

SCIENTIFIC REPORTS



OPEN

Polyphyletic ancestry of expanding Patagonian Chinook salmon populations

Cristian Correa ^{1,2} & Paul Moran³

Chinook salmon native to North America are spreading through South America's Patagonia and have become the most widespread anadromous salmon invasion ever documented. To better understand the colonization history and role that genetic diversity might have played in the founding and radiation of these new populations, we characterized ancestry and genetic diversity across latitude (39–48°S). Samples from four distant basins in Chile were genotyped for 13 microsatellite loci, and allocated, through probabilistic mixture models, to 148 potential donor populations in North America representing 46 distinct genetic lineages. Patagonian Chinook salmon clearly had a diverse and heterogeneous ancestry. Lineages from the Lower Columbia River were introduced for salmon open-ocean ranching in the late 1970s and 1980s, and were prevalent south of 43°S. In the north, however, a diverse assembly of lineages was found, associated with net-pen aquaculture during the 1990s. Finally, we showed that possible lineage admixture in the introduced range can confound allocations inferred from mixture models, a caveat previously overlooked in studies of this kind. While we documented high genetic and lineage diversity in expanding Patagonian populations, the degree to which diversity drives adaptive potential remains unclear. Our new understanding of diversity across latitude will guide future research.

Multiple independent introduction events can lead to shifts in genetic variation relative to native source populations, potentially boosting invasiveness and potential for rapid local adaptation¹. Furthermore, different introduction vectors delivering distinct genetic lineages in different regions can result in a mosaic of populations with varying genetic diversity and evolutionary potential². We investigated self-sustaining Chinook salmon [*Oncorhynchus tshawytscha* (Walbaum)] populations, currently part of a rapid colonization that is sweeping through Patagonia, the binational region of Chile and Argentina at the southern cone of South America. Specifically, we measured genetic diversity and evaluated the most likely phylogenetic origins of four self-sustaining populations spread over a wide geographical range that received multiple distinct introductions from different kinds of artificial propagation.

Beginning in the 1870s, Chinook salmon, native to North America and the North Pacific Ocean, were deliberately introduced into innumerable rivers in all continents except Antarctica³; yet successful naturalization has been rare. Self-sustaining adfluvial (migrating between lake and river) Chinook salmon populations have been established in the North American Great Lakes⁴, but anadromous (migrating between the sea and river) populations outside their native range exist only in New Zealand's South Island and in Patagonia³. The phylogenetic ancestry of New Zealand Chinook salmon was tracked to introductions in the early 1900s from the Sacramento River fall run (seasons characterize typical adult return to freshwater), most likely Battle Creek, California⁵. When studied about 30 salmon generations later, a number of phenotypic traits had evolved apparently in response to local environmental conditions^{5,6}, offering increased local fitness⁷. The monophyletic ancestry and known history of introduction in New Zealand greatly facilitated that research^{5,7}. This rapid evolution in New Zealand underscores the potential of Chinook salmon to adapt to new environments, yet it begs the question of why successful introductions have been rare elsewhere.

¹Facultad de Ciencias Forestales y Recursos Naturales, Instituto de Conservación Biodiversidad y Territorio, Universidad Austral de Chile, Casilla 567, Valdivia, Chile. ²Facultad de Ciencias, Instituto de Ciencias Marinas y Limnológicas, Universidad Austral de Chile, Casilla 567, Valdivia, Chile. ³Conservation Biology Division, Northwest Fisheries Science Center, Seattle, Washington, United States of America. Correspondence and requests for materials should be addressed to C.C. (email: cristiancorrea@gmail.com)

Patagonian Chinook salmon are much less studied than New Zealand Chinook salmon. Patagonian populations are younger than those in New Zealand but much more widespread. The ancestral origin of these populations is unclear, but out of the many introduction attempts beginning in the late 1800s, two relatively recent experimental commercial open-ocean ranching operations were most likely the sources of the introduced populations (Table 1). Self-sustaining populations occur in a vast region on western slopes south of 38°S draining through Chile and into the Pacific Ocean. They are also established on eastern slopes south of 50°S draining through Argentina and into the Atlantic Ocean (Fig. 1).

For the purpose of this study, we distinguish two classes of artificial propagation. In open-ocean ranching, returning adults are spawned artificially in fish hatcheries and the juveniles are released to grow at sea. These free-ranging fish may be harvested at sea but especially in a terminal fishery of maturing adults homing back to the stream of release. The second class is floating net-pen aquaculture. In this case, fish are held captive from hatching until harvest or reproduction (see below). One open-ocean ranching operation took place on Chiloé Island (42°S) during the period 1978–1982 using primarily Cowlitz River Hatchery spring-run brood stock from the Lower Columbia River, Washington, U.S.A.⁸. The other open-ocean ranching operation took place further south in the Magellan region (51°S) during the period 1982–1989 using University of Washington Hatchery fall-run broodstock, Puget Sound, Washington, as well as the progeny of adult fish returning to Chiloé Island^{9–11}.

By 1988, Chilean salmonid propagation shifted entirely from open-ocean ranching to net-pen aquaculture in captivity. Despite the captive design, fish escaped in the thousands every year and many of those escapes were poorly documented¹². The stocks utilized, also poorly documented, were relatively diverse, as ova importations came at least from Washington, Oregon, Vancouver Island, and from New Zealand (the latter with ancestry from the Sacramento River, Central Valley, California)¹³. In contrast to open-ocean ranching, net-pen aquaculture was mostly concentrated in Chile's Lake District region (42°S)¹². Therefore, in contrast to New Zealand, Patagonian Chinook salmon represent a potentially polyphyletic assemblage, offering new and interesting research avenues to study colonization dynamics and contemporary evolution. A summary of Chinook salmon introductions to Patagonia is presented in Table 1, and a complete review of introductions is provided as an online supplement (Supplementary Table S1).

We analysed the genetic diversity and population structure of Chinook salmon in four western Patagonia basins. We then compared these to native North America populations, and finally estimated ancestry and lineage distribution in regions that differed in their history of artificial propagation by open-ocean ranching and net-pen aquaculture. For estimation of ancestry, we used two different classes of genetic mixture analysis, conditional maximum likelihood (CML) and model-based clustering (M-BC). We took advantage of a large, interagency, microsatellite baseline as our known-origin reference against which we compared fish from Patagonia. The International Genetic Analysis of Pacific Salmonids (GAPS) consortium assembled a comprehensive genetic baseline for coast-wide fishery management of mixed-stock fisheries¹⁴. The accuracy and precision of mixture analysis using the GAPS baseline is substantial with either conditional or unconditional Bayesian methods^{14–17} (for an assessment of the accuracy of the CML method, see Supplementary Figure S1), and mixture analysis provides a potentially useful approach for inferring the ancestry of Patagonian Chinook salmon. CML mixture analysis was used in two recent studies of the ancestry of Patagonian Chinook salmon^{18,19}. However, as far as we know, the validity of this application in mixed-origin populations has not been demonstrated. Introduced populations violate the fundamental assumption of the model—that individuals in the unknown sample actually originated from one of the baseline populations. Newly founded populations are isolated and expected to diverge from their ancestors through founder effects and genetic drift¹⁸. An even greater concern is interbreeding of mixed-origin fish in the new range (admixture), producing novel genotypes not accurately attributable to any single source population. Founding effects and drift are not easily accounted for, but neither are they likely to fundamentally confound our analysis. When considering microsatellites or other neutral markers, it is unlikely that founding and drift would result in a Patagonian population looking more like an unrelated population than the true population of origin. Such a result would require parallel patterns of allele frequency convergence at multiple highly polymorphic loci. The problem of interbreeding is more serious. It was unclear to us how the Rannala and Mountain²⁰ CML mixture algorithm would treat individuals of mixed origin. Simulations and sensitivity analysis helped us answer that question, and M-BC allowed an independent evaluation of CML mixture analysis in this unusual application of stock identification of introduced and naturalized populations.

Results

Data quality, neutrality, and genetic structure. Data quality and genotyping success were high, with an average of 12 out of 13 microsatellite loci scored per individual fish. We removed from the analysis 9 Toltén individuals that gave no reliable genotypes (collected as decomposing carcass samples), leaving 87 individual Patagonian Chinook salmon for our study.

Ots213 and *OtsG474* departed significantly from neutral expectation (F_{ST} outlier test) in Patagonian and North American populations, respectively, and were removed from further population genetic analyses due to potential directional selection. These loci were retained for genetic mixture analysis, which is generally robust to departures from neutrality. *Oki100* and *Ots208b* also departed from neutral expectation in Patagonian populations, but were only marginally significant and so were retained. Pairwise allele frequency differences were not significant between sample collection locations within river basins in Patagonia, even including all loci (range of intrabasin P values, 0.0042–0.690, Bonferroni-corrected $\alpha = 0.0014$) and were therefore pooled for population-level analyses. The analyses we present here are based on those river-basin-level aggregates of collections that are intended to represent separate populations, though because our samples are small, we scrutinized that assumption in several ways. For example, mean F_{IS} value for Patagonian populations (0.012, SD = 0.013) was substantially larger than estimates from the North America baseline populations (0.008, SD = 0.004, Supplementary Table S2), and although statistically non-significant, those heterozygote deficits might indicate a Wahlund effect of sampling

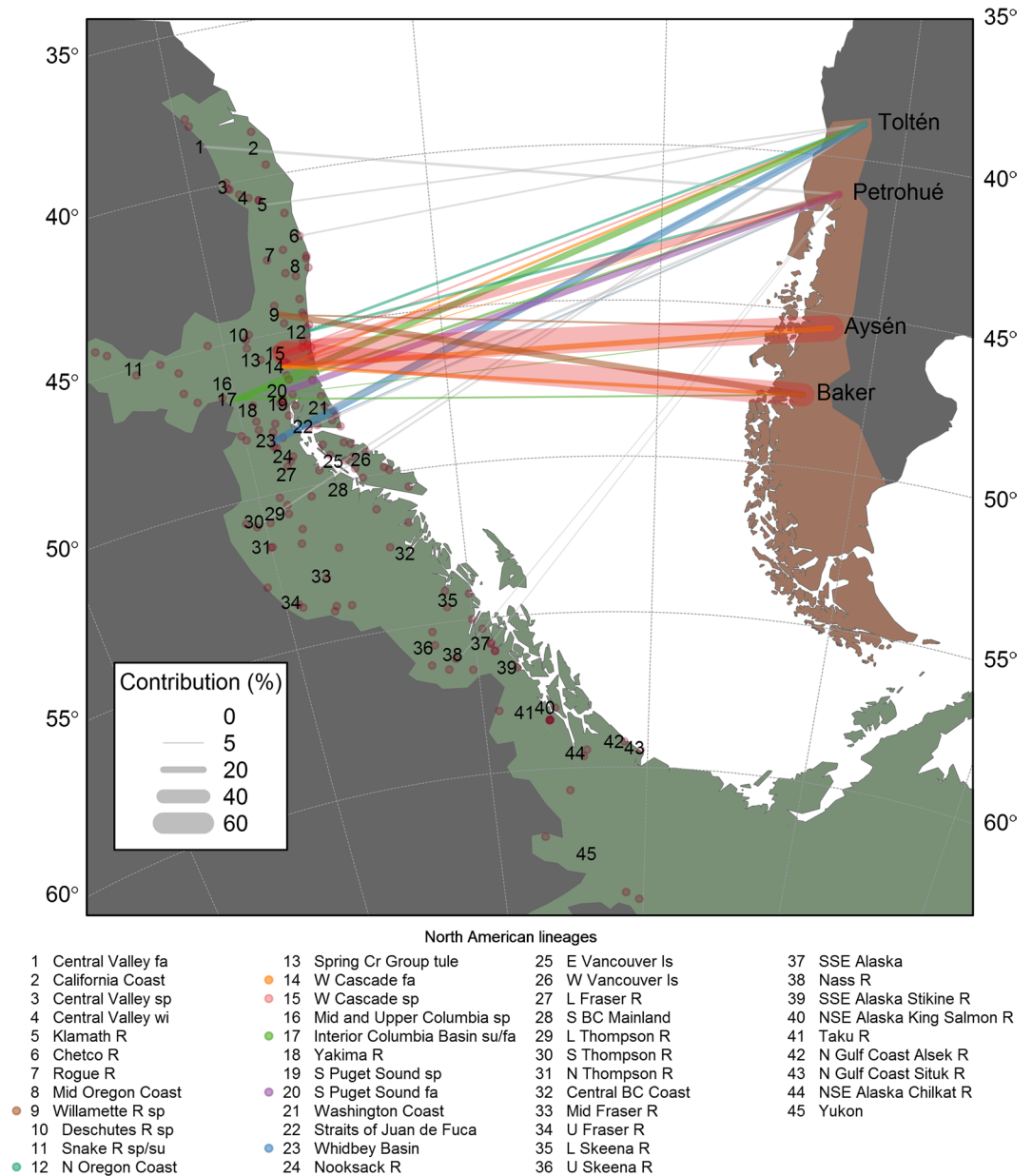


Figure 1. Chinook salmon native (green) and invasive (brown) range in the Americas, and illustration of genetic contribution of North American lineages to South American introduced populations (connecting vectors). The North American continent has been rotated and translated such that the North American and South American Pacific coasts face each other matching latitudes (longitudes were shifted for clarity). Connecting vectors represent average genetic contribution (proportional to line thickness) of the 46 North American baseline reporting groups (lineages) used in conditional maximum likelihood (CML) mixture analysis. Contributions $\geq 10\%$ were colour-coded by lineage, and smaller contributions were displayed in grey. Lineages were characterized by 146 baseline populations (dots). North and South American reference maps⁵⁰ were extracted with function worldHires of R package mapdata⁵¹ whereas schematic representations of distribution ranges were redrawn from references^{35,52}. The figure was produced with the R software environment version 3.3.1⁴⁹.

distinct populations within a river basin. We also found potential evidence of non-equilibrium conditions from fewer-than-expected split probabilities in assignment tests (below, *Split assignments consistent with primary donors*).

Genetic diversity. Average genetic diversity in the Patagonian populations (mean $H_S = 0.83$, $SD = 0.024$) was essentially identical to the average of all North American baseline populations (0.83 , $SD = 0.038$) and close to the populations identified as likely founders (see below, e.g., Puget Sound, Whidbey Basin, West Cascades, and

Country year(s) Basin, Latitude River stocked	Number of Individuals Released [*]	Ontogenetic Stage	Stock Origin	Adult Returns	Comments
Chile, 1970, 1971 Bueno, 40°S Río Chirri ^{1,2}	50,000 (1970) ² ; 270,150 (1971) ²	Subyearlings (8 mo, 15 g, 12 cm in 1970; 5 mo, 5.5–17 g, 6–12 cm in 1971) ²	USA, Green River Hatchery ² (Cowlitz River, lower Columbia River, Washington).	?	Agriculture and Livestock Service of the Government of Chile and the U.S. Peace Corps. First shipment by plane ^{2,3} . Eggs were received at Lautaro Hatchery on December; stockings took place 36 (Sep 1970) or 23 weeks later (May 1971) ² .
Chile, 1978 Coastal, 42°S Chiloé Island, Curaco de Vélez ^{1,4–6}	120,000 (late 1978) ^{4,5,7} ; 170,000 ^{1,6}	Smolts (1+; 70 g) ^{4,5}	USA, Cowlitz River spring-run (lower Columbia River, Washington) ⁷	Yes ^{4,5,7}	Domsea Pesquera Chile Ltd. (Union Carbide Corporation, USA), began salmon ocean- ranching experimentation ^{1,5,6} . In 1979, 334 returning jacks and 2 females were trapped ^{4,5} . In the period 1979–1982, 1050 returnees of this release were recorded ⁷ .
Chile, 1979 Same location ⁵	190,000 ^{4,5}	Smolts (1+) ^{4,5}	Idem ⁷	Yes ^{4,5,7}	Domsea Pesquera Chile Ltd. ^{4,5} . In the period 1980–1982, 228 returnees of this release were recorded ⁷ .
Chile, 1980–1981 Same location ⁷	90,000 (late 1980, early 1981) ⁷	Smolts (1+) ⁷	USA, Bonneville Hatchery fall-run (Columbia River, Washington) ⁷	Yes ⁷	In the period 1981–1982, 260 returnees of this release were recorded ⁷ .
Chile, 1982 Same location ⁷	3000 (early 1982) ⁷	Smolts (1+) ⁷	USA, University of Washington's Hatchery fall-run ⁷	?	In 1981, Domsea Pesquera Chile Ltd. was sold to Fundación Chile (private, non-profit), and renamed Salmenes Antártica Ltd. ^{1,6} . Stocking continued at this location at least during the first year of the new administration.
Chile, 1982 Same location ⁷	>1,000,000 fish were being raised ⁷ , but their fate remains unknown to us.	Ova & subyearlings ⁷	USA, University of Washington's Hatchery fall-run ⁷ and Bonneville Hatchery fall-run ⁷ ; Chile, progeny from local returnees ^{1,6,7}	?	Through October 1982, 1538 adults had returned to the hatchery from previous brood years. Returnee's progeny (F2) was being raised at the facility along with fry from two additional importations ⁷ . However, we found no posterior records of fish release at this location.
Chile, 1982 Coastal, 54°S Río Santa María ^{6–8}	200,000 ^{7,8}	Fry ⁷	USA, University of Washington's Hatchery fall-run ^{6,7,9}	? ^{6,9} . Yes ¹⁰	Fundación Chile through Salmenes Antártica Ltd. launched new facility in the Magellan region subsequently destroyed by storm, and abandoned ⁶ . Jacks seen returning in 1983 ¹⁰ .
Chile, 1983 Prat, 51°S Río Prat ¹¹	5,000 ¹¹	Smolts (1+) ¹¹	USA ¹¹ , University of Washington's Hatchery fall-run ^{9,12} (?); Chile, progeny from returnees at Curaco de Vélez and Astilleros, Chiloe ^{11,12} .	Yes (~2.3%) ¹¹	Fundación Chile through Salmenes Antártica Ltd. launched another facility in the Magellan region with successfully returning spawners ^{6,8} .
Chile, 1987 Same location ¹¹	294,967 ¹¹ (USA origin); 40,042 ¹¹ (Chilean origin)	Smolts (1+) ¹¹	USA ¹¹ , seemingly University of Washington's ⁹ ; Chile, progeny from returnees ¹¹ at Río Prat (local), Curaco de Vélez, and Astilleros ¹³	Yes (~0.07% until 1989) ¹¹	Continuation of the above enterprise. In 1998 Fundación Chile and Salmenes Antártica created Salmotec S.A. ¹⁴ .
Chile, 1989, 1990, 1993 Bueno 40°S Estero Huillin ¹⁵	? (1989–1990); 3347 (Jan 1993) ¹⁵	Smolts (0+)	? (1989–1990); Chile, progeny from 1992 returning adults (38 females + 12 males) ¹⁵	Yes, at least from 1989–1990 stockings ¹⁵	Universidad de Los Lagos' experimentation at Piscicultura Experimental Lago Ranco ¹⁵ .
Chile, 1987–2000 Coast, 39–45°S Inner seas ¹⁶	100,000 ¹⁷	Mostly subadults ¹⁸	USA, Washington Sate ^{19,20} ; USA, Oregon State ²¹ ; Canada, Vancouver Island ²² ; New Zealand ²³ ; USA, Alaska ²⁴ (?); Australia ²⁵ (?)	Yes ²⁶	Chinook stocks were imported primarily to the Lakes District Region for commercial net pen rearing. Last recorded importation in 2000.

Table 1. Chinook salmon deliberate and accidental releases in Patagonia since 1970 (a complete review spanning earlier introductions is available in Supplementary Table S1; modified from reference³⁵). Table footnotes: *The actual number of individuals released may be less than the figure reported due to mortality during transport and handling; pre-release mortality was accounted for whenever possible. Approximate latitude is given at the river mouth. ? = unreported, likely stock origin, or lack of adults return assessment. Follow footnotes indicated by superscript numbers; references were numbered as in the main text: ¹Fundación Chile⁵³; ²Snyder⁵⁴; ³Ellis and Salo (1969) in Basulto⁵⁵; ⁴Lindbergh *et al.*⁸; ⁵Lindbergh²⁸; ⁶Méndez and Munita¹¹; ⁷Lindbergh and Brown⁹; ⁸Basulto⁵⁵; ⁹Donaldson and Joyner¹⁰; ¹⁰Manuel Barros personal communication (2008) in Aedo¹³. At the time, M. Barros worked for Fundación Chile.; ¹¹Salmotec Ltd. in Sakai⁵⁶; ¹²Cristian Jélvez personal communication (2005) in Aedo¹³. C. Jélvez worked for Fundación Chile (1982).; ¹³Fredy Carrasco personal communication (2005) in Aedo¹³. F. Carrasco worked for Fundación Chile (1986).; ¹⁴United Nations⁵⁷; ¹⁵Del Real⁵⁸. Aedo¹³ mentioned other stocking locations (Río Contaco and Río Maicolpué) by Universidad de los Lagos, but we found no further records of these releases.; ¹⁶Primarily marine aquaculture concessions in the Lake District region.; ¹⁷Rough estimate of number of sub-adult Chinook salmon escapees (see main text).; ¹⁸Mostly 1+ year class and older since most escapes were from marine net-pens³².; ¹⁹Follow fragmentary records of ova imported (OI) by the Chilean aquaculture industry in 1987–2000¹³. Some information of suppliers was available for 60% of the imports; we report specific lineages and origins of livestock whenever possible, and ova suppliers and/or geographic origin of shipments otherwise. Additional potential sources of the unaccounted imports were identified from import permits (OP) issued by the Chilean National Fisheries Service (SERNAPESCA), although it remains unclear if these planned importations ever materialized. Sources listed in decreasing order of importance¹³; ²⁰OI: Columbia River. OP: Fish Pro Inc. and University of Washington; ²¹OI: Springfield. OP: Aqua Food, Aquafoods, and Aqua Seed Corp.; ²²OI: Koksilah River. OP: Sea Spring Salmon Farms Ltd., Hardy Sea Farms, Hatfield Consultants Inc., Hatfield International SA., Fishpro, and Aqua Seed; ²³OI: Sanford Waitaki Salmon Hatchery (Kaitan Gata). OP: Big Glory Bay Hatchery, and Kaitan Gata Hatchery and Sanford Waitaki Salmon Hatchery (Stewart Island).; ²⁴OP: Sitka; ²⁵OP: Tasmania; ²⁶This study.

Oregon Coast, 0.86, SD = 0.026, Supplementary Table S2). Similarly, average allelic richness in Patagonian populations ($AR = 9.168$, SD = 1.420) was not significantly different from that observed in all North American populations (9.810, SD = 1.131) nor in the putative founders (10.618, SD = 0.703). The 259 alleles observed in Patagonia included more than half of the 490 observed in North America, from the Yukon River in the Bering Sea to Central Valley California—most of the native species range²¹. Within Patagonia, overall genetic diversity did not differ significantly among river basins, as inferred from heterozygosity and allelic richness (Supplementary Table S2).

Population structure. Between Patagonian river basins, allele frequencies were highly significantly different (mean pairwise $P = 0.00017$, Bonferroni-corrected $\alpha = 0.008$). Similarly, pairwise F_{ST} values were large among Patagonian populations (mean $F_{ST} = 0.038$, SD = 0.004), especially considering the relatively narrow range of genetic differentiation among the putative source populations (0.037, SD = 0.007) relative to all of North America (0.065, SD = 0.011).

Contingency testing, pairwise F_{ST} estimates, and jack knife self-assignment analysis were all consistent in showing the closest genetic relationship between the two southern populations Aysén and Baker ($F_{ST} = 0.01$). Petrohué was similar to Baker and Aysén, whereas Toltén, the northern-most population, was the most distinct of the four.

Origin of Patagonian Chinook salmon (CML mixture analysis). We found evidence of multiple North American lineages in Patagonia. Of the 46 reporting groups, 16 (35%) were identified by CML mixture analysis as possibly represented in Patagonia (Figs 1 and 2). Despite the high accuracy of CML mixture analysis in assigning individual genotypes back to the correct reporting groups, we expected a small rate of misassignments (e.g., 1–3%) to unrelated reporting groups, as a result of random associations between individual genotypes and reporting group's allelic frequencies (Supplementary Figure S1). More importantly, however, based on our simulations, we expected a higher fraction (potentially as high as 10%) of spurious assignments associated with admixture among lineages occurring in the introduced range (see below, *Results: Simulated mixed-origin founding*). Hence, in order to emphasize major contributing lineages and avoid over-interpretation, we focused on putative contributors with *c.* 10% or greater allocation from any source to any single Patagonian population (Fig. 2). Seven potential source lineages satisfied this criterion. Approximately south to north in their native range, these were: Willamette River spring, North Oregon Coast, West Cascade fall, West Cascade spring, Interior Columbia Basin summer/fall, South Puget Sound fall, and Whidbey Basin.

The estimated number of lineages that contributed to our South American populations was higher in the north (Toltén and Petrohué) than in the south of our study area (Aysén and Baker) (Fig. 1). West Cascade fall, and especially West Cascade spring, contributed substantially, especially in the south with 71% and 65% genetic contribution of West Cascade spring to Aysén and Baker populations (Fig. 2). Interior Columbia basin summer/fall lineage contributed to all four Patagonian populations, but especially to those in the north. Other contributors showed more localized effects. For example, South Puget Sound fall in Petrohué (20%), Whidbey Basin in Toltén (20%), and Willamette River spring in Baker (18%). Various other donors had lower contributions, but, again, some of those results might be attributable to misassignments, especially those observed in the simulation of mixed ancestry study.

Whereas the two northern river basins (39°–42°S) appeared highly polyphyletic, the two southern basins (45°–48°S) were nearly monophyletic, attributing nearly all of their ancestry to the closely related lower Columbia River, West Cascade spring and Willamette River spring lineages. Despite small sample sizes and apparent mixed ancestry, we found general agreement between modelled stock composition estimates and the proportions of individual fish assignments (Fig. 2; Supplementary Table S3).

Split assignments consistent with primary donors. The ancestries of most fish were readily identifiable though CML mixture analysis. Most assigned with high probability to one of the baseline reporting groups (average maximum *a posteriori* value = 0.903, SD = 0.145) or reference populations (0.861, SD = 0.168) (Supplementary Figure S2). High assignment probabilities to baseline reporting groups were observed even in basins where spawners from multiple lineages were collected together (*i.e.*, Toltén and Petrohué). Both at the reporting group and population-level, individual fish that did show affinity to multiple source lineages, were invariably associated with the same sources to which other fish in the same collection assigned with high probability. For example, Petrohué and Toltén had individuals that assigned with relatively high probability to North Oregon Coast, West Cascade spring and fall, Interior Columbia Basin summer/fall, and South Puget Sound fall; however other presumed mixed-origin individuals split their assignment probability among those same sources.

Simulated mixed-origin founding. Our simulation result showed that most individual fish in a population derived from multiple sources would assign back to those sources, often splitting their assignment probability between source populations (Supplementary Figure S3). However, we also learned that sometimes a non-trivial number of simulated, mixed-origin fish might assign with high probability to unrelated lineages. Some misassignment was expected to genetically similar reporting groups. Both reporting groups used in our simulation have closely related sister groups in the GAPS baseline. West Cascade spring is genetically similar to West Cascade fall, and South Puget Sound fall is similar to Whidbey Basin (an adjoining inland water body). Most simulated fish (61.3%) assigned to one or the other of the true source reporting groups. Another 28.2% assigned to those closely related sister groups. That level of misassignment was expected based on leave-one-out jackknife analysis of the North American baseline. Approximately 17% of real fish collected from Cowlitz River spring in the West Cascade spring reporting group (one of the seed populations for the simulation) misassigned to Cowlitz River fall in the West Cascade fall reporting group (Supplementary Figure S1).

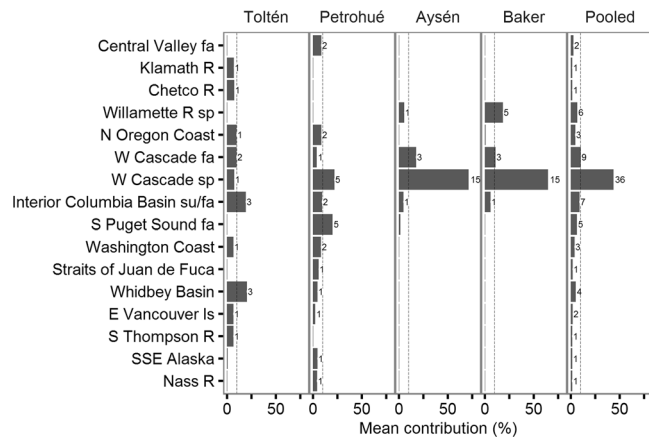


Figure 2. Estimated contribution of North American Chinook salmon lineages to Patagonian populations, as inferred from reporting group-level CML. A strong trend in lineage diversity was apparent, with populations from southern rivers (Aysén and Baker) dominated by the West Cascade spring run lineage, whereas those from northern rivers (Toltén and Petrohué) composed by a broad diversity of lineages. With a few exceptions, contributing lineages were concordant with historical introductions and regional differences in propagation activities. Numbers beside bars of mean proportional genetic contribution indicate number of individuals allocated to a particular lineage. Dashed vertical line indicate 10% genetic contribution.

Expected misassignment to closely related reporting groups contrasted strongly with assignment of simulated individuals to unrelated groups. In our simulation, 10% of individuals assigned to unrelated reporting groups, 44% of these with high assignment probability ($P \geq 0.8$). Interior Columbia Basin summer/fall received about half of these misassignments, whereas the rest were distributed among 14 other unrelated reporting groups. This level of misassignment observed in simulated fish was much higher than the normal level of misassignment observed between closely related reporting groups in North America (Supplementary Figure S1). Note that this result is completely simulated and is not related to mutation or drift in the founded populations or in the baseline.

Our simulation study confirmed the general utility of CML mixture analysis for individual assignment and proportional allocation despite admixture, but also showed the potential for misleading results, even when individuals assign with high probability. Therefore, we interpreted our empirical results of CML mixture analysis of real fish with caution when the estimated proportional contribution of a reporting group approached the level of misassignment we observed in the simulation (i.e., $< 10\%$, see above *Results: Origin of Patagonian Chinook salmon*). Any criterion would be somewhat arbitrary, but for the empirical dataset there seemed to be a break between 0.08 and 0.1, and we knew that values below 0.08 could be relatively strongly influenced by the spurious assignments we revealed in the simulation.

Therefore, the simulations largely confirmed the utility of the CML algorithm for studies of mixed ancestry with the important caveat that a few mixed origin individuals can show high probability of membership to completely unrelated lineages.

Equivocal results regarding expected contributors. Not a single fish we sampled from four locations in Patagonia showed any affinity whatever to the University of Washington Hatchery fall-run stock (zero relative probability of membership). Only five fish, collected in the Petrohué basin, assigned with high probability to another Puget Sound fall-run hatchery stock, Soos Creek (from which University of Washington brood stock was derived), but none were similar to our contemporary sample from University of Washington Hatchery fall-run stock.

Central Valley California was also a suspected source of Patagonian Chinook salmon, however, our results offered very little support for that conclusion at the locations we sampled. Only two fish assigned to Central Valley populations, one to Tuolumne River and the other to Stanislaus River. It was unclear if those two fish demonstrated true Central Valley ancestry or spurious assignments resulting from admixture (see above *Simulated mixed-origin founding*).

Model-based clustering (M-BC). Our reduced reference dataset for M-BC (see *Methods*) included 31 of the GAPS baseline populations²¹ from seven reporting groups (Whidbey Basin, S. Puget Sound fall, Interior Columbia Basin summer/fall, West Cascade fall, West Cascade spring, North Oregon Coast, and Willamette River spring). These reporting groups contributed most (84.13%) of the combined posterior probability in the reporting group-level CML mixture analysis (Fig. 2). Examination of log likelihood of the modelled clusters²² for various numbers of clusters (K) in the North American populations showed a peak in delta log likelihood at $K = 3$ and $K = 5$ (Supplementary Figure S4). We observed the clearest discrimination of the seven reporting groups at $K = 5$ (Supplementary Figure S5), and that value showed high concordance with the genetic stock reporting groups that we used in the CML mixture analysis (Fig. 3a). Concordance between M-BC and CML was also evident in the Patagonian populations where the mean Q -values matched the proportional estimates of North American lineages based on CML (compare Fig. 2 with Fig. 3). For example, from south to north, individual Baker fish showed

Q-value proportions mostly consistent with ancestry from West Cascade spring/fall (Q3), or a combination of West Cascade spring and Willamette River spring runs (Q4). Aysén fish mostly showed affinity with West Cascade spring/fall runs (Q3), but also some possible Willamette River spring (Q4) and little South Puget Sound fall and Whidbey Basin influence (Q5). In Petrohué, West Cascade spring/fall and Willamette River (Q3 and Q4) lineages were still important, but a substantial increase of all other lineages was apparent, particularly South Puget Sound fall and Whidbey Basin (Q5). This trend towards more diverse ancestry in the north of the Patagonian range persisted in Toltén where fish showed affinity to all five ancestral lineages inferred by the M-BC analysis (Q1 to Q5), and by extension the seven lineages distinguished in phylogeographic analysis of Chinook salmon in the native range²¹. Thus, the ancestry of Patagonian Chinook salmon populations was strongly spatially structured (Fig. 3b). In general, model based clustering suggested broader, more uniform contributions from the source populations that were identified with CML mixture analysis. Beyond that, however, results from M-BC were highly concordant with CML and led to nearly identical inferences about the ancestry of naturally-sustaining, Patagonian Chinook salmon populations.

Discussion

We investigated the phylogenetic ancestry of introduced Chinook salmon in Patagonia by conducting two different classes of genetic mixture analysis. Diverse genotypes led to the identification of multiple putative ancestral sources, introduced primarily from the states of Oregon and Washington. Our results were largely consistent with historical records of fish introductions and recent molecular genetic studies. However, we also found interesting differences such as apparent contributions from undocumented introductions, and conversely, lack of evidence supporting the naturalization of well-documented introductions.

Our study is distinguished from previous molecular studies of Patagonian Chinook salmon^{18,19,23–25} in three principal ways: (1) We used the most inclusive baseline dataset possible, including all potential North American donor lineages. Previous studies have been more or less limited by the number of reference lineages available for particular genetic markers. The GAPS microsatellite baseline is the most extensive of its kind and is ideally suited for this application. (2) To our knowledge, ours is the first study of its kind to take into account the potential for spurious allocation of admixed genotypes. Other studies have addressed genetic drift in introduced populations¹⁸, but not admixture between divergent lineages. We suggest that admixture may be more problematic than genetic drift, and simulations of mixed-origin founding offered a cautionary note on potentially misleading results. Our observations are relevant to many studies that seek to characterize small genetic contributions from multiple founding lineages²⁶. (3) Our study included naturalized populations both within and outside salmon aquaculture zones. This helped clarify the poorly documented role of net-pen aquaculture on Chinook salmon introduction to Chilean Patagonia. Thus, our study provides important new insight into the ancestry and history of colonization of Patagonian Chinook salmon.

Mating structure is critical to our understanding of potential for adaptation and persistence of introduced lineages in Patagonia. Even though allele frequencies led us to pool within-river-basin samples into putative populations, we did find some evidence of heterozygote deficits including non-significant yet elevated F_{IS} values within basins. Recognizing limited power for detecting a Wahlund Effect, we also tried to draw inference from CML individual assignment probabilities. Fewer intermediate assignments (split probabilities between reference groups) were observed than expected for a (simulated) mixed-origin population at equilibrium. If real, such a result and the possibly low heterozygosity might indicate assortative mating of two or more genetically distinct lineages within a river basin. Alternatively, strong out-breeding depression²⁷ or insufficient time since introductions could result in fewer than expected mixed-origin and heterozygous individuals. A companion study to ours used the same Patagonian samples and an intensive genome scan of 11,759 single nucleotide polymorphisms (SNPs)²⁵. That study compared results to those from selected North American populations that our study had identified as potential founders. Average F_{IS} values for Patagonia populations were an order of magnitude higher for SNPs than for microsatellites. Yet SNPs estimated similar F_{IS} values between North and South America. This was in contrast to our observation of a two-fold higher F_{IS} in Patagonia. Resolution of these discrepancies and elucidation of specific population structure will require larger, and spatially stratified samples within and among river basins and across multiple generations.

Our genetic CML mixture analyses strongly suggested a diverse array of founding lineages in Patagonia. Moreover, the number of founding lineages was clearly higher in northern Patagonia than in the south, and yet this latitudinal trend was uncorrelated with standard measures of genetic diversity. Overall, population genetic diversity measured by heterozygosity and allelic richness was higher than expected in Patagonia—nearly identical to North American populations. Both of these surprising results were consistent with those from the companion SNP study of the same Patagonian populations²⁵. Initially, we assumed high genetic diversity, i.e., heterozygosity and allelic richness, was a result of mixed ancestry. Previous studies have made similar conclusions^{18,19,24}. CML mixture analysis and M-BC both estimated more founding lineages in the two northern Patagonian sites. Surprisingly, however, populations with the greatest number of founding lineages showed no higher levels of heterozygosity or allelic richness. Nor was there spatial structure for heterozygosity or allelic diversity, as was evident for lineage diversity. It would seem almost axiomatic that a diversity of founders would introduce more alleles relative to monophyletic populations, yet, as mentioned above, other factors might be confounding a clear association between lineage diversity and genetic diversity.

A principal challenge in this study was to distinguish the genetic signal of true ancestry from the noise created by mixed-origin, introgressed populations. In our CML mixture analysis we expected some misassignments between related populations. The patterns of misassignment in the North American reference baseline (Supplementary Figure S1) were well understood based on previous results¹⁶. However, our simulation of mixed-origin founding revealed some spurious genetic allocation not previously reported in studies that used

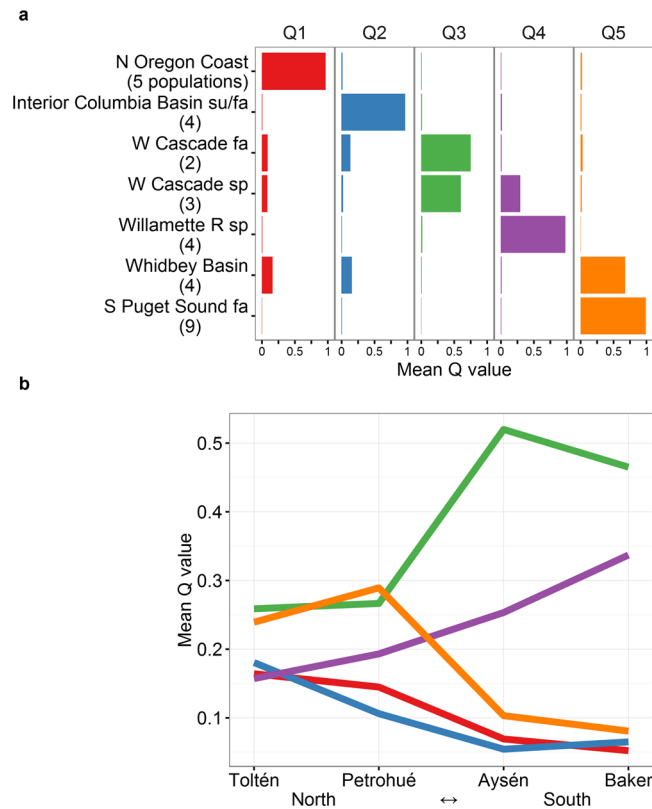


Figure 3. Latitudinal trend in lineage diversity as inferred from model-based clustering (M-BC). The clearest discrimination among the 31 baseline populations selected for this analysis was achieved at five clusters ($K = 5$), which associated (Q1–Q5) to known North American lineages (a). Membership of Patagonian samples to these clusters (i.e., inferred ancestry), showed a decline in lineage diversity from north to south, consistent with CML mixture analysis. Standard ΔK plot, and STRUCTURE plots, are available in Supplementary Figures S4 and S5.

this approach^{18,19,26}. The CML simulation showed us that mixed origin (admixed) individuals usually assign to the true donors, however, some individuals will assign to unrelated populations, sometimes with high assignment probability. Note, however, that the probability values are conditional on available baseline references. A high assignment probability might simply mean that no other reference population could likely have produced a given genotype. This is relevant, for example, to studies that assume a probability threshold (e.g., 0.8) for accurate assignment, and especially to those with incomplete baseline datasets. While a probability threshold such as 0.8 is clearly useful in conventional genetic mixture applications¹⁷, in a mixed-origin founding application, it is almost certainly the case that some high-probability individual assignments are attributable to the misassignment phenomenon that we demonstrated through simulation. Apparently, introgression between donor populations results in unique genotypes that assign arbitrarily to other, unrelated populations. To avoid falsely implicating potential donor lineages we set a lower threshold for genetic contribution below which putative ancestry was viewed with caution and some skepticism. Based on review of both simulated and empirical data, we used a threshold of 10% estimated genetic contribution to any single population in Patagonia. It was not that we rejected as potential contributors lineages below 10%, rather we were less confident in identifying those as founders.

Despite this caveat, our findings clearly confirm the importance of entrepreneurial attempts to establish Chinook salmon open-ocean ranching programs in Chile during the late 1970s and 1980s. In particular, the West Cascade spring and fall lineages, introduced from the Lower Columbia River, were likely associated with the enterprises suspected to have initiated the invasion. In 1978 Domsea Pesquera Chile Ltda. (a subsidiary of Union Carbide Corporation, USA) began yearly stockings of Chinook salmon in a small stream on Quinchao Island, Curaco de Vélez, Chiloé (42°S). Three strains of Chinook salmon were introduced: Cowlitz River spring run, Bonneville Hatchery fall run, and University of Washington Hatchery fall run, in decreasing number of individuals released^{19,28}. The primary stock that appeared to become established was Cowlitz River spring run^{8,28,29}; precisely the same primary ancestral lineage we identified in our samples. The finding of the Interior Columbia River summer/fall lineage in our samples probably reflects the introduced Bonneville Hatchery fall-run. By contrast, there is little evidence of successful University of Washington Hatchery fall-run (see below).

Subsequent open-ocean ranching experiments in the Magellan region (51°S) utilized the University of Washington Hatchery fall-run stock^{9,10}. Here, Salmones Antártica S.A., initiated stockings first in Río Santa María (1982) and then in Río Prat (1983–1989) using University of Washington Hatchery fall-run and the newly established Chilean-based brood stocks (Table 1). Nevertheless, we found little evidence supporting the successful colonization by University of Washington Hatchery fall-run stock. This stock, from the South Puget Sound fall

genetic stock group, was developed at Soos Creek Hatchery (1949–1950s), though exchanges with other populations took place over the years³⁰, mainly within the Puget Sound area. It is puzzling that in our study not a single fish showed any probability of assignment to the University of Washington reference population. Only five fish (17% genetic contribution) from Petrohué River assigned to Soos Creek Hatchery.

Previous studies have inferred ancestral affiliation from the University of Washington Hatchery based on mitochondrial DNA^{23,24}, but owing to the low resolution of those analyses (incomplete baselines; geographically wide-spread haplotypes), those results deserve further scrutiny. Other investigators, using SNPs and a more extensive baseline³¹, found little evidence to support South Puget Sound ancestry, although they still suggested a possible contribution of University of Washington Hatchery stock based on its supposed origin from lower Columbia River lineages, which do appear as major contributors to Patagonian populations^{18,19}. Our results provided no support whatever for that interpretation. The GAPS baseline shows no similarity between Lower Columbia River populations and the most recent sample from University of Washington Hatchery stock—or indeed any other Puget Sound populations. University of Washington Hatchery brood stock is genetically distinct with 80% correct assignment of known-origin fish in a leave-one-out self-assignment test. Moreover, all University of Washington Hatchery fish assign correctly to the South Puget Sound fall run lineage (i.e., no misassignments to West Cascade spring nor to any other population outside South Puget Sound). Further study is obviously needed in the Magellan region where the University of Washington Hatchery brood stock was introduced.

Patagonian Chinook salmon unrelated to the West Cascade spring and fall lineages (hereafter non-WC) accounted for almost half of the fish (45.8%), mostly taken from north of parallel 45°S. The phylogenetic origins of non-WC fish were tracked to 14 diverse and geographically widespread lineages in North America. This seemingly hyperdiverse assemblage is at least partly explained by spurious individual assignments, especially those related to mixed ancestry (consistent with our simulations). However, much of this diversity appears due to accidental escapes from thriving Chilean net-pen salmonid aquaculture operations in the 1990s that imported ova from diverse sources.

Since 1988, Chilean artificial production shifted entirely from ocean ranging to net-pen aquaculture. Although substantial escapes were frequent, they were hard to quantify because they commonly went unreported^{12,32}. Data gathered from insurance companies alone¹² led to an estimate of one million salmon and trout escaping every year, primarily from net-pen aquaculture. Although Chinook salmon farming has always been marginal in Chile relative to other species (<4,000 tons γ^{-1} and <2.5% γ^{-1} of total salmonid production³³), sustained Chinook salmon farming in leaky rearing facilities certainly contributed propagules to the wild^{12,32}. We estimate that with one million escaped salmon and trout per year, adjusting for the contribution of Chinook salmon to annual harvest³³, c. 100,000 farmed sub-adult Chinook salmon might have escaped between the onset of modern salmonid farming (1988) and the date of our sampling (early 2004). During this period, numerous aquaculture companies were established in Chile, propagating Chinook salmon brood stocks to meet increasing demand. Aedo¹³ assembled the fragmentary record of importations and concluded that ova originated at least from Washington, Oregon, Vancouver Island, and New Zealand (of Central Valley ancestry), but specific stocks mostly went undocumented (Table 1).

Although sparse, this information could explain the incidence of Patagonian genotypes related to multiple lineages of the West Coast of U.S.A. Furthermore, the spatial distribution of non-WC fish was concentrated in the north (Toltén and Petrohué), consistent with more intensive propagule pressure of distinct lineages farmed in this region. In fact, during the 1990s, most aquaculture activity (and escapes) was concentrated in the Lakes District region (41–43°S), within 200 km of the Petrohué River. Toltén River (39°S), being a few hundred kilometres further north, might have received colonists dispersing naturally from the Lake District region, escapees from inland hatcheries, or deliberate stockings. Both Toltén and Petrohué are located in well-populated regions easily accessible by paved roads. In sharp contrast, our more remote southern study sites (45–50°S) experienced only incipient aquaculture activity (Aysén River) or none at all (Baker River). Here, accidental escapes or unreported deliberate stockings were less likely, and natural dispersion alone seems the most likely source of founders, especially in the remote and isolated Baker River basin. Other studies have come to similar conclusions by comparing one Chinook salmon population in the Pacific slope of Patagonia (Futaleufú River, 43°S¹⁹) to another population in the Atlantic slope (Santa Cruz River, 50°S¹⁸). The population closer to the Lake District region (Pacific slope) showed affinity to more lineages than the population in the Atlantic slope¹⁹. Thus, escapes from net-pen aquaculture activity in the Lake District region have augmented the genetic diversity initially introduced by open-ocean ranching operations^{19,24}, this study. Similarly, the introgression of cultured rainbow trout genes into pre-existing, naturalized, rainbow trout populations in Patagonia, has also been demonstrated³⁴.

An apparent lack of Central Valley ancestry was similar to the findings of one previous molecular study¹⁸ but different to another¹⁹. It is worth noting that previous studies like ours did not account for potential spurious assignments of mixed-origin fish to unrelated reporting groups. Thus, numerous additional sources were identified as putative contributing lineages. We do not refute the presence of these additional lineages. We simply recognize the limitation we discovered in our genetic mixture analysis regarding mixed-origin fish, and we focus on what we take to be the major lineages that became established at our study sites (those lineages receiving $\geq 10\%$ allocation).

Notable patterns of lineage distribution are emerging at a continental scale. Northern study sites (Toltén and Petrohué) showed unexpectedly high lineage diversity, including substantial estimated contributions from five or more lineages (North Oregon Coast, West Cascade spring, South Puget Sound fall, Whidbey Basin, and Interior Columbia River summer/fall). By contrast, southern study sites (Aysén and Baker) revealed a striking dominance of three closely related lineages from the Lower Columbia River (i.e., West Cascade spring, West Cascade fall and Willamette spring lineages). Similar diversity was also found in the Santa Cruz River draining to the Atlantic Ocean in Argentinian Patagonia¹⁸. A number of possibilities could explain this interesting pattern: (1) As discussed above, higher propagule pressure of distinct lineages escaping from fish farms might have resulted

in higher lineage diversity in northern sites compared to remote southern localities that lacked farms (propagule pressure hypothesis). (2) Most dispersing colonists moved northwards in our study area such that those colonizing Aysén and Baker rivers were derived from stockings in the Magellan region of West Cascade spring and fall ancestry, whereas those colonizing Petrohué and Toltén rivers were derived from fish farms in the Lake District region (propagule+dispersion hypothesis). This pattern would resemble natural ocean migration patterns of Washington and Oregon stocks towards the north in the North Pacific, although it is unclear how the fish would respond to magnetic stimuli of the southern hemisphere. (3) Deliberate stockings initiated the first wave of invasion that eventually reached Aysén and Baker rivers, either from the north or south. Previous establishment of West Cascade spring and fall lineages might have inhibited the expansion of farmed lineages from the north (density-dependent hypothesis). At present, it is hard to gauge the relative importance of these processes, and it is conceivable that each contributed to the observed pattern of lineage distribution.

Complex invasion dynamics and distinct patterns of genetic diversity are likely to have profound evolutionary consequences in the continuing range expansion of Chinook salmon in Patagonia. High genetic diversity, as measured by high allelic and lineage richness¹⁹, this study, high phenotypic plasticity and rapid adaptive evolution observed in Chinook salmon introduced to New Zealand^{5,7}, as well as adaptation to environmental conditions inferred for Patagonian populations^{35,36}, are all features that suggest high adaptive potential of Chinook salmon in Patagonia. Lower-than-expected levels of heterozygosity, if confirmed and maintained over generations, could indicate reproductive isolation between sympatric lineages, which would help maintain lineage identity. Whether this happens is crucial to the evolutionary trajectory of the species in its new range. Lineages could either evolve adapting genetically to local conditions⁷ or they could interbreed and recombine, potentially resulting in new phenotypic variation with unpredictable evolutionary outcomes. Chinook salmon may exert strong ecological impacts in freshwater, estuarine and marine ecosystems³⁵, and different evolutionary outcomes will directly affect these impacts owing to considerable phenotypic variation among lineages.

Materials and Methods

Field collection. Chinook salmon populations were sampled in the fall of 2004 from rivers in the four Andean basins of Toltén, Petrohué, Aysén and Baker (39–48°S), all draining to the South Pacific through Chile (Fig. 1). The results presented here are from 87 fish (after exclusion of 9 fish for which DNA could not be extracted or genotyped). Adult spawners made up 91% of our sample (weight range 2250–21000 g), and 9% were juvenile parr (9–37 g). Collection took place near spawning habitat using a variety of methods including gillnets, hooks, hand-line, and fly-fishing. A fin clip was dissected from each fish and preserved in 96% ethanol for later genetic analyses. A more detailed description of the sampling and sites is available elsewhere³⁵. Fish sample collections were approved by Chilean Subsecretaría de Pesca del Ministerio de Economía, Fomento y Reconstrucción in accordance to their guidelines and regulations (permit number R. Ex. No. 445; 18/02/2004).

Population genetic analysis. DNA was extracted from individual fish tissue samples by using silica-based purification (QIAGEN DNeasy 96 Blood and Tissue Kit, rat tail protocol). Genomic DNA samples were amplified for 13 highly polymorphic microsatellite loci and genotyped by using an Applied Biosystems Incorporated (ABI) 3100 Genetic Analyzer and ABI's Genescan and Genotyper software (v3.7), according to established methods³⁷. We used the GAPS Chinook salmon microsatellite markers following the GAPS consortium conventions for standardization of genotypic data¹⁴. We calculated allelic and genotypic frequencies as well as diversity metrics allelic richness, AR (based on minimum sample size of 12 individuals), and expected heterozygosity, H_S , estimated with FSTAT v2.9.3³⁹. Fixation indices F_{ST} and F_{IS} were represented by Weir and Cockerham's³⁸ multilocus estimators, θ and f , calculated with FSTAT v2.9.3³⁹. F_{ST} was used to evaluate relationships among populations (see below) whereas F_{IS} was used to evaluate departures from Hardy-Weinberg-Castle (HWC) expectations for genotypic proportions within putative populations⁴⁰. We also tested for composite linkage disequilibrium by using the Markov chain Monte Carlo (MCMC) method implemented in the program GENEPOP v4.2⁴¹. Before proceeding with our population genetic analyses, we tested for loci that departed from neutral expectation in order to avoid loci that might bias our parameter estimates and genetic distance estimates (although departures from neutrality would not necessarily bias our genetic mixture analysis). We used the F_{ST} outlier approach⁴² implemented in the LOSITAN software package⁴³. Genetic differences among the South American river basins were analysed in more detail utilizing pairwise F_{ST} values and leave-one-out jackknife self-assignment among introduced populations in order to characterize genetic differences among the introduced Patagonian populations. Because our sample sizes were small, we were especially cautious of non-significant results because we had little power to reject null hypotheses, even when false. However, where we did observe statistical significance, we generally trusted our results as biologically meaningful, despite small sample sizes.

North American baseline dataset. We relied on a baseline dataset of North American reference populations that were selected from the GAPS baseline¹⁴ to represent as closely as possible historical phylogeographic lineages in North American Chinook salmon. Specifically, we used a slight modification of the dataset analysed by Moran *et al.*²¹. We added the following populations because they were identified previously as potential sources of Patagonian Chinook salmon (Table 1): Cowlitz River spring and fall-run, and Green River fall-run from the West Cascade spring and fall-run reporting groups, and University of Washington Hatchery fall-run population from the South Puget Sound fall-run reporting group. Reporting groups are intended to reflect phylogenetic lineages, and here we use either terminology depending on whether the emphasis is on methodology (reporting groups) or phylogeography (lineages). In total, our baseline dataset consisted of 19,973 individuals from 146 populations nested within 46 reporting groups distributed from Central Valley, California (40°N) to the Yukon River in British Columbia (64°N) (Fig. 1). The baseline is widely accepted in the Pacific salmon genetics community as being both

comprehensive and temporally stable¹⁴. Even very small headwater populations show long-term stability of allele frequencies³⁷.

Conditional maximum likelihood mixture analysis. Two different methods were used to explore the likely ancestry of introduced Chinook salmon populations in Chile, conditional maximum likelihood (CML) and model-based clustering (M-BC). As a first approximation of genetic similarity and potential ancestry, we used the Rannala and Mountain²⁰ CML algorithm implemented in the ONCOR software package⁴⁴. That analysis allowed us to estimate the posterior probability distribution for baseline population membership of each individual fish collected in Patagonia. Individual fish were assigned to the population for which they had the maximum *a posteriori* probability of membership. Results were subsequently aggregated by reporting group (genetic lineage), which is the level of inference in essentially all our interpretations. The mean probability across individuals for membership to a particular lineage was taken as an approximation of overall genetic ancestral contribution to the population (normally interpreted as the unbiased estimate of that population's contribution to, e.g., a mixed-stock fishery). Simulation studies helped us evaluate the strength of those assumptions under conditions of mixed ancestry. The CML analysis was used as a first approximation to help narrow the range of contributing lineages to just the major contributors. A reduced set of populations and lineages was then further analysed by M-BC.

Simulated mixed-origin founding. Despite the common use of CML for studies of ancestry of non-native populations^{18,19,26}, its validity in that application has not to our knowledge been demonstrated. As a test of the algorithm's sensitivity to the presence of mixed-origin founding, we simulated a new population derived from large and equal numbers of individuals from two of our North American baseline populations. These were specific populations that appeared to be associated with the introduction to Patagonia (Cowlitz River Hatchery spring-run population in the West Cascade spring-run lineage and Soos Creek Hatchery fall-run population in the South Puget Sound fall-run lineage). That simulated founding involved the creation of an ideal population (at mutation/drift equilibrium) with intermediate allele frequencies to the Cowlitz River and Puget Sound source populations. We then drew 1000 individuals from that new, mixed-origin population, and we estimated the posterior probability distributions for those simulated individuals, the same as we did for the real Patagonian fish. We sought to determine whether the admixed genotypes would assign with high probability to one or the other source population, or split their probability of assignment between the two source populations. These novel genotypes might even assign to other, unrelated populations in the large coast-wide North American baseline (where "assignment" simply reflects the maximum *a posteriori* probability of membership to population or reporting group). We needed to know the extent to which potentially spurious assignments might affect our analysis of the Patagonian populations. As a benchmark for misassignment in the North American reference populations, we conducted a leave-one-out, jackknife analysis for comparison of *a posteriori* probability distributions between simulated mixed-origin fish and real fish, from both North and South America (Supplementary Figure S3).

Model-based clustering. Based on the results of our CML mixture analysis we conducted a more focused model-based cluster analysis (M-BC) of the introduced fish relative to the putative source populations using computer program Structure v2.3.4²². From the North American reference baseline described above, we selected 8,228 fish from 31 populations distributed among seven lineages to include in our analysis. Selected North American lineages for M-BC included the following: Whidbey Basin, S. Puget Sound fall, Interior Columbia Basin summer/fall, Willamette River spring, West Cascade fall, West Cascade spring, and North Oregon Coast. The North American populations and lineages selected for M-BC were those identified in CML as major contributors by virtue of *c.* 10% allocation or higher to any single Patagonian population (North Oregon Coast was included because allocation to that lineage was quite close to our threshold, rounding up to 10%. Including North Oregon Coast as a possible source seemed to be conservative, especially with records of potential introductions from Northern Oregon). The 10% value was selected based on an apparent break in the composition estimates. We also inferred from simulations that allocations of less than 10% might be due at least in part to spurious assignment of mixed-origin fish (see *Methods: Simulated mixed-origin founding*).

For the M-BC analysis, we conducted 110,000 MCMC realizations per chain, discarding the first 10,000 iterations as a burn in. Apparent convergence of diagnostic parameters was observed within the first 10,000 iterations (*i.e.*, α , F , divergence among populations D_{ij} , and the likelihood estimate). We used an admixture model, including location information⁴⁵, with allele frequencies correlated among populations⁴⁶. We assumed population specific F_{ST} values, and updated allele frequencies by using baseline individuals only, thus treating Patagonian samples as having unknown origins. Multiple MCMC chains (10) were constructed for each value of K (number of ancestral clusters). We modelled values of K from two to nine for the North American populations. The most appropriate K value was selected based on likelihood and ΔK ⁴⁷, as well as optimal discrimination of putative North American source populations. Structure output was processed with the STRUCTURE HARVESTER computer program⁴⁸, and plotted in the R computing environment version 3.3.1⁴⁹.

References

- Rius, M. & Darling, J. A. How important is intraspecific genetic admixture to the success of colonising populations? *Trends Ecol. Evol.* **29**, 233–242 (2014).
- Asif, J. H. & Krug, P. J. Lineage distribution and barriers to gene flow among populations of the globally invasive marine mussel *Musculista senhousia*. *Biol. Invasions* **14**, 1431–1444 (2012).
- Crawford, S. & Muir, A. Global introductions of salmon and trout in the genus *Oncorhynchus*: 1870–2007. *Rev. Fish Biol. Fish.* **18**, 313–344 (2007).
- Peck, J. W., Jones, T. S., MacCallum, W. R. & Schram, S. T. Contribution of hatchery-reared fish to Chinook salmon populations and sport fisheries in Lake Superior. *North Am. J. Fish. Manag.* **19**, 155–164 (1999).

5. Quinn, T. P., Kinnison, M. T. & Unwin, M. J. Evolution of chinook salmon (*Oncorhynchus tshawytscha*) populations in New Zealand: pattern, rate, and process. *Genetica* **112–113**, 493–513 (2001).
6. Kinnison, M. T., Unwin, M. J. & Quinn, T. P. Migratory costs and contemporary evolution of reproductive allocation in male chinook salmon. *J. Evol. Biol.* **16**, 1257–1269 (2003).
7. Kinnison, M. T., Unwin, M. J. & Quinn, T. P. Eco-evolutionary vs. habitat contributions to invasion in salmon: experimental evaluation in the wild. *Mol. Ecol.* **17**, 405–414 (2008).
8. Lindbergh, J., Noble, R. & Blackburn, K. *First returns of Pacific salmon to Chile*. 8pp. (C.M. 1981/F:27 International Council for The Exploration of the Sea, 1981).
9. Lindbergh, J. M. & Brown, P. *Continuing experiments on salmon ocean ranching in southern Chile*. (C.M. 1982/M:21 International Council for The Exploration of the Sea, 1982).
10. Donaldson, L. R. & Joyner, T. The salmonid fishes as a natural livestock. *Sci. Am.* **249**, 50–58 (1983).
11. Méndez, R. & Munita, C. *La salmonicultura en Chile*. (Fundación Chile, 1989).
12. Sepúlveda, M., Arismendi, I., Soto, D., Jara, F. & Farias, F. Escaped farmed salmon and trout in Chile: incidence, impacts, and the need for an ecosystem view. *Aquac. Environ. Interact.* **4**, 273–283 (2013).
13. Aedo, E. In *Evaluación cuantitativa del estado trófico de salmonidos de vida libre en el fiordo Aysén, XI región. Informe final proyecto FIP200830*. (eds Niklitschek, E. & Toledo, P.) 120–135 (Universidad Austral de Chile - Subsecretaría de Pesca, 2011).
14. Seeb, L. W. *et al.* Development of a standardized DNA database for Chinook salmon. *Fisheries* **32**, 540–552 (2007).
15. Anderson, E. C., Waples, R. S. & Kalinowski, S. T. An improved method for predicting the accuracy of genetic stock identification. *Can. J. Fish. Aquat. Sci.* **65**, 1475–1486 (2008).
16. Hess, J. E., Whiteaker, J. M., Fryer, J. K. & Narum, S. R. Monitoring stock-specific abundance, run timing, and straying of Chinook salmon in the Columbia River using genetic stock identification (GSI). *North Am. J. Fish. Manag.* **34**, 184–201 (2014).
17. Moran, P., Bromaghin, J. F. & Masuda, M. Use of genetic data to infer population-specific ecological and phenotypic traits from mixed aggregations. *PLoS ONE* **9**, e98470 (2014).
18. Ciancio, J. E., Rossi, C. R., Pascual, M., Anderson, E. & Garza, J. C. The invasion of an Atlantic Ocean river basin in Patagonia by Chinook salmon: new insights from SNPs. *Biol. Invasions* **17**, 2989–2998 (2015).
19. Di Prinzio, C. Y., Rossi, C. R., Ciancio, J., Garza, J. C. & Casaux, R. Disentangling the contributions of ocean ranching and net-pen aquaculture in the successful establishment of Chinook salmon in a Patagonian basin. *Environ. Biol. Fishes* **98**, 1987–1997 (2015).
20. Rannala, B. & Mountain, J. L. Detecting immigration by using multilocus genotypes. *Proc. Natl. Acad. Sci.* **94**, 9197–9201 (1997).
21. Moran, P. *et al.* Divergent life-history races do not represent Chinook salmon coast-wide: the importance of scale in Quaternary biogeography. *Can. J. Fish. Aquat. Sci.* **70**, 415–435 (2013).
22. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
23. Becker, L. A., Pascual, M. A. & Basso, N. G. Colonization of the Southern Patagonia ocean by exotic Chinook salmon. *Conserv. Biol.* **21**, 1347–1352 (2007).
24. Riva-Rossi, C. M. *et al.* The invasion of Patagonia by Chinook salmon (*Oncorhynchus tshawytscha*): inferences from mitochondrial DNA patterns. *Genetica* **140**, 439–453 (2012).
25. Narum, S. R. *et al.* Genomic patterns of diversity and divergence of two introduced salmonid species in Patagonia, South America. *Evol. Appl.* **10**, 311–416 (2017).
26. Signorile, A. L. *et al.* Using DNA profiling to investigate human-mediated translocations of an invasive species. *Biol. Conserv.* **195**, 97–105 (2016).
27. Gharrett, A. J. & Smoker, W. W. Two generations of hybrids between even- and odd-year pink salmon (*Oncorhynchus gorbuscha*): a test for outbreeding depression? *Can. J. Fish. Aquat. Sci.* **48**, 1744–1749 (1991).
28. Lindbergh, J. M. A successful transplant of Pacific salmon to Chile. *Proc. Gulf Caribb. Fish. Inst.* **34**, 81–87 (1982).
29. Lindbergh, J. M., Noble, R. E. & Blackburn, K. M. Salmon ranching in Chile: the private sector. *ICLARM newsletter* (1981).
30. Quinn, T. P., Peterson, J. A., Gallucci, V. F., Hershberger, W. K. & Brannon, E. L. Artificial selection and environmental change: countervailing factors affecting the timing of spawning by coho and Chinook salmon. *Trans. Am. Fish. Soc.* **131**, 591–598 (2002).
31. Clemente, A. J., Crandall, E. D., Garza, J. C. & Anderson, E. C. Evaluation of a single nucleotide polymorphism baseline for genetic stock identification of Chinook Salmon (*Oncorhynchus tshawytscha*) in the California Current large marine ecosystem. *Fish. Bull.* **112**, 112–130 (2014).
32. Soto, D., Jara, F. & Moreno, C. Escaped salmon in the inner seas, southern Chile: facing ecological and social conflicts. *Ecol. Appl.* **11**, 1750–1762 (2001).
33. FAO. Fishery and aquaculture statistics. Global aquaculture production 1950–2012 (Fishstat) In: FAO Fisheries and Aquaculture Department [online or CD-ROM]. Rome. Updated 2014. <http://www.fao.org/fishery/statistics/software/fishstatj/en> (2014).
34. Consuegra, S., Phillips, N., Gajardo, G. & de Leaniz, C. G. Winning the invasion roulette: escapes from fish farms increase admixture and facilitate establishment of non-native rainbow trout. *Evol. Appl.* **4**, 660–671 (2011).
35. Correa, C. & Gross, M. R. Chinook salmon invade southern South America. *Biol. Invasions* **10**, 615–639 (2008).
36. Araya, M., Niklitschek, E. J., Secor, D. H. & Piccoli, P. M. Partial migration in introduced wild Chinook salmon (*Oncorhynchus tshawytscha*) of southern Chile. *Estuar. Coast. Shelf Sci.* **149**, 87–95 (2014).
37. Van Doornik, D. M., Waples, R. S., Baird, M. C., Moran, P. & Berntson, E. A. Genetic monitoring reveals genetic stability within and among threatened Chinook salmon populations in the Salmon River, Idaho. *North Am. J. Fish. Manag.* **31**, 96–105 (2011).
38. Weir, B. S. & Cockerham, C. C. Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358 (1984).
39. Goudet, J. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available at: <http://www.unil.ch/izea/software/fstat.html> (2001).
40. Castle, W. E. The laws of heredity of Galton and Mendel: and some laws governing race improvement by selection. *Proc. Am. Acad. Arts Sci.* **35**, 233–242 (1903).
41. Raymond, M. & Rousset, F. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* **86**, 248–249 (1995).
42. Beaumont, M. A. & Nichols, R. A. Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond. B Biol. Sci.* **263**, 1619–1626 (1996).
43. Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A. & Luikart, G. LOSITAN: A workbench to detect molecular adaptation based on a F_{st} -outlier method. *BMC Bioinformatics* **9**, 323 (2008).
44. Kalinowski, S., Manlove, K. & Taper, M. *ONCOR: a computer program for genetic stock identification*. (Department of Ecology, Montana State University, USA., 2007).
45. Hubisz, M. J., Falush, D., Stephens, M. & Pritchard, J. K. Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* **9**, 1322–1332 (2009).
46. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587 (2003).
47. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
48. Earl, D. A. & vonHoldt, B. M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361 (2012).

49. R Development Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2016).
50. Gorny, A. & Carter, R. *World Data Bank II: General Users Guide*. (Central Intelligence Agency, 1987).
51. Becker, R. A., Wilks, A. R. & Brownrigg, R. mapdata: Extra map databases. R package version 2.2–3, Available at: <https://CRAN.R-project.org/package=mapdata> (2014).
52. Augerot, X. & Foley, D. N. *Atlas of Pacific salmon*. (University of California Press; State of the Salmon, 2005).
53. Fundación, Chile. *El libro del salmón*. (Fundación Chile, 1990).
54. Snyder, B. P. *Supplemental report on inland fresh water resources of central Chile*. 35pp. (United Nations Development Programme, 1971).
55. Basulto, S. *El largo viaje de los salmones. Una crónica olvidada. Propagación y cultivo de especies acuáticas en Chile*. (Maval Ltda., 2003).
56. Sakai, M. *Final report of aquaculture project in Chile*. (Japan International Cooperation Agency (JICA), 1989).
57. United Nations. Transfer of technology for successful integration into the global economy; A case study of the salmon industry in Chile. In *United Nations conference on trade and development* (United Nations, 2006).
58. Del Real, A. Antecedentes sobre el cultivo de salmón Chinook (*Oncorhynchus tshawytscha* W.) en la fase de agua dulce proveniente de reproductores retornantes en la piscicultura experimental Lago Rupanco. (Universidad de Los Lagos, Departamento de Acuicultura y Recursos Acuáticos, 1993).

Acknowledgements

We especially thank Mart Gross for his pivotal support at the initial stages of this study. Comments by Mart Gross, Jim Myers, David Teel, Robin Waples, and Jeff Hard helped us improve earlier drafts of this article. Jon Wittouck kindly provided historical records of fish transfers from the University of Washington Hatchery. Cristián Correa was funded by the Government of Chile through grants CONICYT-PAI N°82130009, and FONDECYT-Iniciación en la Investigación N°11150990.

Author Contributions

C.C. conceived the study, and collected samples from Patagonia. P.M. conducted the genetic analyses. C.C. prepared all figures and tables, and both authors contributed to the interpretation of results and writing of the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-017-14465-y>.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017