The role of historical contingency in shaping geographic pattern of deleterious mutation load in a broadly distributed Pacific Salmon

Quentin Rougemont[1*], Jean-Sébastien Moore[1], Thibault Leroy[2,3], Eric Normandeau[1], Eric B. Rondeau[4;5], Ruth E. Withler[6], Donald M. Van Doornik[7], Penelope A. Crane[8], Kerry A. Naish[9], John Carlos Garza[10] Terry D. Beacham[6], Ben F. Koop[4,5], Louis Bernatchez[1]

[1] Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Québec, Canada QC G1V OA6
[2] ISEM, Univ. Montpellier. CNRS, EPHE, IRD, Montpellier, France
[3] Department of Botany & Biodiversity Research, University of Vienna, Vienna, Austria
[4] Centre for Biomedical Research, University of Victoria, Victoria, BC, Canada V8P 5C2
[5] Department of Biology, University of Victoria, Victoria, BC, Canada V8P 5C2
[6] Department of Fisheries and Ocean, Pacific Biological Station, 3190 Hammond Bay Road, Nanaimo, British Columbia, Canada V9R 5K6
[7] National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Northwest Fisheries Science Center, Manchester Research Station, 7305 Beach Drive East, Port Orchard, Washington 98366, USA
[8] Conservation Genetics Laboratory, U.S. Fish and Wildlife Service, 1011 E. Tudor Road, Anchorage, Alaska 99503
[9] School of Aquatic and Fishery Sciences, University of Washington, 1122 NE Boat Street, Box 355020, Seattle, WA 98195 5020, USA
[10] Fisheries Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries Service and Institute of Marine Sciences, University of California–Santa Cruz, 110 McAllister Way, Santa Cruz, California 95060

*Corresponding author e-mail: quentinrougemont@orange.fr (QR)

Abstract

A thorough reconstruction of historical processes is essential for a comprehensive understanding of processes shaping patterns of genetic diversity. Indeed, past and current conditions influencing effective population size have important evolutionary implications for the efficacy of selection, increased accumulation of deleterious mutations, and loss of adaptive potential. Here, we gather extensive genome-wide data that represent the extant diversity of the Coho salmon (*Oncorhynchus kisutch*) to address two objectives. We demonstrate that a single glacial refugium is the source of most of the present-day genetic diversity, with detectable inputs from a putative secondary micro-refugium. We found statistical support for a scenario whereby ancestral populations located south of the ice sheets expanded in postglacial time, swamping out most of the diversity from other putative micro-refugia. Demographic inferences revealed that genetic diversity was also affected by linked selection in large parts of the genome. Moreover, we demonstrate that the recent demographic history of this species generated regional differences in the load of deleterious mutations among populations, a finding that mirrors recent results from human populations and provides increased support for models of expansion load. We propose that insights from these historical inferences should be better integrated in conservation planning of wild organisms, which currently focuses largely on neutral genetic diversity and local adaptation, with the role of potentially maladaptive variation being generally ignored.

Introduction

Historical climate variation has had a major influence on the current distribution of species genetic diversity [1]. The Pleistocene glaciations, in particular, resulted in major contractions in the geographical distributions of many species into refugia that persisted in unglaciated areas [2–4]. Postglacial range expansions often led to contact between ancestral populations previously segregated in different refugia [2,4]. The effects of long-term climate change combined with recent human-induced population declines in wild animal populations can foster genetic changes including a loss of genetic diversity, increased inbreeding, increased load of deleterious mutations, and a loss of local adaptation [5].

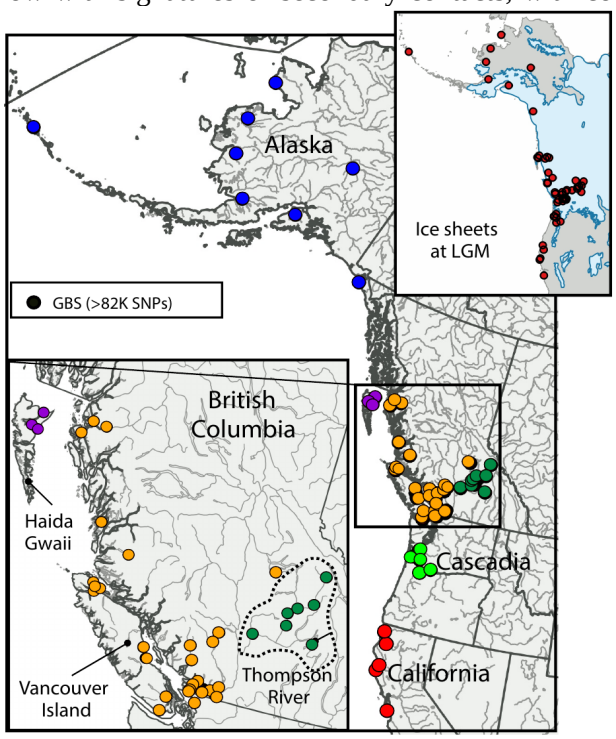
In this context, understanding the impacts of historical climatic oscillations on the demographic history of a given species is crucial. Namely, it is important to understand how historical demographic events influenced the present-day geographic distribution of the within-species genetic diversity across its range. Here it is noteworthy that while the majority of studies on wild populations have focused on patterns of neutral and potentially adaptive genetic diversity, the importance of potentially maladaptive variation (e.g. patterns of deleterious mutation load) has generally been ignored. By disentangling past and current drivers of range-wide genomic diversity, this information can inform management and conservation decisions [6]. Beyond conservation implications, such context provides a unique opportunity to address outstanding questions in evolutionary biology: How is the efficacy of selection affected by historical processes that resulted in population expansion or bottleneck [7,8]? What are the demographic conditions required to generate substantial differences in deleterious load among populations [7–9]?

A major challenge in understanding drivers of genome-wide patterns of diversity is that different demographic processes can lead to similar contemporary genomic footprints [10]. As populations diverge, they accumulate genetic incompatibilities forming barriers to gene flow [11], while the rest of the genome may continue to be freely exchanged. Consequently, the genomic landscape of divergence is expected to vary, with greater differences along the genome at genomic barriers as compared to genomic regions exhibiting ongoing gene flow. However, similar patterns of heterogeneous genome-wide divergence can also be due to genetic hitchhiking of neutral alleles linked to selective sweeps [12,13] or to background selection (BGS;[12,14,15]]). These two processes, collectively referred to as linked selection, result in a reduction in nucleotide diversity in the vicinity of the sites targeted by positive or purifying selection. Due to its diversity-reducing effect, linked selection can be modelled as a local reduction of effective population size (*Ne)* [15]. It is now recognized that neglecting linked selection can bias demographic inferences [16,17] or lead to false adaptive interpretations [18]. Inference methods that incorporate variation in local effective population size and migration rate can help better understand how demography unfolds through time.

An understanding of historical demography is also essential for a sound interpretation of patterns of deleterious mutation load observed among contemporary populations [7,9]. Population bottlenecks are predicted to reduce potential for local adaptation, but also to reduce standing genetic variation and the efficacy of selection [8,19]. In turn, a reduced efficacy of purifying selection leads to an increase in the number of deleterious variants segregating in a population, thus reducing its fitness, and can result in maladaptation and/or reduced adaptive potential in bottlenecked populations. From a conservation standpoint, populations harbouring an elevated number of deleterious variants might need to be monitored more closely to minimize extinction risks [20].

Combining population genomics data with demographic modelling represents a powerful strategy to test alternative hypotheses about historical drivers of existing genomic diversity. Previous studies employing a similar approach have focused mostly on species with a narrow geographic range, such as those inhabiting small islands [21,22], on the verge of extinction [23–26], or that experienced a strong bottleneck [27]. Few studies, however, have used demographic modelling to understand how historical processes have shaped the geographical patterns in the distribution of genomic diversity in a more broadly distributed species. An exception to this observation is the vast literature on demographic reconstructions of human populations. Long-lasting debates in this literature regarding the role of demography in generating mutation load differences among populations [9,28,29] could benefit from studies of species displaying similarly complex demographic histories and broad geographic distributions.

Salmonid fishes are economically important species that have suffered recent demographic declines [30,31]. This is particularly the case for Coho salmon (*Oncorhynchus kisutch*), one of the five anadromous species of Pacific salmon that supports important recreational and indigenous subsistence fisheries, and which has suffered dramatic population declines (> 90%) over the last three decades in parts of its range [32]. A previous study investigated the range-wide population structure and demographic history of the species and found a cline of decreasing diversity from south to north, as well as some endemic diversity in small putative refugia [33] (see also [34]). This study indicated that Coho salmon may have survived the last glacial maximum (LGM, i.e. the Fraser Glaciation in British Columbia, and the McConnell/McCauley Glaciation in Yukon and Alaska; 23 to 18 Ky ago) in unglaciated areas of Haida Gwaii and Beringia in addition to areas south of the ice sheets. This study, however, predates the genomic era and could not eliminate alternative hypotheses regarding the origin and number of glacial refugia during the LGM. In North America, the species is currently distributed from California to Alaska [35]. Unglaciated areas that could potentially serve as glacial refugia persisted both north (e.g. the Beringian refugium in Alaska, the Yukon Territory of Canada and areas of Asia and the Bering Land Bridge) and south (e.g. all of the deglaciated area south of British Columbia, Canada) of the ice sheets [35–37]. Other unglaciated areas (e.g. Haida Gwaii in British Columbia) could also have been micro-refugia [38,39]. In this context, distinct demographic scenarios can be tested. Under a first scenario whereby populations expanded north from a single southern refugium, we predict: *i*) a latitudinal decrease in genetic diversity from south to north along with a pattern of isolation-by-distance (IBD), and *ii*) ancestral populations located in areas south of the ice sheets. Under a second scenario, populations expanded south from a single northern refugium, and we predict the opposite geographic pattern. The third scenario corresponds to the survival of populations in different refugia where we predict: *i)* the existence of clearly distinct genetic clusters, and *ii)* postglacial gene flow with signatures of secondary contacts, with contact zones displaying higher genetic diversity through postglacial admixture between different genomic backgrounds.



In order to test these alternative scenarios, we generated genome-wide data from nearly 2,000 Coho salmon from California to Alaska, one of the most extensive genomic datasets for a non-model vertebrate species to date. First, to resolve the species demographic history, we used a modelling approach that accounts both for barriers to gene flow affecting migration locally, and for linked selection affecting the rate of drift. Finally, we hypothesized that demographic history shaped the pattern of deleterious mutation load, both within and among populations. We hypothesized that postglacial re-colonisation influenced levels of standing genetic variation and favoured the accumulation of deleterious mutations at the expansion front. In these conditions, we predicted that neutral genetic diversity should decrease as a function of the distance from the ancestral populations.

Figure 1) Sampling design
Sampling locations of 58 Coho salmon populations distributed across the species' North American range of distribution. Each dot represents a sampling location. Inset: Map showing the extent of ice-sheet during the Last Glacial Maximum 13 Kya. Data modified from [92].

Results & Discussion

Latitudinal gradient in nucleotide diversity

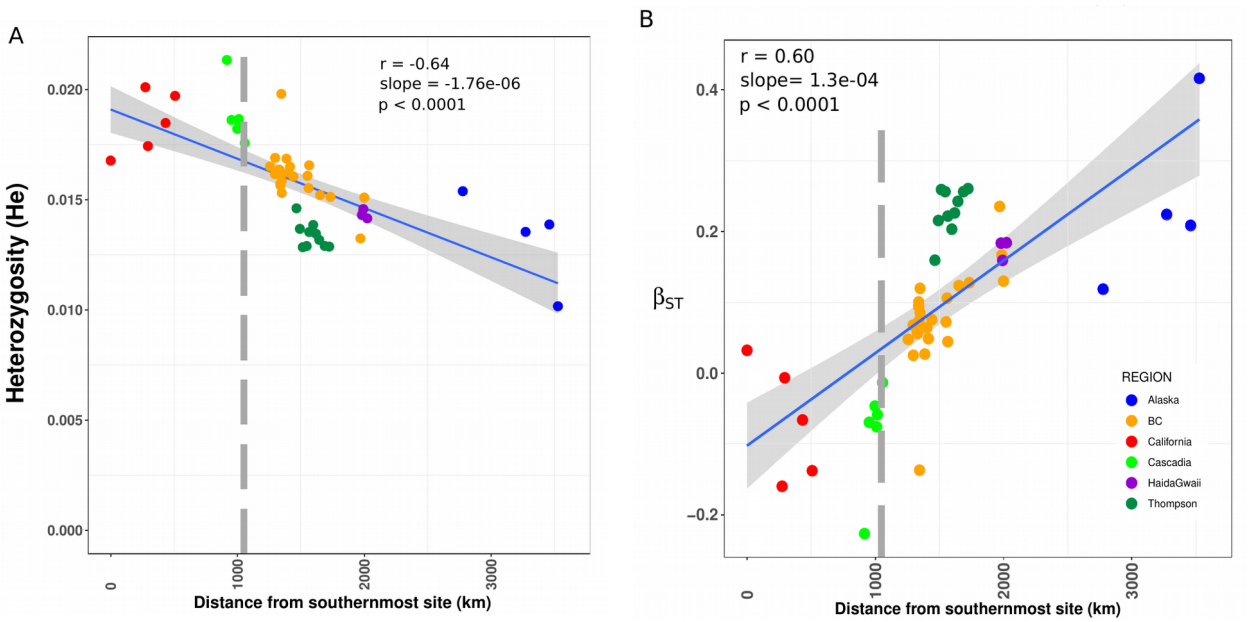


Figure 2) Gradients of genetic diversity and differentiation
A) Linear relationship between expected heterozygosity and distance from the southernmost population located in California. B) Linear increase in genetic differentiation as measured by β_{ST} as a function of the distance from the southernmost population located in California. Negative values indicate the most likely ancestral population. The relationship in A and B was tested using linear models. The grey vertical bar in panel A and B) indicates the approximate location of the southern limit of the ice-sheet at the end of the last glacial maximum.

To investigate the distribution of genetic diversity and population differentiation, we sampled 1,957 Coho salmon from 58 sampling locations throughout their entire North American range from California to Alaska (mean n = 34 fish per location, **Fig. 1a, S1 Table**). Samples were genotyped using a genotyping-by-sequencing (GBS) method that generated 82,772 high quality filtered single nucleotide polymorphisms (SNPs). Consistent with earlier work based on non-genomic data, our analyses indicate that the southern-most populations are the most ancestral and suggest that a single southern refugium explains most of the observed patterns in the distribution of genetic diversity. Under a single refugial expansion scenario, one would expect genetic diversity to be highest in the ancestral refugium and a linear decrease as a function of the distance from the source, as documented in humans [40]. To test this prediction, we plotted the distribution of genetic diversity as a function of the distance to the southernmost sample in our data. As expected, levels of genetic diversity (observed and expected heterozygosity, π_{SNP}) were highest in formerly deglaciated refugial areas in the south (California, Cascadia, **Fig. 2a, Fig 1b**) and decreased as a function of distance from the southernmost site up to Alaska ($r = 0.64$, $p < 0.0001$**, Fig. 2a, S1 Fig, S2 Fig.**). Two populations however deviated from this linear pattern (Clackamas R in Cascadia and Upper Pitt R. in BC) and displayed higher genetic diversity than the Southern samples. We hypothesized that these two rivers may have undergone recent admixture from differentiated genetic clusters, locally increasing genetic diversity. Recent admixture between populations from several refugia can produce increased genetic variation [41], which could explain why most ancestral populations (relict populations from the refugia) are not systematically those displaying the highest genetic diversity. The Thompson River watershed (Thompson R. hereafter) in southern British Columbia was an exception to this latitudinal pattern and displayed the lowest average level of regional genetic diversity of all sampling locations, which we hypothesized resulted from bottlenecks (see below). The remaining samples from British Columbia were intermediate in genetic diversity.

Present-day southern populations are the most ancestral

To help discriminate a single versus two population refugia, we aimed to identify the most ancestral populations among all samples. Under a single scenario refugium, the ancestral populations should be located preferentially in one area (e.g., South or North), whereas under a two-population scenario, we expected ancestral populations to occur in different areas located at both edges of the species' distribution range (Beringia covering most of Alaska and the Yukon being known as an important glacial refugium for many species (reviewed in [2-3]). To discriminate between these alternative scenarios, we used two simple summary statistics, namely (1) the distribution of singletons among populations, and (2) the βst index recently developed in [42]. It was previously shown that older populations are expected to have accumulated a higher density of singletons [43]. Therefore, counting the number of singletons by sampling site and averaging by regional groups helped identify older populations. Cascadian samples contained the highest number of singletons, with a mean of 1,263 singletons per site, consistent with the hypothesis of a refugium being present in this region. Californian samples had the fewest number of singletons (n_{MEAN} = 55) while Alaska harboured an intermediate density (n_{MEAN} = 966). We hypothesized that such reduced number of singletons in a few Californian samples reflects recent and strong population declines in some of these populations [44,45].

Next, we used the β_{ST} coefficient to identify ancestral populations [42]. Unlike F_{ST} estimates [46], β_{ST} can account for the non-independence among populations and negative values are indicative of ancestral populations [42]. More specifically, the β_{ST} coefficient will be negative when a population has accumulated many private alleles at low to intermediate frequencies and the intensity of this value is dependent on the number of sampled populations [42]. As for singletons, older populations and populations of larger effective size have more time to accumulate rare alleles, resulting in negative values, as illustrated in human samples from Africa (see details in 42). Here, the β_{ST} indicated that the likely ancestral populations were in previously unglaciated areas corresponding to Cascadia (n= 5 localities), California (n = 3 localities) as well as one site from southern British Columbia **(Fig. 2b, S2 Table**). A linear decrease in β_{ST} with distance from the southernmost site was observed (r = 0.60, slope = 1.03e-04, p <0.0001) as expected under isolation-by-distance (IBD). Support for this IBD pattern was also observed using F_{ST} (**S3 Fig**, r = 0.66, slope= 4e-05, p < 0.0001) as well as Mantel tests (r = 0.64, p < 0.0001; and r = 0.72, p < 0.0001 when removing Thompson R. populations). As for genetic diversity, the river CLA in Cascadia and UPT in British Columbia departed from the pattern and displayed the most negative coefficients (**Fig. 2b**). Finally, average pairwise F_{ST} across all populations was 0.095 and varied from 0.002 to 0.334 (**S4 Fig**), typical of anadromous species connected by gene flow [47].

To further confirm the number of putative refugia and better understand the extent of divergence among populations, we performed a principal components analysis (PCA). PCA can be used to summarize the distribution of genetic diversity with different expectations under discrete models of population genetic structure (as expected with different refugia) versus IBD under a one- or two-dimensional habitat model. Assuming K clusters, we expect that the top K-1 PCA axes should discriminate all these clusters [48]. On the contrary, under models of IBD, the top two axes should be correlated with geographical axes such as the longitude and latitude of populations' locations [49]. Consistent with this latter hypothesis, we observed that the first axis was strongly correlated with latitude (r = 0.77, *p = 2e-16*), whereas the second axis was strongly correlated with longitude (r = -0.54, p = 2e-16). The result of the PCA (**Fig. 3a, see also S5 Fig**) is therefore consistent with a South to North spatial pattern of genetic structure, with the most divergent samples found in California, and a second geographical gradient from East to West, mainly reflecting the pronounced differentiation of the Thompson R. populations. Along these axes, populations followed an IBD pattern. These results were also supported by an MDS analysis (**S6 Fig**). A model-based analysis of population structure implemented in LEA [50] failed to reveal a clear number of distinct populations (K value). Instead, K values ranging from 30 to 60 all fit the data well (**S7 Fig**), providing little insight on the number of ancestral clusters in the data due most likely to the confounding effect of IBD [51]. While all these summary statistics lend support to the hypothesis of an expansion from a single ancestral refugium, a

formal testing of alternative scenarios is required as the signature of secondary contact can be transient and both models of secondary contact and stepping stone can converge to the same IBD signal and apparent demographic equilibrium [10].

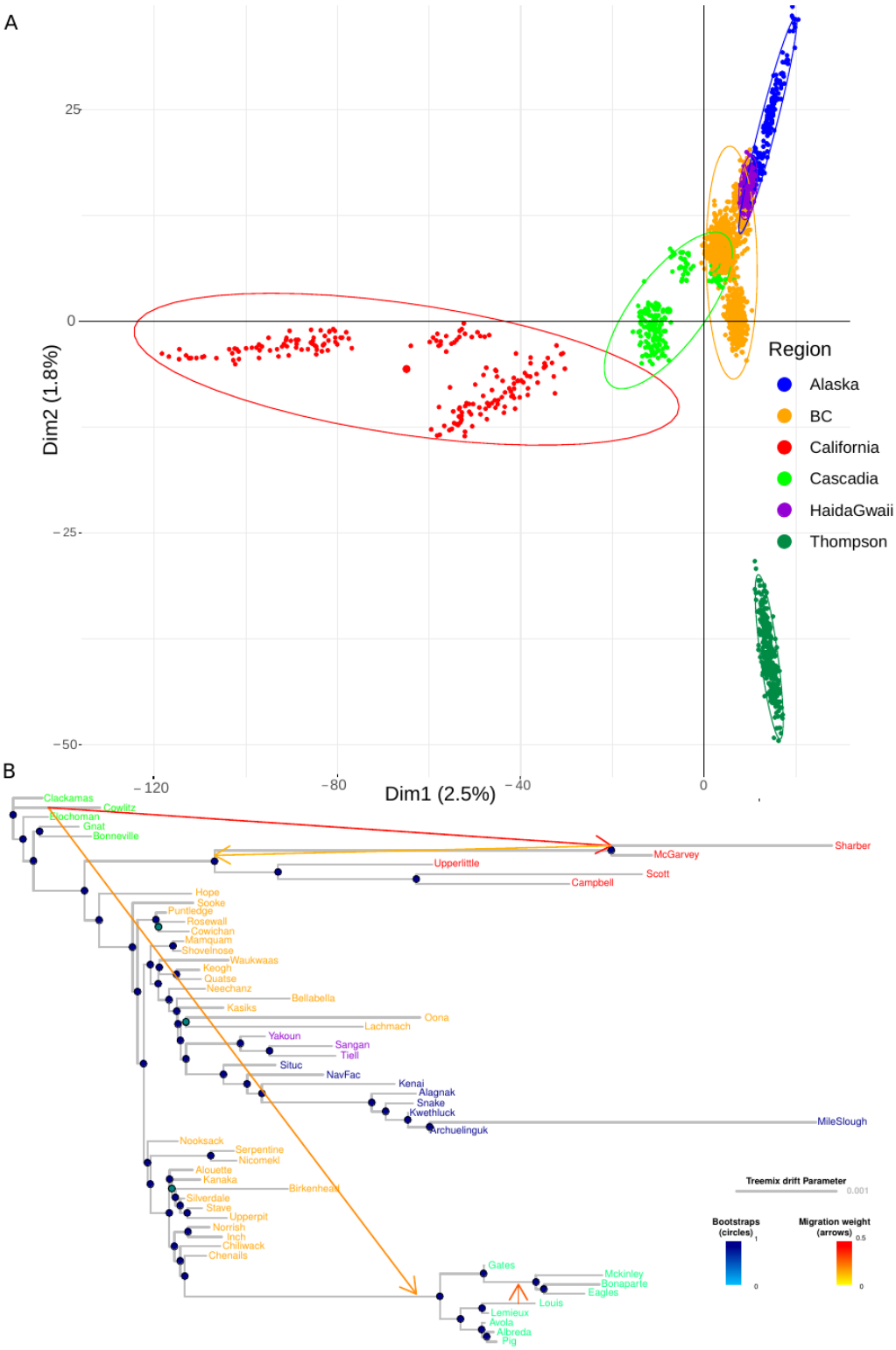


Figure 3) Genetic structure and gene flow

A) Principal Component Analysis (PCA) summarising population genetic structure among 1,957 individuals based on the principal component axis 1 and axis 2. Each point represents an individual and the colours represent the major regional groups. The larger point represents the barycentre of each group. B) Inference of population splits and mixture by Treemix with 4 statistically significant migration events ($p < 0.0001$). Bootstrap supports and migration weights are indicated according to the colour scale.

Evidence for gene flow among Coho salmon populations

In order to infer both population splits and past or ongoing gene flow events irrespective of the phylogeny, we used Treemix [52] (**Fig. 3b**). A first tree assuming no migration (i.e. drift only) explained 98.1% of the variance observed. Including up to four significant migration events ($p < 0.0001$) explained over 99.1% of variance, and the proportion of explained variance subsequently plateaued when adding additional migration events (**S8 Fig**). Populations from Cascadia occupied basal positions. California populations displayed pronounced genetic drift, corroborating the high divergence observed in the PCA. Populations from Alaska (MSL River) and Thompson R. also displayed higher genetic drift, in line with evidence based on analyses of genetic diversity and genetic structure. The two most supported migration events occurred from Cascadia south to California and north to the Thompson R. We note that populations followed the south to north arrangement, with the samples from Cascadia displaying less drift than those located further north.

Colonisation wave from a single major southern refugium

In order to formally assess the occurrence of one or more refugial origins for contemporary populations, we performed explicit model-based inferences of population divergence scenarios using ∂a∂i [53] and Fastsimcoal [54]. ∂a∂i was used to perform pair-wise population comparisons in order to 1) consider the confounding effects of linked selection and barriers to gene flow, and 2) calibrate a multi-population model to be used in Fastsimcoal. Under a model of a single refugium, either northern or southern, we expect to find support for a model of recent or ongoing divergence with gene flow (isolation with migration or IM) when comparing neighboring pairs. Following this hypothesis, IM models should be the best supported because the divergence time between the different populations should be recent, thus resulting in recent inferred divergence times. In addition, these populations are expected to still be connected by ongoing gene flow. Under a scenario with two ancestral refugia, we predict that comparing Alaskan populations (i.e. assuming that Coho survived in the Beringian refugium) to other genetic groups should reveal signals of secondary contact (SC). In addition to this evidence for postglacial gene flow, one can expect to infer a long divergence time between Alaskan populations and the rest of the populations. Finally we completed our inferences with a null model of strict isolation (SI) where populations would never have exchanged gene and a model of ancient migration (AM) where populations exchanged gene at the early stage of divergence only.

In order to reduce the number of pairwise comparisons and thus reduce the possibility of spurious results, we pooled individuals into higher order groups based on the PCA results (**Fig. 2a** and **S9 Fig**). The following six groups were constructed (named according to geographic labels): 1) California group 1, 2) California group 2, 3) Thompson 4) Cascadia 5) British Columbia and 6) Alaska. To avoid over-fitting of complex models, we performed model selection hierarchically. First, we compared all models (AM, IM and SC, SI) represented in **S10 Fig**. Our models account for the confounding effects of selection at linked sites and that of the accumulation of local barriers to gene flow in the genome, two potential confounding factors that can lead to erroneous interpretations when they are not explicitly taken into account [12,17]. We then compared Alaskan samples to all other groups (n = 4 pairwise comparisons). Then we compared the Thompson R. populations, which display an unusual pattern of reduced diversity and divergence from all other groups (n = 5 pairwise comparisons), resulting in a total of 9 pairwise comparisons.

Our inferences provided considerable support for an expansion from a single major refugium. Indeed, our analyses support models of divergence with ongoing gene flow between Alaska and all the other groups in three out of the four pairwise comparisons involving Alaska (**S11 Fig** for model fit and residuals and **S3 Table**; ΔAIC > 10; with the notable exception of Thompson, see below). These results are inconsistent with a Beringian refugium, or at least a contribution of these northern populations to the present-day within-species variation. One out of the 4 comparisons ("California 1" versus "Alaska") provided support for secondary contact. We hypothesize that this might be a false positive that highlights the well-known difficulty of distinguishing models of secondary divergence from primary divergence. Indeed, these two classes of models can quickly converge to the same state as observed here, namely large-scale isolation

by distance and apparent equilibrium [10]. Interestingly, when any group is compared to Thompson R., all inferences find unambiguous support for postglacial secondary contact (**S3 Table**), suggesting that the Thompson R. might have been a second refugium or, alternatively, that the populations from this area were established from another refugium. The relative isolation of the Thompson R. would therefore be consistent with the lower genetic diversity observed in each local population within this area. It is noteworthy that in all models with support for secondary contact, the relative time of secondary contact represented on average 10% of the total divergence time. This proportion provided ideal statistical power to discriminate among models of divergence with continuous gene flow versus allopatric divergence followed by gene flow. Indeed, the signal of allopatric divergence is expected to be lost if gene flow was initiated too long in the past [17,55].

Statistical support for ancestral expansion followed by recent demographic declines across all populations

We added a second layer of complexity by incorporating changes in population size in the models namely, expansions or bottlenecks. We hypothesized that these changes are likely to arise following postglacial expansion or through serial founder events during the recolonization process. We chose the best models identified through weighted AIC (**S3 Table**) and performed a round of parameter estimation. A salient result was that all populations exhibited a signature of expansion from the ancestral refugial population, with the effective population size at the divergence time being approximately 12 times higher than the ancestral population (**Fig. 4**). Then all populations declined to approximately 10-12% of their initial size (**Fig. 4**). More precisely, assuming a generation time of 3 years [56] and a mutation rate of $8e^{-9}$ bp/generation [57], our analysis inferred remarkably similar ancestral effective population sizes across all our pairwise comparisons (approximately 15,000 **Fig. 4, S4 Table**). Then we recovered a signal of expansion, with stronger expansion signals inferred in Cascadia and Alaska (**Fig. 4**). However, we observed some uncertainty in these parameter estimates (**S5-S6 Table**). This was expected because of the limited information available from the site frequency spectrum [58–59]. Moreover, we hypothesized that the life history trait of the species, particularly its strong natal homing behaviour [60] may result in different local evolutionary trajectories. Admittedly, such unmodelled process may further complicate the interpretation of demographic parameters.

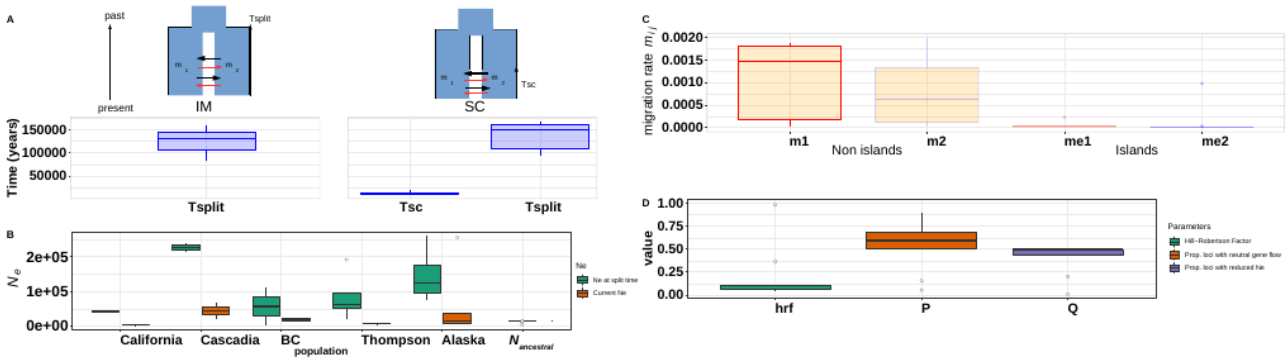


Figure 4) Inferences of demographic history using ∂a∂i
A) Estimates of divergence time (in years) between each major region as inferred by ∂a∂i under the best model (displayed in blue) based on SNP data. B) Estimates of effective population size Ne for each major region as inferred by ∂a∂i under each best demographic model. C) Estimates of migration rate among populations in neutral regions (m) of the genome and in areas of restricted recombination (m_e) or islands of divergence. D) mean Hrf (Hill-Robertson factor) inferred across all comparisons. The Hrf estimates the extent of Ne reduction in areas affected by linked selection. Values are the averaged from multiple pairwise comparison across regions.

Demographic inferences support glacial population isolation and recent postglacial recolonization

Our demographic inferences provided highly consistent divergence time estimates among all models and populations (**Fig. 4)**, with an estimated divergence time of 140 Kya (average = 107 KyA under IM and 137 Kya under SC; **Fig. 4, S5-S6 Table**) and suggesting a nearly simultaneous subdivision and colonization of the Northern range, likely before the Wisconsin glaciation. Accordingly, the median time of secondary

contact (SC) among all models involving the Thompson R. was 13.5 Kya [min 9,200 – max = 20,100], corresponding to the onset of the last glacial retreat (**Fig. 4**), supporting the idea that the Thompson R. has evolved in isolation from the remaining populations for a long period of time.

All models incorporating linked selection and restricted introgression along the genome always received the highest support. Here we modelled the effect of linked selection assuming that it can be approximated as a reduction of *Ne* in regions of low recombination [15]. Our inferences revealed that on average, 40% of the genome was undergoing reduced effective population size (Hill Robertson Factor in **S4 Table, Fig. 4**). Accordingly, the *Ne* values were reduced to approximately 20 percent of their initial values depending on the considered population (i.e. Hrf factor in **Fig. 4**). Obviously, the magnitude of the reduction would require further investigation as quantifying the effect of linked selection is likely more complicated than a simple rescaling of *Ne*. Still, these results provided increased support for a role of linked selection affecting the estimation of demographic parameters [17,61] and potentially shaping the species' genomic landscape. Quantifying how background selection affects the species' genome and shapes the genomic landscape of divergence is beyond the scope of this study and awaits further investigation. Finally, intrinsic barriers to gene flow reduced the estimated migration rate to nearly zero inside genomic islands and affected approximately 47% of the genome (**Fig. 4, S6 Table**).

To consolidate inferences made from ∂a∂i we constructed a global model in Fastsimcoal. To limit the confounding of barriers to gene flow [11], linked selection [12] or gBGC [16], we restricted our jSFS to areas of high recombination, as inferred by LDHAT [62] from whole genome sequences (unpublished data). Results largely validate our ∂a∂i inferences and provide rich information regarding population founder events toward the north followed by expansion as detailed in **Supplementary Note S1 and S12 Fig**.

In summary, our results best supported a scenario whereby contemporary populations mainly originated from a single major ancestral refugium located south of the ice sheets in Cascadia/California, followed by a postglacial expansion and population divergence along the South to North recolonization axis. As a consequence, we expect that the South-North expansion should favour the accumulation of deleterious mutations at the expansion front toward the north [63,64], but also with the possibility for a pulse of deleterious variants from the isolated populations from the Thompson River watershed when they came into secondary contact with other expanding populations.

The recent evolution of Ne shaped the current mutation loads

Our above demographic reconstruction revealed expansion from a major southern refugium. Consequently, we hypothesized that such a history has generated significant differences in the deleterious mutation load among populations, consistent with patterns documented in human populations [63–65]. Therefore, we hypothesized that populations having undergone a strong expansion following a bottleneck should display a higher load than populations of more constant size. Under this hypothesis we predicted a negative correlation between Tajima's D, a classic population genetics statistic capturing population demography changes and the ratio of non-synonymous to synonymous diversity (π_N/π_S), a commonly used proxy for the deleterious mutation load [66]. Accordingly, π_N/π_S was significantly and negatively correlated with Tajima's D (**Fig 5,** R^2 =0.65, *p = 1e-15*, slope = 3.5).

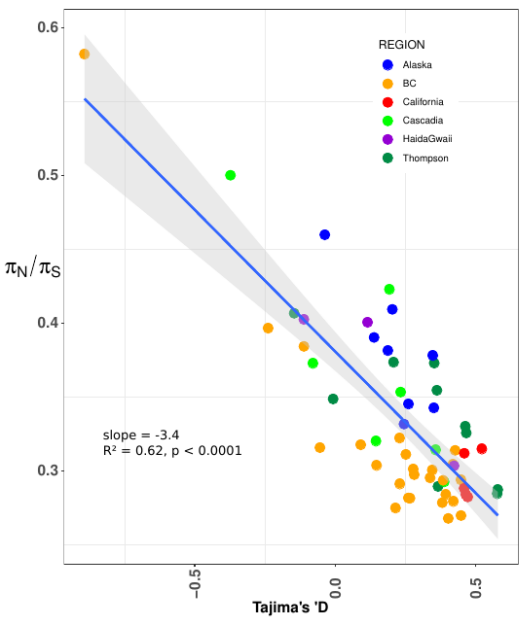


Figure 5) Recent demographic history shaped the deleterious mutation load: *Correlation between Tajima's D and the deleterious load (π_N/π_S).*

Considering our demographic inferences, this result illustrates how the postglacial recolonization changed local effective populations and shaped the present-day deleterious load variation among these populations. Interestingly, even after excluding the two most highly loaded populations (Clakamas and Upper Pitt River in Cascadia and BC respectively), the relationship between the π_N/π_S and the Tajima's D still remain strong (R^2 = 0.42; *p = 4e-8*, slope= -2.7). Our results suggest that the Clackamas River populations exhibit higher levels of admixture with the Thompson R populations (details in **S13 Fig**). As a consequence, we hypothesize that the post-glacial secondary contact with this population contributed to the introgression pulses of maladaptive alleles that are still detectable nowadays. This raises exciting future research questions regarding the role of past introgression and the maintenance of deleterious alleles [67]. It must however be noted that the Tajima's D values were more randomly distributed than expected under a strict model of population expansion from the South (assuming a linear decrease with distance, **S14 Fig**), suggesting additional sources of variation that remain to be identified and their effect evaluated.

Genetic surfing of deleterious mutations during the postglacial recolonization

One expected consequence of a postglacial recolonization is associated with the possibility for neutral and deleterious mutations to surf at the expanding front [63]. This situation occurs because populations at the expansion front exhibit smaller sizes that prevent the efficient purging of deleterious variants. Theory predicts (i) a higher recessive load (measured as the proportion of homozygous derived deleterious mutations) at the expansion front and (ii) an approximately constant total load (which can be approximated by the total number of derived deleterious mutations under an additive model [28,63,65]. We found support for these predictions.

First, we found a nearly linear relationship between the distance to the candidate source (defined as the one with the lowest βST and highest diversity) and the recessive load (linear models, p<0.003; R2= 0.13, **Fig. 6 a**), validating theoretical expectations from range expansion [65]. This latter analysis was performed with PROVEAN [68] based on a total of 1,297 deleterious mutations for which we were able to identify the derived allele using the genome sequences of the 1) Chinook salmon (*Oncorhynchus tshawytscha*) [69], 2) Rainbow trout (*O. mykiss*) [70] and 3) Atlantic salmon (*Salmo salar*) [70] (see methods). Second, and as expected, the total load was not significantly correlated with the distance to the source (R^2 =0.02, p = 0.96, **Fig. 6b**). Since PROVEAN is primarily designed for Human and Mouse (although it has been used successfully in non-mammalian studies [71,72], we also plotted the distribution of π_N/π_S as a function of the distance to the southernmost site. The π_N/π_S ratio was significantly correlated with the distance from the southernmost site **(Fig. 6c**, *p = 0.05;* R^2 = 0.05, for all populations).

After excluding Clackamas and Upper Pitt populations (hypothesized to display increased load due to admixture), the correlation significantly increased (R^2 = 0.2, *p = 0.0003*, slope =2.395e-05*)*. Third, we observed significant differences in the derived allele frequency (DAF) at predicted deleterious non-synonymous mutations among regions (Kruskal-Wallis chi-squared = 100.57, df = 5, *p-value < 2.2e-16*) as well as among populations (Kruskal-Wallis chi-squared = 638.59, df = 57, *p-value < 2.2e-16*, S11 Table) with higher values indicative of less efficient purging (See supplementary **Fig S15-S18 and S9-S10 Table for details**). To the best of our knowledge it is the first time that such correlations between different proxies of the load and distance to the putative refugial origin are reported other than in studies on humans [28,64] and a few in plants [73,74] or bacteria [75]. As such, our two major results regarding Tajima's D and the evolution of expansion load as a function of the distance aligned perfectly with the inferred expansion history and recent change in effective population size in each local population.

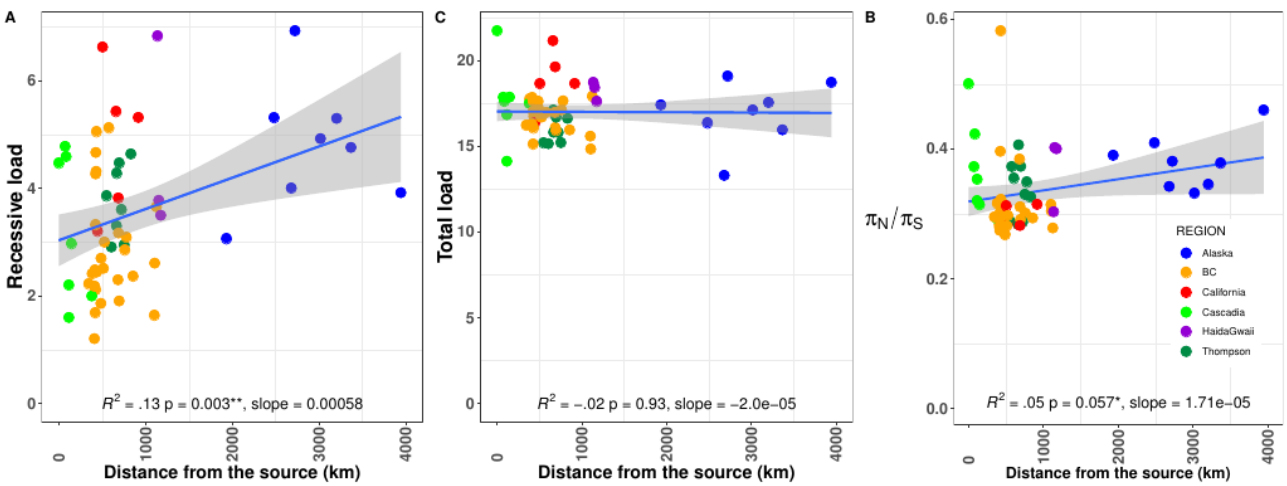


Figure 6) Expansion history and geographic pattern of deleterious mutation load: *Correlation between the distance to the southernmost site and the deleterious load measured as (A) the distribution of homozygous derived putatively deleterious mutations (i.e. recessive load). B) the total load expected to be approximately constant among all populations) and C) the π_N/π_S ratio.*

Finally, a general prediction from population genetic theory is that deleterious mutations in a heterozygous state should be more frequent than in a homozygous state, especially in populations with higher effective sizes, where selection should be more effective at purging mutations with increased homozygosity [19]. Accordingly, we found that 77% of deleterious mutations were maintained in heterozygous states across all samples, as reported in other species [71]. Also, salmon from Alaska, Haida Gwaii, and California populations harboured a significantly higher number of deleterious mutations in a homozygous state when compared to Cascadia or British Columbia (Wilcoxon-test, $p < 0.01$) **(S12a Fig, S11 Table**), a result that is in line with our finding of increased expansion load in **Fig 6a**. When considering the total load of derived deleterious mutations, we found that, on average, there were indeed significantly more putatively deleterious variants per individual in California, Cascadia, and Haida Gwaii populations than in fish from Alaska, British Columbia, or the Thompson R. watershed (**S10 and S11 Table, S12B Fig**, Wilcoxon-test, $p < 0.01$), although these differences were modest. Accordingly, we observed that π_N/π_S varies as a function of the levels of nucleotide diversity (π_S). We indeed found a positive relationship between π_N/π_S and π_S (slope = 175.5; $R^2 = 0.45$, $p < 0.0001$, **S19 Fig**). We note again the same two populations (Clackamas River & Upper Pitt) were strongly contributing to the linear regression. However, the relationship remained significant after their exclusion ($R^2 = 0.10$, $p < 0.0078$). This result provides increased support for our above findings that populations with higher effective population size also displayed a higher deleterious load when considering mildly deleterious variants in heterozygous states (those that are more likely to contribute to π_N). This suggests that strongly that deleterious mutations are efficiently purged in populations of large *Ne* while mildly deleterious mutations would still segregate at appreciable frequency [23], as observed elsewhere [21,22,26]

The role of historical demography in generating differences in the mutation load at the population level remains debated among evolutionary biologists, particularly among human geneticists [7,9,28,29,76]. Our results are however consistent with the nearly neutral theory and the accumulation of deleterious mutations in populations on the expansion front, thus producing a so-called expansion load [63,65]. It is noteworthy how well our findings mirror the 'out-of-Africa' expansion model of human populations [77], despite the fact that the investigated geographic distance for the Coho salmon is an order of magnitude smaller than that involved in the human studies, and that the presence of additional smaller refugia such as Thompson R. (as opposed to the sole African ancestral origin for humans) led to more challenging inferences.

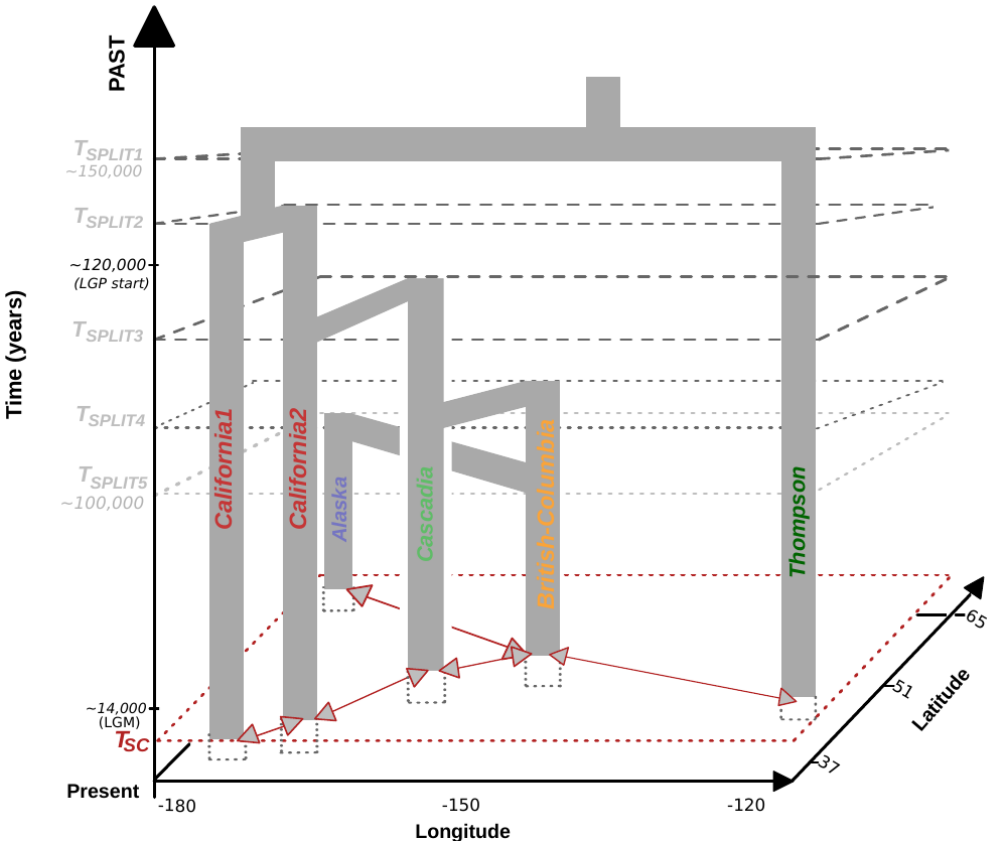


Figure 7: **Simplified spatio temporal representation of the coho salmon demographic history based on ∂a∂i and fastsimcoal results.** Each bar represents a population branch. the width of the branch is not proportional to the population effective size and do not account for historical population size changes. The split time are not to scale and geographic positions are not exact. Tsc = Time of secondary contact. Red arrows reprensents major gene flow events between neighboring populations. Details about gene flow event can be found in table S5 and table S7. LGM = approximate start time of the Last Glacial Period [93]. Beringia was unglaciated during most of the pleistocene [94]. See results for details of the grouping.

Conclusion

In this study, we provided rare evidence of how demographic modelling can provide valuable information for conservation genomics. We found considerable statistical support for a recent postglacial demographic history in the Coho salmon, consistent with a main southern refugium, which generated relatively simple latitudinal gradients in nucleotide diversity and mutation load. Complex demographic processes including past population expansions or periods of isolation and secondary contact, have also been reconstructed with state-of-the-art demographic modelling methods. This was made possible by jointly considering the confounding impact of linked selection for such demographic reconstructions. Altogether, this postglacial history, as well as some additional demographic changes, have influenced the efficacy of selection and favoured the accumulation of deleterious mutations in the most distant populations from the main refugium (Fig 7). Using the Coho salmon as a case study, we demonstrated how these approaches are meaningful for conservation genomics and provide opportunities to disentangle the impacts of historical *vs.* contemporary drivers of population declines. Such investigations also allow assessing the distribution of all selected variants, including not only the beneficial, but also deleterious ones.

Methods

Genotyping by Sequencing

A total of 2,088 individuals were collected from 58 sample sites located along the Pacific coast from California to Alaska (S1 Table and Figure 1). DNA was extracted from all individuals and sequenced using a GBS method (protocol detailed in [78]). Reads were aligned to the Coho salmon reference genome v1 (GCF_002021745.1) using bwa-mem 0.7.13 [79]. Samtools v1.7 [80] was used to keep reads with a mapping quality above 20, remove supplementary alignment and unmapped read. Variants were then

called with Stacks v1.46 [81]. To do so, the module "pstacks" was used with a minimum depth of 5, and up to three mismatches were allowed in catalog assembly. The module "populations" was run to produce a vcf file that was filtered with a custom python script. We performed stringent filtering to remove SNPs that were 1) genotyped in less than 60% of the individuals; 2) at a mean depth of sequencing below 7, and 3) with observed heterozygosity above 0.60, thus resulting in 93,000 SNPs. The pipeline for SNP calling is available on github at https://github.com/enormandeau/stacks_workflow/releases/tag/coho_demography_paper. Next, we removed any individuals with more than 5% missing data and finally only kept SNPs present in at least 95% of the individuals yielding a total of 82,772 filtered SNPs for 1,957 individuals. Remaining filtration was done according to the requirement of each analysis performed below. To perform explicit demographic analyses based on the site frequency spectrum, the likelihood approach implemented in ANGSD v0.930 [82] was used instead of the commonly used Stacks pipeline because it is less biased than the genotype calling approach [83]. We first used the SAMTools model [80] in ANGSD to estimate genotype likelihoods from BAM files using only reads with a minimal base quality score of 20 and a minimal mapping quality score of 30 across all individuals. We further filtered out reads with a depth below 5 and above 100. Given the high residual tetrasomy in the species, the upper bounds enabled removing the high rates of paralogs that would otherwise generate an excess of high frequency shared variants in the jSFS. We verified that such variants were effectively removed in all jSFS before ∂a∂i fitting. We further constructed 1D SFS for all localities to then compute Tajima's D value as a summary of recent effective population size change.

Genetic diversity and ancestral populations

For each sampling location we estimated the observed heterozygosity and π using vcftools 0.1.16 [84] and Hierfstat [85]. The most likely ancestral populations were identified using β_{ST} [42]. A total of 1,000 bootstraps was performed to obtain the 95% confidence intervals around the β_{ST}. Weir and Cockerham's F_{ST} estimator θ [46] was computed in vcftools. We measured the relationship between observed heterozygosity, β_{ST}, F_{ST} and the distance to the southernmost site using linear models with the lm() function implemented in R. We also verified the relationship between F_{ST} and the distance to the southernmost site using Mantel tests with 10,000 permutations. Vcftools was also used to identify singletons (i.e. variants present in one single individual across the whole dataset). Their distributions were then summed in each locality. We then computed the averaged (min, max and median) number of singletons at the regional level as well as levels of private and shared polymorphisms (S2 Fig.). The scripts are available on GitHub at https://github.com/QuentinRougemont/utility_scripts

Population structure, admixture and gene flow

Levels of ancestry and admixture proportions were inferred with the snmf function implemented in the R package LEA [50], allowing only less than 5% of missing data. We then kept a set of SNPs in approximate linkage equilibrium by removing pairs of variants with r^2 greater than 0.2 (option --indep-pairwise 50 10 0.2) resulting in 40,632 SNPs. K-values ranging between 1 and 60 were tested and cross-validations were obtained using the cross-entropy criterion with 5% of masked genotypes. The default value for the regularization parameter was used to avoid forcing individuals into groups and hence underestimating admixture. Similar results were obtained from Admixture [86] and are not presented here. The genetic relationship among all salmon was assessed using a PCA with the R package ade4 [103] based on the LD-pruned dataset (40,632 SNPs). We used a 1% minor allele frequency (MAF) threshold and allowed less than 4% missing data. Formal tests of admixture were performed using Treemix [52] using the LD-pruned dataset of 40,632 SNPs and without any MAF threshold. An MDS was also constructed using plink and plotted with the ggplot2 [87] R package. We ran Treemix allowing up to 20 migration events and performed 500 bootstrap replicates of the best model to obtain robustness of the nodes. The "best" model was inferred after inspecting the relevant migration edges by measuring the percentage of variance explained as migration edges were added to the tree, as well as by assessing the p-value associated to each migration edge. A total 500 bootstrap replicate runs were performed under the "best" model and under a model without migration to

infer the robustness of the nodes. The scripts are available on GitHub at https://github.com/QuentinRougemont/treemix_workflow

Explicit demographic inferences.

We tested alternative hypotheses of divergence from a single southern or northern refugium versus divergence from two or more refugia (e.g. Alaska & California). In the first case, by comparing different regional groups, we expected a model of divergence with ongoing migration (IM) to be supported, or alternatively a model of divergence with ancient migration if gene-flow has stopped recently (AM) following the divergence and recent colonization of rivers. In the case of divergence into multiple refugia we expected models of secondary contact between putative groups to be best supported. We further expected postglacial gene flow between the historical refugia to leave detectable traces that should correspond to the signals of secondary contact. In the unlikely case of no gene flow between regional groups, we expected a model of strict isolation to be most likely. All divergence scenarios are represented in Fig S12 and initially described in [17,88]. The four major models tested included a model of Secondary Contact (SC), a model of Strict Isolation (SI), a model of Ancient Migration (AM) and a model of Isolation with Migration (IM).

The models shared the following parameters: the ancestral populations of size N*anc*, splits at time T_{split} into two daughter populations of size N_1 and N_2. Under the SI model, no gene flow occurs between the two populations. Under AM, gene flow occurred between T_{split} and T_{am} and is followed by a period of strict isolation. Under IM, gene flow occurs at a constant rate at each generation between the two populations. Gene flow can be asymmetric, so that two independent migration rates m_{12} (from population 2 to 1) and m_{21} (from population 1 to 2) were modeled. Under the SC model, the population evolved in strict isolation between T_{split} and until T_{sc} where a secondary contact occurs continuously up to present time. Gene flow is modeled as $M = 2Nref.m$. In ∂a∂i, heterogeneous introgression was modeled using two categories of loci occurring in proportions P (i.e., loci with a migration rates M_{12} and M_{21}) and $1\text{-}P$ (i.e., loci with a reduced effective migration rates Me_{12} and Me_{21}) across the genome. The same procedure was used to account for linked selection by defining two categories of loci with varying effective population sizes (proportion Q of loci with a "neutral N_e" and a proportion 1-Q of loci with a reduced effective population size due to either linked or background selection). To quantify how linked selection affects reduced *Ne*, we used a Hill-Robertson scaling factor (Hrf) to relate the effective population size of loci influenced by selection ($Nr = Hrf * Ne$) to that of neutral loci (N_e). A hierarchical approach was used to avoid over-fitting: first we compared models assuming constant effective population size. Second, the best identified models were modified to incorporate population expansion or decline, as expected given the observed distribution of genetic diversity. Population expansion was implemented using two additional parameters for population 1 and population 2, allowing each population to either grow or decline exponentially.

Models were fitted using the diffusion theory implemented in ∂a∂i [53] and includes the effect of linked selection and barriers to gene flow as detailed in [17,88]. ∂a∂i uses the SFS as a summary of the data. For a given demographic model, the SFS is computed using diffusion approximation and compared to the empirical SFS using AIC. Here, we started from the whole file containing 200,000 SNPs and used one single SNP per GBS locus and filtered the data to minimize missing data. No MAF filter was used and singletons were kept to avoid ascertainment bias in estimates of demographic parameters. For each model, 20 independent replicate runs were performed and only models with the lowest AIC and ΔAIC were kept. A model was classified as "ambiguous" and not used for parameter estimation if ΔAIC between the best model and second-best model was below 10. The Godambe information criteria was initially used to evaluate parameter uncertainties based on 100 bootstrapped datasets, which resulted in large confidence intervals in a few pairwise comparisons. Therefore, we then performed parametric bootstrap. More specifically, we converted ∂a∂i estimates into ms [89] estimates to construct sets of 100 bootstrapped datasets over the different categories of the SFS (i.e. neutral SFS, SFS with reduced rate of migration, SFS in low recombination areas undergoing linked selection, SFS in areas of normal recombination) that were summed and then fitted with ∂a∂i. The new estimates were then converted into demographic units to obtain the 95%

confidence intervals. The whole pipeline is available at https://github.com/QuentinRougemont/DemographicInference.

Inferring a global divergence history

While the ∂a∂i analyses provided valuable results regarding the demographic history and how it is confounded by linked selection and barriers to gene flow, this analytical framework is currently limited to pairwise comparisons, which does not allow painting a global scenario of the species divergence history. Therefore, we build upon the ∂a∂i models and other observed statistics to construct a global scenario of population expansions and secondary contact. This scenario was then tested using Fastsimcoal v2.6 [54] and is described in supplementary note S1.

Genetic load estimation from the GBS data set

Estimating ancestral and derived alleles

We used the derived allele count as an estimator of deleterious allele count. We used the genomes of three outgroup species, the chinook salmon, the rainbow trout, and the Atlantic salmon, to classify SNPs as ancestral or derived. Whole genome data for the chinook salmon (n = 3 individuals) were provided by one of us (B. Koop, unpublished), whereas rainbow trout (n = 5) and Atlantic salmon (n = 5) data were downloaded from NCBI Sequence Read Archive (rainbow trout, SRA, Bioproject : SRP117091; *Salmo salar* SRA Bioproject: SRP059652). Every individual was aligned against the Coho salmon V1 reference genome (GCF_002021745.1) using GATK UnifiedGenotyper and calling every SNP using the EMIT_ALL_SITES modes. We then constructed a Python script and determined the ancestral state of the GBS SNPs if 1) the SNP was homozygous in at least two of the three outgroups, and 2) match one of the two alleles identified in Coho salmon. Otherwise, the site was inferred as missing and not used in subsequent analyses.

Measuring the deleterious load through piN/piS ratio

The ratio of non-synonymous diversity to synonymous diversity is a commonly used metric to quantify the deleterious mutation load [66]. To do so, we first generated aligned fasta sequences and then computed summary statistics. More specifically, for every individual, we ran GATK v4.1.2.0 on the bam file generated above following GATK best practices. We generated individual gVCF files then combined individuals into populations using the CombinedGVCFs module from GATK and finally performed a joint genotyping using the all-site options. We then reconstructed fasta sequences from VCF files for every individual by reconstructing two genomic sequences with a pipeline modified from [90]. We then used cutSeqGff.py to subsample and clean sequences considering several genomic features (CDS, introns, intergenic regions) from gff files. The resulting alignment was then used to compute the π_N/π_S ratio in sequences containing 4-fold codons only. We then tested a number of correlations between the load (π_N/π_S) and 1) π_S used as a proxy of the long-term effective population size; 2) Tajima's D used as a proxy for historical change in population size and 3) distance to the Southernmost site. Significance of the correlation were tested using the lm() function in R. Plots were drawn using ggplot2 [87]. All pipelines to run GATK and compute deleterious load are available at :*https://github.com/QuentinRougemont/gatk_haplotype and https://github.com/QuentinRougemont/piNpiS.*

Measuring damaging impact of non-synonymous alleles

While the computation above provides insights into the total deleterious mutation load, it does not allow distinguishing among recessive and additive load nor testing for differences in allelic frequencies. We therefore further tested differences in mutation load among populations as follows. The software Provean [68] was used to predict whether a given non-synonymous mutation was deleterious with a threshold score of -2.5 or less using the pipeline available at https://github.com/QuentinRougemont/gbs_synonymy_with_genome. We analysed the data in two ways: first we counted the total number of putative homozygous deleterious alleles per individual as well as the total number of deleterious alleles (both in homozygous and heterozygous states) using: $N_{total} = 2 \text{ X } N_{homo} +$

N_{hetero} [34]. These two metrics are expected to be proportional to the recessive and deleterious load respectively [28,65]. These individual values were then averaged per population and major regional group (i.e., California, Cascadia, British Columbia, Haida Gwaii, Thompson, and Alaska). We then computed derived allele frequencies (DAF) in all sampling locations and across mutation categories (synonymous, non-synonymous, and non-synonymous deleterious) and tested for significant allele frequency differences among populations in non-synonymous and non-synonymous deleterious mutations using Wilcoxon rank sum tests. DAF spectra were constructed for all populations separately. For the ease of visualisation, we also constructed DAF spectra by region for a sample of size n = 100 individuals (Fig 7A). This size was chosen according to the smallest sample size of the three combined Haida Gwaii populations. Finally, we used SnpEff v3.4 [91] to obtain the functional annotations of the putatively deleterious variants. The annotations were comprised of missense variants, non-coding transcripts, 3' and 5' untranslated regions, 5kb up- and down-stream variants, intergenic and intronic, splice acceptor and splice region, stop gained and start loss. We found 35% of deleterious mutations to be missense variants, 38% to be non-coding transcripts and 22% to be either upstream or downstream gene variants, with the remaining being spread over the other categories.

Ethic Statement

A permit number SIRUL 111722 was obtained to work on DNA sequences.

Raw data will be deposit on NCBI together with Short Read Archive (SRA) accession number. All relevant vcf and jSFS files will be deposited on dryad.

Acknowledgements:

We thank B. Bougas, A. Perrault-Payette, C. Hernandez for laboratory support and K. Wellband, H. Cayuela, F. Hartmann for their constructive comments. TL and QR thank Benoit Nabholz for discussions regarding the deleterious mutation loads and Hill-Robertson effects. We want to warmly thank two reviewers (Christelle Fraïsse & an anonymous reviewer) and Associate Editor for their thoughtful comments and efforts towards improving a previous version of our manuscript. Computations were performed on Colosse (Calcul Quebec), Graham and Cedar (Compute Canada) servers. This research was carried out in conjunction with EPIC4 (Enhanced Production in Coho: Culture, Community, Catch), a project supported by the government of Canada through Genome Canada, Genome British Columbia, and Genome Quebec. The authors declare no conflicts of interest.

References

1. Provan J, Bennett KD. Phylogeographic insights into cryptic glacial refugia. Trends Ecol Evol (Amst). 2008;23: 564–571. doi:10.1016/j.tree.2008.06.010

2. Hewitt GM. Post-glacial re-colonization of European biota. Biological Journal of the Linnean Society. 1999;68: 87–112. doi:10.1006/bijl.1999.0332

3. Hewitt GM. Genetic consequences of climatic oscillations in the Quaternary. Philos Trans R Soc Lond, B, Biol Sci. 2004;359: 183–195; discussion 195. doi:10.1098/rstb.2003.1388

4. Bernatchez L, Wilson CC. Comparative phylogeography of Nearctic and Palearctic fishes. Molecular Ecology. 1998;7: 431–452. doi:10.1046/j.1365-294x.1998.00319.x

5. Frankham R, Ballou JD, Briscoe DA. Introduction to Conservation Genetics by Richard Frankham. In: Cambridge Core [Internet]. Jan 2010 [cited 2 Jul 2019]. doi:10.1017/CBO9780511809002

6. Funk WC, McKay JK, Hohenlohe PA, Allendorf FW. Harnessing genomics for delineating conservation units. Trends in Ecology & Evolution. 2012;27: 489–496. doi:10.1016/j.tree.2012.05.012

7. Simons YB, Turchin MC, Pritchard JK, Sella G. The deleterious mutation load is insensitive to recent population history. Nature Genetics. 2014;46: 220–224. doi:10.1038/ng.2896

8. Kirkpatrick M, Jarne P. The Effects of a Bottleneck on Inbreeding Depression and the Genetic Load. Am Nat. 2000;155: 154–167. doi:10.1086/303312

9. Simons YB, Sella G. The impact of recent population history on the deleterious mutation load in humans and close evolutionary relatives. Curr Opin Genet Dev. 2016;41: 150–158. doi:10.1016/j.gde.2016.09.006

10. Bierne N, Gagnaire P-A, David P. The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. Curr Zool. 2013;59: 72–86. doi:10.1093/czoolo/59.1.72

11. Barton N, Bengtsson BO. The barrier to genetic exchange between hybridising populations. Heredity. 1986;57: 357. doi:10.1038/hdy.1986.135

12. Cruickshank TE, Hahn MW. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol Ecol. 2014;23: 3133–3157. doi:10.1111/mec.12796

13. Noor M a. F, Bennett SM. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity (Edinb). 2009;103: 439–444. doi:10.1038/hdy.2009.151

14. Charlesworth B. The Effects of Deleterious Mutations on Evolution at Linked Sites. Genetics. 2012;190: 5–22. doi:10.1534/genetics.111.134288

15. Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. Genetics. 1993;134: 1289–1303.

16. Pouyet F, Aeschbacher S, Thiéry A, Excoffier L. Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. Veeramah K, Wittkopp PJ, Gronau I, editors. eLife. 2018;7: e36317. doi:10.7554/eLife.36317

17. Roux C, Fraïsse C, Romiguier J, Anciaux Y, Galtier N, Bierne N. Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence. PLOS Biology. 2016;14: e2000234. doi:10.1371/journal.pbio.2000234

18. Comeron JM. Background selection as null hypothesis in population genomics: insights and challenges from Drosophila studies. Philos Trans R Soc Lond, B, Biol Sci. 2017;372. doi:10.1098/rstb.2016.0471

19. Charlesworth D, Willis JH. The genetics of inbreeding depression. Nat Rev Genet. 2009;10: 783–796. doi:10.1038/nrg2664

20. Kyriazis C, Wayne RK, Lohmueller KE. High genetic diversity can contribute to extinction in small populations | bioRxiv. [cited 10 Mar 2020]. Available: https://www.biorxiv.org/content/10.1101/678524v1

21. Robinson JA, Räikkönen J, Vucetich LM, Vucetich JA, Peterson RO, Lohmueller KE, et al. Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. Sci Adv. 2019;5: eaau0757. doi:10.1126/sciadv.aau0757

22. Robinson JA, Ortega-Del Vecchyo D, Fan Z, Kim BY, vonHoldt BM, Marsden CD, et al. Genomic Flatlining in the Endangered Island Fox. Curr Biol. 2016;26: 1183–1189. doi:10.1016/j.cub.2016.02.062

23. Abascal F, Corvelo A, Cruz F, Villanueva-Cañas JL, Vlasova A, Marcet-Houben M, et al. Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. Genome Biol. 2016;17: 251. doi:10.1186/s13059-016-1090-1

24. Dobrynin P, Liu S, Tamazian G, Xiong Z, Yurchenko AA, Krasheninnikova K, et al. Genomic legacy of the African cheetah, Acinonyx jubatus. Genome Biol. 2015;16: 277. doi:10.1186/s13059-015-0837-4

25. Yang Y, Ma T, Wang Z, Lu Z, Li Y, Fu C, et al. Genomic effects of population collapse in a critically endangered ironwood tree Ostrya rehderiana. Nat Commun. 2018;9: 5449. doi:10.1038/s41467-018-07913-4

26. Xue Y, Prado-Martinez J, Sudmant PH, Narasimhan V, Ayub Q, Szpak M, et al. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science. 2015;348: 242–245. doi:10.1126/science.aaa3952

27. Grossen C, Guillaume F, Keller LF, Croll D. Accumulation and purging of deleterious mutations through severe bottlenecks in ibex. bioRxiv. 2019; 605147. doi:10.1101/605147

28. Henn BM, Botigué LR, Peischl S, Dupanloup I, Lipatov M, Maples BK, et al. Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. PNAS. 2016;113: E440–E449. doi:10.1073/pnas.1510805112

29. Henn BM, Botigué LR, Bustamante CD, Clark AG, Gravel S. Estimating the mutation load in human genomes. Nat Rev Genet. 2015;16: 333–343. doi:10.1038/nrg3931

30. Krkosek M, Ford JS, Morton A, Lele S, Myers RA, Lewis MA. Declining wild salmon populations in relation to parasites from farm salmon. Science. 2007;318: 1772–1775. doi:10.1126/science.1148744

31. Irvine JR, Fukuwaka M. Pacific salmon abundance trends and climate change. ICES J Mar Sci. 2011;68: 1122–1130. doi:10.1093/icesjms/fsq199

32. Gustafson RG, Waples RS, Myers JM, Weitkamp LA, Bryant GJ, Johnson OW, et al. Pacific Salmon Extinctions: Quantifying Lost and Remaining Diversity. Conservation Biology. 2007;21: 1009–1020. doi:10.1111/j.1523-1739.2007.00693.x

33. Smith CT, Nelson RJ, Wood CC, Koop BF. Glacial biogeography of North American coho salmon (Oncorhynchus kisutch). Mol Ecol. 2001;10: 2775–2785.

34. Beacham TD, Wetklo M, Deng L, MacConnachie C. Coho Salmon Population Structure in North America Determined from Microsatellites. Transactions of the American Fisheries Society. 2011;140: 253–270. doi:10.1080/00028487.2011.558782

35. McPhail JD, Lindsey CC. Freshwater fishes of northwestern Canada and Alaska. Fisheries Research Board of Canada : available by mail from the Queen's Printer; 1970.

36. UBC Press | Pacific Salmon Life Histories, By Cornelis Groot, Leo Margolis and Leo Margolis. In: UBC Press [Internet]. [cited 1 Jul 2019]. Available: https://www.ubcpress.ca/pacific-salmon-life-histories

37. Hocutt CH, Wiley EO, editors. The Zoogeography of North American Freshwater Fishes. 1 edition. New York: Wiley-Interscience; 1986.

38. Mee JA, Moore J-S. The ecological and evolutionary implications of microrefugia. Journal of Biogeography. 2014;41: 837–841. doi:10.1111/jbi.12254

39. Warner BG, Mathewes RW, Clague JJ. Ice-free conditions on the queen charlotte islands, british columbia, at the height of late wisconsin glaciation. Science. 1982;218: 675–677. doi:10.1126/science.218.4573.675

40. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. 2008;18: 1851–1858. doi:10.1101/gr.078212.108

41. Petit RJ, Aguinagalde I, Beaulieu J-L de, Bittkau C, Brewer S, Cheddadi R, et al. Glacial Refugia: Hotspots But Not Melting Pots of Genetic Diversity. Science. 2003;300: 1563–1565. doi:10.1126/science.1083264

42. Weir BS, Goudet J. A Unified Characterization of Population Structure and Relatedness. Genetics. 2017;206: 2085–2103. doi:10.1534/genetics.116.198424

43. Cubry P, Vigouroux Y, François O. The Empirical Distribution of Singletons for Geographic Samples of DNA Sequences. Front Genet. 2017;8. doi:10.3389/fgene.2017.00139

44. Gilbert-Horvath EA, Pipal KA, Spence BC, Williams TH, Garza JC. Hierarchical Phylogeographic Structure of Coho Salmon in California. Transactions of the American Fisheries Society. 2016;145: 1122–1138. doi:10.1080/00028487.2016.1201003

45. Williams TH, Lindley ST, Spence BC, Boughton DA. STATUS REVIEW UPDATE FOR PACIFIC SALMON AND STEELHEAD LISTED UNDER THE ENDANGERED SPECIES ACT: SOUTHWEST. : 106.

46. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. Evolution. 1984;38: 1358–1370. doi:10.2307/2408641

47. Ward RD, Woodwark M, Skibinski DOF. A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. Journal of Fish Biology. 1994;44: 213–232. doi:10.1111/j.1095-8649.1994.tb01200.x

48. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006;2: e190. doi:10.1371/journal.pgen.0020190

49. Novembre J, Stephens M. Interpreting principal component analyses of spatial population genetic variation. Nat Genet. 2008;40: 646–649. doi:10.1038/ng.139

50. Frichot E, François O. LEA: An R package for landscape and ecological association studies. Methods in Ecology and Evolution. 2015;6: 925–929. doi:10.1111/2041-210X.12382

51. Meirmans PG. The trouble with isolation by distance. Mol Ecol. 2012;21: 2839–2846. doi:10.1111/j.1365-294X.2012.05578.x

52. Pickrell JK, Pritchard JK. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. PLOS Genetics. 2012;8: e1002967. doi:10.1371/journal.pgen.1002967

53. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. PLOS Genetics. 2009;5: e1000695. doi:10.1371/journal.pgen.1000695

54. Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. Robust Demographic Inference from Genomic and SNP Data. PLOS Genetics. 2013;9: e1003905. doi:10.1371/journal.pgen.1003905

55. Alcala Nicolas, Vuilleumier Séverine. Turnover and accumulation of genetic diversity across large time-scale cycles of isolation and connection of populations. Proceedings of the Royal Society B: Biological Sciences. 2014;281: 20141369. doi:10.1098/rspb.2014.1369

56. COSEWIC assessment and status report on the coho salmon Oncorhynchus kisutch (Interior Fraser population) in Canada - Species at Risk Public Registry. [cited 1 Jul 2019]. Available: https://wildlife-species.canada.ca/species-risk-registry/document/default_e.cfm?documentID=105

57. J. Wang, Personal Communication.

57. Lapierre M, Lambert A, Achaz G. Accuracy of Demographic Inferences from the Site Frequency Spectrum: The Case of the Yoruba Population. Genetics. 2017;206: 439–449. doi:10.1534/genetics.116.192708

58. Baharian S, Gravel S. On the decidability of population size histories from finite allele frequency spectra. Theoretical Population Biology. 2018;120: 42–51. doi:10.1016/j.tpb.2017.12.008

59. Terhorst J, Song YS. Fundamental limits on the accuracy of demographic inference based on the sample frequency spectrum. PNAS. 2015;112: 7677–7682. doi:10.1073/pnas.1503717112

60. Quinn TP. Homing and Straying in Pacific Salmon. In: McCleave JD, Arnold GP, Dodson JJ, Neill WH, editors. Mechanisms of Migration in Fishes. Boston, MA: Springer US; 1984. pp. 357–362. doi:10.1007/978-1-4613-2763-9_21

61. Ewing GB, Jensen JD. The consequences of not accounting for background selection in demographic inference. Molecular Ecology. 2016;25: 135–141. doi:10.1111/mec.13390

62. McVean G, Awadalla P, Fearnhead P. A Coalescent-Based Method for Detecting and Estimating Recombination From Gene Sequences. Genetics. 2002;160: 1231–1241.

63. Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L. On the accumulation of deleterious mutations during range expansions. Molecular Ecology. 2013;22: 5972–5982. doi:10.1111/mec.12524

64. Peischl S, Dupanloup I, Foucal A, Jomphe M, Bruat V, Grenier J-C, et al. Relaxed Selection During a Recent Human Expansion. Genetics. 2018;208: 763–777. doi:10.1534/genetics.117.300551

65. Peischl S, Excoffier L. Expansion load: recessive mutations and the role of standing genetic variation. Molecular Ecology. 2015;24: 2084–2094. doi:10.1111/mec.13154

66. Chen H, Chen H, Hu L, Wang L, Wang S, Wang ML, et al. Genetic diversity and a population structure analysis of accessions in the Chinese cowpea [Vigna unguiculata (L.) Walp.] germplasm collection. The Crop Journal. 2017;5: 363–372. doi:10.1016/j.cj.2017.04.002

67. Kim BY, Huber CD, Lohmueller KE. Deleterious variation shapes the genomic landscape of introgression. PLOS Genetics. 2018;14: e1007741. doi:10.1371/journal.pgen.1007741

68. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS ONE. 2012;7: e46688. doi:10.1371/journal.pone.0046688

69. Christensen KA, Leong JS, Sakhrani D, Biagi CA, Minkley DR, Withler RE, et al. Chinook salmon (Oncorhynchus tshawytscha) genome and transcriptome. PLOS ONE. 2018;13: e0195461. doi:10.1371/journal.pone.0195461

70. Yáñez JM, Naswa S, López ME, Bassini L, Correa K, Gilbey J, et al. Genomewide single nucleotide polymorphism discovery in Atlantic salmon (Salmo salar): validation in wild and farmed American and European populations. Mol Ecol Resour. 2016;16: 1002–1011. doi:10.1111/1755-0998.12503

71. Zhou Y, Massonnet M, Sanjak JS, Cantu D, Gaut BS. Evolutionary genomics of grape (Vitis vinifera ssp. vinifera) domestication. PNAS. 2017;114: 11715–11720. doi:10.1073/pnas.1709257114

72. Nevado B, Wong ELY, Osborne OG, Filatov DA. Adaptive Evolution Is Common in Rapid Evolutionary Radiations. Current Biology. 2019;29: 3081-3086.e5. doi:10.1016/j.cub.2019.07.059

73. Willi Y, Fracassetti M, Zoller S, Van Buskirk J. Accumulation of Mutational Load at the Edges of a Species Range. Mol Biol Evol. 2018;35: 781–791. doi:10.1093/molbev/msy003

74. González-Martínez SC, Ridout K, Pannell JR. Range Expansion Compromises Adaptive Evolution in an Outcrossing Plant. Current Biology. 2017;27: 2544-2551.e4. doi:10.1016/j.cub.2017.07.007

75. Bosshard L, Dupanloup I, Tenaillon O, Bruggmann R, Ackermann M, Peischl S, et al. Accumulation of Deleterious Mutations During Bacterial Range Expansions. Genetics. 2017;207: 669–684. doi:10.1534/genetics.117.300144

76. Lohmueller KE. The distribution of deleterious genetic variation in human populations. Curr Opin Genet Dev. 2014;29: 139–146. doi:10.1016/j.gde.2014.09.005

77. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, et al. Worldwide human relationships inferred from genome-wide patterns of variation. Science. 2008;319: 1100–1104. doi:10.1126/science.1153717

78. Moore J-S, Harris LN, Le Luyer J, Sutherland BJG, Rougemont Q, Tallman RF, et al. Genomics and telemetry suggest a role for migration harshness in determining overwintering habitat choice, but not gene flow, in anadromous Arctic Char. Mol Ecol. 2017;26: 6784–6800. doi:10.1111/mec.14393

79. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997 [q-bio]. 2013 [cited 3 Jul 2019]. Available: http://arxiv.org/abs/1303.3997

80. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25: 2078–2079. doi:10.1093/bioinformatics/btp352

81. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. G3: Genes, Genomes, Genetics. 2011;1: 171–182. doi:10.1534/g3.111.000240

82. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics. 2014;15: 356. doi:10.1186/s12859-014-0356-4

83. Warmuth VM, Ellegren H. Genotype-free estimation of allele frequencies reduces bias and improves demographic inference from RADSeq data. Molecular Ecology Resources. 2019;19: 586–596. doi:10.1111/1755-0998.12990

84. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27: 2156–2158. doi:10.1093/bioinformatics/btr330

85. Goudet J. hierfstat, a package for r to compute and test hierarchical F-statistics. Molecular Ecology Notes. 2005;5: 184–186. doi:10.1111/j.1471-8286.2004.00828.x

86. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19: 1655–1664. doi:10.1101/gr.094052.109

87. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2009. Available: https://www.springer.com/gp/book/9780387981413

88. Rougemont Q, Gagnaire P-A, Perrier C, Genthon C, Besnard A-L, Launey S, et al. Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. Molecular Ecology. 2017;26: 142–162. doi:10.1111/mec.13664

89. Hudson RR. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics. 2002;18: 337–338. doi:10.1093/bioinformatics/18.2.337

90. Leroy T, Anselmetti Y, Tilak M-K, Bérard S, Csukonyi L, Gabrielli M, et al. A bird's white-eye view on neosex chromosome evolution. bioRxiv. 2019; 505610. doi:10.1101/505610

91. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6: 80–92. doi:10.4161/fly.19695

92. Government of Canada NRC. GEOSCAN Search Results: Fastlink [Internet]. 7 Dec 2015 [cited 26 Jul 2019]. Available: https://geoscan.nrcan.gc.ca/starweb/geoscan/servlet.starweb?

path=geoscan/fulle.web&search1=R=214399

93. Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, et al. The Last Glacial Maximum. Science. 2009;325: 710–714. doi:10.1126/science.1172873

94. Elias SA, Brigham-Grette J. GLACIATIONS | Late Pleistocene Events in Beringia. In: Elias SA, editor. Encyclopedia of Quaternary Science. Oxford: Elsevier; 2007. pp. 1057–1066. doi:10.1016/B0-44-452747-8/00132-0

Supporting Information Legend

S1 Fig. Linear decrease in genetic diversity when considering πSNP as a function of the distance to the southernmost sample site. Each points represents a sample site and is colored by region.

S2 Fig. Network of shared and private polymorphisms. The branch (grey) represent shared polymorphism between sample site and are proportional to levels of sharing. Each point represent the number of private polymorphisms and is colored by region. Computation were based on a sample of size 100 in each region to enable comparison. Regional groups were chosen based on the literature regarding expected ancestral refugia.

S3 Fig. Patterns of Isolation By Distance. Increasing F_{ST} as a function of the distance to the southernmost site. Each points represents a sample site and is coloured by region. The F_{ST} was computed between the southernmost site and all other remaining sites.

S4 Fig. Summaries of F_{ST} values.
A. F_{ST}-based Hierarchical tree depicting relationship among samples. Colors represent the major region.

B. Heatmap of F_{ST} values among samples ordered from North to South on the X and Y-axis.

S5 Fig. Principal Component Analysis recapitulating the relationship among individuals. The Axis 3 and axis 4 are displayed.

S6 Fig. Multidimensional Scaling (MDS) plot depicting relationship among individuals. Each points represents an individual site and is colored by region.

S7 Fig. Structure and Admixture inferences. A. Admixture Barplot obtained from LEA for various K-values. B. Progressive decrease of LEA cross-entropy criterion. Lower cross-entropy values indicates the number of cluster compatible with the data (here from 30 to 60).

S8 Fig. Treemix results. A. Proportion of variance explained (y-axis) as a function of the number of migration edge (x-axis) B. Treemix tree inferred without gene flow and residuals. C. Residuals for Treemix tree with four migration edges

S9 Fig. Representations of the major groups used in the demographic inferences.

S10 Fig. Compared Demographic Models
Strict Isolation (SI), Isolation with constant Migration (IM), Ancient Migration (AM) and Secondary Contact (SC). The models shared the following parameters: Tsplit: number of generation of divergence (backwards in time). *Nanc*, N_1, N_2: effective population size of the ancestral population, of the first and second daughter population. M_1 and M_2 represent the effective migration rates per generation with *m* the proportion of population made of migrants from the other population. T*sc* is the number of generations since gene flow started (secondary contact) after a period of isolation. T*am* is the number of generations since the two populations have diverged without gene flow. Each models are declined in alternative version allowing homogeneous or heterogeneous effective size and homogeneous and heterogeneous gene flow to account for the effect of linked selection (affecting *Ne)* and barrier to gene flow (affecting *m*).

S11 Fig: : Models and residuals for each best models
Each plot displays the observed jSFS (data), the modeled jSFS (model) and the residuals. Left part : Best model without population size change. Each best model was inferred using ΔAIC and AIC weights. Right panel: The same model as the left but including the possibility for population size change of the diverging daughter populations.

S12 Fig: Global model tested with Fastsimcoal.

S13 Fig: Summary of admixture coefficient obtained from Fig6. For easier interpretations of admixture coefficient among the 58 samples site a PCA was performed to summarized the distribution of admixture. Populations with lower cos2 contributed weakly to the plot and hence have higher admixture.

S14 Fig: Relationship between Tajima's D and distance to the southernmost site. No significant relationship was observed (p=0.5, r=-0.07), suggesting that each local population has undergone different evolutionary trajectory in post-glacial time.

S15 Fig : **DAF spectrum of synonymous, non-synonymous and putatively deleterious mutation aggregated at the regional level for all samples.** Data are normalized for a sample of size n = 102 corresponding to the smallest size for the combined samples in Haida Gwaii.

S16 Fig. DAF spectrum of synonymous, non-synonymous and putatively deleterious mutation in each localities for all samples.

S17 Fig: A) Distribution of the count of homozygous derived deleterious alleles in each major group. B) Distribution of the count of total derived deleterious alleles in each major group.

S18 Fig. Significant mean derived allele frequencies of deleterious mutation. Displayed are the mean derived allele frequencies of polymorphic deleterious sites in each regions +/- 2 standard deviation.

Table Legend:

S1 Table: Abbreviation, Region and coordinates (Longitude and Latitude) of each river used in the GBS data with the number of individuals provided (nb. Inds).

S2 Table: β_{ST} values along with 95% confidence intervals for each river from the GBS data (82 K SNPs). 95% confidence intervals obtained after 1000 bootstraps.

S3 Table: model choice results for dadi. AIC, ΔAIC and AIC weigths are provided for each pairwise comparison and model. AM = Ancient Migration, IM = Isolation with Migration, SI = Strict Isolation, SC = Secondary Contact, the simplest models assume homogeneous migration and homogeneous effective population size. 2N suffix = heterogeneous effective population size, 2Msuffix = heterogeneous migration.
Model with both suffix assumes that both effective population size and migration are heterogeneous.
Model with a single suffix assumes that either migration or effective population size are heterogeneous. In one cases (Calif2 vs Thompson) the SC2N, SC2m and SC2N2m provided similar fit. For consistency the SC2N2m was considered for parameter estimates. Its parameter estimates were more consistent than those from either SC2N or SC2m.

S4 Table : Parameter estimates obtained under the best demographic model with dadi.
Ne1 and Ne2, effective population size of the compared pair. m1 ← 2 and m2 ← 1, migration from population 2 to population 1 and migration from population 1 into population 2. me12 and me21, effective migration rate estimated in the most differentiated regions of the genome Ts: Split Time of the ancestral population in two population; Tsc: duration of the secondary contact P: proportion of the genome freely exchanged (1-P provides the proportion of the genome non-neutrally exchanged); Q: proportion of the genome with a reduced effective population size due to selection at linked sites; hrf = Hill-Robertson factor representing the reduction of Ne in the region Q with reduced Ne.

S5 Table : Confidence Intervals obtained from the Gobambe Information Matrix (GIM).

S6 Table : Confidence Intervals obtained from 100 dataset constructed with the coalescent simulator ms.

S7 Table : Parameter estimates and confidence intervals obtained from Fastsimcoal.

S8 Table: Tajjim'as D value and piN/piS value observed for each locality.

S9 Table: Summary of deleterious variation by region. 1)Derived Allele Frequency (DAF) of deleterious mutation, after averaging by rivers and then by major regional group. 2) Count of deleterious mutations in each rivers and then averaged by major regional group. 3) Number of homozygous derived deleterious mutations by individual, after averaging by rivers and then by major regional group. 4) Number of heterozygous mutations by individuals, after averaging by rivers and then by major regional group 5) Total load of derived deleterious mutations by individuals, after averaging by rivers and then by major regional group.

S10 Table: Results of Wilcoxon test (Man-Whitney tests) for differences in derived allele frequencies among major groups for a sample of size 100.

S11 Table: Results of Wilcoxon test (Man-Whitney tests) for differences in count of derived homozygous variants and total load among individuals in each region.