NRC Research Press

ARTICLE

Identifying hidden biocomplexity and genomic diversity in Chinook salmon, an imperiled species with a history of anthropogenic influence

Mariah H. Meek, Molly R. Stephens, Alisha Goodbla, Bernie May, and Melinda R. Baerwald

Can. J. Fish. Aquat. Sci. Downloaded from www.nrcresearchpress.com by National Marine Mammal Lab Lib on 02/25/20
For personal use only.

Abstract: Biocomplexity is an important mechanism for population resilience in changing environments. However, we are just beginning to understand how to identify biocomplexity so that species management efforts promote resilience and stability. Genomic techniques are emerging as an important method for identifying biocomplexity. Central Valley (CV) Chinook salmon are an example of a species at risk of extinction if better methods for identifying and protecting biocomplexity are not employed. To address this knowledge gap, we employed restriction site associated DNA sequencing to conduct the first genomic study of all major populations of CV Chinook salmon. We found greater population structure across the Central Valley than previously documented. Additionally, we show evidence for differentiation and adaptation within migratory phenotypes despite high levels of gene flow. We also determined that genomic data can vastly improve our ability to assign individuals to their natal populations, even as they mix during migration, a finding that will assist management practices. These results demonstrate how genomic study can greatly improve our ability to identify and conserve biocomplexity.

Résumé : La biocomplexité est un mécanisme important pour la résilience des populations dans des milieux changeants. Nous ne faisons toutefois que commencer à comprendre comment circonscrire la biocomplexité pour nous permettre de gérer les espèces d'une manière qui favorise leur résilience et leur stabilité. De nouvelles techniques de génomique constituent une approche importante pour circonscrire la biocomplexité. Le saumon quinnat de la vallée centrale (VC) est un exemple d'espèce à risque de disparition en l'absence de meilleures méthodes pour circonscrire et protéger sa biocomplexité. Afin de combler ces lacunes dans les connaissances, nous avons utilisé le séquençage d'ADN associé à des sites de restriction pour réaliser la première étude génomique de toutes les grandes populations de saumons quinnat de la VC. Nous avons ainsi constaté que les populations à la grandeur de la vallée centrale présentent une structure plus importante que ce qui avait déjà été rapporté. Nous présentons également des preuves de différentiation et d'adaptation au sein de phénotypes migratoires, et ce, malgré des niveaux élevés de flux génétique. En ce qui concerne les pratiques de gestion, nous constatons également que les données génomiques peuvent grandement améliorer la capacité d'affecter les individus à leurs populations natales, même s'ils se mélangent durant la migration. Ces résultats démontrent comment l'approche génomique peut améliorer considérablement la capacité de circonscrire et de conserver la biocomplexité. [Traduit par la Rédaction]

Introduction

Biocomplexity reflects the amount of intraspecific diversity within and among populations and is important for the stability and resilience of metapopulations (Hilborn et al. 2003). Identifying the factors that contribute to biocomplexity, such as genetic diversity, increases our understanding of the role these factors play in species resilience and persistence. It is also important for developing sustainable management practices that promote long-term stability (Dedrick and Baskett 2018). However, fine-scale information on genetic diversity can be lacking for species of conservation or management concern, compromising the ability of resource managers to monitor and enhance the genetic variation essential to the biocomplexity of metapopulations (Ouborg et al. 2010; Angeloni et al. 2012).

Pacific salmonids are an excellent example illustrating the importance of biocomplexity for persistence and harvest production. Pacific salmonid stocks in Alaska have been shown to display high biocomplexity, allowing for stability in annual returns (Schindler et al. 2010). This is often referred to as the portfolio effect, where differences among individual populations in response to environmental conditions can drive greater stability of the whole population complex when compared with the stability of each population individually (Hilborn et al. 2003; Schindler et al. 2010). For example, variability in annual returns of sockeye salmon (*Oncorhynchus nerka*) in the Bristol Bay of Alaska is over two times lower than returns would be without the population and life history diversity present in the system (Schindler et al. 2010). In contrast, demographic synchrony has increased over recent decades in Chinook salmon (*Oncorhynchus tshawytscha*) from the Snake River, decreasing the portfolio effect and increasing extinction risk (Moore et al. 2010). This increase in synchrony has been attributed to human activities that have decreased genetic diversity and homogenized habitats, such as hatchery production and dam construction (Moore et al. 2010).

Our ability to monitor changes in biocomplexity is limited by our ability to identify the components of biocomplexity and man-

Received 15 May 2019. Accepted 20 August 2019.

M.H. Meek. Department of Integrative Biology and AgBio Research, Michigan State University, East Lansing, MI 48824, USA.
M.R. Stephens. School of Natural Sciences, University of California Merced, Merced, CA 95343, USA.
A. Goodbla and B. May. Department of Animal Science, University of California Davis, Davis, CA 95616, USA.
M.R. Baerwald. California Department of Water Resources, Division of Environmental Services, 3500 Industrial Boulevard, West Sacramento, CA 95691, USA.
Corresponding author: Mariah H. Meek (email: mhmeek@msu.edu).
Copyright remains with the author(s) or their institution(s). Permission for reuse (free in most cases) can be obtained from RightsLink.

Can. J. Fish. Aquat. Sci. Downloaded from www.nrcresearchpress.com by National Marine Mammal Lab Lib on 02/25/20
For personal use only.

age for fine-scale differences in these components among populations (Hilborn et al. 2003; Moore et al. 2010). Advances in genomic techniques, however, can address this problem by identifying, at a much finer resolution, the components that influence biocomplexity (e.g., Larson et al. 2014; Prince et al. 2017). By applying genomics to identify the components that make up fine-scale biocomplexity, we can then design management activities to protect all the diversity found within species, boosting the portfolio effect and decreasing extinction risk.

In this study, we apply genomic tools to better understand the biocomplexity contained within Central Valley (CV) Chinook salmon. The Central Valley of California is home to four runs of Chinook salmon, each named for the timing of their freshwater spawning migration (fall, late-fall, spring, and winter runs). These runs are differentiated by the timing of major life history transitions, and the Central Valley is the only area in the species range where these four runs co-occur (Williams 2006). Given this diversity in life history, Chinook salmon in the CV should have a robust portfolio. However, CV Chinook have among the weakest portfolio in the species range (Griffiths et al. 2014).

The abundance of wild CV Chinook populations has been in serious decline in recent decades (Yoshiyama et al. 1998), and several of these populations are predicted to go extinct in California 0–100 years if no new actions are taken (Moyle et al. 2017). Sacramento River winter run and Central Valley spring run Chinook salmon are listed under the United States Endangered Species Act (ESA) as endangered and threatened, respectively (Federal Register 1999, 2005), while the fall and late-fall runs are federal species of concern (Myers et al. 1998). Spring and winter runs are each designated as their own evolutionarily significant unit (ESU) and given the protection of "species" under the ESA, while the fall and late-fall runs are grouped into a single ESU (Myers et al. 1998).

A population collapse in the early 2000s resulted in the closing of the fishery in 2008 and 2009. Loss of life history and genetic diversity are cited as making the stocks susceptible to collapse (Lindley et al. 2009). Previous work has shown that much of the biocomplexity in the CV has been lost (Carlson and Satterthwaite 2011). This includes decreased portfolio effects and increases in synchrony among populations and resulting increases in variance in population abundance (Carlson and Satterthwaite 2011; Griffiths et al. 2014). This loss of biocomplexity includes genetic homogenization of the fall run (Williamson and May 2005) and introgression between runs due to hatchery practices (California Hatchery Scientific Review Group 2012).

To protect all the diversity found in a portfolio of stocks and promote population buffering, it is vital to be able to accurately assign individuals to their population of origin. This can be very difficult when multiple populations intermingle along their migration paths, as is the case with CV Chinook salmon. CV Chinook runs can co-occur on their spawning grounds; even more commonly, they co-occur along migration routes, as juveniles in rearing grounds, and in the open ocean (Williams 2006, 2012). For example, all four runs of CV Chinook use the limited floodplain habitat in the San Francisco Estuary to feed and grow before they move out to the ocean, and they can occur in this habitat at overlapping times (Sommer et al. 2001*a*, 2001*b*). Currently, much effort is being invested in understanding how different populations of juveniles use the San Francisco Bay Delta system (Johnson et al. 2017). Because there are no morphological differences among the different runs, this has made accurately monitoring and managing the different populations highly error prone (Harvey et al. 2014).

Hatchery management practices can also disrupt the natural processes that create and maintain population diversification and biocomplexity (Huber and Carlson 2015; Satterthwaite and Carlson 2015). In the CV, hatcheries were created to mitigate the negative effects of dams on Chinook populations, yet we now know that hatcheries can also have strong negative effects on wild populations (Araki et al. 2007; but also see Hess et al. 2012). There are five hatcheries throughout the CV producing fall run populations and one each for the remaining runs. Hatchery practices have included trucking and releasing juveniles downstream of the hatchery or in the San Francisco Estuary to decrease juvenile mortality, which hinders homing during spawning migrations (California Hatchery Scientific Review Group 2012). Additionally, the only CV hatchery to produce both spring and fall run Chinook salmon (the Feather River Hatchery) has caused hybridization between the two runs in their hatchery breeding program by not properly separating the broodstock for the two runs (California Hatchery Scientific Review Group 2012). These practices have led to increased straying among populations and increased synchrony via demographic coupling (Huber and Carlson 2015; Satterthwaite and Carlson 2015). This in turn has resulted in the mixing of gene pools among runs and populations that were previously distinct, and ultimately complicates the identification of management units and assignment of individuals to those units.

Previous genetic studies of CV Chinook populations have reported varying levels of genetic distinctiveness among the runs and locations (reviewed in Lindley et al. 2004; Williams, 2006). Studies have ranged from those that have reported divergence among all four run types (Banks et al. 2000) to studies that have found evidence for and against introgression between the runs (Banks et al. 2000; Garza et al. 2008; O'Malley et al. 2013; Clemento et al. 2014). These conflicting genetic reports make it difficult to design management actions to protect biocomplexity in CV Chinook populations. All these genetic studies used limited marker sets (range: 10–95 markers). Studies that take advantage of advanced next-generation sequencing techniques — sampling a greater proportion of the genome — are needed to clarify the relationships among CV Chinook salmon and elucidate how genomic information can assist in conserving biocomplexity in this species.

We use restriction site associated DNA sequencing (RAD-seq; Baird et al. 2008; Etter et al. 2011) to conduct the first comprehensive genome-wide genetic study of all the major populations of CV Chinook salmon. Of primary importance is identifying the unique genetic diversity contained within and among the CV Chinook salmon population complex. We aim to do this by investigating the patterns of diversification shown across the CV Chinook salmon genome. Specifically, we ask the following questions: (*i*) What is the population structure of CV Chinook salmon and how is diversity partitioned among migratory phenotypes? (*ii*) Are there signals of fine-scale population structure within the fall and spring runs, despite high levels of human-mediated gene flow? (*iii*) Can we assign individuals to their migratory phenotypes (i.e., run type) and populations using genomic data?

Information resulting from these questions will be greatly beneficial for identifying the biocomplexity found within CV Chinook salmon and enabling resource managers to develop strategies that protect genetic variation.

Materials and methods

Sample collection

We obtained fin tissue samples of adult Chinook salmon from the Anadromous Resources Tissue Archive at the California Department of Fish and Wildlife. Samples were originally collected during spawning migrations and come from all major populations within the four Chinook salmon runs (fall, late fall, spring, winter) (Table 1; Fig. 1). We analyzed 28–32 individuals per population. Previous work has shown this sample size to be more than adequate for capturing genetic variation with genomic data (Nazareno et al. 2017). Populations were represented by samples collected from 2 or more years, to capture temporal variation. However, previous work has shown temporal variation within populations of Chinook salmon to be relatively small compared

Can. J. Fish. Aquat. Sci. Downloaded from www.nrcresearchpress.com by National Marine Mammal Lab Lib on 02/25/20
For personal use only.

Table 1. Sample information.

Location	Years sampled	Location abbreviation	Sample size	H_e	H_o	A_r	F_{is}
Fall run							
Butte Creek	2002–2004	F_BUT	32/16	0.246	0.259	1.844	−0.054 (−0.063 to −0.044)
Coleman Hatchery (Battle Creek)	2002	F_COL	30/30	0.245	0.261	1.915	−0.064 (−0.072 to −0.056)
Deer Creek	2002–2004	F_DER	30/13	0.245	0.259	1.826	−0.056 (−0.065 to −0.046)
Feather River Fish Hatchery	2001, 2007–2010	F_FRH	28/21	0.244	0.259	1.870	−0.061 (−0.069 to −0.052)
Merced River	2001, 2004, 2008	F_MER	34/27	0.245	0.262	1.910	−0.072 (−0.080 to −0.063)
Mill Creek	2001–2004	F_MIL	35/14	0.249	0.264	1.847	−0.061 (−0.070 to −0.051)
Mokelumne River Fish Hatchery	2004–2005, 2008	F_MKH	30/18	0.244	0.258	1.854	−0.058 (−0.068 to −0.049)
Merced River Fish Hatchery	2001–2004	F_MRH	30/21	0.245	0.258	1.877	−0.050 (−0.059 to −0.041)
Nimbus Fish Hatchery (American River)	2002–2005	F_NIM	30/25	0.243	0.259	1.886	−0.068 (−0.076 to −0.060)
Stanislaus River	2001, 2002, 2004	F_STN	30/12	0.243	0.259	1.791	−0.067 (−0.077 to −0.058)
Tuolumne River	2004, 2008	F_TOU	33/18	0.246	0.260	1.862	−0.061 (−0.070 to −0.052)
Late-fall run							
Upper Sacramento River	2003–2005	L_USR	30/24	0.245	0.26	1.881	−0.062 (−0.071 to −0.054)
Spring run							
Butte Creek	2008–2009	S_BUT	30/22	0.241	0.250	1.827	−0.034 (−0.043 to −0.026)
Deer Creek	2002, 2005	S_DER	31/20	0.246	0.254	1.864	−0.031 (−0.039 to −0.022)
Mill Creek	2002, 2004, 2005	S_MIL	32/18	0.247	0.261	1.858	−0.056 (−0.065 to −0.048)
Feather River Fish Hatchery	2009–2010	S_FRH	30/30	0.247	0.263	1.910	−0.068 (−0.076 to −0.060)
Winter run							
Upper Sacramento River	2001, 2002	W_USR	30/30	0.194	0.199	1.717	−0.030 (−0.040 to −0.020)

Note: Sample size column gives total number used in RAD-sequencing/total number used in analyses after filtering. H_e, expected heterozygosity; H_o, observed heterozygosity; A_r, mean number of alleles; F_{is}, inbreeding coefficient (95% confidence interval).

with variation among populations (Banks et al. 2000; Beacham et al. 2006; Narum et al. 2008).

Molecular biology

We used the molecular methods outlined in Meek et al. (2016). Briefly, we extracted genomic DNA and constructed RAD libraries using the *Sbf*I restriction enzyme following the protocol of Miller et al. (2012). Each sample was ligated with a unique custom 6 base pair (bp) barcode, and we multiplexed 30–47 individuals per library. We sequenced the libraries as 100 bp single-end reads on an Illumina HiSeq 2000 (Vincent J. Coates Genomics Sequencing Laboratory, Berkeley, California, USA), running a single library per lane.

SNP genotyping

We aligned the sequences from each individual to the RAD loci in Meek et al. (2016) using Bowtie (Langmead et al. 2009) and performed genotyping and quality filtering following the methods detailed in Meek et al. (2016). In summary, we trimmed the reads from the 3′ end to 92 bp and eliminated those with a >20% probability of sequencing error based on Phred scores. We also eliminated those that had one or more ambiguous base calls. We used the genotyping method of Lew et al. (2015). After genotyping, we removed individuals that were genotyped at fewer than the lower confidence interval of genotyped loci per individual (<6621 loci). We then removed loci that were genotyped at <70% of the remaining individuals, followed by removing individuals that were genotyped at <70% of remaining loci.

Alignment to Chinook linkage map

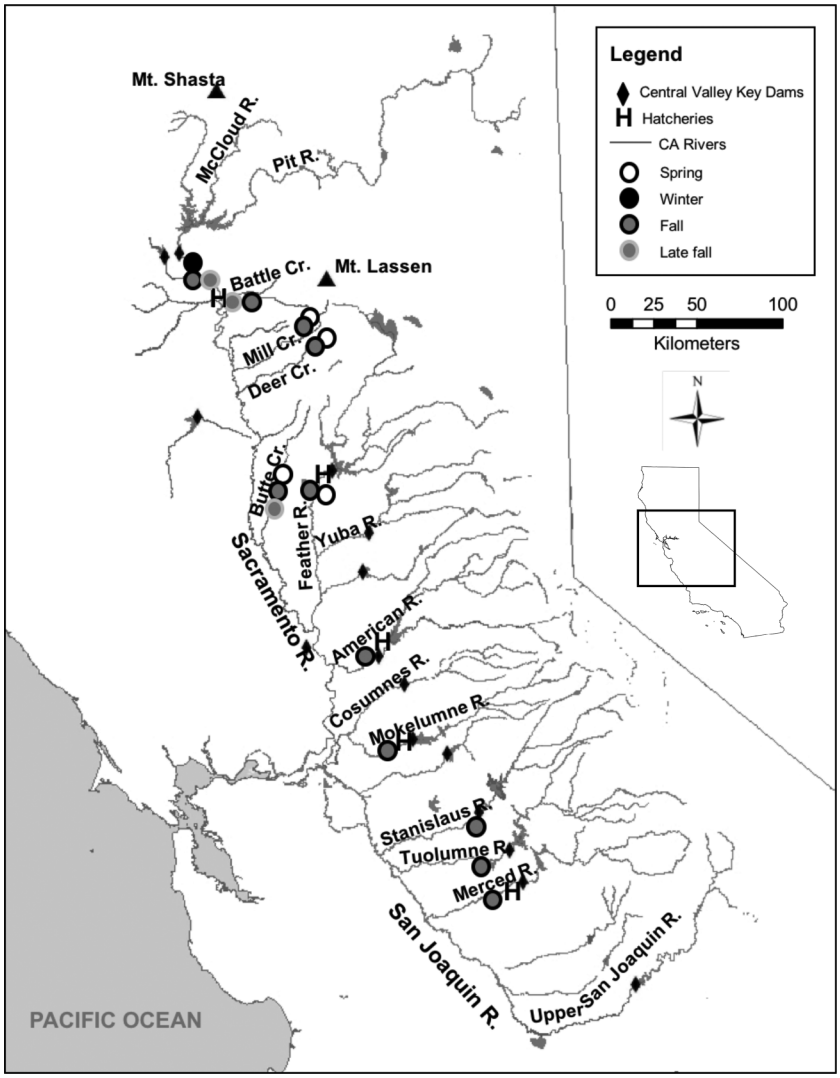
We aligned our RAD loci to the integrated Chinook linkage map, using the consensus female map (McKinney et al. 2016). We used BLAST+ (Camacho et al. 2009) with default settings to make the alignments. We further filtered the BLAST+ output in R (R Development Core Team 2005), only retaining alignments where the length was ≥78 bp, 1 or fewer mismatches, percent identity was ≥95%, and there were no gaps.

Population structure and genetic diversity analyses

We analyzed the data set for population structure using two methods. First, we used the software STRUCTURE (Pritchard et al. 2000) to identify genetically distinct populations in our data set. We ran five iterations of each model with three to eight clusters, with a burn-in of 20 000 steps followed by 750 000 steps. We then employed the programs CLUMPP (Jakobsson and Rosenberg 2007), in greedy mode, and Structure Harvester (Earl and VonHoldt 2012) to average over replicates. We visualized our plots in R using custom code (R Development Core Team 2005; available upon request).

To further explore population structure, we conducted a discriminant analysis of principal components (DAPC), as implemented in the R package adegenet (Jombart and Ahmed 2011). This method allows analysis of group clustering similar to Bayesian methods, such as STRUCTURE (Pritchard et al. 2000), without the assumptions of Hardy–Weinberg and linkage equilibria. DAPC has been shown to be better at finding fine-scale structure than the STRUCTURE software (Jombart et al. 2010; Benestan et al. 2015). We ran the DAPC once to optimize the number of principal components (PCs) that were retained, testing up to 25 clusters and retaining 250 PCs. In this run, the alpha score showed that retaining 11–14 PCs provides the optimal number without overfitting the data. In the final DAPC analysis, we tested up to 10 clusters, retaining all discriminant functions and the optimal number of principal components for each number of clusters (K). Results of the DAPC analysis identified several individuals that appeared to be mislabeled, misidentified in the field, or strays, as they had high probabilities of clustering with a different run. We removed any individual whose membership probability to an alternate population was >0.85 (individuals removed: 1 F_BUT, 2 F_MKH, 2 L_USR, and 1 S_DER; see location abbreviations in Table 1). We then calculated Weir and Cockerham's unbiased estimator of F_{ST} (Weir and Cockerham 1984), as implemented in GENODIVE (Meirmans and Van Tienderen 2004), using 999 permutations and determining significance with a Benjamini and Yekutieli (2001) false discovery rate corrected value ($p < 0.015$), as per Narum (2006). We also calculated allelic richness, observed and expected heterozygosity

Fig. 1. Central Valley Chinook salmon sampling locations. The points represent the rivers from which samples were collected, not the exact sampling location. Figure modified from Meek et al. (2016).



(Nei 1987), and inbreeding coefficients using GENODIVE. We conducted an analysis of molecular variance (AMOVA), using the infinite alleles model and 1000 permutations, to investigate the partitioning of genetic variation across populations and runs in GENODIVE. We calculated effective population size (N_e) with N_eEstimator (Do et al. 2014) using the following parameters: the linkage disequilibrium model with random mating; a minimum allele frequency cutoff of 0.02; and calculating the parametric 95% confidence interval for each estimate. We removed loci from the data set that were found to be potential outliers (see methods below) and only included those SNPs that could be located on the linkage map. We incorporated chromosome information in the model so that linkage disequilibrium calculations were only made in pairwise comparisons between loci on different chromosomes (the "LD locus pairing across chromosomes" option in N_eEstimator).

We determined presence of isolation by distance by testing for correlation between F_{ST} (transformed to $F_{ST}/(1 - F_{ST})$) and river distance using a partial Mantel test in R using the vegan package (Oksanen et al. 2017), accounting for run type in a dissimilarity matrix and using run type as the strata. We calculated river distance by measuring the river distance between river mouths in Google Earth. We also tested for correlations between river distance and F_{ST} within the fall and spring runs using a Mantel test.

Alignment to the *GREB1L* locus

We aligned our RAD loci to the *GREB1L* scaffold shown to be associated with premature and mature run timing in Chinook and steelhead salmon in other parts of the range (Prince et al. 2017; Thompson et al. 2019). We used BLAST+ (Camacho et al. 2009) on the default settings to make the alignments, and then looked for alignments in the regions identified by Prince et al. (2017) and Thompson et al. (2019). Using the same methods, we also aligned our loci to the *GREB1L* to *ROCK1* region identified by Narum et al. (2018) as being related to migration phenotypes.

Assignment testing

We evaluated the ability of the SNP data set to assign individuals to the differently identified groups. We first made a data set of only loci with <10% missing data. We then conducted leave-one-out tests in GENODIVE (settings: 0.005 in place of zero frequencies; likelihood ratio test statistic; alpha level of 0.001; and 500 permutations) to assess the data set's ability to correctly assign individuals to their population of origin. We tested several configurations of unique populations: (*i*) each tributary or hatchery as a unique population; (*ii*) all fall run individuals included in one population; and (*iii*) combining spring run from Mill and Deer creeks in one population, but leaving the other spring run popu-

Can. J. Fish. Aquat. Sci. Downloaded from www.nrcresearchpress.com by National Marine Mammal Lab Lib on 02/25/20
For personal use only.

Can. J. Fish. Aquat. Sci. Downloaded from www.nrcresearchpress.com by National Marine Mammal Lab Lib on 02/25/20
For personal use only.

lations as individual populations. This allowed us to evaluate what tributaries and hatcheries could be uniquely identified with this data set.

Identifying candidate loci under selection

Fall run Chinook salmon is the only run that still occurs in both the Sacramento River basin and the San Joaquin River basin. Previous work has shown that the fall run salmon are genetically indistinguishable across their range in the CV (Banks et al. 2000; Williamson and May 2005). We wanted to explore this finding further using our set of discovered SNPs. To see whether there is hidden genetic differentiation among fall run populations, we conducted an outlier test to identify loci that show signals of being under selection. Our goal was to compare the population structure found using the entire SNP data set with the population structure found among fall run salmon using just the fall run outlier SNPs.

To identify outliers, we employed the FDIST2 approach implemented in Arlequin version 3.5 (Excoffier et al. 2009; Excoffier and Lischer 2010). After running 20 000 coalescent simulations, loci that were outside the 99% quantile were identified as being candidates for selection. We then created a data set of just the loci that were identified as candidates for positive selection. We evaluated population clustering based on the set of loci under positive selection using the aforementioned DAPC methods. We aligned the outliers to the *Salmo salar* (Davidson et al. 2010; ICSASG_v2: GenBank accession No. GCA_000233375.4) and *Oncorhynchus mykiss* genomes (GenBank assembly acc. No. GCA_002163495.1), using Bowtie2 (Langmead and Salzberg 2012) and default settings, to investigate functional annotations of the outliers. The *S. salar* genome provided the most annotations, so we proceeded with this genome for annotating outliers for the spring and fall runs. We used the following alignment criteria for annotation: alignment length > 75 bp; percent identity > 90%; *e* value > 0.0001.

To further investigate the validity of our outlier results and the potential for false positives, we created a null data set of fall run individuals, randomly reorganizing individuals into populations. We randomly assigned individuals (using the random number generator function runif in R) to the same number of populations with the same number of individuals per population as the real data set. We then conducted the same analyses for detecting outliers as described previously. Given that there is no biological meaning behind our randomly created populations, we expected that there should be no outliers detected. If outliers were detected, it would indicate that the outlier detection method is prone to a high number of false positives and that the results would be viewed with less certainty.

We conducted the same outlier analyses on the spring run populations, to see whether there is any signal of selection driving differentiation among the different spring run populations, despite all extant populations being located in the Sacramento River basin.

Results

Sequencing and SNP genotyping

We obtained a mean of 248 595 544 reads per sequencing library (range: 136 417 117 – 460 215 599). The number of unique reads per individual ranged from 15 822 to 24 137 958, with a mean of 2 985 730 per individual. The mean number of genotyped loci for each individual was 18 041 (CI: 6621 – 22 969). We removed six individuals that were typed at fewer than the lower confidence interval (<6621 loci). We used the final set of 11 783 SNP loci, as described in Meek et al. (2016), for genotyping. This SNP data set excluded loci with a minor allele frequency < 0.01 and observed heterozygosity > 0.55 to remove potential paralogs. Table 1 shows the number of individuals per population that remained after also filtering out individuals that were not genotyped in at least 70% of the 11 783 SNP loci (12–30 individuals per population). The mean read depth per locus was 22.4 (95% CI: 22.3–22.6). We aligned 6666 of our loci to the consensus Chinook salmon linkage map.

Table 2. Results of analysis of molecular variance (AMOVA).

Source of variation	Nested in	% variance	*F* statistic	*p*
Within individual	—	1.006	*F*_it	—
Among individual	Population	–0.051	*F*_is	1
Among population	Run	0.005	*F*_sc	0.001
Among run	—	0.039	*F*_ct	0.002

Genetic diversity and population structure

Mean expected heterozygosity (H_e) across all populations was 0.24 (SD = 0.08), with the winter run having the lowest H_e of 0.19 and all other groups having H_e between 0.24 and 0.25 (Table 1). The inbreeding coefficient for all populations was slightly negative (range: −0.072 to −0.03), with all 95% confidence intervals below zero.

The AMOVA showed that the highest amount of between-group genetic variation is partitioned among runs, followed by among populations nested within runs (Table 2). STRUCTURE analyses showed that population structuring with six clusters (K = 6) had both the highest likelihood and the highest delta K (refer to online Supplementary Fig. S1[1] and Supplementary Table S1; Evanno et al. 2005), but the likelihoods plateaued between K = 3 and 7. The six-cluster model grouped individuals by run type, with some substructure within spring run populations (Supplementary Fig. S1[1]). Spring run in Butte Creek clustered uniquely, whereas Mill and Deer creeks clustered together. Spring run from the Feather River Hatchery showed admixture between a unique genetic lineage and the fall run cluster. All of the fall run displayed primary membership in the same unique cluster. The fall run from the Feather River Hatchery showed some admixture with the unique Feather River Hatchery spring run cluster. There is also a signature of a unique late-fall run lineage; however, it is admixed with the fall run cluster. The DAPC analysis showed similar results; however, each run and spring run population showed greater distinction and less individual admixture than in the STRUCTURE analysis (Fig. 2; Supplementary Figs. S2–S3[1]). In the DAPC analysis for K = 6, the late-fall run clustered uniquely, as did the Feather River Hatchery spring run. Additionally, in the K = 7 model, the fall run in the Coleman Hatchery displayed unique cluster membership.

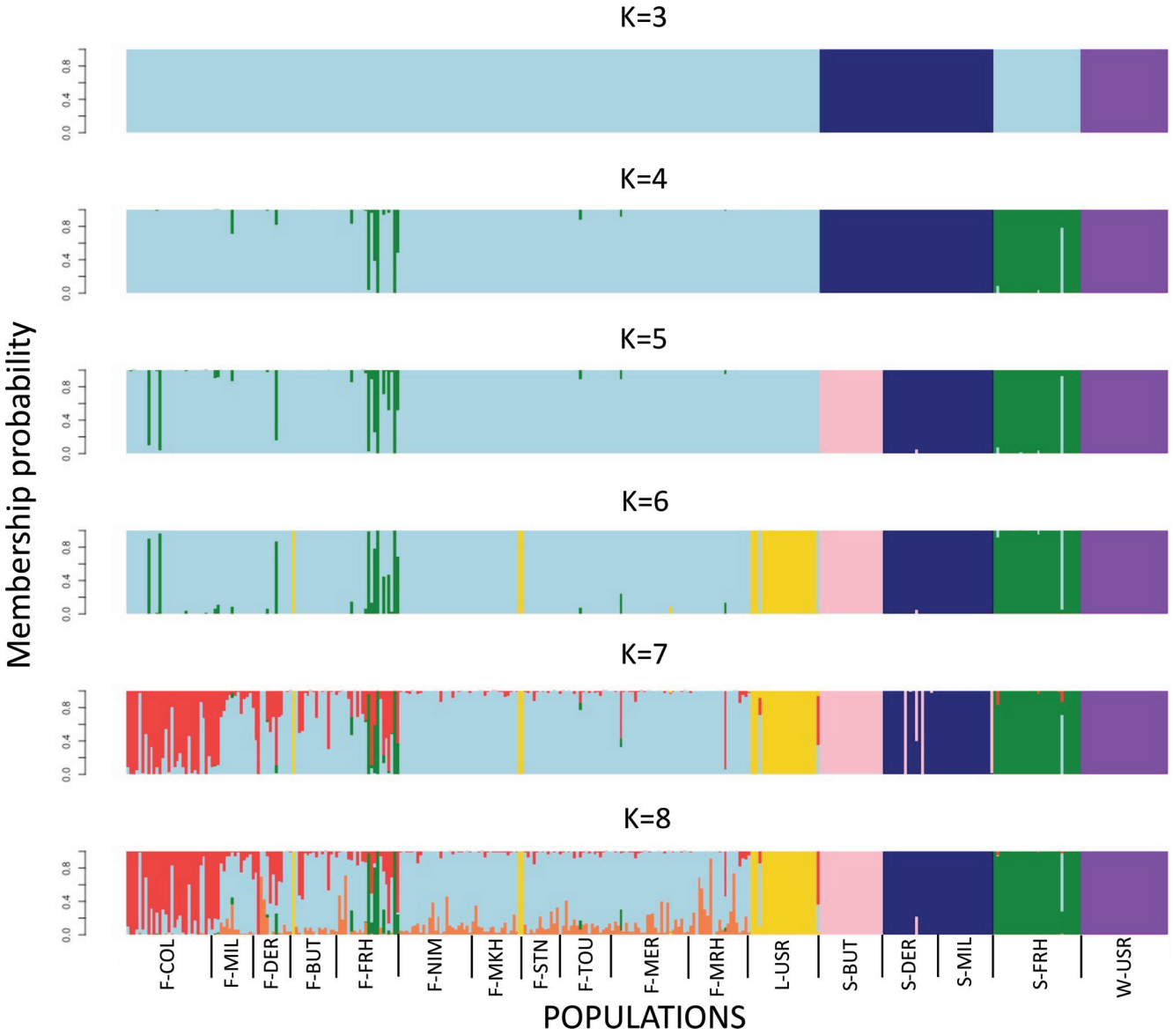
We found significant differentiation within and among the different runs, based on F_{ST} values (range: 0–0.161; Table 3). The highest values were between the winter run and all other populations (range: 0.135–0.161). Within the spring run, F_{ST} values ranged between 0.004 and 0.033, with all values being significant. F_{ST} within the fall run ranged from 0 to 0.005. The fall run from the Coleman Fish Hatchery was significantly different from all other fall run populations, as was the fall run from the Feather River Hatchery. The fall run from Mill Creek was also significantly different from the Nimbus Hatchery population and the Stanislaus and Merced rivers.

Using a data set of 4689 loci that were identified neutral markers and could be placed on the Chinook linkage map, we calculated N_e for the six groupings identified in the DAPC analysis. The confidence interval for the N_e estimate included infinity for all groups, with the exception of the spring run from Mill and Deer

[1]Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/cjfas-2019-0171.

Can. J. Fish. Aquat. Sci. Downloaded from www.nrcresearchpress.com by National Marine Mammal Lab Lib on 02/25/20
For personal use only.

Fig. 2. Discriminant analysis of principal components (DAPC) clustering results, testing different number of groups (K = 3–8) using the full SNP data set. Each bar represents an individual, and the x axis shows sampled location (location abbreviations defined in Table 1). The colour distinguishes clusters, and the y axis quantifies cluster membership.



creeks and the winter run. The N_e estimate for the spring run from Mill and Deer creeks was 591.1 (CI: 558.3–628.0) and for the winter run was 376.4 (CI: 352.6–403.5). Estimates of infinity for N_e are likely due to sampling error outweighing the effects of genetic drift (i.e., sample size too small to capture the effects of genetic drift in large populations; see Waples and Do (2010) for further discussion of this phenomenon).

There was significant isolation by distance when all runs were used in the model and run is accounted for in the partial Mantel test (p = 0.007). Isolation by distance was not significant when limited to only the spring run (p = 0.125), but was highly significant within fall run populations (p = 0.007; Supplementary Fig. S4[1]).

Alignment to *GREB1L* loci

We found one SNP in our final data set that aligned with the run timing associated *GREB1L* locus. This locus (SNP No. R008612) aligned to position 595 079 – 595 161 of the scaffold 79929e described in Prince et al. (2017). The genotypes at this locus were nearly perfectly associated with premature (spring and winter run) and mature (fall and late fall) run timings (Fig. 3). Interestingly, all spring run individuals were homozygous for one allele, with the exception of six individuals from the Feather River Hatchery (representing 7.6% of the spring run), which were heterozygous. Within the fall run, 90% (170 individuals) were homozygous for the alternate allele, with 8.5% (16 individuals) being heterozygous, and 1.6% (three individuals) being homozygous for the spring–winter associated allele. Two of the individuals that were homozygous for the spring–winter allele were from the Feather River Hatchery and one was from Mill Creek. We found 87 SNP loci that aligned to the 203 kb region containing the *GREB1L* and *ROCK1* genes identified by Narum et al. (2018) on Ots28 between 11.022 and 11.225 Mb. We calculated the allele frequency differences for each run at these 87 loci to evaluate the patterns among the different runs (Fig. 4). The same locus identified by aligning to the Prince et al. (2017) scaffold (SNP No. R008612) aligned to this region and clearly shows differences among the premature and mature run timings. No other locus showed as clear a pattern.

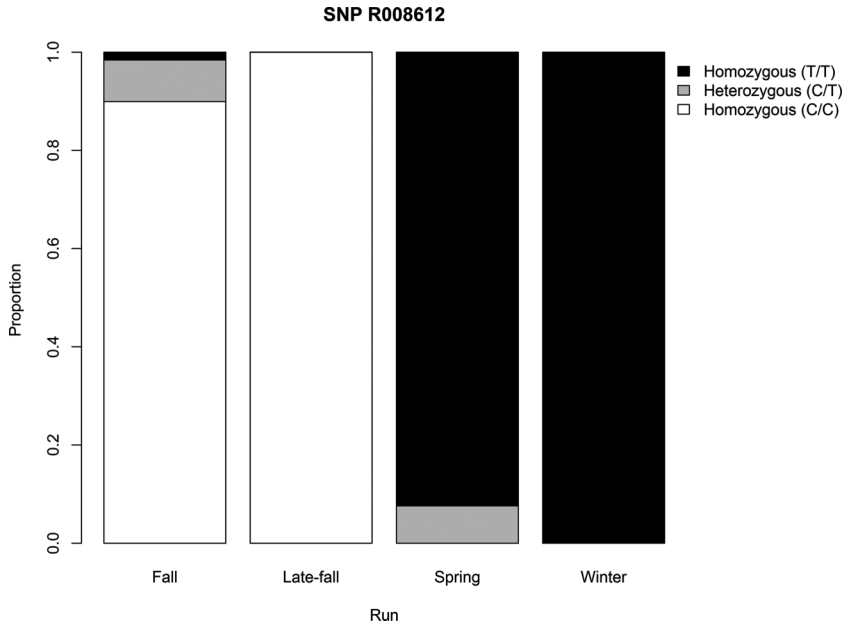
Can. J. Fish. Aquat. Sci. Downloaded from www.nrcresearchpress.com by National Marine Mammal Lab Lib on 02/25/20
For personal use only.

Table 3. F_{ST} values among populations using full, filtered data set.

	F_COL	F_MIL	F_DER	F_BUT	F_FRH	F_NIM	F_MKH	F_STN	F_TOU	F_MER	F_MRH	L_USR	S_BUT	S_DER	S_MIL	S_FRH	W_USR
F_COL	—	0.002	0.009	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
F_MIL	**0.002**	—	0.441	0.309	0.013	0.002	0.039	0.007	0.038	0.002	0.220	0.001	0.001	0.001	0.001	0.001	0.001
F_DER	**0.002**	0	—	0.864	0.010	0.022	0.112	0.037	0.033	0.024	0.106	0.001	0.001	0.001	0.001	0.001	0.001
F_BUT	**0.003**	0	–0.001	—	0.299	0.306	0.589	0.977	0.528	0.534	0.489	0.001	0.001	0.001	0.001	0.001	0.001
F_FRH	**0.003**	**0.001**	**0.001**	0	—	0.001	0.001	0.001	0.003	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
F_NIM	**0.004**	**0.002**	0.001	0	**0.002**	—	0.713	0.558	0.231	0.234	0.029	0.001	0.001	0.001	0.001	0.001	0.001
F_MKH	**0.004**	0.001	0.001	0	**0.002**	0	—	0.524	0.370	0.442	0.038	0.001	0.001	0.001	0.001	0.001	0.001
F_STN	**0.004**	**0.002**	0.002	–0.002	**0.003**	0	0	—	0.703	0.593	0.276	0.001	0.001	0.001	0.001	0.001	0.001
F_TOU	**0.004**	0.001	0.001	0	**0.002**	0	0	0	—	0.231	0.160	0.001	0.001	0.001	0.001	0.001	0.001
F_MER	**0.004**	**0.002**	0.001	0	**0.002**	0	0	0	0	—	0.037	0.001	0.001	0.001	0.001	0.001	0.001
F_MRH	**0.005**	0.001	0.001	0	**0.003**	0.001	0.001	0.001	0.001	0.001	—	0.001	0.001	0.001	0.001	0.001	0.001
L_USR	**0.009**	**0.007**	**0.007**	**0.006**	**0.009**	**0.008**	**0.007**	**0.007**	**0.008**	**0.007**	**0.008**	—	0.001	0.001	0.001	0.001	0.001
S_BUT	**0.037**	**0.035**	**0.038**	**0.038**	**0.035**	**0.041**	**0.040**	**0.041**	**0.040**	**0.040**	**0.041**	**0.042**	—	0.001	0.001	0.001	0.001
S_DER	**0.020**	**0.019**	**0.020**	**0.021**	**0.018**	**0.024**	**0.023**	**0.024**	**0.023**	**0.023**	**0.024**	**0.025**	**0.017**	—	0.002	0.001	0.001
S_MIL	**0.021**	**0.019**	**0.021**	**0.021**	**0.019**	**0.024**	**0.024**	**0.024**	**0.024**	**0.023**	**0.024**	**0.026**	**0.023**	**0.004**	—	0.001	0.001
S_FRH	**0.010**	**0.007**	**0.008**	**0.009**	**0.005**	**0.012**	**0.011**	**0.012**	**0.010**	**0.010**	**0.011**	**0.015**	**0.033**	**0.019**	**0.019**	—	0.001
W_USR	**0.140**	**0.148**	**0.152**	**0.150**	**0.145**	**0.152**	**0.153**	**0.161**	**0.154**	**0.150**	**0.154**	**0.151**	**0.151**	**0.140**	**0.146**	**0.135**	—

Note: Values below the diagonal are F_{ST}. Values above the diagonal are p values. Bold F_{ST} values are significant at $p < 0.015$.

Fig. 3. Proportion of individuals showing each genotype at the *GREB1L* associated SNP, by run timing.



Assignment testing

Assignment testing showed a high probability of correct assignment using the data set of loci with <10% missing data (7829 loci). When fall run populations were grouped together in one general "fall run" reporting group, assignment probabilities were high for the fall run (100%), but the fall run also drew a lot of non-fall run assignments (e.g., individuals from the late-fall and spring runs). The vast majority (96%) of the late-fall run assigned to the fall run population using this data set and assignment of spring run populations ranged from 40% to 100% (Supplementary Table S2[1]). However, when we used each individual fall run population as a unique reporting group, assignment for individual fall run populations ranged wildly (0%–100%), although 95% of the assignments for fall run individuals were assigned to a fall run population. Notably, the fall run from the Merced River, Coleman Hatchery, and Nimbus Hatchery drew almost all assignments from other fall run drainages when using individual reporting groups (Supplementary Table S3[1]). Based on these data, we cannot distinguish individuals who had genetic misassignments to the wrong natal stream from those that were assigned correctly but returned to non-natal streams to spawn (strayed), making them appear to be misassigned. Notably, we achieved 100% accuracy of assignment for the late-fall run when using individual fall run reporting groups. Assignment accuracy was 100% when we combined the spring run from Mill and Deer creeks into a single group (Fig. 5). These data show that using the full SNP data set enables high assignment accuracy of populations.

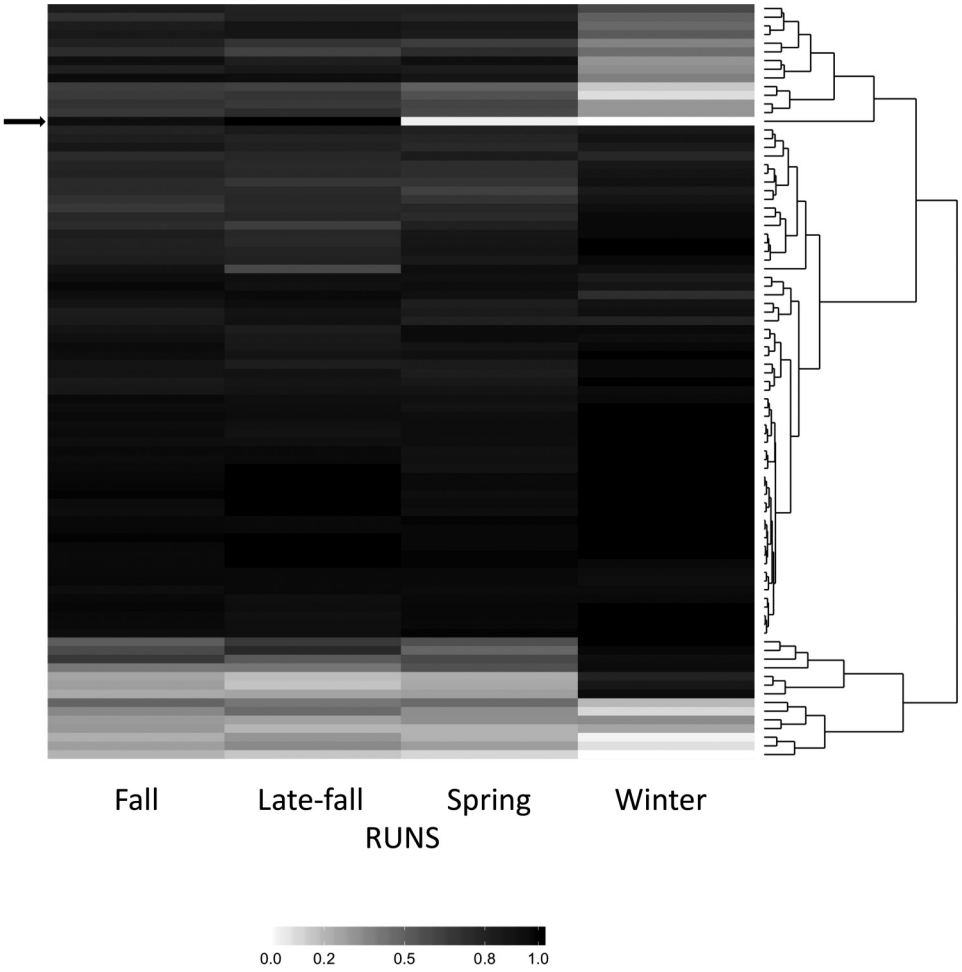
Loci under selection

Outlier analyses performed within the fall run found 829 loci that were significant for showing signatures of diversifying selection using Arlequin. A total of 654 of the outliers aligned to the Chinook salmon linkage map, with the loci spread throughout the genome (between 10 and 38 outliers aligned to each chromosome) (Fig. 6). Twenty-three loci mapped to linkage groups but were not positioned on a chromosome. Forty-five outliers had functional annotations associated with them (Table S4[1]). We found zero outliers in the randomized data set, providing reassurance that our results from analysis of the real data set are meaningful.

The DAPC model that used only these loci and retained 70 PCs found that K = 1–3 had the lowest Bayesian information criterion

Can. J. Fish. Aquat. Sci. Downloaded from www.nrcresearchpress.com by National Marine Mammal Lab Lib on 02/25/20
For personal use only.

Fig. 4. Heatmap of allele frequencies for 87 SNPs aligned to the *GREB1L* to *ROCK1* region of *Omy28* identified in Narum et al. (2018). Each row is one of the 87 SNPs. The arrow points to SNP R008612 shown in Fig. 3, which aligns to the *GREB1L* region identified in Prince et al. (2017). The scale bar denotes the frequency of the first allele at each locus.



(BIC), with K = 4 and 5 slightly higher but with clustering that shows biological relevance (Fig. 7; Supplementary Fig. S5[1]). At K = 2, a unique cluster was formed by several individuals from Mill, Deer, and Butte creeks, the Tuolumne River, and Merced River hatcheries, and one individual from the Merced River population. When the spring and winter run populations are also included, it becomes clear that this unique cluster is formed with the spring run from the Feather River Hatchery (data not shown). Therefore, it is very likely that this cluster is driven by introgression with strays from the spring run in the Feather River Hatchery, rather than being a unique fall run lineage. The fall run from Coleman Hatchery showed unique cluster membership starting at K = 3. At K = 4, two additional groups emerged, with Deer, Butte, and Mill creeks clustering together along with the Merced River and Merced River Hatchery populations, and the Nimbus and Mokelumne hatcheries and Stanislaus and Tuolumne rivers clustering together. At K = 5, the Merced River population was resolved as its own unique cluster, with some individuals in Butte Creek drawing membership from this cluster. These clustering patterns show a general geographic pattern of clustering, with many of the wild Sacramento River basin populations clustering together and many of the San Joaquin River basin populations joining another cluster.

The outlier analyses within the spring run found 940 loci that were significant for signatures of diversifying selection. A total of 776 of the outliers aligned to the Chinook salmon linkage map, with between 12 and 46 outliers aligned to each chromosome. Twenty-nine loci mapped to linkage groups but were not positioned on a chromosome. Fifty-seven outlier loci had annotations associated with them (Supplementary Table S4[1]). No outliers were detected when analyzing the randomized data set.

The DAPC model that used just these loci found K = 3–4 had the lowest BIC (Fig. 8; Supplementary Fig. S6[1]). At K = 3, Deer and Mill creeks grouped together, while at K = 4, each tributary population grouped separately. Overall, 101 loci were identified as being potential outliers in both the fall and spring run analyses.

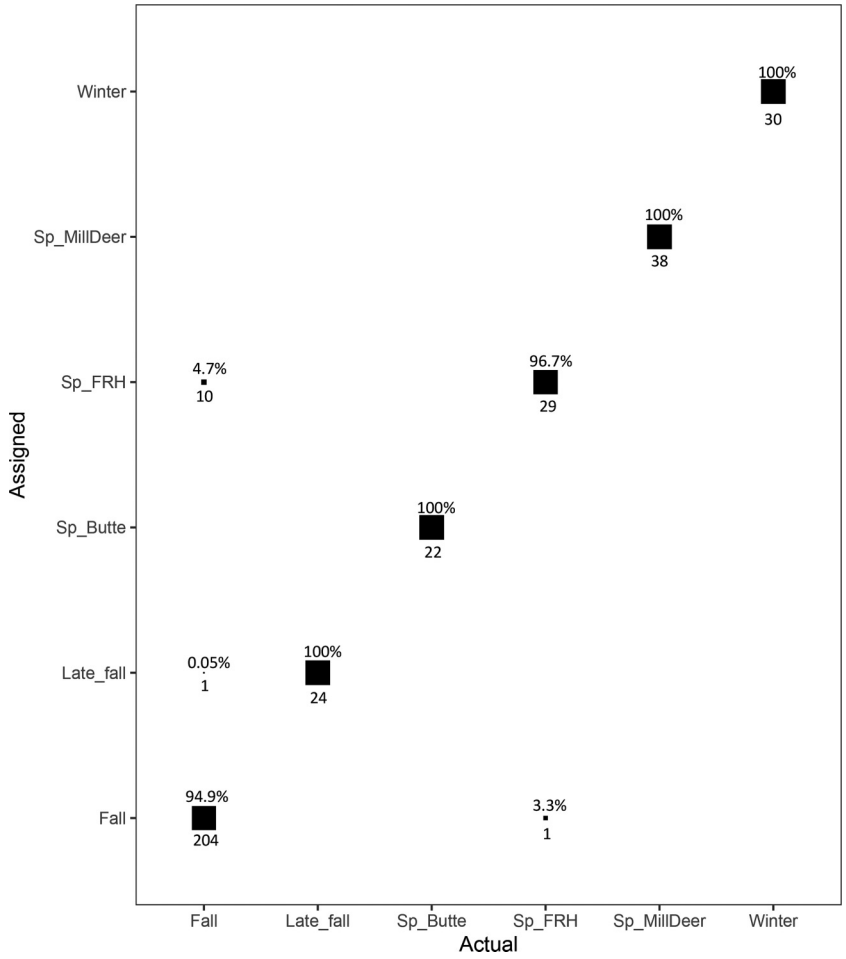
F_{ST} values were much higher among populations when using the outlier data set, and all pairwise comparisons were significant (fall run: mean = 0.035, range = 0.008–0.061; spring run: mean = 0.102, range = 0.036–0.166; Supplementary Tables S5 and S6[1]).

Discussion

Chinook salmon exhibit some of the highest genetic and life history trait diversity of all Pacific salmonids (Waples 2001). In particular, the presence of four distinct spawning runs in the Central Valley of California represents the greatest life history trait diversity observed for Chinook salmon, and presumably corresponds with the highest biocomplexity and portfolio for this species. In reality, CV Chinook salmon display the weakest portfolio in the species range (Griffiths et al. 2014); however, our results indicate that there is greater genetic biocomplexity than previously described. Protection of this remaining biocomplexity

Can. J. Fish. Aquat. Sci. Downloaded from www.nrcresearchpress.com by National Marine Mammal Lab Lib on 02/25/20
For personal use only.

Fig. 5. Assignment accuracies of each run using 7892 SNPs with <10% missing data and individual fall run reporting groups. Points are scaled to the assignment value. The value on top of each point is the percentage assigned to each group, while the value under each point is the number of individuals that were assigned to each group.



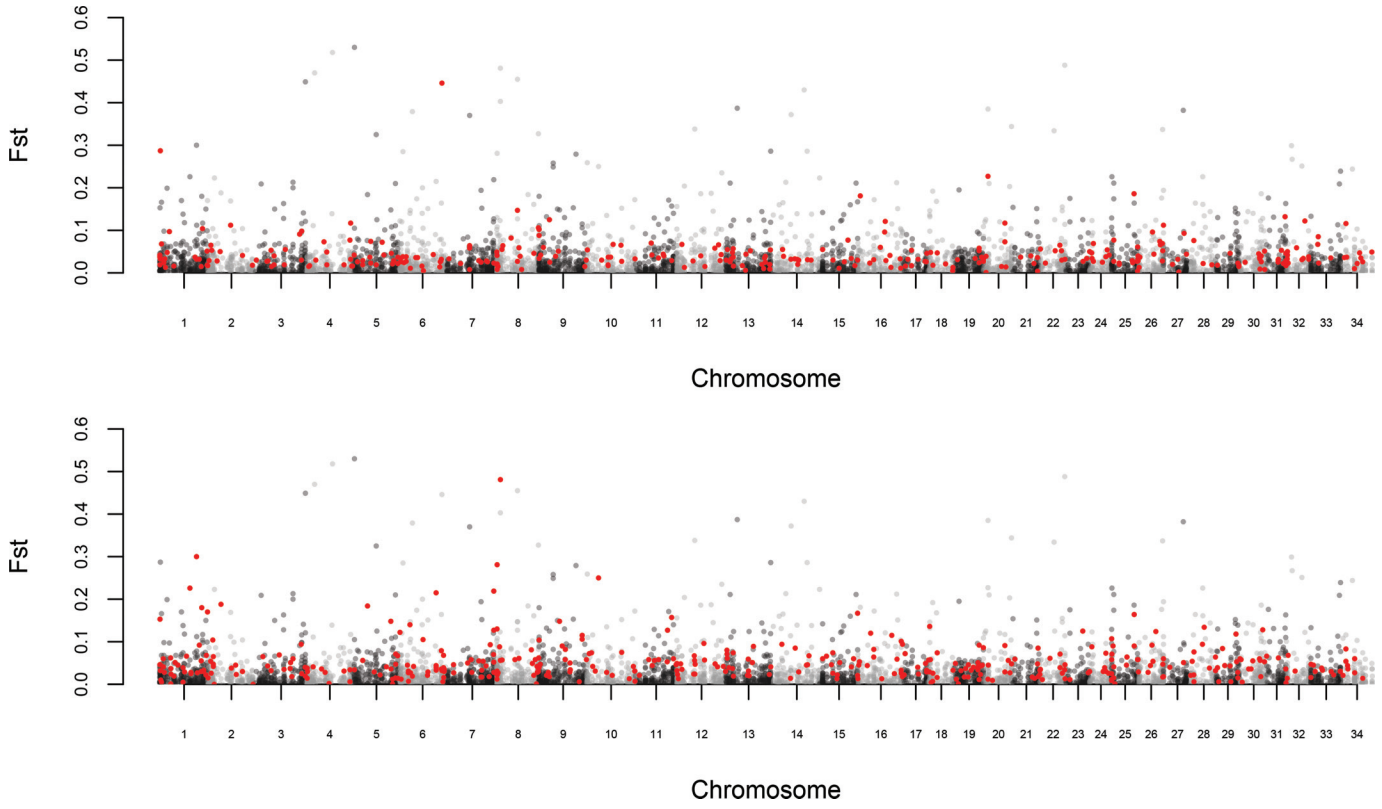
may be important for sustainable management of the CV Chinook salmon stock complex and restoration of the portfolio.

This study is the first genomic investigation into the population structure and genomic diversity of CV Chinook salmon. By genotyping hundreds of individuals at thousands of SNPs distributed across the genome, we found evidence for greater population structuring beyond the level of run, including fine-scale population structure for spring run fish and a pattern of isolation-by-distance among fall run populations. We also found genetic distinctiveness of fall and late-fall runs, despite conclusions from previous scientific studies and current management that defined them as a single ESU (Lindley et al. 2004). Our results provide managers with information necessary to not only conserve the diversity of runs in the CV, but also maintain and enhance life history diversity within runs for maximal buffering from environmental variation and overall portfolio performance (Satterthwaite et al. 2017).

We show that the spring run contains finer-scale substructure, with evidence for three distinct populations: (*i*) Mill and Deer creeks, (*ii*) Butte Creek, and (*iii*) the spring run in the Feather River Hatchery. This latter finding is quite remarkable, as previous work has shown that the spring run in the Feather River Hatchery is genetically indistinguishable from the fall run (Lindley et al. 2004; Garza et al. 2008). Past hatchery practices include inadequate separation of the brood stock for the fall and spring runs, leading to introgression between the two runs, and the prevailing wisdom has been that these runs are genetically indistinguishable (California Hatchery Scientific Review Group 2012). O'Malley et al. (2007, 2013) found no differentiation between the fall and spring runs in the Feather River Hatchery based on nine neutral microsatellite loci, but they did find clear distinction at three circadian clock genes thought to control run timing. Our results show that if the full genome is surveyed, additional distinctions between the Feather River Hatchery fall and spring runs are observed. We do see evidence of past introgression at the *GREB1L* associated locus, with the spring run from the Feather River Hatchery being the only spring run population that contained heterozygous individuals. These results show that some of the genetic diversity found in the Feather River Hatchery spring run is unique and has not been eliminated from introgression with the fall run. Additionally, these results support the use of revised hatchery practices that aim to better separate the two runs in the hatchery (Baerwald et al. 2011).

Our results demonstrate that with the full data set of ~12 000 SNPs, not only can we distinguish the spring run in the Feather River Hatchery from the other spring run populations and from the fall run, but we can also distinguish the late-fall run from the fall run populations. Our demonstrated ability to distinguish individuals from these runs is a major leap forward in our ability to monitor and manage these populations separately, enabling protection of biocomplexity in the system. It is vital that all of the genetic diversity present in this system is protected to maintain a healthy portfolio of life histories and genetic diversity (Schindler et al. 2010; Carlson and Satterthwaite 2011).

Fig. 6. Manhattan plot of outliers of SNPs mapped to the Chinook linkage map (McKinney et al. 2016). Each point represents an SNP. The *x* axis shows the position of the SNP on the linkage map. Alternating chromosomes are shaded in black or grey. The top panel has the fall run outlier loci highlighted in red, while the bottom panel has the spring run outlier loci highlighted in red. [Colour online.]



Recently, a high-throughput CV-specific 80 SNP panel was developed that goes a long way toward improving our ability to distinguish the different runs (Meek et al. 2016). The benefit of this panel is that it is available as a Fluidigm SNP Type Assay (Fluidigm Corp., San Francisco, California, USA) allowing very fast turnaround from tissue sample to genotype and run assignment. This fast turnaround is required for some management questions (e.g., real-time take determinations). The trade-off, however, is that this SNP panel is unable to distinguish the spring run in the Feather River Hatchery from the fall run populations and it has lower assignment accuracies. We show that, with almost 12 000 SNPs, we can distinguish the different populations with high accuracy. Therefore, when run assignment is not needed for real-time analyses, genotyping individuals via a sequence-based approach will allow finer-scale and more accurate management of the different populations (Meek and Larson 2019). Our work also shows that by looking at the *GREB1L* associated locus, we can easily distinguish the ESA listed (spring and winter run) from the unlisted (fall and late fall) individuals. Future work will involve exploring this *GREB1L* locus further and developing these loci into Rapture baits (Ali et al. 2016), to allow very high throughput and cost-effective genotyping via sequencing of thousands of samples.

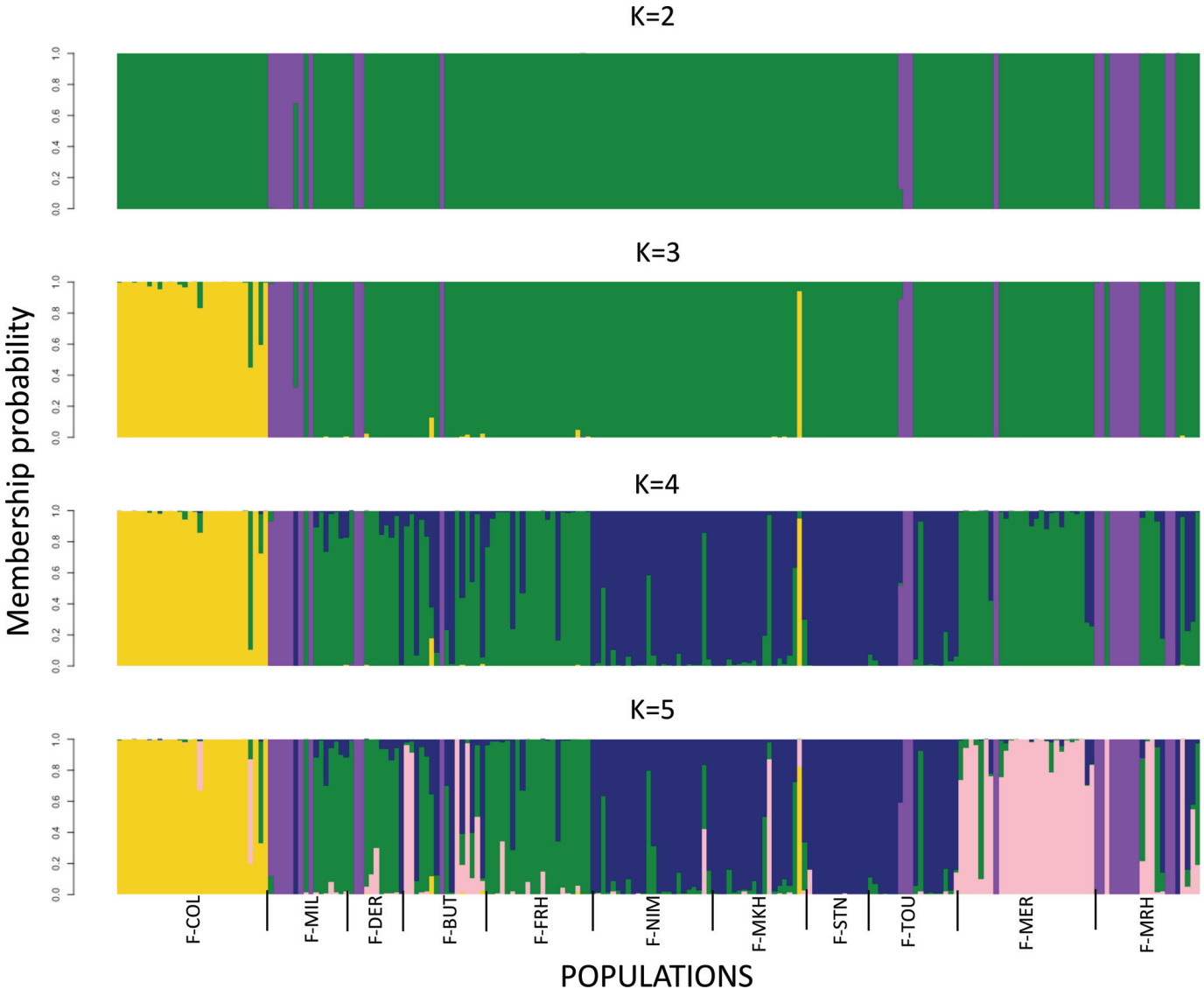
Our work also shows that the winter run is the most genetically distinct population in the CV, but it also has the lowest genetic diversity (e.g., observed and expected heterozygosity; see Table 1). We show that the winter run has a low effective population size (N_e = 376) and may be experiencing high levels of genetic drift associated with decreased population sizes. Frankham et al. (2014) suggest that an N_e < 500 is too low to retain evolutionary potential in perpetuity, and that a target of N_e = 1000 for wild populations is more appropriate. Following this guideline, the spring run from the Mill–Deer creeks complex is also at risk of reduced long-term viability given that N_e = 591. These results are not surprising, especially given the endangered and threatened status of the winter and spring runs, but it highlights the importance of management efforts aimed at protecting the portfolio of diversity present in these runs specifically and in CV Chinook as a whole.

Dispersal among populations is one of the main mechanisms for increased synchrony among populations and weakening of the portfolio effect (Liebhold et al. 2004). Human activities, such as artificial production, can increase dispersal among populations, decreasing locally adapted life history variation and increasing synchrony (McClure et al. 2008; Moore et al. 2010). It is estimated that straying in hatchery-reared Chinook can be >70% (California Hatchery Scientific Review Group 2012; Palmer-Zwahlen and Kormos 2015). This is a drastic increase compared with the often <5% straying that occurs in wild salmonid populations (Quinn 2005). It is thought that this increased straying owing to hatchery production and the lack of migration path imprinting from trucking has led to the genetic homogenization and decreased portfolio effect among the fall run in the CV (Huber and Carlson 2015; Satterthwaite and Carlson 2015). In fact, Dedrick and Baskett (2018) found that genetic homogenization in the CV fall run plays a larger role in decreasing the portfolio effect than demographic synchronization. The role of trucking and increased straying could be playing a large role in the genetic homogenization we see in the fall run. Our work here, however, shows that despite increased straying, there are still some remaining genetic differences among populations. This is evident in the structuring we see with the putative loci under selection. This work also demonstrates the importance of altering hatchery and management practices now to protect extant diversity while it still exists in the system. Maintenance of biocomplexity in a system can counteract homogenizing forces and promote population asynchrony (Moore et al. 2014).

Can. J. Fish. Aquat. Sci. Downloaded from www.nrcresearchpress.com by National Marine Mammal Lab Lib on 02/25/20
For personal use only.

Can. J. Fish. Aquat. Sci. Downloaded from www.nrcresearchpress.com by National Marine Mammal Lab Lib on 02/25/20
For personal use only.

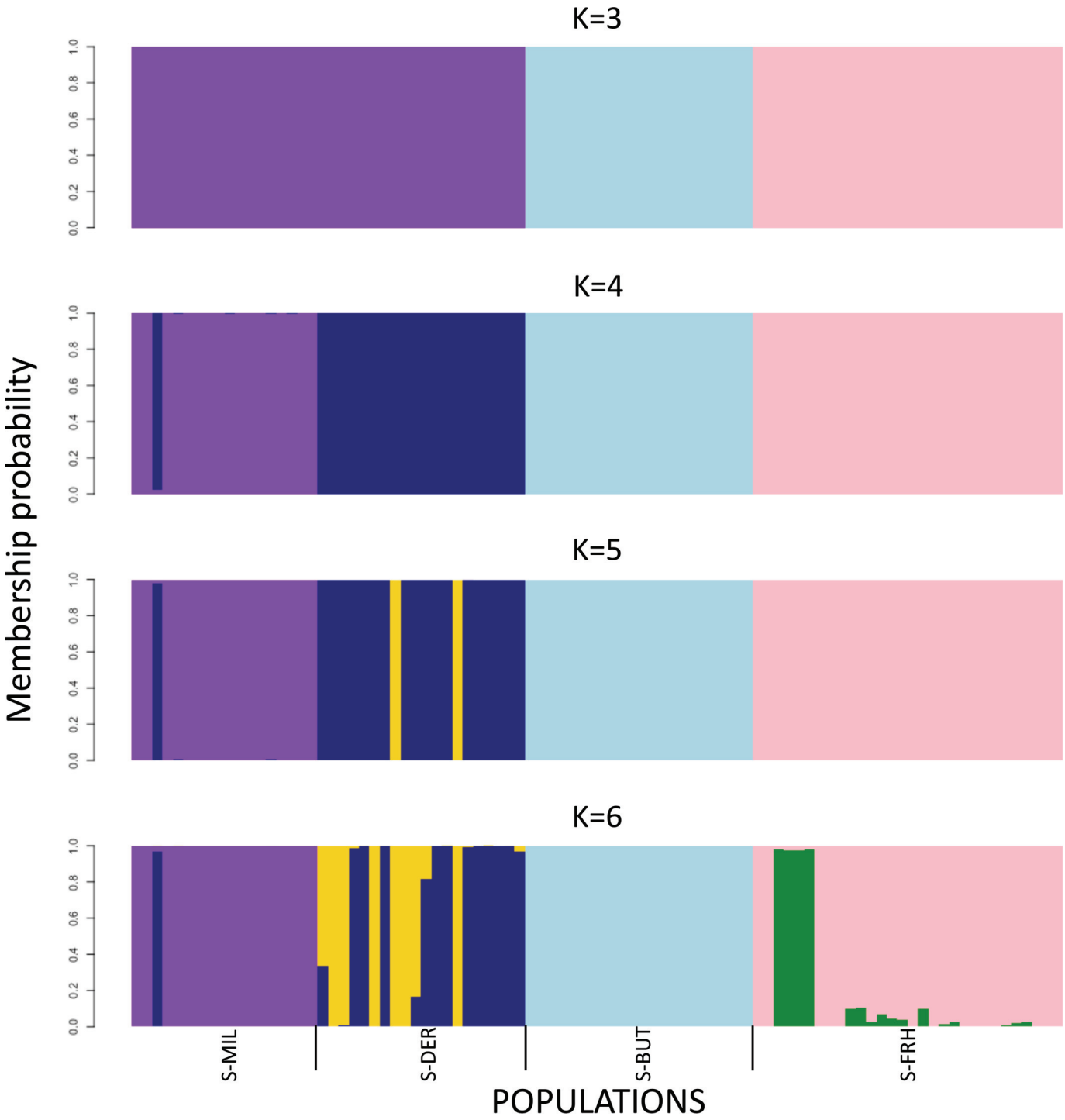
Fig. 7. Discriminant analysis of principal components (DAPC) clustering results for fall run populations using 852 candidate outlier loci for diversifying selection. Each bar represents an individual. The *x* axis shows the location sampled (location abbreviations defined in Table 1). The colour distinguishes clusters, and the *y* axis quantifies cluster membership.



Our work shows how insights about population differentiation can be gained by evaluating genomic loci presumably under selection. When we evaluated population structure using loci that are only candidates for selection, we observed finer-scale structuring of both the spring and fall runs than we observed using the full data set. All the spring run populations show differentiation, and differences among groups of the fall run populations are apparent. The fall run population in the Coleman Hatchery is clearly distinct from the rest of the fall run based on the outlier data set, suggesting that some aspect of Coleman Hatchery practices may be driving differentiation. Additionally, the fall run from the Nimbus Hatchery, Mokolumne Hatchery, Stanislaus River, and Tuolumne River also separate into one group that is differentiated from the rest of the fall run populations. It is possible that fish in these tributaries are facing similar selection pressures driving this pattern. At $K = 5$, we also start to see the fall run spawning in the Merced River appear as distinct from the rest of fall run. Interestingly, the fall run spawning in the Merced River Hatchery show very mixed cluster membership, with representation from other clusters found in the CV. It is possible that there is unique selection pressure experienced by fish spawning in the river habitat versus in the hatchery that drives this difference (Vasemägi et al. 2016), or it could be that the Merced River Hatchery receives a lot of strays from other tributaries. Indeed, a 2012 study of hatchery returns found only 12.7% of returns to the Merced River Hatchery were from Merced River Hatchery raised fish (Palmer-Zwahlen and Kormos 2015).

There is some debate regarding the use of adaptive loci in conservation planning. Allendorf et al. (2010) caution against focusing on adaptive loci for conservation, stating that loci identified as being adaptive might not be the loci crucial for future adaptation. Additionally, the authors warn that focusing on detectable adaptive differences may result in overlooking, and potentially losing, important genetic diversity in other regions of the genome. Funk et al. (2012), however, highlight the importance of using information from both neutral and non-neutral genetic data to identify an appropriate management approach for species. We agree with both Funk et al. and Allendorf et al. and believe there is value in incorporating both neutral and non-neutral genetic information in conservation strategies. We believe that both these cautions highlight the need to conserve as much genetic diversity as is feasible, because it is very difficult to predict what diversity will be

Fig. 8. Discriminant analysis of principal components (DAPC) clustering results for spring run populations using 940 candidate outlier loci for diversifying selection. Each bar represents an individual. The *x* axis shows the location sampled (location abbreviations defined in Table 1). The colour distinguishes clusters, and the *y* axis quantifies cluster membership.



"important" both now and in the future. We believe this to be particularly true for migratory species. Because they are subject to selective forces across the migratory pathway, it can be very difficult to decipher what is driving selective differences, and selective forces will change over time (Dingle 2014). For example, the differences we see in selected loci among CV Chinook salmon populations may be due to selection differences among the spawning grounds, seasonal differences among migration life histories, or differences in selective forces in separate regions of the ocean. Additionally, our study is not an exhaustive study of all possible forms of biocomplexity in this system. For example, we know that there is also diversity in timing of juvenile outmigration that is important for biocomplexity. Our study highlights the importance of identifying and protecting as much diversity as possible so there is a full portfolio of genetic diversity present in the system.

In conclusion, these results have strong implications for the management of CV Chinook salmon and provide important les-

Can. J. Fish. Aquat. Sci. Downloaded from www.nrcresearchpress.com by National Marine Mammal Lab Lib on 02/25/20
For personal use only.

Can. J. Fish. Aquat. Sci. Downloaded from www.nrcresearchpress.com by National Marine Mammal Lab Lib on 02/25/20
For personal use only.

sons for using genomics to identify biocomplexity and apply resulting inferences to the conservation of migratory species. Genomics is an increasingly important tool for managing species. It provides the ability to identify biological diversity present in a system, and to assign individuals to a population of origin despite spatiotemporal mixing, such as on migratory pathways. It is vital that we improve our ability to identify and protect biocomplexity in management and conservation so that we can improve species and metapopulation stability and resilience.

Data archiving statement

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

Acknowledgements

We thank Morten Limborg and Wes Larson for helpful discussions that improved the manuscript and Nadya Mamoozadeh and anonymous reviewers for suggestions and edits. We thank Mike Miller and Dan Prince for sharing their data and insights on the *GREB1L* locus. We also thank Lea Koerber and the California Department of Fish and Wildlife Tissue Archive for supplying Chinook tissue samples. All tissues were held under the authorization of ESA Section 10(a)(1)(A) permit (Pacific fish) No. 15926, a Memorandum of Understanding with the California Department of Fish and Wildlife, and California Department of Fish and Wildlife Scientific Collecting Permits Nos. 8871, 3561, 8862, 11381, and 1751. This work was made possible by funding from the California Department of Fish and Wildlife contract No. P0740017. M.H.M. was supported by the David H. Smith Conservation Research Fellowship Program. This work used the Vincent J. Coates Genomics Sequencing Laboratory at University of California Berkeley, supported by the National Institutes of Health S10 Instrumentation Grants S10RR029668 and S10RR027303.

References

Ali, O.A., O'Rourke, S.M., Amish, S.J., Meek, M.H., Luikart, G., Jeffres, C., and Miller, M.R. 2016. Rad capture (Rapture): Flexible and efficient sequence-based genotyping. Genetics, **202**: 389–400. doi:10.1534/genetics.115.183665. PMID:26715661.

Allendorf, F.W., Hohenlohe, P.A., and Luikart, G. 2010. Genomics and the future of conservation genetics. Nat. Rev. Genet. **11**: 697–709. doi:10.1038/nrg2844. PMID:20847747.

Angeloni, F., Wagemaker, N., Vergeer, P., and Ouborg, J. 2012. Genomic toolboxes for conservation biologists. Evol. Appl. **5**: 130–143. doi:10.1111/j.1752-4571.2011.00217.x. PMID:25568036.

Araki, H., Cooper, B., and Blouin, M.S. 2007. Genetic effects of captive breeding cause a rapid, cumulative fitness decline in the wild. Science, **318**: 100–103. doi:10.1126/science.1145621. PMID:17916734.

Baerwald, M.R., Stephens, M.R., Bork, K.S., Meek, M.H., Tomalty, K., and May, B.P. 2011. Spring-run Chinook salmon genetic management plan–San Joaquin River Restoration Program.

Baird, N., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., et al. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE, **3**: e3376. doi:10.1371/journal.pone.0003376. PMID:18852878.

Banks, M.A., Rashbrook, V.K., Calavetta, M.J., Dean, C., and Hedgecock, D. 2000. Analysis of microsatellite DNA resolves genetic structure and diversity of chinook salmon (*Oncorhynchus tshawytscha*) in California's Central Valley. Can. J. Fish. Aquat. Sci. **57**: 915–927. doi:10.1139/f00-034.

Beacham, T., Jonsen, K., Supernault, J., Wetklo, M., Deng, L., and Varnavskaya, N. 2006. Pacific rim population structure of Chinook salmon as determined from microsatellite Analysis. Trans. Am. Fish. Soc. **135**: 1604–1621. doi:10.1577/T06-071.1.

Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., and Bernatchez, L. 2015. RAD-genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species; the American lobster (*Homarus americanus*). Mol. Ecol. **24**: 3299–3315. doi:10.1111/mec.13245.

Benjamini, Y., and Yekutieli, D. 2001. The control of false discovery rate under dependency. Ann. Stat. **29**: 1165–1188. doi:10.1214/aos/1013699998.

California HSRG. 2012. California hatchery review report. California Hatchery Scientific Review Group. Prepared for the US Fish and Wildlife Service and Pacific States Marine Fisheries Commission.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. 2009. BLAST+: architecture and applications. BMC Bioinf. **10**: 421. doi:10.1186/1471-2105-10-421.

Carlson, S.M., and Satterthwaite, W.H. 2011. Weakened portfolio effect in a collapsed salmon population complex. Can. J. Fish. Aquat. Sci. **68**: 1579–1589. doi:10.1139/f2011-084.

Clemento, A.J., Crandall, E.D., Garza, J.C., and Anderson, E.C. 2014. Evaluation of a single nucleotide polymorphism baseline for genetic stock identification of Chinook Salmon (*Oncorhynchus tshawytscha*) in the California Current large marine ecosystem. Fish. Bull. **112**: 112–130. doi:10.7755/FB.112.2-3.2.

Davidson, W.S., Koop, B.F., Jones, S.J.M., Iturra, P., Vidal, R., Maass, A., et al. 2010. Sequencing the genome of the Atlantic salmon (*Salmo salar*). Genome Biol. **11**: 403. doi:10.1186/gb-2010-11-9-403. PMID:20887641.

Dedrick, A.G., and Baskett, M.L. 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. PMID:30016162.

Dingle, H. 2014. Migration: the biology of life on the move. Second Edition. Oxford University Press, New York.

Do, C., Waples, R.S., Peel, D., Macbeth, G.M., Tillett, B.J., and Ovenden, J.R. 2014. NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. Mol. Ecol. Resour. **14**: 209–214.

Earl, D.A., and VonHoldt, B.M. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. **4**: 359–361. doi:10.1007/s12686-011-9548-7.

Etter, P.D., Bassham, S., Hohenlohe, P.A., Johnson, E., and Cresko, W.A. 2011. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *In* Molecular Methods for Evolutionary Genetics. *Edited by* V. Orgogozo and M. Rockman. Humana Press, New York. pp. 157–178.

Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. **14**: 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x. PMID:15969739.

Excoffier, L., and Lischer, H. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. **10**: 564–567. doi:10.1111/j.1755-0998.2010.02847.x. PMID:21565059.

Excoffier, L., Hofer, T., and Foll, M. 2009. Detecting loci under selection in a hierarchically structured population. Heredity, **103**: 285–298. doi:10.1038/hdy.2009.74. PMID:19623208.

Federal, Register. 1999. Endangered and threatened species: Threatened status for two Chinook salmon evolutionarily significant units (ESUs) in California. Office of the Federal Register, National Archives and Records Administration, **64**: 50394–50415.

Federal Register. 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. Office of the Federal Register, National Archives and Records Administration, **70**: 37160–37204.

Frankham, R., Bradshaw, C.J.A., and Brook, B.W. 2014. Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. Biol. Conserv. **170**: 56–63.

Funk, W.C., McKay, J.K., Hohenlohe, P.A., and Allendorf, F.W. 2012. Harnessing genomics for delineating conservation units. Trends Ecol. Evol. **27**: 489–496. doi:10.1016/j.tree.2012.05.012. PMID:22727017.

Garza, J.C., Blankenship, S.M., Lemaire, C., and Charrier, G. 2008. Genetic population structure of Chinook salmon (*Oncorhynchus tshawytscha*) in California's Central Valley. Final report for CalFed Project "Comprehensive evaluation of population structure and diversity for Central Valley Chinook salmon".

Griffiths, J.R., Schindler, D.E., Armstrong, J.B., Scheuerell, M.D., Whited, D.C., Clark, R.A., et al. 2014. Performance of salmon fishery portfolios across western North America. J. Appl. Ecol. **51**: 1554–1563. doi:10.1111/1365-2664.12341. PMID:25552746.

Harvey, B.N., Jacobson, D.P., and Banks, M.A. 2014. Quantifying the uncertainty of a juvenile Chinook salmon race identification method for a mixed-race stock. North Am. J. Fish. Manage. **34**: 1177–1186. doi:10.1080/02755947.2014.951804.

Hess, M.A., Rabe, C.D., Vogel, J.L., Stephenson, J.J., Nelson, D.D., and Narum, S.R. 2012. Supportive breeding boosts natural population abundance with minimal negative impacts on fitness of a wild population of Chinook salmon. Mol. Ecol. **21**: 5236–5250. doi:10.1111/mec.12046. PMID:23025818.

Hilborn, R., Quinn, T.P., Schindler, D.E., and Rogers, D.E. 2003. Biocomplexity and fisheries sustainability. Proc. Natl. Acad. Sci. U.S.A. **100**: 6564–6568. doi:10.1073/pnas.1037274100. PMID:12743372.

Huber, E.R., and Carlson, S.M. 2015. Temporal trends in fall Chinook hatchery releases from California's Central Valley. San Fr. Estuary Watershed Sci. **13**: 1–23.

Jakobsson, M., and Rosenberg, N.A. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics, **23**: 1801–1806. doi:10.1093/bioinformatics/btm233. PMID:17485429.

Johnson, R.C., Windell, S., Brandes, P.L., Conrad, J.L., Ferguson, J., and Goertler, P.A.L. 2017. Science Advancements Key to Increasing Management Value of Life Stage Monitoring Networks for Endangered Sacramento River Winter-Run Chinook Salmon in California. San Fr. Est. Watershed Sci. **15**(3). Retrieved from https://escholarship.org/uc/item/6751j957.

Jombart, T., and Ahmed, I. 2011. adegenet 1. 3-1: new tools for the analysis of genome-

Can. J. Fish. Aquat. Sci. Downloaded from www.nrcresearchpress.com by National Marine Mammal Lab Lib on 02/25/20
For personal use only.

wide SNP data. Bioinformatics, **27**: 3070–3071. doi:10.1093/bioinformatics/btr521. PMID:21926124.

Jombart, T., Devillard, S., and 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. PMID:20950446.

Langmead, B., and Salzberg, S. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Meth. **9**: 357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Gen. Biol. **10**(3): R25.

Larson, W.A., Seeb, L.W., Everett, M.V., Waples, R.K., Templin, W.D., and Seeb, J.E. 2014. Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). Evol. Appl. **7**: 355–369. doi:10.1111/eva.12128. PMID:24665338.

Lew, R.M., Finger, A.J., Baerwald, M.R., Goodbla, A., May, B., and Meek, M.H. 2015. Using next-generation sequencing to assist a conservation hatchery: a single-nucleotide polymorphism panel for the genetic management of endangered delta smelt. Trans. Am. Fish. Soc. **144**: 767–779. doi:10.1080/00028487.2015.1037016.

Liebhold, A., Koenig, W.D., and Bjørnstad, O.N. 2004. Spatial synchrony in population dynamics. Annu. Rev. Ecol. Evol. Syst. **35**: 467–490. doi:10.1146/annurev.ecolsys.34.011802.132516.

Lindley, S.T., Schick, R., May, B.P., Anderson, J.J., Greene, S., Hanson, C., et al. 2004. Population structure of threatened and endangered Chinook salmon ESUs in California's Central Valley basin. NOAA Tech. Memo. NOAA-TM-NM.

Lindley, S.T., Grimes, C.B., Mohr, M.S., Peterson, W., Stein, J., Anderson, J.T., et al. 2009. What caused the Sacramento River fall chinook stock collapse? NOAA Tech. Memo. NMFS-SWFSC, **447**: 1–61.

McClure, M.M., Utter, F.M., Baldwin, C., Carmichael, R.W., Hassemer, P.F., Howell, P.J., et al. 2008. Evolutionary effects of alternative artificial propagation programs: implications for viability of endangered anadromous salmonids. Evol. Appl. **1**: 356–375. doi:10.1111/j.1752-4571.2008.00034.x. PMID:25567637.

McKinney, G.J., Seeb, L.W., Larson, W.A., Gomez-Uchida, D., Limborg, M.T., Brieuc, M.S.O., 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. PMID:26490135.

Meek, M.H., and Larson, W.A. 2019. The future is now: Amplicon sequencing and sequence capture usher in the conservation genomics era. Mol. Ecol. Resour. **19**: 795–803.

Meek, M.H., Baerwald, M.R., Stephens, M.R., Goodbla, A., Miller, M.R., Tomalty, K.M.H., and 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. PMID:30128122.

Meirmans, P.G., and Van Tienderen, P.H. 2004. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Mol. Ecol. Notes, **4**: 792–794. doi:10.1111/j.1471-8286.2004.00770.x.

Miller, M.R., Brunelli, J.P., Wheeler, P.A., Liu, S., Rexroad, C.E., Palti, Y., et al. 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. PMID:21988725.

Moore, J.W., McClure, M.M., Rogers, L.A., and Schindler, D.E. 2010. Synchronization and portfolio performance of threatened salmon. Conserv. Lett. **3**: 340–348. doi:10.1111/j.1755-263X.2010.00119.x.

Moore, J.W., Yeakel, J.D., Peard, D., Lough, J., and Beere, M. 2014. Life-history diversity and its importance to population stability and persistence of a migratory fish: steelhead in two large North American watersheds. J. Anim. Ecol. **83**: 1035–1046. doi:10.1111/1365-2656.12212.

Moyle, P.B., Lusardi, R.A., Samuel, P.J., and Katz, J.V.E. 2017. State of the salmonids: status of California's emblematic fishes. Center for Watershed Sciences, University of California, Davis and California Trout, San Francisco, Calif.

Myers, J.M., Kope, G.H., Bryant, G.J., Teel, D., Lierheimer, L.J., Wainwright, T.C., et al. 1998. Status review of Chinook salmon from Washington, Idaho, Oregon and California. US Deptartment Commer. Commer. NOAA Tech. Memorandum NMFS-NWFSC-35 443.

Narum, S.R. 2006. Beyond Bonferroni: Less conservative analyses for conservation genetics. Conserv. Genet. **7**: 783–787. doi:10.1007/s10592-005-9056-y.

Narum, S.R., Banks, M.A., Beacham, T.D., Bellinger, M.R., Campbell, M.R., Dekoning, J., et al. 2008. Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. Mol. Ecol. **17**: 3464–3477. PMID:19160476.

Narum, S.R., Di Genova, A., Micheletti, S.J., and 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.

Nazareno, A.G., Bemmels, J.B., Dick, C.W., and Lohmann, L.G. 2017. Minimum sample sizes for population genomics: An empirical study from an Amazonian plant species. Mol. Ecol. Resour. **17**: 1136–1147. doi:10.1111/1755-0998.12654. PMID:28078808.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Oksanen, J., Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. 2017. vegan: Community Ecology Package. R package version 2.4-4 [online]. Available from https://CRAN.R-project.org/package=vegan.

O'Malley, K.G., Camara, M.D., and Banks, M.A. 2007. Candidate loci reveal genetic differentiation between temporally divergent migratory runs of Chinook salmon (*Oncorhynchus tshawytscha*). Mol. Ecol. **16**: 4930–4941.

O'Malley, K.G., Jacobson, D.P., Kurth, R., Dill, A.J., and Banks, M.A. 2013. Adaptive genetic markers discriminate migratory runs of Chinook salmon (*Oncorhynchus tshawytscha*) amid continued gene flow. Evol. Appl. **6**: 1184–1194. doi:10.1111/eva.12095. PMID:24478800.

Ouborg, N.J., Pertoldi, C., Loeschcke, V., Bijlsma, R., and Hedrick, P.W. 2010. Conservation genetics in transition to conservation genomics. Trends Genet. **26**: 177–187. doi:10.1016/j.tig.2010.01.001. PMID:20227782.

Palmer-Zwahlen, M., and Kormos, B. 2015. Recovery of coded-wire tags from Chinook salmon in California's Central Valley escapement, inland harvest, and ocean harvest in 2012. California Department of Fisheries and Wildlife Fisheries Administrative Report 2015-4.

Prince, D.J., O'Rourke, S.M., Thompson, T.Q., Ali, O.A., Arciniega, M., Lyman, H.S., Saglam, I.K., Clemento, A.J., Hotaling, T.J., Kinziger, A.P., Spidle, A.P., Garza, J.C., Pearse, D.E., and Miller, M.R. 2017. The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. Sci. Adv. **3**: e1603198.

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics, **155**: 945–959. PMID:10835412.

Quinn, T.P. 2005. The behavior and ecology of Pacific salmon and trout. University of Washington Press, Seattle, Wash.

R Development Core Team. 2005. R: A language and environment for statistical computing, reference index version 2.2.1 [online]. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.R-project.org.

Satterthwaite, W.H., and Carlson, S.M. 2015. Weakening portfolio effect strength in a hatchery-supplemented Chinook salmon population complex. Can. J. Fish. Aquat. Sci. **72**: 1860–1875. doi:10.1139/cjfas-2015-0169.

Satterthwaite, W.H., Carlson, S.M., and Criss, A. 2017. Ocean size and corresponding life history diversity among the four run timings of California central valley chinook salmon. Trans. Am. Fish. Soc. **146**: 594–610. doi:10.1080/00028487.2017.1293562.

Schindler, D.E., Hilborn, R., Chasco, B., Boatright, C.P., Quinn, T.P., Rogers, L.A., and Webster, M.S. 2010. Population diversity and the portfolio effect in an exploited species. Nature, **465**: 609–612. doi:10.1038/nature09060. PMID:20520713.

Sommer, T.R., Harrell, W.C., Nobriga, M.L., Brown, R., Moyle, P.B., Kimmerer, W., and Schemel, L.E. 2001*a*. California's Yolo Bypass: Evidence that flood control can be compatible with fisheries, wetlands, wildlife, and agriculture. Fisheries, **26**(8): 6–16. doi:10.1577/1548-8446(2001)026<0006:CYB>2.0.CO;2.

Sommer, T.R., Nobriga, M.L., Harrell, W.C., Batham, W., and Kimmerer, W.J. 2001*b*. Floodplain rearing of juvenile chinook salmon: evidence of enhanced growth and survival. Can. J. Fish. Aquat. Sci. **58**: 325–333. doi:10.1139/f00-245.

Thompson, T.Q., Bellinger, M.R., O'Rourke, S.M., Prince, D.J., Stevenson, A.E., Rodrigues, A.T., Sloat, M.R., Speller, C.F., Yang, D.Y., Butler, V.L., Banks, M.A., and Miller, M.R. 2019. Anthropogenic habitat alteration leads to rapid loss of adaptive variation and restoration potential in wild salmon populations. Proc. Natl. Acad. Sci. **116**(1): 177–186. doi:10.1073/pnas.1811559115.

Vasemägi, A., Kahar, S., and Ozerov, M.Y. 2016. Genes that affect Atlantic salmon growth in hatchery do not have the same effect in the wild. Funct. Ecol. **30**: 1687–1695. doi:10.1111/1365-2435.12635.

Waples, R.S. 2001. Characterizing diversity in salmon from the Pacific Northwest. J. Fish Biol. **59**: 1–41. doi:10.1111/j.1095-8649.2001.tb01376.x.

Waples, R.S., and Do, C. 2010. Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. Evol. Appl. **3**: 244–262.

Weir, B.S., and Cockerham, C.C. 1984. Estimating F-Statistics for the analysis of population structure. Evolution, **38**: 1358–1370. doi:10.2307/2408641. PMID:28563791.

Williams, J.G. 2006. Central Valley salmon: A perspective on Chinook and Steelhead in the Central Valley of California. San Fr. Estuary Watershed Sci. 4.

Williams, J.G. 2012. Juvenile Chinook salmon (*Oncorhynchus tshawytscha*) in and around the San Francisco Estuary. San Fr. Estuary Watershed Sci. 10.

Williamson, K.S., and May, B.P. 2005. Homogenization of Fall-run Chinook salmon gene pools in the Central Valley of California, USA. North Am. J. Fish. Manage. **25**: 993–1009. doi:10.1577/M04-136.1.

Yoshiyama, R.M., Fisher, F.W., and Moyle, P.B. 1998. Historical abundance and decline of Chinook salmon in the Central Valley region of California. North Am. J. Fish. Manage. **18**: 487–521. doi:10.1577/1548-8675(1998)018<0487:HAADOC>2.0.CO;2.