Vol. 14: 329–342, 2022
https://doi.org/10.3354/aei00444

AQUACULTURE ENVIRONMENT INTERACTIONS
Aquacult Environ Interact

Published December 15

Genetic structure and origin of non-native, free-living Atlantic salmon *Salmo salar* along a latitudinal gradient in Chile, South America

Rodrigo Marín-Nahuelpi[1,2], José M. Yáñez[1,2], Selim S. Musleh[2,3,4], Diego Cañas-Rojas[2,3], Juan Carlos Quintanilla[5], Sergio Contreras-Lynch[5], Gonzalo Gajardo[6], Maritza Sepúlveda[2,7], Chris Harrod[2,8,9], Daniel Gomez-Uchida[2,3,*]

[1]Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, La Pintana, Santiago, Chile
[2]Núcleo Milenio INVASAL, Concepción, Chile
[3]Genomics in Ecology, Evolution & Conservation Laboratory (GEECLAB), Departamento de Zoología, Facultad Ciencias Naturales y Oceanográficas, Universidad de Concepción, Concepción, Chile
[4]Departamento de Evaluación de Recursos, Instituto de Fomento Pesquero (IFOP), Valparaíso, Chile
[5]Departamento de Salud Hidrobiológica, División de Investigación en Acuicultura, Instituto de Fomento Pesquero (IFOP), Puerto Montt, Chile
[6]Laboratorio de Genética, Acuicultura & Biodiversidad, Departamento de Ciencias Biológicas y Biodiversidad, Universidad de Los Lagos, Osorno, Chile
[7]Centro de investigación y Gestión de Recursos Naturales (CIGREN), Facultad de Ciencias, Universidad de Valparaíso, Valparaíso, Chile
[8]Instituto de Ciencias Naturales Alexander von Humboldt, Universidad de Antofagasta, Antofagasta, Chile
[9]Universidad de Antofagasta Stable Isotope Facility, Instituto Antofagasta, Universidad de Antofagasta, Antofagasta, Chile

ABSTRACT: Limited stocking efforts to introduce Atlantic salmon *Salmo salar* into Chilean rivers and streams were unsuccessful during the 20th century. Following the arrival of the aquaculture industry during the 1980s, escaped Atlantic salmon have presented an ecological risk to native taxa through predation, competition, and transmission of pathogens or parasites. However, whether commercial aquaculture strains represent the likely source of free-living Atlantic salmon in marine and freshwater environments is unclear. We used 272 single nucleotide polymorphisms to characterize free-living Atlantic salmon (n = 80) captured from 12 marine and freshwater locations in southern Chile. These were compared with 8 reference collections, 6 known commercial strains, and 2 wild populations of Atlantic salmon. We evaluated genetic structure among free-living Atlantic salmon and assessed individual ancestry and origin by assigning mixture samples to reference collections. We found evidence for genetic structure (number of clusters, $K = 3$) among free-living salmon unexplained by geography, environment, or life stage, but consistent with the number of clusters among commercial aquaculture strains. Most free-living Atlantic salmon had a close ancestry with farmed Norwegian strains, the most widely used by the industry, pointing to recent aquaculture escapes as their origin. Yet recent establishment of self-sustaining populations weakly differentiated from aquaculture broodstock cannot be ruled out. We propose increasing monitoring efforts of free-living Atlantic salmon in remote sites as well as in watersheds located in densely stocked aquaculture areas.

KEY WORDS: Invasive species · Genetic stock identification · Single nucleotide polymorphisms · SNPs · Salmonids

*Corresponding author: dgomezu@udec.cl

Publisher: Inter-Research · www.int-res.com

1. INTRODUCTION

Net-pen aquaculture of Atlantic salmon *Salmo salar* has become a global industry amid many environmental issues that arise from their production. Atlantic salmon are only native to North Atlantic Ocean basins and have been introduced to several continents outside their natural distribution to support global aquaculture. Chile is currently the second-largest producer of Atlantic salmon globally (FAO 2020). Since the late 1990s, Chile's aquaculture industry growth has mainly relied on Atlantic salmon production and market value, which currently exceeds 70% of total salmonid production (Lhorente et al. 2019) and reached an annual harvest of 700 000 t in 2019 (SERNAPESCA 2019). However, high stocking densities and a growing number of facilities have increased the likelihood of escapes — either chronic small leakages or episodic large escapes — which represents one of the principal environmental risks associated with Chile's salmon aquaculture industry (Quiñones et al. 2019). Salmon escapes may occur following damages to net pens from multiple causes: attacks from marine mammals, collisions with watercraft, theft or vandalism, negligence during fish handling or maintenance, and severe weather conditions (Jensen et al. 2010, Vilata et al. 2010, Sepúlveda et al. 2013).

Escaped salmon can displace native fishes due to predatory and interference competition, severely impacting aquatic biodiversity and species richness (Garcia de Leaniz et al. 2010, Sepúlveda et al. 2013). Atlantic salmon may prey on forage fish such as southern sprat *Sardinops sagax* and Chilean silverside *Odontesthes regia*, which play key roles in marine food webs and fisheries (Soto et al. 1997, Niklitschek & Toledo 2011). They also can be vectors of pathogens, parasites, and chemicals (Garcia de Leaniz et al. 2010, Jensen et al. 2010, Sepúlveda et al. 2013). Epidemiological studies conducted in the Northern Hemisphere suggest that the occurrence of diseases (e.g. rickettsial septicemia and sea lice) in salmonids, but also in native fish, could be directly related to the presence of farmed fish in high concentrations in Norway (Naylor et al. 2005, Krkošek 2010).

Pacific salmon species — Chinook salmon (Correa & Gross 2008), rainbow trout (Pascual et al. 2001, Soto et al. 2006, Arismendi et al. 2011), and more recently, coho salmon (Chalde et al. 2019, Maldonado-Márquez et al. 2020) — have established self-sustaining populations in South America following heavy stocking efforts. In comparison, efforts to stock Atlantic salmon in rivers and lakes were few and far between during the 20th century (Table 1) (Basulto 2003, Arismendi et al. 2014), but several million Atlantic salmon have escaped in freshwater systems and at sea (Table 1) (Soto et al. 2001, 2006, Sepúlveda et al. 2013, Gomez-Uchida et al. 2018). To date, there is little evidence for successful natural reproduction of Atlantic salmon or self-sustaining populations, especially anadromous populations (Nikli-

Table 1. Introduction events of Atlantic salmon, including stocking in freshwater and records of escapes from marine net pens in Chile. NA: not available

Events	Year	No. of individuals	Location	Origin and reference
Stocking in freshwater	1905	100 000 eggs	Río Blanco Hatchery	Germany (Basulto 2003, Arismendi et al. 2014)
	1907	20 000 juveniles	Several streams from Santiago to Valdivia	Europe (Basulto 2003)
	1910	80 000 eggs	Puerto Varas Hatchery	Europe (Basulto 2003, Arismendii et al. 2014)
	1916	360 000 eggs	Lautaro Hatchery	Europe (Basulto 2003)
	1927–1928	400 000 eggs	Southern Patagonia and Tierra del Fuego	Argentina via Maine, USA (Basulto 2003)
	1941	NA	San Rafael Lake	Unknown (Basulto 2003)
Escaped fish from marine net pens	1993–1997	1 497 000 adults	Los Lagos, Aysén and Magallanes regions	Sepúlveda et al. (2013)
	2004–2008	1 865 100 adults	Los Lagos, Aysén and Magallanes regions	Sepúlveda et al. (2013)
	2009–2012	398 000 adults	Los Lagos, Aysén and Magallanes regions	Sepúlveda et al. (2013)
	2010–2020	4 947 464 adults	Los Lagos, Aysén and Magallanes regions	SERNAPESCA (2021)

tschek et al. 2013, Arismendi et al. 2014, Monzón-Argüello et al. 2014, Quiñones et al. 2019). However, this apparent lack of success should not be assumed as permanent due to high propagule pressure provided by escapes from marine net pens and the high density and wide distribution of aquaculture activities (Pascual & Ciancio 2007).

Given the potential impacts of escapee fish on receiving ecosystems and wild fish populations, researchers in the Northern Hemisphere have used forensic approaches to identify the origin of putative wild-living salmon (Glover 2010). In order to discriminate between farmed, wild, and naturalized salmon, a range of different methodologies have been used, including carotenoid pigment differentiation (Poole et al. 2000), comparison of growth rates (Stokesbury et al. 2001), scale microchemistry (Adey et al. 2009), and stable isotope analysis (Schröder & Garcia de Leaniz 2011, Wang et al. 2018). Another approach is to use molecular tools, given the limited genetic variation associated with aquaculture salmon in their non-native distribution. Genetic-based identification using single nucleotide polymorphisms (SNPs) has recently been used as part of a monitoring program for free-living rainbow trout in Chile, which are distributed from the tropical border with Peru to sub-Antarctic waters of Tierra del Fuego (Benavente et al. 2015, Canales-Aguirre et al. 2018). Such tools are readily available from genetic studies of native and farmed populations of Atlantic salmon. Yáñez et al. (2016) developed a 159K SNP ascertainment panel that included wild populations and farmed strains from different origins to be applied for genetic studies that require moderate- to high-resolution genomic information.

Despite the availability of genomic resources for Atlantic salmon and an increased number of free-living individuals, to our knowledge there are no studies aimed at inferring the origin of free-living Atlantic salmon captured in Chile. We hypothesized that genetic structure and ancestry between free-living and aquaculture Atlantic salmon should be similar if free-living salmon originated from recent escapes. Here, we inferred genetic structure among free-living Atlantic salmon and compared these samples to 8 reference collections — 6 aquaculture strains used in Chile and 2 wild populations — using genotypes obtained from 500 highly informative SNP markers. The main objectives of the current study were to (1) evaluate genetic structure of free-living Atlantic salmon from historical collections, in both freshwater and marine locations (~2000 km coverage) using individual-based inference; and (2) assess individual ancestry among free-living salmon by assignment to key reference collections represented by farmed strains used in Chile's Atlantic salmon aquaculture. First, we discuss the implications of assignment results, predominantly to commercial aquaculture strains, especially from Norway. We then consider the limitations of our data set to further evaluate the invasion success of Atlantic salmon and offer insights into where monitoring of this species should be strengthened.

2. MATERIALS AND METHODS

2.1. Sampling, genotyping, and quality control

Tissue samples from free-living Atlantic salmon adults and juveniles (n = 95) captured via gillnetting or electrofishing, respectively, were collected during 2007–2017 from marine and freshwater surveys spanning nearly 2000 km of geographic coverage (Table 2, Fig. 1). Muscle or fin clip tissues were stored in 95% ethanol until laboratory analyses. Genomic DNA from fin clips was extracted using a commercial kit NucleoSpin® Tissue (Qiagen), following the manufacturer's instructions. DNA quantification was performed using a NanoDrop 2000 spectrophotometer. All samples were screened through a panel of 500 SNPs, a subset of the 159K SNP array developed by Yáñez et al. (2016). SNPs were selected aiming at an even distribution of markers per chromosome, a minor allele frequency (MAF) > 0.01, and a unique position in the reference genome (GenBank accession no. GCA_000233375.4) (Yoshida et al. 2018).

Genotyping was outsourced to LGC group laboratories (www.lgcgroup.com) using KASP technology (He et al. 2014). Genotype quality control (QC) was assessed using PLINK (Purcell et al. 2007) in order to remove SNPs with a genotype call rate lower than 0.90, MAF < 0.01, and departure from Hardy-Weinberg equilibrium (HWE) ($p < 1 \times 10^{-6}$). MAF and HWE QC filters were performed for each population separately. We also removed individuals with a sample call rate < 80% from further analyses.

2.2. Reference collections

We compared free-living genotypes to 8 reference collections of Atlantic salmon (n = 456) previously genotyped for the 159K SNP panel containing the 500 SNP subset (Yáñez et al. 2016, Yoshida et al.

Table 2. Sampling sites (north to south) and description of Atlantic salmon samples collected. n = number of samples per site; NA: not available

Site number	Site description	Collection year	Environ -ment	Latitude (° S)	Longitude (° W)	n	Life stage	Size range (mm)	Reproductive status
1	Lago Colico	2017	Freshwater	39.05	72.02	1	Adult	590	Immature
2	Lago Villarrica	2017	Freshwater	39.21	72.15	1	Adult	570	Immature
3	Río Quimán – Lago Ranco	2009	Freshwater	40.11	72.34	12	Parr–smolt	130–180	Immature
4	Río Pichichanlelfu – Lago Puyehue	2007	Freshwater	40.42	72.19	10	Parr–smolt	116–150	Immature
5	Río Pescadero – Lago Puyehue	2008	Freshwater	40.72	72.40	12	Parr–smolt	115–145	Immature
6	Lago Rupanco	2017	Freshwater	40.77	72.67	1	Adult	395	Immature
7	Lago Rupanco	2007	Freshwater	40.85	72.50	15	Parr–smolt	125–160	Immature
8	Puerto Fonck – Lago Llanquihue	2008	Freshwater	41.00	72.70	11	Parr–smolt	NA	NA
9	Lago Llanquihue	2017	Freshwater	41.25	73.00	2	Adult	430–440	Immature
10	Río Blanco – Correntoso	2009	Freshwater	41.39	72.64	13	Parr–smolt	81–156	Immature
11	Macrozona 1	2017	Marine	41.51	72.78	1	Smolt	200	Immature
12	Río Pichicolo – Hornopirén	2009	Freshwater	41.98	72.56	1	Smolt	195	Immature
13	Río Negro – Hualaihue	2009	Freshwater	42.00	72.68	1	Smolt	157	Immature
14	Chiloé Sur	2017	Marine	43.20	73.58	2	Adult	310–330	Immature
15	Macrozona 8	2017	Marine	44.94	73.10	3	Adult	390–530	Immature
16	Fiordo Aysén	2017	Marine	45.40	72.82	2	Adult	560–820	Mature
17	Macrozona 7	2017	Marine	45.60	73.25	1	Adult	560	Mature
18	Río Cochrane	2017	Freshwater	47.24	72.53	1	Adult	540	Mature
19	Puerto Natales	2017	Marine	52.00	72.96	5	Adult	325–605	Mature

2018). These collections originated from North America and Europe and included 6 commercial aquaculture strains used in Chile (F1–F6) and 2 wild populations (W1, W2); genetic metrics of these collections can be found elsewhere (Yáñez et al. 2016). These reference collections covered a broad geographic range and genetic divergence of the extant variation present in aquaculture and wild populations: F1, F2, and F3 = farmed Norway; F4 = farmed Ireland; F5 = farmed Canada; F6 = farmed Scotland; W1 = wild Scotland; and W2 = wild Canada (Table 2; see also Bourret et al. 2013, López et al. 2019). Inclusion of reference individuals originating from wild Atlantic salmon populations from Europe and North America provided outgroups to delimit maximum levels of genetic differentiation among free-living individuals captured in Chile.

2.3. Genetic structure analyses

We assessed genetic structure using a Bayesian clustering approach from STRUCTURE v.2.3.4 (Pritchard et al. 2000) implemented in the 'ParallelStructure' v.1.0 (Besnier & Glover 2013) package in R v.4.1.0 (R Core Team 2020). STRUCTURE assigns individuals into K genetic clusters using Markov chain Monte Carlo (MCMC) sampling algorithms considering a random-cross model, maximizing HWE and minimizing linkage disequilibrium. We performed clustering analyses in 2 steps: (1) we analyzed only free-living individuals; (2) we analyzed free-living and reference collections combined. Ten independent runs were performed for each K between 1 and 8, with a burn-in period of 50 000 MCMC iterations and keeping 250 000 iterations of the MCMC algorithm. To infer the most likely K value, we used Evanno et al.'s (2005) method via STRUCTURE HARVESTER v.0.6.94 (Earl & vonHoldt 2012). We provided a consensus run from multiple independent simulations using CLUMPAK v.1.1 software (Kopelman et al. 2015). We then used the 'LargeKGreedy' algorithm (Earl & vonHoldt 2012) to return an average run for each cluster and a custom R script for a graphical display of the results.

We additionally performed a discriminant analysis of principal components (DAPC) on both free-living and combined (free-living plus reference collections) genotype data sets using the R package 'adegenet' v.2.1.4 (Jombart et al. 2010, Jombart & Ahmed 2011). One of the advantages of DAPC is that it provides probabilities for each individual's membership in the different groups based on the retained discriminant functions, free of assumptions such as HWE, which may be common among recently founded, non-native salmonid populations (Benavente et al. 2015, Canales-Aguirre et al. 2018, Musleh et al. 2020). Synthetic variables were constructed first by transforming genotype data using principal component analysis and second by performing a discriminant analysis on the retained principal components. We defined the number of clusters using the 'find.clusters'

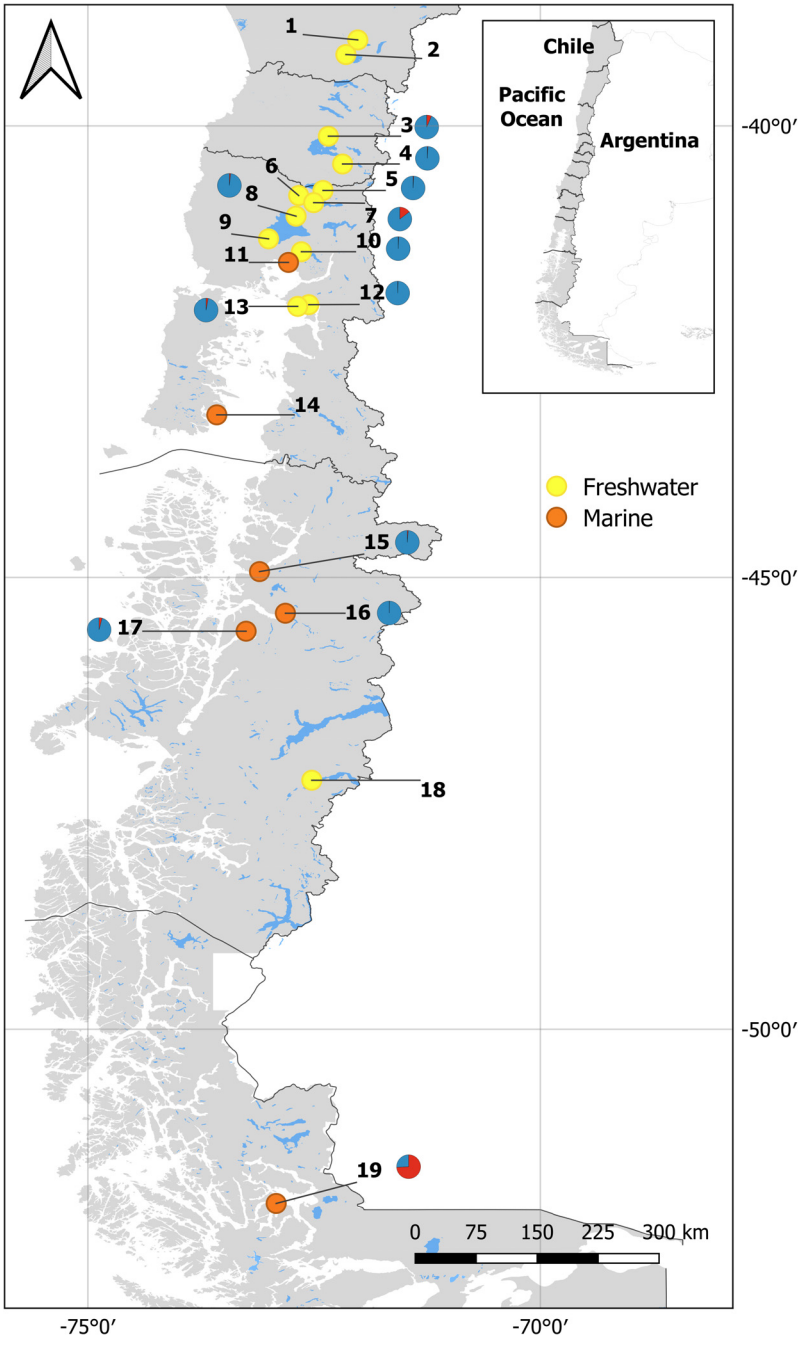

Fig. 1. Collection sites of free-living Atlantic salmon in freshwater and marine locations. Pie charts from sites that passed quality control show genetic membership to 2 clusters: European (blue) or North American (red) ancestry

function, which uses a *k*-means clustering algorithm that maximizes genetic differences between groups and minimizes genetic differences within groups over 10^6 iterations. The optimum number of retained principal components was inferred by a cross-validation method using the 'xvalDapc' function. Cluster modeling was performed among reference collections. Subsequently, we used the 'predict.DAPC' function to predict which cluster free-living individuals were most closely associated with. To find clusters in the free-living data set, 80 principal components were retained and 3 clusters (K = 3) were chosen following the aforementioned criteria. To find clusters in the combined data set, 362 principal components were retained, and 2 clusters (K = 2) were chosen based on the K position relative to an elbow pattern displayed in a Bayesian information criterion curve (Fig. S1 in Supplement 1 at www.int-res.com/articles/suppl/q014p329_supp1.pdf). Finally, a total of 50 principal components were used in the DAPC analysis based on the results of cross-validation (Fig. S2).

2.4. Mixed-stock analysis and individual assignment of free-living individuals to reference collections

We performed mixture analyses and individual assignment of free-living individuals to reference collections using the R package RUBIAS v.0.3.3 (Moran & Anderson 2019) to identify reference collections as likely sources of free-living Atlantic salmon. RUBIAS uses hierarchical Bayesian inference taking into account population structure and resolves possible biases associated with building reference groups (Anderson et al. 2008, Hasselman et al. 2016).

First, we carried out simulations using the function 'assess_reference_loo' to evaluate self-assignment or how much accuracy could be achieved from the data set of reference collections using a leave-one-out approach. We used 6 reference collections (see Section 2.2) and 2 reference groups ('Europe' and 'North America') following results from genetic clusters inferred using STRUCTURE or DAPC (see above). We then simulated 50 mixture samples of size n = 200 considering the known simulated proportions of each reference group.

Second, we assessed free-living Atlantic salmon origin by analyzing both mixture proportions and individual posterior membership probabilities to reference collections and reference groups. We used the function 'infer_mixture' with 200 000 MCMC iterations and a 40 000 burn-in period. RUBIAS additionally provided *Z*-score tests contrasting maximum *a posteriori* probability assignments to random normal to ascertain whether any fish from the mixture sample of free-living Atlantic salmon had an origin in potentially unsampled reference groups.

3. RESULTS

3.1. Genotyping quality

Descriptive results of population genetic estimates for the different groups are shown in Table 3. All individuals from the reference populations passed QC. A total of 80 free-living Atlantic salmon genotypes passed sample call-rate QC; individuals from Sites 1, 2, 6, 9, 11, 14, and 18 were removed from further analyses. From a total of 500 SNPs, 484 amplified consistently and were available for QC; of these, we kept 272 SNPs that were shared among all collections (Table 3).

3.2. Genetic clustering analyses

We found significant genetic structure among free-living Atlantic salmon (STRUCTURE K = 3: Figs. 2, S3 & S4). Geography, environment, and life stage failed to explain patterns of genetic differentiation. Salmon from Sites 3, 5, 10, 12, and 13 (all juvenile parr–smolt from freshwater) showed membership to the same cluster, while Sites 15, 16, and 17 (adults at sea) showed similar membership to Sites 4–7 and 10 (juveniles and adults). Site 19 (adults at sea) diverged significantly from the others and formed a third cluster. DAPC performed on free-living salmon was also consistent with 3 clusters (Fig. 3), with a proportion of conserved variance of 0.86.

Genetic structure among free-living salmon was consistent with genetic structure found among farmed broodstock after adding reference collections of Atlantic salmon for a combined data set (Fig. 4), with Evanno et al.'s (2005) method suggesting K = 2 (Fig. S5). Most free-living individuals (72 of the 80; Q-value > 0.85) showed ancestry similar to one cluster of Norway strains and one Scotland wild population (F1–F4, F6, and W1), while 8 of 80 (Q-value > 0.4) shared ancestry between North American farmed and wild salmon. Of these 8, 3 were found in Site 19. When we additionally explored K = 3 in STRUCTURE, we found that F4 (= farmed Ireland) significantly diverged from other farmed Norway and North American farmed stocks, while the wild population from Scotland (W1) appeared admixed (Fig. 5A). Free-living salmon genotypes continued to show predominantly farmed Norway ancestry (Fig. 5B).

DAPC performed on the combined data set revealed 2 clear clusters defined by the first discriminant functions: one European and one North American (Fig. 6A). The second discriminant function re-

Table 3. Genetic summary statistics of quality control (QC) for reference and free-living collections of Atlantic salmon. n_{QC}: number of samples that passed QC. MAF > 0.01: single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF) > 0.01. HWE: SNPs that fit Hardy-Weinberg Equilibrium (HWE) proportions (Bonferroni correction was applied). Heterozygosity: estimates using 272 SNPs shared among all individuals that passed QC (H_o, H_e: observed, expected heterozygosity, respectively). ARUL: average reporting unit scaled likelihood—posterior probability of assigning the fish to the inferred collection given an equal prior on every collection in the reference

	Population	Origin	n_{QC}	MAF > 0.01		HWE		Heterozygosity		Self-assignment
				n	%[a]	n	%	H_o	H_e	ARUL
Reference	F1	Norway	86	484	96.8	484	96.8	0.48	0.47	0.901
	F2	Norway	91	484	96.8	484	96.8	0.51	0.49	0.957
	F3	Norway	74	484	96.8	435	87	0.54	0.45	0.88
	F4	Ireland	43	484	96.8	484	96.8	0.38	0.37	1
	F5	Canada	37	273	54.6	272	54.4	0.32	0.29	0.99
	F6	Scotland	40	484	96.8	484	96.8	0.45	0.45	0.99
	W1	Scotland	41	484	96.8	484	96.8	0.45	0.44	0.99
	W2	Canada	44	418	83.6	418	83.6	0.27	0.26	1
Free-living	FL	Chile	80	483	96.6	468	93.6	0.39	0.46	–

[a]Percentage of SNPs out of 500 initially genotyped

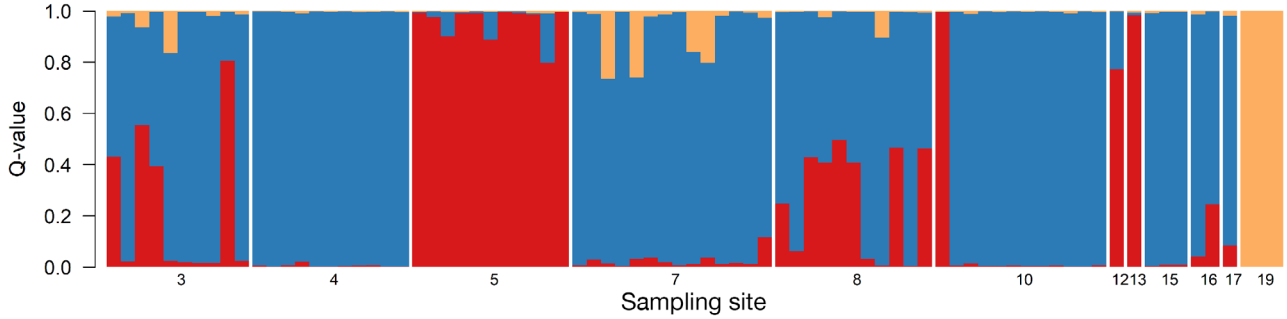

Fig. 2. Genetic structure of free-living Atlantic salmon using STRUCTURE $K = 3$. Estimated genetic structure of 80 individuals distributed by sampling site, revealed by 272 single nucleotide polymorphism markers for $K = 3$. The most prominent cluster is represented in blue, followed by red and light brown. Each individual is represented by a single vertical line with the estimated individual proportions of cluster membership

vealed 3 clusters distinguishing Atlantic salmon European strains: farmed Norway (F1–F3), farmed Ireland (F4), and farmed Scotland (F6; Fig. 6B). Farmed Norway stocks were indistinguishable from one another within the second discriminant function. Among free-living Atlantic salmon, 72 of 80 had close European ancestry, while 8 had hybrid European–North American ancestry. These were the same individuals identified as admixed by STRUCTURE in Site 19. The proportion of conserved variance was 0.592.

3.3. Mixed-stock analysis and individual assignment of free-living individuals to reference collections

Self-assignment of individuals following simulations revealed that genotypes originating from reference collections were accurately assigned to their reference group between 88% (F3) and 100% (F4, W2) of the time (Table S1 in Supplement 2 at www.int-res.com/articles/suppl/q014p329_supp2.xlsx). Also, simulations indicated that mixture proportions were highly accurate for all reference groups when 6 or 2 reporting units were used (Fig. S6). RUBIAS mean estimates of mixed proportions among free-living individuals was over 99% credible interval (CI) (95% CI = 0.979–1.00) of 'European' origin assuming 2 reference groups and 97% (95% CI = 0.917–0.997) 'Norway' origin assuming 6 reference groups (Table 4). Posterior probabilities of assignment indicated that most free-living Atlantic salmon originated from 'European' reference group sources; notice that assignments varied by reference collection, with farmed strains F1, F2, and F3 representing the most probable sources (Table S1). One individual from sampling Site 5 was assigned to W1 (= wild Scotland). Some *Z*-scores from RUBIAS suggested an incongruence between maximum *a posteriori* proba-

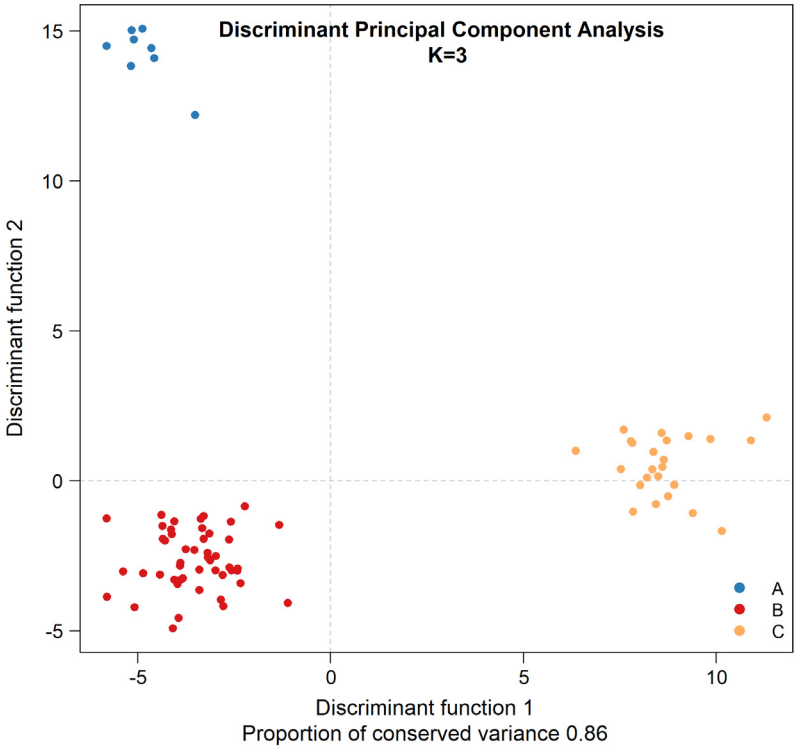

Fig. 3. Discriminant analysis of principal components of free-living Atlantic salmon genotypes. The first discriminant function distinguished between A and B from C clusters, whereas the second discriminant function distinguished A from B and C clusters. Individuals in cluster A correspond to admixed individuals, while individuals in clusters B and C correspond to animals of predominantly European ancestry

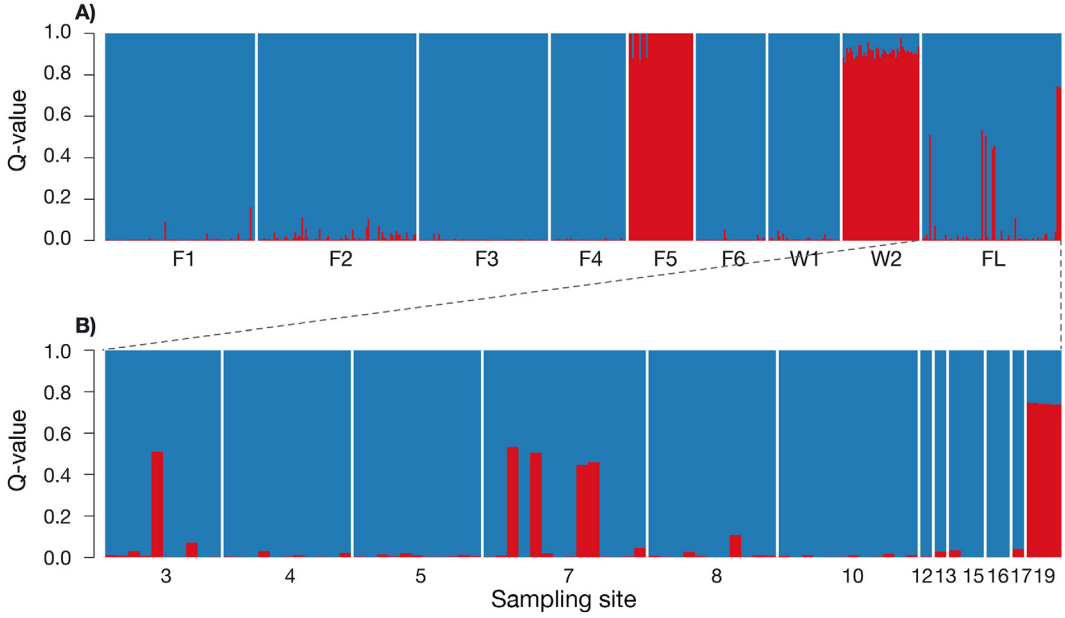

Fig. 4. Genetic structure of reference collections and free-living Atlantic salmon using STRUCTURE K = 2. (A) Q-values among reference collections (F1, F2, and F3: farmed Norway; F4: farmed Ireland; F5: farmed Canada; F6: farmed Scotland; W1: wild Scotland; W2: wild Canada) and free-living Atlantic salmon (FL); (B) Q-values from each sampling site of free-living Atlantic salmon. Red and blue colors represent ancestry to North American and European clusters, respectively

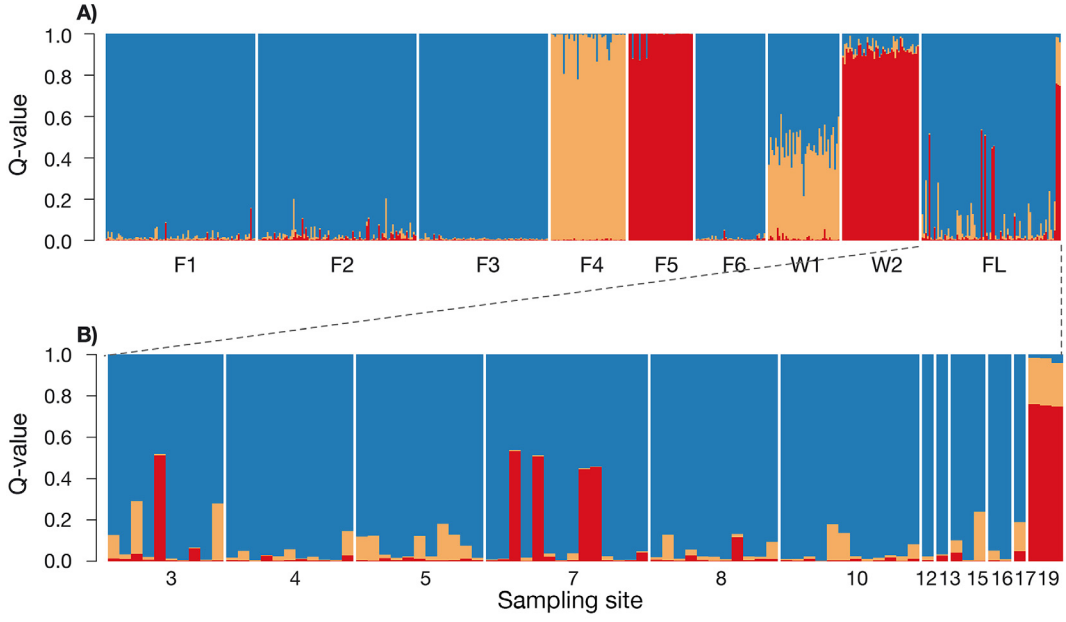

Fig. 5. Genetic structure of reference collections and free-living Atlantic salmon using STRUCTURE K = 3. (A) Q-values among reference collections (F1, F2, and F3: farmed Norway; F4: farmed Ireland; F5: farmed Canada; F6: farmed Scotland; W1: wild Scotland; W2: wild Canada) and free-living Atlantic salmon (FL); (B) Q-values from each sampling site of free-living Atlantic salmon. Red and blue colors represent North American and European clusters, respectively. Yellow represents an Ireland cluster, which shows a common ancestry with the wild Scotland population

bilities and random normal expectations (Kolmogorov-Smirnov test, p = 0.006), suggesting a potential violation of assumptions of mixed-stock analyses that are discussed below.

4. DISCUSSION

This study originated out of a marked need to evaluate genetic structure of free-living Atlantic

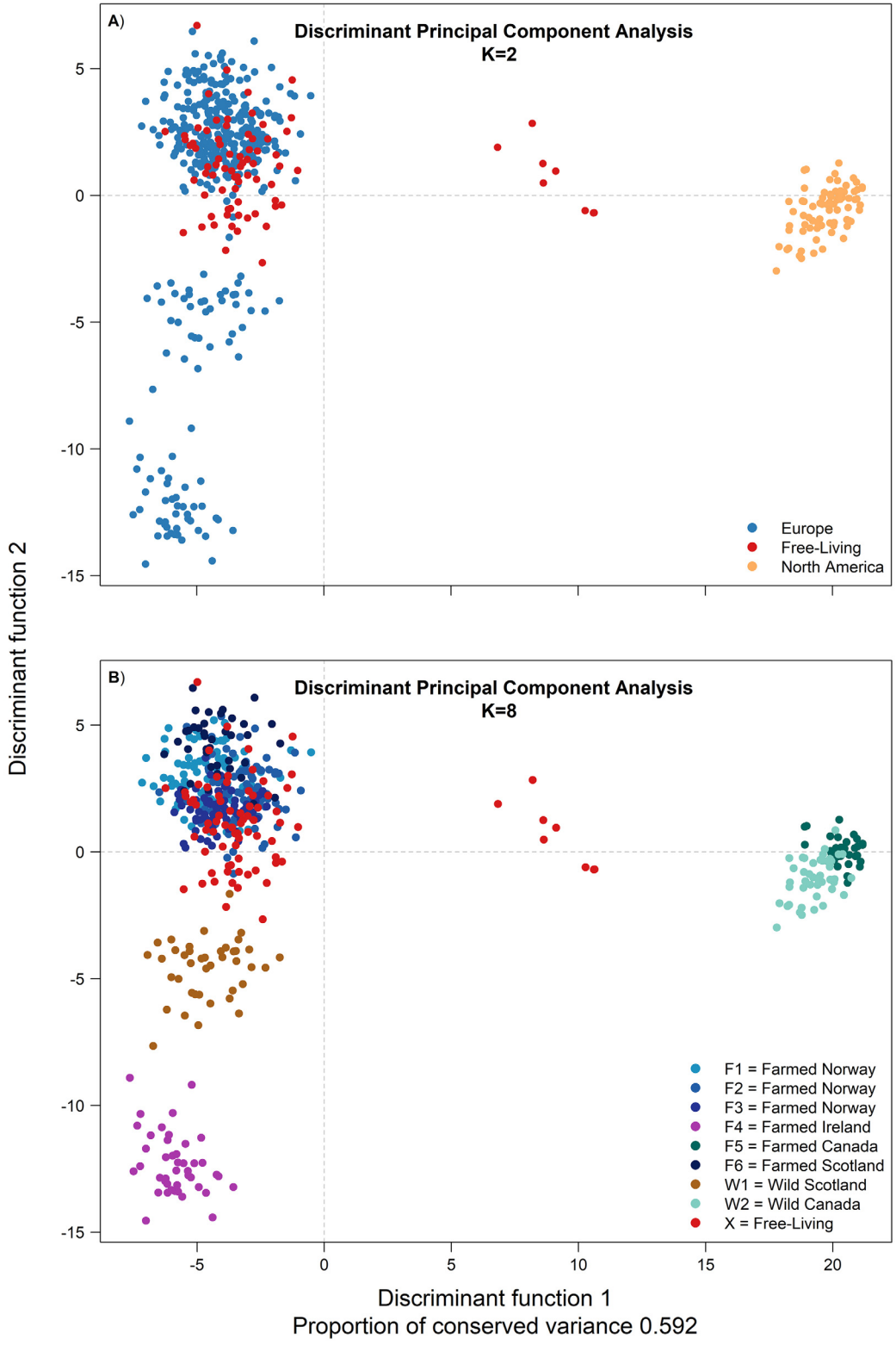

Fig. 6. Discriminant analysis of principal components of free-living and reference collections of Atlantic salmon. (A) First discriminant function distinguishes between European and North American populations; 73 free-living individuals were positioned within the European cluster, whereas 7 were positioned at an intermediate distance from the North American cluster. (B) Second discriminant function distinguishes between farmed Norway, farmed Ireland, and wild Scotland collections

Table 4. Estimates of mixing proportion $p(m)$ among free-living Atlantic salmon with their 95% credible intervals (CI) between (A) 2 or (B) 6 reference groups. The latter column corresponds to the bootstrap-corrected estimates of mixing proportions among reference groups

		$p(m)$	Low CI	High CI	Bootstrap-corrected reference group
(A)	European	0.997	0.979	1.00	0.999
	North American	0.003	3.37 e−9	0.021	0.00074
(B)	Farmed Canada	0.00152	1.18 e−15	0.0137	0
	Farmed Ireland	0.00155	1.39 e−15	0.0140	0.0004
	Farmed Norway	0.971	0.917	0.997	0.98
	Farmed Scotland	0.00158	1.11 e−15	0.0144	0
	Wild Canada	0.00155	1.32 e−15	0.0139	0.00024
	Wild Scotland	0.0230	9.35 e−4	0.0726	0.02

salmon and identify their possible sources. Atlantic salmon may become an invasive species in Chile (Gomez-Uchida et al. 2018), but there is hitherto limited information about their origin (Soto et al. 2001, Arismendi et al. 2014). We evaluated genetic structure and ancestry among historical samples of free-living Atlantic salmon using a panel of SNPs. Our sampling design spanned nearly 2000 km of geographic coverage in freshwater and marine locations and life stages, including juveniles and adults. We additionally compared genotypes from free-living Atlantic salmon with genotypes from reference collections composed of 6 commercial aquaculture strains and 2 wild populations. Overall, our results pointed to commercial aquaculture strains as likely sources of free-living Atlantic salmon. We discuss these findings in further detail below and contextualize the limitations of our data set to inform other facets of the process of Atlantic salmon invasion.

4.1. Commercial aquaculture as the source of free-living Atlantic salmon

Individual-based clustering analyses (STRUCTURE and DAPC) revealed significant genetic structure among free-living Atlantic salmon (3 gene pools; K = 3), where geographical distribution (latitude), environment (marine versus freshwater), or life stage (juveniles versus adults) played no role in explaining genetic divergence patterns. Subsequent analyses incorporating reference collections strongly indicated that the genetic structure of free-living salmon was consistent with divergent gene pools found among commercial aquaculture strains and admixture between these collections. We found significant support for K = 2 gene pools in STRUCTURE, representing deep divergence (>1 million yr) between wild ancestors of European and North American commercial strains of Atlantic salmon (Rougemont & Bernatchez 2018). Most free-living individuals clustered within the European cluster and likely derived from Norway strains, as these are the most prevalent worldwide, including in Chile's aquaculture industry (Solar 2009, Yáñez et al. 2016). The exception was marine Site 19, showing admixture between European and North American strains. Results from DAPC confirmed the presence of admixed (hybrid) individuals in Site 19, explaining why they appeared genetically distinct from the rest. Crossing strains for improving salmon production through hybrid vigor seems to be frequent in the aquaculture industry (López et al. 2019), suggesting hybrid genotypes could be escapees from crossings made in aquaculture facilities. O'Reilly et al. (2006) and Bradbury et al. (2022) also showed evidence of admixture between local North American broodstock of Atlantic salmon and unapproved European broodstock. Bradbury et al. (2022) suggested that escaped Atlantic salmon of different origins may hybridize in the wild, a possibility hard to rule out in our study.

Mixed-stock analyses and individual assignment performed in RUBIAS suggested that free-living Atlantic salmon had European ancestry represented by aquaculture strains F1, F2, and F3 (6 reference groups) (Table S1), confirming our expectations. However, RUBIAS differed from results obtained by clustering using STRUCTURE and DAPC when assigning admixed individuals (e.g. Sites 5 and 19). This discrepancy was expected, since assumptions for mixed-stock and clustering analyses vary and this might result in different outcomes. Mixed-stock analysis assumes that individuals are purely of one reference group, and that reference collections or groups are discrete (Moran & Anderson 2019); under this assumption, the method has shown good performance even with small data sets (e.g. 91 SNPs; Kuismin et al. 2020). This is consistent with Z-scores from RUBIAS pointing to incongruence between maximum *a posteriori* probabilities and random normal expectations as a result of violations of assumptions induced by hybrid, admixed individuals in mixture samples. Indeed, 3 hybrid individu-

als from Site 19 showed *Z*-scores that departed the most from the normal curve and were assigned to F3 collection with a high probability. We suggest that assignment results using RUBIAS for Atlantic salmon from Sites 5 and 19 showing admixed/hybrid individuals need to be interpreted with caution. STRUCTURE seems best suited in this scenario to evaluate the ancestry of such individuals as the method also evaluates admixture coefficients (Pritchard et al. 2000).

4.2. What can genetic data of free-living Atlantic salmon tell us about patterns of establishment and naturalization?

Free-living salmon captured during the last 5–15 yr in freshwater and marine locations had a close ancestry with commercial aquaculture strains of Atlantic salmon, consistent with recent escapes from aquaculture facilities (Sepúlveda et al. 2013). Yet, our data need to be interpreted with caution as definitive evidence for Atlantic salmon establishment and naturalization is beyond the scope of our study. First, small sample sizes per site may be unable to detect very recent and subtle genetic structure among naturalized salmon populations, only permitting global assignment analyses of the entire data set as presented here. Indeed, recently founded populations will be weakly differentiated from donor aquaculture sources, and if there is ongoing gene flow between naturalized and aquaculture-escaped fish, this may prevent genetic divergence even further (Consuegra et al. 2011). Second, and assuming some stocking events in freshwater were successful during the early 20th century (Table 1), we either lack reference populations (e.g. Germany) or historical information about donor populations is not specific enough to compare free-living genotypes to (Basulto 2003). This is likely to obscure any results that may be pertinent to the success of Atlantic salmon early stocking events from a genetic perspective.

Several reviews of the literature indicate that early stocking events of Atlantic salmon were most likely unsuccessful in establishing naturalized, anadromous populations in Chile (Basulto 2003, Arismendi et al. 2014), even though there are documented records of established non-anadromous, resident or landlocked, Atlantic salmon populations in 4 watersheds from northern and southern Patagonia, but on the other side of the Andean range in Argentina (Vigliano & Alonso 2007). Two hypotheses have been discussed as to why Atlantic salmon have failed to establish self-sustaining, anadromous populations, despite continuous propagule pressure from aquaculture escapes: (1) they experience low survival in the wild (Soto et al. 2001) or (2) they experience intense competition with other invasive salmonids (Arismendi et al. 2014). Consistent with these hypotheses, occurrence and abundance of free-living Atlantic salmon have been historically low; both freshwater and marine surveys indicate that salmonid species such as rainbow trout, Chinook salmon, or brown trout, have often outnumbered Atlantic salmon (Soto et al. 2001, 2006, Arismendi et al. 2009, authors' unpubl. data).

Ultimately, testing for establishment of Atlantic salmon populations will require an integration of ecological and genetic approaches. First, increasing the breadth and intensity of freshwater surveys to remote watersheds will fill the gap of unsampled populations in genetic inference, some of which may reveal large numbers of juveniles, which would empirically suggest Atlantic salmon became established. Second, enhanced sampling efforts in rivers or streams located nearby areas with high densities of Atlantic salmon net pens at sea may be the best strategy to monitor establishment events following escapes (Fisher et al. 2014, Gomez-Uchida et al. 2018). Third, implementation of environmental DNA metabarcoding can help identify incipient populations and pinpoint susceptible habitats at low abundance (Deiner et al. 2017), whereas fatty acid profiles and stable isotope analyses can help reveal whether free-living salmon have transitioned from farmed to wild environments (Schröder & Garcia de Leaniz 2011, Skilbrei et al. 2015). Only then will we have robust evidence to determine whether massive aquaculture escapes of Atlantic salmon during recent times (Gomez-Uchida et al. 2018) and in the past (Soto et al. 2001) have resulted in self-sustainable populations.

5. CONCLUSIONS

Free-living Atlantic salmon have been consistently reported in marine and freshwater environments since the onset of commercial salmon aquaculture nearly 40 yr ago. Many authors have hypothesized that their presence is the result of continuous escapes rather than natural reproduction (Niklitschek et al. 2013, Arismendi et al. 2014). Our genetic inference — based on genetic structure, mixed-stock, and assignment methods from historical samples broadly

distributed from marine and freshwater sites and reference collections used in aquaculture — suggests this is the case given the consistency between genetic structure in free-living and aquaculture strains of Atlantic salmon. Yet recent establishment of self-sustaining populations of Atlantic salmon, thus weakly differentiated from aquaculture broodstock, cannot entirely be ruled out and highlights the need for increased monitoring efforts of free-living Atlantic salmon in remote sites as well as in watersheds located in densely stocked aquaculture areas. Only by integrating ecological and genetic analyses will we have the confidence to assess whether massive escapes in current and past times have resulted in Atlantic salmon establishing self-sustainable populations.

Acknowledgements. We thank the staff of Instituto de Fomento Pesquero (IFOP) for their sampling efforts during 'Evaluación y seguimiento de la situación sanitaria de peces silvestres en agua dulce y mar', a research program led by Departamento de Salud Hidrobiológica, funded by Chile's Ministerio de Economía, Programa ASIPA from Subsecretaría de Pesca y Acuicultura. This research was supported by Núcleo Milenio INVASAL funded by ANID–Programa Iniciativa Científica Milenio–NCN16_034 and NCN2021_056 from Chile's Ministerio de Ciencia, Tecnología, Conocimiento & Innovación. D.G.U. received additional support from ANID–FONDECYT 1191256.

LITERATURE CITED

Adey EA, Black KD, Sawyer T, Shimmield TM, Trueman CN (2009) Scale microchemistry as a tool to investigate the origin of wild and farmed *Salmo salar*. Mar Ecol Prog Ser 390:225–235

Anderson EC, Waples RS, Kalinowski ST (2008) An improved method for predicting the accuracy of genetic stock identification. Can J Fish Aquat Sci 65: 1475–1486

Arismendi I, Soto D, Penaluna B, Jara C, Leal C, León-Muñoz J (2009) Aquaculture, non-native salmonid invasions and associated declines of native fishes in Northern Patagonian lakes. Freshw Biol 54:1135–1147

Arismendi I, Sanzana J, Soto D (2011) Seasonal age distributions and maturity stage in a naturalized rainbow trout (*Oncorhynchus mykiss* Walbaum) population in southern Chile reveal an ad-fluvial life history. Ann Limnol 47: 133–140

Arismendi I, Penaluna B, Dunham J, García de Leaniz C and others (2014) Differential invasion success of salmonids in southern Chile: patterns and hypotheses. Rev Fish Biol Fish 24:919–941

Basulto S (2003) El largo viaje de los salmones: una crónica olvidada. Propagación y cultivo de especies acuáticas en Chile. Editorial Maval, Santiago

Benavente JN, Seeb LW, Seeb JE, Arismendi I and others (2015) Temporal genetic variance and propagule-driven genetic structure characterize naturalized rainbow trout (*Oncorhynchus mykiss*) from a Patagonian lake impacted by trout farming. PLOS ONE 10: e0142040

Besnier F, Glover KA (2013) ParallelStructure: a R package to distribute parallel runs of the population genetics program STRUCTURE on multi-core computers. PLOS ONE 8:e70651

Bourret V, Kent MP, Primmer CR, Vasemägi A and others (2013) SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). Mol Ecol 22:532–551

Bradbury IR, Lehnert SJ, Kess T, Van Wyngaarden M and others (2022) Genomic evidence of recent European introgression into North American farmed and wild Atlantic salmon. Evol Appl 15:1436–1448

Canales-Aguirre CB, Seeb LW, Seeb JE, Cádiz MI and others (2018) Contrasting genetic metrics and patterns among naturalized rainbow trout (*Oncorhynchus mykiss*) in two Patagonian lakes differentially impacted by trout aquaculture. Ecol Evol 8:273–285

Chalde T, Nardi CF, Fernández DA (2019) Early warning: detection of exotic coho salmon (*Oncorhynchus kisutch*) by environmental DNA and evidence of establishment at the extreme south of Patagonia. Can J Fish Aquat Sci 76: 2343–2349

Consuegra S, Phillips N, Gajardo G, de Leaniz CG (2011) Winning the invasion roulette: escapes from fish farms increase admixture and facilitate establishment of non-native rainbow trout. Evol Appl 4:660–671

Correa C, Gross MR (2008) Chinook salmon invade southern South America. Biol Invasions 10:615–639

Deiner K, Bik HM, Machler E, Seymour M and others (2017) Environmental DNA metabarcoding: transforming how we survey animal and plant communities. Mol Ecol 26: 5872–5895

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361

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

FAO (2020) The state of world fisheries and aquaculture 2020. Sustainability in action. FAO, Rome

Fisher AC, Volpe JP, Fisher JT (2014) Occupancy dynamics of escaped farmed Atlantic salmon in Canadian Pacific coastal salmon streams: implications for sustained invasions. Biol Invasions 16:2137–2146

Garcia de Leaniz C, Gajardo G, Consuegra S (2010) From best to pest: changing perspectives on the impact of exotic salmonids in the southern hemisphere. Syst Biodivers 8:447–459

Glover KA (2010) Forensic identification of fish farm escapees: the Norwegian experience. Aquacult Environ Interact 1:1–10

Gomez-Uchida D, Sepúlveda M, Ernst B, Contador TA, Neira S, Harrod C (2018) Chile's salmon escape demands action. Science 361:857–858

Hasselman DJ, Anderson EC, Argo EE, Bethoney ND and others (2016) Genetic stock composition of marine bycatch reveals disproportional impacts on depleted river herring genetic stocks. Can J Fish Aquat Sci 73:951–963

He C, Holme J, Anthony J (2014) SNP genotyping: the KASP assay. Methods Mol Biol 1145:75–86

Jensen Ø, Dempster T, Thorstad EB, Uglem I, Fredheim A (2010) Escapes of fishes from Norwegian sea-cage aquaculture: causes, consequences and prevention. Aquacult Environ Interact 1:71–83

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27: 3070–3071

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11: 94

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

Krkošek M (2010) Host density thresholds and disease control for fisheries and aquaculture. Aquacult Environ Interact 1:21–32

Kuismin M, Saatoglu D, Niskanen AK, Jensen H, Sillanpää MJ (2020) Genetic assignment of individuals to source populations using network estimation tools. Methods Ecol Evol 11:333–344

Lhorente JP, Araneda M, Neira R, Yáñez JM (2019) Advances in genetic improvement for salmon and trout aquaculture: the Chilean situation and prospects. Rev Aquacult 11:340–353

López ME, Linderoth T, Norris A, Lhorente JP, Neira R, Yáñez JM (2019) Multiple selection signatures in farmed Atlantic salmon adapted to different environments across hemispheres. Front Genet 10:901

Maldonado-Márquez A, Contador T, Rendoll-Cárcamo J, Moore S, Pérez-Troncoso C, Gomez-Uchida D, Harrod C (2020) Southernmost distribution limit for endangered peladillas (*Aplochiton taeniatus*) and non-native coho salmon (*Oncorhynchus kisutch*) coexisting within the Cape Horn biosphere reserve, Chile. J Fish Biol 96: 1495–1500

Monzón-Argüello C, Consuegra S, Gajardo G, Marco-Rius F, Fowler DM, Defaveri F, Garcia de Leaniz C (2014) Contrasting patterns of genetic and phenotypic differentiation in two invasive salmonids in the Southern Hemisphere. Evol Appl 7:921–936

Moran BM, Anderson EC (2019) Bayesian inference from the conditional genetic stock identification model. Can J Fish Aquat Sci 76:551–560

Musleh SS, Seeb LW, Seeb JE, Ernst B, Neira S, Harrod C, Gomez-Uchida D (2020) Mixed-stock analyses of migratory, non-native Chinook salmon at sea and assignment to natal sites in fresh water at their introduced range in South America. Biol Invasions 22:3175–3182

Naylor R, Hindar K, Fleming IA, Goldburg R and others (2005) Fugitive salmon: assessing the risks of escaped fish from net-pen aquaculture. Bioscience 55: 427–437

Niklitschek E, Toledo P (eds) (2011) Evaluación cuantitativa del estado trófico de salmónidos de vida libre en el fiordo Aysén, XI región. Informe Final Proyecto FIP 2008-30. Subsecretaría de Pesca, Coyhaique

Niklitschek EJ, Soto D, Lafon A, Molinet C, Toledo P (2013) Southward expansion of the Chilean salmon industry in the Patagonian Fjords: main environmental challenges. Rev Aquacult 5:172–195

O'Reilly P, Carr JW, Whoriskey FG, Verspoor E, (2006) Detection of European ancestry in escaped farmed Atlantic salmon, *Salmo salar* L., in the Magaguadavic River and Chamcook Stream, New Brunswick, Canada. ICES J Mar Sci 63:1256–1262

Pascual MA, Ciancio JE (2007) Introduced anadromous salmonids in Patagonia: risks, uses, and a conservation paradox. In: Bert TM (ed) Ecological and genetic implications of aquaculture activities. Springer, Dordrecht, p 333–353

Pascual M, Bentzen P, Rossi CR, Mackey G, Kinnison M, Walker R (2001) First documented case of anadromy in a population of introduced rainbow trout in Patagonia, Argentina. Trans Am Fish Soc 130:53–67

Poole WR, Webb JH, Matthews MA, Youngson AF (2000) Occurrence of canthaxanthin in Atlantic salmon, *Salmo salar* L., fry in Irish rivers as an indicator of escaped farmed salmon. Fish Manag Ecol 7:377–385

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

Purcell S, Neale B, Todd-Brown K, Thomas L and others (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575

Quiñones RA, Fuentes M, Montes RM, Soto D, León-Muñoz J (2019) Environmental issues in Chilean salmon farming: a review. Rev Aquacult 11:375–402

R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Rougemont Q, Bernatchez L (2018) The demographic history of Atlantic salmon (*Salmo salar*) across its distribution range reconstructed from approximate Bayesian computations. Evolution 72:1261–1277

Schröder V, Garcia de Leaniz C (2011) Discrimination between farmed and free-living invasive salmonids in Chilean Patagonia using stable isotope analysis. Biol Invasions 13:203–213

Sepúlveda M, Arismendi I, Soto D, Jara F, Farias F (2013) Escaped farmed salmon and trout in Chile: incidence, impacts, and the need for an ecosystem view. Aquacult Environ Interact 4:273–283

SERNAPESCA (Servicio Nacional de Pesca y Acuicultura) (2019) Anuario Estadístico de Pesca y Acuicultura. http://www.sernapesca.cl/informacion-utilidad/anuarios-estadisticos-de-pesca-y-acuicultura

SERNAPESCA (2021) Escape de peces en la salmonicultura. http://www.sernapesca.cl/sites/default/files/escape_de_peces_de_salmonicultura_2010_al_2020_20210203.pdf

Skilbrei OT, Normann E, Meier S, Olsen RE and others (2015) Use of fatty acid profiles to monitor the escape history of farmed Atlantic salmon. Aquacult Environ Interact 7:1–13

Solar II (2009) Use and exchange of salmonid genetic resources relevant for food and aquaculture. Rev Aquacult 1:174–196

Soto D, Jara F, Guerrero A, Moreno C and others (1997) Evaluación de Salmónidos de vida libre existentes en las aguas interiores de las Regiones X y XI. Informes Final Proyecto FIP-IT/95-31. Universidad Austral de Chile, Valdivia

Soto D, Jara F, Moreno C (2001) Escaped salmon in the inner seas, southern Chile: facing ecological and social conflicts. Ecol Appl 11:1750–1762

Soto D, Arismendi I, González J, Sanzana J and others (2006) Southern Chile, trout and salmon country: invasion patterns and threats for native species. Rev Chil Hist Nat 79:97–117

Stokesbury MJW, Lacroix GL, Price EL, Knox D, Dadswell MJ (2001) Identification by scale analysis of farmed Atlantic salmon juveniles in southwestern New Brunswick rivers. Trans Am Fish Soc 130:815–822

Vigliano PH, Alonso MF (2007) Salmonid introductions in Patagonia: a mixed blessing. In: Bert TM (ed) Ecological and genetic implications of aquaculture activities. Springer, Dordrecht, p 315–331

Vilata J, Oliva D, Sepúlveda Vilata M, Vilata J, María Bayarri J, Oliva SD, Sepúlveda M (2010) The predation of farmed salmon by South American sea lions (*Otaria flavescens*) in southern Chile. ICES J Mar Sci 67: 475–482

Wang YV, Wan AHL, Lock EJ, Andersen N, Winter-Schuh C, Larsen T (2018) Know your fish: a novel compound-specific isotope approach for tracing wild and farmed salmon. Food Chem 256:380–389

Yáñez JM, Naswa S, López ME, Bassini L and others (2016) Genomewide single nucleotide polymorphism discovery in Atlantic salmon (*Salmo salar*): validation in wild and farmed American and European populations. Mol Ecol Resour 16:1002–1011

Yoshida GM, Carvalheiro R, Lhorente JP, Correa K, Figueroa R, Houston RD, Yáñez JM (2018) Accuracy of genotype imputation and genomic predictions in a two-generation farmed Atlantic salmon population using high-density and low-density SNP panels. Aquaculture 491:147–154

Editorial responsibility: Ian Fleming, St. John's, Newfoundland and Labrador, Canada
Reviewed by: 2 anonymous referees, and previous version also reviewed by C. Correa.

Submitted: August 6, 2022
Accepted: October 13, 2022
Proofs received from author(s): December 7, 2022