conservation biology. Trends Ecol. Evol. **15**: 290–295.

Crête-Lafrenière A., Weir L. K., Bernatchez L., 2012 Framing the Salmonidae Family Phylogenetic Portrait: A More Complete Picture from Increased Taxon Sampling. PLoS ONE **7**: e46662.

Dietrich M. O., Zimmer M. R., Bober J., Horvath T. L., 2015 Hypothalamic Agrp Neurons Drive Stereotypic Behaviors beyond Feeding. Cell **160**: 1222–1232.

Dizon A. E., Lockyer C., Perrin W. F., Demaster D. P., Sisson J., 1992 Rethinking the stock concept: a phylogeographic approach. Conserv. Biol. **6**: 24–36.

Felsenstein J., 1989 PHYLIP-phylogeny inference package (version 3.2). Cladistics **5**: 163–166.

Fraser D. J., Bernatchez L., 2001 Adaptive evolutionary conservation: towards a unified concept for defining conservation units. Mol. Ecol. **10**: 2741–2752.

Fraser D. J., Jones M. W., McParland T. L., Hutchings J. A., 2006 Loss of historical immigration and the unsuccessful rehabilitation of extirpated salmon populations. Conserv. Genet. **8**: 527–546.

Fumagalli M., Vieira F. G., Korneliussen T. S., Linderoth T., Huerta-Sánchez E., Albrechtsen A., Nielsen R., 2013 Quantifying population genetic differentiation from next-generation sequencing data. Genetics **195**: 979–992.

Fumagalli M., Vieira F. G., Linderoth T., Nielsen R., 2014 ngsTools: methods for population genetics analyses from next-generation sequencing data. Bioinforma. Oxf. Engl. **30**: 1486–1487.

Funk W. C., McKay J. K., Hohenlohe P. A., Allendorf F. W., 2012 Harnessing genomics for delineating conservation units. Trends Ecol. Evol. **27**: 489–496.

Garner B. A., Hand B. K., Amish S. J., Bernatchez L., Foster J. T., Miller K. M., Morin P. A., Narum S. R., O'Brien S. J., Roffler G., Templin W. D., Sunnucks P., Strait J., Warheit K. I., Seamons T. R., Wenburg J., Olsen J., Luikart G., 2016 Genomics in Conservation: Case Studies and Bridging the Gap between Data and Application. Trends Ecol. Evol. **31**: 81–83.

Gustafson R. G., Waples R. S., Myers J. M., Weitkamp L. A., Bryant G. J., Johnson O. W., Hard J. J., 2007 Pacific Salmon Extinctions: Quantifying Lost and Remaining Diversity. Conserv. Biol. **21**: 1009–1020.

Hearsey J. W., Kinziger A. P., 2014 Diversity in sympatric chinook salmon runs: timing, relative fat content and maturation. Environ. Biol. Fishes **98**: 413–423.

Henry F. E., Sugino K., Tozer A., Branco T., Sternson S. M., 2015 Cell type-specific transcriptomics of hypothalamic energy-sensing neuron responses to weight-loss. eLife **4**.

Hess J. E., Zendt J. S., Matala A. R., Narum S. R., 2016 Genetic basis of adult migration timing in anadromous steelhead discovered through multivariate association testing. Proc R Soc B **283**:

bioRxiv preprint doi: https://doi.org/10.1101/056853; this version posted June 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

20153064.

Hohenlohe P. A., Bassham S., Etter P. D., Stiffler N., Johnson E. A., Cresko W. A., others, 2010 Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet **6**: e1000862.

Hudson R. R., 2002 Generating samples under a Wright–Fisher neutral model of genetic variation. Bioinformatics **18**: 337–338.

Kareiva P., Marvier M., McClure M., 2000 Recovery and Management Options for Spring/Summer Chinook Salmon in the Columbia River Basin. Science **290**: 977–979.

Kim S. Y., Lohmueller K. E., Albrechtsen A., Li Y., Korneliussen T., Tian G., Grarup N., Jiang T., Andersen G., Witte D., Jorgensen T., Hansen T., Pedersen O., Wang J., Nielsen R., 2011 Estimation of allele frequency and association mapping using next-generation sequencing data. BMC Bioinformatics **12**: 231.

Kinziger A. P., Hellmair M., Hankin D. G., Garza J. C., 2013 Contemporary Population Structure in Klamath River Basin Chinook Salmon Revealed by Analysis of Microsatellite Genetic Data. Trans. Am. Fish. Soc. **142**: 1347–1357.

Korneliussen T. S., Moltke I., Albrechtsen A., Nielsen R., 2013 Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. BMC Bioinformatics **14**: 289.

Korneliussen T. S., Albrechtsen A., Nielsen R., 2014 ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics **15**.

Krauss J., Bommarco R., Guardiola M., Heikkinen R. K., Helm A., Kuussaari M., Lindborg R., Ockinger E., Pärtel M., Pino J., Pöyry J., Raatikainen K. M., Sang A., Stefanescu C., Teder T., Zobel M., Steffan-Dewenter I., 2010 Habitat fragmentation causes immediate and time-delayed biodiversity loss at different trophic levels. Ecol. Lett. **13**: 597–605.

Levin P., Schiewe M., 2001 Preserving Salmon Biodiversity. Am. Sci. **89**: 220.

Li H., Durbin R., 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma. Oxf. Engl. **25**: 1754–1760.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., 1000 Genome Project Data Processing Subgroup, 2009 The Sequence Alignment/Map format and SAMtools. Bioinforma. Oxf. Engl. **25**: 2078–2079.

Li H., Durbin R., 2010 Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinforma. Oxf. Engl. **26**: 589–595.

Li H., 2011 A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics **27**: 2987–2993.

Lischer H. E. L., Excoffier L., Heckel G., 2014 Ignoring heterozygous sites biases phylogenomic estimates of divergence times: implications for the evolutionary history of microtus voles. Mol. Biol. Evol. **31**: 817–831.

Millennium Ecosystem Assessment, 2005 *Ecosystems and human well-being*. Island Press Washington, DC.

Miller M. R., Dunham J. P., Amores A., Cresko W. A., Johnson E. A., 2007 Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. **17**: 240–248.

Miller M. R., Brunelli J. P., Wheeler P. A., Liu S., Rexroad C. E., Palti Y., Doe C. Q., Thorgaard G. H., 2012 A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. Mol. Ecol. **21**: 237–249.

Mohammed H., D'Santos C., Serandour A. A., Ali H. R., Brown G. D., Atkins A., Rueda O. M., Holmes K. A., Theodorou V., Robinson J. L. L., Zwart W., Saadi A., Ross-Innes C. S., Chin S.-F., Menon S., Stingl J., Palmieri C., Caldas C., Carroll J. S., 2013 Endogenous purification

bioRxiv preprint doi: https://doi.org/10.1101/056853; this version posted June 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

reveals GREB1 as a key estrogen receptor regulatory factor. Cell Rep. **3**: 342–349.

Moritz C., 1994 Defining "evolutionarily significant units" for conservation. Trends Ecol. Evol. **9**: 373–375.

Moyle P. B., 2002 *Inland Fishes of California*. University of California Press, Berkeley.

Moyle P. B., Israel J. A., Purdy S. E., 2008 *Salmon, Steelhead, and Trout in California: Status of an Emblematic Fauna*. Center for Watershed Sciences, University of California, Davis.

Muñoz N. J., Farrell A. P., Heath J. W., Neff B. D., 2015 Adaptive potential of a Pacific salmon challenged by climate change. Nat. Clim. Change **5**: 163–166.

Myers J. M., Kope R. G., Bryant G. J., Teel D., Lierheimer L. J., Wainwright T. C., Grant W. S., Waknitz F. W., Neely K., Lindley S. T., Waples R. S., 1998 *Status review of chinook salmon from Washington, Idaho, Oregon, and California*. US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service.

Narum S. R., Buerkle C. A., Davey J. W., Miller M. R., Hohenlohe P. A., 2013 Genotyping-by-sequencing in ecological and conservation genomics. Mol. Ecol. **22**: 2841–2847.

Nielsen J. L., Fountain M. C., 1999 Microsatellite diversity in sympatric reproductive ecotypes of Pacific steelhead (Oncorhynchus mykiss) from the Middle Fork Eel River, California. Ecol. Freshw. Fish **8**: 159–168.

Nielsen R., Korneliussen T., Albrechtsen A., Li Y., Wang J., 2012 SNP Calling, Genotype Calling, and Sample Allele Frequency Estimation from New-Generation Sequencing Data. PLoS ONE **7**: e37558.

NOAA, 2006 Endangered and Threatened Species: Final Listing Determinations for 10 Distinct Population Segments of West Coast Steelhead. Fed. Regist. **71**: 834–862.

Pearse D. E., Miller M. R., Abadía-Cardoso A., Garza J. C., 2014 Rapid parallel evolution of standing variation in a single, complex, genomic region is associated with life history in steelhead/rainbow trout. Proc. Biol. Sci. **281**: 20140012.

Pennock D. S., Dimmick W. W., 1997 Critique of the Evolutionarily Significant Unit as a Definition for "Distinct Population Segments" under the U.S. Endangered Species Act. Conserv. Biol. **11**: 611–619.

Pimm S. L., Jenkins C. N., Abell R., Brooks T. M., Gittleman J. L., Joppa L. N., Raven P. H., Roberts C. M., Sexton J. O., 2014 The biodiversity of species and their rates of extinction, distribution, and protection. Science **344**: 1246752.

Quinn T., 2005 *The Behavior and Ecology of Pacific Salmon and Trout*. University of Washington Press, Bethesda, Md.

Quinn T. P., McGinnity P., Reed T. E., 2015 The paradox of "premature migration" by adult anadromous salmonid fishes: Patterns and hypotheses. Can. J. Fish. Aquat. Sci.

Ryder O. A., 1986 Species conservation and systematics: the dilemma of subspecies. Trends Ecol. Evol. **1**: 9–10.

Shafer A. B. A., Wolf J. B. W., Alves P. C., Bergström L., Bruford M. W., Brännström I., Colling G., Dalén L., Meester L. D., Ekblom R., Fawcett K. D., Fior S., Hajibabaei M., Hill J. A., Hoezel A. R., Höglund J., Jensen E. L., Krause J., Kristensen T. N., Krützen M., McKay J. K., Norman A. J., Ogden R., Österling E. M., Ouborg N. J., Piccolo J., Popović D., Primmer C. R., Reed F. A., Roumet M., Salmona J., Schenekar T., Schwartz M. K., Segelbacher G., Senn H., Thaulow J., Valtonen M., Veale A., Vergeer P., Vijay N., Vilà C., Weissensteiner M., Wennerström L., Wheat C. W., Zieliński P., 2015 Genomics and the challenging translation into conservation practice. Trends Ecol. Evol. **30**: 78–87.

Shafer A. B. A., Wolf J. B. W., Alves P. C., Bergström L., Colling G., Dalén L., De Meester L., Ekblom R., Fior S., Hajibabaei M., Hoezel A. R., Hoglund J., Jensen E. L., Krützen M., Norman A. J., Österling E. M., Ouborg N. J., Piccolo J., Primmer C. R., Reed F. A., Roumet M., Salmona J., Schwartz M. K., Segelbacher G., Thaulow J., Valtonen M., Vergeer P., Weissensteiner M.,

bioRxiv preprint doi: https://doi.org/10.1101/056853; this version posted June 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Wheat C. W., Vilà C., Zieliński P., 2016 Reply to Garner et al. Trends Ecol. Evol. **31**: 83–84.
Skotte L., Korneliussen T. S., Albrechtsen A., 2012 Association testing for next-generation sequencing data using score statistics. Genet. Epidemiol. **36**: 430–437.
Tajima F., 1983 Evolutionary Relationship of Dna Sequences in Finite Populations. Genetics **105**: 437–460.
Tajima F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**: 585–595.
Taylor E. B., 1991 A review of local adaptation in Salmonidac, with particular reference to Pacific and Atlantic salmon. Aquaculture **98**: 185–207.
Thorgaard G. H., 1983 Chromosomal Differences among Rainbow Trout Populations. Copeia **1983**: 650–662.
USFWS, NOAA, 1996 Policy regarding the recognition of distinct vertebrate population segments under the Endangered Species Act. Fed. Regist. **61**: 4722–4725.
Wainwright T. C., Waples R. S., 1998 Prioritizing Pacific Salmon Stocks for Conservation: Response to Allendorf et al. Conserv. Biol. **12**: 1144–1147.
Waples R. S., 1991 Pacific salmon, Oncorhynchus spp., and the definition of "species" under the Endangered Species Act. Mar. Fish. Rev. **53**: 11–22.
Waples R. S., 1998 Evolutionarily Significant Units, Distinct Population Segments, and the Endangered Species Act: Reply to Pennock and Dimmick. Conserv. Biol. **12**: 718–721.
Waples R. S., Teel D. J., Myers J. M., Marshall A. R., 2004 Life-history divergence in Chinook salmon: historic contingency and parallel evolution. Evol. Int. J. Org. Evol. **58**: 386–403.
Watterson G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7**: 256–276.
Williams T. H., Garza J. C., Hetrick N. J., Lindley S. T., Mohr M. S., Myers J. M., O'Farrell M. R., Quiñones R. M., 2013 *Upper Klamath and Trinity river Chinook Salmon biological review team report*. US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Southwest Fisheries Science Center.
Zhivotovsky L. A., 2015 Genetic history of salmonid fishes of the genus Oncorhynchus. Russ. J. Genet. **51**: 491–505.

Figure legends

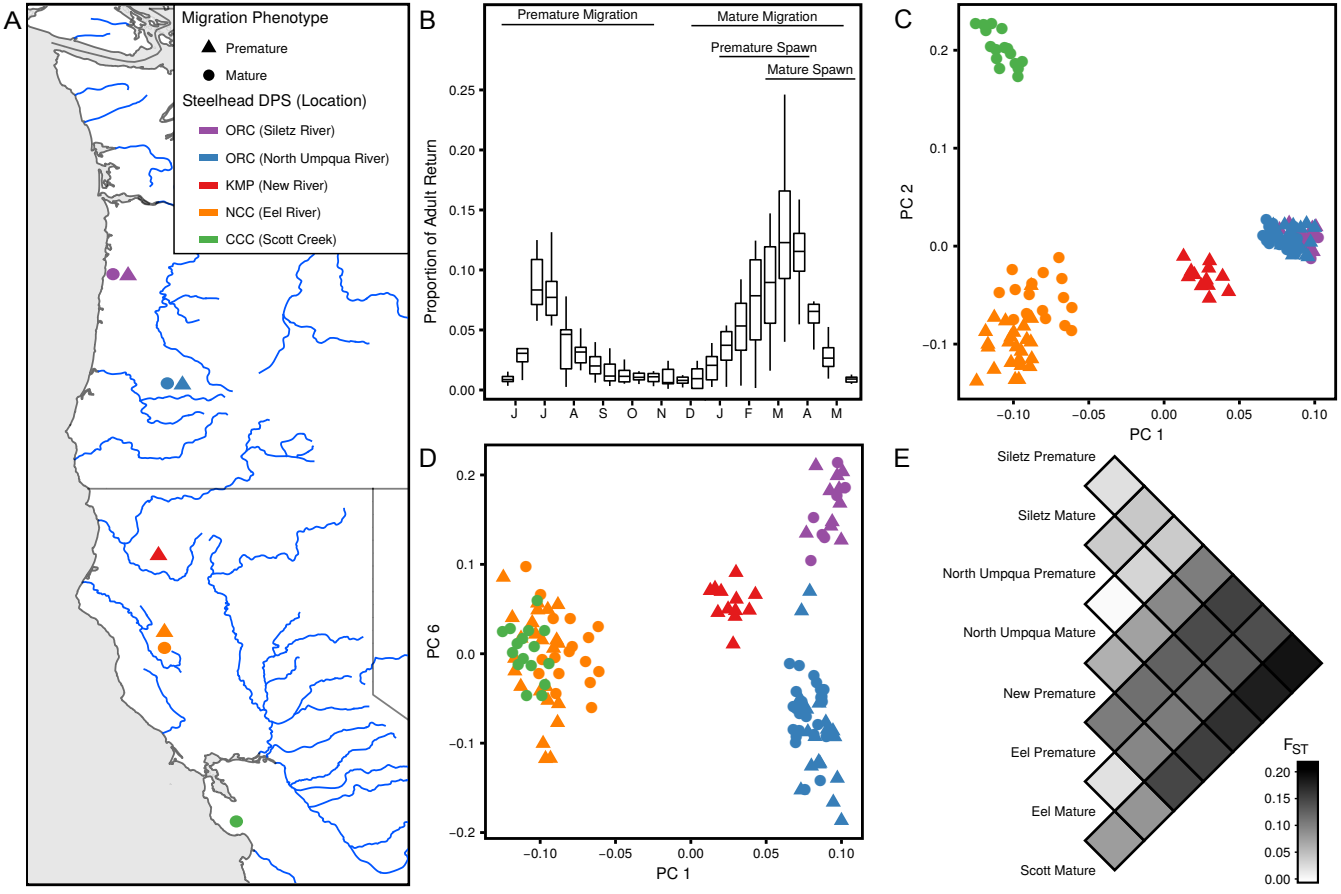
Figure 1. Genetic structure of premature and mature migrating steelhead populations. A) Map of steelhead sample locations and migration phenotypes; color indicates location, shape indicates migration phenotype. B) Bimonthly proportion of annual adult steelhead return over Winchester Dam on the North Umpqua River (2003 - 2013); horizontal bars depict migration and spawn timing of premature and mature migrating populations. C) and D) Principal component analysis and E) pairwise F_{ST} estimates using genome-wide SNP data.

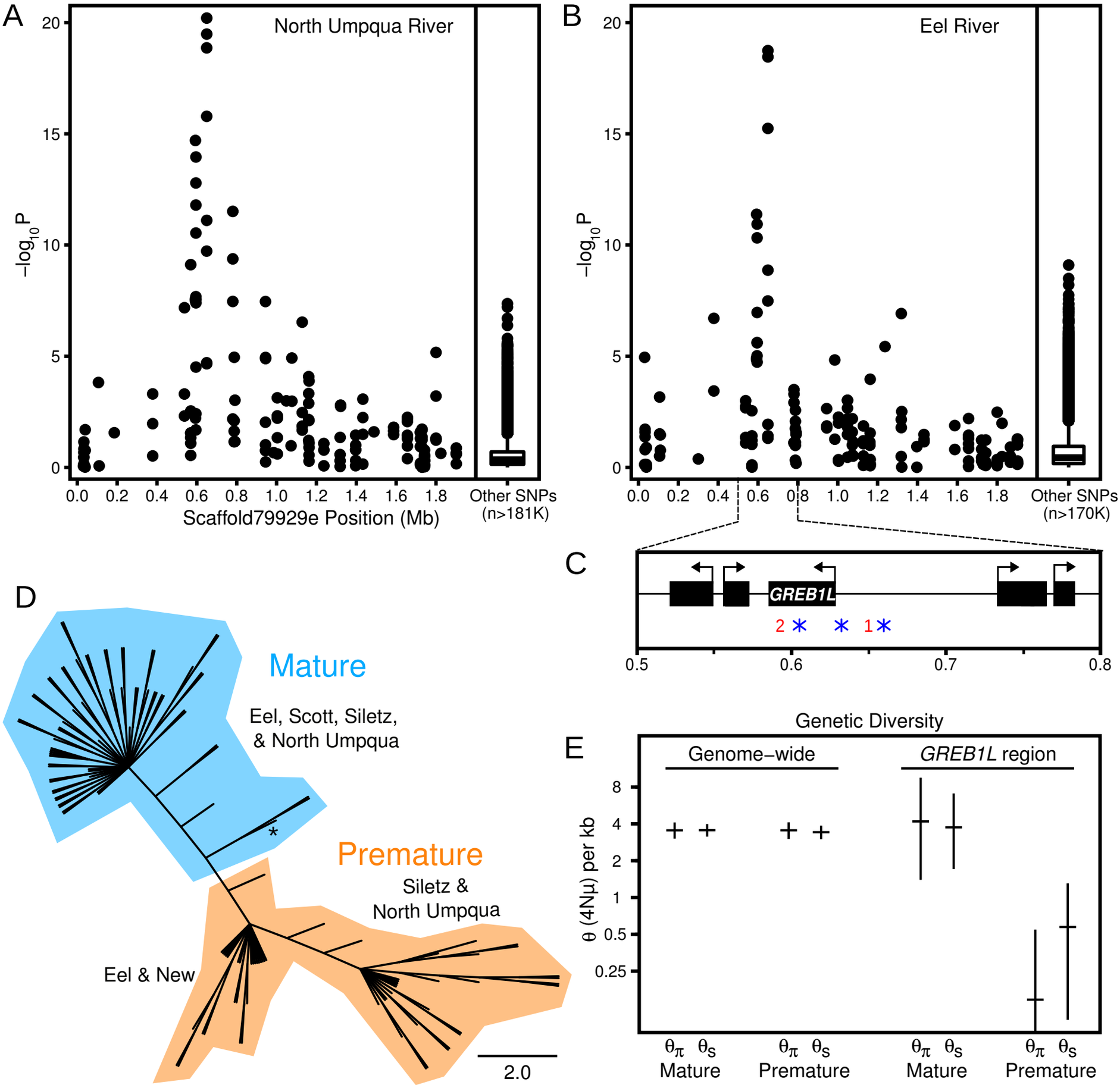
Figure 2. Genetic and evolutionary basis of premature migration in steelhead. Association mapping of migration category in A) North Umpqua River and B) Eel River steelhead. C) Gene annotation of region with strong association; red numbers indicate genomic positions of two restriction sites flanked by strongest associated SNPs, blue asterisks indicate positions of amplicon sequencing. D) Phylogenetic tree depicting maximum parsimony of amplicon sequences from all individuals; asterisk denotes potentially heterozygous individuals. E) Genome-wide and *GREB1L* region diversity estimates in North Umpqua for each migration category with 95% confidence intervals from coalescent simulations.

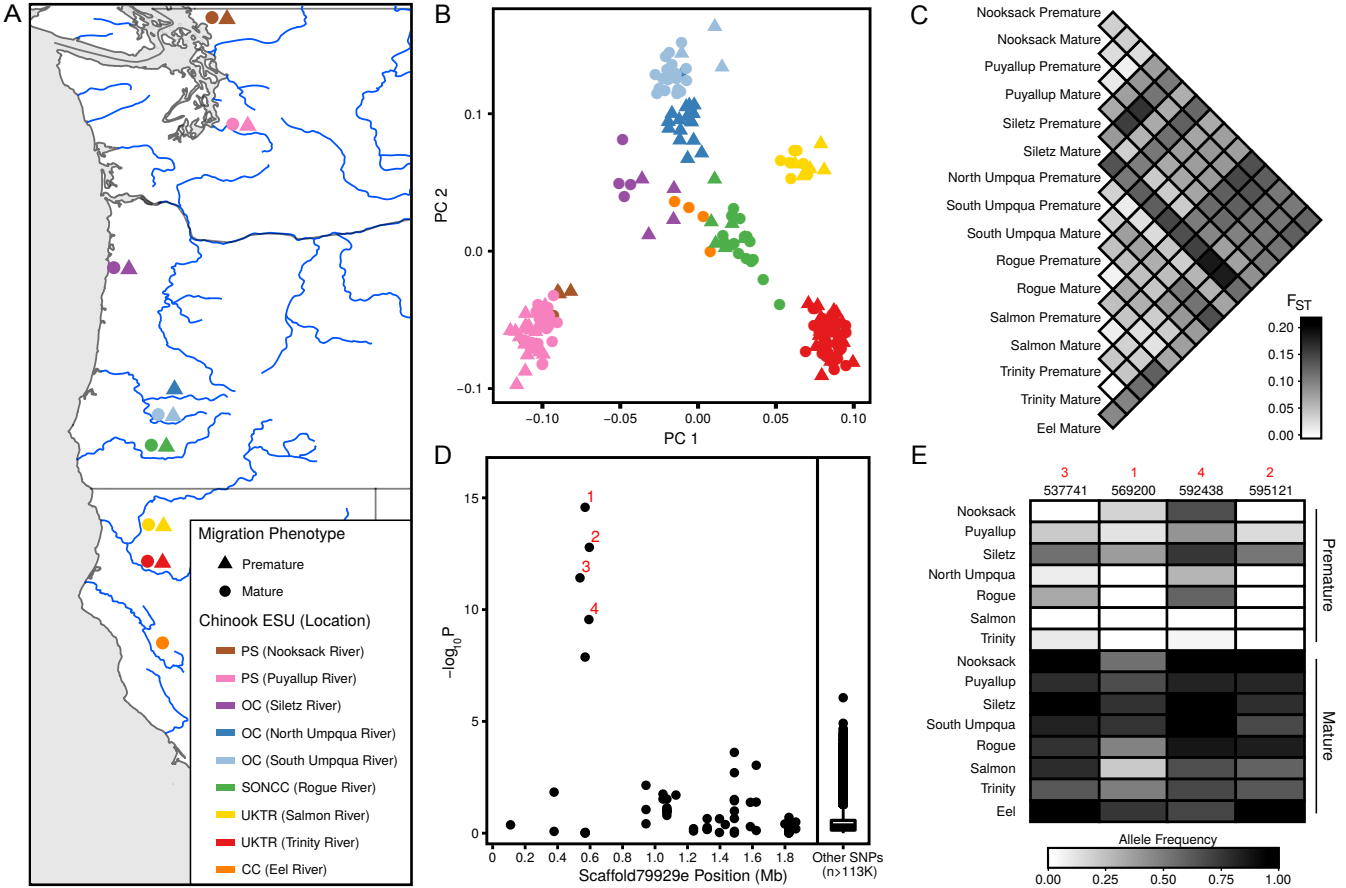
Figure 3. Genetic and evolutionary basis of premature migration in Chinook. A) Map of Chinook sample locations and migration phenotypes; color indicates location, shape indicates migration

bioRxiv preprint doi: https://doi.org/10.1101/056853; this version posted June 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

category. B) Principal component analysis and C) pairwise F_{ST} estimates using genome-wide SNP data. D) Association mapping of migration category in Chinook; red numbers indicate significant SNPs. E) Allele frequency shift at significant SNPs between premature and mature migrating populations. Black numbers indicate SNP position on scaffold79929e.


A
Migration Phenotype
Premature
Mature
Steelhead DPS (Location)
ORC (Siletz River)
ORC (North Umpqua River)
KMP (New River)
NCC (Eel River)
CCC (Scott Creek)
B
Premature Migration
Mature Migration
Premature Spawn
Mature Spawn
Proportion of Adult Return
0.25
0.20
0.15
0.10
0.05
0.00
J J A S O N D J F M A M
C
PC 2
0.2
0.1
0.0
−0.1
−0.10 −0.05 0.00 0.05 0.10
PC 1
D
PC 6
0.2
0.1
0.0
−0.1
−0.2
−0.10 −0.05 0.00 0.05 0.10
PC 1
E
Siletz Premature
Siletz Mature
North Umpqua Premature
North Umpqua Mature
New Premature
Eel Premature
Eel Mature
Scott Mature
F_{ST}
0.20
0.15
0.10
0.05
0.00


A
North Umpqua River
−log10 P
Scaffold79929e Position (Mb)
Other SNPs
(n>181K)
B
Eel River
−log10 P
Other SNPs
(n>170K)
C
GREB1L
2
1
D
Mature
Eel, Scott, Siletz,
& North Umpqua
Premature
Siletz &
North Umpqua
Eel & New
2.0
E
Genetic Diversity
Genome−wide
GREB1L region
θ (4Nμ) per kb
θπ θs
Mature
Premature


A
Migration Phenotype
Premature
Mature
Chinook ESU (Location)
PS (Nooksack River)
PS (Puyallup River)
OC (Siletz River)
OC (North Umpqua River)
OC (South Umpqua River)
SONCC (Rogue River)
UKTR (Salmon River)
UKTR (Trinity River)
CC (Eel River)
B
PC 2
PC 1
0.1
0.0
−0.1
−0.10
−0.05
0.00
0.05
0.10
C
Nooksack Premature
Nooksack Mature
Puyallup Premature
Puyallup Mature
Siletz Premature
Siletz Mature
North Umpqua Premature
South Umpqua Premature
South Umpqua Mature
Rogue Premature
Rogue Mature
Salmon Premature
Salmon Mature
Trinity Premature
Trinity Mature
Eel Mature
F_{ST}
0.20
0.15
0.10
0.05
0.00
D
$-\log_{10}P$
15
10
5
0
1
2
3
4
0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
Scaffold79929e Position (Mb)
Other SNPs
(n>113K)
E
3
537741
1
569200
4
592438
2
595121
Nooksack
Puyallup
Siletz
North Umpqua
Rogue
Salmon
Trinity
Premature
Nooksack
Puyallup
Siletz
South Umpqua
Rogue
Salmon
Trinity
Eel
Mature
Allele Frequency
0.00
0.25
0.50
0.75
1.00