Population structure of chum salmon and selection on the markers collected for stock identification

Shuichi Kitada[1*] and Hirohisa Kishino[2,3]
[1] Tokyo University of Marine Science and Technology, Tokyo 108-8477, Japan.
[2] Graduate School of Agriculture and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan.
[3] Current address: The Research Institute of Evolutionary Biology, Tokyo 138-0098, Japan
[1*]kitada@kaiyodai.ac.jp, [2,3] hirohisa.kishino@gmail.com

Abstract

Genetic stock identification (GSI) is a major management tool of Pacific salmon (*Oncorhynchus* Spp.) that has provided rich genetic baseline data of allozymes, microsatellites, and single nucleotide polymorphisms (SNPs) across the Pacific Rim. Here, we analyzed published data sets for adult chum salmon (*Oncorhynchus keta*), namely 10 microsatellites, 53 SNPs, and a mitochondrial DNA locus (mtDNA3, control region and NADH-3 combined) from 495 locations in the same distribution range (n = 61,813). TreeMix analysis of the microsatellite loci identified the highest level of genetic drift towards Japanese/Korean populations and suggested two admixture events from Japan/Korea to Russia and the Alaskan Peninsula. The SNPs had been purposively collected from rapidly evolving genes to increase the power of GSI. The highest expected heterozygosity was observed in Japanese/Korean populations for microsatellites, whereas it was highest in Western Alaskan populations for SNPs, reflecting the SNP discovery process. By regressing the SNP population structures on those of the microsatellites, we estimated the selection on the SNP loci according to deviations from the predicted structures. Specifically, we matched the sampling locations of the SNPs with those of the microsatellites according to geographical information and performed regression analyses of SNP allele frequencies on the two coordinates of multi-dimensional scaling (MDS) of matched locations obtained from microsatellite pairwise F_{ST} values. The MDS first axis indicated a latitudinal cline in American and Russian populations, whereas the second axis found a differentiation of Japanese/Korean populations. The top five outlier SNPs were mtDNA3 (combined locus of the control region and NADH-3), U502241 (unknown), GnRH373, ras1362, and TCP178, which were consistently identified by principal component analysis. We summarized the functions of the 53 nuclear SNPs and mtDNA3 locus by referring to a gene database system and discussed the functions of the outlier SNPs and fitness of chum salmon.

Keywords

demographic history, hatchery, microsatellite, mtDNA, population structure, SNPs

1 | INTRODUCTION

Chum salmon (*Oncorhynchus keta*) has a wide distribution range from the Arctic to California in North America, and from Siberia to Japan and Korea in Asia (Salo, 1991). The ability to adapt to the varied environments of spawning rivers in the past and at present may have expanded its distribution range. Chum salmon also comprise the world's largest hatchery release program (Amoroso et al., 2017; Flagg, 2015; Kitada, 2018; Naish et al., 2007), and every year, more than 3 billion hatchery-reared chum salmon are released in the North Pacific. Therefore, chum salmon is one of the best species to study the structure and history of populations, which could be underpinned by environmental adaptation and anthropogenic selection pressures.

The genetic population structure of chum salmon has traditionally been studied in the context of genetic stock identification (GSI). GSI has been a major management tool of Pacific salmon (*Oncorhynchus* Spp.) since the early 1990s (Beacham et al. 2008a-d, 2009b, 2020; Larson et al., 2014; Shaklee, 1999; Utter, 1991), and it has significantly contributed to high-seas and coastal migration studies (Myers et al., 2007; Seeb et al., 2004). In Pacific salmon GSI, samples from mixed-stock fishery and forensic studies were analyzed with the general objective of providing the optimal resolution of mixing proportions of stocks at a reasonable cost (Beacham et al., 2020). GSI studies have provided genetic baseline data for salmon populations across the Pacific Rim, and these data have contributed to studies into population structure, mixed-stock fisheries, and genetic interactions between hatchery and wild salmon (Waples et al. 2020).

Genetic markers for GSI have progressed from allozymes to microsatellites and single-nucleotide polymorphisms (SNPs) (Beacham et al. 2020; Bernatchez et al., 2017). Allozyme loci often have a small number of alleles. To improve the power of GSI resolution for the high gene flow salmonids, microsatellites were developed because the number of alleles is generally much larger than that of isozymes and much more information can be included. However, standardizing hundreds of microsatellite alleles across sampling points in different countries is difficult (Seeb et al., 2011). To avoid the standardization problem, genotyping of microsatellites of salmon species was generally performed by a single laboratory (Beacham et al. 2008a,b, 2009a,b; Seeb et al., 2011). In contrast, calibrating SNP genotyping is more straightforward because genotype data can be stored in a unified format and can be accessed by different laboratories on different continents (Waples et al. 2020).

Chum salmon is one of the major salmonid species that have been studied for GSI, and SNPs with large allele frequency differences among populations have been genotyped in three different studies to improve the power of GSI (Elfstrom et al., 2007; Smith et al., 2005a,b). The genes that were selected for SNP typing were originally identified as rapidly evolving genes (Elfstrom et al., 2007; Seeb et al., 2011) that showed positive selection in humans and chimpanzees (Nielsen et al., 2005). They included genes that were associated with fatty acid synthesis, testis-specific expression, olfactory receptors, immune responses, and cell growth and differentiation (Elfstrom et al., 2007; Smith et al., 2005a,b). The population structure determined using the SNPs selected for the GSI was biased not only by the selection of the genes but also by the SNP discovery process. Specifically, the three studies were focused on Western Alaska, which was the area of the authors' interest (Seeb et al., 2011). Consistently, the SNP allelic richness and heterozygosity is high in Alaskan

populations.

Neutral and adaptive markers and combinations of them can be useful in establishing optimal management strategies (Funk et al., 2012). Population structures inferred using neutral markers reflect gene flow and genetic drift (Waples & Gaggiotti, 2006), which affect within and among population variations and can lead to adaptive divergence in the genome (Funk et al., 2012). To integrate adaptive markers into the definition of conservation units, Funk and colleagues proposed a framework of comparing population structures inferred from putatively neutral and adaptive loci. The inclusion of information on putatively adapted loci can help to understand mechanisms of local adaptation and is useful for conservation and management of the species (Moore et al., 2014).

Here, we analyzed the published data sets of microsatellites and the SNPs genotyped for chum salmon GSI. First, we inferred the chum salmon population structure and its demographic history using the microsatellite data in a distribution range. Then, we matched the sampling locations of the SNP genotyping studies with those of the microsatellite data according to the available geographical information. By regressing the SNP population structure on the microsatellite population structure, we estimated the selection on the SNPs as deviations from the predicted structure.

2 | MATERIAL AND METHODS

2.1 | Screening of population genetics data for chum salmon

We screened population genetics studies of chum salmon in the literature published after 1990 using the Google Scholar search system with keyword searches of “mixed-stock fisheries”, “population structure”, “salmon”, and “stock identification”. We also added studies that were known to us. The population genetics of chum salmon in the North Pacific has been studied extensively over the last four decades. International sampling of chum salmon has been conducted cooperatively across its distribution range to establish baseline genotype data for effective GSI (Beacham et al., 2008a,b, 2009a,b; Seeb & Crane, 1999; Seeb et al., 1995; Urawa et al., 2005, 2009; Wilmot et al., 1994). The population structure of chum salmon has also been studied extensively using various markers, such as isozymes (Kijima & Fujio, 1979; Okazaki, 1982; Sato & Urawa, 2015; Seeb & Crane, 1999; Seeb et al., 1995; Wilmot et al., 1994; Winans et al., 1994), mitochondrial DNA (mtDNA) (Garvin et al., 2010; Park et al., 1993; Sato et al., 2004; Yoon et al., 2008), minisatellites (Taylor et al., 1994), and microsatellites (Beacham et al., 2008a,b, 2009a,b; Olsen et al., 2008. More recently, SNPs were developed for accurate stock identification and have also been used for population genetics (Garvin et al., 2013; Petrou et al., 2014; Saito et al., 2020; Sato et al., 2014; Seeb et al., 2011; Small et al., 2015). A high-throughput panel was developed for chum salmon and patterns of linkage disequilibrium have been examined (McKinney et al., 2020).

Through the data screening, we found several publicly available data sets that covered the distribution range, including Japan, of chum salmon. These data sets comprised microsatellite allele frequencies (Beacham et al., 2009a), nuclear SNP genotypes with a combined mtDNA locus (Seeb et al., 2011), isozyme allele frequencies (Winans et al., 1994; Seeb et al., 1995), and mtDNA control regions (Sato et al., 2004). In this study, we analyzed the microsatellite (Beacham et al., 2009a) and SNP (Seeb et al., 2011) data sets. All Japanese samples were caught in hatcheries and weirs in hatchery-enhanced rivers and

were therefore hatchery-reared fish and/or hatchery descendants.

2.2 | Microsatellite and SNP data sets analyzed in this study

Microsatellite allele frequencies of chum salmon were retrieved from the Molecular Genetics Lab Online Data, Fisheries and Oceans Canada website (https://www.pac.dfo-mpo.gc.ca/science/facilities-installations/pbs-sbp/mgl-lgm/data-donnees/index-eng.html, accessed 01 August 2020). The data consisted of the allele frequencies at 14 loci from chum salmon populations at 381 localities in a distribution range (n = 51,355) (Table S1; Figure S1) (Beacham et al., 2009a). These data were used to infer the population structure and demographic history of chum salmon.

SNP genotypes were retrieved from the Dryad data repository (Seeb et al, 2011). The data consisted of 58 SNPs collected from 114 locations in a distribution range (Table S2; Figure S2) (Seeb et al, 2011). We excluded four loci, *Oke_U401-220*, *Oke_GHII-2943*, *Oke_IL8r-272*, and *Oke_U507-87*, to avoid pseudo-replication following the original study. A total of 53 nuclear SNPs and a combined mtDNA locus (mtDNA3) were included in our analysis (n = 10,458). The mtDNA3 locus contained three gene loci, *Oke-Cr30* and *Oke-Cr386* (mtDNA control region), and *Oke-ND3-69*（mtDNA NADH-3).

2.3 | Accounting for ascertainment bias in the data sets

Significant deviations from Hardy–Weinberg equilibrium at four loci (*Oke*3, *Ots*103, *One*111 and *OtsG68*) from a 14-microsatellite locus panel were found in Japanese, Russian, and representative North American populations (Beacham et al., 2008a,b). These deviations were likely because of ascertainment bias (Seeb et al., 2011), and therefore we excluded these four loci and used the remaining 10 loci in our analysis.

For the SNP data, the highest levels of allelic richness detected in parts of the Alaskan region may be affected by ascertainment bias (Seeb et al., 2011). Allele/site frequency spectra have been used to examine potential ascertainment bias in SNPs (Nielsen et al., 2004). We calculated allele frequency spectra in six geographical areas according to the major lineage used in the original study (Seeb et al., 2011), namely Western Alaska, Yukon/Kuskokwim upper, Alaskan Peninsula, Southeast Alaska (SEA)/British Columbia (BC)/Washington (WA), Russia, and Japan/Korea.

2.4 | Inferring population structure and demographic history

We used the microsatellite allele frequencies (Beacham et al., 2009a) to infer the population structure of chum salmon. In previous coalescent simulations, we found that the bias-corrected G_{ST} moment estimator (Nei & Chesser, 1983) performed better than other F_{ST} estimators when estimating pairwise F_{ST} values (Kitada et al., 2017). Therefore, in this study, we used the Nei & Chesser G_{ST} moment estimator (NC83) to compute pairwise F_{ST}. Because only allele frequency data could be accessed, we used the 'read.frequency' function and computed pairwise F_{ST} values averaged over loci using the 'GstNC' function in the R package FinePop ver.1.5.1 on CRAN. A multi-dimensional scaling (MDS) analysis was performed on the pairwise F_{ST} distance matrix using the 'cmdscale' function in R. We used the cumulative contribution ratio up to the kth axis $(j = 1, \ldots, k, \ldots, K)$ as the explained variation measure, which was computed in the R function as $C_k = \sum_{j=1}^{k} \lambda_j / \sum_{j=1}^{K} \lambda_j$, where λ_j is the eigenvalue and $\lambda_j = 0$ if $\lambda_j < 0$. Following the original study (Beacham et al., 2009a), the populations were classified into eight

geographical areas: Japan, Korea, Russia, Alaskan Peninsula, Western Alaska, Yukon/Tanana/Upper Alaska, SEA/BC, and WA (Table S1). A neighbor-joining (NJ) tree (Saitou & Nei, 1987) was constructed using the pairwise F_{ST} distance matrix using the 'nj' function in the R package 'ape'.

To infer demographic history with admixture, we applied TreeMix (Pickrell &Pritchard, 2012) for the microsatellite allele frequencies (Beacham et al., 2009a). The analyses were performed using six regional genetic groups, where Japan/Korea and SEA/BC/WA were combined to reduce the number of parameters to be estimated under the limited number of loci. On the basis of the allele frequencies and numbers of repeats, we computed the mean and variance in length at each microsatellite locus and used them to run TreeMix v1.13. We tested up to five migration events and found the best tree was based on the Akaike Information Criterion (AIC; Akaike, 1973); $AIC = -2\log\hat{L} + 2 \times s$, where $\hat{L}$ is the maximum composite likelihood and *s* is the number of parameters to be estimated, which, in this study, was three per migration event, namely, migration edge, branch length, and migration weight. For example, for the model with two migration events, the number of parameters was six. TreeMix maximizes composite likelihood and the 2×log likelihood ratio is expected to follow an approximate χ^2 distribution with 3 degrees of freedom. Because TreeMix calculates composite likelihoods, the penalty in AIC needs to account for the correlation in the likelihoods. To avoid unconscious bias towards selection of parameter-rich models, we kept all the candidate models except those where the AIC values were decisively different.

2.5 | Gene flow and genetic diversity

We recorded approximate longitudes and latitudes of the sampling sites according to the names of rivers and/or areas and maps from the original studies using Google Maps. Sampling locations were plotted on the map using the 'sf' package in R. Sampling points with pairwise F_{ST} values lower than a given threshold were connected by yellow lines to visualize the population connectivity. Under the assumption of Wright's island model $F_{\mathrm{ST}} \approx \frac{1}{4N_e m+1}$, where N_e is the effective population size and *m* is the average migration rate between populations (Slatkin, 1987), a threshold of $F_{\mathrm{ST}} = 0.01$ refers to $4N_e m \approx 99$ migrants per generation (see, Whitlock & Mccauley, 1999; Waples & Gaggiotti, 2006).

We computed expected heterozygosity (*He*) values for the data sets. For microsatellites, we computed *He* for each population at each locus based on the allele frequencies. We then averaged *He* values over loci. For the 53 nuclear SNPs, we converted the original genotype data to genepop format (Rousset, 2008), and read the data using the 'read.GENEPOP' function in the R package FinePop2 ver.0.4 on CRAN, where *He* values were computed automatically. The mtDNA3 locus had five alleles, and *He* was computed using the allele frequencies. We also computed NC83 pairwise F_{ST} values using 53 nuclear SNPs using the 'pop_pairwiseFST' function in FinePop2.

Sampling points were visualized by a color gradient of population-specific *He* values. The color of population *i* was rendered as RGB $(1 - H_{e0,i}, 0, H_{e0,i})$, where $H_{e0,i} = (H_{e,i} - \min H_e)/(\max H_e - \min H_e)$. This conversion represents the standardized magnitude of an *He* value at the sampling point, with a continuous color gradient ranging from blue to red for the lowest and highest *He* values, respectively. Population-specific *He* values of

microsatellites and SNPs were grouped by seven geographical areas: Japan/Korea, Russia, Alaskan Peninsula, Western Alaska, Yukon/Tanana/Kuskokwim, SEA/BC, and WA. Differences in average *He* values between all pairs of the groups were tested by one-way ANOVA, where *p*-values were adjusted for the multiple comparisons using the 'TukeyHSD' function in R.

2.6| Identifying highly differentiated SNPs beyond the neutral population structure

To compare the genetic differentiation of SNP allele frequencies with the population structure inferred using the microsatellite markers, we analyzed 10 microsatellite loci and 53 nuclear SNPs loci plus the combined mtDNA3 locus. The mtDNA3 locus had five alleles and the major allele (second one) was used for the meta-analysis. The functions of all 54 analyzed gene loci were confirmed by referring to the GeneCards database system (https://www.genecards.org/) and published literature.

First, we matched the 114 sampling locations of the SNPs (Figure S2) to the nearest location of the 381 sampling points of the microsatellites (Figure S1). The same locations were not necessarily used for genotyping. When we did not find the same location names, we assigned the closest locations of the microsatellites to those of SNPs using the longitudes and latitudes of the sampling locations. Among the 114 sampling locations for SNPs, 56 locations also had samples for microsatellite genotyping. We assigned the closest locations of microsatellites to the other 58 locations for SNPs. Then, we identified loci with allele frequencies that were significantly differentiated from the microsatellite population structure in the pairwise F_{ST}. Specifically, we performed regression analyses of allele frequencies of the 53 nuclear SNPs and the mtDNA3 locus on the two coordinates (axes) of the MDS of the pairwise F_{ST} estimated using the microsatellite allele frequencies as:

$$p_j^{(i)} = \beta_0^{(i)} + \beta_1^{(i)} mds1_j + \beta_2^{(i)} mds2_j + \varepsilon_j^{(i)}, \varepsilon_j^{(i)} \sim N(0, \sigma^2),$$

where, $p_j^{(i)}$ is the allele frequency of SNP $i\ (= 1, \ldots, I)$ in population $j\ (= 1, \ldots, J)$. The explanatory variables $mds1_j$ and $mds2_j$ are the 1st and 2nd MDS coordinates of pairwise F_{ST} values estimated from the microsatellite allele frequencies for the matched populations $j\ (= 1, \ldots, J)$. These explanatory variables are population specific ($J = 114$). The regression analysis was performed for each marker ($i = 1, \ldots, 54$) and 54 sets of regression coefficients ($\beta_0^{(i)}, \beta_1^{(i)}, \beta_2^{(i)}$) were obtained. The *p*-values for the coefficients were adjusted for the multiple comparison (*q*-values) by the method of Benjamini & Hochberg (1995) using the 'p.adjust' function in R. Given the population structure predicted from the neutral microsatellites, the outlier loci from the set of 53 nuclear SNPs and mtDNA3 locus that deviated largely from the predicted population structure are likely to be under natural selection and/or anthropogenic selection.

To identify genes that best characterize geographical areas, we conducted a principal components analysis (PCA) for the allele frequencies of the 53 nuclear SNPs and mtDNA3 locus using the 'prcomp' function in R. Finally, we visualized the geographical distributions of the identified highly differentiated genes using boxplots and geographical maps of the allele frequencies and *He* values. The sampling locations were mapped with a color gradient of allele frequencies for population *j* rendered as RGB $(p_{j0}, 0, 1 - p_{j0})$, where $p_{j0} = (p_j - \min p)/(\max p - \min p)$. This conversion represents the standardized

allele frequency values at the sampling point, with a continuous color gradient ranging from blue to red for the lowest and highest allele frequencies, respectively.

3 | RESULTS

3.1 | Population structure and demographic history of chum salmon

Pairwise F_{ST} values based on the microsatellite allele frequencies were 0.019 ± 0.010 (SD, standard deviation). The MDS of pairwise F_{ST} values (Figure 1a) indicated a latitudinal cline in the American and Russian populations and a separation of Alaskan populations from the others along the first axis (mds1) explaining 29% of the variance. The second axis (mds2, 16% of variance) showed a differentiation of Japanese/Korean populations from the others, although they were closely related to southern Russian populations from Sakhalin, Amur, and Primorye. Plots of mds1 vs mds3, and mds1 vs mds4 produced similar results, showing most differentiation was driven by divergence from a latitudinal cline (Figure S3). The mds1 to mds4 explained 60% variation in the pairwise F_{ST} distant matrix (Figure S4). In the unrooted NJ tree generated from this data set (Figure S5a), five large regional population groupings were inferred: (i) Alaskan Peninsula, (ii) Western Alaska/Yukon (Canada), (iii) Russia/Japan/Korea, (iv) SEA/Northern BC, and (v) Southern BC/WA. The TreeMix analysis indicated two periods of admixture (migration events), from Japan/Korea to Russia and the Alaskan Peninsula (Figure 1b; Figure S6). Japanese/Korean populations had the highest level of genetic drift among all the groups, whereas the drift was the lowest in the Alaskan Peninsula and SEA/BC/WA populations.

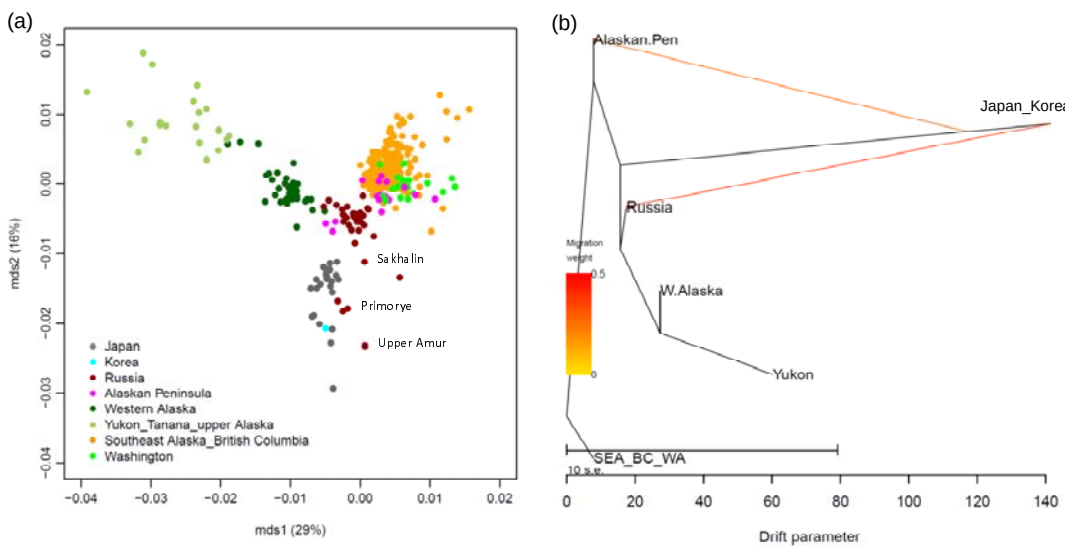

FIGURE 1 Population structure and admixture events of chum salmon based on microsatellite markers. (a) Multi-dimensional scaling (MDS) plots (mds1 vs mds2) of the population structure of chum salmon based on pairwise F_{ST} values estimated from 10 microsatellite loci in 381 populations. (b) Admixture graph inferred using TreeMix with two migration events. The scale bar (drift parameter) is 10 times the average standard error of the sample covariance matrix of allele frequencies between populations (Pickrell & Pritchard, 2012).

3.2 | Ascertainment bias in the SNP data

The allele frequency spectra (Figure S7) showed a similar pattern in all areas except Western Alaska, where a much lower frequency was found for the lowest allele frequency in Western Alaska.

3.3 | Gene flow and genetic diversity

Using the criterion of pairwise F_{ST} <0.01 ($4N_e m \approx 99$), substantial gene flow between American and Asian populations was identified for the microsatellites (Figure 2a). Japanese populations were found to be connected to Russian and Korean populations. For the SNPs, using the same criterion, we detected substantial gene flow within Alaska, but the SEA and WA populations appeared to be isolated and gene flow was very little between American and Russian populations (Figure 2b). Japanese and Korean populations were isolated from all the others.

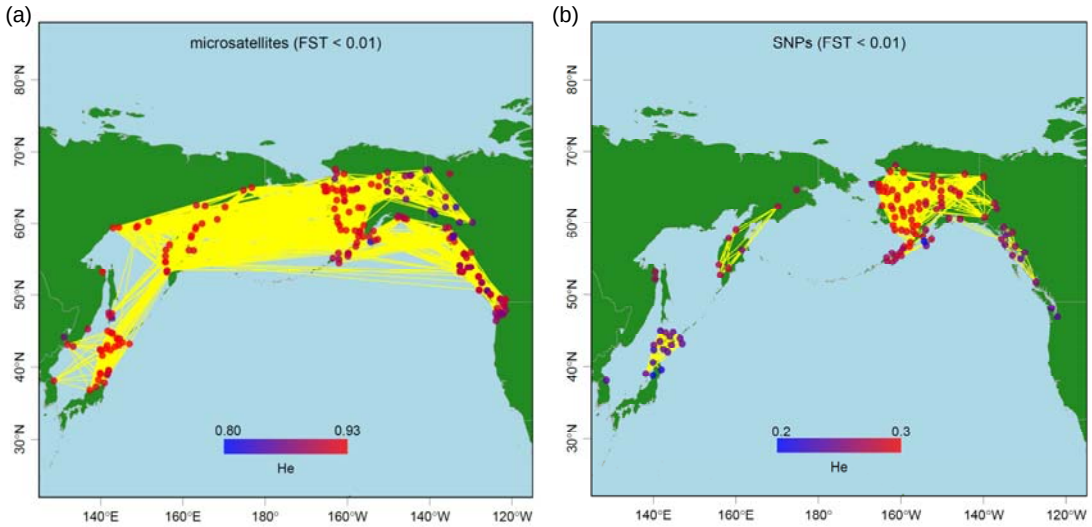

FIGURE 2 Visualization of genetic diversity and population connectivity of chum salmon based on microsatellite and SNP markers. (a) Map obtained using 10 microsatellite loci of 381 populations (n = 51,355) (Beacham et al., 2009a). (b) Map obtained using 53 SNPs of 114 populations (n = 10,458) (Seeb et al., 2011). The color of each population reflects the magnitude of the expected heterozygosity (H_e) values, with a continuous color gradient ranging from blue to red for the lowest and highest *He* values, respectively. The yellow lines connect populations with pairwise F_{ST} <0.01.

For the microsatellites, the *He* values were generally high in almost all the populations (Figure 2a, Figure S8a). The average *He* value over loci was 0.89±0.02 (range 0.81–0.93). High *He* values were found in Japanese, Russian, and Alaskan populations, whereas *He* values were slightly lower in SEA/BC, WA, and Canadian Yukon populations (Figure S5b). Japanese populations had the highest *He* values, but the mean was similar to those in Russian and Western Alaskan populations, lower in Alaskan Peninsular, SEA/BC, and WA populations, and extremely low in the Yukon population (Figure S8a).

In contrast, for the SNPs, the highest *He* values were found in Western Alaska populations, followed by Yukon, Alaskan Peninsula, Russia, and SEA populations (Figure 2b, Figure

S8b). The average *He* value over loci was 0.27±0.02 (range 0.20–0.30). Japan and WA populations had the lowest *He* values. The *He* values were similar in populations from Western Alaska, Yukon, Alaskan Peninsula, and Russia (Figure S8b) and, although Russian populations had higher *He* values than the SEA/BC and WA populations, the differences between these populations were not significant. Japanese populations had the lowest *He* values with no significant difference between it and the WA populations.

3.4 | Differentiated SNPs beyond the neutral population structure

We summarized 53 nuclear SNPs and the mtDNA3 locus with functions (if known), locus names, and raw data locus names used in our analysis (Table S3). Five of the outlier SNPs were identified as potential candidates for selection in the original study (Seeb et al., 2011). We found 31 SNPs in functional genes and 23 SNPs were unknown.

The results of our regression analysis are summarized in Table 1. Most of the SNPs had large coefficients for both mds1 and mds2 with highly significant *q*-values, showing that most of SNPs were differentiated beyond the neutral population structure. The top five outlier loci (Table 1) with the highest regression coefficients and *q*-values were mtDNA3, U502241, GnRH373, ras1362, and TCP178, which characterized mds2 of the pairwise F_{ST}. The scatter plot of mds1 and mds2 (Figure 3a) identified the outlier SNPs that diverged beyond the scale of mds1. The PCA (Figure 3b) identified differences between Japanese/Korean and other populations as a primary component (PC1, 43% variance), whereas the second component (PC2, 20% variance) corresponded to the latitudinal cline among the Russian/American populations. Biplots of PCA1 versus PCA3 and PCA4 described outliers (Susitna, and Sturgeon River on Kodiak Island), but most differentiation was caused by divergence of the Japanese/Korean populations (Figure S9). PCA1 to PCA4 explained 76% of variance of the SNP allele frequencies (Figure S10). The PCA (Figure 3b) confirmed that differentiation of Japanese/Korean populations were characterized mainly by the top five outlier SNPs detected by the regression analysis (Figure 3a).

The allele frequencies of the top five outlier loci were distinct in Japanese/Korean populations in the distribution range (Figure 4). The major allele of mtDNA3 was fixed at close to 1.0 in American and Russian populations, whereas the allele frequencies were significantly lower (0.21± 0.11) in Japanese/Korean populations. The U502241 allele frequencies were much higher in Japanese/Korean and Washington populations than they were in the other populations. GnRH373, ras1362, and TCP178 had similar geographical distributions of allele frequencies, and in Japanese/Korean populations they were differentiated from those in the other populations. The average genetic diversity (*He*) was significantly higher in Japanese/Korean populations than that it was in Western Alaska for mtDNA3 (Welch two sample *t* test, $t = 30.5, \mathrm{df} = 17.5, p = 2.2 \times 10^{-16}$) and TCP178 loci ($t = 36.8, \mathrm{df} = 51.3, p = 2.2 \times 10^{-16}$), whereas it was similar in Western Alaska for U502241 ($t = 0.9, \mathrm{df} = 19.7, p = 0.37$), and ras1362 ($t = 0.6, \mathrm{df} = 35.7, p = 0.57$). GnRH373 had lower *He* values in Japanese/Korean populations (0.33) than in Western Alaskan populations (0.40) ($t = -3.0, \mathrm{df} = 21.2, p = 0.006$).

Table 1 Regression analysis of SNP allele frequencies on abs(mds1) and abs(mds2) of the microsatellite pairwise F_{ST} values

locus raw name	locus name	abs(mds1)		abs(mds2)	
		estimate[a]	$-(\log_{10} q)$[b]	estimate[a]	$-(\log_{10} q)$[b]
mtDNA3	mtDNA3[c]	9.83	6.13	41.92	14.92
U502241*	*Oke_U502-241*	7.04	2.68	34.78	18.61
GnRH373	*Oke_GnRH-373*	0.43	0.12	26.25	25.07
ras1362	*Oke_ras1-249*	3.84	2.26	18.84	15.88
TCP178	*Oke_TCP1-78*	3.50	3.52	18.44	25.07
U22	*Oke_u200-385*	3.62	1.56	17.69	11.43
Tf278*	*Oke_Tf-278*	13.27	13.06	17.32	11.66
U506110*	*Oke_U506-110*	13.07	14.42	16.06	11.87
IL8r406	*Oke_IL8r-406*	3.90	4.76	15.62	22.51
DM20548	*Oke_DM20-548*	0.27	0.09	15.34	13.59
U217	*Oke_u217-172*	9.31	13.02	14.63	15.02
GHII3129	*Oke_GHII-3129*	13.05	22.84	14.36	16.23
U505112	*Oke_U505-112*	3.94	3.85	12.93	14.46
u1519	*Oke_u1-519*	9.95	10.38	12.44	8.49
GPDH	*Oke_GPDH-191*	7.97	11.28	11.93	12.21
AHR178	*Oke_AhR1-78*	0.75	0.31	11.89	12.10
U212	*Oke_u212-87*	0.88	0.89	11.31	25.07
CKS389	*Oke_CKS-389*	2.79	1.19	10.00	5.15
IL1RA	*Oke_il-1racp-67*	2.11	1.38	9.67	9.21
RFC2*	*Oke_RFC2-618*	15.49	23.68	8.75	6.26
U504228	*Oke_U504-228*	9.09	8.52	8.26	4.08
COPA*	*Oke_copa-211*	0.45	0.11	8.05	3.35
ZAN132	*Oke_U305-130*	2.50	1.35	8.01	4.95
U507286	*Oke_U507-286*	4.61	2.66	7.98	3.65
serpin	*Oke_serpin-140*	8.49	7.99	7.33	3.54
CTS1627	*Oke_PP2A-635*	6.62	4.07	7.16	2.62
TSHA1	*Oke_Tsha1-196*	2.53	1.30	6.89	3.63
U514150	*Oke_U514-150*	1.47	1.02	6.63	6.42
RH1OP	*Oke_RH1op-245*	7.08	15.95	6.21	7.85
U216	*Oke_u216-222*	1.91	3.23	6.01	11.65
CCT3143	*Oke_U401-143*	2.97	5.92	5.95	10.35
HP182	*Oke_HP-182*	3.77	2.92	5.94	3.47
GPH105	*Oke_GPH-105*	2.03	0.70	5.31	1.69
U511271	*Oke_U511-271*	3.10	3.83	5.05	4.78
U302195	*Oke_U302-195*	5.48	3.96	4.91	1.85
ctgf105	*Oke_ctgf-105*	3.47	6.46	4.17	4.93
hnRNPL239	*Oke_hnRNPL-239*	2.07	2.44	3.93	3.90
MOESIN	*Oke_Moesin-160*	3.55	11.04	3.72	6.77
U509219	*Oke_U509-219*	4.66	2.88	3.32	0.99
MARCKS362	*Oke_MARCKS-362*	13.58	11.23	3.09	0.66
GnRH527	*Oke_GnRH-527*	4.47	4.19	3.03	1.30
ARF	*Oke_arf-319*	4.30	5.07	2.81	1.42
GPH78	*Oke_GPH-78*	0.76	0.31	2.53	1.06

U510204	*Oke_U510-204*	2.30	2.19	2.43	1.35
FARSLA242	*Oke_FARSLA-242*	5.23	9.87	1.86	1.07
EIF4EB	*Oke_eif4ebp2-64*	5.64	13.79	1.85	1.35
MAPK1135	*Oke_MAPK1-135*	3.70	5.41	1.78	0.99
ZAN311	*Oke_U305-307*	4.70	9.02	1.52	0.88
IGFI1	*Oke_hsc71-199*	1.42	3.25	1.22	1.42
U503272	*Oke_U503-272*	2.76	13.06	1.16	1.84
U202	*Oke_u202-131*	5.24	12.89	1.07	0.65
KPNA287	*Oke_KPNA2-87*	2.36	0.98	0.75	0.15
AhR1278	*Oke_AhR1-278*	0.26	0.26	0.61	0.53
HSP90BA299	*Oke_HSP90BA-299*	0.70	2.25	0.39	0.57

The markers are ordered by the regression coefficients for mds2, which characterized the differentiation of Japanese/Korean populations.
[a] Absolute values of the regression coefficients.
[b] *p*-values adjusted by the Benjamini & Hochberg's method.
[c] *Oke-Cr30*, *Oke-Cr386*, *Oke-ND3-69* were combined (Seeb et al., 2011).
[*] Five outlier loci found in the original study (Seeb et al., 2011).

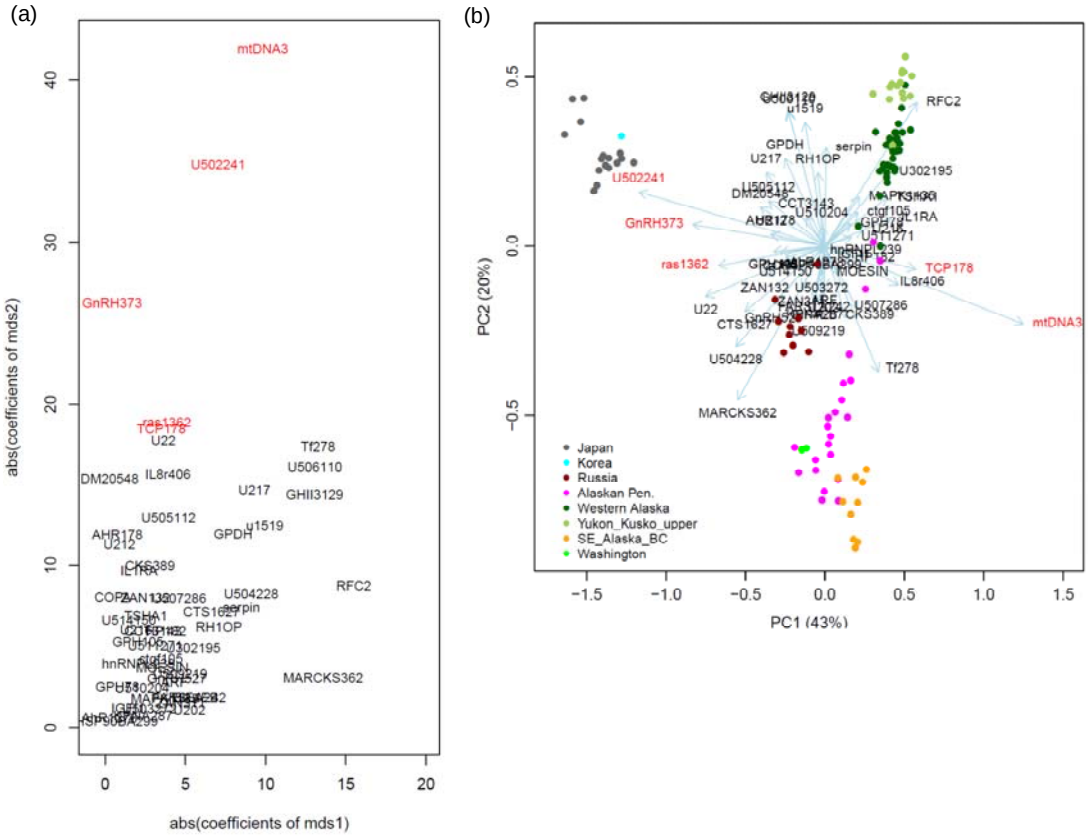

FIGURE 3 SNPs that characterized the distinctiveness of Japanese/Korean chum salmon populations. (a) Absolute values of the regression coefficients of mds1 vs those of mds2 on the same scale. Red indicates the outlier SNPs that highly diverged beyond the neutral population structure. (b) Principal component analysis (PCA) biplot, based on 53 nuclear SNPs and the combined mtDNA3 locus.

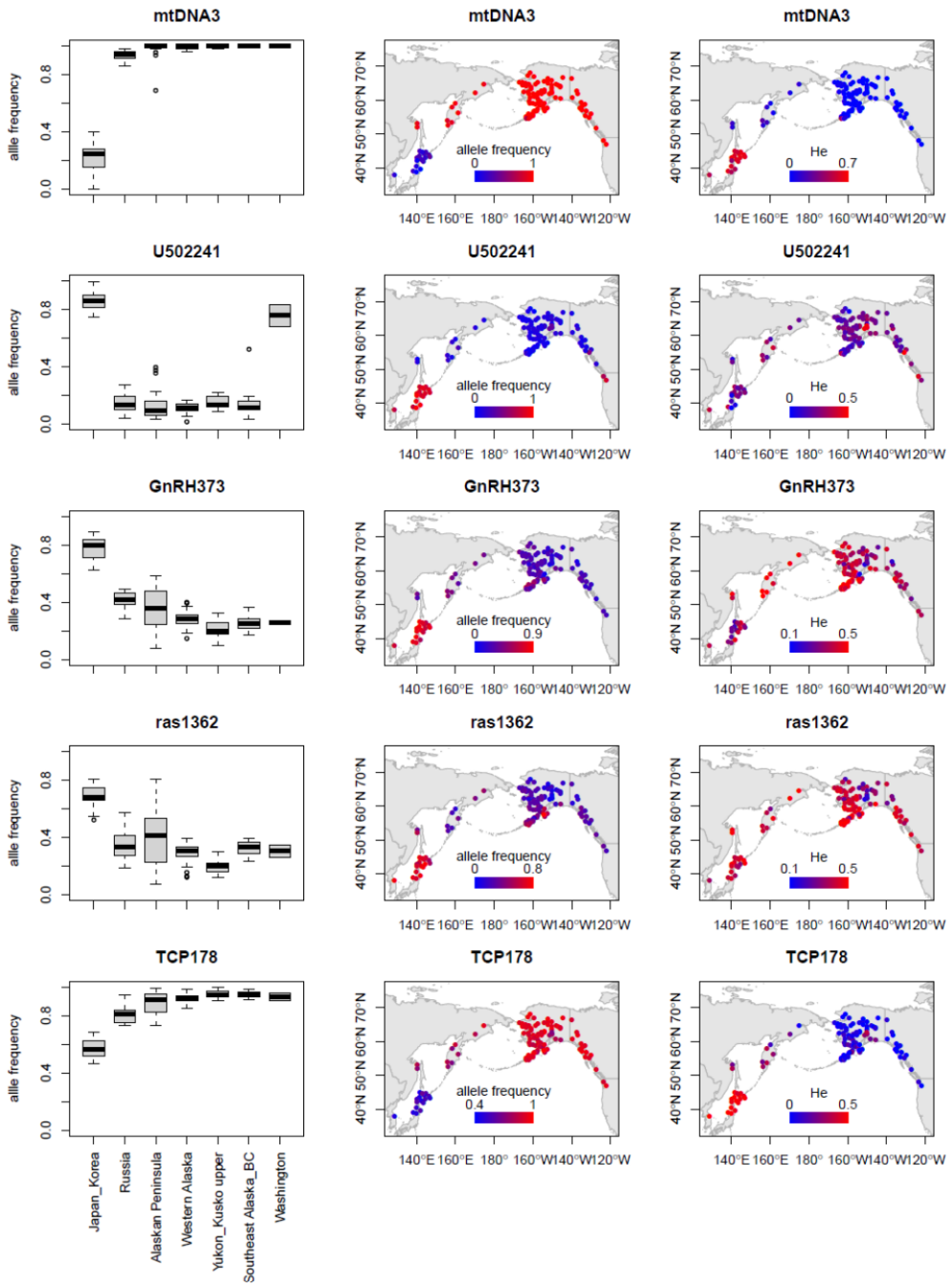

FIGURE 4 Geographical distribution of the top five outlier SNPs that characterized the distinctiveness of Japanese/Korean chum salmon populations. The points on the maps show allele frequencies (center) and expected heterozygosity (*He*) (right) at sampling locations, with a continuous color gradient ranging from blue to red for the lowest and highest *He* values, respectively.

4 | DISCUSSION

The inferred population structure based on microsatellites was consistent with the one obtained by the TreeMix analysis (Figure 1). The mds1 of pairwise F_{ST} identified the

north–south cline in American and Russian populations, and mds2 clearly showed divergence of Japanese/Korean populations, in agreement with the original study (Beacham et al., 2009a). TreeMix indicated that modern chum salmon populations may be recolonized from source populations between the Alaskan Peninsula and SEA/BC/WA, and expanded to the north and to Asia. However, a higher level of drift among the Japanese/Korean populations was clear relative to the Russian and American populations, suggesting that the Japanese/Korean populations were the most diverged from the ancestral population and that they diverged between Russia and the Alaskan Peninsula. Two inferred admixture events may reflect the stray of Japanese/Korean populations into Russian and Alaskan rivers.

The SNP allele frequency spectra (Figure S7) showed a similar pattern in all areas except Western Alaska. Because the entire range was represented for the SNP selection from Japan to WA (Seeb et al., 2011), ascertainment bias caused by cut-off of minor alleles may have occurred in Western Alaska, as indicated in Figure S7. This bias may have increased in intermediate allele frequencies and genetic diversity in the Western Alaska samples relative to other samples, leading to an increase in the genetic diversity of SNPs in the Western Alaska samples (Figure S8b), although the effects of the ascertainment bias were expected to be minimal within Alaska (Seeb et al., 2011).

Our regression analysis identified the top five outlier gene loci (Table 1, Figures 3, 4) that were significantly differentiated beyond the neutral population structure in Japanese/Korean populations, namely, mtDNA3, U502241, GnRH373, ras1362, and TCP178. The scatter plot of mds1 and mds2 indicated that most of the SNPs were differentiated beyond the neutral population structure. Selection can be induced by natural and anthropogenic pressures. The mds1 described the north–south cline of population expansion (Figure 1). The maximum regression coefficient of mds1 was 15.5 for RFC2 (Table 1). In contrast, the top five outlier gene loci had mds2 coefficients >18.4, which suggested that the allele frequencies of the top five gene loci were the result of anthropogenic selection pressures in Japanese/Korean populations.

mtDNA3 (a combined locus of the control region and NADH-3, Smith et al., 2005b) was the most differentiated gene locus (Figures 3, 4). mtDNA encodes some of the proteins in the oxidative phosphorylation enzymatic complex and plays a key role in aerobic ATP production, by contributing to the ability to respond to endurance exercise training (reviewed by Stefàno et al., 2019). The mtDNA control region functions in dNTP metabolism (Nicholls & Minczuk, 2014) and oxygen consumption (Kong et al., 2020). Subunits of the NADH dehydrogenase complex encoded by mtDNA are involved in the pathway from NADH to oxygen (Rivera et al., 1998). A significant reduction in mtDNA3 allele frequencies might reduce the efficiency of aerobic exercise ability and endurance, energy metabolism, and oxygen consumption. These results are consistent with the lower swimming endurance of Japanese hatchery-born chum salmon fry measured in a stamina measurement experimental device in the Japanese oldest Chitose Hatchery, Hokkaido (Kobayashi & Ohkuma, 1983). Wild chum salmon fry had a 1.4-fold higher swimming ability (56.6 ± 11.1 cm/s) than hatchery-reared fry (41.4 ± 12.3) (t = 2.45, df = 8.7, p = 0.038) in the early 1980s.

U502241 allele frequencies were significantly higher in Japanese/Korean and Washington

populations (Figure 4). Elevated levels of U502241 allele frequencies in Japanese/Korean and Washington populations were similar ($t = 1.3, \mathrm{df} = 1.1, p = 0.39$), suggesting a common effect. The original study (Seeb et al. 2011) found this phenomenon and detected U502241 as an outlier SNP. No function has been associated with the chum U502241 locus so far (Elfstrom et al., 2007), but the Atlantic salmon immunoglobulin IgH locus B genomic sequence was the most likely match for U502241 (Seeb et al. 2011). Immunoglobulins initiate immune responses (Gene Cards).

GnRH (gonadotropin-releasing hormone) is a key regulator of vertebrate reproduction, including that of salmonids (Khakoo et al., 1994). TCP1 (T-Complex 1) is related to sperm–zona–pellucida interaction and spermatozoan fertilization ability (Dun et al., 2011). The *ras* gene has a central role in cell growth (Rotchell et al., 2001). These outliers related to reproduction and growth are key parameters in hatchery production. In chum salmon, GnRH is involved in gonadal maturation during the early and final phases of upstream migration (Kudo et al., 1996). GnRH also improved stream odor discrimination in adult fish (Ueda, 2019), and expression of GnRH increased in the brain during homing migration (Ueda et al., 2016). The differentiation of GnRH373 (Figure 4) might influence homing migration timing, potentially leading to the current introgression from Japanese into Russian populations, as suggested by our TreeMix analysis.

Apparently different distributions in *He* values were observed in microsatellites and SNPs in the Japanese/Korean populations (Figure S8). In the Japanese populations, the highest microsatellite *He* values may have been influenced by the transplantation history in Japanese river populations (Beacham et al., 2008a), whereas the lowest SNP *He* values may be a consequence of artificial selection in the hatcheries. The *He* values in Japanese/Korean and Western Alaskan populations were similar for U502241 and ras1362, whereas the *He* values for GnRH373 were smaller in the Japanese/Korean populations than in the Western Alaskan populations as shown in Figure 4. In contrast, the *He* values in Japanese/Korean populations were significantly higher than those in Western Alaskan populations for mtDNA3 and TCP178, suggesting increases in heterozygosity. These results suggested genotype-specific effects on fitness.

The analyses in this study relied on limited data sets. Clearly, further genomic studies are needed to obtain more precise views of the demographic history, and environmental and anthropogenic selection of chum salmon. Population genomics studies can also provide finer resolution of GSI and will be useful for management and conservation of chum salmon.

ACKNOWLEDGEMENTS

This research was made possible by the SNP genotypes of chum salmon throughout the distribution range maintained by Lisa W. Seeb and her colleagues, and by the North Pacific baseline microsatellite allele frequencies maintained by Terry D. Beacham and his colleagues at Fisheries and Oceans Canada. We also appreciate researchers and staff for their substantial efforts to obtain baseline samples. This study was supported by the Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research KAKENHI (nos. 18K0578116 to SK and 19H04070 to HK).

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY STATEMENT

The authors affirm that all data necessary for confirming the conclusions of this article are present within the article, figures, and supporting information.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proceedings of the 2nd International Symposium on Information Theory*, pp. 267–281. Ed. by B. N. Petrov, & F. Caski. Akadimiai Kiado, Budapest.

Amoroso, R. O., Tillotson, M. D., & Hilborn, R. (2017). Measuring the net biological impact of fisheries enhancement: Pink salmon hatcheries can increase yield, but with apparent costs to wild populations. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(8), 1233–1242. https://doi.org/10.1139/cjfas-2016-0334

Beacham, T. D., Sato, S., Urawa, S., Le, K. D., & Wetklo, M. (2008a). Population structure and stock identification of chum salmon *Oncorhynchus keta* from Japan determined by microsatellite DNA variation. *Fisheries Science*, 74(5), 983–994. https://doi.org/10.1111/j.1444-2906.2008.01616.x

Beacham, T. D., Varnavskaya, N. V., Le, K. D., & Wetklo, M. H. (2008b). Determination of population structure and stock composition of chum salmon (*Oncorhynchus keta*) in Russia determined with microsatellites. *Fishery Bulletin*, 106(3), 245–256. http://fishbull.noaa.gov/1063/beacham.pdf

Beacham, T. D., Wetklo, M., Wallace, C., Olsen, J. B., Flannery, B. G., Wenburg, J. K., ... & Seeb, L. W. (2008c). The application of microsatellites for stock identification of Yukon River Chinook salmon. *North American Journal of Fisheries Management*, 28(1), 283–295. https://doi.org/10.1577/M06-253.1

Beacham, T. D., Winther, I., Jonsen, K. L., Wetklo, M., Deng, L., & Candy, J. R. (2008d). The application of rapid microsatellite based stock identification to management of a Chinook salmon troll fishery off the Queen Charlotte Islands, British Columbia. *North American Journal of Fisheries Management*, 28(3), 849–855. https://doi.org/10.1577/M06-167.1

Beacham, T. D., Candy, J. R., Le, K. D., & Wetklo, M. (2009a). Population structure of chum salmon (*Oncorhynchus keta*) across the Pacific Rim, determined from microsatellite analysis. *Fishery Bulletin*, 107(2), 244–260. http://fishbull.noaa.gov/1072/beacham.pdf

Beacham, T. D., Candy, J. R., Wallace, C., Urawa, S., Sato, S., Varnavskaya, N. V., Le, K. D., & Wetklo, M. (2009b). Microsatellite stock identification of chum salmon on a Pacific Rim basis. *North American Journal of Fisheries Management*, 29(6), 1757-1776. https://doi.org/10.1577/M08-188.1

Beacham, T. D., Wallace, C., Jonsen, K., McIntosh, B., Candy, J. R., Rondeau, E. B., ... & Withler, R. E. (2020). Accurate estimation of conservation unit contribution to coho salmon mixed-stock fisheries in British Columbia, Canada, using direct DNA sequencing for single nucleotide polymorphisms. *Canadian Journal of Fisheries and Aquatic Sciences*, 77(8), 1302–1315. https://doi.org/10.1139/cjfas-2019-0339

Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M., Beacham, T. D., ... & Withler, R. E. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends in Ecology & Evolution*, 32(9), 665–680. https://doi.org/10.1016/j.tree.2017.06.010

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Dun, M. D., Smith, N. D., Baker, M. A., Lin, M., Aitken, R. J., & Nixon, B. (2011). The chaperonin containing TCP1 complex (CCT/TRiC) is involved in mediating sperm-oocyte interaction. *Journal of Biological Chemistry*, 286(42), 36875–36887. https://doi.org/10.1074/jbc.M110.188888

Elfstrom, C. M., Smith, C. T., & Seeb, L. W. (2007). Thirty□eight single nucleotide polymorphism markers for high□throughput genotyping of chum salmon. *Molecular Ecology Notes*, 7(6), 1211–1215. https://doi.org/10.1111/j.1471-8286.2007.01835.x

Flagg, T. A. (2015). Balancing conservation and harvest objectives: a review of considerations for the management of salmon hatcheries in the US Pacific Northwest. *North American Journal of Aquaculture*, 77(3), 367–376. https://doi.org/10.1080/15222055.2015.1044058

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, 27(9), 489–496. https://doi.org/10.1016/j.tree.2012.05.012

Garvin, M. R., Saitoh, K., Churikov, D. Y., Brykov, V. A., & Gharrett, A. J. (2010). Single nucleotide polymorphisms in chum salmon (*Oncorhynchus keta*) mitochondrial DNA derived from restriction site haplotype information. *Genome*, 53(7), 501–507. https://doi.org/10.1139/G10-026

Garvin, M. R., Kondzela, C. M., Martin, P. C., Finney, B., Guyon, J., Templin, W. D., DeCovich, N., Gilk-Baumer, S., & Gharrett, A. J. (2013). Recent physical connections may explain weak genetic structure in western Alaskan chum salmon (*Oncorhynchus keta*) populations. *Ecology and Evolution*, 3(7), 2362–2377. https://doi.org/10.1002/ece3.628

Khakoo, Z., Bhatia, A. Gedamu, L., & Habibi, H. R. (1994). Functional specificity for salmon gonadotropin-releasing hormone (GnRH) and chicken GnRH-II coupled to the gonadotropin release and subunit messenger ribonucleic acid level in the goldfish pituitary. *Endocrinology*, 134(2), 838–847. https://doi.org/10.1210/en.134.2.838

Kijima, A., & Fujio, Y. (1979). Geographical distribution of IDH and LDH isozymes in chum salmon population. *Nippon Suisan Gakkaishi*, 45(3), 287–295. (In Japanese with English abstract) https://doi.org/10.2331/suisan.45.287

Kitada, S. (2018). Economic, ecological and genetic impacts of marine stock enhancement and sea ranching: A systematic review. *Fish and Fisheries*, 19(3), 511–532. https://doi.org/10.1111/faf.12271

Kitada, S., Nakamichi, R., & Kishino, H. (2017). The empirical Bayes estimators of fine scale population structure in high gene flow species. *Molecular Ecology*, 17(6), 1210–1222. https://doi.org/10.1111/1755-0998.12663

Kobayashi, T., & Ohkuma, K. (1983). On the device for stamina measurement of salmon fry. *Scientific Reports of the Hokkaido Salmon Hatchery*, 37, 41–44. (in Japanese) http://salmon.fra.affrc.go.jp/kankobutu/srhsh/data/srhsh297.htm

Kong, M., Xiang, H., Wang, J., Liu, J., Zhang, X., & Zhao, X. (2020). Mitochondrial DNA Haplotypes Influence Energy Metabolism across Chicken Transmitochondrial Cybrids. *Genes*, 11(1), 100. https://doi.org/10.3390/genes11010100

Kudo, H., Hyodo, S., Ueda, H., Hiroi, O., Aida, K., Urano, A., & Yamauchi, K. (1996). Cytophysiology of gonadotropin-releasing-hormone neurons in chum salmon (*Oncorhynchus keta*) forebrain before and after upstream migration. *Cell and Tissue Research*, 284(2), 261–267. https://doi.org/10.1007/s004410050586

Larson, W. A., Seeb, J. E., Pascal, C. E., Templin, W. D., & Seeb, L. W. 2014. Singlenucleotide polymorphisms (SNPs) identified through genotyping-by-sequencing improve genetic stock identification of Chinook salmon (*Oncorhynchus tshawytscha*) from western Alaska. *Canadian Journal of Fisheries and Aquatic Sciences*, 71(5), 698–708. https://doi.org/10.1139/cjfas-2013-0502

McKinney, G., McPhee, M. V., Pascal, C., Seeb, J. E., & Seeb, L. W. (2020). Network analysis of linkage disequilibrium reveals genome architecture in Chum Salmon. *G3: Genes, Genomes, Genetics*, 10(5), 1553–1561. https://doi.org/10.1534/g3.119.400972

Moore, J. S., Bourret, V., Dionne, M., Bradbury, I., O'Reilly, P., Kent, M., ... & Bernatchez, L. (2014). Conservation genomics of anadromous Atlantic salmon across its North American range: outlier loci identify the same patterns of population structure as neutral loci. *Molecular Ecology*, 23(23), 5680–5697. https://doi.org/10.1111/mec.12972

Myers, K. W., Klovach, N. V., Gritsenko, O. F., Urawa, S., & Royer, T. C. (2007). Stock-specific distributions of Asian and North American salmon in the open ocean, interannual changes, and oceanographic conditions. *North Pacific Anadromous Fish Commission Bulletin*, 4, 159–177. Available at https://npafc.org

Naish, K. A., Taylor III, J. E., Levin, P. S., Quinn, T. P., Winton, J. R., Huppert, D., & Hilborn, R. (2007). An evaluation of the effects of conservation and fishery enhancement hatcheries on wild populations of salmon. *Advances in Marine Biology*, 53, 61–194. https://doi.org/10.1016/S0065-2881(07)53002-6

Nei, M., & Chesser, R. K. (1983). Estimation of fixation indices and gene diversities. *Annals of Human Genetics*, 47(3), 253–259. https://doi.org/10.1111/j.1469-1809.1983.tb00993.x

Nicholls, T. J., & Minczuk, M. (2014). In D-loop: 40 years of mitochondrial 7S DNA. *Experimental Gerontology*, 56, 175–181. https://doi.org/10.1016/j.exger.2014.03.027

Nielsen, R., Hubisz, M. J., & Clark, A. G. (2004). Reconstituting the frequency spectrum of ascertained single-nucleotide polymorphism data. *Genetics*, 168(4), 2373–2382. https://doi.org/10.1534/genetics.104.031039

Nielsen, R., Bustamante, C., Clark, A. G., Glanowski, S., Sackton, T. B., Hubisz, M. J., Fledel-Alon, A., Tanenbaum. D.M., Civello, D., White, T.J., Sninsky, J. J., Adams, M.D., Cargill, M. (2005). A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol*, 3(6), e170. https://doi.org/10.1371/journal.pbio.0030170

Okazaki, T. (1982) Geographical distribution of allelic variations of enzymes in chum salmon *Oncorhynchus keta*, river populations of Japan and the effects of transplantation. *Nippon Suisan Gakkaishi*, 48(11), 1525–1535. https://doi.org/10.2331/suisan.48.1525

Olsen, J. B., Flannery, B. G., Beacham, T. D., Bromaghin, J. F., Crane, P. A., Lean, C. F., Dunmall, K.M., & Wenburg, J. K. (2008). The influence of hydrographic structure

and seasonal run timing on genetic diversity and isolation-by-distance in chum salmon (*Oncorhynchus keta*). *Canadian Journal of Fisheries and Aquatic Sciences*, 65(9), 2026–2042. https://doi.org/10.1139/F08-108

Park, L. K., Brainard, M. A., Dightman, D. A., & Winans, G. A. (1993). Low levels of intraspecific variation in the mitochondrial DNA of chum salmon (*Oncorhynchus keta*). *Molecular Marine Biology and Biotechnology*, 2(6), 362–370. PMID: 7910770

Petrou, E. L., Seeb, J. E., Hauser, L., Witteveen, M. J., Templin, W. D., & Seeb, L. W. (2014). Fine-scale sampling reveals distinct isolation by distance patterns in chum salmon (*Oncorhynchus keta*) populations occupying a glacially dynamic environment. *Conservation Genetics*, 15(1), 229–243. DOI:10.1007/s10592-013-0534-3

Pickrell, J., & Pritchard, J. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, 8(11), e1002967. https://doi.org/10.1038/npre.2012.6956.1

Rivera, M. A., Wolfarth, B., Dionne, F. T., Chagnon, M., Simoneau, J. A., Boulay, M. R., ... & Bouchard, C. (1998). Three mitochondrial DNA restriction polymorphisms in elite endurance athletes and sedentary controls. *Medicine and Science in Sports and Exercise*, 30(5), 687–690. https://doi.org/10.1097/00005768-199805000-00007

Rousset, F. (2008). genepop'007: a complete re□implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, 8(1), 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Rotchell, J. M., Lee, J. S., Chipman, J. K., & Ostrander, G. K. (2001). Structure, expression and activation of fish ras genes. *Aquatic Toxicology*, 55(1-2), 1–21. https://doi.org/10.1016/S0166-445X(01)00214-4

Saito, T., Honda, K., Sasaki, K., Watanabe, K., Suzuki, K., Hirabayashi, Y., ... & Sato, S. (2020). Stock composition of adult chum salmon *Oncorhynchus keta* caught in a setnet fishery estimated using genetic identification, scale patterns, and otolith thermal marking. *Fisheries Science*, 86(2), 271–286. https://doi.org/10.1007/s12562-020-01401-9

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4), 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Salo, E. O. (1991). Life history of chum salmon (*Oncorhynchus keta*). *Pacific salmon life histories*, pp.231–309. Ed. by C. Groot, & L. Margolis.UBC Press, Vancouver.

Sato, S., Kojima, H., Ando, J., Ando, H., Wilmot, R. L., Seeb, L. W., Efremov, V., LeClair, L., Buchholz, W., Jin, D. H., Urawa, S., Kaeriyama, M., Urano, A., & Abe, S. (2004). Genetic population structure of chum salmon in the Pacific Rim inferred from mitochondrial DNA sequence variation. *Environmental Biology of Fishes*, 69(1), 37–50. https://doi.org/10.1023/B:EBFI.0000022881.90237.aa

Sato, S., Templin, W. D., Seeb, L. W., Seeb, J. E., & Urawa, S. (2014). Genetic structure and diversity of Japanese chum salmon populations inferred from single nucleotide polymorphism markers. *Transactions of the American Fisheries Society*, 143(5), 1231–1246. https://doi.org/10.1080/00028487.2014.901251

Sato, S., & Urawa, S. (2015). Genetic structure of chum salmon populations in Japan. *Bulletin of Fishery Research Agency*, 39, 21–470. (in Japanese with English abstract) Available at http://www.fra.affrc.go.jp/bulletin/thesis.html#bull

Seeb, L. W., Crane, P. A. & Gates, R. B. (1995). Progress Report of Genetic Studies of Pacific Rim Chum Salmon, and Preliminary Analysis of the 1993 and 1994 South Unimak June Fisheries. Anchorage, USA: Alaska Department of Fish and Game.

http://www.adfg.alaska.gov/fedaidpdfs/rir.5j.1995.07.pdf
Seeb, L. W., & Crane, P. A. (1999). High genetic heterogeneity in chum salmon in western Alaska, the contact zone between northern and southern lineages. *Transactions of the American Fisheries Society*, 128(1), 58–87. https://doi.org/10.1577/1548-8659(1999)128<0058:HGHICS>2.0.CO;2
Seeb, L. W., Crane, P. A., Kondzela, C. M., Wilmot, R. L., Urawa, S., Varnavskaya, N. V., & Seeb, J. E. (2004). Migration of Pacificrim chum salmon on the high seas: insights from genetic data. *Environmental Biology of Fishes*, 69(1–4), 21–36. https://doi.org/10.1007/978-94-007-0983-6_3
Seeb, L. W., Templin, W. D., Sato, S., Abe, S., Warheit, K., Park, J. Y., & Seeb, J. E. (2011). Single nucleotide polymorphisms across a species' range: implications for conservation studies of Pacific salmon. *Molecular Ecology Resources*, 11 (Suppl. 1), 195–217. https://doi.org/10.1111/j.1755-0998.2010.02966.x
Shaklee, J. B., Beacham, T. D., Seeb, L., & White, B. A. (1999). Managing fisheries using genetic data: case studies from four species of Pacific salmon. *Fisheries Research*, 43(1-3), 45–78. https://doi.org/10.1016/S0165-7836(99)00066-1
Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science*, 236(4803), 787–792. https://doi.org/10.1126/science.3576198
Small, M. P., Rogers Olive, S. D., Seeb, L. W., Seeb, J. E., Pascal, C. E., Warheit, K. I., & Templin, W. (2015). Chum salmon genetic diversity in the northeastern Pacific Ocean assessed with single nucleotide polymorphisms (SNPs): Applications to fishery management. *North American Journal of Fisheries Management*, 35(5), 974–987. https://doi.org/10.1080/02755947.2015.1055014
Smith, C. T., Elfstrom, C. M., Seeb, L. W., & Seeb, J. E. (2005a). Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. *Molecular Ecology*, 14(13), 4193–4203. https://doi.org/10.1111/j.1365-294X.2005.02731.x
Smith, C. T., Baker, J., Park, L., Seeb, L. W., Elfstrom, C. M., Abe, S., Seeb, J. E. (2005b) Characterization of 13 single nucleotide polymorphism markers for chum salmon. *Molecular Ecology Notes*, 5(2), 259–262. https://doi.org/10.1111/j.1471-8286.2005.00903.x
Stefàno, E., Marsigliante, S., Vetrugno, C., & Muscella, A. (2019). Is mitochondrial DNA profiling predictive for athletic performance?. *Mitochondrion*, 47, 125–138. https://doi.org/10.1016/j.mito.2019.06.004
Taylor, E. B., Beacham, T. D., & Kaeriyama, M. (1994). Population structure and identification of North Pacific Ocean chum salmon (*Oncorhynchus keta*) revealed by an analysis of minisateliite DNA variation. *Canadian Journal of Fisheries and Aquatic Sciences*, 51(6), 1430–1442. https://doi.org/10.1139/f94-143
Ueda, H. (2019). Sensory mechanisms of natal stream imprinting and homing in *Oncorhynchus* spp. *Journal of Fish Biology*, 95(1), 293–303. https://doi.org/10.1111/jfb.13775
Ueda, H., Nakamura, S., Nakamura, T., Inada, K., Okubo, T., Furukawa, N., ... & Watanabe, M. (2016). Involvement of hormones in olfactory imprinting and homing in chum salmon. *Scientific Reports*, 6(1), 1–13. https://doi.org/10.1038/srep21102
Urawa, S., Azumaya, T., Crane, P. A., & Seeb, L. W. (2005). Origins and distribution of chum salmon in the central Bering Sea. *North Pacific Anadromous Fish Commission Technical Report*, 6, 67–70.
Urawa, S., Sato, S., Crane, P. A., Agler, B., Josephson, R., & Azumaya, T. (2009).

Stock-specific ocean distribution and migration of chum salmon in the Bering Sea and North Pacific Ocean. *North Pacific Anadromous Fish Commission Bulletin*, 5, 131–146. https://npafc.org

Utter, F. M. (1991). Biochemical genetics and fishery management: an historical perspective. *Journal of Fish Biology*, 39, 1–20. https://doi.org/10.1111/j.1095-8649.1991.tb05063.x

Waples, R. S., & Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, 15(6), 1419–1439. https://doi.org/10.1111/j.1365-294X.2006.02890.x

Waples, R. S., Naish, K. A., & Primmer, C. R. (2020). Conservation and Management of Salmon in the Age of Genomics. *Annual Review of Animal Biosciences*, 8, 117–143. https://doi.org/10.1146/annurev-animal-021419-083617

Whitlock, M. C., & Mccauley, D. E. (1999). Indirect measures of gene flow and migration: $F_{\mathrm{ST}} \neq 1(4Nm + 1)$. *Heredity*, 82(2), 117–125. https://doi.org/10.1046/j.1365-2540.1999.00496.x

Wilmot, R. L., Everett, R. J., Spearman, W. J., Baccus, R., Varnavskaya, N. V., & Putivkin, S. V. (1994). Genetic stock structure of western Alaska chum salmon and a comparison with Russian Far East stocks. *Canadian Journal of Fisheries and Aquatic Sciences*, 51(S1), 84–94. https://doi.org/10.1139/f94-297

Winans, G. A., Aebersold, P. B., Urawa, S., & Varnavskaya, N. V. (1994). Determining continent of origin of chum salmon (*Oncorhynchus keta*) using genetic stock identification techniques: status of allozyme baseline in Asia. *Canadian Journal of Fisheries and Aquatic Sciences*, 51(S1), 95–113. https://doi.org/10.1139/f94-298

Yoon, M., Sato, S., Seeb, J. E., Brykov, V., Seeb, L. W., Varnavskaya, N. V., Wilmot, R. L., Jin, D.H., Urawa, S., Urano, A., & Abe, S. (2008). Mitochondrial DNA variation and genetic population structure of chum salmon *Oncorhynchus keta* around the Pacific Rim. *Journal of Fish Biology*, 73(5), 1256–1266. https://doi.org/10.1111/j.1095-8649.2008.01995.x