SAN FRANCISCO ESTUARY & WATERSHED SCIENCE
Sponsored by the Delta Stewardship Council and UC Davis Institute of the Environment

RESEARCH

Every Cog and Wheel: Identifying Biocomplexity at the Genomic and Phenotypic Level in a Population Complex of Chinook Salmon

Shannon J. O'Leary[1], Tasha Q. Thompson[2], Mariah H. Meek*[2]

DATA ACCESSIBILITY

Data for this study are available from Meek et al. (2019): *https://doi.org/10.5061/dryad.tht76hdvt*). Appendix A and B contain fully reproducible code that supports the analysis; the research compendium is available at *https://github.com/sjoleary/ONC_GenDiv/* and contains both original Rmarkdown files and corresponding rendered html files.

ABSTRACT

Genetic diversity is the fundamental building block of biodiversity and the necessary ingredient for adaptation. Specifically, the intraspecific diversity (biocomplexity) comprising phenotypic and genetic variation partitioned within and among populations can determine the ability of a species to respond to changing environmental conditions. Here, we explore the biocomplexity of California's Central Valley Chinook Salmon (*Oncorhynchus tshawytscha*) population complex at the genomic level by quantifying population genomic diversity among and within migration life-history phenotypes. Notably, despite apparent gene flow among populations with the same migration (life-history) phenotypes that inhabit different tributaries, each group is characterized by a distinct component of unique genomic diversity. While enumerating biodiversity contained within individual hierarchical levels is informative, it is important to consider inter- and intraspecific diversity simultaneously, because there may be emergent properties at higher levels as a result of the presence of diversity at lower ones. Our results emphasize the importance of formulating conservation goals focused on maintaining biocomplexity at both the phenotypic and genotypic level. Doing so will preserve the species' adaptive potential and increase the probability of persistence of the population complex despite changing environmental pressures.

KEY WORDS

conservation genomics, intraspecific diversity, portfolio effect

SFEWS Volume 22 | Issue 4 | Article 5
https://doi.org/10.15447/sfews.2024v22iss4art5

* Corresponding author: *mhmeek@msu.edu*

1 Department of Biology
Saint Anselm College
Manchester, NH 03102 USA

2 Department of Integrative Biology, AgBioResearch, and Ecology, Evolution, and Behavior Program
Michigan State University
East Lansing, MI, 48824 USA

INTRODUCTION

> *“The last word in ignorance is the man who says of an animal or plant, ‘What good is it?’[…] If the biota, in the course of aeons, has built something we like but do not understand, then who but a fool would discard seemingly useless parts? To keep every cog and wheel is the first precaution of intelligent tinkering.”*
> *— Aldo Leopold*

As the fundamental building block of biodiversity, genetic diversity forms the cogs and wheels that comprise a population’s adaptive potential. However, identifying and quantifying biodiversity at each hierarchical level of diversity—including genes, species, and ecosystems—is a complex undertaking, with distinct frameworks to enumerate compositional, structural, and functional diversity (Petchey and Gaston 2002; Duelli and Obrist 2003; Péru and Dolédec 2010). Functional genetic diversity includes sequence polymorphisms and differences in gene expression, which together shape the phenotypic diversity that comprises differences in morphology, physiology, and life-history characteristics present within a population. Under current environmental conditions, some components of functional diversity will be selectively neutral, though future shifts in ecological pressures may result in certain phenotypes and underlying genotypes that appeared unimportant under previous conditions becoming critical to the persistence of a species (Messer et al. 2016a). Therefore, to sustain biodiversity across scales, sound conservation and management strategies should seek to conserve genetic diversity that may prove critical for future adaptation (CBD 2011; Hoelzel et al. 2019; Mable 2019).

Increasingly, the importance of intraspecific diversity (biocomplexity) has been recognized as a determining factor for the stability and resilience of biological systems (Hilborn et al. 2003; Des Roches et al. 2021a). For example, at an ecosystem level, the number of trophic levels and number of species at each level determines the stability of the food web; while at a species level, the diversity of life-history strategies can be critical for maintaining a temporally stable population through risk partitioning. This effect is described as the portfolio effect (Hilborn et al. 2003, Schindler et al. 2010). However, biocomplexity of a given level of biodiversity cannot always be confined to a single metric. For example, while intraspecific diversity at a population level can be measured as phenotypic diversity, a genetic component underlies some phenotypic differences. While it is more straightforward to assess biodiversity at each individual hierarchical level, it is important to acknowledge that the nested complexity of diversity harbored within and across hierarchical levels has its own emergent properties. Therefore, because selection acts on standing variation, forward-looking conservation strategies must stress the importance of maintaining a diverse genetic portfolio within and among populations to maintain adaptive potential in changing environmental conditions (Mimura et al. 2017; Mable 2019; Hoban et al. 2020; Des Roches et al. 2021a).

Chinook Salmon, *Oncorhynchus tshawytscha*, in the Central Valley of California are emerging as a model system to explore the importance of biocomplexity for the persistence of a population complex that faces multiple external threats, including habitat fragmentation, over-exploitation, and climate change (Moyle et al. 2017). Chinook are anadromous with a distinct life history that spans both freshwater and marine ecosystems. Eggs are laid in the tributaries, where juveniles rear for a period of time before migrating out to the ocean. There, they spend 1 to several years growing in the ocean before migrating back to their natal river to spawn and die, providing an important source of oceanic nutrients to ecosystems as well as supporting recreational, commercial, and indigenous fisheries (Quinn 2018). The tributaries of the Central Valley contain four distinct run types (migration phenotypes) named for the time of year adults enter freshwater systems to spawn: winter (endemic to the Central Valley), spring, fall, and late-fall. The same tributary may support multiple runs (Williams 2006). Early migrating runs (winter, spring) make the tradeoff of migrating earlier at a smaller size, leaving behind the nutrient-rich oceanic habitats to access spawning sites higher in the watershed that remain cool over the summer, where they complete maturation in a fasted state, spawn, and die (Quinn et al. 2016). By contrast, late-migrating salmon (fall, late-fall) remain in the ocean until relatively mature before making their spawning run. This life-history diversity results in a partitioning of spawning habitat in space and time. And as a result, biodiversity is not straightforwardly partitioned across this population complex, because individuals that share a run type are more similar to each other compared to other individuals in the same tributary.

This asynchronicity in run timing of adults creates a "portfolio of stocks" that stabilizes the population complex overall by buffering against spatio-temporally variable environmental conditions and anthropogenic effects. For example, while environmental conditions within a given year may be poor for early migrating adults, in that same year they may be optimal for their late-migrating counterparts in the same tributary, thus promoting interannual stability and increasing resilience overall (Carlson and-Satterthwaite 2011; Griffiths et al. 2014). However, this buffering ability is threatened when one run type is consistently negatively affected across tributaries, and the level of homogenization among population segments increases. Specifically, dams and other anthropogenic factors have disproportionately affected historical early migratory habitat in much of the Central Valley and, as a result, both spring run and winter run Chinook Salmon have experienced severe declines in abundance and are listed under the Endangered Species Act (NMFS 2005).

Additionally, human activities have homogenized habitat and substantially increased interbreeding between spring- and fall-run Chinook Salmon in many locations (Waples et al. 2022). This has affected the biocomplexity of the Central Valley population complex as a whole, as it has shifted to primarily comprising fall-run individuals. In turn, this has reduced the portfolio effect as the early-run phenotypes have become increasingly rare, thus making the population complex more vulnerable (Carlson and Satterthwaite 2011). However, wild spring-run populations in California's Central Valley still access habitat very distinct from their fall-run counterparts, and, overall, a great deal of habitat heterogeneity exists *within* both the spring and fall runs. Thus, the Central Valley provides a unique opportunity to examine the extent to which unique genetic variation still exists—despite these homogenizing factors—between runs, and among populations within a run where historical spatio-temporal separation between runs is relatively intact.

Collectively, the demographic changes described above are likely to result in an erosion of genetic diversity within and among components of this population complex, and consequently adaptive potential, which may otherwise have proved important for persistence under future environmental conditions. Measuring the genetic diversity has long been used as a proxy to quantify the "future adaptive potential" of populations (Reed and Frankham 2001; Reed and Frankham 2003). However, despite a plethora of tools to quantify genetic diversity, consensus on which metrics of genetic diversity to focus on for conservation remains elusive. For example, for salmonids, neutral genomic diversity has long been used to identify intrinsic markers to understand structural diversity, i.e., how the diversity of the whole complex is partitioned among individual groups, with a focus on the ability to identify demographic groups. These analyses are less interested in genomic diversity itself, and more on how it functions as a marker to delineate groups. More recently, the direct interrogation of functional biodiversity at the genomic level has also gained attention. Indeed, recent studies have identified a single chromosomal region (GREB1L to ROCK1) that underlies adult migration timing in Chinook and other salmonids (Prince et al. 2017; Narum et al. 2018; Thompson et al. 2020). Additionally, increasing the availability of genomic datasets holds promise for identifying genes associated with polygenic traits (Ouborg et al. 2010; Sinclair–Waters et al. 2020).

Ultimately, identifying the composition of both known functional diversity and presumed neutral genetic diversity of individual groups—rather than just how the diversity is partitioned *among* groups—will be vital to inform conservation actions that can protect all levels of diversity, improve species persistence, and promote adaptive potential. Here, we aim to understand the genomic biocomplexity contained within and among the four Central Valley Chinook Salmon run types to help inform conservation actions. To achieve this, we leverage a previously published dataset (Meek et al. 2019). The value of this dataset is that it includes all the major populations of Central Valley Chinook Salmon, including winter run, which is on the verge of extinction. The analysis of Meek et al. (2019) demonstrated greater population structure across the Central Valley than had been previously described using several thousand biallelic single-nucleotide polymorphisms (SNPs). This approach is effective in identifying the structural diversity of demographic groups, and effectively identified differences among run types and tributaries (population-level biocomplexity) that were previously overlooked, which was the focus of the initial analysis of this dataset.

In this expansion of the original analysis, we resolve the limitations of the original analysis of biallelic SNPs by using the recently completed Chinook Salmon genome (Christensen et al. 2018) and a novel microhaplotype analysis. This analysis enables efficiently phasing SNPs on the same restriction site-associated (RAD)-tag into haplotypes (Willis et al. 2017) to identify >10,000 multi-allelic loci distributed throughout the coding and non-coding parts of the Chinook genome, and additionally allows us to assess additional sequence-based metrics, despite sequencing only a subset of the genome. We build out the initial findings of Meek et al. (2019) of hidden biocomplexity at a population level by identifying hidden biocomplexity at an even finer scale by focusing on assessing genomic diversity within and among population components and migration phenotypes. Haplotyping loci unlocks a range of additional metrics of genomic diversity (e.g., number of alleles at a locus and the evenness of their distribution, private haplotypes and loci only variable in a limited number of populations, along with sequence-based diversity metrics) compared to the original SNP dataset (Willis et al. 2017; O'Leary et al. 2018). This enables a shift from comparing structural diversity (how diversity is partitioned across populations of the Central Valley Complex) mainly at a phenotypic level, toward an emphasis on compositional diversity, especially unique genomic diversity present in individual populations and migration phenotypes of the entire complex. These metrics allow us to quantify the amount of (unique) genetic variation contained in each population, to understand what may be lost in the population complex as a whole if individual sub-populations go extinct. Overall, our results not only provide new insight about this important and highly imperiled species, but also demonstrate the power of applying new advances to existing datasets to gain vital biological understanding, without the need to re-sample imperiled populations.

MATERIALS AND METHODS

Sample Collection and Sequencing

We obtained fin clips from all four run types (fall, late-fall, winter, spring; see Figure 1) from adult Chinook Salmon from major tributaries in the Central Valley during their spawning migrations (Meek et al. 2019). Our goal is to explore differences among individuals within and among run types. Therefore, we have explored patterns for individuals grouped by run type within each tributary throughout this analysis, and additionally distinguish between wild and hatchery-reared individuals. It should be noted that while fall-run Chinook Salmon are represented by eleven tributaries, and spring-run individuals by four, late-fall-run and winter-run populations are represented by one tributary (upper Sacramento River) each, when considering patterns described at a run level. Genomic DNA was extracted and digested using *SbfI* to construct RAD libraries following Miller et al. (2012). Fifteen libraries consisting of 30 to 47 individuals each were sequenced on single Illumina HiSeq 2000 lanes (100 bp, single end).

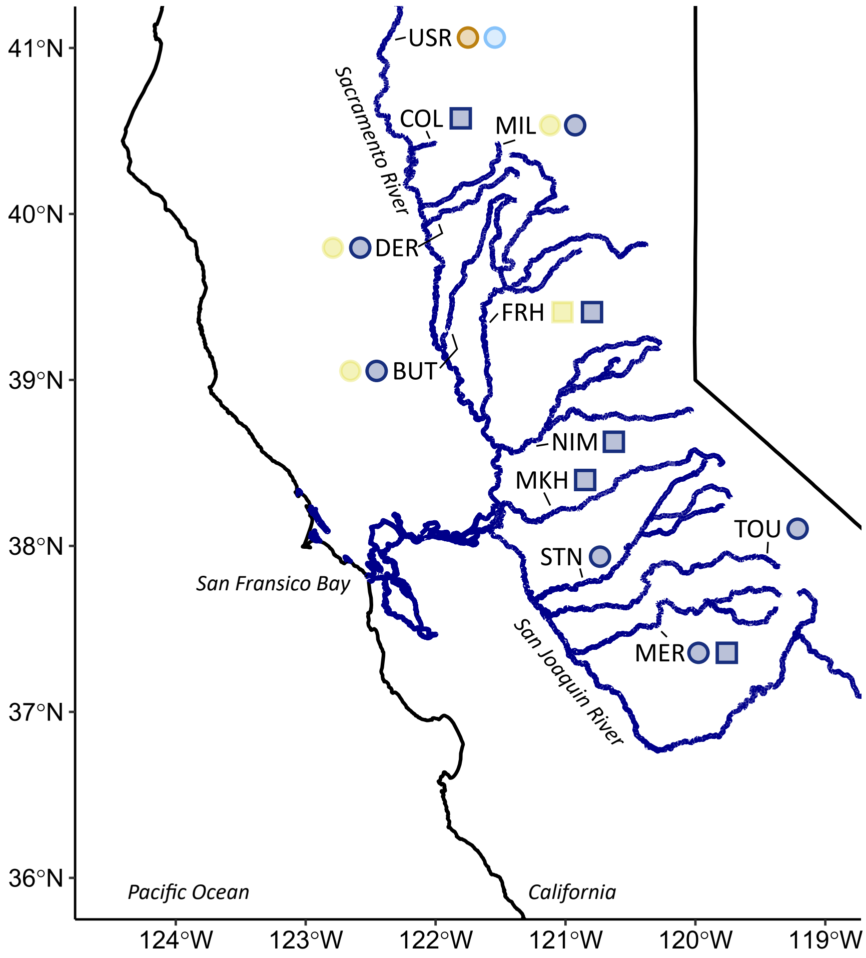

Figure 1 Tributaries of the California's Central Valley sampled for this study. Tributaries are labeled with the abbreviations used throughout: USR (upper Sacramento River), COL (Coleman Hatchery/Battle Creek), MIL (Mill Creek), DER (Deer Creek), BUT (Butte Creek), FRH (Feather River Hatchery), NIM (Nimbus Hatchery/American River), MKH (Mokelumne River Hatchery), STN (Stanislaus River), TOU (Tuolumne River), MER (Merced River). Samples from each tributary comprising hatchery individuals are indicated by *squares*, and wild populations by *circles*. *Colors* represent the sampled run type at each location (spring = yellow, winter = brown, fall = dark blue, late-fall = light blue). Carcasses of adults after spawning where sampled in multiple locations throughout the tributaries.

Genotyping

Raw sequences were de-multiplexed using *process_radtags* (Catchen et al. 2011), quality trimmed using *fastp* (Chen et al. 2018) and mapped to a Chinook Salmon reference genome (Christensen et al.-2018) using *BWA-mem* (Li and Durbin 2009). Reads with mapping quality >5 were concatenated into a single bam file using *bedtools* (Quinlan 2014). Based on the expected size range of RAD-tags, we extracted mapping intervals >25 bp and <500 bp with coverage >50 reads to enable downstream haplotyping of SNPs on the same RAD-tag. Once we compiled sequences (putative loci) into a reduced-representation reference consisting of only recovered RAD-tags, we mapped reads to this reference and called SNPs using *freebayes* (maximum allowed gap -E = 3, minimum mapping and base quality = 5, otherwise default parameters settings, see Garrison and Marth [2012]) to generate a dataset consisting of SNPs called on individual RAD-tags to produce multi-allelic loci of similar length downstream.

We rigorously filtered the raw dataset following principles set forth in O'Leary et al. (2018): Genotypes with <5 reads or quality (confidence in genotype call) <20 were coded as missing. Loci were required to have a mean minimum depth of 15 and be called in at least 50% of individuals of a given library, 85% of individuals of a run type, and 90% of individuals overall to be retained in the final dataset. To remove potential artifacts that resulted from polymerase chain reaction (PCR) or sequencing error, we removed singletons with a depth <10 and doubletons with a depth <20. We further filtered the remaining loci for allele balance, mapping quality ratio of reference vs.

alternate allele, depth/quality ratio, and excess heterozygosity to remove paralogs, loci impacted by allelic dropout and other technical artifacts (O'Leary et al. 2018). (See Appendix A for details on implementation.) Finally, we collapsed SNPs on the same RAD-tag into haplotypes using *rad_haplotyper* (Willis et al. 2017). This newly compiled dataset with multi-allelic loci has increased power to assess genomic diversity and biocomplexity compared to the original analysis (Meek et al 2019) because, in general, biallelic SNPs contain less information per locus compared to multi-allelic loci (Morin et al. 2009). Further, the necessity of thinning SNPs to ensure loci are independent observations (Kaeuffer et al. 2007) reduces the information content because the power of a dataset resides in the number of independent alleles rather than the number of loci (Kalinowski 2002). By contrast, haplotyping preserves the information content of all SNPs in the dataset, resolves physical linkage artefacts, and results in more inferential power per locus (Willis et al. 2017; Baetscher et al. 2018).

Assessment of Population Structure and Differentiation

We explored population structure using a clustering analysis based on genetic similarity and an assessment of population differentiation among individuals grouped *a priori*. In the first method, we first performed a Principal Component Analysis (PCA) to reduce the dataset's dimensionality, and projected the genotypes into two-dimensional (2-D) space to identify patterns of similarities and differences among individuals. Then, we clustered individuals into K = 1–10 groups using k-means clustering based on a PCA-transformed genotype matrix (no assumptions regarding Hardy–Weinberg/linkage disequilibrium), followed by a Discriminant Analysis of Principle Components (DAPC) using *adegenet* (Jombart 2008) to determine membership probabilities of every sample to each inferred cluster. To ensure sufficient variance was retained to discriminate among groups but not overfit the data, we determined the optimum number of principle components to retain using a stratified cross-validation.

For the second method, we grouped individuals by run type within tributary and calculated Weir and Cockerham's unbiased estimator of F_{ST} (Weir and Cockerham 1984) to assess population differentiation. Then we calculated global F_{ST} to test for genetic differentiation across all groups and calculated pairwise F_{ST} to test for pairwise differentiation among groups. We determined significance using 95% confidence intervals around each estimate generated by resampling loci 1,000 times using *assigner* (Gosselin et al.- 2016).

Assessment of Genomic Diversity

We calculated all measures of genomic diversity for individuals grouped by run type within each tributary; when present in the same tributary, wild and hatchery individuals were treated as separate groups. We assessed four types of parameters, (1) measures of heterozygosity (observed/expected heterozygosity, inbreeding coefficient F_{IS}), (2) measures of allelic diversity (allelic richness/evenness), (3) sequence-based parameters (nucleotide diversity, Tajima's D), and (4) measures of unique variation (fixed loci, singletons, private alleles/polymorphisms).

For the first three sets of parameter types, we determined whether measures of genetic diversity varied among groups (significant heterogeneity) using a Friedman's rank sum test followed by a Wilcoxon signed-rank test for significance of pairwise differences between groups; we corrected *p*-values for multiples comparisons assuming a false discovery rate (FDR) of 0.05 (Benjamini and Hochberg 1995). We measured the observed heterozygosity (Ho) as the proportion of heterozygote genotypes per locus (Nei 1987), and the expected heterozygosity (gene diversity, Hs) as the proportion of heterozygous genotypes expected under Hardy–Weinberg Equilibrium (Nei 1987). The inbreeding coefficient F_{IS} was calculated as 1–(Ho/He) (Weir and Cockerham 1984). A positive F_{IS} indicates an excess of homozygotes, generally interpreted as the result of individuals more likely to mate with individuals with similar genotypes (e.g., related individuals). By contrast, a negative F_{IS} indicates an excess of heterozygotes from outbreeding or other forms of disassortative mating.

To account for differences in sample size, we determined allelic richness as rarefied allele counts. We calculated the evenness of allelic diversity at a given locus as the ratio of the number of abundant to the number of more rare genotypes using the ratio of the Stoddart and Taylor index (diversity index weighted for more abundant alleles) and Shannon–Wiener index (diversity weighted for more rare alleles) as implemented in *poppr* (Kamvar et al. 2014)*;* lower values indicate the prevalence of more rare alleles and uneven distributions of allele frequencies.

We calculated nucleotide diversity (π, Nei 1987) as the sum of the number of pairwise differences between haplotypes of a given locus over the number of comparisons made as implemented in *pegas* (Paradis and Barrett 2010). Tajima's D was calculated as the difference between estimating the population-scaled mutation rate Θ as nucleotide diversity π and as the number of segregating sites standardized by dividing it by the square root of its variance. Because the nucleotide diversity π under-estimates the number of mutations that are rare in the population, Tajima's D can be used to test the neutral mutation hypothesis. Tajima's D = 0 occurs when the observed and expected variation are similar; this is generally interpreted as the population evolving per mutation-drift equilibrium. A negative Tajima's D occurs when there are fewer haplotypes than expected, given the number of segregating sites; this can indicate loci under directed selection or a population expanding after a bottleneck. Finally, a positive value is the result of more heterozygosity than expected, and indicates loci affected by balancing selection or population declines. To identify whether the observed genome-wide distributions for each population reflect the expected distribution under the neutral mutation hypothesis, we generated a genome-wide distribution of Tajima's D for a set of neutral loci reflecting the composition of the haplotyped empirical dataset consisting of the same number of loci with the same distribution of segregating sites using MS (Hudson 2002). We performed a Kolmogorov–Smirnov test to identify significant differences among the empirical and simulated distribution for each population.

Finally, we assessed patterns of unique diversity by comparing (1) the number of fixed loci (loci not polymorphic within a group), (2) the number of loci with singletons (allele observed in only one individual) and the number of singletons per individual by run/tributary group, (3) the number of private polymorphisms, and (4) the number of private alleles. Private polymorphisms are defined as loci that are variable only in a single group while all other groups are fixed at this locus. By contrast, private alleles are alleles found in only a single group, though other groups exhibit more than one allele at that locus. We used rarefied allele counts to identify private polymorphisms as loci consistently being polymorphic in only one population independent of the subset of individuals included. Private alleles were identified by sub-sampling 15 individuals from all populations except Feather River Hatchery and identifying private polymorphism 100 times. To compare whether identified loci are randomly distributed across chromosomes, we generated null distributions by shuffling chromosome designations across loci 1,000 times to determine whether the observed values fall outside the null distribution. Additionally, we determined private polymorphisms and alleles for each run by randomly drawing 21 individuals (number of late-fall-run individuals) from the pooled set of wild individuals present in the dataset for each run, to account for the fact that spring and fall individuals are represented by multiple tributaries in the dataset, and therefore the alleles private to an entire run will not be identified when the dataset is split into individual populations based on tributary.

RESULTS

Genotyping

After decomposing indels and removing low-quality individuals (mean coverage <3 reads, >75% missing data, affected by library effects) and retaining only biallelic SNPs, the raw dataset consisted of 456 individuals and 1,738,126 loci. Details on the number of individuals and loci filtered at each step are documented in Appendix A. The SNP dataset that was haplotyped comprised 31,897 SNPs and 413 individuals that produced 14,654 multi-allelic loci, which were further filtered to remove loci with low haplotyping success and possible paralogs, as well as individuals with low haplotyping success and/or large proportions of flagged loci. The final filtered dataset consisted of 386 individuals genotyped for 12,983 multi-allelic loci with a total of 30,037 alleles. The number of alleles per locus ranged from two to nine.

Assessment of Population Structure and Differentiation

The clustering analysis and comparisons of pairwise F_{ST} recovered the same groups as identified in Meek et al (2019), with the four runs forming distinct groups, with the exception of spring Feather River Hatchery individuals clustering with fall individuals (Figure 2). Principal component 1 (2.5% of variation) separates winter-run individuals as being most distinct from other groups. Principle component 2 (1.1% of variation) identifies clusters of fall and late-fall individuals. Notably, fall and spring Feather hatchery individuals form a cluster between the fall and spring groups. (Figure 2A). We recover the same intraspecific groupings within wild spring-run individuals (Mill and Deer Creek cluster together, while Butte Creek clusters separately) as in Meek et al 2019. When comparing pairwise F_{ST} values, additionally, Coleman Hatchery showed genetic distinction within fall-run individuals (Figure 2, Table S2; Figure S2).

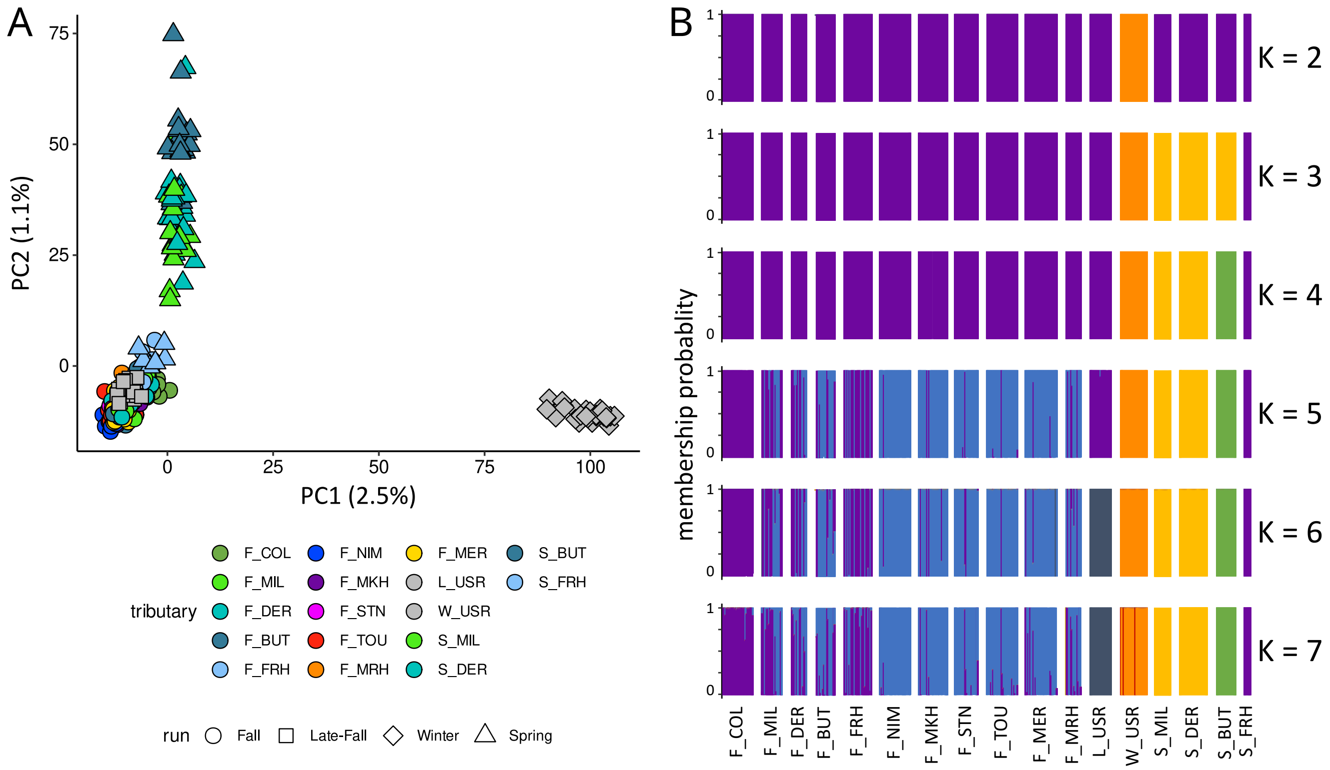

Figure 2 Comparison of genetic similarity based on Principle Component Analysis and Discriminant Analysis of Principle Components: (**A**) Biplot of clustering of individuals on principle components 1 and 2. Individuals from the same tributary indicated by the same *color*. Run type is indicated by the *shape* (spring = triangle, winter = diamond, fall = circle, late-fall = square). (**B**) Membership probability of each individual to clusters identified using -means hierarchical clustering for K = 2–7. Tributaries are labeled with the abbreviations used throughout: USR (upper Sacramento River), COL (Coleman Hatchery/Battle Creek), MIL (Mill Creek), DER (Deer Creek), BUT (Butte Creek), FRH (Feather River Hatchery), NIM (Nimbus Hatchery/American River), MKH (Mokelumne River Hatchery), STN (Stanislaus River), TOU (Tuolumne River), MER (Merced River).

Assessment of Genomic Diversity: Measures of Heterozygosity

The mean observed population-level heterozygosity was lowest for winter-run individuals (mean = 0.1272), followed by spring-run from Butte Creek (mean = 0.1587), and highest for fall-run from Coleman and Feather River hatcheries (0.1713 and 0.1741, respectively, Table S3; Figure 3A). Similarly, mean expected heterozygosity was lowest for upper Sacramento River winter-run (mean = 0.1285), and highest for Feather River Hatchery fall-run and Mill Creek fall-run (mean = 0.1718 and 0.1714, respectively, Table S4; Figure 3B). For both observed and expected heterozygosity, spring-run tributaries exhibited a wider range of distributions compared to fall-run tributaries, despite the smaller number of spring-run tributaries in the dataset (Figure 3A and 3B). Late-fall-run values fell into the range of distributions observed among fall-run tributaries (Figure 3A and 3B).

The lowest F_{IS} values are observed in populations from the Coleman, Feather River, Nimbus, and Mokelumne hatcheries (mean = -0.0194 to -0.0131); Merced River Hatchery is the exception among hatcheries, with a positive mean F_{IS} value (mean = 0.0149), indicative of excess heterozygotes observed in these populations relative to expectations. By contrast, the highest mean F_{IS}

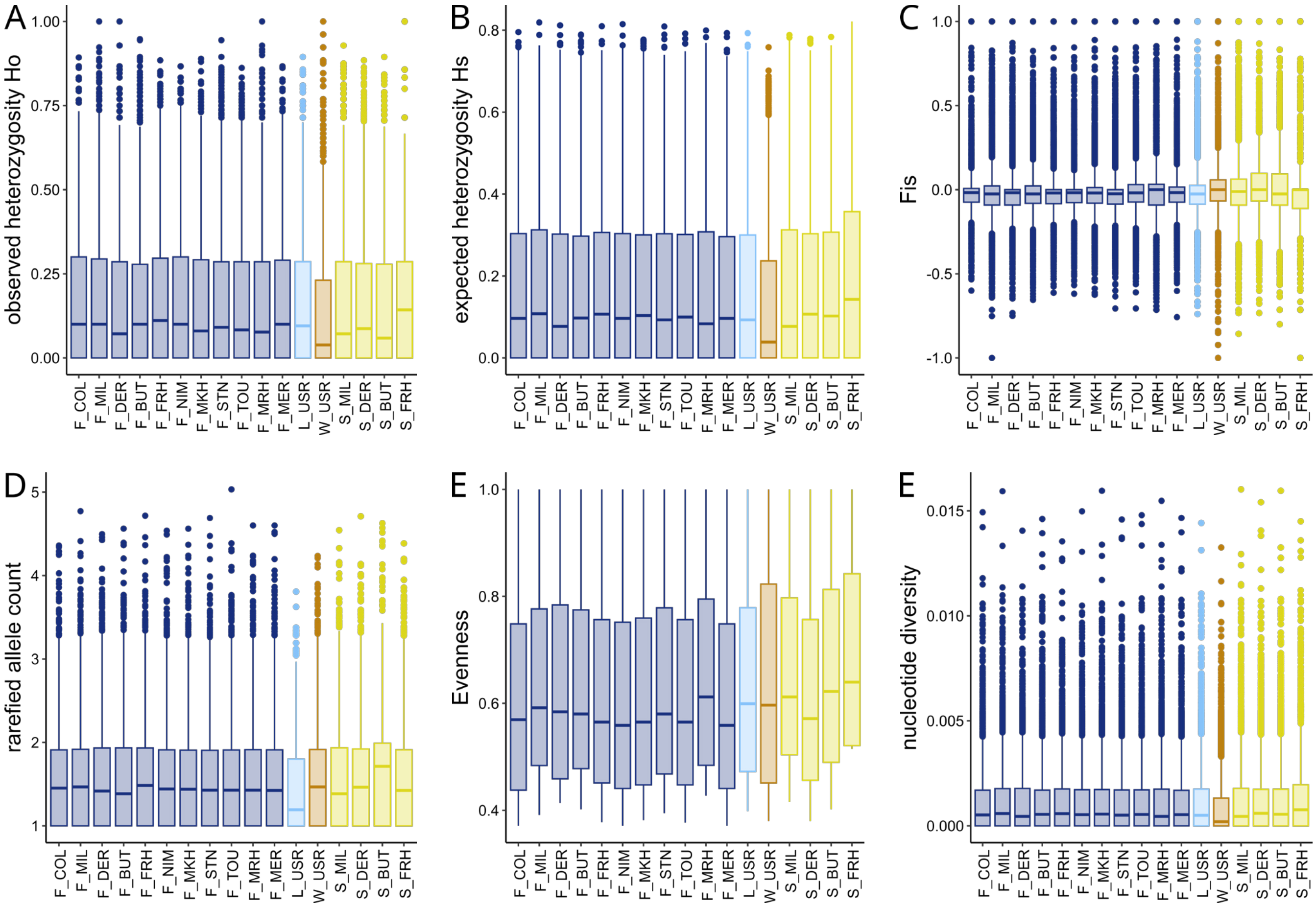

Figure 3 Assessment of the distribution of genomic diversity across all loci for individuals grouped by run type within tributaries using heterozygosity-based parameters: (**A**) observed heterozygosity Ho, **B.** expected heterozygosity Hs, **C.** inbreeding coefficient Fis), measures of allele diversity (**C.** Allelic richness, **D.** Evenness), and sequence-based parameters (**E.** observed nucleotide diversity). Tributaries are labeled with the abbreviations used throughout: USR (upper Sacramento River), COL (Coleman Hatchery/Battle Creek), MIL (Mill Creek), DER (Deer Creek), BUT (Butte Creek), FRH (Feather River Hatchery), NIM (Nimbus Hatchery/American River), MKH (Mokelumne River Hatchery), STN (Stanislaus River), TOU (Tuolumne River), MER (Merced River). *Colors* represent the sampled run type at each location (spring = yellow, winter = brown, fall = dark blue, late-fall = light blue).

https://doi.org/10.15447/sfews.2024v22iss4art5

values were identified for upper Sacramento River winter-run individuals and spring-run wild populations (mean = 0.0104 – 0.0370, Figure 3C; Table S5); this is indicative of excess homozygotes present in these populations.

Assessment of Genomic Diversity: Measures of Allelic Diversity

Mean values of rarefied allele counts were comparable across tributary/run groups, ranging from 1.51 to 1.52 alleles per locus for all run/tributary groups except upper Sacramento River late-fall (1.36) and Deer Creek spring-run (1.48), which exhibited significantly lower mean values (Tables S6 and S7). Despite similar mean values, most pairwise comparisons were significant (Table S6), indicating that even though there was a relatively consistent global number of alleles per locus, the patterns of which loci are variable are consistently different across run/tributary groups. For example, Butte Creek spring-run exhibited a pattern of rarefied allele counts significantly different from all other locations, and, despite a mean value comparable to most other groups, also exhibited the highest median value (1.71; Figure 3D). Overall, spring-run tributaries exhibited more variation among tributaries compared to fall-run tributaries (median = 1.38 to 1.71; all pairwise comparisons are significant, Table S6). Notably, this pattern is driven by wild spring-run in Mill, Deer, and Butte creeks, which have avoided the extensive fall-run introgression experienced by the Feather River hatchery spring-run. By contrast, distributions were more similar to each other across fall tributaries. Butte Creek fall individuals exhibited the lowest (1.38) values, while Feather River Hatchery individuals exhibited the highest (1.48) median values (Figure 3D), perhaps as a result of introgression with spring individuals in the hatchery. Notably, individuals from different run types from the same tributary may exhibit quite different patterns. For example, fall-run individuals from Mill Creek and the Feather River Hatchery exhibited higher allele counts compared to spring-run individuals from the same tributary. By contrast, fall-run individuals from Deer and Butte creeks had lower allele counts compared to their spring-run counterparts. The strongest contrast was Butte Creek, where spring-run individuals exhibited the highest allele counts overall, while fall-run individuals exhibited the lowest allelic richness among all fall tributaries (Figure 3D).

Overall, winter-run individuals from upper Sacramento River exhibited statistically significant lower evenness of allelic richness across loci, indicating that many loci were characterized by rare alleles (Figure 3E). By contrast, Feather River Hatchery fall-run individuals exhibited the highest median evenness (0.78) compared to all other run/tributary groups (Table S7). Within fall-run tributaries, mean levels of evenness were all approximately 0.76, though there were some significant differences in their overall distributions (Figure 4E). Again, spring-run tributaries exhibited a wider range of mean and median values for evenness compared to fall-run tributaries (Figure 3E).

Assessment of Genomic Diversity: Sequence-Based Parameters

Upper Sacramento River winter-run and Butte Creek spring-run individuals exhibited lower nucleotide diversity compared to all other locations (Table S8; Figure 3E). By contrast, fall-run Feather River Hatchery individuals exhibited the highest mean nucleotide diversity; similarly, among spring tributaries, Feather River Hatchery individuals exhibited the highest nucleotide diversity (Table S8; Figure 3E); again, this result is expected, give the known level of introgression within the Feather River.

Upper Sacramento River winter-run individuals and Butte Creek spring-run individuals are the only populations to exhibit positive mean values for Tajima's D (Table 1; Table S10). These groups show the largest differences in their median values compared to their simulated null distribution under mutation-drift equilibrium (Figure 5; Table S11); similarly, fall-run individuals from Merced River Hatchery, Deer Creek, Mill Creek, Mokelumne Hatchery, and Stanislaus River, as well as late-fall individuals, and spring individuals from Butte Creek and Deer Creek, exhibit median values that are

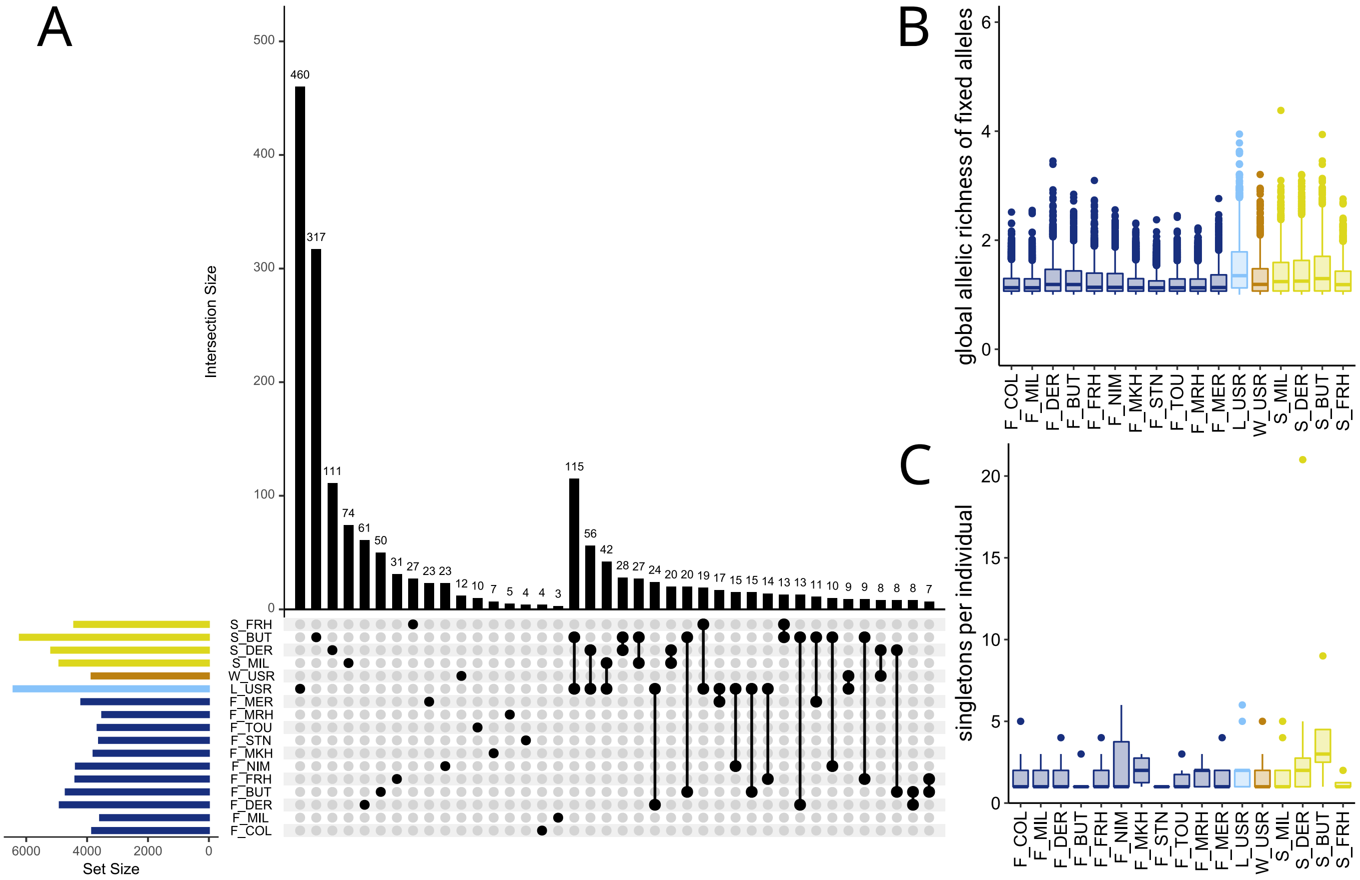

Figure 4 Assessment of fixed loci and singletons: (**A**) Comparisons of fixed loci across run/tributary groups. The set size (*horizontal bars*) indicates the total number of fixed loci in a given group, the intersect size (*vertical black bars*) correlates to the number of loci fixed in a single group (*single black dot*) or in two (*black dots connected by line*); (**B**) Distribution of global allelic richness of loci fixed in a given group; (**C**) Distribution of the number of singletons per individual for each run/tributary group. The *color of horizontal bars and boxplots* represent the sampled run type at each location (spring = yellow, winter = brown, fall = dark blue, late-fall = light blue). Tributaries are labeled with the abbreviations used throughout: USR (upper Sacramento River), COL (Coleman Hatchery/Battle Creek), MIL (Mill Creek), DER (Deer Creek), BUT (Butte Creek), FRH (Feather River Hatchery), NIM (Nimbus Hatchery/American River), MKH (Mokelumne River Hatchery), STN (Stanislaus River), TOU (Tuolumne River), MER (Merced River).

significantly more positive compared to the distribution of simulated neutral loci. By contrast, fall-run individuals from Mill Creek, Nimbus Hatchery, Butte Creek, Merced River, Tuolumne River, and Coleman Hatchery exhibit distribution of Tajima's D that are significantly more negative compared to the corresponding simulated dataset. Notably, despite the distribution of observed and simulated datasets themselves being different for both fall-run and spring-run Feather River individuals, their median values are the same (Table S11).

Assessment of Genomic Diversity: Measures of Unique Diversity

Upper Sacramento late-fall-run and Butte Creek spring-run individuals exhibited the highest number of fixed loci (6,416 and 6,210, respectively); for all other run/tributary groups, 3,500 to 5,000 fixed loci were identified (Figure 4A). The number of fixed loci in spring-run tributaries was generally larger compared to fall-run tributaries, with the exception of Feather River spring-run. By far the two largest intersects were loci fixed exclusively in a single location: upper Sacramento late-fall-run (460) and Butte Creek spring-run (317). All other intersects were

Table 1 Summary of sample locations, abbreviations, sample sizes and metrics of genomic diversity used in this study. Where mean values are reported, the standard deviation is included in parentheses below the mean.

Run	Tributary/Hatchery	Abbrev.	N	Ho	Hs	Fis	AR	Even.	π	Taj. D	fixed	Singlet.	priv. polym.	priv. alleles
Fall	Coleman Hatchery	F_COL	30	0.1713 (0.1186)	0.1693 (0.1798)	-0.017 (0.182)	1.51 (0.48)	0.7645 (0.1556)	0.001058 (0.001275)	-0.0714 (0.9426)	3835	1.55 (1.00)	57	140.21 (9.71)
	Mill Creek	F_MIL	20	0.1699 (0.1922)	0.1714 (0.1845)	-0.001 (0.232)	1.52 (0.48)	0.7681 (0.1568)	0.001056 (0.001298)	-0.0932 (0.9536)	3574	1.38 (0.65)	83	130.8 (6.84)
	Deer Creek	F_DER	15	0.1674 (0.1939)	0.1685 (0.1850)	-0.002 (0.247)	1.51 (0.50)	0.7631 (0.1612)	0.001048 (0.001294)	-0.1072 (0.9625)	4898	1.67 (1.12)	49	140.63 (6.53)
	Butte Creek	F_BUT	21	0.1641 (0.1884)	0.1662 (0.1827)	0.007 (0.221)	1.51 (0.49)	0.7617 (0.1602)	0.001046 (0.001292)	-0.0955 (0.9564)	4702	1.33 (0.82)	18	90.48 (7.57)
	Feather River Hatchery	F_FRH	27	0.1741 (0.1908)	0.1718 (0.1810)	-0.019 (0.188)	1.52 (0.49)	0.7697 (0.1540)	0.001045 (0.001272)	-0.0760 (0.9461)	4385	1.71 (0.92)	47	139.5 (13.8)
	Nimbus Hatchery	F_NIM	30	0.1688 (0.1874)	0.1678 (0.1804)	-0.010 (0.184)	1.51 (0.49)	0.7631 (0.1582)	0.001043 (0.001271)	-0.0596 (0.9516)	4369	2.21 (1,67)	33	123.39 (11.65)
	Mokelumne River Hatchery	F_MKH	28	0.1689 (0.1884)	0.1680 (0.1810)	-0.013 (0.183)	1.51 (0.48)	0.7632 (0.1580)	0.001042 (0.001276)	-0.0589 (0.9531)	3788	2 (0.87)	54	115 (11.75)
	Stanislaus River	F_STN	23	0.1673 (0.1905)	0.1670 (0.1827)	-0.006 (0.210)	1.51 (0.48)	0.7631 (0.1596)	0.001038 (0.001281)	-0.0809 (0.9626)	3609	1 (0.00)	55	104.04 (7.96)
	Tuolumne River	F_TOU	30	0.1659 (0.1849)	0.1676 (0.1807)	0.006 (0.194)	1.51 (0.48)	0.762 (0.1585)	0.001035 (0.001275)	-0.0683 (0.9516)	3655	1.36 (0.63)	82	94.75 (8.86)
	Merced River Hatchery	F_MRH	15	0.1658 (0.1938)	0.1695 (0.1870)	0.014 (0.263)	1.52 (0.48)	0.7648 (0.1614)	0.001035 (0.001311)	-0.0883 (0.9595)	3504	1.71 (0.76)	89	132.01 (6.33)
	Merced River	F_MER	31	0.167 (0.1846)	0.1676 (0.1799)	-0.003 (0.182)	1.51 (0.48)	0.7634 (0.1581)	0.001034 (0.001274)	-0.0636 (0.9539)	4191	1.47 (0.80)	46	93.19 (8.60)
Late-Fall	Upper Sacramento River	L_USR	21	0.1654 (0.1888)	0.1668 (0.1831)	0.003 (0.214)	1.36 (0.45)	0.7609 (0.1613)	0.001033 (0.001289)	-0.0465 (0.9567)	6416	2.23 (1.54)	70	134.58 (9.58)
Winter	Upper Sacramento River	W_USR	26	0.1272 (0.1840)	0.1285 (0.1789)	0.010 (0.198)	1.52 (0.48)	0.6925 (0.1988)	0.001029 (0.001212)	0.0942 (1.0350)	3853	1.78 (1,39)	153	122.99 (9.40)
Spring	Mill Creek	S_MIL	16	0.1651 (0.1932)	0.1695 (0.1877)	0.020 (0.258)	1.51 (0.50)	0.7663 (0.1615)	0.001028 (0.001315)	-0.0507 (0.9637)	4910	1.82 (1.40)	70	193.20 (10.07)
	Deer Creek	S_DER	27	0.1628 (0.1802)	0.1699 (0.1808)	0.037 (0.208)	1.48 (0.50)	0.7633 (0.1600)	0.001026 (0.001275)	-0.0621 (0.9457)	5179	3 (4.64)	97	203.8 (38.64)
	Butte Creek	S_BUT	19	0.1587 (0.1888)	0.1642 (0.1874)	0.027 (0.227)	1.51 (0.55)	0.7578 (0.1695)	0.001012 (0.001308)	0.0491 (0.9545)	6210	4 (3.46)	44	185.43 (24.75)
	Feather River Hatchery	S_FRH	7	0.1695 (0.2157)	0.1695 (0.1983)	-0.014 (0.325)	1.50 (0.49)	0.7715 (0.1577)	0.00079 (0.001392)	-0.0874 (0.9773)	4423	1.25 (0.50)	84	59.46 (4.87)

< 115 loci. Apart from Feather River hatchery spring-run (27), spring-run tributaries had more loci fixed in a given tributary (74 to 317) compared to fall-run tributaries where 10 to 61 loci were fixed among individuals from a single tributary. Notably, among intersects of loci fixed in two locations, the three largest intersects were all a combination of upper Sacramento River late-fall-run population and a wild spring-run population (42 to 115 loci). Overall, about a third of intersects of two run/tributary combinations were loci fixed among upper Sacramento River late-fall run and a second location. There was no observed pattern of loci more likely to be fixed among tributaries in geographic proximity (Figure 4A).

Loci fixed in upper Sacramento late-fall and Butte Creek spring runs also exhibited the highest global allele diversity (mean = 1.46 and 1.41, respectively; Table S11), i.e., loci that are fixed in these groups are more variable when alleles were tabulated across individuals from all runs/tributaries (Figure 4B). By contrast, the global diversity of fixed alleles is lowest in fall-

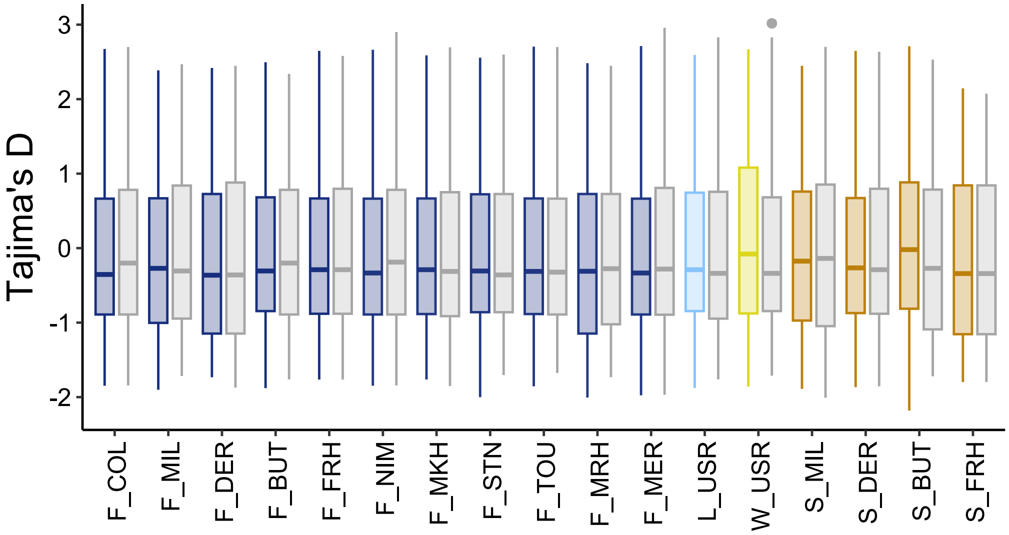

Figure 5 Comparison of observed and simulated distribution of Tajima's D assuming mutation-drift equilibrium. Empirical data sets are colored according to run type (spring = yellow, winter = brown, fall = dark blue, late-fall = light blue), simulated data distributions are depicted in *grey*. Tributaries are labeled with the abbreviations used throughout: USR (upper Sacramento River), COL (Coleman Hatchery/Battle Creek), MIL (Mill Creek), DER (Deer Creek), BUT (Butte Creek), FRH (Feather River Hatchery), NIM (Nimbus Hatchery/American River), MKH (Mokelumne River Hatchery), STN (Stanislaus River), TOU (Tuolumne River), MER (Merced River).

run groups, and overall levels are more similar across tributaries (median = 1.13 to 1.14, with the exception of Deer Creek and Butte Creek fall runs at 1.19) compared to spring tributaries, where those distributions of global diversity were higher and more variable (median = 1.28-1.41; Table S11). The distribution of global diversity among spring-run and fall-run tributaries varied, with some run/tributaries exhibiting much tighter ranges than others (Figure 4B). In general, the proportion of loci that were fixed for a run/tributary group was consistent across chromosomes, and there was no distinct, non-random pattern (Tables S12 and S13).

Three hundred forty-seven (1.2% of the total) loci exhibited at least one singleton. A comparison of individuals grouped by run/tributary demonstrates that Butte Creek and Deer Creek spring-run individuals (mean = 4.0 and 3.0, respectively) and upper Sacramento River late-fall individuals (mean = 2.23) exhibited the highest mean number of singletons per individual. In general, fall tributaries exhibited a lower mean number of singletons per individual compared to other run/tributary groups, indicating that they were comparatively less characterized by rare alleles (Figure 4C; Table S14).

At a population level, spring run from Deer Creek had the highest number of loci with private polymorphisms (153); all other groups had <100 loci with private polymorphisms (Table 1). Spring run from Feather River Hatchery exhibited the lowest number of loci exclusively polymorphic among individuals of a group (18). Eighty-four loci with private polymorphisms were identified in late-fall-run individuals. This number was higher than observed for all but one fall-run group (Coleman Hatchery). Notably, individuals from hatcheries fall along the entire range of private polymorphisms; Nimbus Hatchery fall-run individuals were on the low end (33) and Feather River Hatchery spring-run individuals on the high end (84). Comparing chromosomal positions of private polymorphisms indicated that Deer Creek spring run had the highest mean proportion of loci on a chromosome that are fixed in all other groups (0.10). The second highest mean proportion of loci per chromosome is 0.07, observed in Coleman Hatchery fall-run, Feather River fall-run, and Butte Creek spring-run individuals. The mean was lowest for Feather River spring run (Table 1, Table S13). While there were chromosomes with significantly more/fewer than expected numbers of loci with private polymorphisms, no consistent non-random patterns stood out (Table S14). For wild individuals grouped by run type, the largest number of private polymorphisms was observed in fall-run individuals (1,348) compared to 580 private polymorphisms observed only in spring-run individuals. By contrast, only 122 and 95 private polymorphisms were observed in late-fall-run and winter-run individuals, respectively, though it should be noted that only fall and spring runs are represented by multiple tributaries.

While there was no distinct pattern of hatchery individuals that exhibited more/fewer private alleles compared to wild individuals in general, notably, wild spring-run tributary populations exhibit the highest number of private alleles (185 to 204; Table 1) and Feather River Hatchery spring-run individuals exhibit the lowest number (60). Upper Sacramento River winter-run and Butte Creek spring-run individuals exhibited the most "common" private alleles, carried in 25 and ten individuals, respectively (Table S16). Comparing the chromosomal positions of loci with private alleles in a single run/tributary group indicated that Deer Creek spring-run individuals had the highest mean proportion of loci with private alleles on a chromosome (Table S17). While there were chromosomes with higher or lower than expected number of private alleles for run/tributary groups, no distinct patterns emerged (Tables S18 and S19). For wild individuals grouped by run type, the largest number of private alleles was observed in spring-run individuals (mean = 1,094.6). Fall-run and late-fall-run individuals exhibited 861.22 and 807.24 private alleles, respectively. Winter-run individuals exhibited the lowest number of private alleles (385.09).

DISCUSSION

Overall, our re-assessment using multi-allelic loci supports the population-level findings of Meek et al. (2019; Figure 2), while also allowing discovery of the fine-scale patterns of genomic diversity and a fuller understanding of the biocomplexity contained within and among Central Valley Chinook Salmon populations. The initial analysis in Meek et al. (2019) of the dataset presented here focused on assessing population structure within and among runs present in the Central Valley; we extend this analysis with an in-depth exploration of genomic diversity using the wide range of metrics presented here, which adds critical context to the ongoing discussion about how to determine conservation goals to ensure the persistence of a population complex such as Central Valley Chinook Salmon. We identified significant differentiation among and within run types and the tributaries they inhabit, and low levels of genetic diversity in populations that have experienced recent demographic declines. Notably, each run/tributary group was indeed characterized by a distinct component of unique genomic diversity, despite apparent gene flow between individuals from different tributaries that shared the same run type. Maintaining this unique diversity present in distinct population segments is likely important to the overall genetic health of both individual populations and the population complex as a whole (portfolio effect), and is thus vital to identify and consider in conservation efforts. Overall, our results emphasize the importance of not only maintaining life-history (phenotypic) diversity within and among groups, but also maintaining the genetic diversity of each run and tributary to enhance the portfolio effect, maintain adaptive potential, and ensure the long-term persistence of Chinook Salmon in the Central Valley.

The population structure analysis presented here (Figure 2; Table S2 and Figure S2) confirms that fall and late-fall runs within the Central Valley are genetically more similar to each other than either is to the winter or spring runs, and they are managed under a single Evolutionarily Significant Unit (ESU). However, the comprehensive assessment of genomic diversity using multi-allelic haplotyped loci adds to increasing evidence that fall and late-fall individuals also exhibit genetic distinctness. In addition, we found the late-fall-run individuals to have overall lower diversity compared to fall-run populations, which likely reflects differences in natural and/or anthropogenic forces acting on the two runs. Though it is important to acknowledge that our late-fall run comprises individuals from only one tributary, this finding highlights that fall and late-fall represent distinct components of both genetic and phenotypic diversity within their ESU as well as the Central Valley population complex as a whole. Given that late-fall-run fish occupy more constrained habitat (currently, spawning primarily occurs only in the tributaries of the upper Sacramento River), have a smaller population size, and lower genetic diversity relative to fall-run fish, conservation actions aimed at specifically monitoring and promoting the distinctiveness of late-fall-run fish may be warranted and necessary to preserve this diversity, especially in the light of the widespread homogenization that has occurred throughout the fall run. Similarly, the Feather River Hatchery spring-run individuals highlight the effects of management practices. The Feather River Hatchery produces both spring-run and fall-run individuals, which has resulted in hybridization between the two runs (Huber and Carlson 2015). As a result~~,~~ spring-run Feather River Hatchery individuals cluster with fall-run individuals (Figure 2), and in general they show higher levels of genetic diversity across most metrics assessed here (Figure 3).

Decreased levels of heterozygosity are the result of the erosion of genetic diversity in small populations, as the rate of some alleles being lost and others fixed as a result of drift increases, and individuals become increasingly likely to mate with individuals with similar genotypes. Accordingly, we observed the highest mean F_{IS} values in the winter-run and spring-run populations. These positive values are the result of an excess of homozygous genotypes, likely the result of genetically similar (closely related) individuals mating, which becomes increasingly likely as population sizes decrease (Figure 3). Additionally, both winter-run and late-fall-run populations represented here exhibit Tajima's D distributions that are more positive compared to corresponding datasets simulated under mutation-drift equilibrium (Figure 5). This indicates a lack of rare alleles, which are more likely to be lost when populations experience drastic declines and genetic drift increases. If drift is the primary evolutionary force shaping the genetic diversity within the declining late-fall-run, winter-run, and spring-run populations, we would expect to observe not only low levels of heterozygosity (Figure 3) but also comparatively higher numbers of fixed alleles (Figure 4). Indeed, though late-fall individuals typically look similar to fall-run populations when metrics that assess levels of heterozygosity and allelic diversity are compared, we see that they exhibit the highest number of loci fixed in only one location, and the distribution of Tajima's D indicates a population bottleneck that results in a lack of rare alleles in this population (Figures 3 through 5). Similarly, winter-run individuals, which have experienced precipitous demographic declines as a result of habitat modifications, exhibit the lowest levels of heterozygosity and other measures of genetic diversity compared to all other groups. Next to late-fall individuals, the highest number of fixed loci is found in spring-run groups (Figure 4); and, overall, we find much wider distributions across all measures of diversity among spring-run groups (Figure 3). This underscores the stochasticity of genetic drift. Not only is drift accelerated in smaller populations, but *how* each population segment is affected will differ. Thus, groups may diverge from each other by chance alone. Additionally, the differences in the environmental conditions that characterize each tributary can be expected to contribute to each tributary experiencing a different selection regime, which results in disparate effects of decoupled demographic and environmental stochastic events affecting each population.

The assessment of private polymorphisms and private alleles reveals that within each tributary each run exhibits unique components of genetic diversity. Despite upper Sacramento River winter-run individuals having the lowest level of diversity when measures related to heterozygosity and allelic diversity are compared, they exhibit the highest level of private polymorphisms, though they exhibit less diversity overall, i.e., diversity that is present is unique compared to all the other runs. By contrast, Butte Creek spring-run individuals exhibit levels of heterozygosity and allelic diversity similar to winter-run individuals, but also harbor a low number of private polymorphisms, which indicates existing differences in the demographic and evolutionary forces that shape genetic diversity across groups (Table 1, Figure 3). As a result, the genome-wide intraspecific diversity is unique within each group. This suggests that there are differences in the standing differences within each group beyond loci controlling migration timing upon which selective pressure may act. This could include differences in hydrology, thermal regimes, and varying levels of anthropogenic effect on habitat, among other things. Additionally, our new analysis reveals clear distinctions in the unique diversity harbored by late-fall-run individuals from the upper Sacramento River compared to fall-run groups, despite fall-run and late-fall-run individuals sharing GREB1L genotypes and being managed as a single ESU. The late-fall-run population has lower overall allele counts compared to fall-run populations, and the number of private polymorphisms is higher for late-fall-run individuals than almost all of the fall-run groups. Further assessment of differences in the evolutionary trajectories of the run types using a larger number of late-fall-run samples is warranted.

The observed presence of unique diversity among and within individual components of the Central Valley population complex underscores the importance of a management strategy that seeks to maintain a robust portfolio at both a phenotypic and genotypic level. While it is important to acknowledge that the (neutral) genetic diversity of a population is not always correlated with functional diversity (Reed and Frankham 2001), the variation of genotypes among individuals has been demonstrated to be a suitable proxy to predict fitness of individuals as well as the ability of populations and ecosystems to respond to changes in environmental conditions (Vazquez–Dominguez et al.1999; Reed and Frankham 2003; Reusch et al. 2005; Hoffman et al. 2014). Furthermore, examples from translocation and genetic rescue efforts have demonstrated that heterozygosity and genetic diversity can be more efficient predictors of success than the ability to match (neutral) genotypes as closely as possible to individuals already present in the population (Coleman et al. 2013; Scott et al. 2020). Losing early-run populations—which are currently the most imperiled in the Central Valley—therefore runs the risk of losing both early-run alleles (i.e., the ability to recover the early-run phenotype) and the more cryptic yet likely important unique components of genetic diversity harbored among and within migration phenotypes.

Spatially and temporally heterogeneous environments promote and maintain polymorphisms and high levels of standing genetic variation that form a diverse portfolio of genetic and phenotypic diversity which enables the population to persist (Gulisija and Kim 2015; Svardal et al. 2015; Bertram and Masel 2019). Therefore, the loss of diversity and increasing genetic homogenization may be a more important factor that drives the loss of the portfolio than demographic synchronization itself (Dedrick and Baskett 2018; Des Roches et al. 2021b). Because of their complex life history, environmental pressures differ widely across salmonid salmon life stages such that the genotypes and phenotypes that confer higher survival probability at one life stage do not necessarily translate into the genotypes and phenotypes that best match conditions during a different life stage. Additionally, climate change will affect environmental conditions in individual tributaries differently, again necessitating genomic diversity across the Central Valley to allow adaption to changing conditions (Yates et al. 2008). Important phenotypic traits—including growth, temperature tolerance, and stress responses—are likely polygenic traits, controlled by many loci of small effects, and populations characterized by the presence of a large proportion of polygenic traits are more likely to adapt to new conditions, and therefore increase population viability with rapidly fluctuating environmental conditions (Kardos and Luikart 2019). This points to the importance of conserving genetic diversity as a whole to ensure adaptive potential is maintained.

Despite a history of anthropogenic stressors expected to erode and homogenize intraspecific genetic diversity, our study shows that each population of Central Valley Chinook Salmon still harbors unique variation that adds to the biocomplexity of the whole system. While we cannot quantify the amount of genetic diversity lost with the data on hand, our results demonstrate the need to preserve the existing standing genetic diversity along with previously identified functional diversity. Therefore, preserving this standing genetic diversity along with identified functional diversity (e.g., GREB1L) is critical to provide Central Valley populations with the best chance at a healthy portfolio and the ability to adapt to changing conditions (Hairston et al. 2005; Richardson et al. 2014; Messer et al. 2016b). Even though anthropogenic effects significantly alter the composition and structure of both neutral and functional diversity at a genetic level, the conservation of intraspecific genetic diversity is frequently overlooked (Laikre et al. 2010; Des Roches et al. 2021b), despite its serving as the fundamental building block of biodiversity. For example, the remaining wild spring-run Chinook Salmon populations in the Central Valley still access habitat that is very distinct from their fall-run counterparts. In addition, a great deal of habitat heterogeneity exists within the spring-run populations. This could facilitate the development of important local adaptive differences both between runs and between the different spring-run tributaries. Indeed, our finding of unique diversity within each run/tributary group that comprises the population complex of Chinook Salmon in California's Central Valley underscores the importance of monitoring intraspecific genomic diversity at multiple levels (across and within locations and life-history phenotypes). To ensure persistence in a rapidly changing and increasingly unfavorable environment, genetic monitoring should inform conservation and management policies that counteract genetic homogenization and conserve the biodiversity and biocomplexity of the population complex as a whole.

CONCLUSION

This study highlights the importance of understanding genetic diversity at fine scales to promote overall species biocomplexity. Our work brings to light that unique variation is found both between the major demographic groups of the Central Valley and between different sub-populations within those groups. Notably, unique variation was found in both late-fall-run and even across the many fall-run populations (which are often currently managed as if they are genetically homogenous)—something which has been previously unreported. To promote the full portfolio of diversity in the Central Valley, we must identify and monitor diversity at this level or run the risk of losing important variation that is foundational to population and species persistence. This is particularly true given rapid environmental change, when it is very difficult to predict what variation is going to be needed to withstand and thrive in future conditions. Overall, our results emphasize the necessity of maintaining biocomplexity at multiple scales, because this is an important factor that determines resilience to changing environmental pressures. Management actions that avoid artificial homogenization of populations, and maintain unique components of genetic diversity within populations, are needed to maintain this important biocomplexity.

ACKNOWLEDGEMENTS

The authors would like to thank the members of the Meek Lab for helpful discussions and Melinda Baerwald for helpful feedback on early versions of the manuscript. Additionally, we would like to thank Pascale Goertler for important discussions related to the project.

REFERENCES

Baetscher DS, Clemento AJ, Ng TC, Anderson EC, Garza JC. 2018. Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. Mol Ecol Resour. [accessed 2021 Apr 20];18(2):296–305. *https://doi.org/10.1111/1755-0998.12737*

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc. [accessed 2018 Aug 27];57(1):289–300. *https://doi.org/10.1111/j.2517-6161.1995.tb02031.x*

Bertram J, Masel J. 2019. Different mechanisms drive the maintenance of polymorphism at loci subject to strong versus weak fluctuating selection. Evolution. [accessed 2020 Dec 29];73(5):883–896. *https://doi.org/10.1111/evo.13719*

Carlson SM, Satterthwaite WH. 2011. Weakened portfolio effect in a collapsed salmon population complex. Can J Fish Aquat Sci. [accessed 2020 Dec 21];68(9):1579–1589. *https://doi.org/10.1139/f2011-084*

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. *Stacks*: building and genotyping loci *de novo* from short-read sequences. G3: Genes|Genomes|Genetics. [accessed 2015 Mar 31];1(3):171–182. *https://doi.org/10.1534/g3.111.000240*

[CBD] Convention on Biological Diversity. 2011. Strategic plan for biodiversity 2011–2020, including Aichi biodiversity targets. [accessed 2020 Dec 21]. Nagoya (Japan): Convention on Biological Diversity. Montreal, Canada: Secretariat of the Convention on Biological Diversity.

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. [accessed 2019 Aug 27];34(17):i884–i890. *https://doi.org/10.1093/bioinformatics/bty560*

Christensen KA, Leong JS, Sakhrani D, Biagi CA, Minkley DR, Withler RE, Rondeau EB, Koop BF, Devlin RH. 2018. Chinook Salmon (*Oncorhynchus tshawytscha*) genome and transcriptome. PLOS ONE 13(4):e0195461. [accessed 2019 Oct 24]. *https://doi.org/10.1371/journal.pone.0195461*

Coleman RA, Weeks AR, Hoffmann AA. 2013. Balancing genetic uniqueness and genetic variation in determining conservation and translocation strategies: a comprehensive case study of threatened dwarf galaxias, *Galaxiella pusilla* (Mack) (Pisces: Galaxiidae). Mol Ecol. [accessed 2021 Mar 25];22(7):1820–1835. *https://doi.org/10.1111/mec.12227*

Dedrick AG, Baskett ML. 2018. Integrating genetic and demographic effects of connectivity on population stability: the case of hatchery trucking in salmon. Am Nat. [accessed 2020 Dec 29];192(2):E62–E80. *https://doi.org/10.1086/697581*

Duelli P, Obrist MK. 2003. Biodiversity indicators: the choice of values and measures. Agric Ecosyst Environ. [accessed 2020 Dec 29];98(1–3):87–98. *https://doi.org/10.1016/S0167-8809(03)00072-0*

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. PLOS ONE. [accessed 2020 Dec 29];11(3):e0151651. *https://doi.org/10.48550/arXiv.1207.3907*

Gosselin T, Anderson EC, Bradbury IR. 2016. assigner: assignment analysis with GBS/RAD data using R. R package. [accessed 2020 Dec 29]. *https://doi.org/10.5281/zenodo.51453*

Griffiths JR, Schindler DE, Armstrong JB, Scheuerell MD, Whited DC, Clark RA, Hilborn R, Holt CA, Lindley ST, Stanford JA, et al. 2014. Performance of salmon fishery portfolios across western North America. J App Ecol. [accessed 2020 Dec 21];51(6):1554–1563. *https://doi.org/10.1111/1365-2664.12341*

Gulisija D, Kim Y. 2015. Emergence of long-term balanced polymorphism under cyclic selection of spatially variable magnitude. Evolution. [accessed 2020 Dec 29];69(4):979–992. *http://doi.org/10.1111/evo.12630*

Hairston NG, Ellner SP, Geber MA, Yoshida T, Fox JA. 2005. Rapid evolution and the convergence of ecological and evolutionary time. Ecol Lett. [accessed 2020 Dec 29];8(10):1114–1127. *https://doi.org/10.1111/j.1461-0248.2005.00812.x*

Hilborn R, Quinn TP, Schindler DE, Rogers DE. 2003. Biocomplexity and fisheries sustainability. Proc Natl Acad Sci. [accessed 2020 Dec 29];100(11):6564–6568. *https://doi.org/10.1073/pnas.1037274100*

Hoban S, Bruford M, D'Urban Jackson J, Lopes–Fernandes M, Heuertz M, Hohenlohe PA, Paz–Vinas I, Sjögren–Gulve P, Segelbacher G, Vernesi C, et al. 2020. Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. Biol Conserv. [accessed 2024 Sep 25];248:108654. *https://doi.org/10.1016/j.biocon.2020.108654*

Hoelzel AR, Bruford MW, Fleischer RC. 2019. Conservation of adaptive potential and functional diversity. Conserv Genet. [accessed 2024 Sep 25];20(1):1–5. *https://doi.org/10.1007/s10592-019-01151-x*

Hoffman JI, Simpson F, David P, Rijks JM, Kuiken T, Thorne MAS, Lacy RC, Dasmahapatra KK. 2014. High-throughput sequencing reveals inbreeding depression in a natural population. Proc Natl Acad Sci. 111(10):3775–3780. [accessed 2020 Dec 25]. *https://doi.org/10.1073/pnas.1318945111*

Hudson RR. 2002. Generating samples under a Wright–Fisher neutral model of genetic variation. Bioinformatics. 18(2):337–338. [accessed 2019 Feb 7]. *https://doi.org/10.1093/bioinformatics/18.2.337*

Jombart T. 2008. adegenet: an R package for the multivariate analysis of genetic markers. Bioinformatics. [accessed 2019 Feb 7];24(11):1403–1405. *https://doi.org/10.1093/bioinformatics/btn129*

Kaeuffer R, Réale D, Coltman DW, Pontier D. 2007. Detecting population structure using STRUCTURE software: effect of background linkage disequilibrium. Heredity. [accessed 2021 Jun 30];99(4):374–380. *https://doi.org/10.1038/sj.hdy.6801010*

Kalinowski S. 2002. How many alleles per locus should be used to estimate genetic distances? Heredity. [accessed 2021 Jun 30];88(1):62–65. *https://doi.org/10.1038/sj.hdy.6800009*

Kamvar ZN, Tabima JF, Grünwald NJ. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ. [accessed 2018 Aug 27];2:e281. *https://doi.org/10.7717/peerj.281*

Kardos M, Northwest Fisheries Science Center, Luikart G. 2020. The genetic architecture of fitness drives population viability during rapid environmental change. bioRxiv. [accessed 2020 Aug 7];660803. *https://doi.org/doi:10.1101/660803*

Laikre L, Schwartz MK, Waples RS, Ryman N, The GeM Working Group. 2010. Compromising genetic diversity in the wild: unmonitored large-scale release of plants and animals. Trends Ecol Evol. [accessed 2020 Aug 7];25(9):520–529. *https://doi.org/10.1016/j.tree.2010.06.013*

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. [accessed 2018 Aug 27];25(14):1754–1760. *https://doi.org/10.1093/bioinformatics/btp324*

Mable BK. 2019. Conservation of adaptive potential and functional diversity: integrating old and new approaches. Conserv Gen. [accessed 2020 Aug 7];20(1):89–100. *https://doi.org/10.1007/s10592-018-1129-9*

Meek MH, Stephens MR, Goodbla A, May BP, Baerwald MR. 2019. Identifying hidden biocomplexity and genomic diversity in Chinook Salmon, an imperiled species with a history of anthropogenic influence. Can J Fish Aquat Sci. [accessed 2019 Nov 18];77(3). *https://doi.org/10.1139/cjfas-2019-0171*

Messer PW, Ellner SP, Hairston NG. 2016a. Can population genetics adapt to rapid evolution? Trends Gen. [accessed 2020 Aug 7];32(7):408–418. *https://doi.org/10.1016/j.tig.2016.04.005*

Miller MR, Brunelli JP, Wheeler PA, Liu S, Rexroad CE, Palti Y, Doe CQ, Thorgaard GH. 2012. A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. Mol Ecol. [accessed 2018 Sep 15];21(2):237–249. *https://doi.org/10.1111/j.1365-294X.2011.05305.x*

Mimura M, Yahara T, Faith DP, Vázquez–Domínguez E, Colautti RI, Araki H, Javadi F, Núñez–Farfán J, Mori AS, Zhou S, et al. 2017. Understanding and monitoring the consequences of human impacts on intraspecific variation. Evol Appl. [accessed 2020 Aug 7];10(2):121–139. *https://doi.org/10.1111/eva.12436*

Morin PA, Martien KK, Taylor BL. 2009. Assessing statistical power of SNPs for population structure and conservation studies. Mol Ecol Resour. [accessed 2021 Apr 20];9(1):66–73. *https://doi.org/10.1111/j.1755-0998.2008.02392.x*

Moyle PB, Lusardi RA, Samuel PJ, Jacob MA, Katz VE. 2017. State of the salmonids: status of California’s emblematic fishes. A report commissioned by California Trout. [accessed 2020 Aug 7]. San Francisco (CA): Center for Watershed Sciences, University of California–Davis and California Trout.

Narum SR, Genova A Di, Micheletti SJ, Maass A. 2018. Genomic variation underlying complex life-history traits revealed by genome sequencing in Chinook Salmon. Proc Royal Soc B: Bio Sci. 285(1883):20180935. [accessed 2019 Nov 20]. *https://doi.org/10.1098/rspb.2018.0935*

[NMFS] National Marine Fisheries Service. 2005. Endangered and threatened species: final listing determinations for 16 ESUs of west coast salmon, and final 4(d) protective regulations for threatened salmonid ESUs. Fed Regist. [accessed 2020 Dec 21];70(123):37160–37204. Available from: *https://www.federalregister.gov/documents/2005/06/28/05-12351/endangered-and-threatened-species-final-listing-determinations-for-16-esus-of-west-coast-salmon-and*

Nei M. 1987. Molecular evolutionary genetics. New York (NY): Columbia University Press. 514 p.

O'Leary SJ, Puritz JB, Willis SC, Hollenbeck CM, Portnoy DS. 2018. These aren't the loci you're looking for: principles of effective SNP filtering for molecular ecologists. Mol Ecol. [accessed 2020 Dec 21];27(16):3193–3206. *https://doi.org/10.1111/mec.14792*

Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW. 2010. Conservation genetics in transition to conservation genomics. Trends Gen. [accessed 2020 Dec 21];26(4):177–187. *https://doi.org/10.1016/j.tig.2010.01.001*

Paradis E, Barrett J. 2010. pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics. [accessed 2022 Jun 28];26(3):419–420. *https://doi.org/doi:10.1093/bioinformatics/btp696*

Péru N, Dolédec S. 2010. From compositional to functional biodiversity metrics in bioassessment: a case study using stream macroinvertebrate communities. Ecol Indic. [accessed 2020 Dec 21];10(5):1025–1036. *https://doi.org/10.1016/j.ecolind.2010.02.011*

Petchey OL, Gaston KJ. 2002. Functional diversity (FD), species richness and community composition. Ecol Lett. [accessed 2021 Feb 19];5(3):402–411. *https://doi.org/10.1046/j.1461-0248.2002.00339.x*

Prince DJ, O'Rourke SM, Thompson TQ, Ali OA, Lyman HS, Saglam IK, Hotaling TJ, Spidle AP, Miller MR. 2017. The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. Sci Adv. [accessed 2020 Feb 5];3(8):e1603198. *https://doi.org/10.1126/sciadv.1603198*

Quinlan AR. 2014. BEDTools: the Swiss-Army tool for genome feature analysis. Curr Protoc Bioinformatics. [accessed 2019 Sep 3];2014(1):11.12.1–11.12.34. *https://doi.org/10.1002/0471250953.bi1112s47*

Quinn TP. 2018. The behavior and ecology of Pacific salmon and trout. 2nd ed. Seattle (WA): University of Washington Press. 562 p.

Quinn TP, McGinnity P, Reed TE. 2016. The paradox of "premature migration" by adult anadromous salmonid fishes: patterns and hypotheses. Can J Fish Aquat Sci. [accessed 2020 Aug 7];73(7):1015–1030. *https://doi.org/10.1139/cjfas-2015-0345*

Reed DH, Frankham R. 2001. How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. Evolution. [accessed 2020 Aug 7];55(6):1095–1103. *https://doi.org/10.1111/j.0014-3820.2001.tb00629.x*

Reed DH, Frankham R. 2003. Correlation between fitness and genetic diversity. Conserv Biol. [accessed 2010 Jul 10];17(1):230–237. *https://doi.org/10.1046/j.1523-1739.2003.01236.x*

Reusch TBH, Ehlers A, Hämmerli A, Worm B. 2005. Ecosystem recovery after climatic extremes enhanced by genotypic diversity. Proc Natl Acad Sci. 102(8):2826–2831. [accessed 2020 Dec 25]. *https://doi.org/10.1073/pnas.0500008102*

Richardson JL, Urban MC, Bolnick DI, Skelly DK. 2014. Microgeographic adaptation and the spatial scale of evolution. Trends Ecol Evol. 29(3):165–176. [accessed 2018 Sep 15]. *https://doi.org/10.1016/j.tree.2014.01.002*

Des Roches S, Pendleton LH, Shapiro B, Palkovacs EP. 2021a. Conserving intraspecific variation for nature's contributions to people. Nat Ecol Evol. [accessed 2020 Dec 25];5(2021):574–582. *https://doi.org/10.1038/s41559-021-01403-5*

Schindler DE, Hilborn R, Chasco B, Boatright CP, Quinn TP, Rogers LA, Webster MS. 2010. Population diversity and the portfolio effect in an exploited species. Nature. [accessed 2020 Dec 25];465(7298):609–612. *https://doi.org/10.1038/nature09060*

Scott PA, Allison LJ, Field KJ, Averill–Murray RC, Bradley Shaffer H. 2020. Individual heterozygosity predicts translocation success in threatened desert tortoises. Science. [accessed 2021 Mar 25];370(6520):1086–1089. *https://doi.org/10.1126/science.abb0421*

Sinclair–Waters M, Ødegård J, Korsvoll SA, Moen T, Lien S, Primmer CR, Barson NJ. 2020. Beyond large-effect loci: large-scale GWAS reveals a mixed large-effect and polygenic architecture for age at maturity of Atlantic salmon. Gen Sel Evol. [accessed 2020 Aug 7];52(1):9. *https://doi.org/10.1186/s12711-020-0529-8*

Svardal H, Rueffler C, Hermisson J. 2015. A general condition for adaptive genetic polymorphism in temporally and spatially heterogeneous environments. Theor Popul Biol. [accessed 2020 Dec 25];99:76–97. *https://doi.org/10.1016/j.tpb.2014.11.002*

Thompson NF, Anderson EC, Clemento AJ, Campbell MA, Pearse DE, Hearsey JW, Kinziger AP, Garza JC. 2020. A complex phenotype in salmon controlled by a simple change in migratory timing. Science. [accessed 2021 Jan 15];370(6516):609–613. *https://doi.org/10.1126/science.aba9059*

Thompson TQ, Bellinger MR, O'Rourke SM, Prince DJ, Stevenson AE, Rodrigues AT, Sloat MR, Speller CF, Yang DY, Butler VL, et al. 2018. Anthropogenic habitat alteration leads to rapid loss of adaptive variation and restoration potential in wild salmon populations. Proc Natl Acad Sci. [accessed 2020 Aug 7];116(1):177–186. *https://doi.org/10.1073/pnas.1811559115*

Vazquez–Dominguez E, Pinero D, Ceballos G. 1999. Linking heterozygosity, demography, and fitness of tropical populations of *Liomys pictus*. J Mammal. [accessed 2020 Dec 25];80(3):810–822. *https://doi.org/10.2307/1383250*

Waples RS, Ford MJ, Nichols K, Kardos M, Myers J, Thompson TQ, Anderson EC, Koch IJ, McKinney G, Miller MR, et al. 2022. Implications of large-effect loci for conservation: a review and case study with Pacific salmon. J Heredity. [accessed 2020 Dec 25];113(2):121–144. *https://doi.org/10.1093/jhered/esab069*

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution. [accessed 2018 Aug 27];38(6):1358–1370. *https://doi.org/10.2307/2408641*

Williams JG. 2006. Central Valley salmon: a perspective on Chinook and Steelhead in the Central Valley of California. San Franc Estuary Watershed Sci. [accessed 2020 Dec 21];4(3). *https://doi.org/10.15447/sfews.2006v4iss3art2*

Willis SC, Hollenbeck CM, Puritz JB, Gold JR, Portnoy DS. 2017. Haplotyping RAD loci: an efficient method to filter paralogs and account for physical linkage. Mol Ecol Resour. [accessed 2020 Dec 21];17(5):955–965. *https://doi.org/10.1111/1755-0998.12647*

Yates D, Galbraith H, Purkey D, Huber–Lee A, Sieber J, West J, Herrod–Julius S, Joyce B. 2008. Climate warming, water storage, and Chinook Salmon in California's Sacramento Valley. Clim Change. [accessed 2020 Dec 29];91(3–4):335–350. *https://doi.org/10.1007/s10584-008-9427-8*