Journal of Heredity, 2021, 602–613
https://doi.org/10.1093/jhered/esab060
Original Article
Advance Access publication October 7, 2021

OXFORD

Original Article

Genome-Wide Investigation of the Multiple Origins Hypothesis for Deep-Spawning Kokanee Salmon (*Oncorhynchus nerka*) across its Pan-Pacific Distribution

Farida Samad-zada, Kouji Nakayama, and Michael A. Russello

From the Department of Biology, University of British Columbia, Kelowna, BC, Canada (Samad-zada and Russello); the Division of Applied Biosciences, Kyoto University, Kyoto, Japan (Nakayama).

Address correspondence to M. Russello at the address above, or e-mail: michael.russello@ubc.ca.

Received March 25, 2021; First decision May 15, 2021; Accepted October 1, 2021.

Corresponding Editor: Kim Andrews

Abstract

Salmonids have emerged as important study systems for investigating molecular processes underlying parallel evolution given their tremendous life history variation. Kokanee, the resident form of anadromous sockeye salmon (*Oncorhynchus nerka*), have evolved multiple times across the species' pan-Pacific distribution, exhibiting multiple reproductive ecotypes including those that spawn in streams, on lake-shores, and at lake depths >50 m. The latter has only been detected in 5 locations in Japan and British Columbia, Canada. Here, we investigated the multiple origins hypothesis for deep-spawning kokanee, using 9721 single nucleotide polymorphisms distributed across the genome analyzed for the vast majority of known populations in Japan (Saiko Lake) and Canada (Anderson, Seton, East Barrière Lakes) relative to stream-spawning populations in both regions. We detected 397 outlier loci, none of which were robustly identified in paired-ecotype comparisons in Japan and Canada independently. Bayesian clustering and principal components analyses based on neutral loci revealed 6 distinct clusters, largely associated with geography or translocation history, rather than ecotype. Moreover, a high level of divergence between Canadian and Japanese populations, and between deep- and stream-spawning populations regionally, suggests the deep-spawning ecotype independently evolved on the 2 continents. On a finer level, Japanese kokanee populations exhibited low estimates of heterozygosity, significant levels of inbreeding, and reduced effective population sizes relative to Canadian populations, likely associated with transplantation history. Along with preliminary evidence for hybridization between deep- and stream-spawning ecotypes in Saiko Lake, these findings should be considered within the context of on-going kokanee fisheries management in Japan.

Keywords: genomic scans, life history evolution, Pacific salmon, trait polymorphisms

Parallel evolution, the independent evolution of the same trait in closely related lineages, has been documented across a wide range of taxa (Haldane 1932; Wood et al. 2005). For example, short-winged ecomorphs of Japanese scorpionfly have independently evolved in geographically isolated populations as an adaptation to high-altitude conditions (Suzuki et al. 2019), adult dwarfism

has repeatedly emerged in Icelandic charr inhabiting volcanic springs (Macqueen et al. 2011), and sympatric, microhabitat-associated ecotypes of marine snails have independently originated several times (Quesada et al. 2007). Molecular mechanisms underlying parallel evolution have been traditionally demonstrated for phenotypic traits that are regulated by specific candidate genes or sets of genes, the functions of which have been known *a priori*. Examples include insecticide resistance in red flour beetles, where different point mutations in a resistance-conferring gene have evolved in geographically distinct populations (Andreev et al. 1999), and the depigmentation phenotype present in isolated populations of Mexican blind cavefish resulting from different mutations within the melanocortin receptor gene (Gross et al. 2009). In contrast, deciphering molecular mechanisms responsible for repeated evolution of phenotypic traits for which no prior genetic information is available only became feasible with advances in next-generation sequencing that made it possible to access loci spanning the entire genome, using such approaches as reduced representation and whole-genome sequencing (Fraser and Whiting 2020; Stern 2013). Over the past decade, genomic scans have been commonly employed to identify candidate genes responsible for the evolution of the same phenotypes in replicated populations (Deagle et al. 2012; Westram et al. 2014; Nichols et al. 2016; Veale and Russello 2017b).

Salmonids, including sockeye salmon (*Oncorhynchus nerka*), have become classic examples of parallel evolution of divergent ecotypes in response to ecological pressures, as demonstrated by their unusually diverse life history variation. For instance, multiple independent lineages of freshwater resident *O. nerk*a (known as kokanee) have evolved from anadromous sockeye salmon across the species' pan-Pacific distribution, spanning from western North America to East Asia from Russia (Kamchatka) to Japan (Kaeriyama et al. 1995; Taylor et al. 1996). Kokanee populations are genetically distinct from sympatric populations of anadromous sockeye salmon and can be further distinguished based on their habitat preferences, physiological traits, morphology, life history, and behavior (Foote et al. 1989; Taylor et al. 1996; Lemay and Russello 2015; Veale and Russello 2016; Veale and Russello 2017b). Moreover, these differences appear to be adaptive and have evolved due to differential selection pressures attributed to variation in their associated environments and life cycles (Foote et al. 1999). Both sockeye salmon and kokanee can be further subdivided into reproductive ecotypes, depending on spawning location (shore-, stream- and deep-spawning), which have repeatedly evolved in different parts of the world (Taylor et al. 1996; Moreira and Taylor 2015; Veale and Russello 2017b).

Among these ecotypes, deep-spawning kokanee (also known as black kokanee or kunimasu) can be distinguished from the more common shore- and stream-spawning kokanee based on its morphology, unusual spawning depth and pigmentation at maturity. While stream-spawners and, to a lesser extent, shore-spawners transition to bright or brown-red coloration upon maturation, deep-spawning kokanee remain olive-black, even during the spawning season (Nakabo et al. 2011, 2014; Moreira and Taylor 2015). For this ecotype, spawning occurs in shoreline habitat, but more than 50 m below the lake surface in the lower part of the profundal zone and continues throughout the winter months (Nakabo et al. 2014; Moreira and Taylor 2015). Moreover, deep-spawning kokanee can be differentiated from stream-spawning kokanee based on the number of pyloric caeca, fin rays, and gill rakers (Nakabo et al. 2014).

Geographically isolated populations of deep-spawning kokanee that share similar behavioural and morphological traits occur in Saiko and Motosu Lakes, Yamanashi Prefecture, Japan (Nakabo et al. 2011), as well as Anderson, Seton (Moreira and Taylor 2015), and East Barrière Lakes (Andrusak and Morris 2004) in British Columbia, Canada, leading to questions surrounding the origin and genetic signatures of parallel evolution associated with this reproductive form. To date, a handful of studies have investigated the morphology, genetics, and behavior of deep-spawning kokanee (Muto et al. 2013; Nakabo et al. 2011, 2014; Veale and Russello 2017b; Nakayama et al. 2018), only one of which combined genetic data for both Japanese and Canadian populations (Moreira and Taylor 2015). In Saiko Lake, genetic distinctiveness of sympatric deep- and stream-spawning kokanee has been previously demonstrated based on 5 microsatellite markers [*One102*, *One108*, *One110*, *One114*, *One115*; (Muto et al. 2013)]. Likewise, deep-spawning kokanee in Anderson and Seton Lakes were found to be genetically distinct from anadromous sockeye salmon in the connecting Portage Creek, based on 9 microsatellites [*Omy77*, *Ots100*, *Ots103*, *Ots108*, *Oki10*, *Oki29*, *One103*, *One108*, *One110*] and mitochondrial DNA (Moreira and Taylor 2015), as well as genome-wide analyses of single nucleotide polymorphisms (SNPs) (Veale and Russello 2017b). The only known incidences of hybridization between deep- and stream-spawning kokanee have been documented in Motosu Lake, Yamanashi Prefecture, Japan (located less than 20 km away from Saiko Lake) (Nakayama et al. 2018). On a broader scale, deep-spawning kokanee populations in Japan and Canada are not monophyletic, as demonstrated by microsatellite and mitochondrial DNA analysis, suggesting that this ecotype evolved as a result of multiple independent divergence events (i.e., multiple origins hypothesis; Moreira and Taylor 2015).

Here, we employed genotyping-by-sequencing to investigate the origin and reconstruct the evolutionary history of deep-spawning kokanee, including the vast majority of documented populations worldwide (Saiko Lake, Japan; Anderson, Seton and East Barrière Lakes, Canada) (Figure 1). Using genome-wide SNP data, we conducted genomic scans to identify outlier loci between paired deep- and stream-spawning populations in Canada and Japan to investigate putative genetic signatures of parallel evolution and ecotype divergence. We further reconstructed genome-wide patterns of within-population variation and among-population structure of the 4 deep-spawning populations relative to each other as well as to stream-spawning kokanee populations in their respective basins to explicitly test the multiple origins hypothesis. In addition, we genotyped all samples at a SNP within the leucine-rich repeat-containing protein 9 (LRRC9) gene, at which genotypes have previously been found to be associated with spawning behavior in *O. nerka* (Veale and Russello 2017b). We hypothesized that genotypically predicted spawning types would match the observed reproductive ecotypes within these populations (Table 1). Lastly, we investigated whether there was evidence for hybridization between deep- and stream-spawning kokanee in Saiko Lake, one of only 2 locations in the world where these ecotypes are known to co-occur.

Methods

Study Design and Sampling

Our study included deep-spawning kokanee (n = 20) and stream-spawning kokanee (n = 17) previously sampled in 2010–2011 from Saiko Lake, Yamanashi Prefecture, Japan, as well as stream-spawning kokanee (n = 21) from Akan Lake, Hokkaido Prefecture, Japan (Nakabo et al. 2011). The samples were originally collected by a combination of gill netting and angling, and were comprised of muscle tissue preserved in 100% ethanol (Nakabo et al. 2011). For

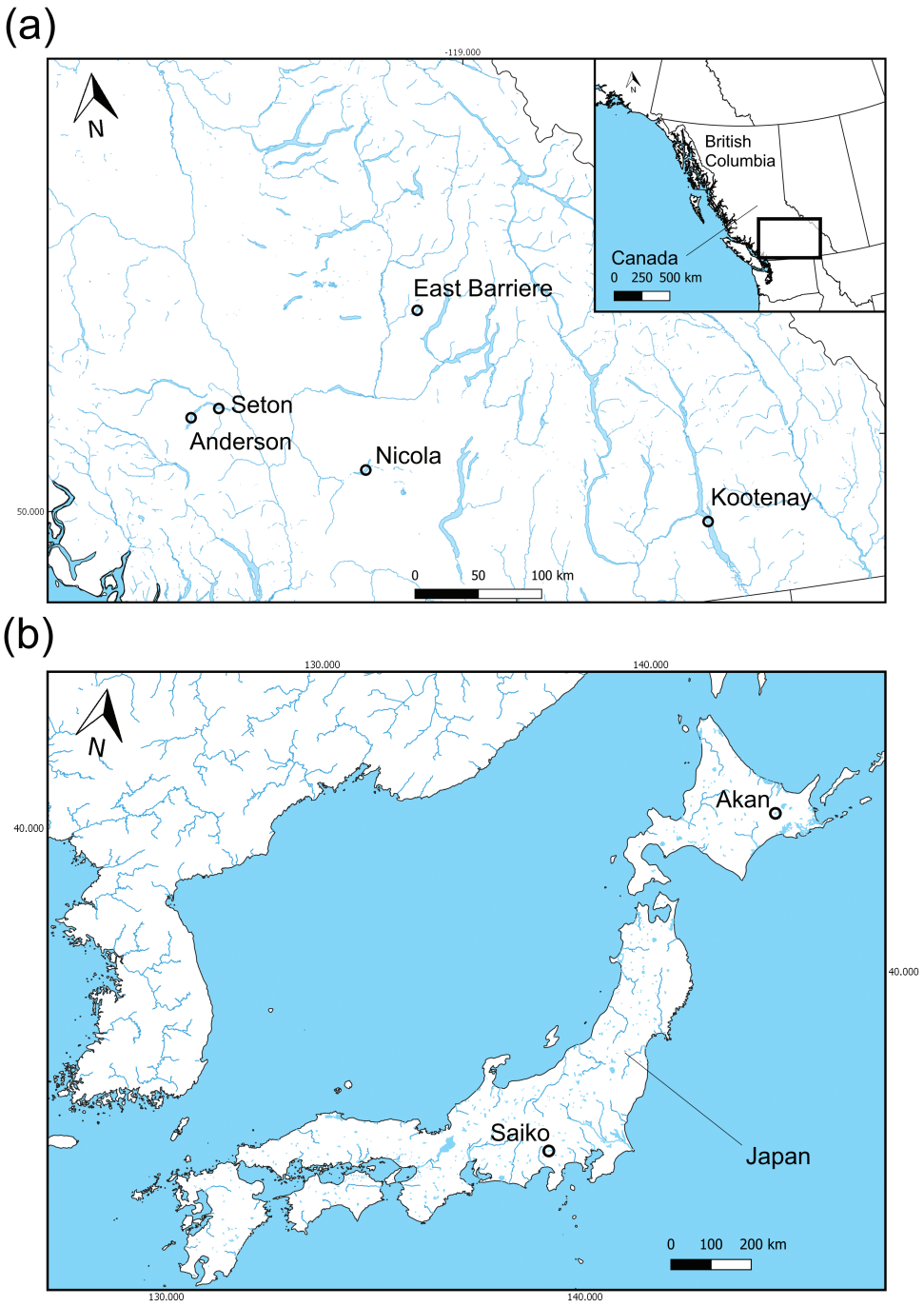

Figure 1. Locations of the study lakes. Maps produced using QGIS Geographic Information System, Open Source Geospatial Foundation Project (http://qgis.org). (**a**) Map of the Fraser River and Columbia River drainages, British Columbia, Canada. Shapefiles were retrieved from B.C. Data Catalogue (https://catalogue.data.gov.bc.ca/dataset). (**b**) Map of Japan. Shapefiles were retrieved from FAO Geonetwork (http://www.fao.org/geonetwork/srv/en/main.home).

deep- and stream-spawning kokanee in British Columbia, Canada we used previously collected genomic data from individuals from East Barrière (n = 31 plus 2 replicates) and Nicola (n = 25) Lakes (Samadzada et al. 2021), as well as Anderson (n = 22), Seton (n = 23), and Kootenay (n = 22) Lakes (Veale and Russello 2017b). Information regarding sample collection, DNA extraction, and library construction can be found in the original papers; complete distribution of kokanee populations used in this study, along with their respective ecotypes, is provided in Table 1 and Figure 1.

Library Preparation

Genomic DNA was extracted from muscle tissue using the Qiagen DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's protocol with the addition of 4 µL of 100 mg/mL 7000U RNase (Qiagen) prior to ethanol precipitation. We used restriction site-associated DNA sequencing (RADseq) to simultaneously identify and genotype SNPs within the *O. nerka* samples from Japan. Specifically, we employed a *SbfI* RADseq protocol following Baird et al. (2008) as modified in Lemay and Russello (2015) in order

Table 1. Sample size, spawner type, summary of diversity statistics, and estimated N_e for the 8 populations examined in this study

Population	Country	Spawner type	N_S	N_R	H_o	H_e	F_{IS}	N_e (95% CI)
Anderson	Canada	Deep-spawning	22	22	0.2153	0.212	−0.01573	2287.8 (1765.9–3245.4)
Seton	Canada	Deep-spawning	23	23	0.2174	0.2145	−0.01314	1463.4 (1247.2–1769.9)
East Barrière	Canada	Deep-spawning	31	31	0.2017	0.1921	−0.04993	814.5 (753.7–886.0)
Nicola	Canada	Stream-spawning	25	25	0.1685	0.1679	−0.00369	4194.2 (2594.0–10914.4)
Kootenay	Canada	Stream-spawning	22	22	0.1869	0.1888	0.01050	2101.7 (1615.0–3006.2)
Akan	Japan	Stream-spawning	21	15	0.0769	0.09	0.14569[a]	223.3 (196.6–258.1)
Saiko	Japan	Deep-spawning	20	20	0.1126	0.1212	0.07087[a]	30.4 (30.1–30.6)
Saiko	Japan	Stream-spawning	17	9	0.0737	0.0862	0.14437[a]	33.2 (32.1–34.4)

N_S, number of sequenced samples; N_R, number of samples retained after full filtering; CI, confidence interval; H_e, expected heterozygosity; H_o, observed heterozygosity; F_{IS}, inbreeding coefficient; N_e, effective population size.

[a]F_{IS} values that were significantly different from 0 at $\alpha = 0.01$.

to ensure direct compatibility with previously collected data from East Barrière and Nicola Lakes (Samad-zada et al. 2021), as well as Anderson, Seton, and Kootenay Lakes (Veale and Russello 2017b). We constructed 2 RADseq libraries that in total contained 58 unique individuals, 4 of which were replicated (2 per library) to estimate genotyping error within each library (Tintle et al. 2009). The constructed libraries were sequenced at the McGill University and Génome Québec Innovation Centre, using one lane each of Illumina HiSeq 2500 PE125 (2 lanes in total).

Genotyping and Filtering

We combined the newly sequenced reads from Akan and Saiko Lakes with those previously generated for samples from Anderson, Seton, and Kootenay Lakes (Illumina HiSeq 2000 SE110 sequencing; Veale and Russello 2017b) and East Barrière and Nicola Lakes (Illumina HiSeq 4000 PE150 sequencing; (Samad-zada et al. 2021) (Table 1). To ensure compatibility with previously sequenced samples, raw paired-end reads were demultiplexed and trimmed via the *process radtags* command (*-r -c -q -t 100*) in STACKS v2.41 (Catchen et al. 2011); all new data were trimmed to the same length (100 bp) as those available from the previously analyzed populations from Veale and Russello (2017b). After demultiplexing, samples containing less than 2 million reads were not used in the construction of the main SNP dataset, however, they were included in the Saiko-specific workflow (see below). Identical reads generated due to PCR amplification were removed using the *clone filter* command in STACKS v2.41 (Catchen et al. 2011). Demultiplexed, trimmed, and filtered reads were aligned against the *O. nerka* reference genome (*Oner_1*, GenBank Assembly Accession ID: GCA_006149115.1, Christensen et al. 2020) using the *bwa mem* algorithm in BWA (Li and Durbin 2009) and sorted using *samtools* v1.9 (Li et al. 2009). The resulting bam files together with those produced earlier for the samples from Anderson, Seton, East Barrière, Nicola, and Kootenay Lakes were used to identify loci and call SNPs via the *gstacks* command. Next, we processed the resulting loci through the *populations* module in STACKS v2.41 (Catchen et al. 2011). We performed a sensitivity analysis on the retained individuals by running the *populations* module in STACKS v2.41 (Catchen et al. 2011) with varying sets of parameters (*r, --min-maf)* to determine the optimal set for SNP ascertainment by comparing the number of retained SNPs and samples with the mean depth and missing data percentage (Supplementary Table S1). We ultimately chose a set of parameters that maximized the number of retained SNPs, while still ensuring that mean depth remained high and missing data percentage remained low. Specifically, we only retained loci observed in ≥70% (r70) of the individuals and present in all 8 populations, with a minimum allele frequency of 0.03 and a maximum observed heterozygosity of 0.50 (Supplementary Table S1). The *--write-single-snp* flag was used to only retain one SNP per locus in order to minimize linkage. Next, we calculated mean coverage per individual using VCFtools v0.1.15 (Danecek et al. 2011) and removed individuals with mean coverage <6× or mean missing data percentage >30%, and re-ran *populations* using the optimal parameters. We further processed this dataset through VCFtools v0.1.15 (Danecek et al. 2011) to only include sites with a minimum mean depth of 10× and a maximum mean depth of 100×, and exclude sites with >10% missing data (*--max-missing 0.9*). Additionally, we excluded loci identified as putative paralogs as described in McKinney et al. (2017) using the HDplot function available from https://github.com/gjmckinney/HDplot. Then, we estimated the genotyping error rate using a custom python script (https://github.com/bsjodin/genoerrorcalc) that compared genotypes for each SNP between the 2 replicates and calculated the percentage of SNPs that did not have the same genotype. Lastly, we removed replicate samples from our dataset.

Outlier Detection

In order to test for putative signatures of parallel evolution, we conducted outlier scans on "paired" (deep- vs. stream-spawning) populations from Canada and Japan. Canada-specific analyses were conducted on the following pairs of populations: Nicola-East Barrière, Nicola-Anderson, and Nicola-Seton, where the Nicola Lake population is stream-spawning, and the rest of the populations are deep-spawning kokanee. Japan-specific outlier scans compared Saiko deep-spawning to Saiko stream-spawning; and Saiko deep-spawning to Akan stream-spawning kokanee (see Supplementary Table S2 for details). Given the high false-positive rates associated with outlier detection approaches and the hierarchical structure of our dataset, we employed 3 different analyses including the F_{ST}-based approaches implemented in Bayescan v2.0 (Foll and Gaggiotti 2008) and Arlequin v3.5 (Excoffier and Lischer 2010), which is an extension of the FDIST method of Beaumont and Nichols (1996), as well as the principal component analysis (PCA)-based approach implemented in *pcadapt* (Luu et al. 2017). For BayeScan v2.0 (Foll and Gaggiotti 2008), we used a pairwise approach comparing allele frequencies between the above-mentioned pairs of populations ($n = 5$ comparisons; Supplementary Table S2). The analyses were run for 100 000 iterations with a 50 000 burn-in period and Prior Odds set to 10; loci with q values <0.05 were marked as outliers. For Arlequin, we used the hierarchical island model (Slatkin and Voelm 1991) that allows for higher migration rates within the group than

between the groups. We performed 20 000 simulations, with the number of simulated demes set to 100, number of simulated groups set to 10, and simulated derived allele frequency set to 0.05. We then corrected *P* values for each SNP; loci in the first quantile with FDR <0.05 were identified as outliers. Lastly, we inferred the number of genetic clusters through analyses of principal components (PCs), as implemented in *pcadapt* v4.1.0 (Luu et al. 2017), and Cuttel's rule to infer the most likely number of PCs that explained the genetic structure within the dataset. The resulting *P* values were corrected for multiple comparisons using the method of Benjamini-Yekutieli (2001), and loci with adjusted *P* values of less than 0.05 were considered outliers. We considered outliers to be robust if they were detected by 2 or more methods, an approach that has been commonly employed to decrease error rates (Alves-Pereira et al. 2020; de Villemereuil et al. 2014; Jordan et al. 2017). Robust outliers detected in Canada-specific outlier scans were compared to those detected in the Japan-specific outlier scans, and all outliers were functionally annotated using the *blastn* function in BLAST v2.9.0 (Altschul et al. 1990) with the *e*-value threshold of 1e-28, and *-entrez_query* set to "*Oncorhynchus*."

Additionally, we identified genes located within 100 kb of robust outliers (distance following Larson et al. 2017) using *bedtools* v2.26.0 (Quinlan and Hall 2010) to extract linkage group positions and BEDOPS v2.4.38 (Neph et al. 2012) *bedmap --echo-map-id-uniq* utility to locate genes within the specified ranges in the *O. nerka* genome annotation file (GCF_006149115.1_Oner_1.0); outliers located on unplaced scaffolds were excluded from this analysis. To obtain a list of gene descriptions, we used the NCBI Entrez direct *efetch* tool (Kans 2013). Gene symbols and/or gene descriptions were compared against the zebrafish (*Danio rerio*) genome annotation (GCF_000002035.6) and those that generated an exact match were used in gene ontology (GO) enrichment analysis. The *D. rerio* genome was chosen as a reference because this species contains the most up-to-date GO information, as well as the most functionally annotated genome of all teleost fishes (Babin et al. 2017). Next, 2 unranked gene sets were compared in GOrilla (Eden et al. 2009), where matching genes were treated as a target set, while all *D. rerio* genes were treated as a background set. The GOrilla analysis was conducted separately for Canada- and Japan-specific robust outlier sets.

Lastly, we specifically examined genotypes for all Japanese samples at a SNP in the LRRC9 gene (Veale and Russello 2017a) that was previously demonstrated to be a candidate locus under divergent selection associated with reproductive ecotype (shore- or stream-spawning). Genotypes at the LRRC9 gene have been previously reported for all Canadian populations examined in this study (Samad-zada et al. 2021; Veale and Russello 2017b).

Population Genetic Analyses

In order to help ensure a putatively neutral dataset for population genetic analyses, we conducted additional outlier detection analyses for all populations using the same 3 outlier detection approaches (BayeScan, *pcadapt,* Arlequin) and identical parameters as above (see Supplementary Table S2 for details); we removed any locus identified as an outlier in any comparison (i.e., Canada-specific, Japan-specific, all). Following outlier removal, we removed loci that significantly (*-h* 0.05) deviated from Hardy–Weinberg Equilibrium in 50% or more of the populations using the filter_hwe_by_pop.pl script from https://github.com/jpuritz/dDocent/tree/master/scripts.

Using the resulting putatively neutral SNP dataset, we calculated locus-specific observed (H_o) and expected (H_e) heterozygosity following Nei (1987) using the *basic.stats* function in *hierfstat* (Goudet and Jombart 2015), averaged across loci for each population. Additionally, we computed inbreeding coefficients (F_{IS}) for every population using Genetix v.4.05.2 (Belkhir et al. 2004) and 1000 permutations to determine levels of significance. We estimated effective population sizes (N_e) employing the linkage disequilibrium method (Waples and Do 2008) as implemented in NeEstimator v.2 (Do et al. 2014), with the lowest allele frequency set to 0.05. To examine the levels of pairwise population differentiation, we calculated Weir and Cockerham's (1984) θ between all pairs of populations, performing 1000 permutations in Genetix v.4.05.2 (Belkhir et al. 2004).

To assess the number of genetic clusters representing the populations in our dataset, we employed several approaches. First, we used the Bayesian method of Pritchard et al. (2000), as implemented in STRUCTURE v2.3.4. We set the run length to 100 000 Markov Chain Monte Carlo (MCMC) iterations after a burn-in period of 100 000 MCMC iterations, using correlated allele frequencies under the admixture model and the LOCPRIOR option. We varied the number of clusters (*K*) from 1 to 10, with 10 replicate runs for each. We then evaluated the output in STRUCTURE Harvester (Earl and vonHoldt 2012), where we inferred the optimal *K* value by employing a combination of the ΔK method (Evanno et al. 2005) and plotting of the log probability of the data (Pritchard et al. 2000) to estimate where ln Pr (*X*|*K*) plateaued. The results of the STRUCTURE output were visualized in CLUMPAK (Kopelman et al. 2015). Second, we conducted a PCA analysis in order to visualize the relationships among populations using SNPRelate v1.14.0 (Zheng et al. 2012).

In order to examine the relationship among the populations in our dataset, we used TreeMix (Pickrell and Pritchard 2012) that implements a model accounting for both population splitting and gene flow, and hence is a more appropriate method for estimating relationships among multiple populations of the same species compared to simple bifurcating trees. Given that TreeMix requires a non-linked set of markers without any missing data, this analysis was performed on a reduced dataset, where we removed SNPs that did not map to a Linkage Group (i.e., those mapped to the unplaced scaffolds on the reference genome), and only retained those that did not contain any missing data for any of the individuals [--max-missing 1 in VCFtools v0.1.15 (Danecek et al. 2011)]. To maximize the number of retained SNPs for this analysis, we removed 4 individuals that had more than 20% missing data. Lastly, SNPs were pruned using PLINK, with the window size set to 50 kb, step size set to 10, and r^2 threshold of 0.2 (Purcell et al. 2007).

Saiko Lake Hybridization

To assess the level of hybridization between deep- and stream-spawning kokanee in Saiko Lake, we re-ran *gstacks* and *populations* in STACKS v2.41 using the same parameters (*-r* 0.7, *-p* 2, *--min-maf* 0.03, *--max-obs-het* 0.50) and the same filtering steps that were employed in generating the main SNP dataset. Here, we used only Saiko Lake deep- and stream-spawning samples ($n = 41$), including those individuals ($n = 9$) that were originally discarded due to a low number of reads in the broad-scale analyses. We removed samples with mean coverage <6× and discarded loci with mean missing data >10%, reducing the number of retained loci in order to maximize the number of samples, while still maintaining a robustly filtered SNP dataset.

We used the resulting Saiko-specific SNP dataset and the approach implemented in NewHybrids v 2.0 (Anderson and Thompson 2002)

to investigate evidence for admixture between the co-occurring deep- and stream-spawning kokanee in Saiko Lake. As NewHybrids is known to experience underflow errors with larger datasets, we used a dataset of 500 SNPs that had the highest F_{ST} following Elliott and Russello (2018). We defined 6 genotype frequency classes (deep-spawning kokanee pure stock, stream-spawning kokanee pure stock, F1, F2, B2 deep-spawning kokanee backcross; B2 stream-spawning kokanee backcross) and ran 10 000 sweeps during, and 50 000 sweeps after the burn-in period. Individuals were subsequently assigned into the genotype class that had the highest probability.

Results

Dataset Quality

Following demultiplexing and quality filtering, newly sequenced samples from Japan (n = 62; 58 individuals plus 4 replicates) averaged 9 189 554 retained forward and reverse reads. Combining the new data with the previously published data from the Canadian populations (Veale and Russello 2017b; Samad-zada et al. 2021), yielded 187 samples with an average of 93.95% of reads that successfully mapped to the reference genome; however, only samples (n = 173) with >2 000 000 reads were used to generate the main SNP dataset (see Methods). After full filtering, 9721 SNPs were retained for 167 individuals, with a mean depth of 31.42× and mean missing percentage of 2.45%. The mean within-library genotyping error rate was 1.89%.

Outlier Detection, Mapping, and Annotation

Outlier scans on "paired" (deep- vs. stream-spawning) populations from Canada and Japan identified 397 outliers (Supplementary Figure S1). Outliers detected within a region by more than one method were considered robust, and their annotations are provided in Supplementary Tables S3 and S4. For Canadian deep-spawners (East Barrière, Anderson, Seton Lakes) versus stream-spawners (Nicola Lake), we detected the following numbers of outliers: Arlequin (n = 6); BayeScan (n = 8); and *pcadapt* (n = 148; optimal PCs retained = 3); 8 of these outliers were detected by more than one method (Supplementary Figure S1). Of these 8 robust outliers, 6 revealed BLAST annotations (Supplementary Table S3). For Japanese deep-spawners (Saiko Lake) versus stream-spawners (Saiko, Akan Lakes), we detected the following numbers of outliers: Arlequin (n = 94); BayeScan (n = 0); and *pcadapt* (n = 201; optimal PCs retained = 2); 44 of these outliers were detected by more than one method (Supplementary Figure S1). Of these 44 robust outliers, 35 revealed BLAST annotations (Supplementary Table S4). No shared robust outliers were independently detected in Japan and Canada.

Significant BLAST annotations were revealed for 6 of 8 robust outliers in Canada (Supplementary Table S3) and 35 of 44 robust outliers in Japan (Supplementary Table S4), 4 of which may be indicative of some overlap between potential functions across the 2 geographic regions. SZNR01014727.1_80153 [unplaced scaffold], detected in stream- versus deep-spawning comparisons in British Columbia, and SZNR01027293.1_1297100 [Linkage Group 17], detected in stream- versus deep-spawning comparisons in Japan, annotated to the same genomic region [*Oncorhynchus mykiss* SYPG1 (SYPG1), PHF1 (PHF1), and RGL2 (RGL2) genes, complete cds; DNaseII pseudogene, complete sequence; LGN-like, PBX2 (PBX2), NOTCH-like, TAP1 (TAP1), and BRD2 (BRD2) genes, complete cds; and MHCII-alpha and Raftlin-like pseudogenes, complete sequence; Supplementary Tables S3 and S4]. Moreover, 2 different outliers (SZNR01019859.1_12357 [unplaced scaffold] and SZNR01017145.1_7778989 [Linkage Group 20]) that were independently detected in Canada and Japan, respectively, annotated to the same genomic region: PREDICTED: *Oncorhynchus kisutch* inactive phospholipid phosphatase 7 (LOC109868517), mRNA (Supplementary Tables S3 and S4).

Canada- and Japan-specific robust outliers were located within 100 kb of 27 and 199 unique genes, respectively (Supplementary Tables S5 and S6). Of these, 64 genes (Canada: n = 10; Japan: n = 54) had an exact annotation match found within *D. rerio* genome (Supplementary Tables S7 and S8). No significant GO enrichment terms were identified for Japan-specific outliers, whereas Canada-specific analysis identified 2 GO terms that were potentially overrepresented: GO:0034199, activation of protein kinase A activity (P value = 4.93E-04, FDR = 1.00E+00) and GO:0021576, hindbrain formation (P value = 9.86E-04, FDR = 1.00E+00) (Supplementary Table S9).

All Japanese kokanee samples, regardless of ecotype or location, were fixed for the "GG" genotype at SZNR01010580.1_848156 (LRRC9) that has been associated with shore-spawning behavior in *O. nerka* (Veale and Russello 2017a). In Canada, all kokanee populations were genetically assigned to the observed reproductive ecotypes based on LRRC9 genotypes. Nicola and Kootenay Lake stream-spawning populations contained no individuals with the GG genotype, while in East Barrière, Anderson, and Seton Lake kokanee populations, the G allele frequencies were between 0.95 and 1.00, consistent with observed shore-spawning (both shore- and deep-spawners spawn in a similar habitat, albeit at different depths).

Diversity Statistics and Population Structure Estimation

For genomic scans conducted across all populations, we detected the following numbers of outliers: Arlequin (n = 10); BayeScan (n = 24 unique outliers across all pairwise comparisons); and *pcadapt* (n = 641 unique outliers across all regional comparisons). We subsequently removed all SNPs that were identified as outliers (n = 700) across all analyses; none of the remaining SNPs deviated from HWE. Based on this putatively neutral dataset of 9021 SNPs, H_o, H_e, and F_{IS} values were similar across all Canadian populations, with both H_o and H_e, ranging from 0.17 to 0.22 and none of the F_{IS} values significantly differed from zero. Levels of heterozygosity were consistently lower for the Japanese populations (0.07–0.12), all of which exhibited heterozygote deficit and positive F_{IS} values that were significantly different from zero (Table 1). Both deep- and stream-spawning kokanee populations in Japan demonstrated substantially lower N_e estimates (Saiko = 30.4 [deep-spawning] and 33.2 [stream-spawning]; Akan = 223.3) than those detected in Canada, which ranged from 814.5 (East Barrière) to 4194.2 (Nicola).

Pairwise θ estimates indicated that the Saiko Lake deep-spawning population was the most highly differentiated relative to all others in both Canada and Japan, with the largest values observed in comparison with stream-spawning kokanee from Akan Lake (θ = 0.70) and Nicola Lake (θ = 0.63) (Table 2). The Akan Lake and Saiko Lake stream-spawning kokanee populations were not genetically distinct (θ = –0.00031). The Canadian populations all demonstrated significant pairwise θ estimates largely associated with geographic proximity, ranging from 0.00955 (Anderson Lake deep-spawning vs. Seton Lake deep-spawning) to 0.25524 (Kootenay Lake stream-spawning vs. Nicola Lake stream-spawning) (Table 2).

Table 2. Weir and Cockerham's (1984) θ estimates between each pair of populations based on 1000 permutations, as calculated by Genetix v.4.05.2 (Belkhir et al. 2004)

Population	Seton (deep)	East Barrière (deep)	Nicola (stream)	Kootenay (stream)	Akan (stream)	Saiko (deep)	Saiko (stream)
Anderson (deep)	0.00955[a]	0.17221[a]	0.21146[a]	0.13696[a]	0.35993[a]	0.56557[a]	0.34399[a]
Seton (deep)	—	0.17046[a]	0.20593[a]	0.13518[a]	0.35616[a]	0.562[a]	0.34044[a]
East Barrière (deep)	—	—	0.25303[a]	0.22313[a]	0.4079[a]	0.58897[a]	0.39593[a]
Nicola (stream)	—	—	—	0.25524[a]	0.46237[a]	0.62815[a]	0.44937[a]
Kootenay (stream)	—	—	—	—	0.40332[a]	0.59703[a]	0.38911[a]
Akan (stream)	—	—	—	—	—	0.7005[a]	−0.00031
Saiko (deep)	—	—	—	—	—	—	0.69688[a]

[a]Values that were significant at $\alpha = 0.01$.

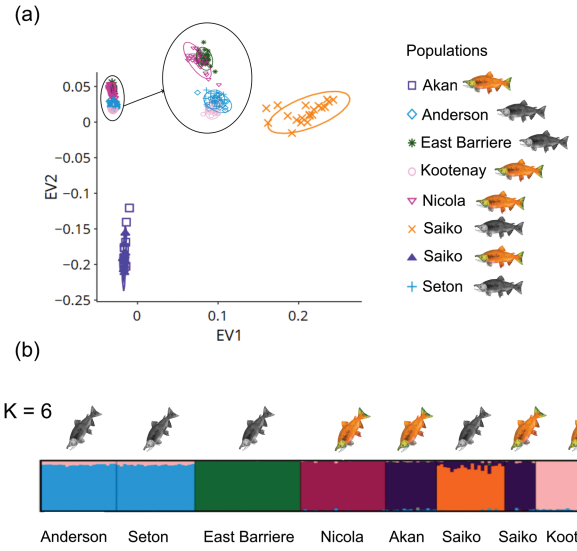

Figure 2. (**a**) PCA results produced using R package *SNPRelate* (Zheng et al. 2012) showing population clustering based on the first 2 eigenvectors that explain 44.4% and 13.6% of the variation, respectively. (b) Results of Bayesian clustering analysis, as implemented in STRUCTURE v3.4 (Pritchard et al. 2000). Output results represent the optimal *K* value (*K* = 6), as determined by the ΔK method (Evanno et al. 2005), as implemented in STRUCTURE HARVESTER (Earl and vonHoldt 2012). Visualized using CLUMPAK (Kopelman et al. 2015). Orange fish represent stream-spawning kokanee, black fish represent deep-spawning kokanee. Fish illustrations are courtesy of Eileen Klatt.

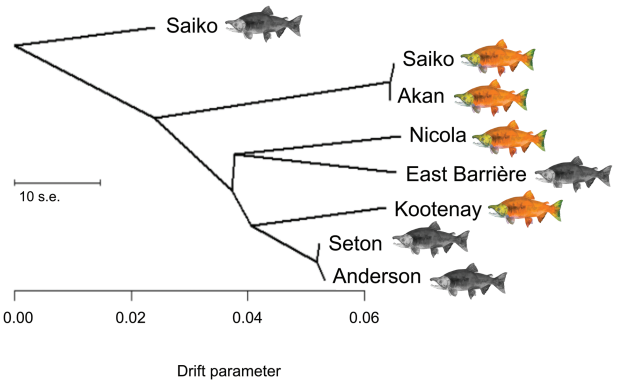

Figure 3. Maximum likelihood (ML) tree generated by TreeMix (Pickrell and Pritchard 2012) based on the pruned dataset of 901 putatively neutral SNPs. The tree was rooted with the Saiko deep-spawning population. Orange fish represent stream-spawning kokanee, black fish represent deep-spawning kokanee. Fish illustrations are courtesy of Eileen Klatt.

The STRUCTURE analysis revealed evidence for 6 clusters that best explained the genetic variation within our dataset (Figure 2, Supplementary Figure S2 and Supplementary Table S10). Saiko Lake deep-spawning kokanee was identified as a distinct cluster starting from *K* = 2, and Akan Lake and Saiko Lake stream-spawning kokanee formed a separate cluster starting from *K* = 3, separating Japanese kokanee from populations in Canada. East Barrière Lake deep-spawning kokanee and Nicola Lake stream-spawning kokanee each formed distinct clusters at *K* = 5, and Kootenay Lake stream-spawning kokanee separated from Anderson Lake and Seton Lake deep-spawning kokanee at *K* = 6 (Figure 2). This clustering persisted even at higher values of *K*; Anderson Lake and Seton Lake deep-spawning kokanee never separated into distinct clusters. The PCA results largely agreed with those produced by STRUCTURE, revealing differentiation by geography (Figure 2). The first PC explained the largest percent of variation (44.4 %) and separated Saiko Lake deep-spawning kokanee from all other populations. The second PC explained 13.6% of the variation and separated the tightly clustered Akan Lake and Saiko Lake stream-spawning kokanee from the Canadian populations. In Canada, all populations formed distinct clusters, with the exception of the Anderson Lake and Seton Lake deep-spawning populations.

The maximum likelihood tree generated in TreeMix based on the reduced dataset of 901 SNPs was consistent with patterns revealed by STRUCTURE and PCA analyses (Figure 3). Saiko Lake deep-spawning kokanee showed the highest divergence relative to all other populations, followed by stream-spawning kokanee from Akan Lake and Saiko Lake. In Canada, population split estimates reflected geographic proximity, with the exception of Kootenay Lake stream-spawning kokanee that was most closely related to Anderson Lake and Seton Lake deep-spawning kokanee populations.

Hybridization Analysis of Saiko Lake Co-occurring Ecotypes

After full filtering, 2165 SNPs and 34 samples (n = 20 deep-spawning, n = 14 stream-spawning) were retained for the Saiko-specific dataset; we retained only the 500 SNPs with the highest F_{ST} for the NewHybrids analysis (see Methods). In total, 33 of 34 individuals were assigned to their pure stock of origin based on NewHybrids analysis. There was one individual that was assigned as a B2 deep-spawning kokanee backcross with a probability of 0.998.

Discussion

The evolutionary history of kokanee has been the subject of continued study, revealing polyphyletic origins (Taylor et al. 1996;

Frazer and Russello 2013), genetic distinctiveness from anadromous sockeye salmon (Beacham and Withler 2017), and a genomic basis for reproductive ecotype variation across their distribution (Lemay and Russello 2015; Veale and Russello 2017a, 2017b). To complement this growing body of literature, our study contributes important insights into population history and divergence of the less studied deep-spawning kokanee ecotype across its pan-Pacific range. We demonstrate that population structure of deep-spawning kokanee is not associated with ecotype, but reflects geographic and translocation history. Furthermore, the low number of shared outliers detected between deep- and stream-spawning kokanee, along with the high level of divergence between Canadian and Japanese kokanee populations, suggest that the deep-spawning ecotype likely evolved multiple times.

Parallel Genetic Evolution

In salmonids, diverse life history variation results from environmental fluctuations that create conditions for local adaptation (Waples et al. 2008). When distinct populations adapt to similar habitats, selective forces might act on the same set of traits, resulting in the evolution of similar phenotypes (Østbye et al. 2006). For example, in whitefish, adaptation to benthic or pelagic niches is associated with divergence in multiple traits, including diet composition, morphological parameters (e.g., number of gill rakers), growth rate, as well as age and size at maturity (Østbye et al. 2006). For *Oncorhynchus* spp., spawning habitat preference depends on many environmental variables, such as flow of oxygenated water and presence of substrate necessary for redd formation (Arostegui and Quinn 2019). Although limited information exists on spawning ground availability in lakes examined in this study, all deep-spawning kokanee populations share one defining characteristic in this respect: spawning depth. For this ecotype, the loss of bright red pigmentation in mature fish has been proposed to be an adaptation to decrease visibility to predators in deep waters with diminished light penetration, or might have evolved due to trade-offs in resource allocation between pigmentation and other costly metabolic processes (Moreira and Taylor 2015). However, the genetic basis of traits associated with deep-spawning behavior has not yet been examined.

Identification of candidate genetic regions underlying parallelisms is an approach that can enhance our understanding of the repeatability of evolution, as well as aid in predicting the adaptive response of wild populations to different environmental conditions (Stinchcombe and Hoekstra 2008; Erickson et al. 2016; van Boheemen and Hodgins 2020). In the last 2 decades, there has been an upsurge in the number of studies that have employed genomic scans to identify molecular mechanisms responsible for the evolution of similar phenotypic traits in distinct lineages (Fraser and Whiting 2020). In some systems, genomic scans have revealed specific regions associated with ecotype divergence: for instance, in sticklebacks, outliers that were highly divergent between limnetic and benthic ecotypes were localized to genes associated with immune response (Jones et al. 2012). Likewise, genotype frequencies at LRRC9 have been associated with reproductive ecotype divergence in *O. nerka* (Nichols et al. 2016; Veale and Russello 2017a, 2017b). Despite these examples, in a large proportion of studies, overlapping outliers identified between pairs of ecotypes remain rare, with the majority of outliers specific to geographic differentiation or population history of the system in question (Russello et al. 2012; Frazer and Russello 2013; Perrier et al. 2013; Rougemont et al. 2017; Salisbury et al. 2020).

Similarly, our study found little overlap in outliers underlying ecotype divergence in deep- and stream-spawning kokanee in Japan and Canada. The 2 instances in which outlier loci detected in Japan and Canada independently annotated to the same region most likely lack biological significance (Supplementary Tables S3 and S4). One of these annotations consists either of genes for which no association with spawning could be determined and/or pseudogenes [*Oncorhynchus mykiss* SYPG1 (SYPG1), PHF1 (PHF1), and RGL2 (RGL2) genes, complete cds; DNaseII pseudogene, complete sequence; LGN-like, PBX2 (PBX2), NOTCH-like, TAP1 (TAP1), and BRD2 (BRD2) genes, complete cds; and MHCII-alpha and Raftlin-like pseudogenes, complete sequence]. The other annotation was predicted to fall within a gene coding for an enzyme exhibiting no detectable activity [*Oncorhynchus kisutch* inactive phospholipid phosphatase 7 (LOC109868517)], with no clear (or tangential) association with spawning behavior.

Furthermore, we identified only 2 GO terms potentially associated with deep-spawning behavior: activation of protein kinase A activity (GO:0034199) and hindbrain formation (GO:0021576). Although it is challenging to directly associate protein kinase A activity with deep-spawning behavior, further exploration of genes involved in hindbrain formation may be of interest for future studies. For example, studies have shown that in zebrafish, the hindbrain contains neurons involved in nonvisual photo-sensation (Kokel et al. 2013), which might be of relevance for deep-spawning kokanee that spawn in low-light conditions. Yet, this analysis had a number of limitations and the association should be interpreted with caution. For instance, we compared *O. nerka* gene descriptions with those of *D. rerio* given its superior annotation, however, this approach eliminated a large proportion of identified genes. In addition, the lack of a GO association file for *O. nerka* impeded the use of other approaches to allow comparison of results with those from the GOrilla analysis to improve stringency.

Taken together, the low number of outliers and GO term associations related to deep-spawning behavior can have multiple explanations. First, it is possible that *O. nerka* diversification into deep-spawning kokanee happened independently and through different molecular processes that resulted in the evolution of morphologically and behaviorally similar populations. Moreover, if many genes with a small effect, rather than a single locus with a large effect, are responsible for the evolution of a certain phenotypic trait, identification of such genes might not be feasible via genomic scans (Rockman 2012). Additionally, reduced representation sequencing methods, such as RADseq, only capture a fraction of the genome; given that the sockeye salmon genome is estimated to be 2.4-2.6 Gbp (Christensen et al. 2020), the probability that a gene or genes underlying a certain phenotypic trait will be found among the several thousand examined markers is low. The absence of any lakes where deep-spawning and stream-spawning kokanee naturally co-occur further complicates the situation. Outlier detection between the 2 ecotypes in Saiko Lake constitutes a comparison of 2 populations that were artificially introduced into the ecosystem (Nakabo et al. 2011; Muto et al. 2013), and hence might be confounded (see below). Lastly, a small effective population size and reduced genetic diversity of deep-spawning kokanee in Saiko Lake might have amplified the impact of genetic drift in this population, masking the presence of shared outliers (Fraser and Whiting 2020).

In addition to the lack of common outlier loci, analyses based on the putatively neutral dataset of 9021 SNPs demonstrated that deep-spawning kokanee populations are more closely related to

stream-spawning kokanee on the same continent, than they are to deep-spawning kokanee populations across the Pacific Ocean. These findings are consistent with the multiple origins hypothesis suggested by Moreira and Taylor (2015), which is in line with the polyphyletic history of kokanee more broadly (Taylor et al. 1996).

Regional Genetic Diversity and Divergence of Deep-Spawning Kokanee

In Canada, *O. nerka* population divergence primarily reflects differentiation by geography. The 3 known deep-spawning kokanee populations in Canada constitute natural stocks that have not been subject to translocations, and hence represent a unique opportunity to study divergence of this ecotype in the absence of human intervention. Based on putatively neutral data alone, differentiation between deep-spawning populations in Canada is best explained by geography rather than life history, as evidenced by distinct STRUCTURE clustering and reconstructed relationships based on the TreeMix analysis. Geographically proximate populations are less genetically distinct (e.g., Anderson and Seton Lake kokanee). In contrast, populations that have the same reproductive ecotype, but are geographically isolated, demonstrated a high level of differentiation (Figures 2 and 3). Kootenay Lake kokanee represent the only deviation from this pattern, as this population is more closely related to Anderson and Seton Lake, despite currently being located in the Columbia Basin. This finding is consistent with previous work by Taylor et al. (1996), which hypothesized that the Kootenay Lake kokanee population is more closely related to Fraser Basin kokanee due to fluctuations in historic river connectivity during a deglaciation period that resulted in the Lower Fraser River being obstructed, and the Upper Fraser River draining through the Columbia Basin (Taylor et al. 1996). Previous studies have proposed that some of the genetically similar *O. nerka* populations in these 2 river basins might have been established during that period (Taylor et al. 1996). In general, our results are concordant with earlier population genetic and genomic studies of kokanee reproductive ecotypes on multiple scales (Nichols et al. 2016; Beacham and Withler 2017; Veale and Russello 2017b; Samad-zada et al. 2021).

In Japan, on the other hand, *O. nerka* divergence patterns are more consistent with the documented history of kokanee transplantations (Kogura et al. 2011). Kokanee in Japan constitute a valuable commercial fishery and, due to their economic and recreational value, are subject to on-going translocation and supplementation efforts (Yamamoto et al. 2011). Here, we analyzed 3 Japanese kokanee populations, spanning 2 different lakes and representing 2 ecotypes, all of which were artificially introduced into the lakes they currently inhabit. All "red" kokanee in Japan have their origin in Akan Lake (stream-spawning) (Hokkaido Government Office Fisheries Division 1900) or Chimikeppu Lakes (shore-spawning) (Japan Fisheries Resource Conservation Association 2005). Saiko Lake stream-spawning kokanee, for example, were transplanted from Towada or Shikotsu Lakes (Muto et al. 2013), but they were initially introduced to these source lakes from Akan Lake (Kogura et al. 2011). This transplantation history likely explains the absence of genetic differentiation between Saiko Lake and Akan Lake stream-spawning kokanee. Furthermore, the present population of Akan Lake is thought to have lost some of its original ecological and/or genetic characteristics due to frequent reciprocal transplantation among hatcheries and natural habitats (Yamamoto 2015). Similarly, deep-spawning kokanee in Saiko Lake are the progeny of transplants that were relocated from Tazawa Lake in 1935 (Nakabo et al. 2011). Given this context, it is not surprising that our fine-scale analyses of population structure revealed high differentiation between stream- and deep-spawning kokanee in Saiko Lake (Figures 2–3), results that are concordant with previous findings in this system based on microsatellite analyses (Muto et al. 2013). Previous research has attributed this divergence to the temporal and spatial distribution of the 2 ecotypes (Muto et al. 2013), which is a known driver of reproductive isolation in *O. nerka* (Arostegui and Quinn 2019). Yet, the presence of one individual that was assigned to the B2 deep-spawning kokanee backcross demonstrates that there exists a potential for hybridization. These results, however, should be considered within the context of the dataset, as SNP panels can vary in their assignment accuracy and power (Elliott and Russello 2018). A recent study in a different system demonstrated that the 500 highest F_{ST} SNP dataset was sufficient to correctly assign backcrosses (Elliott and Russello 2018) between co-occurring anadromous sockeye salmon and kokanee in the Okanagan Basin, but the applicability of this approach to Saiko Lake kokanee requires additional study. A high degree of hybridization (64% of sampled individuals) was previously reported in Motosu Lake (Nakayama et al. 2018), the only other known location in the world where deep- and stream-spawning kokanee co-occur. Although the degree of hybridization is low (if any) within Saiko Lake, our results should be considered within the context of on-going fisheries management.

In terms of genetic diversity, all 3 Japanese kokanee populations exhibited unusually low estimates of N_e and H_o, compared to populations in Canada. Moreover, all Japanese populations demonstrated evidence for inbreeding, as revealed by positive F_{IS} coefficients that were significantly different from 0 (Table 1). These values are indicative of reduced genetic diversity that is often observed in transplanted populations, even in cases when a viable population is successfully established. Loss of diversity is characteristic of a founder effect, as the subset chosen for transplantation rarely represents the diversity of an entire donor stock and not all transplants survive in the new locality (Quinn et al. 1998).

Lastly, we found that all Japanese populations were genotypically shore-spawners based on results at the LRRC9 locus. Limited information exists regarding the original reproductive ecotypes of kokanee in the Japanese lakes studied here, especially given their repeated transplantation histories. While a shore-spawning genotype for deep-spawning kokanee was expected based on previous work (Veale and Russello 2017b), several factors might explain why the stream-spawners also exhibited the GG genotype (Moreira and Taylor 2015). For example, "red" kokanee have been observed spawning in both streams and along the shore of Akan Lake (Nakayama K, personal observation), which was one source of transplants into Saiko Lake. Yet, it is unclear which ecotype was originally transplanted to Saiko Lake. The first egg collection from Akan Lake was from stream-spawners, however, when Akan Lake individuals were transplanted to Shikotsu Lake, shore-spawning was observed near the mouth of the river (Hokkaido Government Office Fisheries Division 1900). In addition to practical considerations associated with transplant sourcing, some studies propose that acidification of lacustrine waters can impact the reproductive behavior of kokanee during spawning (Ikuta et al. 2003). If valid, such changes in water chemistry may introduce additional selection pressures that could drive shifts in ecotype frequency, perhaps even to fixation. Nevertheless, information on the spawning behavior of Japanese kokanee remains scarce, which prevents us from making any definitive conclusions regarding the history of reproductive (shore- or stream-spawning) ecotype(s) in these systems.

Limitations and Future Studies

Although we detected no evidence for parallel genetic signatures underlying deep-spawning kokanee ecotype divergence in Canada and Japan, we recognize that the use of reduced representation sequencing might not be suitable for identifying candidate genes under selection in this system. Therefore, a more comprehensive understanding of the genomic basis of deep-spawning in kokanee likely requires whole-genome analysis. Furthermore, genomic scans alone might not be sufficient for identifying statistically significant gene candidates if landscape and demographic factors are not considered (Fraser and Whiting 2020). Alternatively, direct investigation of the functional significance of targeted genes hypothesized to play a role in deep-spawning kokanee morphology or behavior might shed more light on the evolution of this ecotype. For example, the beta-carotene oxygenase 2-like gene is associated with red pigmentation in Chinook salmon (Lehnert et al. 2019) and thus might be a potential candidate for divergent selection in deep-spawning kokanee. Divergent selection might also act on genes associated with adaptations to deep water conditions, such as hemoglobins that are involved in oxygen-binding and transport, or rhodopsins that affect visual perception in low-light conditions (Hahn et al. 2017).

Conclusion

Overall, we demonstrate that the observed structure among the 8 kokanee populations examined in this study is associated with geographic isolation or translocation history, rather than ecotype. Our results show a high level of divergence between kokanee populations in Canada and Japan, as well as between deep- and stream-spawning kokanee populations regionally, suggesting the independent evolution of the deep-spawning ecotype on the 2 continents. We also detected low estimates of heterozygosity, significant levels of inbreeding, and reduced effective population sizes in all Japanese kokanee populations, likely associated with transplantation history, which should be considered within the context of on-going fisheries management in these systems.

Supplementary Material

Supplementary data are available at *Journal of Heredity* online.

Acknowledgments

Although no new samples were collected for this work, we are grateful to all involved in original sample collection from the previous studies cited that helped make this research possible, including, but not limited to: S. Harris and staff from the BC Ministry of Environment and Climate Change and the BC Ministry of Forests, Lands, Natural Operations and Rural Development; A. Moreira and R. Taylor (UBC); and Y. Watanabe (Lake Saiko Fisheries Cooperative), T. Nakabo, N. Muto, A. Tohkairin, A. Yoshikawa (Kyoto University). We also thank B. van Poorten and S. Pakpour for comments on an earlier version of this article.

Funding

This work was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant # RGPIN-2019-04621 to M.R., and computational resources were made available by Compute Canada through the Resources for Research Groups program.

Data Availability

All Illumina raw reads are available from the NCBI sequence read archive (BioProject ID: PRJNA768875). RAD loci sequences and SNP genotypic data are deposited in DRYAD (doi.org/10.5061/dryad.mw6m905xk).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*. 215:403–410.

Alves-Pereira A Clement CR, Picanço-Rodrigues D, Veasey EA, Dequigiovanni G, Ramos SLF, Zucchi MI. 2020. A population genomics appraisal suggests independent dispersals for bitter and sweet manioc in Brazilian Amazonia. *Evol Appl*. 13:342–361. doi:10.1111/eva.12873

Anderson EC, Thompson EA. 2002. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*. 160:1217–1229.

Andreev D, Kreitman M, Phillips TW, Beeman RW, ffrench-Constant RH. 1999. Multiple origins of Cyclodiene insecticide resistance in *Tribolium castaneum* (Coleoptera: Tenebrionidae). *J Mol Evol*. 48:615–624.

Andrusak H, Morris A. 2004. *Final survey results of select lakes in the Kamloops area inhabited by kokanee—Year 4*. Nelson, BC: Redfish Consulting Ltd.

Arostegui MC, Quinn TP. 2019. Reliance on lakes by salmon, trout and charr (*Oncorhynchus*, *Salmo* and *Salvelinus*): an evaluation of spawning habitats, rearing strategies and trophic polymorphisms. *Fish Fish*. 20:775–794.

Babin C, Gagnaire P-A, Pavey SA, Bernatchez L. 2017. RAD-seq reveals patterns of additive polygenic variation caused by spatially-varying selection in the American eel (*Anguilla rostrata*). *Genome Biol Evol*. 9:2974–2986.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*. 3:e3376. doi:10.1371/journal.pone.0003376

Beacham TD, Withler RE. 2017. Population structure of sea-type and lake-type sockeye salmon and kokanee in the Fraser River and Columbia River drainages. *PLoS ONE*. 12:e0183713.

Beaumont MA, Nichols RA. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proc Biol Sci*. 263:1619–1626.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F. 2004. *GENETIX 4.05, Windows TM software for population genetics*. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000.

Benjamini Y, Yekutieli D. 2001. The control of the false discovery rate in multiple testing under dependency. *Ann Stat*. 29:1165–1188.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. *Stacks*: building and genotyping loci *de novo* from short-read sequences. *Genes Genom Genet*. 1:171–182.

Christensen KA, Rondeau EB, Minkley DR, Sakhrani D, Biagi CA, Flores A-M, Koop BF. 2020. The sockeye salmon genome, transcriptome, and analyses identifying population defining regions of the genome. *PLoS ONE*. 15:e0240935. doi:10.1371/journal.pone.0240935

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA. 2011. 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics*. 27:2156–2158. doi:10.1093/bioinformatics/btr330

Deagle BE, Jones FC, Chan YF, Absher DM, Kingsley DM, Reimchen TE. 2012. Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proc Biol Sci*. 279:1277–1286.

de Villemereuil P, Frichot É, Bazin É, François O, Gaggiotti OE. 2014. Genome scan methods against more complex models: when and how much should we trust them? *Mol Ecol*. 23:2006–2019.

Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR. 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 DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour*. 4:359–361.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*. 10:48.

Elliott L, Russello MA. 2018. SNP panels for differentiating advanced-generation hybrid classes in recently diverged stocks: a sensitivity analysis to inform monitoring of sockeye salmon re-stocking programs. *Fish Res*. 208:339–345.

Erickson PA, Glazer AM, Killingbeck EE, Agoglia RM, Baek J, Carsanaro SM, Miller CT. 2016. Partially repeatable genetic basis of benthic adaptation in threespine sticklebacks: repeatable evolution in benthic sticklebacks. *Evolution*. 70:887–902. doi:10.1111/evo.12897

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*. 14:2611–2620.

Excoffier L, Lischer HEL. 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.

Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*. 180:977–993.

Foote CJ, Moore K, Stenberg K, Craig K, Wenburg J, Wood CC. 1999. Genetic differentiation in gill raker number and length in sympatric anadromous and nonanadromous morphs of sockeye salmon, *Oncorhynchus nerka*. *Environ Biol Fish*. 54:263–274.

Foote CJ, Wood CC, Withler RE. 1989. Biochemical genetic comparison of sockeye salmon and kokanee, the anadromous and nonanadromous forms of *Oncorhynchus nerka*. *Can J Fish Aquat Sci*. 46:149–158.

Frazer KK, Russello MA. 2013. Lack of parallel genetic patterns underlying the repeated ecological divergence of beach and stream-spawning kokanee salmon. *J Evol Biol*. 26:2606–2621.

Fraser BA, Whiting JR. 2020. What can be learned by scanning the genome for molecular convergence in wild populations? *Ann N Y Acad Sci*. 1476:23–42.

Goudet J, Jombart T. 2015. *hierfstat*: Estimation and tests of hierarchical F-statistics. R package version 0.04-22. Available from: https:// CRAN.R-project.org/package=hierfstat.

Gross JB, Borowsky R, Tabin CJ. 2009. A novel role for Mc1r in the parallel evolution of depigmentation in independent populations of the cavefish *Astyanax mexicanus*. *PLoS Genet*. 5:e1000326.

Hahn C, Genner MJ, Turner GF, Joyce DA. 2017. The genomic basis of cichlid fish adaptation within the deepwater "twilight zone" of Lake Malawi: genomic basis of deep-water cichlid fish adaptation. *Evol Lett*. 1:184–198.

Haldane JBS. 1932. *The causes of evolution*. London: Longmans Green.

Hokkaido Government Office Fisheries Division. 1900. *The report of the Chitose salmon hatchery*. pp. 32–41. (in Japanese)

Ikuta K, Suzuki Y, Kitamura S. 2003. Effects of low pH on the reproductive behavior of salmonid fishes. *Fish Physiol Biochem*. 28:407–410.

Japan Fisheries Resource Conservation Association. 2005. *Report on the project to develop basic information on the lake environment, Lake Shikotsu*. pp. 69–79. (in Japanese)

Jones FC, Chan YF, Schmutz J, Grimwood J, Brady SD, Southwick AM, Kingsley DM. 2012. A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Curr Biol*. 22:83–90. doi:10.1016/j.cub.2011.11.045

Jordan R, Hoffmann AA, Dillon SK, Prober SM. 2017. Evidence of genomic adaptation to climate in *Eucalyptus microcarpa*: implications for adaptive potential to projected climate change. *Mol Ecol*. 26:6002–6020.

Kaeriyama M, Urawa S, Fukuwaka M. 1995. Variation in body size, fecundity, and egg size of sockeye and kokanee salmon, *Oncorhynchus nerka*, released from hatchery. *Sci Rep Hokkaido Salmon Hatchery*. 49:1–9.

Kans J. 2013. Entrez Direct: e-utilities on the Unix command line. In: *Entrez programming utilities help* [Internet]. Bethesda (MD): National Center for Biotechnology Information.

Kogura Y, Seeb JE, Azuma N, Kudo H, Abe S, Kaeriyama M. 2011. The genetic population structure of lacustrine sockeye salmon, *Oncorhynchus nerka*, in Japan as the endangered species. *Environ Biol Fish*. 92:539–550.

Kokel D, Dunn TW, Ahrens MB, Alshut R, Cheung CYJ, Saint-Amant L, Peterson RT. 2013. Identification of nonvisual photomotor response cells in the vertebrate hindbrain. *J Neurosci*. 33:3834–3843. doi:10.1523/JNEUROSCI.3689-12.2013

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across *K*. *Mol Ecol Resour*. 15:1179–1191.

Larson WA, Limborg MT, McKinney GJ, Schindler DE, Seeb JE, Seeb LW. 2017. Genomic islands of divergence linked to ecotypic variation in sockeye salmon. *Mol Ecol*. 26:554–570.

Lehnert SJ, Christensen KA, Vandersteen WE, Sakhrani D, Pitcher TE, Heath JW, Devlin RH. 2019. Carotenoid pigmentation in salmon: variation in expression at BCO2-l locus controls a key fitness trait affecting red coloration. *Proc Biol Sci*. 286:20191588. doi:10.1098/rspb.2019.1588

Lemay MA, Russello MA. 2015. Genetic evidence for ecological divergence in kokanee salmon. *Mol Ecol*. 24:798–811.

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

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N; 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25:2078–2079. doi:10.1093/bioinformatics/btp352

Luu K, Bazin E, Blum MGB. 2017. *pcadapt*: an R package to perform genome scans for selection based on principal component analysis. *Mol Ecol Resour*. 17:67–77.

Macqueen DJ, Kristjánsson BK, Paxton CGM, Vieira VLA, Johnston IA. 2011. The parallel evolution of dwarfism in Arctic charr is accompanied by adaptive divergence in mTOR-pathway gene expression: gene expression evolution in fish populations. *Mol Ecol*. 20:3167–3184.

McKinney GJ, Waples RK, Seeb LW, Seeb JE. 2017. Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Mol Ecol Resour*. 17:656–669.

Moreira AL, Taylor EB. 2015. The origin and genetic divergence of "black" kokanee, a novel reproductive ecotype of *Oncorhynchus nerka*. *Can J Fish Aquat Sci*. 72:1584–1595.

Muto N, Nakayama K, Nakabo T. 2013. Distinct genetic isolation between "Kunimasu" (*Oncorhynchus kawamurae*) and "Himemasu" (*O. nerka*) in Lake Saiko, Yamanashi Prefecture, Japan, inferred from microsatellite analysis. *Ichthyol Res*. 60:188–194.

Nakabo T, Nakayama K, Muto N, Miyazawa M. 2011. *Oncorhynchus kawamurae* "Kunimasu," a deepwater trout, discovered in Lake Saiko, 70 years after extinction in the original habitat, Lake Tazawa, Japan. *Ichthyol Res*. 58:180–183.

Nakabo T, Tohkairin A, Muto N, Watanabe Y, Miura Y, Miura H, Kai Y. 2014. Growth-related morphology of "Kunimasu" (*Oncorhynchus kawamurae*: Family Salmonidae) from Lake Saiko, Yamanashi Prefecture, Japan. *Ichthyol Res*. 61:115–130. doi:10.1007/s10228-013-0382-7

Nakayama K, Tohkairin A, Yoshikawa A, Nakabo T. 2018. Detection and morphological characteristics of "Kunimasu" (*Oncorhynchus kawamurae*)/"Himemasu" (*O. nerka*) hybrids in Lake Motosu, Yamanashi Prefecture, Japan. *Ichthyol Res*. 65:270–275.

Nei M. 1987. *Molecular Evolutionary Genetics*. New York (NY): Columbia University Press.

Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, Stamatoyannopoulos JA. 2012. BEDOPS: high-performance genomic feature operations. *Bioinformatics*. 28:1919–1920. doi:10.1093/bioinformatics/bts277

Nichols KM, Kozfkay CC, Narum SR. 2016. Genomic signatures among *Oncorhynchus nerka* ecotypes to inform conservation and management of endangered Sockeye Salmon. *Evol Appl*. 9:1285–1300.

Østbye K, Amundsen P-A, Bernatchez L, Klemetsen A, Knudsen R, Kristoffersen R, Hindar K. 2006. Parallel evolution of ecomorphological traits in the European whitefish *Coregonus lavaretus* (L.) species complex during postglacial times: parallel evolution in European whitefish. *Mol Ecol*. 15:3983–4001. doi:10.1111/j.1365-294X.2006.03062.x

Perrier C, Bourret V, Kent MP, Bernatchez L. 2013. Parallel and nonparallel genome-wide divergence among replicate population pairs of freshwater and anadromous Atlantic salmon. *Mol Ecol*. 22:5577–5593.

Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*. 8:e1002967.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics*. 155:945–959.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 81:559–575. doi:10.1086/519795

Quesada H, Posada D, Caballero A, Morán P, and Rolán-Alvarez E. 2007. Phylogenetic evidence for multiple sympatric ecological diversification in a marine snail. *Evolution*. 61:1600–1612.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 26:841–842.

Quinn TP, Graynoth E, Wood CC, Foote CJ. 1998. Genotypic and phenotypic divergence of sockeye Salmon in New Zealand from their ancestral British Columbia populations. *Trans Am Fish Soc*. 127:517–534.

Rockman MV. 2012. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution*. 66:1–17.

Rougemont Q, Gagnaire P-A, Perrier C, Genthon C, Besnard A-L, Launey S, Evanno G. 2017. Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Mol Ecol*. 26:142–162.

Russello MA, Kirk SL, Frazer KK, Askey PJ. 2012. Detection of outlier loci and their utility for fisheries management: outlier loci for fisheries management. *Evol Appl*. 5:39–52.

Salisbury SJ, McCracken GR, Perry R, Keefe D, Layton KKS, Kess T, Ruzzante DE. 2020. Limited genetic parallelism underlies recent, repeated incipient speciation in geographically proximate populations of an Arctic fish (*Salvelinus alpinus*). *Mol Ecol*. 29:4280–4294. doi:10.1111/mec.15634

Samad-zada F, van Poorten BT, Harris S, Godbout L, Russello MA. (in press). Genome-wide analysis reveals demographic and life-history patterns associated with habitat modification in landlocked, deep-spawning sockeye salmon (*Oncorhynchus nerka*). *Ecol Evol*. 11:13186–13205. doi:10.1002/ece3.8040

Slatkin M, Voelm L. 1991. FST in a hierarchical island model. *Genetics*. 127:627–629.

Stern DL. 2013. The genetic causes of convergent evolution. *Nat Rev Genet*. 14:751–764.

Stinchcombe JR, Hoekstra HE. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*. 100:158–170.

Suzuki T, Suzuki N, Tojo K. 2019. Parallel evolution of an alpine type ecomorph in a scorpionfly: independent adaptation to high-altitude environments in multiple mountain locations. *Mol Ecol*. 28:3225–3240.

Taylor EB, Foote CJ, Wood CC. 1996. Molecular genetic evidence for parallel life-history evolution within a Pacific salmon (sockeye salmon and kokanee, *Oncorhynchus nerka*). *Evolution*. 50:401–416.

Tintle N, Gordon D, Van Bruggen D, Finch S. 2009. The cost effectiveness of duplicate genotyping for testing genetic association. *Ann Hum Genet*. 73:370–378.

van Boheemen LA, Hodgins KA. 2020. Rapid repeatable phenotypic and genomic adaptation following multiple introductions. *Mol Ecol*. 29:4102–4117.

Veale AJ, Russello MA. 2016. Sockeye salmon repatriation leads to population re-establishment and rapid introgression with native kokanee. *Evol Appl*. 9:1301–1311.

Veale AJ, Russello MA. 2017a. An ancient selective sweep linked to reproductive life history evolution in sockeye salmon. *Sci Rep*. 7:1747.

Veale AJ, Russello MA. 2017b. Genomic changes associated with reproductive and migratory ecotypes in sockeye salmon (*Oncorhynchus nerka*). *Genome Biol Evol*. 9:2921–2939.

Waples RS, Do C. 2008. LDNE: a program for estimating effective population size from data on linkage disequilibrium: computer programs. *Mol Ecol Resour*. 8:753–756.

Waples RS, Pess GR, Beechie T. 2008. Synthesis: evolutionary history of Pacific salmon in dynamic environments: evolutionary history of Pacific salmon in dynamic environments. *Evol Appl*. 1:189–206.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution*. 38:1358–1370.

Westram AM, Galindo J, Rosenblad AM, Grahame JW, Panova M, Butlin RK. 2014. Do the same genes underlie parallel phenotypic divergence in different *Littorina saxatilis* populations? *Mol Ecol*. 23:4603–4616.

Wood TE, Burke JM, Rieseberg LH. 2005. Parallel genotypic adaptation: when evolution repeats itself. *Genetica*. 123:157–170.

Yamamoto S. 2015. Kokanee—an important recreational and commercial salmonid fish with frequent translocations among hatcheries and natural habitats over several decades. *Jpn J Ichthyol*. 62:195–198. (in Japanese)

Yamamoto S, Kitamura S, Sakano H, Morita K. 2011. Genetic structure and diversity of Japanese kokanee *Oncorhynchus nerka* stocks as revealed by microsatellite and mitochondrial DNA markers. *J Fish Biol*. 79:1340–1349.

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*. 28:3326–3328.