Title

Every cog and wheel: Unraveling biocomplexity at the genomic and phenotypic level in a population complex of Chinook salmon.

Authors

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

Affiliations

[1]Saint Anselm College, Department of Biology, 100 Saint Anselm Drive, Manchester, NH 03102, United States

[2]Department of Integrative Biology, AgBioResearch, and Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, MI, United States

*corresponding author: mhmeek@msu.edu

Abstract

Genomic diversity is the fundamental building block of biodiversity and the necessary ingredient for adaptation. Our rapidly increasing ability to quantify functional, compositional, and structural genomic diversity of populations forces the question of how to balance conservation goals – should the focus be on important functional diversity and key life history traits or on maximizing genomic diversity as a whole? Specifically, the intra-specific diversity (biocomplexity) comprised of phenotypic and genetic variation can determine the ability of a population to respond to changing environmental conditions. Here, we explore the biocomplexity of California's Central Valley Chinook salmon population complex at a genomic level. Notably, despite apparent gene flow among individuals of different migration (life history) phenotypes inhabiting the same tributaries, each group is characterized by a component of unique genomic diversity. Our results emphasize the importance of formulating conservation goals focused on maintaining biocomplexity at both the phenotypic and genotypic level. Doing so will maintain the adaptive potential to increase the probability of persistence of the population complex despite changing environmental pressures.

Keywords

Portfolio effect, intraspecific diversity, conservation genomics

Background

> *"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

Genetic diversity represents the fundamental building block of biodiversity – from genes to species, to communities and ecosystems. The genetic diversity contained within populations forms the cogs and wheels comprising the adaptive potential of a population. Protecting this diversity is a central task of conservation biology. However, identifying and quantifying biodiversity is a complex undertaking with distinct frameworks for enumerating compositional, structural, and functional diversity [1–3]. Functional genetic diversity includes both sequence polymorphisms and differences in gene expression, which together shape the phenotypic diversity composed of differences in morphology, physiology, and life history characteristics present within a population. Under current environmental conditions, it is expected that some components of functional diversity will be selectively neutral, but future shifts in conditions and evolutionary pressures may result in certain phenotypes and genotypes that appeared unimportant under previous conditions becoming critical to the persistence of a species or population [4]. Therefore, in order to maintain critical evolutionary and ecological processes that sustain biodiversity across scales, sound conservation and management strategies require a fundamental understanding of both past and present patterns of biodiversity to enable the protection of diversity that will form the building blocks for future adaptation [5–7].

Increasingly, the importance of intra-specific diversity (biocomplexity) has been recognized as a determining factor for the stability and resilience of biological systems [8,9]. 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 diversity of life history strategies can be critical for maintaining a temporally stable population through risk partitioning (i.e., the portfolio effect [8,10]). While biocomplexity is frequently viewed as a horizontal measure within a given level of biodiversity, it cannot always be neatly confined to a single hierarchical level. For example, while intra-specific diversity at a population-level is comprised of phenotypic diversity, there is a genetic component dictating some phenotypic differences. Similarly, species interactions not only shape an ecosystem but simultaneously act as evolutionary pressures determining the fitness of individual phenotypes. Therefore, because selection acts on standing variation, forward-looking conservation strategies must stress the importance of maintaining adequate levels of genetic diversity within a population to maintain adaptive potential in changing environmental conditions [6,9,11,12].

Chinook salmon in the Central Valley are emerging as a case study to understanding the importance of biocomplexity for the persistence of a population complex facing multiple external threats, including habitat fragmentation, overexploitation, and climate change. Here, life history diversity resulting from

different migration phenotypes creates a “portfolio of stocks” buffering against spatiotemporally variable environmental conditions and anthropogenic impacts, which in turn increases resilience and promotes interannual stability [13,14]. Chinook are anadromous with a distinct lifestyle spanning both freshwater and marine ecosystems. Eggs are laid and hatched in the tributaries, where juveniles rear for a period of time before migrating out to the ocean. There, they spend one 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 and supporting recreational, commercial, and heritage fisheries [15]. The tributaries of the Sacramento-San Joaquin River system 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 [16]. Early-migrating runs (Winter, Spring) make the trade-off 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 while relying on fat stores, spawn, and die [17]. By contrast, late-migrating individuals (Fall, Late-Fall) remain in ocean until relatively mature before making their spawning run.

This asynchronicity in run timing of adults stabilizes the population complex overall, as temporal and spatial variation in challenges to population components will vary. 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 buffering the population complex overall. However, this buffering ability is threatened when one run-type is consistently negatively impacted across tributaries, specifically, dams and other anthropogenic factors have disproportionately affected historical early migrator habitat in much of the Central Valley. As a result both Spring and Winter run are listed under the Endangered Species Act [18] and the dramatic declines in their abundance and distribution has resulted in a loss of biocomplexity across run types, thus lowering the portfolio effect and making the population complex as a whole more vulnerable [13].

These demographic changes are also likely to result in an erosion of genetic diversity (and consequently adaptive potential) due to low effective sizes, which may otherwise have proved important for persistence under future environmental conditions. Measuring the genetic diversity of a population has long been used as a proxy to quantify the “future adaptive potential” of populations [19,20]. However, despite a plethora of tools to quantify genetic diversity, consensus on which part of the genetic diversity to focus on for conservation remains elusive. By focusing on the structural biodiversity at a genomic level, tools such as diagnostic SNP panels for Chinook can be used as intrinsic markers to identify (hierarchical sets of) demographic groups, quantify biocomplexity across a species’ range, and describe changes in the portfolio of stocks [21]. Recent technological advances now also allow for the direct

interrogation of functional biodiversity at the genomic level. For example, recent studies have identified a single chromosomal region (GREB1L to ROCK1) underlying adult migration timing in Chinook and other salmonids[22–25]. In addition, the increasing availability of genomic data sets holds promise for identifying genes associated with polygenic traits [26,27]. Finally, focusing on the compositional biodiversity at a genomic level will lead to an increased understanding of the differences in genetic diversity and unique components among populations within a complex.

The increased availability of data quantifying functional, compositional, and structural genomic diversity of populations poses questions about what is important to consider for conservation. There is a need for balancing goals focused on preserving important functional diversity and key life history traits with goals aimed at preserving adaptive potential that may be more cryptic, especially when putatively neutral and adaptive markers display divergent signals [28]. This tradeoff is highlighted recent studies that uncovered the genomic basis of run timing in Chinook salmon and have focused on the need to preserve the genomic basis for run timing rather than the genetic diversity contained throughout the genome (Thompson et al. 2020). However, for conservation, a singular focus on run timing risks ignoring important biocomplexity at a genomic level underlying other traits that comprise additional diversity among individuals within the same migration phenotype. Given the myriad selective pressures facing Chinook throughout their life history (e.g. juvenile outmigration, ocean survival, and feeding) and the large degree of environmental heterogeneity experienced by different populations, it is unlikely that the chromosomal region associated with run timing is the only component of unique genetic diversity characterizing individuals across populations. Rather, it is more probable that each component of a population complex has been shaped by a unique combination of evolutionary forces over time, and as a result each harbors unique components of genetic diversity that may prove critical for conservation management. Here, we aim to understand the distribution and abundance of genomic diversity, beyond the GREB1L locus, contained within and among the four Chinook salmon run types, using the populations of the Central Valley of California as a study system. We genotyped each individual at > 10,000 SNP-containing loci throughout the coding and non-coding parts of the genome to assess patterns of (1) population structure and differentiation, and (2) heterozygosity and allelic diversity, and (3) distribution of unique diversity within and among run types in tributaries of the Central Valley.

Methods

Genotyping

The data set presented in Meek et al. 2019 was re-analyzed to capitalize on newer genomic methods and recently developed resources to better understand the patterns of genomic diversity present among Chinook salmon from all four run types present in the Central Valley. In the previous assessment, loci were assembled *de novo* and only Restriction-site Associated DNA (RAD)-tags that contained a single

SNP were used for the analysis, which focused the analyses on broad levels of differentiation among run types. Here, reads were aligned to a species-specific assembled genome and SNPs on the same RAD-tag were collapsed into haplotypes, resulting in multi-allelic loci and allowing for a more fine-scale assessment of patterns of genomic diversity.

As reported in Meek et al. 2019, fin clips from all four run types (Fall, Late-Fall, Winter, Spring,

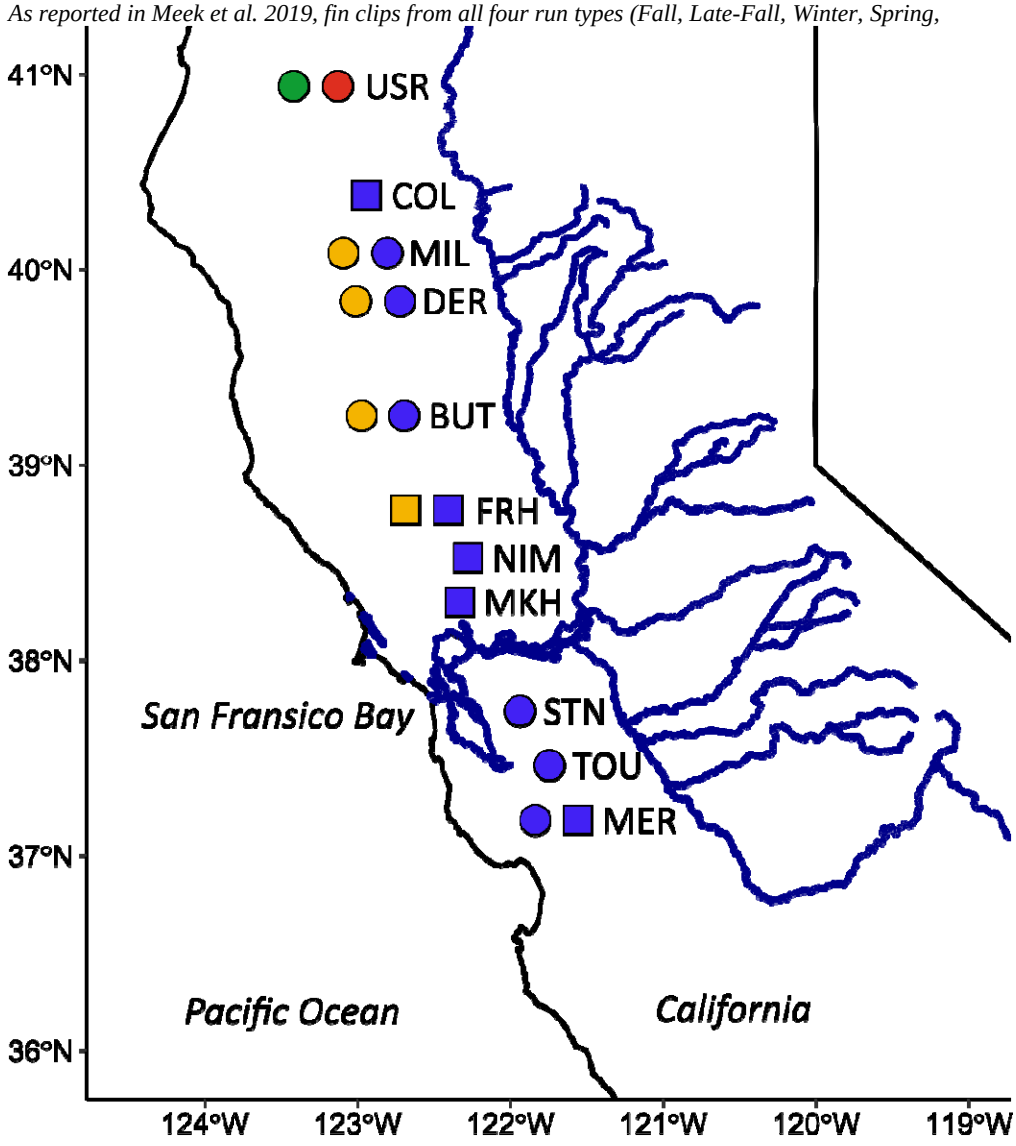

Figure 1) were obtained from adult Chinook Salmon from each major tributary in the Central Valley during their spawning migrations. Genomic DNA was extracted and digested using *SbfI* to construct RAD libraries following Miller et al. [30]. Fifteen libraries consisting of 30 – 47 individuals each were sequenced on single Illumina HiSeq 2000 lanes (100bp, single end).

For the new analyses presented here, raw sequences were demultiplexed using *process_radtags* [31] and quality trimmed using *fastp* [32]. Quality trimmed reads were mapped to a Chinook reference genome [33] using *BWA-mem* [34]. Reads with a mapping quality < 5 were removed and filtered bam files were

concatenated into a single bam file to query coverage per interval using *bedtools* [35]. Mapping intervals > 25bp and < 500bp with coverage > 50 were extracted to create a reduced-representation reference consisting of the recovered RAD-tags, reads were re-mapped to the reduced-representation reference and SNPs were called using *freebayes;*the maximum allowed gap (-E) was set to three and minimum mapping and base quality set to five, otherwise default parameters were used [36].

The raw data set was rigorously filtered to remove low quality genotypes, loci, and samples following principles set forth in O'Leary et al. (2018). In short, genotypes with < 5 reads or quality < 20 were coded as missing, retained loci had a mean minimum depth of 15, and were called in at least 50% of individuals of a given library, 85% of individuals of a run type, or 90% of individuals overall. Loci were further filtered for allele balance, mapping quality ratio of reference vs alternate allele, depth/quality ratio, and excess heterozygosity to remove paralogs and other technical artifacts. Finally, *rad_haplotyper* [38] was used to collapse SNPs on the same RAD-tag into haplotypes. Detailed processing steps and threshold values are contained in Supplementary Material 1.

Assessment of population structure and differentiation

Population structure was explored using two methods, a clustering analysis based on genetic similarity and an assessment of population differentiation among individuals grouped *a priori* based on run type and tributary. In the first method, individuals were clustered into K = 1 – 10 groups using k-means clustering based on the PCA-transformed genotype matrix (i.e. no assumptions regarding Hardy-Weinberg or linkage disequilibrium) followed by a discriminant analysis of principle components (DAPC) to determine the membership probabilities of each sample to each inferred cluster as implemented in *adegenet* [39]. During DAPC, the genotype matrix is first PCA-transformed, then samples are partitioned into a within- and between-group component to maximize discrimination between groups using linear discriminant analysis, and finally, the membership probability of each sample to each cluster is calculated. To ensure sufficient variance was retained to discriminate among groups but not overfit the data, the optimum number of principle components to retain was determined using a stratified cross-validation of DAPC.

In the second method, Weir & Cockerham's unbiased estimator of F_{ST} [40] was used to calculate population differentiation among individuals grouped *a priori* by run type within tributary. To test for genetic heterogeneity in the data set, global F_{ST} was calculated across all groups, then, pairwise F_{ST} was calculated as a *post hoc* test for pairwise differences among groups. Significance was determined using 95% confidence intervals around each estimate generated by resampling loci 1,000 times using *assigner* [41].

Estimates of effective population size

Estimates of effective population size *N*e were determined using *LinkNe* [42] an extension of the linkage disequilibrium method (LD) designed for data sets with loci of known linkage relationships, for example because they are on the same chromosome. Loci with elevated LD need to be accounted for to not downwardly bias *N*e estimates. To estimate *N*e, SNP data sets were generated for each tributary/run group. Only SNPs placed on a chromosome were retained. Additionally, SNPs with a minor allele frequency < 0.01 were removed and only the first SNP on each RAD-tag retained. Genomic distances (bp) were converted to recombination rates (cM) using the size of the female Chinook linkage map (3,118 cM, [43] =) and the length of the genome (2.4 Gbp, Christensen et al. 2018)) resulting in an estimate of approx. 770kb being equivalent to one cM.

Assessment of Genomic Diversity

All measures of genomic diversity were made for individuals grouped by run type within each tributary, with wild and hatchery individuals treated as separate groups. Samples sizes among groups are comparable, though only seven Spring Feather River individuals are present in the filtered data set [29]. Four types of parameters were assessed, (1) measures of heterozygosity, (2) measures of allelic diversity, (3) sequence-based parameters, and (4) measures of unique variation.

For the first three sets of parameter types, significant heterogeneity was determined using a Friedman's rank sum test followed by a *post hoc* Wilcoxon signed-rang test to test for significant pairwise differences between groups; *p*-values were corrected for multiples comparisons assuming an FDR of 0.05 [44]. The observed heterozygosity (Ho) was measured as the proportion of heterozygote genotypes per locus [45] and the expected heterozygosity (gene diversity) (Hs) as the proportion of heterozygous genotypes expected under Hardy-Weinberg Equilibrium [45]. To account for differences in sample size, allelic richness was measured as rarefied allele counts. The evenness of allelic diversity at a given locus was calculated as the ratio of the number of abundant to the number of more rare genotypes using the ration of the Stoddart & Taylor index (diversity index weighted for more abundant alleles) and Shannon-Wiener index (diversity weighted for more rare alleles) as implemented in *poppr* [46]*;* lower values indicate prevalence of more rare alleles and uneven distributions of allele frequencies. The nucleotide diversity statistic π [45] was calculated as the sum of the number of pairwise differences between haplotypes of a given nucleotide over the number of comparisons made; this parameter is biased toward alleles segregating at intermediate rates and will underestimate genetic diversity when many rare alleles are present.

Finally, patterns of unique diversity were assessed by comparing (1) the number of fixed loci, i.e. all individuals in a given group are homozygous for the same allele along with the global allele counts for fixed loci across all individuals, (2) the number of loci with singletons 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 where more than one allele is found only in a single group (all other groups are fixed for a single allele). By contrast, private alleles are alleles found in only a single group, though other groups exhibit more than one allele at that locus. To compare whether identified loci are randomly distributed across chromosomes, null distributions were generated by shuffling chromosome designations across loci 1,000 times to determine whether the observed values fall outside the null distribution.

Results

Genotyping

The final filtered data set consisted of 386 individuals genotyped for 12,983 SNP-containing loci (hereafter loci) with a total of 30,037 alleles.

Assessment of population structure and differentiation

Minimum AIC was observed for K = 4 (Figure S1); cross-validation indicated that the optimum number of principle components to retain was 20 – 150 (retaining 11.6 – 75.1% of variance; Table S1). The mean optimum success of assignment declined from 100% (K = 2 – 3) to 96% (K = 7) and then dropped to 89% for K = 8 (Table S1). Figure 1 summarizes membership plots for K = 2 – 8. For K = 2 Winter run individuals form a single cluster set against individuals from all other run types and tributaries. Similarly, for K = 3, Winter run from a single cluster, a second cluster consists of Spring run individuals from Butte, Deer, and Mill Creek, while Spring Feather River Hatchery and Late-Fall individuals form the final cluster along with all Fall groups. Overall, Late-Fall individuals do not emerge as their own distinct cluster until K = 6, while Spring Feather River individuals continue to cluster with Fall individuals from Coleman and Feather River Hatchery for K = 7 – 8. Overall, Spring run individuals form at most three clusters, with Deer and Mill Creek individuals always clustering together. In general, Fall run individuals are assigned to two or three clusters, here, Coleman Hatchery individuals generally form one cluster along with most Feather River Hatchery individuals (both Fall and Spring) and some Deer and Mill Creek individuals; increasing K to 8 results in Coleman Hatchery individuals starting to form a more distinct cluster of their own.

For individuals grouped by run type within tributaries, global F_{ST} = 0.0319 [0.0309 – 0.0328] indicates significant heterogeneity among groups. Pairwise comparisons of Winter individuals from the Upper Sacramento River and all other run/tributary groups result in the highest observed pairwise F_{ST}-values (0.14 - 0.161). By contrast, all non-significant comparisons, with CIs including zero, were comparisons among Fall run tributary populations. Notably, Fall Coleman, Feather River Hatchery, and Stanislaus River individuals are the only Fall tributaries significantly different from all other Fall tributaries (Table S2, Figure S2). Wild and Hatchery individuals from the Merced River are not

significantly different from each other and overall, there is no consistent pattern of Fall Hatchery individuals from different tributaries consistently having more or less observed significant comparisons to other Fall tributaries (Table S2, Figure S2).

Late-Fall Upper Sacramento River individuals are significantly different from all other run/tributary groups. Pairwise F_{ST}-values for comparisons with Fall tributaries were the lowest, ranging from 0.001 - 0.01. For Late-Fall and Winter individuals pairwise F_{ST} is higher at 0.016, despite inhabiting the same tributary. Finally, Late-Fall individuals are most distinct from Spring individuals (F_{ST} = 0.029 – 0.044 for wild individuals, 0.016 for Feather River Hatchery individuals; Table S2, Figure S2).

Spring Feather River Hatchery individuals stand out as having lower pairwise F_{ST}-values in comparison to Fall (0.003 - 0.013) and Late-Fall (0.016) tributaries than to other Spring tributaries (0.018 - 0.036). Pairwise F_{ST}-values between Spring Butte Creek individuals and Fall tributaries range from 0.034 - 0.044 and are lower for Deer and Mill Creek individuals (0.02 - 0.029, for both tributaries). The pairwise comparison for Deer and Mill Creek (F_{ST} = 0.004) is lowest of all within Spring comparisons.

Notably, for all Spring/Fall tributary comparisons the lowest observed value is for the comparison to Fall Feather River Hatchery individuals (0.003 - 0.034). In general, comparisons of Fall and Spring individuals from the same tributary are intermediate to the range of observed pairwise F_{ST} values for Fall/Spring tributaries, overall. By contrast, Fall and Spring Feather River Hatchery have an observed pairwise F_{ST}-value of 0.003 which is the lowest value in the range of comparisons to other Fall/Spring comparisons (Table S2, Figure S2).

*Estimates of effective population size N*e

For Fall individuals, *N*e estimates ranged from 1,125 (CI = 940 – 1,399; Merced River Hatchery) to 7,432 (CI = 5,147 – 13,351; Nimbus Hatchery) with the exception of Coleman Hatchery where the point estimate was lower (*N*e = 615, CI = 594 – 637) and Merced River where it was higher (11,436 CI = 6,768; 36,804; Table 1). Estimates for spring groups were lower, ranging from 85 (CI = 84 – 86; Mill Creek) to 561 (CI = 529 – 597; Butte Creek), while *N*e = 3,114 (CI = 2,398 – 4,435) and 174 (CI = 171 – 177) for Late-Fall and Winter individuals, respectively (Table 1). For Fall Deer and Mill Creek and Spring Feather River Hatchery estimates were negative. This occurs when the LD attributed to sampling variation is larger than the LD attributed to drift. Theoretically, this should be interpreted as an “infinite” population size. However, because LD attributed to sampling variation is a function of sample size, negative estimates can be obtained for small sample sizes even when the real *N*e is small. Here, the groups with negative estimates do in fact fall among the lowest sample sizes present in the data set (Table 1).

Assessment of Genomic Diversity

MEASURES OF HETEROZYGOSITY

The mean observed heterozygosity is lowest for Winter run individuals from the Upper Sacramento River (mean = 0.1272), followed by Spring individuals from Butte Creek (mean = 0.1587) and highest for Fall individuals from Coleman and Feather River Hatcheries (0.1713 and 0.1741, respectively, Table S3, Figure 3A). Similarly, mean expected heterozygosity is lowest for Upper Sacramento River winter run individuals (mean = 0.1285), and highest for Fall Feather River Hatchery and Mill Creek individuals (mean = 0.1718 and 0.1714, respectively, Table S4, Figure 3B). For both observed and expected heterozygosity, Spring tributaries exhibit a wider range of distributions compared to Fall tributaries, despite the smaller number of Spring tributaries in the data set (Figure 3A, B). Late-Fall values fall into the range of distributions observed among Fall tributaries (Figure 3A, B).

ALLELIC RICHNESS

The mean values of rarefied allele counts are comparable across tributary/run groups, ranging from 1.51 - 1.52 alleles per locus for all run/tributary groups except Late-Fall Upper Sacramento River (1.36) and Spring Deer Creek individuals (1.48), which exhibit significantly lower mean values (Table S5, Table S6). Despite similar mean values, most pairwise comparisons are significant (Table S5), indicating that even though there is a relatively consistent global number of alleles per locus (most loci have < 3 alleles) the patterns of which loci are variable are consistently different across run/tributary groups. For example, Spring Butte Creek individuals exhibit a pattern of rarefied allele counts significantly different from all other locations, and, despite a mean value comparable to most other groups, also exhibit the highest median value (1.71; Figure 3C). Overall, Spring tributaries exhibit more variation among tributaries compared to Fall tributaries (median = 1.38 - 1.71; all pairwise comparisons are significant, Table S5, S6). Distributions are more similar to each other across Fall tributaries, with Butte Creek individuals exhibiting the lowest (1.38) and Feather River Hatchery individuals the highest (1.48) median values (Figure 3C). Notably, individuals from different run types from the same tributary may exhibit quite different patterns. For example, Fall individuals from Mill Creek and the Feather River Hatchery exhibit higher allele counts compared to Spring individuals from the same tributary. By contrast, Fall individuals from Deer and Butte Creek have lower allele counts compared to their Spring counter parts. The strongest contrast is Butte Creek, where Spring individuals exhibit the highest allele counts overall while Fall individuals exhibit the lowest allelic richness among all Fall tributaries (Figure 3C).

EVENNESS OF ALLELIC DIVERSITY

Overall, Winter individuals from Upper Sacramento River exhibit significantly lower evenness of allelic richness across loci, indicating that many loci are characterized by rare alleles (Figure 3D). By contrast, Fall Feather River individuals exhibit the highest median evenness (0.78) compared to all other

run/tributary groups (though not all comparisons are significant; Table S7). Within Fall tributaries, mean levels of evenness are all approx. 0.76 though there are some significant differences in their overall distributions (Figure 3D). Again, Spring tributaries exhibit a wider range of mean and median values for evenness compared to Fall (Figure 3D).

NUCLEOTIDE DIVERSITY

Winter Upper Sacramento River and Spring Butte Creek individuals exhibit lower nucleotide diversity compared to all other locations (0.00079 and 0.001012, respectively; Table S8, Figure 3E). By contrast, Fall Feather River Hatchery exhibit the highest mean nucleotide diversity (0.001058); similarly, among Spring tributaries Feather River Hatchery individuals exhibit the highest nucleotide diversity (0.001048; Table S8, Figure 3E).

FIXED LOCI

Late-Fall Upper Sacramento individuals and Spring Butte Creek individuals exhibit the highest number of fixed loci (6,416 and 6,210, respectively); for all other run/tributary groups 3,500 - 5,000 fixed loci were identified (Figure 4A). The number of fixed loci in Spring tributaries is generally larger compared to Fall tributaries with the exception of Spring Feather River. By far the two largest intersects are loci fixed exclusively in a single location, Late-Fall Upper Sacramento individuals (460) and Spring Butte Creek individuals (317). All other intersects are < 115 loci. Apart from Spring Feather River hatchery individuals (27), Spring tributaries have more loci fixed in a given tributary (74 - 317) compared to Fall tributaries where 10 - 61 loci are fixed among individuals from a single tributary. Notably, among intersects of loci fixed in two locations, the three largest intersects are all a combination of Late-Fall Upper Sacramento River and a wild spring population (42 – 115 loci). Overall, about a third of intersects of two run/tributary combinations are loci fixed among Late-Fall Upper Sacramento River individuals and a second location. There is no observed pattern of loci more likely to be fixed among tributaries in geographic proximity (Figure 4A).

Loci fixed in Late-Fall Upper Sacramento and Spring Butte Creek individuals also exhibit the highest global allele diversity (mean = 1.46 and 1.41, respectively; Table S9), i.e. loci that are fixed in these groups are more variable when alleles are tabulated across individuals from all runs/tributaries (Figure 4B). By contrast, the global diversity of fixed alleles is lowest in Fall tributary groups, and overall levels are more similar across tributaries (median = 1.13-1.14 with the exception of Fall Deer Creek and Fall Butte Creek at 1.19) compared to spring tributaries where those distributions of global diversity are higher and more variable (median = 1.28 - 1.41; Table S9). The distribution of global diversity among Spring and Fall tributaries varies with some run/tributaries exhibiting much tighter ranges than others (Figure 4B).

In general, the proportion of loci that are fixed for a run/tributary group is consistent across chromosomes and there is no distinct, non-random pattern. However, for some run/tributary groups, one or two chromosomes do exhibit patterns outside the simulated null distribution (Table S10). Chromosome 9 has a significantly higher than expected number of fixed loci compared the random null distribution for five Fall tributaries while chromosome 19 has a significantly lower than expected number of fixed loci for one Fall and one Late-Fall location, and chromosome 33 has a significantly lower number for two Fall and the Winter run location. None of the Spring run tributary groups exhibit chromosomes with significantly higher or lower than expected number of fixed loci. By contrast, the Late-Fall and Winter run individuals from the Upper Sacramento River have two and one chromosomes, respectively, with lower than expected numbers of fixed loci (Table S11). Notably, all the Fall individuals from hatcheries in the Sacramento River exhibit chromosomes with observed numbers of fixed loci outside the simulated distribution (Coleman hatchery (2 chromosomes), Nimbus hatchery (1 chromosome), and Feather River hatchery (1 chromosome)); in addition Mill, Deer Creek and Tuolumne River wild individuals exhibit one chromosome with a significant number of fixed loci that shows a similar pattern for at least one other set of fall individuals Table S11).

SINGLETONS

Three hundred forty-seven (1.2% of the total) loci exhibit at least one singleton. A comparison of individuals grouped by run/tributary demonstrates that Spring Butte Creek and Deer Creek individuals (mean = 4.0 and 3.0, respectively) and Late-Fall Upper Sacramento River individuals (mean = 2.23), exhibit the highest mean number of singletons per individual. In general, Fall tributaries exhibit a lower mean number of singletons per individual compared to other run/tributary groups, indicating that they are comparatively less characterized by rare alleles (Figure 4C; Table S12).

PRIVATE POLYMORPHISMS

Winter run individuals from Upper Sacramento River have the highest number of loci (N = 153) with private polymorphisms; all other groups have < 100 loci with private polymorphisms (Figure 5). Spring individuals from Butte Creek exhibit the lowest number of loci exclusively polymorphic among individuals of a group (N = 18). Across Late-Fall individuals, 70 loci with private polymorphisms were identified. This number is higher than observed for eight of eleven Fall groups. Notably, individuals from hatcheries fall along the entire range of private polymorphisms; Fall Nimbus Hatchery individuals are on the low end (N = 33) and Spring Feather River Hatchery individuals on the high end (N = 84). Comparing the chromosomal positions of private polymorphisms indicates that Winter Upper Sacramento has 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 Fall Mill Creek individuals, Spring

Deer Creek individuals, and Merced River Hatchery. The mean is lowest for Spring Butte Creek individuals (Figure 5, Table S13).

Chromosomes 16, 30, and 31 are the only chromosomes with significantly higher than expected numbers of loci with private polymorphisms (for one location each; Table S14); interestingly each of these three chromosomes also has a significantly lower than expected number in different run/tributary groups. Overall, chromosome 21 has significantly less than expected private polymorphisms for ten of 17 run/tributary groups. Chromosome 27 and 22 similarly stand out for exhibiting significantly lower than expected exclusive polymorphisms in nine and eight groups, respectively. Fall Tuolumne River and Nimbus River hatchery individuals along with Late-Fall Upper Sacramento River are the only groups which have a single chromosome exhibiting significantly less than the expected numbers of loci with exclusive polymorphisms (Table S15). By contrast, of the six run/tributary groups with the highest number of chromosomes exhibiting less than the expected number of loci with private polymorphisms, four are Fall run hatcheries, though Nimbus Hatchery individuals group with spring run groups which all sit on the lower range of the number of flagged chromosomes. The top nine are all Fall run groups (Table S15).

PRIVATE ALLELES

While most run/tributary groups exhibit 80 - 180 private alleles, Spring Deer Creek individuals exhibit the highest number of private alleles (267) while Fall Butte Creek and Spring Feather River Hatchery individuals exhibit the lowest number of private alleles (34; Figure 5). There is no distinct pattern of hatchery individuals exhibiting more/less private alleles compared to wild individuals, or different run types having consistently higher/lower number of private alleles compared to others; though all groups do exhibit private alleles (Figure 5).

Most commonly, private alleles are only carried in 1 - 3 individuals, i.e. they are singletons or very rare alleles. Winter Upper Sacramento individuals exhibit the highest mean number of individuals carrying private alleles (N = 4.07) compared to other run/tributary groups, followed by Spring Butte Creek individuals (N = 1.34), for all other groups private alleles are found in a mean of 1.01 - 1.17 individuals. Similarly, these two groups have the most “common” private alleles, carried in 25 and ten individuals, respectively (Table S11).

Comparing the chromosomal positions of loci with private alleles in a single run/tributary group indicates that Spring Deer Creek has the highest mean proportion of loci with private alleles on a chromosome (Table S12). Fall Coleman Hatchery, Butte and Deer Creek individuals are the only groups for which there is a single chromosome with a larger than expected number of loci with private alleles (Table S13). Fall Deer Creek individuals stand out as having 11 chromosomes with less than the expected number of loci with private alleles. Similarly, for Spring Feather River Hatchery individuals nine

chromosomes exhibit less than the expected number if these loci were randomly distributed among chromosomes. A comparison of the proportion of loci with private alleles identified for each run/tributary group across loci demonstrates differences among chromosomes with a much wider range of means compared to comparing by run/tributaries, i.e. there are consistent patterns of some chromosomes exhibiting a proportion of private alleles across groups compared to others (Table S14).

Discussion

The patterns of genetic diversity observed in populations are the product of past evolutionary and ecological processes. Intra-specific diversity determines the standing variation upon which evolutionary forces act, and levels of phenotypic and genetic variation determine the ability of a population to respond to changes in environmental conditions, the resilience and persistence of a species, and the stability of the ecosystems they inhabit [47,48]. Therefore, assessing the biocomplexity of a population complex can help predict the range of possible responses to changing conditions. Overlooking the importance of biocomplexity at a genetic and phenotypic level and failing to preserve the adaptive potential of populations could have irreversible consequences for the health and sustainability of populations and the goods and services they provide. Here, we presented a fine-scale assessment of the biocomplexity contained at the genomic level in California's Central Valley Chinook salmon population complex. We found significant differentiation among and within run types and the tributaries they inhabit, effective population sizes severely below suggested critical thresholds to avoid erosion of genetic diversity and loss in fitness, and corresponding low levels of genetic diversity in populations that have experienced recent demographic declines. Finally, despite apparent gene flow among individuals of the same run type across tributaries, each run/tributary group was characterized by a component of unique diversity. 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.

Populations become increasingly vulnerable to environmental, demographic, and genetic stochastic effects as their size decreases. Thus, monitoring Ne of individual components of a population complex is critical for identifying differences in the vulnerability of different groups to genetic issues, evaluating population viability, and predicting evolutionary trajectories to guide conservation managers in their decision-making. Here, individuals from the early migrating run types (Winter, Spring) for each tributary exhibit the lowest Ne values, below recommended targets of Ne = 1,000 to maintain adaptive potential [49]. Further, our results corroborate previous work showing the endemic Winter run stands out as genetically the most distinct from all three other run types. Additionally, we record the lowest estimate of

effective population size for the Winter run (Ne = 174) to date. These results highlight the extreme and alarming risk of extinction facing Winter run [29].

Similarly, we show dangerously low effective population sizes within Spring run populations, with Mill and Deer Creek values being similar to those of Winter run (Ne = 85 and 188, respectively) and Butte Creek being just above 500. There is fine-scale structuring within Spring run. Among wild populations Butte Creek is significantly distinct from Mill and Deer Creeks, which are geographically closer and environmentally more similar to each other. Further, all three wild Spring tributaries are significantly distinct from Feather River Hatchery Spring run. Notably, Spring Feather River Hatchery individuals group more closely with Fall individuals from Coleman and Feather River Hatcheries, pointing towards introgression between runs due to hatchery practices [50].

Contrary to the current practice of managing Late-Fall and Fall runs as a single ESU, our results add to the increasing evidence that Late-Fall is genetically distinct from Fall run [29]. While there is evidence of some distinctions between Fall individuals from more northern tributaries, completely delineating wild Fall run individuals from different tributaries is not possible with this data set, indicating a loss of biocomplexity with increasing homogenization of Fall run individuals [51]. In contrast to the early migrating runs (Spring and Winter), *N*e is robust in the late-migrating Fall and Late-Fall groups and at levels sufficient to maintain adaptive potential [52]. The exception to this is Coleman Hatchery Fall individuals, which had an *N*e an order of magnitude lower than other Fall run populations. Also, of note, the *N*e from the Merced River Hatchery Fall run population was also orders of magnitude lower than the *N*e for the putatively wild fish caught in the Merced River. Both of these populations are concerning from a conservation perspective, as maintaining sufficiently large *N*e is a major concern for stocking programs, especially when introgression with wild populations is likely [53].

The erosion of genetic diversity due to both genetic drift and inbreeding are inversely proportional to *N*e. Both processes lead to a decreased level of heterozygosity as the rate of alleles being lost due to drift increases and individuals become increasingly likely to mate with individuals with similar genotypes. Accordingly, if drift is the primary force shaping the genetic diversity within the declining early migrating populations, we would expect to see low levels of heterozygosity and comparatively higher numbers of fixed alleles. Indeed, Upper Sacramento Winter run individuals exhibit the lowest levels of heterozygosity and other measures of genetic diversity compared to all other groups, and the highest number of fixed loci is found in the Spring run groups. Interestingly, among Spring run groups we find much wider distributions across all measures of diversity – this underscores the stochasticity of genetic drift. Not only is the effect stronger in smaller populations (resulting in an accelerated loss of diversity), but how each component is affected can differ and groups diverge from each other by chance alone. In addition, each tributary experiences a different selection regime, driven by environmental differences, resulting in an

increased impact of decoupled demographic and environmental stochastic events affecting each population.

The assessment of private polymorphisms (loci only variable in a single group) and private alleles (alleles only found in a single group) reveals that within each tributary, each run exhibits unique components of genetic diversity. Notably, despite Winter run individuals from the Upper Sacramento River having the lowest level of diversity when comparing measures related to heterozygosity and allelic diversity, they exhibit the highest level of private polymorphisms, i.e. they have less diversity but the diversity present is unique compared to all the other runs. This is different from Butte Creek Spring individuals which exhibit low levels of heterozygosity and allelic diversity similar to Winter run individuals, but also have a low number of private polymorphisms, indicating that there are differences in the demographic and evolutionary forces that have shaped genetic diversity in these groups. As a result, despite Spring and Winter run types sharing a similar early migration phenotype and an underlying genotype related to the early migration phenotype (demonstrated for this data set in Meek et al. 2019), their intra-specific diversity is unique and suggests differences in standing variation for selective pressures to act upon. Similarly, despite Fall and Late-Fall individuals being managed as a single ESU and sharing similar late migrating phenotypes/genotype, we see important distinctions in the unique diversity harbored by Late-Fall individuals from the Upper Sacramento compared to Fall/tributary groups. Late-Fall exhibits similar levels of heterozygosity, but lower allele counts compared to Fall population components. But again, the number of observed private polymorphisms is higher for the Late-Fall run compared to eight of the eleven Fall run groups, indicating differences in past processes shaping these run types and their evolutionary trajectories. While all groups exhibit 80 – 180 private alleles, it is groups other than Fall run that tend to be characterized by rare alleles. Since rare alleles are expected to be lost first during bottlenecks, the expectation would be that Fall run groups (which have relatively large *Ne*) would instead have a larger number of rare alleles than the groups with smaller *Ne* if genetic drift were the main process affecting allele frequencies. Other processes that could produce these patterns include increased homogenization resulting in an exchange of alleles among tributaries, historical hatchery practices, and/or selection removing rare alleles. By contrast, spatially and temporally heterogeneous environments promote and maintain polymorphisms and high levels of standing genetic variation [54–56]. Indeed, Late-Fall and Spring Deer Creek individuals exhibit the highest numbers of singletons, and Winter Upper Sacramento and Spring Butte Creek individuals exhibit the highest mean number of individuals carrying alleles found only in that population. Even Spring Feather River Hatchery individuals, which have introgressed with Fall individuals in the past, carry private polymorphisms. Overall, these patterns support the conclusion that processes such as balancing selection are occurring to maintain diversity, despite declining population sizes.

This 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 and/or adaptive potential [19], the variation of genotypes within and among individuals has been demonstrated to be a suitable proxy to predict fitness of individuals and the ability of populations and ecosystems to respond to changes in environmental conditions [20,57–59]. Examples from translocation and genetic rescue efforts have demonstrated that heterozygosity and genetic diversity can be more efficient predictors of success over the ability to match genotypes as closely as possible to individuals already present in the population [60,61]. As differences in early and late run timings have been shown to be controlled in part by a single locus of large effect [22–24,62] it may be tempting to limit the focus of conservation efforts to ensuring the maintenance of both run timing alleles while ignoring the protection of overall genomic diversity. While the locus that affects run timing is undoubtedly important and of critical conservation concern [63], there would be drawbacks to emphasizing only functional diversity that has been identified as associated with relatively well-characterized adaptations at the cost of losing cryptic diversity harbored across the genome that may prove to already be vital for persistence or become critical for survival in the future [52].

Focusing management solely on preserving migration phenotypes may ensure that the genetic components controlling adult run timing is maintained, however it runs the risk of losing unique components of genetic diversity harbored by each population within migration phenotypes. This loss of diversity, and increasing genetic homogenization may be a more important factor driving the loss of the portfolio than demographic synchronization itself [9,64]. Because of their complicated life history, environmental pressures differ widely across 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 best matching conditions during a different life stage. Additionally, climate change will impact environmental conditions in individual tributaries differently, again necessitating genomic diversity across the Central Valley to allow adaptation to changing conditions [65]. 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 phenotypes are more likely to adapt to new conditions and therefore increase population viability with rapidly changing and fluctuating environmental conditions [66]. Provided sufficient standing genetic diversity exists and is protected in the Central Valley, the intraspecific variation present in Central Valley Chinook may allow better adaption to changing conditions [4,67,68]. Even though anthropogenic impacts 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 [9,69] despite serving as the fundamental building block of biodiversity. Indeed, finding unique diversity within each run/tributary group comprising 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 run at the level of individual tributaries) to inform conservation and management policies that counteract genetic homogenization and conserve biocomplexity at a genetic level. Additionally, our results highlight the necessity of managing population complexes as a whole with a focus on maintaining biocomplexity as an important factor determining the resilience and resistance to changing environmental pressures.

Data Accessibility

Data for this study are available from https://doi.org/10.5061/ dryad.tht76hdvt.

Supplementary Material 1 and 2 contain fully reproducible code supporting the analysis.

Author contributions

SJO designed and carried out data analysis and the drafting of the initial manuscript. TQT provided feedback on data analysis and critically revised the manuscript. MHM assisted in designing the data analysis, and critically revised the manuscript. All authors gave final approval and take full responsibility for the manuscript.

Competing Interests

We declare we have no competing interests.

Funding

This work was supported by a grant from the Delta Stewardship Council (#18209) to MHM.

Acknowledgements

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

References

1. Petchey OL, Gaston KJ. 2002 Functional diversity (FD), species richness and community composition. *Ecol. Lett.* **5**, 402–411. (doi:10.1046/j.1461-0248.2002.00339.x)
2. Péru N, Dolédec S. 2010 From compositional to functional biodiversity metrics in bioassessment: A case study using stream macroinvertebrate communities. *Ecol. Indic.* **10**, 1025–1036. (doi:10.1016/j.ecolind.2010.02.011)
3. Duelli P, Obrist MK. 2003 Biodiversity indicators: The choice of values and measures. *Agric.*

Ecosyst. Environ. **98**, 87–98. (doi:10.1016/S0167-8809(03)00072-0)

4. Messer PW, Ellner SP, Hairston NG. 2016 Can Population Genetics Adapt to Rapid Evolution? *Trends Genet.* **32**, 408–418. (doi:10.1016/j.tig.2016.04.005)
5. CBD. 2011 Strategic plan for biodiversity 2011–2020, including Aichi Biodiversity Targets.Nagoya, Japan: The Convention on Biological Diversity.
6. Mable BK. 2019 Conservation of adaptive potential and functional diversity: integrating old and new approaches. *Conserv. Genet.* **20**, 89–100. (doi:10.1007/s10592-018-1129-9)
7. Hoelzel AR, Bruford MW, Fleischer RC. 2019 Conservation of adaptive potential and functional diversity. *Conserv. Genet.* **20**, 1–5. (doi:10.1007/s10592-019-01151-x)
8. Hilborn R, Quinn TP, Schindler DE, Rogers DE. 2003 Biocomplexity and fisheries sustainability. *Proc. Natl. Acad. Sci.* **100**, 6564–6568. (doi:10.1073/pnas.1037274100)
9. Des Roches S, Pendleton LH, Shapiro B, Palkovacs EP. 2021 Conserving intraspecific variation for nature's contributions to people. *Nat. Ecol. Evol.* , 1–9. (doi:10.1038/s41559-021-01403-5)
10. 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* **465**, 609–612. (doi:10.1038/nature09060)
11. Hoban S *et al.* 2020 Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. *Biol. Conserv.* **248**, 108654. (doi:10.1016/j.biocon.2020.108654)
12. Mimura M *et al.* 2017 Understanding and monitoring the consequences of human impacts on intraspecific variation. *Evol. Appl.* **10**, 121–139. (doi:10.1111/eva.12436)
13. Carlson SM, Satterthwaite WH. 2011 Weakened portfolio effect in a collapsed salmon population complex. *Can. J. Fish. Aquat. Sci.* **68**, 1579–1589. (doi:10.1139/f2011-084)
14. Griffiths JR *et al.* 2014 Performance of salmon fishery portfolios across western North America. *J. Appl. Ecol.* **51**, 1554–1563. (doi:10.1111/1365-2664.12341)
15. Quinn TP. 2018 *The Behavior and Ecology of Pacific Salmon and Trout*. University of Washington press.
16. Williams JG. 2006 Central Valley Salmon: A Perspective on Chinook and Steelhead in the Central Valley of California. *San Fr. Estuary Watershed Sci.* **4**. (doi:10.15447/sfews.2006v4iss3art2)
17. 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.* **73**, 1015–1030. (doi:10.1139/cjfas-2015-0345)
18. 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.* **70**, 37160–37204.

19. Reed DH, Frankham R. 2001 How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution (N. Y).* **55**, 1095–1103. (doi:10.1111/j.0014-3820.2001.tb00629.x)
20. Reed DH, Frankham R. 2003 Correlation between fitness and genetic diversity. *Conserv. Biol.* **17**, 230–237. (doi:10.1046/j.1523-1739.2003.01236.x)
21. Meek MH, Baerwald MR, Stephens MR, Goodbla A, Miller MR, Tomalty KMH, May B. 2016 Sequencing improves our ability to study threatened migratory species: Genetic population assignment in California's Central Valley Chinook salmon. *Ecol. Evol.* **6**, 7706–7716. (doi:10.1002/ece3.2493)
22. 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.* **3**, e1603198. (doi:10.1126/sciadv.1603198)
23. Thompson TQ *et al.* 2019 Anthropogenic habitat alteration leads to rapid loss of adaptive variation and restoration potential in wild salmon populations. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 177–186. (doi:10.1073/pnas.1811559115)
24. 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 (80-.).* **370**, 609–613. (doi:10.1126/SCIENCE.ABA9059)
25. 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. R. Soc. B Biol. Sci.* **285**, 20180935. (doi:10.1098/rspb.2018.0935)
26. 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. *Genet. Sel. Evol.* **52**, 9. (doi:10.1186/s12711-020-0529-8)
27. Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW. 2010 Conservation genetics in transition to conservation genomics. *Trends Genet.* **26**, 177–187. (doi:10.1016/j.tig.2010.01.001)
28. Waples RS, Lindley ST. 2018 Genomics and conservation units: The genetic basis of adult migration timing in Pacific salmonids. *Evol. Appl.* **11**, 1518–1526. (doi:10.1111/eva.12687)
29. 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.* , cjfas-2019-0171. (doi:10.1139/cjfas-2019-

0171)

30. 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.* **21**, 237–249. (doi:10.1111/j.1365-294X.2011.05305.x)

31. 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* **1**, 171–182. (doi:10.1534/g3.111.000240)

32. Chen S, Zhou Y, Chen Y, Gu J. 2018 Fastp: An ultra-fast all-in-one FASTQ preprocessor. In *Bioinformatics*, pp. i884–i890. Narnia. (doi:10.1093/bioinformatics/bty560)

33. 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**, e0195461. (doi:10.1371/journal.pone.0195461)

34. Li H, Durbin R. 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760. (doi:10.1093/bioinformatics/btp324)

35. Quinlan AR. 2014 BEDTools: The Swiss-Army tool for genome feature analysis. *Curr. Protoc. Bioinforma.* **2014**, 11.12.1-11.12.34. (doi:10.1002/0471250953.bi1112s47)

36. Garrison E, Marth G. 2012 Haplotype-based variant detection from short-read sequencing. *PLoS One* **11**, e0151651. (doi:arXiv:1207.3907 [q-bio.GN])

37. 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.* **27**, 3193–3206. (doi:10.1111/mec.14792)

38. 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.* (doi:10.1111/1755-0998.12647)

39. Jombart T, Devillard S, Balloux F. 2010 Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94. (doi:10.1186/1471-2156-11-94)

40. Weir BS, Cockerham CC. 1984 Estimating F-Statistics for the Analysis of Population Structure. *Evolution (N. Y).* **38**, 1358. (doi:10.2307/2408641)

41. Gosselin T, Anderson EC, Bradbury IR. 2016 assigner: Assignment Analysis with GBS/RAD Data using R. *R Packag.* (doi:doi□: 10.5281/zenodo.51453)

42. Hollenbeck CM, Portnoy DS, Gold JR. 2016 A method for detecting recent changes in contemporary effective population size from linkage disequilibrium at linked and unlinked loci. *Heredity (Edinb).* **117**, 207–216. (doi:10.1038/hdy.2016.30)

43. Mckinney GJ *et al.* 2016 An integrated linkage map reveals candidate genes underlying adaptive variation in Chinook salmon (Oncorhynchus tshawytscha). *Mol. Ecol. Resour.* **16**, 769–783. (doi:10.1111/1755-0998.12479)

44. Benjamini Y, Hochberg Y. 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300. (doi:10.2307/2346101)

45. Nei M. 1987 *Molecular Evolutionary Genetics*. (doi:10.7312/nei-92038)

46. 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* **2**, e281. (doi:10.7717/peerj.281)

47. Prieto I, Violle C, Barre P, Durand JL, Ghesquiere M, Litrico I. 2015 Complementary effects of species and genetic diversity on productivity and stability of sown grasslands. *Nat. Plants* **1**. (doi:10.1038/nplants.2015.33)

48. Siefert A *et al.* 2015 A global meta-analysis of the relative extent of intraspecific trait variation in plant communities. *Ecol. Lett.* **18**, 1406–1419. (doi:10.1111/ele.12508)

49. Frankham R, Bradshaw CJA, Brook BW. 2014 Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biol. Conserv.* **170**, 56–63. (doi:10.1016/j.biocon.2013.12.036)

50. California-HSRG. 2012 California Hatchery Scientific Review Group - California Hatchery Review Report. Prepared for the US Fish and Wildlife Service and Pacific States Marine Fisheries Commission.

51. Williamson KS, May B. 2005 Homogenization of Fall-Run Chinook Salmon Gene Pools in the Central Valley of California, USA. *North Am. J. Fish. Manag.* **25**, 993–1009. (doi:10.1577/m04-136.1)

52. Allendorf FW, Hohenlohe PA, Luikart G. 2010 Genomics and the future of conservation genetics. *Nat. Rev. Genet.* **11**, 697–709. (doi:10.1038/nrg2844)

53. Ryman N, Laikre L. 1991 Effects of Supportive Breeding on the Genetically Effective Population Size. *Conserv. Biol.* **5**, 325–329. (doi:10.1111/j.1523-1739.1991.tb00144.x)

54. Gulisija D, Kim Y. 2015 Emergence of long-term balanced polymorphism under cyclic selection of spatially variable magnitude. *Evolution (N. Y).* **69**, 979–992. (doi:10.1111/evo.12630)

55. Svardal H, Rueffler C, Hermisson J. 2015 A general condition for adaptive genetic polymorphism in temporally and spatially heterogeneous environments. *Theor. Popul. Biol.* **99**, 76–97. (doi:10.1016/j.tpb.2014.11.002)

56. Bertram J, Masel J. 2019 Different mechanisms drive the maintenance of polymorphism at loci subject to strong versus weak fluctuating selection. *Evolution (N. Y).* **73**, 883–896.

(doi:10.1111/evo.13719)

57. Vazquez-Dominguez E, Pinero D, Ceballos G. 1999 Linking Heterozygosity, Demography, and Fitness of Tropical Populations of Liomys pictus. *J. Mammal.* **80**, 810–822. (doi:10.2307/1383250)

58. 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. U. S. A.* **111**, 3775–3780. (doi:10.1073/pnas.1318945111)

59. Reusch TBH, Ehlers A, Hämmerli A, Worm B. 2005 Ecosystem recovery after climatic extremes enhanced by genotypic diversity. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 2826–2831. (doi:10.1073/pnas.0500008102)

60. Scott PA, Allison LJ, Field KJ, Averill-Murray RC, Bradley Shaffer H. 2020 Individual heterozygosity predicts translocation success in threatened desert tortoises. *Science (80-.).* **370**, 1086–1089. (doi:10.1126/science.abb0421)

61. 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.* **22**, 1820–1835. (doi:10.1111/mec.12227)

62. Hess JE, Ackerman MW, Fryer JK, Hasselman DJ, Steele CA, Stephenson JJ, Whiteaker JM, Narum SR. 2016 Differential adult migration-timing and stock-specific abundance of steelhead in mixed stock assemblages. *ICES J. Mar. Sci. J. du Cons.* **73**, 2606–2615. (doi:10.1093/icesjms/fsw138)

63. Ford M *et al.* 2020 Reviewing and synthesizing the state of the science regarding associations between adult run timing and specific genotypes in Chinook salmon and steelhead. , 0–47. (doi:10.25923/mv3n-zb79)

64. Dedrick AG, Baskett ML. 2018 Integrating genetic and demographic effects of connectivity on population stability: The case of hatchery trucking in salmon. *Am. Nat.* **192**, E62–E80. (doi:10.1086/697581)

65. 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* **91**, 335–350. (doi:10.1007/s10584-008-9427-8)

66. Kardos M, Luikart G. 2019 The genetic architecture of fitness drives population viability during rapid environmental change. *bioRxiv* , 660803. (doi:10.1101/660803)

67. Hairston NG, Ellner SP, Geber MA, Yoshida T, Fox JA. 2005 Rapid evolution and the convergence of ecological and evolutionary time. *Ecol. Lett.* **8**, 1114–1127. (doi:10.1111/j.1461-

0248.2005.00812.x)

68. Richardson JL, Urban MC, Bolnick DI, Skelly DK. 2014 Microgeographic adaptation and the spatial scale of evolution. *Trends Ecol. Evol.* **29**, 165–176. (doi:10.1016/j.tree.2014.01.002)

69. Laikre L, Schwartz MK, Waples RS, Ryman N. 2010 Compromising genetic diversity in the wild: Unmonitored large-scale release of plants and animals. *Trends Ecol. Evol.* **25**, 520–529. (doi:10.1016/j.tree.2010.06.013)

Tables

Table 1: Sample sizes and effective population size Ne for all run/tributary groups. Negative point estimates occur when LD attributed to sampling variation is larger than LD attributed to drift. This can either be interpreted as an "infinite" population size (drift is negligible) or due to sample size being too small for an accurate estimate (LD attributed to sampling variation is a function of sample size).

Run	**Tributary/Hatchery**	**Sample Size**	***Ne* (95% CI)**
Fall	Coleman Hatchery	30	615 (594; 637)
	Mill Creek	20	*-6,801 (-56,952; -3,618)*
	Deer Creek	15	*-3,939 (-9,323; -2,497)*
	Butte Creek	21	3,802 (2,708; 6,378)
	Feather River Hatchery	27	5,172 (3,823; 7,988)
	Nimbus Hatchery	30	7,432 (5,147 – 13,351)
	Mokelumne River Hatchery	28	4,106 (3,207; 5,704)
	Stanislaus River	23	2,567 (2,040; 3,460)
	Tuolumne River	30	6,353 (4,383;11,538)
	Merced River Hatchery	15	1,125 (940; 1,399)
	Merced River	31	11,436 (6,768; 36,804)
Late-Fall	Upper Sacramento River	21	3,114 (2,398; 4,435)
Winter	Upper Sacramento River	26	174 (171; 177)
Spring	Mill Creek	16	85 (84; 86)
	Deer Creek	27	188 (185; 191)
	Butte Creek	19	561 (529; 597)
	Feather River Hatchery	7	*-94 (-98; -91)*

Figures

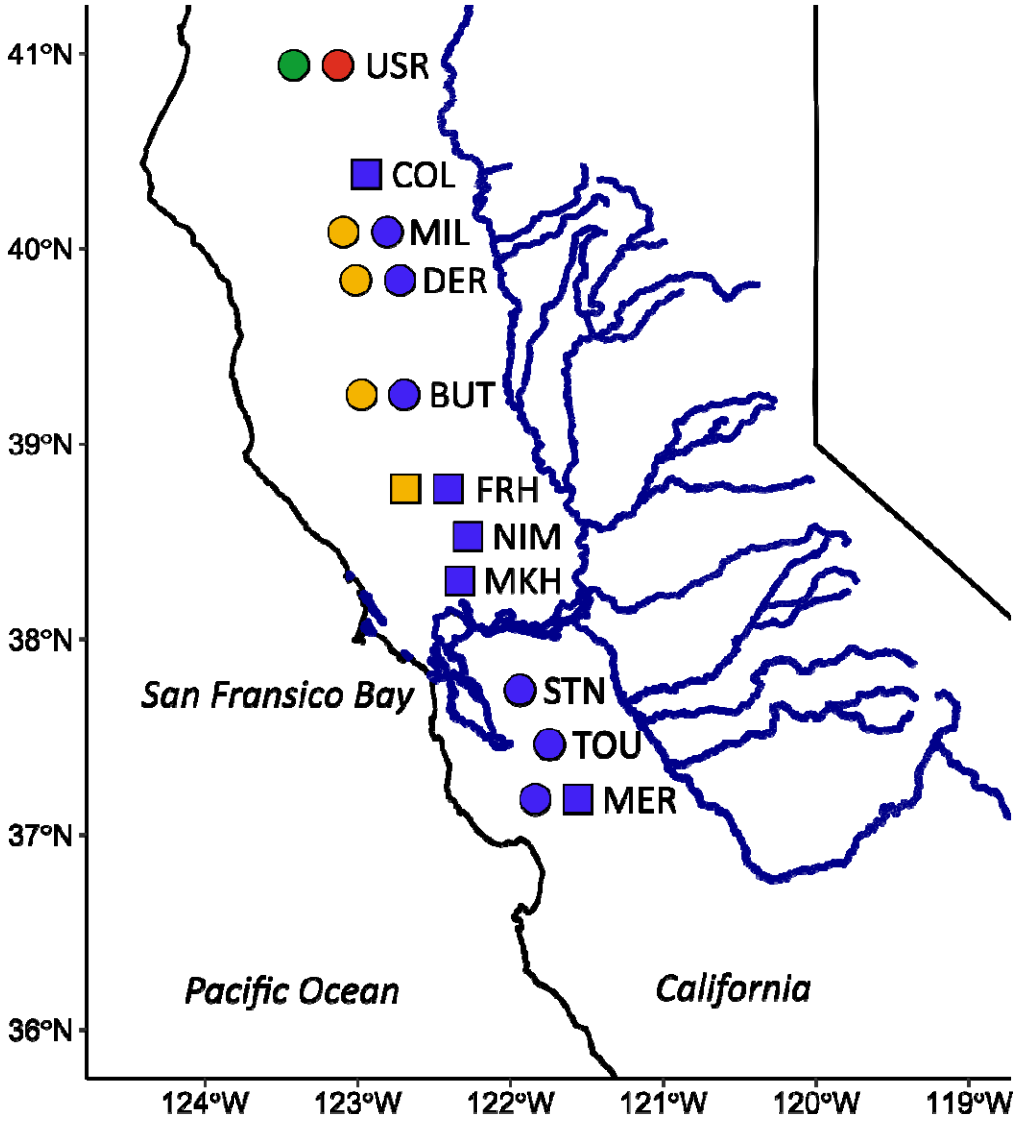

Figure 1: Tributaries of the California's Central Valley sampled for this study. Hatchery individuals are represented by squares and wild populations by circles. Colors represent the sampled run type at each location (Spring = green, Fall = blue, Late-Fall = red, Winter = yellow). Abbreviations for tributaries 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).

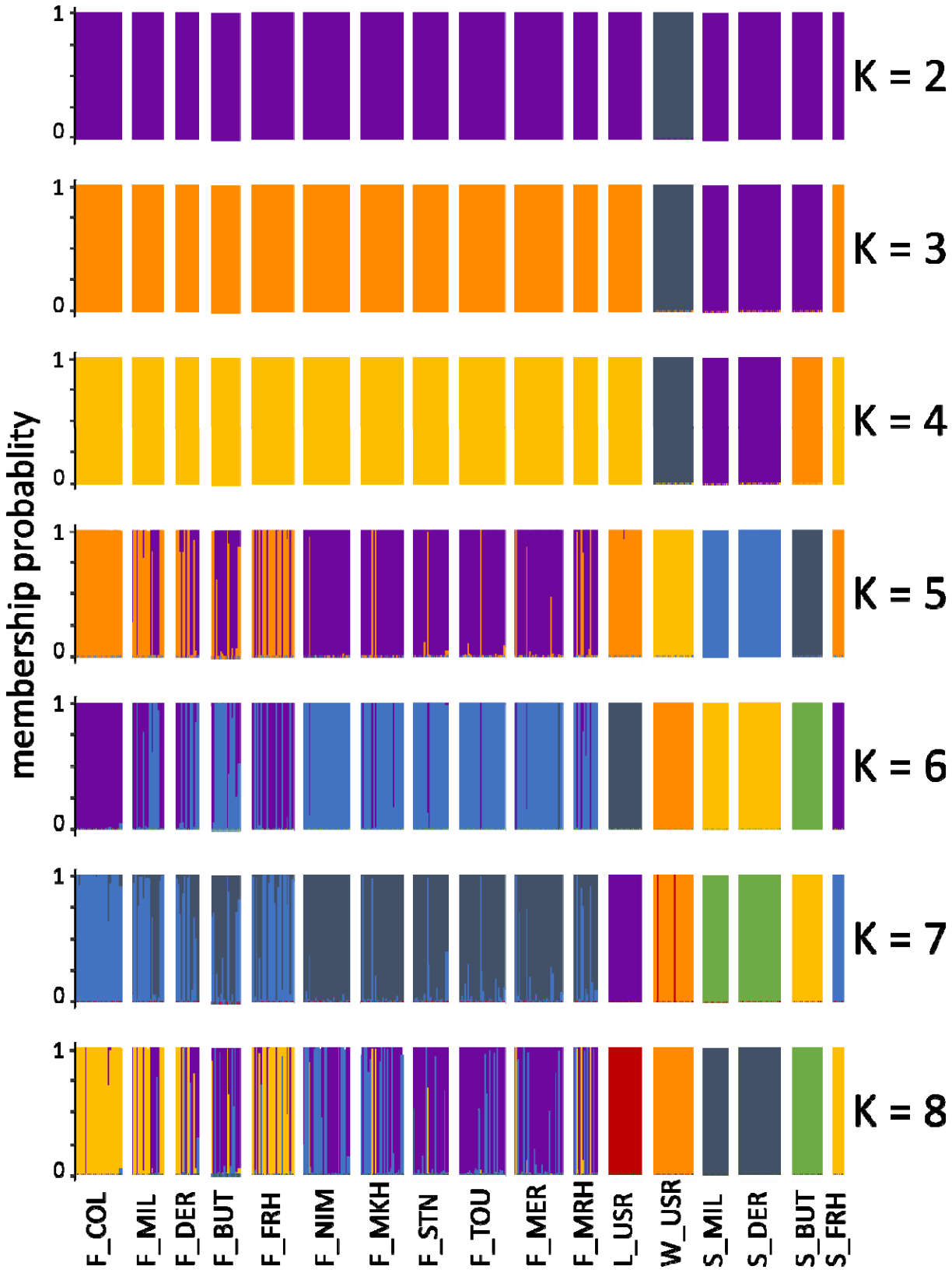

Figure 2: Membership probability of each individual to clusters identified using k-means hierarchical clustering for K = 2 – 8.

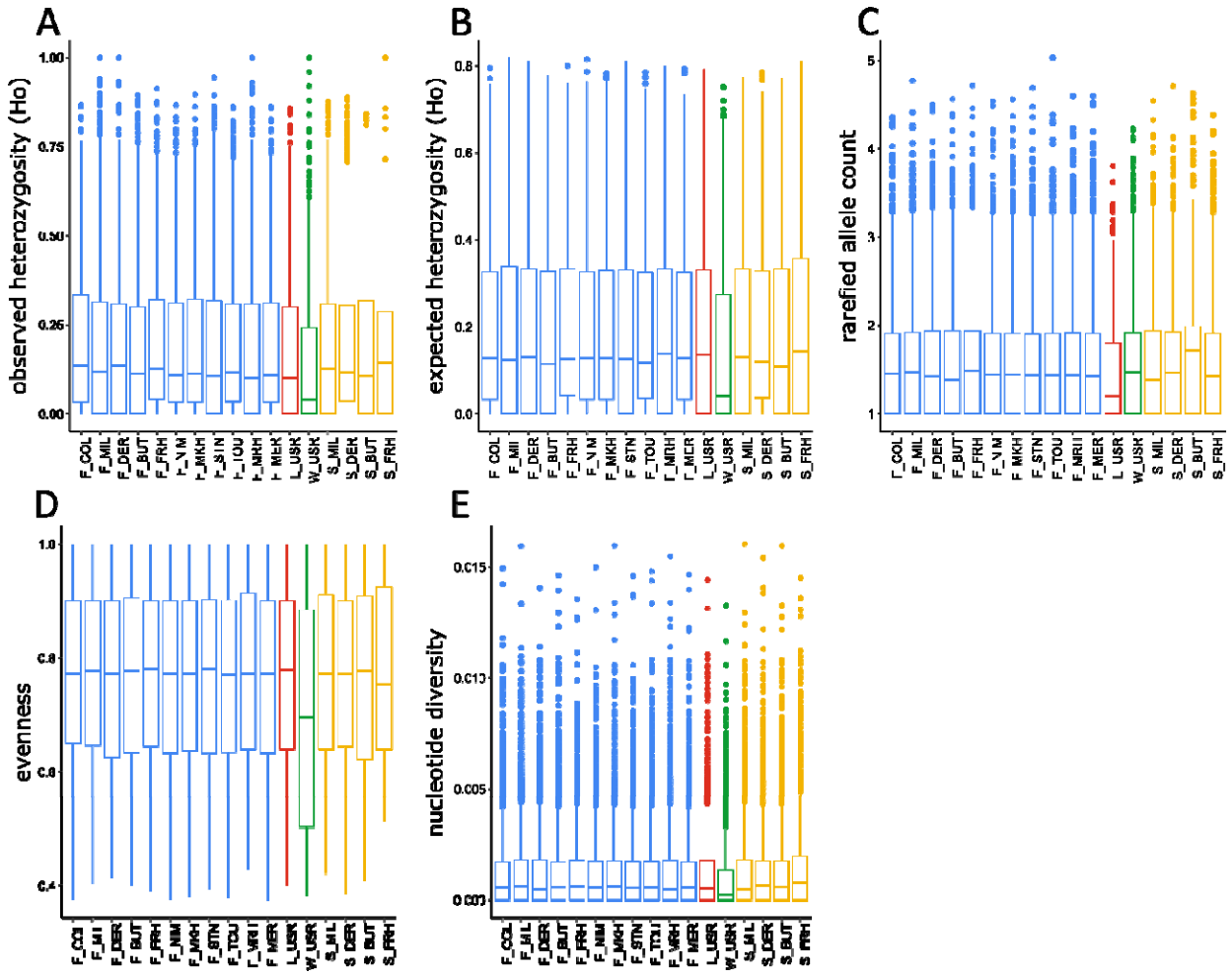

Figure 3: Assessment of the distribution of genomic diversity for individuals grouped by run type within tributaries using heterozygosity-based parameters (A. observed heterozygosity Ho, B. expected heterozygosity Hs), measures of allele diversity (C. Allelic richness, D. Evenness), and sequence-based parameters (E. observed nucleotide diversity). Fall individuals are in blue, Late-Fall in red, Winter depicted in green, and Spring in yellow.

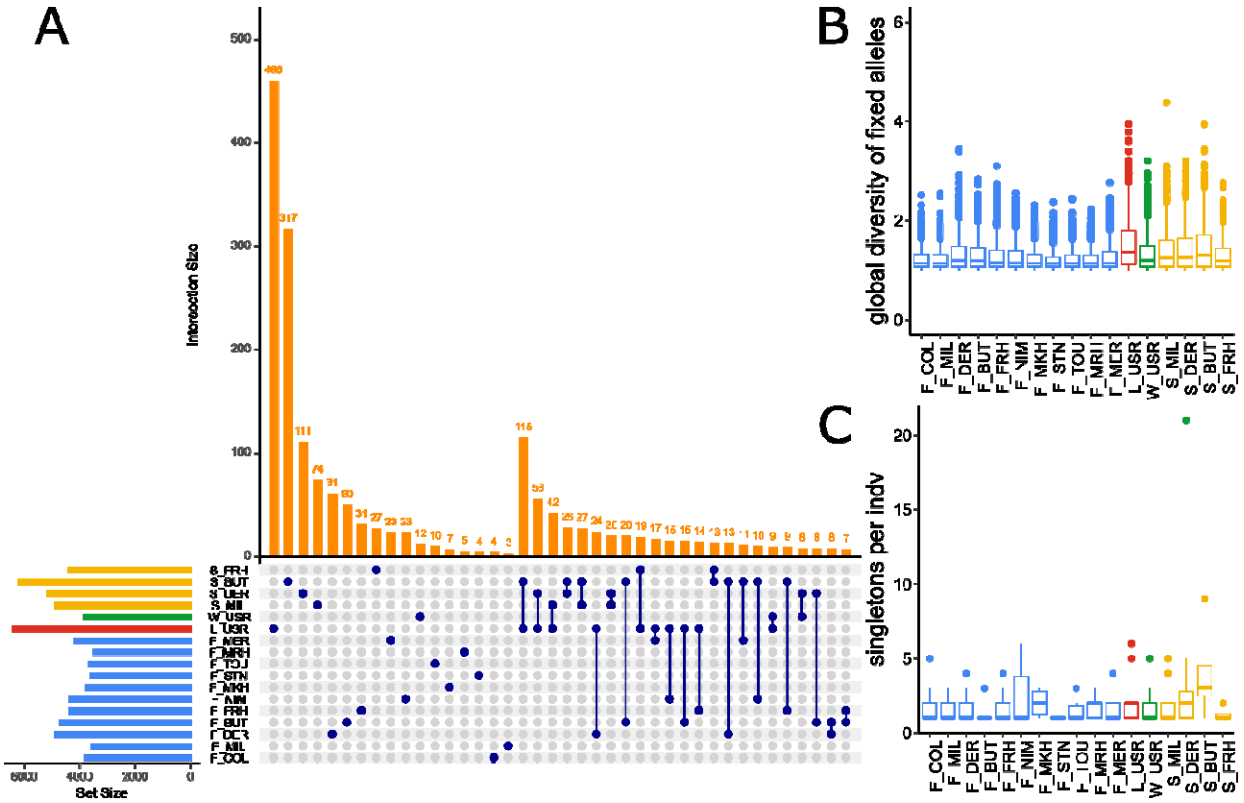

Figure 4: Assessment of fixed loci and singletons. A. Comparisons of fixed loci across run/tributary groups. The set size (horizontal bars) indicate the total number of fixed loci in a given group, the intersect size (vertical orange bars) the number of loci fixed in a single group (single blue dot) or in two (blue 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. Fall individuals are in blue, Late-Fall in red, Winter depicted in green, and Spring in yellow.

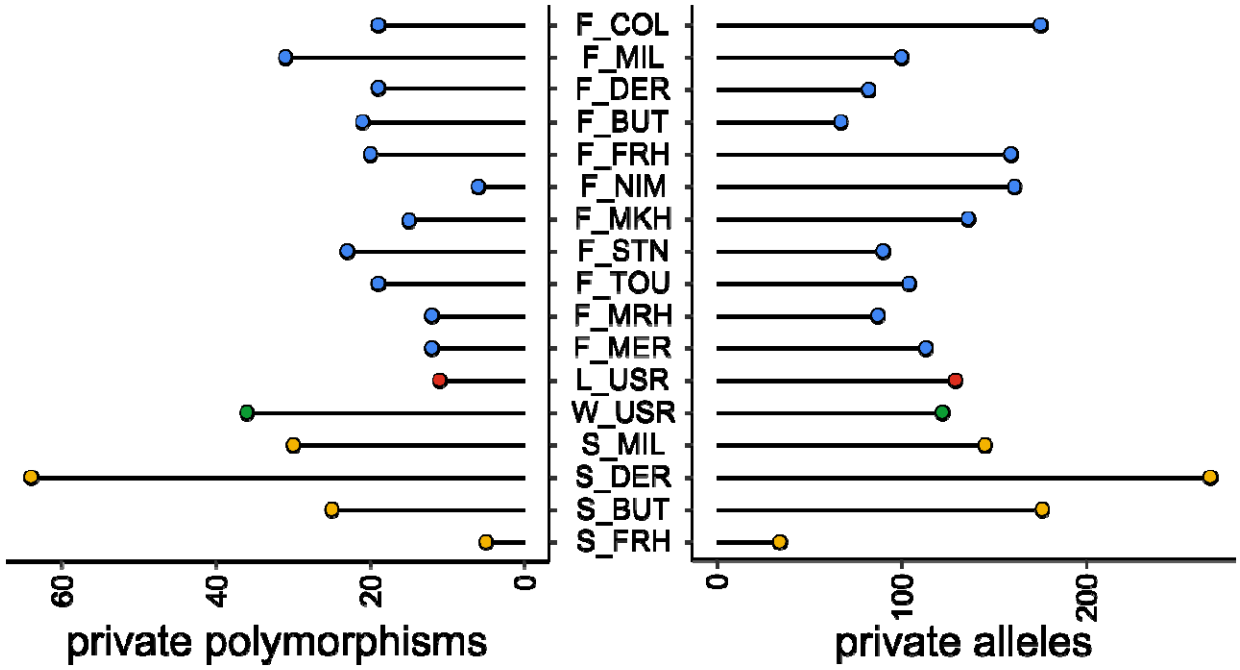

Figure 5: Assessment of unique genomic diversity. Left axis indicates the number of private polymorphisms (loci only variable in a single group), right axis indicates the number of private alleles (alleles found only in a single group) for individuals grouped by run and tributary. Fall individuals are in blue, Late-Fall in red, Winter depicted in green, and Spring in yellow.

Supplementary material

Supplementary Material 1: Data acquisition and processing (Genotyping).

Supplementary Material 2: Assessment of population differentiation and patterns of genomic diversity containing R code used for data analysis and supplementary figures and tables cited in the manuscript.