Check for updates

Received: 30 November 2023 | Revised: 23 April 2024 | Accepted: 30 April 2024

DOI: 10.1111/mec.17459

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Historical and contemporary processes driving the origin and structure of an admixed population within a contact zone between subspecies of a north temperate diadromous fish

Amy Liu[1,2] | Armando Geraldes[1,2] | Eric B. Taylor[1,2,3]

[1]Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada

[2]Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, Canada

[3]Beaty Biodiversity Museum, University of British Columbia, Vancouver, British Columbia, Canada

Correspondence
Amy Liu, Department of Zoology, and Biodiversity Research Centre, University of British Columbia, 6270 University Blvd., Vancouver, BC V6T 1Z4, Canada.
Email: tetradon@zoology.ubc.ca

Funding information
Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: RGPIN-2014-02341 and RGPIN-2018-03926

Handling Editor: Nick Hamilton Barton

Abstract

Hybridization between divergent lineages can result in losses of distinct evolutionary taxa. Alternatively, hybridization can lead to increased genetic variability that may fuel local adaptation and the generation of novel traits and/or taxa. Here, we examined single-nucleotide polymorphisms generated using genotyping-by-sequencing in a population of Dolly Varden char (Pisces: Salmonidae) that is highly admixed within a contact zone between two subspecies (*Salvelinus malma malma*, Northern Dolly Varden [NDV] and *S. m. lordi*, Southern Dolly Varden [SDV]) in southwestern Alaska to assess the spatial distribution of hybrids and to test hypotheses on the origin of the admixed population. Ancestry analysis revealed that this admixed population is composed of advanced generation hybrids between NDV and SDV or advanced backcrosses to SDV; no F_1 hybrids were detected. Coalescent-based demographic modelling supported the origin of this population about 55,000 years ago by secondary contact between NDV and SDV with low levels of contemporary gene flow. Ancestry in NDV and SDV varies within the watershed and ancestry in NDV was positively associated with distance upstream from the sea, contingent on habitat-type sampled, and negatively associated with the number of migrations that individual fish made to the sea. Our results suggest that divergence between subspecies over hundreds of thousands of years may not be associated with significant reproductive isolation, but that elevated diversity owing to hybridization may have contributed to adaptive divergence in habitat use and life history.

KEYWORDS
adaptation, cline analysis, hybridization and introgression, secondary contact

1 | INTRODUCTION

The spatial extent, composition and evolutionary outcomes of hybrid zones are dynamic and often unpredictable. The fate of hybrid zones is influenced by the evolutionary history of the parental populations, intrinsic factors, such as barrier loci contributing to reproductive isolation, and extrinsic factors, such as natural or sexual selection acting on hybrids (Moran et al., 2021). The fate

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2024 The Author(s). *Molecular Ecology* published by John Wiley & Sons Ltd.

of hybrid zones often falls along a continuum from species collapse (the formation of hybrid swarms) to species persistence despite some gene flow (bimodal hybrid zones) (Abbott et al., 2013; Haines et al., 2019; Jiggins & Mallet, 2000; Taylor et al., 2006). Identification of the structure and genotypic composition of hybrid zones is a crucial step in understanding the evolutionary history and outcomes of contact between divergent populations, yet such outcomes and the evolutionary processes shaping them are relatively rarely addressed (Schumer et al., 2018). Hybridization may be associated with: 'genetic conflicts' between divergent lineages with negative consequences for either or both lineages; creative forces that promote adaptation by hybrids to novel environments and perhaps the evolution of new species via hybrid speciation or some combination of these outcomes (Arnold, 1997; Meier et al., 2017, 2023; Muhlfeld et al., 2009; Rieseberg et al., 1995; Schumer et al., 2018). Genome-level investigations provide powerful approaches to unravel hybrid zone structure and history, the extent and basis of reproductive isolation and/or introgression, and the identity of genomic regions associated with adaptation, reproductive isolation and hybrid speciation (e.g. Seehausen et al., 2014; Schumer et al., 2018).

The western North American char species complex (Pisces: Salmonidae, *Salvelinus* spp.) provides opportunities to test hypotheses about the history and consequences of secondary contact and hybridization. Many of these fishes inhabit a wide variety of aquatic habitats across the North Pacific, exhibit different life histories and engage in various degrees of hybridization (Taylor, 2016). This system consists of phenotypically variable species, subspecies and populations that occur in allopatry and sympatry (Taylor, 2016). For example, F_1 or later generation hybrids between sympatric Arctic char (*Salvelinus alpinus*, AC) and Dolly Varden (*Salvelinus malma*, DV) are rarely found, as gene flow appears to be highly restricted (Gharrett et al., 1991; May-McNally et al., 2015; Taylor et al., 2008). This result contrasts with other contact zones, such as those between DV and bull trout (*Salvelinus confluentus*, BT), that contain parental genotypes and a broad range of hybrid genotypes, including presumptive F_1 hybrids (Redenbach & Taylor, 2003; Taylor et al., 2001). These different outcomes of hybridization are thought to result from sympatric AC/DV adopting different life histories and spawning areas (lake resident, stream resident and/or and sea run respectively; May-McNally et al., 2015; Taylor et al., 2008), whereas an overlap in spawning location likely facilitates a more permeable contact zone between BT/DV (Redenbach & Taylor, 2003; Taylor, 2016).

Taylor and May-McNally (2015) documented a range of hybrid genotypes in a putative contact zone between subspecies of DV in southwestern Alaska and parts of the Gulf of Alaska: northern DV (*Salvelinus malma malma*, NDV), found north of the Alaska Peninsula and the Aleutian Islands, west to Russia and east to the Mackenzie River and southern DV (*S. malma lordi*, SDV), found in western Washington north to the southern margin of the Alaska Peninsula and the Aleutian Islands. A similar result has been observed using thousands of genomic markers (E.B. Taylor, A. Geraldes, & J. Shen, unpubl. data). There is, however, little known of the fine-scale geographic patterns of introgression between NDV/SDV, or if certain regions of the genome are more, or less, susceptible to introgression between the two subspecies in the contact zone. In the most intensively sampled area of the contact zone, the Chignik Lake watershed, admixed SDV/NDV occur in lake, river and estuarine habitats (Taylor & May-McNally, 2015). Bond, et al. (2014) found ecological and genetic differences between juvenile Chignik Lake watershed DV (CDV) populations collected from upstream tributaries of freshwater habitats and downstream tributaries draining directly to the sea. The authors argued that these differences were likely driven by life-history distinctions in migration timing and habitat use, but did not frame their work in the context of admixture between subspecies.

Here, we take advantage of this unique CDV population to generate genome-wide polymorphism data using genotyping by sequencing (GBS, Elshire et al., 2011) of fish sampled across the watershed to investigate the spatial distribution of admixed individuals and the potential role of selection in influencing the composition and distribution of hybrids. This contact zone also provides a basis for comparative analysis of the extent, timing, and factors involved in the evolution of reproductive isolation within a genus of closely related taxa. First, the genetic structure of admixed CDV was characterized in terms of the extent of NDV and SDV ancestry. Further, we characterized the generational classes of admixed CDV to understand whether the complex is the result of contemporary hybridization (e.g. presence of recent generation hybrids and parental NDV and SDV) or historical hybridization (e.g. absence of recent generation hybrids and parental NDV and SDV). We also modelled different historical scenarios for the origin of the admixed population and to explicitly test if it originated by secondary contact and subsequent gene flow. Next, we tested for an association between habitats sampled (estuarine, lake, river) and distance from the estuary, and genetic composition (i.e. proportion of NDV vs. SDV ancestry). The two subspecies have morphological, genetic and historical biogeographic differences between them (Kowalchuk et al., 2010; Phillips et al., 1999; Taylor & May-McNally, 2015), so it is plausible that the contact zone may, in part, be structured by different habitat preferences. Alternatively, any spatial patterns of introgression may simply be a function of geographic distance from the estuary. Then, given evidence of a link between life-history traits and genetic differentiation in salmonids (Barson et al., 2015; Pearse et al., 2019; Strait et al., 2021), we tested for an association between NDV/SDV ancestry and several life-history traits previously quantified in the CDV (see Bond, 2013; Bond, et al., 2014). We then used Bayesian genomic cline analysis (Gompert & Buerkle, 2011) of admixed DV to test for departures from neutral expectations for introgression across the genome and thus identify outlier genomic regions that may contribute to reproductive isolation or adaptive introgression. Finally, we used gene ontology (GO) analysis (Young et al., 2010) to see if outlier genomic regions detected from the genomic cline analysis were associated with biological processes that might be linked to life-history variation in Chignik DV. Overall, our work investigates the potential evolutionary processes influencing the genomic structure of a recently described contact zone.

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17459 by ERIC B TAYLOR , Wiley Online Library on [12/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2 | MATERIALS AND METHODS

2.1 | Sample collection

The Chignik Lake watershed in Alaska has a surface area of 1535 km^2 (Narver, 1966) and consists of habitats from headwaters to the estuary including Black Lake, Alec River, Black River, Chignik Lake, Chignik River and Chignik Lagoon (Figure 1). Chignik Lagoon is a semi-enclosed estuarine environment covering ~21 km^2 at high tide. The 2-km long Chignik River drains Chignik Lake and flows into the brackish water of Chignik Lagoon (Westley et al., 2010). Chignik Lake is a small (surface area = 22 km^2) and deep (mean depth = 64 m) lake, while Black Lake is shallower (mean depth = 1.5 m, maximum depth, 3 m), but larger (surface area = 41 km^2). Black Lake is joined to Chignik Lake by the 12 km long Black River which has two major tributaries, including the West Fork Chignik River to the west and Chiaktuak Creek to the east (Conrad, 1984; Westley et al., 2010). The Alec River drains into Black Lake from the northeast.

Whole DV fin clip samples were collected by the University of Washington Alaska Salmon Program from different habitats and locations throughout the Chignik Lake watershed across 20 sample sites in June, July and August between 2007 and 2010 (Figure 1, Table 1, Table S1; $N = 185$; see Bond, 2013). These sites were grouped into nine major localities (e.g. Chignik Lake, Black Lake) and five major regions within the watershed from low in the watershed (e.g. Chignik Bay, Chignik Lagoon) to its upper reaches (lower freshwater and upper freshwater). Samples spanned a range of body sizes (fork length; 81–461 mm) as a proxy for age and maturity and included a mix of juvenile, maturing and mature fish (Figure S1; Bond, 2013).

2.2 | DNA extractions and genotyping-by-sequencing library construction

Genomic DNA was extracted using Qiagen's DNeasy Blood & Tissue kit (Qiagen, Inc., Valencia, CA, USA). Concentration and

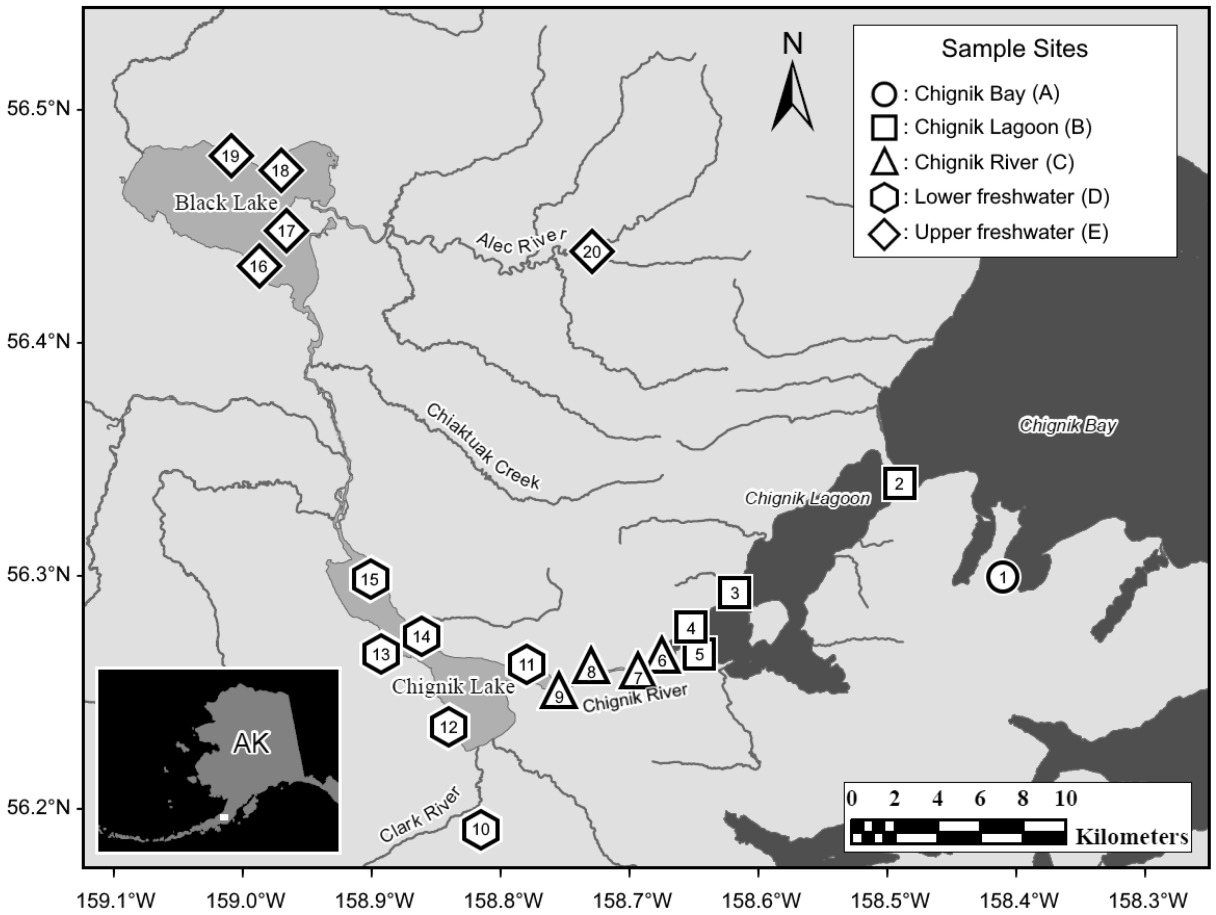

FIGURE 1 Samples of Dolly Varden (*Salvelinus malma*) collected from throughout the Chignik Lake watershed, southwestern Alaska. Numbers correspond to sample sites of fish, shapes correspond to location groups and letters correspond to the major regions (Table 1). Dark grey shading indicates brackish water (Chignik Lagoon) and saltwater (Chignik Bay). Medium grey shading indicates fresh water. Light grey shading indicates land. In the bottom left inset map, the white square shows the position of the Chignik Lake watershed on the Alaska Peninsula. Sample site names are: 1 = Indian Creek, 2 = Lagoon Spit, 3 = Lagoon Hume Point, 4 = Lagoon Alpha, 5 = Lagoon Pillar Rock, 6 = Chignik River Peahi, 7 = Chignik River Fish and Game Weir, 8 = Chignik River Braids, 9 = Chignik River FRI Camp, 10 = Clark River Bay, 11 = Chignik Lake Outlet, 12 = Chignik Lake Clark Bay, 13 = Chignik Lake North Hatchery, 14 = Chignik Lake Cucumber, 15 = Chignik Lake Delta, 16 = Black Lake Cabin Point, 17 = Black Lake Spit, 18 = Black Lake Alec Bay, 19 = Black Lake, 20 = Alec River.

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17459 by ERIC B TAYLOR , Wiley Online Library on [12/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Chignik Lake watershed Dolly Varden (*Salvelinus malma*) sampled.

Major regions (*N*)	Location group	Habitat type	Latitude and longitude	Mean fork length for location group in mm (SD, range)
A. Chignik Bay (14)	Chignik Bay (14/1/1)	Stream	56.3006, −158.4152	22.5 (4.8, 15–35)
B. Chignik Lagoon (58)	Lagoon Spit (14/1/2)	Estuary	56.3402, −158.4916	280.4 (73.1, 116–425)
	Lagoon Hume Point (15/1/3)	Estuary	56.2949, −158.6188	265.0 (65.7, 169–404)
	Lagoon Alpha-Pillar Rock (29/2/4-5)	Estuary	56.2790, −158.6496	268.5 (66.6, 150–401)
C. Chignik River (37)	Chignik River (37/4/6-9)	River	56.2668, −158.6747	249.3 (79.5, 133–461)
D. Lower freshwater (37)	Clark River (12/1/10)	River	56.1967, −158.8133	263.8 (90.5, 152–384)
	Chignik Lake (25/5/11-15)	Lake	56.2609, −158.7714	237.2 (78.5, 133–377)
E. Upper freshwater (39)	Black Lake (15/4/16-19)	Lake	56.4331, −158.9745	214.3 (108.6, 130–432)
	Alec River (24/1/20)	River	56.4387, −158.7246	381.9 (43.2, 305–449)

Note: Under location group numbers in parentheses are: sample size/number of sample sites within each location group/number for location in Figure 1.

quality (260/280 and 260/230 ratios) of DNA were measured using the Qubit dsDNA broad range kit (Thermo Fisher Scientific, Waltham, MA, USA), a Nanodrop 8000 spectrophotometer (Thermo Fisher Scientific) and by visualization using gel electrophoresis in 2% agarose gel and 1% SYBR safe DNA gel stain (Thermo Fisher Scientific).

Genotyping-by-sequencing libraries were created consisting of 185 CDV DNAs placed in microtiter plates along with negative controls (i.e. plate wells that contained all reagents but no DNA or no barcode). The GBS libraries were prepared using a modified protocol (Alcaide et al., 2014; Geraldes et al., 2019; Irwin et al., 2016; Toews et al., 2016). Briefly, the GBS protocol included: (i) DNA digestion using the restriction endonuclease (*Pst*I), (ii) ligation of adapters (barcode and common adapters) to samples, (iii) PCR amplification of samples, (iv) pooling of samples into two libraries, (v) size selection of DNA fragments by excision of the 400–700 base pair (bp) section from a 2% agarose gel, (vi) extraction and purification of the DNA from the agarose and (vii) analysis of the concentration and size of the isolated DNA using a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). The two libraries (mean fragment sizes of 411–449 bp) were sequenced by Genome Québec using two lanes of Illumina NovaSeq 6000 SP 150 bp.

2.3 | Bioinformatics and sequence filtering

The filtering pipeline was adapted from Irwin et al. (2016). The command process_radtags from the program Stacks (Catchen et al., 2013) was used to sort and remove barcode sequences from their respective samples. The program Trimmomatic (Bolger et al., 2014) was used to trim reads for quality by removing bases from the end of a read if their sequence quality fell below a phred score of 3 (TRAILING 3) and when the average quality threshold across a sliding window size of four bases dropped below a phred score of 10 (SLIDINGWINDOW 4:10). Whole reads were removed if below a minimum length of 30 bases (MINLEN 30). The trimmed reads were mapped to the reference genome representing *Salvelinus* sp., accession: GCA_002910315.2 (Christensen et al., 2018, 2021; a *S. malma malma* sequence with about 6% admixture with AC – Shedko, 2019; E.B. Taylor, A. Geraldes, & J. Shen, unpubl. data), using the Burrows–Wheeler Aligner (BWA) software and algorithm BWA–MEM (Li & Durbin, 2009). Aligned sequences in Sequence Alignment Map (SAM) files were converted into Binary Alignment Map (BAM) files using the function AddOrReplaceReadGroups.jar of Picard (Picard Toolkit. 2019. Broad Institute, GitHub Repository. https://broadinstitute.github.io/picard/). The SAMtools (Li et al., 2009) depth command was used to calculate average sequencing depth at each site. The program BEDtools 2.3 (Quinlan & Hall, 2010) and the command -genomecov was used to determine the percentage of the genome sequenced for reads of a given depth.

We used the command HaplotypeCaller from the program Genome Analysis ToolKit (GATK4, Van der Auwera & O'Connor, 2020) to identify variable sites from each sample's alignment BAM file and thus create genome variant calling files (GVCF) for each sample. We used GenomicDBImport to import each single-sample GVCF into a GenomicsDB workspace before jointly genotyping all samples to create a single master variant calling file (VCF) with the genotypes (and associated quality metrics) for all samples at all identified polymorphic sites. The function '-hets' was changed from the default setting of 0.001 based on human populations to 0.01 to account for the potential for hybrid individuals to occur in our dataset.

To place the fish sampled within the Chignik Lake watershed into the broader context of NDV and SDV, 21 allopatric NDV and 23 allopatric SDV from across the North Pacific basin were used as subspecies reference samples (see Taylor & May-McNally, 2015; E.B. Taylor, A. Geraldes, & J. Shen, unpubl. data, Table S1 and Figure S2). These samples comprised at least 99.99% NDV or SDV ancestry determined from admixture analysis (Alexander et al., 2009; E.B. Taylor, A. Geraldes, & J. Shen, unpubl. data). The GBS data from E.B. Taylor, A. Geraldes, & J. Shen (unpubl. data) were merged with the current

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17459 by ERIC B TAYLOR , Wiley Online Library on [12/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

data using the functions GenomicDBImport and GenotypeGVCFs within GATK4 to create one VCF file containing samples from both datasets: 185 newly sequenced DV from the Chignik Lake watershed, 21 reference NDV and 23 reference SDV sequences and 5 CDV sequences from E.B. Taylor, A. Geraldes, & J. Shen (unpubl. data). The five sequences of CDV from E.B. Taylor, A. Geraldes, & J. Shen (unpubl. data) were included to verify that the newly generated Chignik Lake sequences merged correctly with the data from the E.B. Taylor, A. Geraldes, & J. Shen library but they were not used in any subsequent analyses. Six of the newly-sequenced Chignik DV samples with more than 65% of missing loci (F_Miss > 0.65) and four other Chignik DV samples with a second degree or higher relationship (indicating at least 25% shared alleles) with another sample as estimated with the R program SNPrelate and the function snpgdsIBDKING with default parameters (Zheng et al., 2012) were removed from the dataset (Tian et al., 2008). The final dataset consisted of 175 CDV plus the 21 and 23 references NDV and SDV ($N = 219$ total; Table S1; Liu et al., 2023).

Next, the VCF file with the above 219 samples was filtered using the program VCFtools (Danecek et al., 2011). First, insertion and deletion sites (indels) were removed and only bi-allelic SNPs were kept (--remove-indels, --min-alleles 2, --max-alleles 2). Second, sites with high heterozygosity (>0.6), which may indicate paralogous variation that complicates read mapping (McKinney et al., 2017; Torkamaneh et al., 2016) were removed using a custom perl script (Owens et al., 2016). This master VCF file had 4,908,039 candidate SNPs and was the basis for all subsequent filtering steps to obtain SNP files tailored to different analysis. For analyses of population structure, we filtered the master file to obtain a POPSTRUCT VCF where sites with more than 30% missing genotypes (--max-missing 0.7), a genotype quality below 10 (--minGQ 10), minor allele frequency of less than 5% (--maf 0.05) were removed (resulting in a total of 135,205 SNPs). Plink v1.9 (Chang et al., 2015) was used to filter out SNPs in close linkage with other SNPs by removing one of any two SNPs with a value of r^2 equal to or greater than 0.2 within a sliding window of 50 SNPs, moving 10 base pairs at a time. This resulted in an 'LD-trimmed' POPSTRUCT VCF dataset of 44,548 SNPs across 219 fish. For demographic and diversity analyses filtering the dataset in ways that truncate the allele frequency spectrum is not advisable (Linck & Battey, 2019). Instead, we filtered our master VCF file to eliminate genotypes with GQ below 30 (--minGQ 30), depth below 15 (--minDP 15), more than 30% missing genotypes (--max-missing 0.7) and that were not variable after this filtering (--mac 1) resulting in a DEMODIV dataset with 431,502 SNPs. To eliminate SNPs in strong pairwise LD without truncating the site frequency spectrum we required that SNPs were at least 1 kb distant from one another (--thin 1000) resulting in an 'LD-trimmed' DEMODIV dataset with 54,590 SNPs.

2.4 | Dolly Varden hybrid composition and genetic structure within the Chignik Lake watershed

We produced estimates of genetic variation within the reference groups of NDV and SDV and within CDV using the program populations found in the Stacks pipeline (Catchen et al., 2013). The DEMODIV dataset with 431,502 SNPs was used for this analysis. To account for the larger sample size of newly-sequenced CDV (175) relative to NDV and SDV (21 and 23 respectively), we created eight files each with randomly-selected subsets of CDV individuals (N of each ranged from 21 to 23) while ensuring that each SNP was present in all groups (NDV, SDV and the eight subgroups of CDV). We then used the resultant 55,744 SNPs to calculate the diversity values within each group.

We used several complementary approaches to examine the genetic structure of Dolly Varden within the Chignik Lake watershed. First, we calculated Weir and Cockerham's (1984) weighted F_{ST} between NDV, SDV and CDV using Stacks and the same dataset used for estimates of genetic diversity described above. Second, we performed a principal component analysis (PCA) with the LD-trimmed POPSTRUCT dataset using SNPrelate (Zheng et al., 2012). Third, the number of genetic groups (K) within the Chignik Lake watershed and the proportion of ancestry (Q) within each of the K groups was assessed using Admixture v1.3.0 using the LD-trimmed POPSTRUCT dataset (Alexander et al., 2009). Each $K = 1$–10 group model was run five times and was terminated after the difference in log-likelihood between successive runs fell below 1×10^{-9} (Geraldes et al., 2019) using the unsupervised mode (i.e. ancestry of each individual was not fixed to any K group, Alexander et al., 2009). The K values were evaluated with respect to their associated cross-validation errors (CVE) across runs. A model of $K = 2$ (see Results) was then run with 1000 bootstrap replications to determine standard errors and 95% confidence intervals of estimated admixture proportions (i.e. NDV, SDV) of each CDV sample under an unsupervised model (using a supervised, with reference NDV and SDV, model did not improve estimates).

Finally, to assess whether there is evidence of current gene flow between NDV and SDV as indicated by the presence of parental and F_1 hybrids, hybrid triangle plots were used to plot estimated inter-specific heterozygosity (H_o) relative to the estimated hybrid index (H_i), as measured by NDV or SDV ancestry, using the R package HIest (Fitzpatrick, 2012). Values of H_o and H_i were estimated using maximum likelihood functions using the function HIest and maximum likelihood values for the most likely classification to hybrid class were estimated using the function HIclass (Fitzpatrick, 2012). To calculate H_o and H_i, we identified SNPs that were putatively diagnostic between the reference NDV and SDV using the VCFtools command --weir-fst-pop (i.e. markers with $F_{ST} = 1$) allowing up to 2% missing data (--max missing 0.98) resulting in 265 SNPs. Reference allele frequencies of 0.95 for one parent and 0.05 for the other – instead of the default of 0 and 1 – were assumed for the diagnostic SNPs in the HIest calculations to account for the limited sample size of our references NDV and SDV (Fitzpatrick, 2012). The direct correspondence between hybrid allele frequencies and hybrid generational class (e.g. F_2) applies only to the first two generations of interbreeding; consequently, later generational classes are more realistically considered as 'F_2-like' or 'backcross-like' (Fitzpatrick, 2012).

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17459 by ERIC B TAYLOR , Wiley Online Library on [12/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.5 | Historical demographic analyses

We tested various models for the origin of the CDV using the LD-trimmed DEMODIV dataset with 54,590 SNPs, e.g. whether the population originated via divergence from either NDV or SDV in a sequential isolation-by-distance manner or if its origin was due to secondary contact and admixture between NDV and SDV. We also evaluated these scenarios under potential pre- and/or post-divergence/admixture gene flow as well a model of a trifurcation with and without any gene flow (26 models overall; Figure S3).

Demographic scenarios were evaluated by estimating the likelihoods of the observed site frequency spectrum (SFS) under each scenario via simulation using fastsimcoal2 (version 2.7.0.9.3, Excoffier et al., 2013). The folded SFS was generated using easySFS.py script (I. Overcast, https://github.com/isaacovercast/easySFS, see also Gutenkunst et al., 2009). Because the SFS cannot be built with missing data, we projected our data down to 127 CDV samples (52,376 loci), 6 NDV samples (36,888 loci) and 6 SDV samples (34,856 loci). The resulting 3DSFS was used as the input for the fastsimcoal2 analyses. Using only polymorphic SNPs necessitated the fixing of the effective population size for one of either NDV, SDV or CDV in the fastsimcoal2 analyses. We used an estimate of N_e for NDV (33,000) derived from the relationship $N_e=(\pi/4\mu)$ with π set to 0.00079 as obtained from a sample of six NDV and $\mu=5.97\times10^{-9}$ (Bergeron et al., 2023; E.B. Taylor, A. Geraldes, & J. Shen, unpubl. data).

Each scenario was tested with 100 independent runs of fastsimcoal2 across 100,000 simulations and 40 expectation conditional maximization cycles (Excoffier et al., 2013). The Akaike information criterion (AIC) was calculated for the run with the highest likelihood within each scenario's set of 100 replicates. Subsequently, the ΔAIC between the model with the highest likelihood and each alternative each model was calculated. Relative likelihoods of each model (RL) were then calculated using the expression, $RL=e^{(-0.5\Delta AIC)}$ (Burnham & Anderson, 2002). Akaike weights, *w*, were calculated for each model as the RL of each model divided by the sum of RLs across all models (Burnham & Anderson, 2002; Excoffier et al., 2013).

Finally, we calculated confidence intervals for parameter estimates under the most likely model by generating 100 parametric bootstrap replicates of the SFS and re-estimating parameters 100 times for each replicate SFS using the parameter values from the best run of the best model as the initial values of each bootstrapping run (cf. Excoffier et al., 2023; James et al., 2021). To convert fastsimcoal2 parameter estimates into absolute values, we employed a DV generation time of 5 years and divided N_e estimates by two to account for diploidy of DV.

2.6 | Associations between admixture levels and habitat, geographic distance and life-history traits

To assess if DV ancestry was associated with habitat, a contingency table analysis was used to test for non-random association between Q_{NDV} (four bins of values) and habitat-type sampled (Table 1; river/stream, lake and estuary) with a Fisher's exact test (fisher. test function in R, 2000 replicates; R Core Team, 2021) because some cells contained fewer than 10 expected values (McHugh, 2013). We also tested for an association between Q_{NDV} ($K=2$ model) and distance upstream from Chignik Bay using piecewise linear regression (package segmented in R) after visual inspection of the data suggested that more than one linear model might be appropriate. The Q_{NDV} values were averaged within the nine *location groups* (Table 1) to account for spatial correlation among *sample sites* within *location groups*. To test if differences in life history were associated with ancestry, non-piecewise linear models were fit between the mean Q_{NDV} values and the means of values of several life-history traits for fish sampled from each locality. Size (fork length and mass), recent somatic growth measured as plasma insulin-like growth factor 1 concentration (IGF1) by age, gonad mass (calculated as a ratio to body size in mass), the number of times a fish had migrated to sea by age and whether the female parent of each fish sampled had migrated to sea during the summer prior to spawning ('maternal anadromy') were obtained from Bond (2013; Supplementary Methods).

2.7 | Bayesian genomic clines analysis

To test for candidate genomic regions that might contain loci contributing to reproductive isolation between NDV and SDV or loci that might be adaptive and introgress preferentially, we conducted a Bayesian genomic cline analysis using bgc v.1.03 (Gompert & Buerkle, 2012). The key parameters estimated in Bayesian genomic cline analysis for each locus are α and β. The parameter α denotes the position of the cline centre and infers the direction of introgression as excess ancestry from one reference population (positive α) or the other reference population (negative α). Loci that are identified as outliers for the α parameter may be involved in directional selection of homozygous genotypes from one reference population or selection against heterozygotes (Gompert & Buerkle, 2011; Oswald et al., 2019). The parameter β denotes the slope/shape of the cline and estimates the rate of transition from one reference population to the other as either faster than expected (positive β indicates a narrow cline) or slower than expected (negative β indicates a wide cline) (Gompert & Buerkle, 2012; Janoušek et al., 2015; McFarlane et al., 2021). Loci that are identified as outliers for the β parameter may be involved in reproductive isolation (positive β) or adaptive introgression (negative β).

The input files for bgc comprised the dataset of 21 reference NDV, 23 reference SDV and 175 admixed Chignik Dolly Varden ('LD-trimmed' POPSTRUCT dataset of 44,548 SNPs). Five independent runs of bgc using the genotype certainty model were conducted. Each run had 75,000 MCMC steps, discarding the first 50,000 samples as burn-in and sampling every 25th iteration for a total of 1000 MCMC samples. Convergence for each run was assessed by visual inspection of the distribution of log-likelihood values from

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17459 by ERIC B TAYLOR , Wiley Online Library on [12/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

each MCMC as approximately randomly distributed (Gompert & Buerkle, 2012; McFarlane & Pemberton, 2019). The five runs were merged and the widest confidence intervals from the five runs were used to conservatively determine loci with significant α and β values (Janoušek et al., 2015; McFarlane & Pemberton, 2019).

The detection of outlier loci was performed in two steps. First, SNPs with α or β values that were significantly different from zero (i.e. expected for neutral loci) were identified as those with confidence intervals that did not overlap zero. Second, from those non-zero α or β SNPs, putative outlier SNPs were identified as those that were statistically 'unlikely', given the distribution of values across all loci; i.e. those with mean non-zero α or β values within the top and bottom 2.5% quantiles of a normal distribution of all SNPs assuming the same mean and standard deviation (Martin et al., 2021; McFarlane & Pemberton, 2019).

2.8 | Gene ontology: Putative biological processes for genomic regions of interest

A GO enrichment analysis was used to identify potential biological processes associated with sequences within outlier genomic regions of CDV using GOrilla (Eden et al., 2009). Windows of 100,000 bp encompassing each outlier region were used to query the *S. m. malma* genome for overrepresented gene and protein classes. The program BEDtools 2.3 (Quinlan & Hall, 2010) and the function -intersect were used to extract all target gene and protein IDs and annotations. The background set of all *S. m. malma* genes was downloaded from NCBI's Entrez Gene database (Maglott et al., 2011).

The GO database is species neutral, meaning that gene annotation information can be transferred across species and their sequences (Primmer et al., 2013). Many *S. m. malma* gene IDs do not have assigned Entrez gene IDs. Consequently, *S. m. malma* annotation descriptions were used to search against the same descriptions from three other species (*Homo sapiens*, *Mus musculus*, and a salmonid fish, *Oncorhynchus mykiss*; Grummer et al. 2021) to obtain as many matches as possible to Entrez gene IDs.

Two different background gene lists were used to run the enrichment analysis for biological processes with the target set of genes IDs: one independent run contained a background set that included target gene IDs and the other contained a background set that excluded the target set of gene IDs. Enrichment scores are calculated for each gene ID and corresponding GO term using a minimum hypergeometric test (mHG) statistical framework (Eden et al., 2007). To obtain a p value for the scores for each GO term occurrence, the scores are then compared to another mHG score under the null hypothesis that all given background genes have the same probability of occurring. A maximum threshold of $p < 0.001$ was used to declare significance (Eden et al., 2009). To summarize GO terms by grouping functionally similar categories using a representative GO term guided by p values, the output from GOrilla was visualized using 'treemaps' with the program REVIGO (Supek et al., 2011).

3 | RESULTS

3.1 | Genetic composition and structure within Chignik Lake Dolly Varden

The mean sequence quality across 2,160,982,102 raw reads was 36 (Q; inferred base call accuracy of 99.9%). The mean depth per sequenced and mapped locus was 23.7×. An average 1.12% of the genome was mapped with a coverage of at least 10× sequencing depth per locus.

The five Chignik Lake DV previously sequenced by E.B. Taylor, A. Geraldes, & J. Shen (unpubl. data) fell well within the PC1 and PC2 space of the newly sequenced Chignik verifying that the newly-sequenced Chignik Lake population dataset merged correctly with sequences of SDV and NDV generated by E.B. Taylor, A. Geraldes, & J. Shen (unpubl. data) (Figure S4).

After sub-sampling the CDV to create samples with similar sizes to those in the reference NDV and SDV, genetic variation was consistently higher in CDV (Table S2). Pairwise weighted F_{ST} was slightly higher between CDV and SDV (0.058) than between CDV and NDV (0.055) and was highest between NDV and SDV (0.152).

Across all the newly sequenced samples and the reference NDV and SDV, the first principal component axis (8.1% total variance) separated NDV and SDV; CDV were intermediate in position along PC1 with a slight bias towards SDV (Figure 2). The second and third PCs (2.3% and 1.3% of total variance, respectively) separated fish from upper and lower freshwater regions and a few from Chignik Lagoon and Chignik River from fish sampled from Chignik Lagoon and Chignik Bay (Figure 2; Figure S5). The fourth PC (0.92% of total variance) showed the strongest degree of segregation among samples from within the Chignik Lake watershed; fish from Chignik Bay, Chignik Lagoon and upper freshwater regions were diverged from fish from the lower freshwater and Chignik River regions (Figure S5).

Unsupervised admixture models for 175 CDV and 21 reference NDV and 23 reference SDV indicated that models of K from 2 to 5 produced a 'trough' of low CVE values (Figure S6). A model of $K=2$ was associated with a major difference between NDV and SDV (Figure 3). The Chignik Lake watershed samples fell within a range of $Q_{NDV}=0.25$ and 0.67, with a mean of 0.51 (SD = 0.38, median = 0.58), consistent with their intermediate position relative to NDV and SDV along the PC1 axis (Figure 2). Models of $K=3$–4 produced the lowest CVEs (about 1.7% lower; Figure S6) and revealed further structure within CDV. For instance, the $K=3$ model identified a third group exclusive to the Chignik Lake watershed (labelled as 'CDV' in Figure S7), with little NDV ancestry and decreasing contribution of SDV from Chignik Bay to the upper freshwater region. Finally, the $K=4$ admixture model resolved four different groups: NDV, SDV and two groups of CDV, one predominating in upstream freshwater portions of the watershed and one in the estuary portions, with some admixture between the latter two groups (Figure S7). In subsequent analyses, we used the Q_{NDV} values for each fish calculated from the admixture analyses under a model of $K=2$ representing the major subdivision between NDV and SDV.

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17459 by ERIC B TAYLOR , Wiley Online Library on [12/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

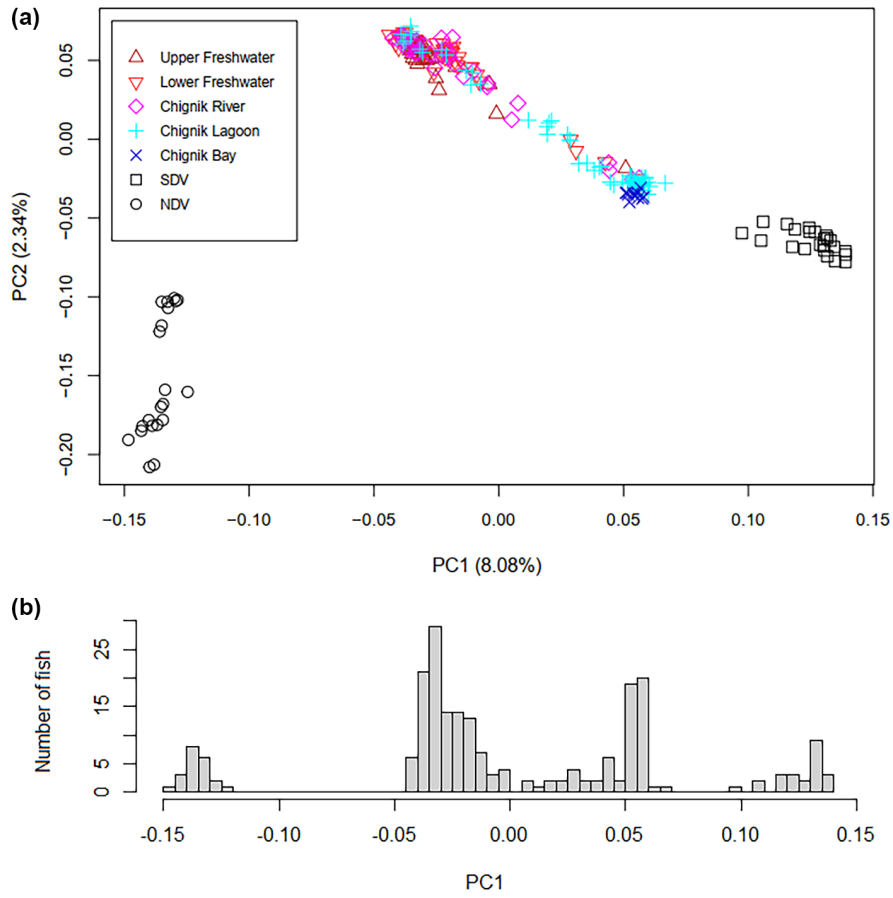

FIGURE 2 Principal components analysis with the LD-trimmed POPSTRUCT dataset with 44,548 SNPs including 175 Chignik Lake watershed Dolly Varden (*Salvelinus malma*), 21 reference northern Dolly Varden (NDV, *S. m. malma*) and 23 reference southern Dolly Varden SDV, (*S. m. lordi*) for (a) principal component (PC) axes one and two with amount of variation in the dataset explained along each axis shown in parentheses and (b) histogram of PC1 values plotted against number of fish showing the sample clusters along PC1. Values along PC1 between −0.15 and −0.10 represent those of reference NDV and between 0.09 and 0.15 those of reference SDV.

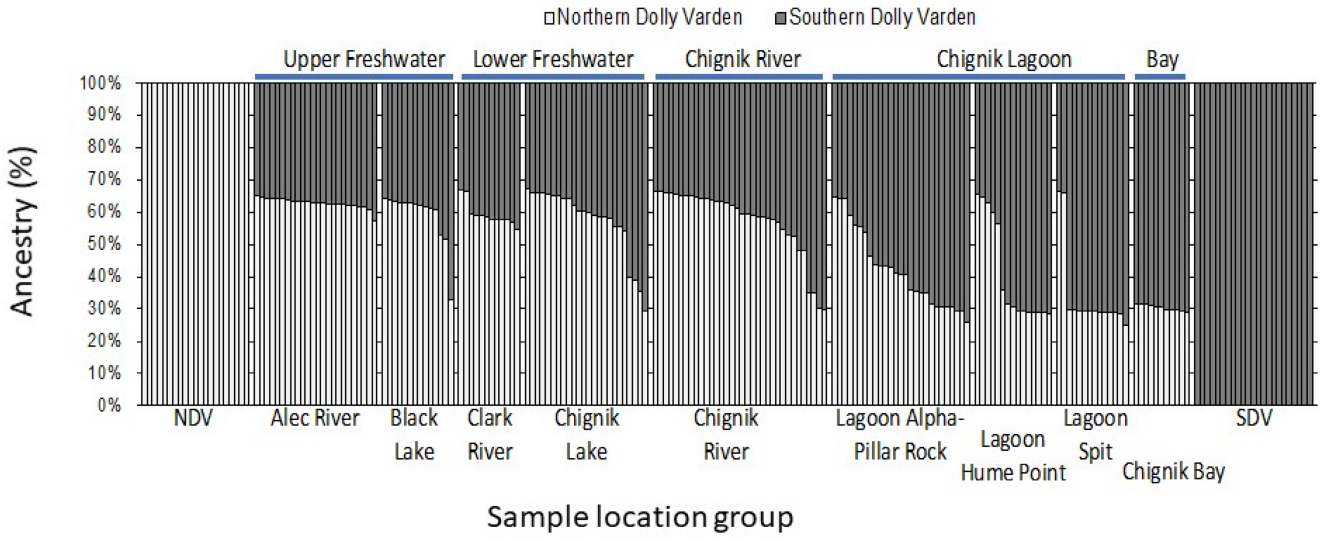

FIGURE 3 Unsupervised admixture model of 175 Chignik Lake (*Salvelinus malma*), 21 reference northern Dolly Varden (NDV, *S. m. malma*) and 23 reference southern Dolly Varden (SDV, *S. m. lordi*) by percentage of ancestry (*y*-axis) explained with $K = 2$ genetic groups corresponding to NDV (light grey) and SDV (dark grey). Chignik Lake Dolly Varden sample location groups are indicated below the plot, major location groups (see Table 1) are indicated above the plot and both are arranged from freshwater headwaters (left) to the ocean (right).

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17459 by ERIC B TAYLOR , Wiley Online Library on [12/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The triangle plot using the 265 unlinked SNPs diagnostic of reference NDV and SDV (up to 2% missing data) suggested that the 175 samples from the Chignik Lake watershed were probably F_2-like or later generation hybrids ($N = 150$), or had genotypes expected for SDV backcross-like samples ($N = 25$); all had moderate inter-specific heterozygosity over a limited range (0.38–0.59, Figure 4). By contrast, no parental NDV, SDV or F_1 generation hybrids were detected in the Chignik Lake watershed from the triangle plots or using maximum likelihood classifications (Figure 4). Analyses with allele frequencies set at the estimated values (1 and 0) or assuming much lower allele frequency differences (0.85 and 0.15) also resulted in the absence of NDV, SDV or F1 genotypes in the CDV samples. Across different regions in the watershed (Chignik Bay, Chignik Lagoon, Chignik River, middle freshwater and upper freshwater), NDV ancestry increased with distance from Chignik Bay to upper freshwater (Figure 4).

3.2 | Historical demographic analyses

Tests of different demographic hypotheses for the origin of the CDV indicted that a model of secondary contact followed by admixture between NDV and SDV with contemporary gene flow among NDV, SDV and CDV provided the best fit to our data (Table S3; model Akaike weight, $w = 0.949$). All other models, except for one that involved NDV–SDV admixture and both historical and contemporary gene flow ($w = 0.051$) received no support (all $w = 0$).

Demographic parameters estimated under the favoured model ranged from current effective population sizes (N_e) of about 27,000 (SDV) to 2,700,000 (CDV) and population size changes from the common ancestor estimate of about 38,000 (Table 2). Secondary contact and admixture between NDV and SDV to generate CDV was estimated to have occurred about 55,000 years ago (ya; range 50,170–60,780 ya) and divergence time between NDV and SDV

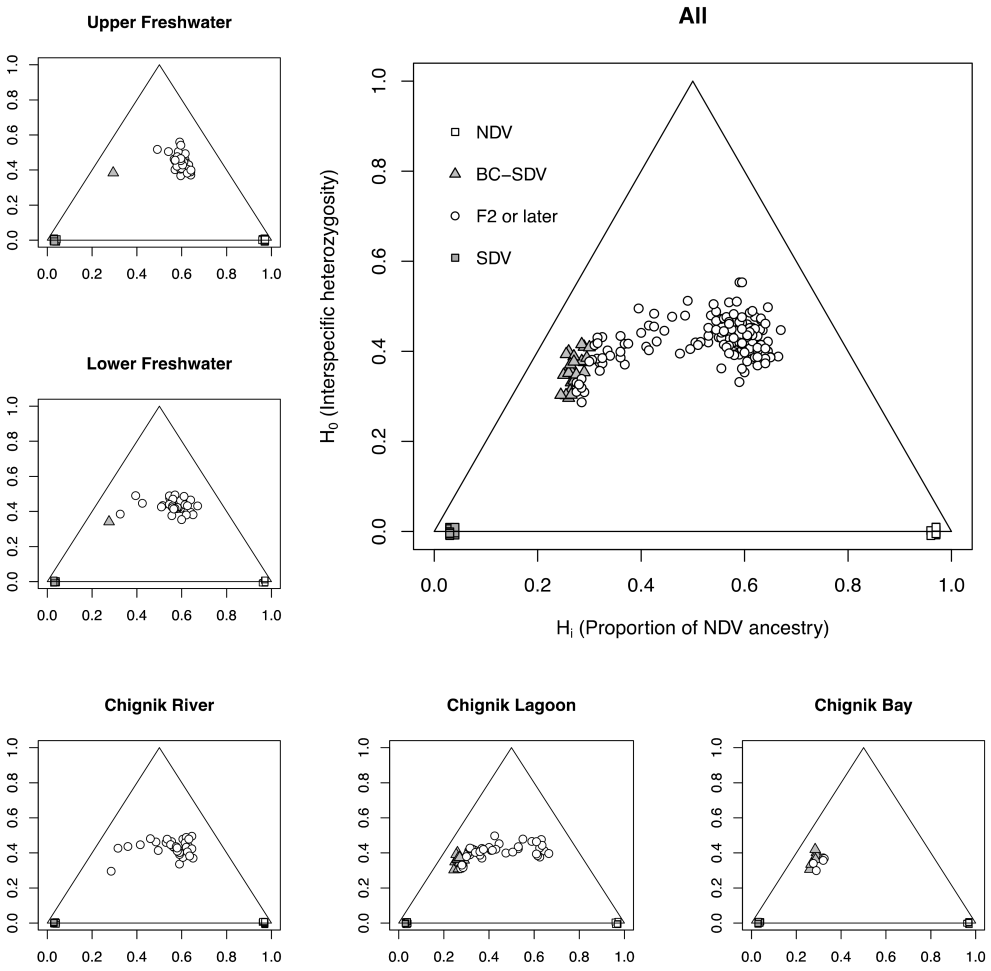

FIGURE 4 Triangle plots of inter-specific heterozygosity (H_o) by proportion of northern Dolly Varden (*Salvelinus malma malma*) ancestry (H_i) estimated using 265 unlinked loci with fixed differences and maximum 2% missing data between southern and northern Dolly Varden. All plots include 21 reference northern Dolly Varden (NDV) and 23 reference southern Dolly Varden (SDV, *S. m. lordi*). The plot labelled 'All' includes all 175 Chignik Lake watershed DV (*Salvelinus malma*) samples and smaller plots show only subsets of samples from the different major regions of the Chignik Lake watershed (See Table 1 for details). Shapes represent best hybrid generational class identified by maximum likelihood analysis: Backcross to SDV in grey triangles and F2 or later generation hybrids in open circles.

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17459 by ERIC B TAYLOR , Wiley Online Library on [12/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ancestral lineages was estimated to have occurred more than 10 times earlier at about 600,000 ya (range 523,330–731,140 ya). Estimates of contemporary migration probabilities per generation were very low; the highest occurred from SDV to CDV (~1.0×10^{-5}) – almost three times the estimated level of migration from NDV to CDV (Table 2). The lowest migration probabilities occurred from CDV both into SDV and NDV.

TABLE 2 Demographic parameter estimates under the admixture between northern and southern Dolly Varden (NDV, *Salvelinus malma malma* and SDV, *S. m. lordi* respectively) with contemporary gene flow (among NDV, SDV and Chignik Lake DV, CDV) model for the origin of Chignik Lake watershed admixed DV from fastsimcoal2 analyses.

Parameter	Lower bound	Mean point estimate (range)	Upper bound
N_eANC	37,469	38,651 (19,193–57,013)	39,834
N_eNDV	NA	NA	NA
N_eCDV	2,652,478	2,688,252 (2,305,185–3,232,212)	2,724,027
N_eSDV	27,043	27,313 (23,364–31,233)	27,582
T_{Admix}	55,050	55,488 (50,170–60,780)	55,927
$T_{Diverge\_NDV\_SDV}$	590,035	597,171 (523,330–731,140)	604,308
$cm_{NDV\text{-}SDV}$	9.3×10^{-7}	9.7×10^{-7} (2.7×10^{-8}–1.5×10^{-6})	1.1×10^{-6}
cm_{SDV_NDV}	4.3×10^{-7}	4.6×10^{-7} (8.0×10^{-10}–8.5×10^{-7})	4.9×10^{-7}
$cm_{CDV\text{-}SDV}$	1.0×10^{-5}	1.1×10^{-5} (6.5×10^{-6}–2.0×10^{-5})	1.2×10^{-5}
$cm_{SDV\text{-}CDV}$	7.7×10^{-8}	1.1×10^{-7} (3.8×10^{-10}–9.2×10^{-7})	1.4×10^{-7}
$cm_{NDV\text{-}CDV}$	4.4×10^{-8}	1.0×10^{-7} (1.8×10^{-10}–2.7×10^{-6})	1.6×10^{-7}
$cm_{CDV\text{-}NDV}$	4.67×10^{-6}	4.8×10^{-6} (2.5×10^{-6}–7.2×10^{-6})	5.0×10^{-6}

Note: No estimate is provided for contemporary effective population (N_e) size in NDV as it was fixed at 33,000 to facilitate the estimation of all other parameters. All times (T) are in years before present, c = contemporary probability of migration per generation (m), ANC = common ancestor of all DV. The 95% confidence intervals were calculated by estimating parameters from 100 parametric bootstrapped site frequency spectrum matrices using fastsimcoal2 under the admixture with contemporary gene flow model.

3.3 | Associations between ancestry proportion and habitat, geographic distance and life-history traits

The numbers of fish within different Q_{NDV} bins was significantly associated with habitat sampled (Fisher's exact test, $p<.001$); CDV with more NDV ancestry were found in upper freshwater areas, whereas those with more SDV ancestry were found closer to Chignik Lagoon and Chignik Bay. Distance (km) of sample sites upstream from Chignik Bay was positively correlated with Q_{NDV} and suggested a break in the relationship at about 33 km upstream (piecewise regression, $p<.01$, adjusted $r^2=.89$, df = 5; Figure S8). Changes in Q_{NDV} were the greatest within the first 30 km upstream from the sea corresponding to the transition between the Chignik Bay/Lagoon samples and the Chignik River. Values of Q_{NDV} showed smaller changes both between bay and lagoon areas and between the Chignik River and upstream.

The number of seaward migrations at age of sampling was negatively associated with Q_{NDV} ($p<.015$, adjusted $r^2=.59$, df = 6); fish sampled from locations with lower Q_{NDV} had a higher and more variable number of seaward migrations (0–5), whereas a higher Q_{NDV} was associated with a lower number of migrations (0–2) (Figure S9).

Whether the female parent migrated to the sea the summer prior to spawning, recent somatic growth (concentration of IGF1), fish mass, fork length and gonad mass as a ratio to body mass were non-significantly associated with Q_{NDV} ($p=.08$–$.96$, df = 6; e.g. Figure S10; Liu, 2021).

3.4 | Bayesian genomic cline analysis

Across the 44,548 LD-trimmed SNPs analysed with five independent runs of bgc, there were 7488 SNPs with significantly positive α values (representing excess ancestry in NDV), of which 3869 also fell within the top 2.5% distribution of all loci and thus were classified as outliers (Table 3; e.g. Figure 5). Moreover, 1952 SNPs had significantly negative α values (representing excess ancestry in SDV), but only one SNP was found within the bottom 2.5% of the distribution of all SNPs and thus considered an outlier. In contrast, only one SNP was found to have a significantly negative β value within the bottom 2.5% of the distribution of all SNPs, thus indicating a relatively wide cline (Table 3). No SNPs were associated with significantly positive β values.

TABLE 3 Bayesian genomic cline analysis in Chignik Lake watershed admixed northern and southern Dolly Varden (NDV, *Salvelinus malma malma* and SDV, *S. m. lordi*, respectively) showing the number of single-nucleotide polymorphisms (SNPs) that were identified as having α or β estimates with 95% confidence intervals (CI) that did not include 0 and that were also within the top or bottom 2.5% quantiles (out of a total 44,548 SNPs).

Category	Number of SNPs (CI not including 0)	Number of SNPs (top/bottom 2.5% quantiles)	Interpretation
Positive α	7488	3869	Excess NDV ancestry
Negative α	1952	1	Excess SDV ancestry
Positive β	0	0	Associated with reproductive isolation (steep cline)
Negative β	1	1	Associated with adaptive introgression (broad cline)

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17459 by ERIC B TAYLOR , Wiley Online Library on [12/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

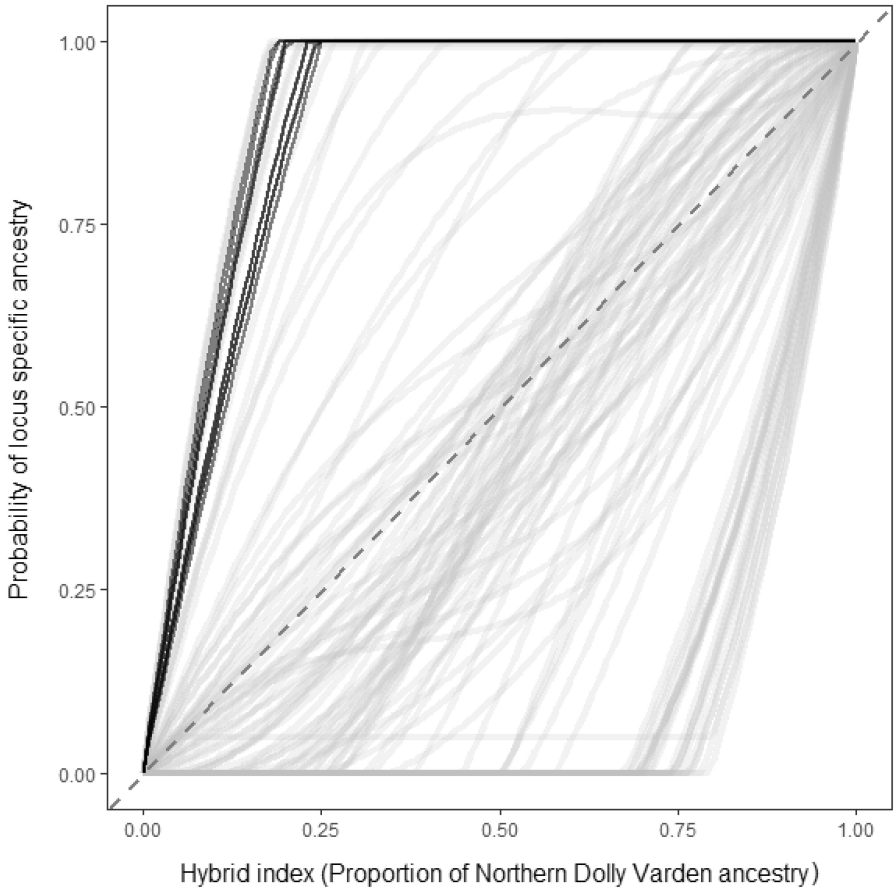

FIGURE 5 Representation of Bayesian genomic cline analysis for 50 randomly selected outlier loci that had significantly positive α values (representing excess ancestry in northern Dolly Varden, *S. m. malma*, black lines) against 100 randomly selected non-significant loci (grey lines) out of a total 44,548 SNPs. Dotted line represents a model of $\alpha = 0$ and $\beta = 0$.

3.5 | Gene ontology: Putative biological processes for genomic regions of interest

The 3869 loci classified as positive α outliers (excess NDV ancestry) were identified as 2745 unique *S. m. malma* gene IDs within a 100,000 bp sliding window. Of these, 2155 were used as the target set of unique Entrez gene IDs for the GO enrichment analysis after querying uncharacterized gene IDs against the gene descriptions of *H. sapiens*, *M. musculus* and *O. mykiss*. The *S. m. malma* genome contains 46,253 genes IDs and annotations with most (36,445) being uncharacterized genes. Of those, 34,250 were matched to an Entrez ID after gene description query and a final 19,296 unique Entrez gene IDs were used as the background set (subsequent results were identical regardless of whether the background set included or excluded the target genes). The top 50 GO terms for biological processes/gene functions organized by functional groups suggested that outlier SNPs were associated with: intracellular signal transduction ($p = 7.28\ e^{-08}$), neuron projection morphogenesis ($p = 5.29\ e^{-07}$), vesicle-mediated transport ($p = 1.83\ e^{-05}$) and cyclic guanosine monophosphate (cGMP) biosynthetic process ($p = 6.78\ e^{-05}$; Table S4, Figure S11).

4 | DISCUSSION

Contact zones between recently diverged populations that hybridize provide the chance to assess the extent and causes of reproductive isolation and the factors that might impact whether distinct lineages persist over time and space. For subspecies, defined as diagnosable populations of species that inhabit different geographic areas, but that can successfully interbreed when in contact, the outcomes of such interbreeding range from the formation of hybrid swarms (Forbes & Allendorf, 1991) to production of few hybrids and the potential elevation of such subspecies to species (e.g. Lovell et al., 2021). Even in cases of hybrid swarm formation, the production of recombinant genotypes provides opportunities to study the evolutionary fate of admixed individuals as well as that of single and multiple loci (Kagawa & Seehausen, 2020; Malek et al., 2012; Matute et al., 2020). Such fates may include adaptation to novel environments (e.g. Arnold, 1997; Meier et al., 2023) and/or the evolution of reproductive isolation from parental lineages, e.g. hybrid speciation (Rieseberg et al., 1995; Schumer et al., 2018). Here, we studied a population complex of Dolly Varden char within a contact zone between two well-differentiated subspecies with implications for

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17459 by ERIC B TAYLOR , Wiley Online Library on [12/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

inferring the processes generating and structuring hybrid zones and for the timing and extent of the evolution of reproductive isolation.

4.1 | Genetic composition, structure and origin of DV of the Chignik Lake watershed

The Chignik Lake watershed, a relatively large watershed on the Alaska Peninsula, is part of a known contact zone between NDV and SDV in North America (Taylor & May-McNally, 2015). Our PCA and admixture analysis found no fish in the Chignik Lake watershed that were characteristic of reference NDV or SDV (e.g. Q values > 95% NDV or SDV). Instead, our results showed that the CDV were genetically intermediate between NDV and SDV. Indeed, the variation in CDV PC1 scores was substantial relative to the variation in NDV and SDV sampled across a much greater range of geography. Similarly, CDV included no fish classified as parental NDV, SDV or F_1 hybrids. Rather, most fish were inferred to be F_2 and later generation hybrids and backcross like. That no CDV were classified as NDV, SDV or F_1 hybrids rests on accurate estimate of allele frequencies in the reference populations, i.e. that they are fixed or nearly so and the HIclass procedure is robust to some uncertainly in parental allele frequencies (Fitzpatrick, 2012). Even so, our results are consistent with those of Taylor and May-McNally (2015), who found no F_1 hybrids in the Chignik Lake watershed using microsatellite DNA in an independent group of samples.

Several lines of evidence suggest that the admixed DV in the Chignik Lake watershed represent a zone of historical introgressive hybridization accompanied by many generations of recombination following secondary contact between NDV and SDV. First, after accounting for differences in sample size, the CDV show greater genetic diversity than either NDV or SDV reference populations – a pattern consistent with hybridization within a contact zone between two widely separated populations (e.g. Muniz et al., 2022; Szymura & Barton, 1986). Second, from the admixed Chignik Lake watershed, there is a transition to fish of mostly non-admixed NDV found a maximum of 340 km further west along the Aleutian Islands (e.g. in Russell Creek and Frosty Creek), and to fish of mostly SDV ancestry beginning about 300 km northeast of Chignik Lake on Kodiak Island (Taylor & May-McNally, 2015; E.B. Taylor, A. Geraldes, & J. Shen, unpubl. data). In marked contrast to the typical distribution of non-admixed populations of NDV and SDV found across thousands of kilometres ranging from Yukon west and south to Bristol Bay and west to Russia and from Washington State north and west to at least southeast Alaska, respectively (Taylor & May-McNally, 2015; E.B. Taylor, A. Geraldes, & J. Shen, unpubl. data), the transitions to non-admixed NDV and SDV populations are relatively abrupt. Third, our demographic modelling found that a model of secondary contact followed by admixture between NDV and SDV and contemporary gene flow (among NDV, SDV and CDV) received by far the greatest support among all those evaluated to account for current patterns of genomic variation. No model involving a scenario of sequential divergence among NDV, SDV and CDV received any support. The admixture event was estimated to have occurred about 55,000 ya following divergence between NDV and SDV that was estimated to have occurred about 10 times as long ago (~600,000 ya). Such a pronounced difference in the divergence time between NDV and SDV is perhaps expected given the many opportunities for isolation and re-contact provided by the many Pleistocene glaciations that impacted the North Pacific. For instance, the estimated divergence time between NDV and SDV corresponds to a period of several major and minor glaciations in the Northern Hemisphere between 0.85 and 0.41 million ya (Ehlers et al., 2018). Although the spatial patterns of these early glaciations are poorly known, during the most recent Wisconsinan glaciation (~75,000–12,000 ya), the Alaska Peninsula and adjacent islands were heavily glaciated, creating more pronounced isolation between the northern (Bering Sea) and southern (North Pacific) margins of the peninsula (Kaufman & Manley, 2004). It became wider by a 100-km or more and the seafloor among eastern Aleutian Islands became exposed illustrating the potential for allopatric divergence within DV distributed across this region (Kaufman & Manley, 2004). The estimated timing of admixture between NDV and SDV (~55,000 ya) overlaps with the mid-Wisconsinan glaciation. Consequently, it is possible that environmental changes associated with fluctuations in glacier extent at this time brought previously isolated NDV and SDV into contact and promoted hybridization and introgression (e.g. Moore et al., 2015; Paun et al., 2006). For instance, Knappen (1929) argued that Black Lake and the West Fork Chignik River once flowed to the Bering Sea and that their current flow to the Pacific was established post-glacially. Although the Pacific drainage capture of the upper Chignik Lake watershed was established post-glacially, it illustrates the potential for changes in watershed orientations that could provide opportunities for contact between NDV (concentrated in Bering Sea drainages) and SDV (Pacific basin drainages).

Finally, secondary contact between subspecies of DV also appears to occur between the Asian SDV (*S. m. krascheninnikovi*) and NDV across the Sakhalin/Hokkaido islands and the Kamchatkan Peninsula and the Sea of Okhotsk in the northwestern Pacific, with mitochondrial DNA evidence showing historical hybridization between subspecies (Salmenkova & Omelchenko, 2013; Yamamoto et al., 2021). A history of population structuring around the Alaska Peninsula involving secondary contact has also been suggested for beluga whales, *Delphinapterus leucas* (O'Corry-Crowe et al., 1997; Rugh et al., 2010) and chum salmon, *Oncorhynchus keta* (Seeb & Crane, 1999; Petrou et al., 2013).

Basin-wide, Chignik Lake DV had genome-wide ancestry biased towards SDV rather than NDV along the PC1 axis and in admixture analyses, especially in downstream portions of the watershed (near the lagoon and estuary). This pattern was also present in the piecewise regression of Q_{NDV} versus distance upstream from Chignik Bay and the contingency analysis across habitats. Fish sampled from the Chignik River showed the most variation in Q_{NDV} suggesting the river is a transitional zone between these areas. Ancestry values of Chignik DV may not be uniformly distributed across the watershed

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17459 by ERIC B TAYLOR , Wiley Online Library on [12/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

owing to the action of selection on fish of different admixture values in different habitats, likely enhanced by the well-known behaviour of salmonid fishes to home to natal habitats for spawning. Bond, et al. (2014) found strong differentiation in microsatellite allele frequencies among CDV sampled from different habitats and argued that the differentiation was mainly driven by selection for alternative life histories. Bond, et al. (2014), however, could not assess the role of secondary contact and admixture between subspecies because they did not study reference NDV or SDV from outside the watershed. Alternatively, perhaps not enough time has passed since secondary contact and hybridization for dispersal to homogenize ancestry values across the watershed.

Bond, et al. (2014) found that the most significant differentiation in the watershed was between a freshwater (Bear Creek, a tributary of the Chignik River) and an estuarine sample (Metrofania Creek draining to Chignik Lagoon). These authors also found that there was a significantly positive correlation between inter-locality salinity differences and genetic differentiation after controlling for geographic distance (Bond, et al., 2014). Further, sockeye salmon (*Oncorhynchus nerka*), pygmy whitefish (*Prosopium coulterii*) and threespine stickleback (*Gasterosteus aculeatus*) within the Chignik Lake watershed also show habitat-based differentiation (Creelman et al., 2011; Gowell et al., 2012; Simmons et al., 2013; Taugbøl et al., 2014).

Our data showed that the total number of seaward migrations at age of sampling was significantly negatively correlated with Q_{NDV}. Studies of other salmonid fishes have shown that admixture values may be associated with migration behaviour (Bourret et al., 2022; Strait et al., 2021; Yates et al., 2015). Altogether, these observations, the current analysis and those of Bond, et al. (2014) and Taylor and May-McNally (2015), support fundamental genetic distinctions between fish from streams draining to the estuary and those further upriver in freshwater environments or with different life histories, and that such distinctions may result from a combination of secondary contract, admixture between subspecies and ecologically based selection in the Chignik Lake watershed.

4.2 | Genomic introgression and enrichment analysis

Almost all outlier loci identified in the Bayesian genomic cline analysis were associated with excess ancestry in NDV, i.e. a Chignik DV with 50% ancestry both from NDV and SDV genome-wide is more likely to be homozygous for NDV alleles at the outlier loci, rather than heterozygous or homozygous for SDV alleles. Excess NDV ancestry at outlier loci suggests that some NDV alleles could be favoured by selection, though this interpretation is made cautiously given the high variability of estimates of α with the modest sample sizes we had (<500 fish, Gompert & Buerkle, 2011; McFarlane et al., 2021).

Only one significant β-outlier locus representing a SNP with relatively broad cline was detected. Hybrid zones in chub fishes (*Macrhybopsis*; Alex Sotola et al., 2019) and North American pine (*Pinus*; Menon et al., 2018) exhibited few to no early-generation hybrids, many more α outlier loci than β outliers (see also Oswald et al., 2019; Pulido-Santacruz et al., 2018) and may also stem from weak reproductive barriers and asymmetrical introgression promoted by ecologically based selection. The virtual absence of β outliers, along with the prevalence of positive α outliers, in Chignik DV may stem from weak to no selection against interbreeding between NDV and SDV, accompanied by ecological selection favouring some recombinant genotypes with excess NDV ancestry across habitats (Bond, et al., 2014; Gompert et al., 2012; Menon et al., 2018). The lack of non-admixed genotypes in the Chignik watershed, however, may have reduced the power for the bgc analysis to detect significant β loci. Further, the low percentage of the genome covered by reduced representation sequencing methods like GBS may constrain genomic cline analysis especially given that reproductive isolation is typically influenced by many loci (e.g. Payseur & Rieseberg, 2016). Our analyses simply raise the hypothesis of adaptive introgression that could be tested in various ways to rule out genetic drift as an alternative explanation (Arnold & Martin, 2009; Gompert et al., 2017; Gompert & Buerkle, 2011; Oswald et al., 2019; Pulido-Santacruz et al., 2018; Schield et al., 2015).

Indeed, our GO enrichment analysis and the associations between DV ancestry and life-history traits that we documented raise the hypothesis that some of the 3869 outlier α loci could signal gene functions under selection related to osmoregulation and immune function. Both osmoregulation and immune function are critically important to the physiology of Chignik DV that migrate between fresh water and the sea (Norman et al., 2011). For instance, enriched GO terms were associated with biological processes related to intracellular signalling and communication, regulation and transport. Ion transmembrane transport is known to be important for osmoregulation (Lee et al., 2020) and migration timing (Hess et al., 2016) in salmonids, which supports this hypothesis. Additionally, cyclic guanosine monophosphate (cGMP) serves as a regulator for ion channels and acts as a secondary messenger for visual processing in the brain (Lane Brown et al., 2006). These associations suggest potential genes or gene families that could be the subject of functional and expression studies to test hypotheses concerning the mechanistic basis to any physiological distinctions among DV exhibiting different migration life histories. In addition, genotype–environmental association (GEA) analyses have been successfully conducted in salmonid fishes, including *Salvelinus*, as an alternative way to generate hypotheses about how inter-locality environmental differences may shape adaptive genomic variation (e.g. Micheletti et al., 2018; Salisbury et al., 2023). Such GEA analyses could be a fruitful way to test the idea that loci associated with excess NDV ancestry at specific locations in the Chignik Lake watershed are associated with particular environmental features, but the relatively small scale of the watershed may limit the power to detect such associations.

Northern DV and SDV are well-differentiated subspecies when examined in allopatry; they differ in average sizes, age at maturity, spawning mortality rates, gill raker and vertebral numbers and karyology (Armstrong & Morrow, 1980; Behnke, 1980;

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17459 by ERIC B TAYLOR , Wiley Online Library on [12/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

McPhail, 1961; Phillips et al., 1999; Sawatzky & Reist, 2021), yet there appears to be little evidence for reproductive isolation when they come into contact in the Chignik Lake watershed (and also in the nearby Karluk Lake watershed, Taylor & May-McNally, 2015). By contrast, we detected an excess of NDV alleles that may reflect positive selection for these alleles in Chignik watershed DV which is consistent with the evidence for isolation by adaptation in the same watershed (Bond, et al., 2014). Consequently, the system presents the intriguing possibility that hybridization between the subspecies following secondary contact has produced novel allelic combinations that may have promoted isolation by adaptation (e.g. Arnold, 1997; Bond, et al., 2014; Gallone et al., 2019; Montejo-Kovacevich et al., 2022; Selz & Seehausen, 2019). Further and when compared to sympatric DV and AC and sympatric DV and BT, the CDV system provides insight relevant to the evolution of reproductive isolation. Recent SNP-based phylogenies of *Salvelinus* suggest that DV and AC last shared a common ancestor about 1.2–1.7 mya (95% CI) and DV/AC and BT about 4.1–5.9 mya (Lecaudey et al., 2018). These estimates compare favourably to those of E.B. Taylor, A. Geraldes, & J. Shen (unpubl. data; 1.3-1.7 mya and 4.5–5.7 mya respectively) who further estimated the divergence between NDV and North American SDV to date to about 1.2-1.6 mya (and ~0.6 mya in our demographic analysis). In sympatry, AC/DV and DV/BT pairs of species show pronounced differences in levels of ecological and reproductive segregation between species (especially aspects of premating isolation) that while associated with bimodal hybrid zones in both species pairs, also show a higher incidence of F_1 hybrids in BT/DV systems (May-McNally et al., 2015; Redenbach & Taylor, 2003). These species-level distinctions stand in marked contrast to the apparent lack of reproductive barriers and relatively similar ecology exhibited between NDV and SDV. Although differences in mutation rates, demography and gene flow within and across taxa constrain firm conclusions (Stankowski & Ravinet, 2021), these observations suggest that even relatively long divergence times (>0.5 my) may be associated with no detectable reproductive isolation in some lineages of *Salvelinus* and that the ecological context is key to speciation in the group (Orr & Smith, 1998; Rieseberg & Willis, 2007; Schluter, 2000). Finally, the CDV, in addition to other similarly-distributed marine taxa, highlight the Alaska Peninsula as a unique and model area for studying genetic diversity and evolutionary trajectories resulting from post-glacial secondary contact.

AUTHOR CONTRIBUTIONS

AL, AG and EBT conceived and planned the study. AL collected the data and conducted most of the analyses with advice and assistance from AG and advice from EBT. AG and EBT conducted the demographic analyses. AG and EBT contributed data from an ongoing study (E.B. Taylor, A. Geraldes, & J. S et al., unpubl. data). AL wrote the initial manuscript which was commented on and revised by AG and EBT.

ACKNOWLEDGEMENTS

We thank Drs Morgan Bond and Tom Quinn for providing the Chignik Lake watershed samples through their work with the University of Washington's Alaska Salmon Program. Funding for this work was provided by Discovery and Equipment grants from the Natural Sciences and Engineering Research Council of Canada awarded to EBT. We appreciate the assistance of the Zoology Computing Unit at UBC, Dr. J-S. Légaré in conducting the demographic analyses and Drs Erin McFarlene and Jared Grummer for assistance with the BGC and functional analyses, respectively. Helpful comments by the subject editor and reviewer are appreciated.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts.

DATA AVAILABILITY STATEMENT

Raw sequence reads will be deposited in the National Centre for Biotechnology Information Short Read Achieve (NCBI SRA, BioProject PRJNA1116813). The admixed *Salvelinus* genome (Dolly Varden x Arctic char) sequencing and assembly are available on NCBI under BioProject PRJNA348349. Phenotypic data for life-history–genetic association analyses, PC and admixture scores are available in Table S1. The master vcf file and files for analyses in fastsimcoal2 are deposited in the UBC Dataverse repository (10.5683/SP3/CFMSB2).

BENEFIT SHARING STATEMENT

Benefits from this research result from sharing of the data and results on public databases. In addition, our work contributes to a better understanding of the distribution and relationship between two subspecies both in Canada and the United States and for which one of the subspecies is a species at risk in Canada (Northern Dolly Varden).

ORCID

Amy Liu https://orcid.org/0000-0002-7216-8414
Armando Geraldes https://orcid.org/0000-0001-7239-4529
Eric B. Taylor https://orcid.org/0000-0002-3974-6315

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*, 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Alcaide, M., Scordato, E. S. C., Price, T. D., & Irwin, D. E. (2014). Genomic divergence in a ring species complex. *Nature*, *511*(7507), 83–85. https://doi.org/10.1038/nature13285

Alex Sotola, V., Ruppel, D. S., Bonner, T. H., Nice, C. C., & Martin, N. H. (2019). Asymmetric introgression between fishes in the Red River basin of Texas is associated with variation in water quality. *Ecology and Evolution*, *9*, 2083–2095. https://doi.org/10.1002/ece3.4901

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17459 by ERIC B TAYLOR , Wiley Online Library on [12/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Armstrong, R. H., & Morrow, J. E. (1980). The Dolly Varden Charr, *Salvelinus malma*. In E. Balon (Ed.), *Charrs: Salmonid fishes of the genus Salvelinus* (pp. 99–140). Springer.

Arnold, M. L. (1997). *Natural hybridization and evolution*. Oxford University Press.

Arnold, M. L., & Martin, N. H. (2009). Adaptation by introgression. *Journal of Biology*, *8*, 1–3.

Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., Jacq, C., Jensen, A. J., Johnston, S. E., Karlsson, S., Kent, M., Moen, T., Niemelä, E., Nome, T., Næsje, T. F., Orell, P., Romakkaniemi, A., Sægrov, H., Urdal, K., … Primmer, C. R. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature*, *528*(7582), 405–408. https://doi.org/10.1038/nature16062

Behnke, R. J. (1980). A systematic review of the genus *Salvelinus*. In E. K. Balon (Ed.), *Charrs, salmonid fishes of the genus Salvelinus* (pp. 441–481). The Hague, W. Junk.

Bergeron, L. A., Besenbacher, S., Zheng, J., Li, P., Bertelsen, M. F., Quintard, B., Hoffman, J. I., Li, Z., St. Leger, J., Shao, C., & Stiller, J. (2023). Evolution of the germline mutation rate across vertebrates. *Nature*, *615*(7951), 285–291.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bond, H. (2013). *Diversity in migration, habitat use, and growth of Dolly Varden char in Chignik Lakes, Alaska*. PhD thesis, University of Washington, Seattle.

Bond, M. H., Crane, P. A., Larson, W. A., & Quinn, T. P. (2014). Is isolation by adaptation driving genetic divergence among proximate Dolly Varden char populations? *Ecology and Evolution*, *4*, 2515–2532.

Bourret, S. L., Kovach, R. P., Cline, T. J., Strait, J. T., & Muhlfeld, C. C. (2022). High dispersal rates in hybrids drive expansion of maladaptive hybridization. *Proceedings of the Royal Society B*, *289*(1986), 20221813.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. Springer.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 1–16. https://doi.org/10.1186/s13742-015-0047-8

Christensen, K. A., Rondeau, E. B., Minkley, D. R., Leong, J. S., Nugent, C. M., Danzmann, R. G., Ferguson, M. M., Stadnik, A., Devlin, R. H., Muzzerall, R., Edwards, M., Davidson, W. S., & Koop, B. F. (2018). The Arctic Charr (*Salvelinus alpinus*) genome and transcriptome assembly. *PLoS One*, *13*, e0204076. https://doi.org/10.1371/journal.pone.0204076

Christensen, K. A., Rondeau, E. B., Minkley, D. R., Leong, J. S., Nugent, C. M., Danzmann, R. G., Ferguson, M. M., Stadnik, A., Devlin, R. H., Muzzerall, R., Edwards, M., Davidson, W. S., & Koop, B. F. (2021). Retraction. The Arctic Charr (*Salvelinus alpinus*) genome and transcriptome assembly. *PLoS One*, *16*(2), e0247083. https://doi.org/10.1371/journal.pone.0247083

Conrad, R. H. C. N.-I. (1984). *Management applications of scale pattern analysis methods for the sockeye salmon runs to Chignik, Alaska*. Alaska Department of Fish and Game, Division of Commercial Fisheries. http://www.adfg.alaska.gov/FedAidPDFs/afrbil.233.pdf

Creelman, E. K., Hauser, L., Simmons, R. K., Templin, W. D., & Seeb, L. W. (2011). Temporal and geographic genetic divergence: Characterizing sockeye Salmon populations in the Chignik watershed, Alaska, using single-nucleotide polymorphisms. *Transactions of the American Fisheries Society*, *140*(3), 749–762. https://doi.org/10.1080/00028487.2011.584494

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Eden, E., Lipson, D., Yogev, S., & Yakhini, Z. (2007). Discovering motifs in ranked lists of DNA sequences. *PLoS Computational Biology*, *3*(3), 508–522. https://doi.org/10.1371/journal.pcbi.0030039

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*, 1–7. https://doi.org/10.1186/1471-2105-10-48

Ehlers, J., Gibbard, P. L., & Hughes, P. D. (2018). Quaternary glaciations and chronology. In J. Menzies & J. J. M. van der Meer (Eds.), *Past glacial environments* (2nd ed., pp. 77–101). Elsevier.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, *6*, e19379. https://doi.org/10.1371/JOURNAL.PONE.0019379

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, *9*(10), e1003905.

Excoffier, L., Marchi, N., & Sousa, V. C. (2023). *Fastsimcoal 2.8 manual*. University of Berne. http://cmpg.unibe.ch/software/fastsimcoal2/man/fastsimcoal28.pdf

Fitzpatrick, B. M. (2012). Estimating ancestry and heterozygosity of hybrids using molecular markers. *BMC Evolutionary Biology*, *12*(131), 1–14.

Forbes, S. H., & Allendorf, F. W. (1991). Associations between mitochondrial and nuclear genotypes in cutthroat trout hybrid swarms. *Evolution; International Journal of Organic Evolution*, *45*(6), 1332–1349. https://doi.org/10.1111/j.1558-5646.1991.tb02639.x

Gallone, B., Steensels, J., Mertens, S., Dzialo, M. C., Gordon, J. L., Wauters, R., Theßeling, F. A., Bellinazzo, F., Saels, V., Herrera-Malaver, B., Prahl, T., White, C., Hutzler, M., Meußdoerffer, F., Malcorps, P., Souffriau, B., Daenen, L., Baele, G., Maere, S., & Verstrepen, K. J. (2019). Interspecific hybridization facilitates niche adaptation in beer yeast. *Nature Ecology & Evolution*, *3*(11), 1562–1575. https://doi.org/10.1038/s41559-019-0997-9

Geraldes, A., Askelson, K. K., Nikelski, E., Doyle, F. I., Harrower, W. L., Winker, K., & Irwin, D. E. (2019). Population genomic analyses reveal a highly differentiated and endangered genetic cluster of northern goshawks (*Accipiter gentilis laingi*) in Haida Gwaii. *Evolutionary Applications*, *12*(4), 757–772. https://doi.org/10.1111/eva.12754

Gharrett, A. J., Goto, A., & Yamazaki, F. (1991). A note on the genetic contrast of sympatric Dolly Varden (*Salvelinus malma*) and Arctic charr (*S. alpinus*) in the Karluk River system, Alaska. In *Reproductive biology and population genetics of Dolly Varden (Salmonidae)*. Report of overseas work supported by Grant-in-aid for overseas scientific survey of the Ministry of Education, Science, and Culture of Japan.

Gompert, Z., & Buerkle, C. A. (2011). Bayesian estimation of genomic clines. *Molecular Ecology*, *20*(10), 2111–2127. https://doi.org/10.1111/j.1365-294X.2011.05074.x

Gompert, Z., & Buerkle, C. A. (2012). bgc: Software for Bayesian estimation of genomic clines. *Molecular Ecology Resources*, *12*(6), 1168–1176. https://doi.org/10.1111/1755-0998.12009.x

Gompert, Z., Mandeville, E. G., & Buerkle, C. A. (2017). Analysis of population genomic data from hybrid zones. *Annual Review of Ecology, Evolution, and Systematics*, *48*(1), 207–229. https://doi.org/10.1146/annurev-ecolsys-110316-022652

Gompert, Z., Parchman, T. L., & Buerkle, C. A. (2012). Genomics of isolation in hybrids. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *367*(1587), 439–450. https://doi.org/10.1098/rstb.2011.0196

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17459 by ERIC B TAYLOR , Wiley Online Library on [12/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Gowell, C. P., Quinn, T. P., & Taylor, E. B. (2012). Coexistence and origin of trophic ecotypes of pygmy whitefish, *Prosopium coulterii*, in a south-western Alaskan lake. *Journal of Evolutionary Biology*, *25*, 2432–2448. https://doi.org/10.1111/JEB.12011

Grummer, J. A., Whitlock, M. C., Schulte, P. M., & Taylor, E.B. (2021). Growth genes are implicated in the evolutionary divergence of sympatric piscivorous and insectivorous rainbow trout (*Oncorhynchus mykiss*). *BMC Ecology and Evolution*, *21*, 1-13. doi.org/10.1186/s12862-021-01795-9

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, *5*, e1000695. https://doi.org/10.1371/journal.pgen.1000695

Haines, M. L., Luikart, G., Amish, S. J., Smith, S., & Latch, E. K. (2019). Evidence for adaptive introgression of exons across a hybrid swarm in deer. *BMC Evolutionary Biology*, *19*(1), 199. https://doi.org/10.1186/s12862-019-1497-x

Hess, J. E., Zendt, J. S., Matala, A. R., & Narum, S. R. (2016). Genetic basis of adult migration timing in anadromous steelhead discovered through multivariate association testing. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1830), 20153064. https://doi.org/10.1098/rspb.2015.3064

Irwin, D. E., Alcaide, M., Delmore, K. E., Irwin, J. H., & Owens, G. L. (2016). Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. *Molecular Ecology*, *25*(18), 4488–4507. https://doi.org/10.1111/mec.13792

James, M. E., Arenas-Castro, H., Groh, J. S., Allen, S. L., Engelstädter, J., & Ortiz-Barrientos, D. (2021). Highly replicated evolution of parapatric ecotypes. *Molecular Biology and Evolution*, *38*(11), 4805–4821.

Janoušek, V., Munclinger, P., Wang, L., Teeter, K. C., & Tucker, P. K. (2015). Functional organization of the genome may shape the species boundary in the house mouse. *Molecular Biology and Evolution*, *32*(5), 1208–1220. https://doi.org/10.1093/molbev/msv011

Jiggins, C. D., & Mallet, J. (2000). Bimodal hybrid zones and speciation. *Trends in Ecology & Evolution*, *15*(6), 250–255. https://doi.org/10.1016/S0169-5347(00)01873-5

Kagawa, K., & Seehausen, O. (2020). The propagation of admixture-derived adaptive radiation potential. *Proceedings of the Royal Society B*, *287*(1934), 20200941. https://doi.org/10.1098/rspb.2020.0941

Kaufman, D. S., & Manley, W. F. (2004). Pleistocene maximum and late Wisconsinan glacier extents across Alaska, USA. *Developments in Quaternary Sciences*, *2*, 9–27.

Knappen, R. S. (1929). Geology and mineral resources of the Aniakchak district. *U.S. Geological Survey Bulletin*, *767*, 166–227.

Kowalchuk, M. W., Sawatzky, C. D., & Reist, J. D. (2010). *A review of the taxonomic structure within Dolly Varden, Salvelinus malma (Walbaum 1792), of North America*. Dept. Fisheries and Oceans Canada Science Advisory Secretariat Research Document. 2010/013. vi +, 3848, 16. http://www.dfo-mpo.gc.ca/csas/

Lane Brown, R., Strassmaier, T., Brady, J., & Karpen, J. (2006). The pharmacology of cyclic nucleotide-gated channels: Emerging from the darkness. *Current Pharmaceutical Design*, *12*(28), 3597–3613. https://doi.org/10.2174/138161206778522100

Lecaudey, L. A., Schliewen, U. K., Osinov, A. G., Taylor, E. B., Bernatchez, L., & Weiss, S. J. (2018). Inferring phylogenetic structure, hybridization and divergence times within Salmoninae (Teleostei: Salmonidae) using RAD-sequencing. *Molecular Phylogenetics and Evolution*, *124*, 82–99. https://doi.org/10.1016/j.ympev.2018.02.022

Lee, S. Y., Lee, H. J., & Kim, Y. K. (2020). Comparative transcriptome profiling of selected osmotic regulatory proteins in the gill during seawater acclimation of chum salmon (*Oncorhynchus keta*) fry. *Scientific Reports*, *10*(1), 1–14. https://doi.org/10.1038/s41598-020-58915-6

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, *19*(3), 639–647.

Liu, A., Geraldes, A., & Taylor, E. B. (2023). *Dolly Varden single nucleotide polymorphism and demographic analysis replication data*. Draft Version. https://doi.org/10.5683/SP3/CFMSB2

Lovell, S. F., Lein, M. R., & Rogers, S. M. (2021). Cryptic speciation in the warbling vireo (*Vireo gilvus*). *Ornithology*, *138*(1), uaa071. https://doi.org/10.1093/ornithology/ukaa071

Maglott, D., Ostell, J., Pruitt, K. D., & Tatusova, T. (2011). Entrez gene: Gene-centered information at NCBI. *Nucleic Acids Research*, *39*(Suppl. 1), 52–57. https://doi.org/10.1093/nar/gkq1237

Malek, T. B., Boughman, J. W., Dworkin, I., & Peichel, C. L. (2012). Admixture mapping of male nuptial colour and body shape in a recently formed hybrid population of threespine stickleback. *Molecular Ecology*, *21*(21), 5265–5279. https://doi.org/10.1111/J.1365-294X.2012.05660.X

Martin, B. T., Chafin, T. K., Douglas, M. R., & Douglas, M. E. (2021). ClineHelpR: An R package for genomic cline outlier detection and visualization. *BMC Bioinformatics*, *22*, 501. https://doi.org/10.1186/S.12859-021-04423-x

Matute, D. R., Comeault, A. A., Earley, E., Serrato-Capuchina, A., Peede, D., Monroy-Eklund, A., Huang, W., Jones, C. D., Mackay, T. F. C., & Coyne, J. A. (2020). Rapid and predictable evolution of admixed populations between two *Drosophila* species pairs. *Genetics*, *214*(1), 211–230. https://doi.org/10.1534/genetics.119.302685

May-McNally, S. L., Quinn, T. P., & Taylor, E. B. (2015). Low levels of hybridization between sympatric Arctic char (*Salvelinus alpinus*) and Dolly Varden char (*Salvelinus malma*) highlights their genetic distinctiveness and ecological segregation. *Ecology and Evolution*, *5*(15), 3031–3045. https://doi.org/10.1002/ece3.1583

McFarlane, S. E., & Pemberton, J. M. (2019). Detecting the true extent of introgression during anthropogenic hybridization. *Trends in Ecology & Evolution*, *34*(4), 315–326. https://doi.org/10.1016/j.tree.2018.12.013

McFarlane, S. E., Senn, H. V., Smith, S. L., & Pemberton, J. M. (2021). Locus-specific introgression in young hybrid swarms: Drift may dominate selection. *Molecular Ecology*, *30*(9), 2104–2115. https://doi.org/10.1111/mec.15862

McHugh, M. L. (2013). The chi-square test of independence. *Biochemia Medica*, *23*(2), 143–149. https://doi.org/10.11613/BM.2013.018

McKinney, G. J., Waples, R. K., Seeb, L. W., & Seeb, J. E. (2017). Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Molecular Ecology Resources*, *17*(4), 656–669. https://doi.org/10.1111/1755-0998.12613

McPhail, J. D. (1961). A systematic study of the *Salvelinus alpinus* complex in North America. *Journal of the Fisheries Research Board of Canada*, *18*, 793–816. https://doi.org/10.1139/f61-053

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, *8*, 1–11. https://doi.org/10.1038/ncomms14363

Meier, J. I., McGee, M. D., Marques, D. A., Mwaiko, S., Kishe, M., Wandera, S., Neumann, D., Mrosso, H., Chapman, L. J., Chapman, C. A., Kaufman, L., Taabu-Munyaho, A., Wagner, C. E., Bruggmann, R., Excoffier, L., & Seehausen, O. (2023). Cycles of fusion and fission enabled rapid parallel adaptive radiations in African cichlids. *Science*, *381*(6665), eade2833. https://doi.org/10.1126/science.ade2833

Menon, M., Bagley, J. C., Friedline, C. J., Whipple, A. V., Schoettle, A. W., Leal-Sàenz, A., Wehenkel, C., Molina-Freaner, F., Flores-Rentería, L., Gonzalez-Elizondo, M. S., Sniezko, R. A., Cushman, S. A., Waring,

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17459 by ERIC B TAYLOR , Wiley Online Library on [12/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

K. M., & Eckert, A. J. (2018). The role of hybridization during ecological divergence of southwestern white pine (*Pinus strobiformis*) and limber pine (*P. flexilis*). *Molecular Ecology*, *27*(5), 1245–1260. https://doi.org/10.1111/mec.14505

Micheletti, S. J., Matala, A. R., Matala, A. P., & Narum, S. R. (2018). Landscape features along migratory routes influence adaptive genomic variation in anadromous steelhead (*Oncorhynchus mykiss*). *Molecular Ecology*, *27*(1), 128–145.

Montejo-Kovacevich, G., Meier, J. I., Bacquet, C. N., Warren, I. A., Chan, Y. F., Kucka, M., Salazar, C., Rueda-m, N., Montgomery, S. H., McMillan, W. O., Kozak, K. M., Nadeau, N. J., Martin, S. H., & Jiggins, C. D. (2022). Repeated genetic adaptation to altitude in two tropical butterflies. *Nature Communications*, *13*(4676), 1–16. https://doi.org/10.1098/rspb.2019.1621

Moore, J. S., Bajno, R., Reist, J. D., & Taylor, E. B. (2015). Post-glacial recolonization of the north American Arctic by Arctic char (*Salvelinus alpinus*): Genetic evidence of multiple northern refugia and hybridization between glacial lineages. *Journal of Biogeography*, *42*(11), 2089–2100.

Moran, B. M., Payne, C., Langdon, Q., Powell, D. L., Brandvain, Y., & Schumer, M. (2021). The genomic consequences of hybridization. *eLife*, *10*, 1–33. https://doi.org/10.7554/ELIFE.69016

Muhlfeld, C. C., Kalinowski, S. T., McMahon, T. E., Taper, M. L., Painter, S., Leary, R. F., & Allendorf, F. W. (2009). Hybridization rapidly reduces fitness of a native trout in the wild. *Biology Letters*, *5*(3), 328–331. https://doi.org/10.1098/rsbl.2009.0033

Muniz, A. C., Pimenta, R. J. G., Cruz, M. V., Rodrigues, J. G., Buzatti, R. S. D. O., Heuertz, M., Lemos-Filho, J. P., & Lovato, M. B. (2022). Hybrid zone of a tree in a Cerrado/Atlantic Forest ecotone as a hotspot of genetic diversity and conservation. *Ecology and Evolution*, *12*(1), e8540.

Narver, D. W. (1966). *Pelagial ecology and carrying capacity of sockeye Salmon in the Chignik Lakes, Alaska*. PhD thesis, University of Washington, Seattle. https://doi.org/10.1016/j.jaci.2012.05.050

Norman, J. D., Danzmann, R. G., Glebe, B., & Ferguson, M. M. (2011). The genetic basis of salinity tolerance traits in Arctic charr (*Salvelinus alpinus*). *BMC Genetics*, *12*, 1–12. https://doi.org/10.1186/1471-2156-12-81

O'Corry-Crowe, G. M., Suydam, R. S., Rosenberg, A., Frost, K. J., & Dizon, A. E. (1997). Phylogeography, population structure and dispersal patterns of the beluga whale *Delphinapterus* leucas in the western Nearctic revealed by mitochondrial DNA. *Molecular Ecology*, *6*, 955–970. https://doi.org/10.1046/j.1365-294X.1997.00267.x

Orr, M. R., & Smith, T. B. (1998). Ecology and speciation. *Trends in Ecology & Evolution*, *13*(12), 502–506.

Oswald, J. A., Harvey, M. G., Remsen, R. C., Foxworth, D. P. U., Dittmann, D. L., Cardiff, S. W., & Brumfield, R. T. (2019). Evolutionary dynamics of hybridization and introgression following the recent colonization of glossy ibis (Aves: *Plegadis falcinellus*) into the New World. *Molecular Ecology*, *28*(7), 1675–1691. https://doi.org/10.1111/mec.15008

Owens, G. L., Baute, G. J., & Rieseberg, L. H. (2016). Revisiting a classic case of introgression: Hybridization and gene flow in Californian sunflowers. *Molecular Ecolology*, *25*, 2630–2643. https://doi.org/10.1111/mec.13569

Paun, O., Stuessy, T. F., & Hörandl, E. (2006). The role of hybridization, polyploidization and glaciation in the origin and evolution of the apomictic *Ranunculus cassubicus* complex. *New Phytologist*, *171*(1), 223–236.

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, *25*(11), 2337–2360. https://doi.org/10.1111/mec.13557

Pearse, D. E., Barson, N. J., Nome, T., Gao, G., Campbell, M. A., Abadía-Cardoso, A., Anderson, E. C., Rundio, D. E., Williams, T. H., Naish, K. A., Moen, T., Liu, S., Kent, M., Moser, M., Minkley, D. R., Rondeau, E. B., Brieuc, M. S. O., Sandve, S. R., Miller, M. R., … Lien, S. (2019). Sex-dependent dominance maintains migration supergene in rainbow trout. *Nature Ecology & Evolution*, *3*(12), 1731–1742. https://doi.org/10.1038/s41559-019-1044-6

Petrou, E. L., Hauser, L., Waples, R. S., Seeb, J. E., Templin, W. D., Gomez-Uchida, D., & Seeb, L. W. (2013). Secondary contact and changes in coastal habitat availability influence the nonequilibrium population structure of a salmonid (*Oncorhynchus keta*). *Molecular Ecology*, *22*, 5848–5860. https://doi.org/10.1111/mec.12543

Phillips, R. B., Gudex, L. I., Westrich, K. M., & DeCicco, A. (1999). Combined phylogenetic analysis of ribosomal ITS1 sequences and new chromosome data supports three subgroups of Dolly Varden char (*Salvelinus malma*). *Canadian Journal of Fisheries and Aquatic Sciences*, *56*(8), 1504–1511.

Primmer, C. R., Papakostas, S., Leder, E. H., Davis, M. J., & Ragan, M. A. (2013). Annotated genes and nonannotated genomes: Cross-species use of gene ontology in ecology and evolution research. *Molecular Ecology*, *22*(12), 3216–3241. https://doi.org/10.1111/mec.12309

Pulido-Santacruz, P., Aleixo, A., & Weir, J. T. (2018). Morphologically cryptic Amazonian bird species pairs exhibit strong postzygotic reproductive isolation. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1874), 20172081. https://doi.org/10.1098/rspb.2017.2081

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Redenbach, Z., & Taylor, E. B. (2003). Evidence for bimodal hybrid zones between two species of char (Pisces: *Salvelinus*) in northwestern North America. *Journal of Evolutionary Biology*, *16*(6), 1135–1148. https://doi.org/10.1046/j.1420-9101.2003.00619.x

Rieseberg, L. H., Van Fossen, C., & Desrochers, A. M. (1995). Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature*, *375*(6529), 313–316.

Rieseberg, L. H., & Willis, J. H. (2007). Plant speciation. *Science*, *317*(5840), 910–914.

Rugh, D. J., Shelden, K. E. W., & Hobbs, R. C. (2010). Range contraction in a beluga whale population. *Endangered Species Research*, *12*, 69–75. https://doi.org/10.3354/esr00293

Salisbury, S. J., Perry, R., Keefe, D., McCracken, G. R., Layton, K. K., Kess, T., Bradbury, I. R., & Ruzzante, D. E. (2023). Geography, environment, and colonization history interact with morph type to shape genomic variation in an Arctic fish. *Molecular Ecology*, *32*(12), 3025–3043.

Salmenkova, E. A., & Omelchenko, V. T. (2013). Genetic divergence and taxonomic status of chars of the genus *Salvelinus*. *Biology Bulletin Reviews*, *3*, 481–492.

Sawatzky, C. D., & Reist, J. D. (2021). Life history types and stages of northern form Dolly Varden, *Salvelinus malma malma* (Walbaum, 1792). *Canadian Manuscript Report of Fisheries and Aquatic Sciences*, *3215*, vii + 39.

Schield, D. R., Card, D. C., Adams, R. H., Jezkova, T., Reyes-Velasco, J., Proctor, F. N., Spencer, C. L., Herrmann, H. W., Mackessy, S. P., & Castoe, T. A. (2015). Incipient speciation with biased gene flow between two lineages of the Western diamondback rattlesnake (*Crotalus atrox*). *Molecular Phylogenetics and Evolution*, *83*, 213–223. https://doi.org/10.1016/j.ympev.2014.12.006

Schluter, D. (2000). *The ecology of adaptive radiation*. Oxford University Press.

Schumer, M., Rosenthal, G. G., & Andolfatto, P. (2018). What do we mean when we talk about hybrid speciation? *Heredity*, *120*, 379–382.

Seeb, L. W., & Crane, P. A. (1999). High genetic heterogeneity in chum salmon in western Alaska, the contact zone between northern and

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17459 by ERIC B TAYLOR , Wiley Online Library on [12/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

southern lineages. *Transactions of the American Fisheries Society*, *28*, 58–87. https://doi.org/10.1577/1548-8659(1999)128<0058:hghics>2.0.co;2

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G. P., Bank, C, Brännström, Å., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Foote, A. D., Franchini, P., Jiggins, C. D., … Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, *15*(3), 176–192. https://doi.org/10.1038/nrg3644

Selz, O. M., & Seehausen, O. (2019). Interspecific hybridization can generate functional novelty in cichlid fish. *Proceedings of the Royal Society B*, *286*(1913), 20191621. https://doi.org/10.1098/rspb.2019.1621

Shedko, S. V. (2019). *Assembly ASM291031v2 (Genbank: GCA_002910315.2) identified as assembly of the Northern Dolly Varden (Salvelinus malma malma) genome, and not the Arctic char (S. alpinus) genome*. 2, 1–15. http://arxiv.org/abs/1912.02474

Simmons, R. K., Quinn, T. P., Seeb, L. W., Schindler, D. E., & Hilborn, R. (2013). Role of estuarine rearing for sockeye salmon in Alaska (USA). *Marine Ecology Progress Series*, *481*, 211–223. https://doi.org/10.3354/MEPS10190

Stankowski, S., & Ravinet, M. (2021). Defining the speciation continuum. *Evolution*, *75*(6), 1256-1273. doi.org/10.1111/evo14215

Strait, J. T., Eby, L. A., Kovach, R. P., Muhlfeld, C. C., Boyer, M. C., Amish, S. J., Smith, S., Lowe, W. H., & Luikart, G. (2021). Hybridization alters growth and migratory life-history expression of native trout. *Evolutionary Applications*, *14*(3), 821–833. https://doi.org/10.1111/eva.13163

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*, *6*(7), e21800. https://doi.org/10.1371/JOURNAL.PONE.0021800

Szymura, J. M., & Barton, N. H. (1986). Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. *Evolution*, *40*(6), 1141–1159.

Taugbøl, A., Junge, C., Quinn, T. P., Herland, A., & Vøllestad, L. A. (2014). Genetic and morphometric divergence in threespine stickleback in the Chignik catchment, Alaska. *Ecology and Evolution*, *4*(2), 144–156. https://doi.org/10.1002/ece3.918

Taylor, E. B. (2016). The Arctic char (*Salvelinus alpinus*) "complex" in North America revisited. *Hydrobiologia*, *783*(1), 283–293. https://doi.org/10.1007/s10750-015-2613-6

Taylor, E. B., Boughman, J. W., Groenenboom, M., Sniatynski, M., Schluter, D., & Gow, J. L. (2006). Speciation in reverse: Morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology*, *15*(2), 343–355. https://doi.org/10.1111/j.1365-294X.2005.02794.x

Taylor, E. B., Lowery, E., Lilliestråle, A., Elz, A., & Quinn, T. P. (2008). Genetic analysis of sympatric char populations in western Alaska: Arctic char (*Salvelinus alpinus*) and Dolly Varden (*Salvelinus malma*) are not two sides of the same coin. *Journal of Evolutionary Biology*, *21*(6), 1609–1625. https://doi.org/10.1111/j.1420-9101.2008.01603.x

Taylor, E. B., & May-McNally, S. L. (2015). Genetic analysis of Dolly Varden (*Salvelinus malma*) across its North American range: Evidence for a contact zone in southcentral Alaska. *Canadian Journal of Fisheries and Aquatic Sciences*, *72*(7), 1048–1057. https://doi.org/10.1139/cjfas-2015-0003

Taylor, E. B., Redenbach, Z. A., Costello, A. B., Pollard, S. J., & Pacas, C. J. (2001). Nested analysis of genetic diversity in northwestern North American char, Dolly Varden (*Salvelinus malma*) and bull trout (*Salvelinus confluentus*). *Canadian Journal of Fisheries and Aquatic Sciences*, *58*(2), 406–420. https://doi.org/10.1139/cjfas-58-2-406

Tian, C., Gregersen, P. K., & Seldin, M. F. (2008). Accounting for ancestry: Population substructure and genome-wide association studies. *Human Molecular Genetics*, *17*(R2), 143–150. https://doi.org/10.1093/hmg/ddn268

Toews, D. P., Brelsford, A., Grossen, C., Milá, B. and Irwin, D. E. (2016). Genomic variation across the Yellow-rumped Warbler species complex. *The Auk*, *133*(4), 698-717. doi.org/10.1642/Auk-16-61.1

Torkamaneh, D., Laroche, J., & Belzile, F. (2016). Genome-wide SNP calling from genotyping by sequencing (GBS) data: A comparison of seven pipelines and two sequencing technologies. *PLoS One*, *11*(8), e0161333. https://doi.org/10.1371/JOURNAL.PONE.0161333

Van der Auwera, G. A., & O'Connor, B. D. (2020). *Genomics in the Cloud. Using Docker, GATK, and WDL in Terra* (1st ed.). O'Reilly Media.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*, 1358–1370.

Westley, P. A. H., Schindler, D. E., Quinn, T. P., Ruggerone, G. T., & Hilborn, R. (2010). Natural habitat change, commercial fishing, climate, and dispersal interact to restructure an Alaskan fish metacommunity. *Oecologia*, *163*(2), 471–484. https://doi.org/10.1007/s00442-009-1534-3

Yamamoto, S., Morita, K., Sahashi, G., Maekawa, K., Oleinik, A., Bondar, E., & Brykov, V. (2021). Introgressive hybridization between southern Asian Dolly Varden, *Salvelinus curilus*, and northern Dolly Varden, *S. malma malma*, on Sakhalin Island. *Russian Journal of Genetics*, *57*, 361–370. https://doi.org/10.1134/S.1022795421030145

Yates, M. C., Debes, P. V., Fraser, D. J., & Hutchings, J. A. (2015). The influence of hybridization with domesticated conspecifics on alternative reproductive phenotypes in male Atlantic salmon in multiple temperature regimes. *Canadian Journal of Fisheries and Aquatic Sciences*, *72*(8), 1138–1145. https://doi.org/10.1139/cjfas-2014-0527

Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biology*, *11*(2), R14.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Liu, A., Geraldes, A., & Taylor, E. B. (2024). Historical and contemporary processes driving the origin and structure of an admixed population within a contact zone between subspecies of a north temperate diadromous fish. *Molecular Ecology*, *00*, e17459. https://doi.org/10.1111/mec.17459

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.17459 by ERIC B TAYLOR , Wiley Online Library on [12/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License