Allele surfing causes maladaptation in a Pacific salmon of conservation concern

Quentin Rougemont[1*], Thibault Leroy[2], Eric B. Rondeau[3], Ben Koop[4], Louis Bernatchez[5]

[1.] Centre d'Ecologie Fonctionnelle et Evolutive, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France
[2.] INRAE, UMR INRAE/UCA 1095 GDEC, 5 chemin de Beaulieu, 63000 Clermont-Ferrand, France
[3.] Department of Fisheries and Ocean, Pacific Biological Station, 3190 Hammond Bay Road, Nanaimo, British Columbia, Canada Or V9T 6N7
[4] Department of Biology, University of Victoria, Victoria, BC, Canada
[5.] Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Québec, Canada

*** Corresponding Author :** quentinrougemont@orange.fr

Abstract

The Anthropocene threatens worldwide biodiversity, with a potential to induce species decline and range contraction. In this context, it is important to quantify species' adaptive potential and how demographic changes may impact selection efficacy and the burden of deleterious mutations in different populations. Here we show that key evolutionary processes, including variation in effective population size through postglacial change in demography and recombination rates have affected the efficacy of selection and impacted the load in Coho salmon (*Oncorhynchus kisutch*), a widely distributed salmonid species on the west coast of North America. Using whole genome resequencing (30x coverage) data from populations at different latitudinal distances from their southern glacial refugium, we found support for postglacial gene surfing, with reduced *Ne* at the range recolonization front, thus inducing both a reduction of the adaptive potential and a surf of deleterious alleles in populations evolving at low N_e. This inference was robust to various proxy of the load, namely ratios of non-synonymous to synonymous diversity, the number of missense and "loss of function" mutations. In addition, comparing residual tetrasomic and re-diploidizing regions of the salmon genome, we found support for a prime role of recombination rates in shaping the within-genome variation of the load. Overall, our empirical results are remarkably consistent with expectations under the nearly neutral theory of molecular evolution. We discuss the fundamental and applied implications of these findings for evolutionary and conservation genomics.

Introduction

Demographic events, including population contraction, expansion and secondary contacts have profound impacts on the spatial distribution of genetic diversity within species [1]. Range expansions can be accompanied by multiple founder events when few individuals colonize new areas. At the wave front, this reduced number of colonizers results in increased drift effects, a loss of genetic diversity and increased levels of genetic differentiation [2–4] This process favours allele surfing, *i.e.* the fixation of alleles, including neutral,

deleterious or advantageous mutations [2,5,6]. Allele surfing is expected to have two main negative consequences at the wave front *i*) an increase of the deleterious load (called expansion load [7,8]) and *ii*) a loss of adaptive potential [9]. Another demographic process, population bottleneck and long term reduction in population size may increase the inbreeding load, due to the segregation of recessive deleterious alleles, inducing inbreeding depression and fitness decline due to the accumulation of deleterious mutations [10].

Ratios of genetic diversity at non-synonymous over synonymous mutations (π_N/π_S) is a common proxy for the mutation load [11]. Due to the lower efficacy of natural selection in low N_e populations, a greater fraction of slightly deleterious mutations is expected to be present in populations at the wave front (i.e. elevated π_N/π_S). A more direct strategy is based on counting the distribution of variants in coding sequence and classifying them according to their predicted fitness effect (i.e. synonymous, non-synonymous, Loss-of-Function) and whether they contribute to the recessive load or additive load [10,12]. Under the nearly neutral theory of molecular evolution, the reduced efficacy of purifying selection in small populations should increase the fixation rate of slightly deleterious mutations [13]. On the other hand, increased homozygosity of deleterious mutations should enable their removal more efficiently through genetic purging, leading to a reduction of the recessive and additive load [14]. Depending on the methods used to investigate these evolutionary phenomena, results may differ. For instance, mixed evidence has been reported concerning the increase of the human genetic load associated with the Out-of-Africa bottleneck [12,15]. Then, π_N/π_S-based studies generally found support for an increased load [15–17], while studies based on derived allele counts found no differences [12,18,19] following the hypothesis that this bottleneck was too short and/or too recent to substantially impact the load [18]. Beyond the specific case of humans, these approaches have been applied to other expanding populations in plants and bacteria with either support for an increased load [20,21] or no evidence of increased load [22].

The reduction in *Ne* due to founder effects and bottlenecks is theoretically predicted to lead to a reduction in the rate of adaptive substitutions given that the supply of new mutations is proportional to N_e *[23]*. Unless compensatory mutations counteract this effect in low N_e populations [24] one expects a lower proportion of substitutions driven by positive selection as measured by the parameter α originally defined [25] as α = 1 - DsPs/DnPn (with Dn, Ds, Ps, and Pn representing the numbers of non-synonymous substitutions, synonymous substitutions, non-synonymous polymorphism and synonymous polymorphism per gene, respectively). Sophistication of the method allows to dissociate the rate of adaptive substitutions (ω_A) from the rate of non-adaptive amino-acid substitution relative to neutral divergence, a metric called ω_{NA} [26]. Given that large N_e populations generally exhibit higher ω_A [13,23,27] we predicted elevated α values in populations from the ancestral refugium (i.e. source populations) and a decrease as a function of increasing geographic distance to the source, with lowest values at the limit of the distribution range.

In addition to the geographic scale at which the load unfolds, the genomic scale is also important, with

variation in both deleterious load and selection efficacy varying along the genome as a function of recombination rate [28,29]. Indeed, non recombining region (e.g. supergene, sex-chromosome) will more freely accumulate deleterious mutation, a process called Muller's Ratchet [30]. Moreover, Hill-Robertson interference, a process whereby competing alleles interfere with each other to become fixed (e.g. hard selective sweeps) will increase the fixation rate of deleterious segregating variants linked to positively selected sites. Such a process will be exacerbated in the absence of recombination [28,31]. How demographic factors and recombination rate jointly affect selection efficacy and the deleterious load has been rarely investigated (see [32]).

Understanding the consequences of demography and recombination and their putative fitness effect is particularly relevant in Pacific salmon, a group of species of culturally and economically major importance, but in which many populations have been severely declining over the last decades (reviewed in [33]). This is particularly true for the Coho salmon (*Oncorhynchus kisutch)*. From a conservation perspective, this species has undergone recent decline in multiple parts of its native range [34] raising the question of how deleterious mutations may increase in frequency and of their fitness effect. In addition to human induced perturbations, it has undergone a series of postglacial expansion from its main glacial refugium (Cascadia/California) with multiple founder events toward Alaska leading to isolation by distance and a gradient of population structure and decreasing genetic diversity [35,36]. Accordingly, these expansions may result in an increased expansion load, raising important questions relative to the role of demography in generating a significant load and whether this may have fitness consequences for populations at the expansion front.

A salient feature of all salmonids is that their ancestor have undergone a whole genome duplication ~80-100MyA [37] and is still on its path to rediploidization. As a consequence, approximately 8% of the Coho genome displays residual tetraploidy [38]. Theoretically, these regions should display an increased effective population size (4 N_e instead of 2 N_e) and therefore a higher efficacy of selection. We thus predict that these regions should display a lower load compared to diploid regions of the genome due to increased N_e.

The evolutionary consequences of demography as well as human induced demographic changes regarding mutation load and its fitness effects, remains a hotly debated topic, especially from a conservation standpoint (e.g. [39–42]). Moreover, the role of recombination, and its interplay with demography to build up the load across the genome is virtually unknown. Finally, residual tetraploidy offers a unique opportunity to investigate the direct effect of the genomic variation of *Ne* on the efficacy of selection. These fundamental questions could bring insights into how selection efficacy operates across the geographic and genomic spaces.

To address these questions, we aimed at testing the following general hypotheses: i) whether demographic expansion and bottleneck has led to an increased load and decreased selection efficacy, ii) whether

heterogeneous recombination levels induce genome-wide variation in the variation of load as a result of Hill-Robertson effects and iii) whether whole genome duplication results in an increased efficacy of selection and reduced deleterious load through increased recombination and *Ne*.

Results and Discussion

Strong postglacial population expansion revealed through whole genome sequences

We generated high quality (30X coverage) whole genome resequencing data for 14 populations of Coho salmon from California to some of the most upstream populations in Alaska (e.g. Porcupine River in Yukon), resulting in a dataset of 20,9 millions biallelic SNPs after variant calling and filtration. We also included several outgroup species namely the Sockeye salmon (*O. nerka*), Chinook salmon (*O. tshawytscha*), Pink sakmon *(O. Gorbuscha)*, Rainbow trout (*O. mykiss*) and Atlantic salmon (*Salmo salar*) used for polarizing our variant and/or testing correlation between genomic load estimates and recombination rate (**Table S01-S03**).

All Coho populations were differentiated nearly at the river/watershed scale based on PCA and admixture analyses (**Fig S1A, B**). The PCA also demonstrates that populations are spatially structured following both latitude and longitudinal gradients, as previously reported in studies of lower genomic resolution [31,34,35]. Our data revealed a signal of population expansion along the colonization axis from the South to the North based on Tajima's D values (**Fig 1C, Fig S2**). $\boldsymbol{\beta}_{ST}$ - a F_{ST} analog that can be informative of ancestrality (Weir & Goudet, 2017) - was also positively correlated with distance to the southernmost site (R^2 = 0.75, p <2e-16, **Fig S2A**). Genetic diversity (H_S) was negatively correlated with distance to the southernmost site (R^2 = 0.77, p<2e-16, **Fig S2B**). The Salmon River from the Thompson area (green point **Fig 1A** and **Fig S2A**) exhibit lower diversity levels, as compared to all non-Alaskan populations. Excluding this outlier increased the correlation with $\boldsymbol{\beta}_{ST}$ and H_S respectively to R^2 = 0.88 and R^2 = 0.90 (p<2e-16 for both tests), suggesting a population-specific demographic trajectory in this population, which seems consistent with a strong bottleneck and inbreeding.

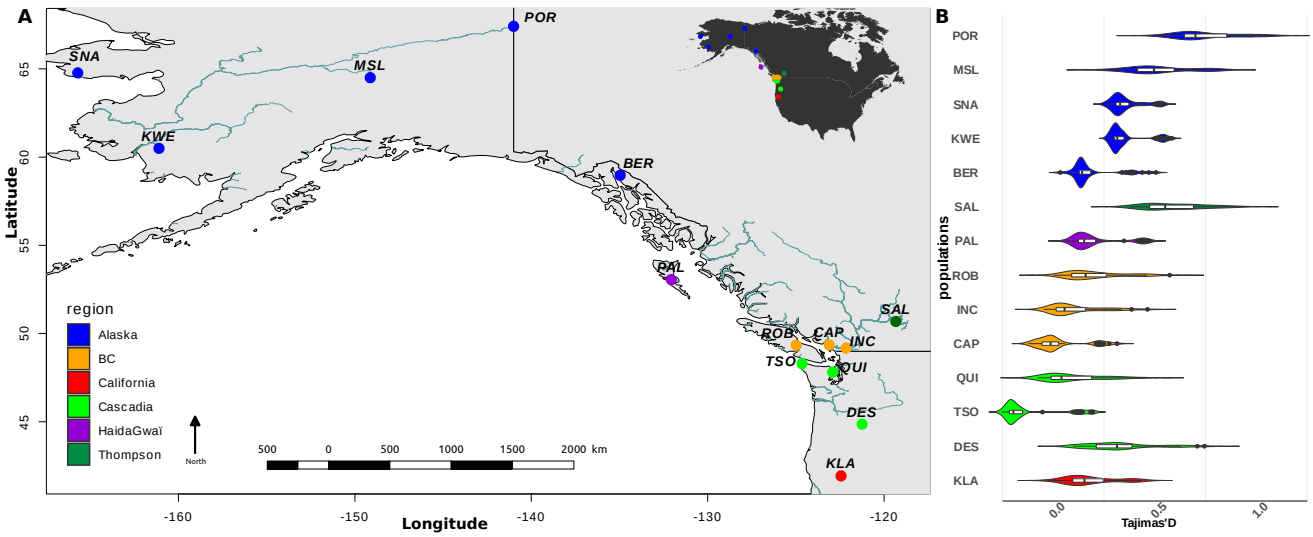

Figure 1: Sampling design and population size change inferred through whole genome sequences. *A) Sample site along the Pacific coast of North America. B) Average Tajima's D values computed along the genome and summarized for each sample locality. Populations are ordered from distance to the river mouth and from the South to the North (i) California, (ii) Cascadia, (iii) British-Columbia (BC), (iv) Haida-Gwaii, (v) Salmon River (Fraser Basin) and (vi) Alaska Abbreviation for each site is provided in Table S02. Tajima's D values vary positively as a function of the distance from the southernmost site (R^2 = 0.28, p <2e-16) and as a function of the distance to the river mouth (R^2 = 0.47, p <2e-16) reflecting the history of founder events from downstream to upstream sites.*

In line with the above observations, linkage disequilibrium (LD) extends to long distance in remote Alaskan populations (Porcupine and Mile Slough Rivers) and in the Salmon R. population located in the Thompson R. watershed **(Fig S2C)**. Accordingly, the LD decay, through a r^2 reaching half of its maximum value, is observed at 24.6 kb, 11.9kb and 6kb in these three populations, contrasting with values in the 0.5kbp range observed in populations with higher Ne as in the Tsoo-Yess River for instance (**Fig S2C**). These results show that there is a large variance in effective population size, likely associated with the strength of founder events. Distance to the southernmost site is not the only factor to consider, distance of spawning (reproductive) grounds to the sea is also highly significant (R^2 = *0.47, p <2e-16* for Tajima's D*)*. This suggests that more genetic drift occurs in upstream populations (Porcupine (POR), Mile Slough (MSL), Thompson (SAL) and Deschutes (DES)). These extended LD may reduce the efficacy of purifying selection and contribute to the accumulation of mildly deleterious alleles [43].

The analysis of Tajima's D (**Fig 1B**) aligned well with the above results, with decreased values observed in the South, and increased values towards the North which correlates with the distance to the South (R^2 = 0.28 p<2e-16). As for the **β**st, He, and LD decay, it is noteworthy that the southernmost populations (Klamath and Deschutes) displayed signatures consistent with lower N_e (**Fig S3**) resulting in both lower genetic diversity, higher LD and higher differentiation than populations further north (e.g. Tsoo-Yess R.). This likely reflects recent population decline, but also loss of genetic diversity associated with upstream river colonization.

We then investigated historical change in effective population size using the Sequentially Markovian Coalescent in SMC++ [44] based on the 71 whole genome sequences (**Fig S2D)** and both a median mutation rate of $0.8e^{-9}$ mutations/bp/generation and a mean of $1.25e^{-8}$ mutations/bp/generation as estimated for a

closely related salmonid; the Atlantic salmon (J. Wang, personal communication). This analysis revealed: i) an expansion of most populations approximately 12-20 KyA, concomitant with postglacial recolonization, ii) a slow and steady decline in the Thompson R. (**Fig S2D)**), and iii) a split time between all pairwise combinations of populations (median = 16,3 KyA, range = 6,7KyA - 61KyA, **Fig S3**) compatible with the onset of postglacial population expansion following glacial retreat [38]. Using the mean mutation rate yielded more recent estimates of split times (median = 9,6 KyA, **Fig S3**) (min = 5 KyA – max = 18 KyA). Overall, SMC++ results indicate that that until recently, all populations shared a similar demographic history (see also **Note S1**).

Range expansion and effective population size explained the mutation load

After confirming the main conclusions from [35], we estimated the mutation load of the populations using three different metrics (π_N/π_S, ω_{NA}, ω_A, and α). Remarkably, we observed a strongly significant and positive correlation of the π_N/π_S ratio of each local population and the distance to the southernmost site ($R^2 = 0.73$, $p<0.0001$, **Fig 2A**). This result aligns well with expectations under a model of expansion load (e.g. [17]). To further explore how the recent demographic history of Coho salmon populations may have affected the efficacy of selection we tested the hypothesis that range expansion led to a loss of the adaptive potential and an increase of the mutation load. We recovered a decreased efficacy of purifying selection associated with the multiple founding event and bottleneck which resulted in a decrease rate of adaptive substitutions (ω_A, $R^2 = 0.37$, $p = 0.013$) and increased rate of non-adaptive substitutions (ω_{NA}) as a function of the distance to the South ($R^2 = 0.57$, $p = 0.0011$, **Fig 2B**). The Salmon R. population (Thompson R. watershed), which has undergone a strong decline in recent population size, supposedly due to the sustained release of hatchery fish derived from few individuals and the Ryman-Laikre effect [45](see discussion below), also displayed a high rate of non-adaptive non-synonymous substitutions (ω_{NA}).

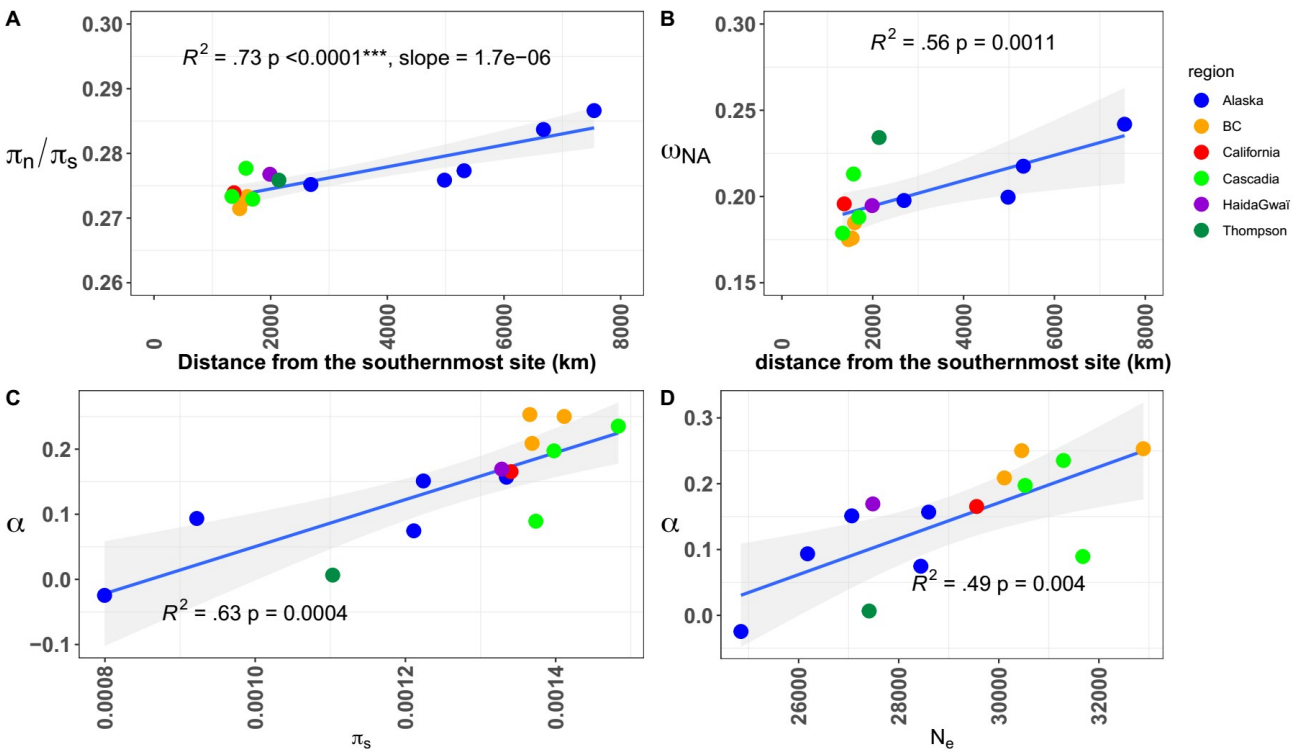

Figure 2: Demographic processes and life history explain variation in the genomic load in Coho salmon.
A) *correlation between* π_N/π_S *and the distance to the southernmost site in Coho Salmon (top).* ***B)*** *increased rate of nonsynonymous non-adaptive substitutions* (ω_{NA}) *in distant population of Coho salmon and in the bottlenecked Thompson River (dark green).* ***C)*** *Correlation between the rate of adaptation* α and synonymous diversity (π_S) as a proxy for effective population size. **D**) *Correlation between* α and *historical variation in the coalescent effective population size estimated from SMC++. Note that the distance from the southernmost site, was computed considering a Californian sample from our previous study, for which no WGS data were available.*

Based on these demographic observations, we also predicted a higher proportion of adaptive non-synonymous substitutions ($\alpha = \omega_A/(d_N/d_S)$) in populations with higher *Ne* and, conversely, a decreased value of α in populations with lower *Ne*. We observed a strong positive correlation between α and the synonymous nucleotide diversity (π_S) of each local population, which represents a good predictor of the long term (coalescent) *Ne* (R^2 *= 0.63, p = 0.0004, Fig 2C and Fig S5A for the* $\pi_N/\pi_S \sim \pi_S$ *relation)*. To more directly test the association with Ne, we correlated SMC++ based estimates of *Ne* (averaged over a 200KyA window) with πN/πS or α and recovered significant correlations for both (R^2 = 0.53, p <0.01, **Fig S5B; and** R^2 = 0.49, p = 0.004, **Fig 2D, respectively**). It is noteworthy that a good correlation between N_e from SMC++ and π_S was also observed (R^2 = 0.68, p = 0.00017). Significant correlations were also observed when considering the relationship between ω_A and π_S (R^2 =0.31, p = 0.0235), or ω_{NA} and π_S (R^2 = 0.34, p = 0.0172). Taken all together, these results shed light on the evolutionary consequences of allele surfing at expanding range margins, in particular regarding the loss of adaptive potential and the mutation burden.

One limitation associated with the use of π_N/π_S is that in populations that are not at mutation-drift equilibrium (say after a bottleneck) π_N will reach an equilibrium value faster than π_S because selected alleles

will be subject to higher negative selection and undergo a faster turnover [12,46,47]. Thus, the π_N/π_S may not always be the best predictor of the total burden of deleterious mutations as it is potentially affected by several factors including demography [12].

Highly vs. weakly deleterious loads in marginal populations

To circumvent this limitation and more directly infer the load, we counted and plotted the distribution of non-synonymous mutations classified according to i) their impact and expected consequences of fitness and ii) segregation patterns (i.e. total load composed of heterozygous and derived homozygous genotypes or fixed load composed of homozygous derived genotype) [10](**Fig. 3**). We identified a total of 671,032 SNPs in genes, among which 180,450 are synonymous, 199,476 missense and 6,390 Loss-of-Function (LoF) mutations. The total number of missense mutations was significantly more abundant in the more southern populations (i.e California, Cascadia, **Table S04** for Tukey-HSD p-values) than in northern ones (**Fig 3A**) while it was approximately constant from the South to the North regarding LoF mutations (**Fig 3B)** with no significant differences among populations **(Table S04**). This suggests a more efficient purifying selection against the more deleterious LoF mutations across all populations. In both cases, the southernmost populations displayed a higher load of mutations in heterozygous state especially when considering missense mutations (**Fig 3**), as expected due to their higher historical effective population size, favouring the segregation of recessive mutations hidden in a heterozygous states [35,39]. This contrast among southern vs northern-most populations was least pronounced when considering LoF mutations, suggesting again an efficient purifying selection against these more deleterious mutations. In this case, the most extreme Alaskan populations (Mile Slough R. and Porcupine R. from the Yukon and Salmon River from Fraser basin) were clearly the least loaded, reflecting a more efficient purging. Finally, the fixed load (i.e. count of derived homozygous sites) increased from the south to the north for both missense and LoF with the most extremes samples, including Mile Slough, Porcupine and Salmon rivers being the most significantly loaded (**Table S04**), as expected due to founder events and allele surfing for Alaskan samples and due to bottlenecks for the Salmon river. We also found a highly significant association with the derived load and the distance to the southernmost sites both for missense mutations ($R^2 = 0.790$, $p < 0.001$, **Fig 3A**) and for LoF mutation ($R^2 = 0.777$, $p < 0.001$, **Fig 3B**). Overall, even if these results support a slightly lower deleterious load on strongly deleterious mutations at the range margin, the most northern populations of Alaska exhibit a greater deleterious load considering slightly deleterious mutations, which is consistent with a reduced efficacy of selection in low N_e populations.

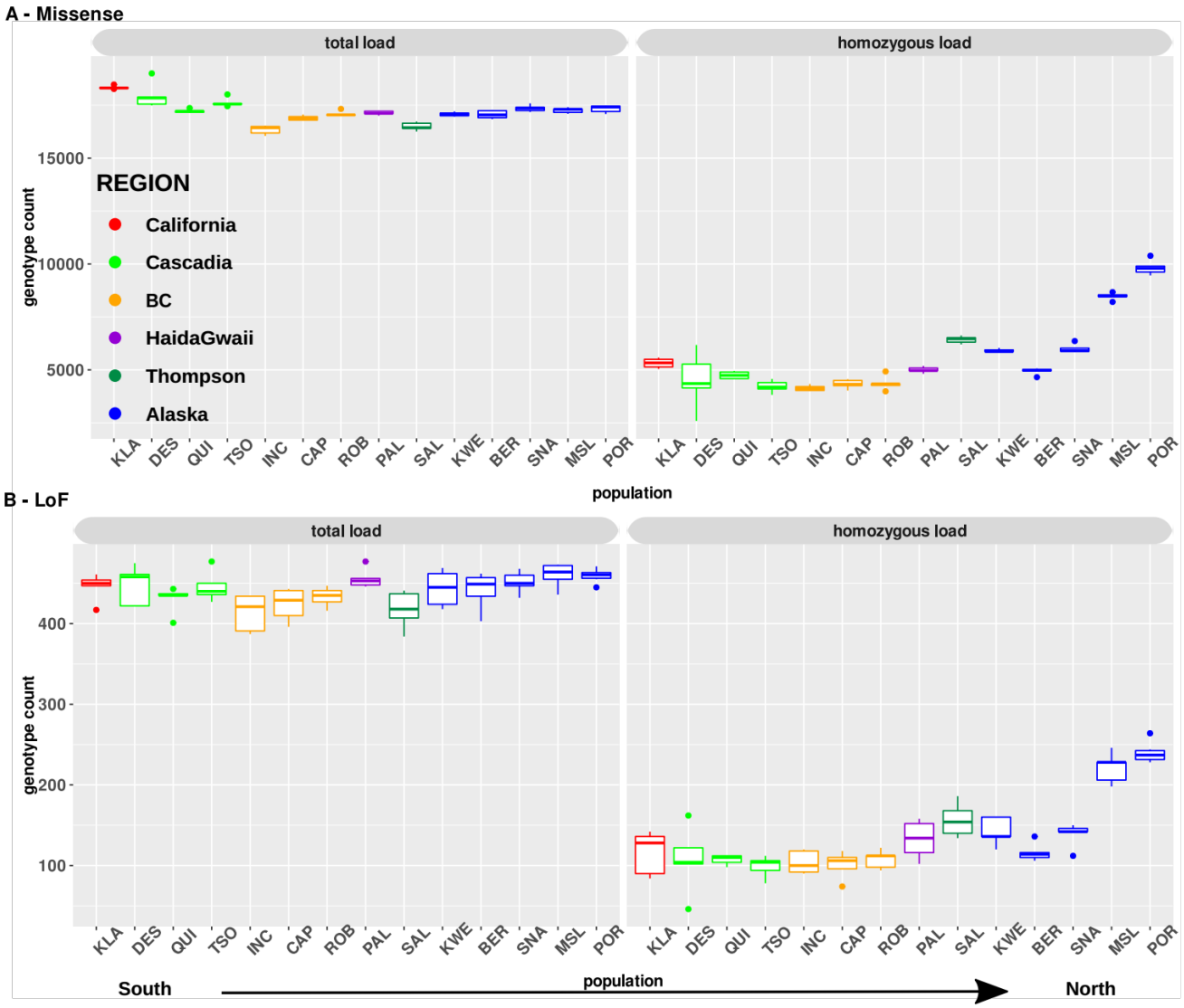

Figure 3: Number of deleterious alleles per river sorted from the South to the North *showing the total load (left panel), and fixed derived homozygous (recessive) load (right panel) for A) missense mutation and B) Loss-Of-Function (LoF) mutations. No strong differences are observed in the total load among populations. In contrast, significant differences were observed for the fixed load in populations at the expansion front both for missense mutations and LoF mutation and this was significantly correlated with the distance to the southernmost site. Each color represents a major regional group.* Abbreviation for each site is provided in Table S01.

Recombination rates shape the deleterious mutational landscape

In addition to the spatial structure associated to the postglacial recolonization, we investigated the genome-wide variation in mutation load. For instance, such variation could be associated with the occurrence of structural variants (e.g. chromosomal inversions) which may incur a significant load because deleterious recessive mutations may freely accumulate in the absence of recombination, as observed for instance in sex chromosome and related supergene-like architecture [48]. To test this hypothesis, we used the GC content at third codon position (GC3 hereafter) as a proxy of the rate of recombination [27,49] (see **Sup. Note S2 for an explanation**). We observed strong correlations (R^2 range = 0.938 – 0.955, p<0.001, except again for MSL, with R^2 = 0.252) between levels of GC3 and π_N/π_S ratio among all Coho salmon populations (**Fig 4A**). An

analysis focused on GC conservative sites (not affected by gGBC) revealed similarly strong patterns across all Coho salmon populations (**Fig 4B**, 0.665< R^2 < 0.876). π_N/π_S ratios considering all sites or GC-conserved only are highly correlated (r = 0.909). We tested the generality of this relationship using four other salmon species and found strikingly similar pattern in Chinook salmon (R^2 GC3 ~ π_N/π_S = 0.9338, p = 1e-11) Sockeye and Kokanee salmon (range: 0.81 < R^2 < 0.96, p <0.001), as well as Rainbow trout populations (range: 0.95 < R^2< 0.96, p <0.001, **Fig 4C**).

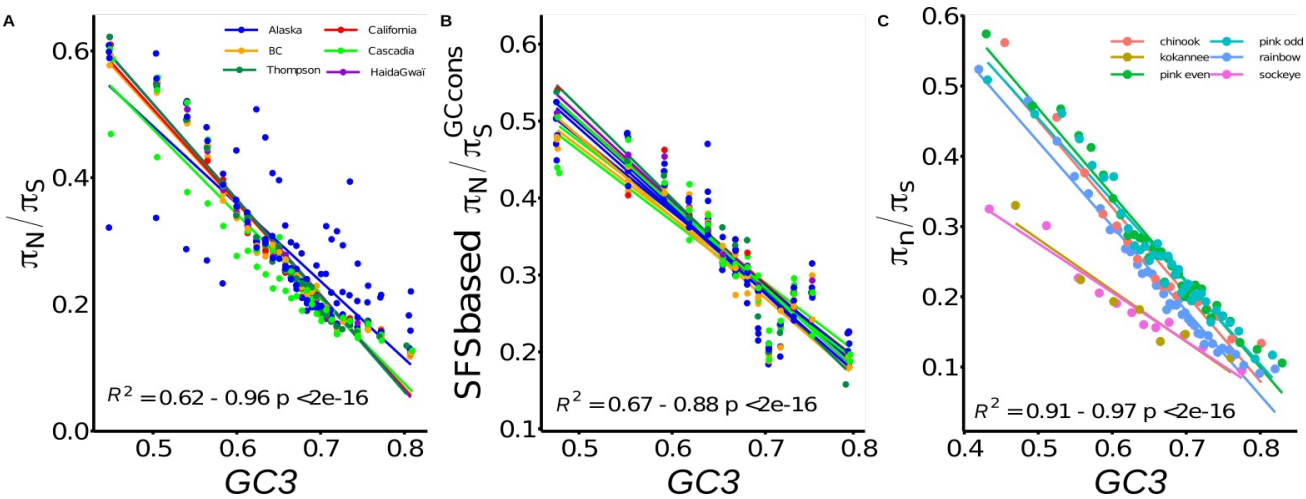

Figure 4) The deleterious load is determined by variation in GC3 content in multiple salmonids
*A) Correlation between the deleterious load (π_N/π_S) and GC content at third codon position in multiple Coho salmon populations. Each line represents the values color by major regional group (See **Fig S8A** for detail by population). B) Correlation between the deleterious load (π_N/π_S) and GC content at third codon position in the same Coho salmon populations controlling for gGBC by using GC-conservative site based on site frequency spectrum. Each line represents the values color by major regional group. C) Correlation between the deleterious load (π_N/π_S) and GC content at third codon position in sockeye salmon ecotypes (sockeye and kokanee), chinook, pink salmon and in rainbow trout. Averages are provided for rainbow trout, salmon and kokanee at the species scale, see **Fig S8B-C** for detail by population.*

Moreover, this also highlights the effect of recombination on the efficiency of purifying selection [50]. The samples at the expansion front (Mile Slough R., Porcupine R., Snake R.) in Alaska were the populations with the lowest correlations between GC3 and π_N/π_S. It is therefore possible that in these populations with reduced effective population size, the efficacy of selection tends to be reduced even more strongly in regions of low recombination, leading to a decoupling of the relationship between GC3 and π_N/π_s. Accordingly, there was no correlation between the maximum π_N/π_s value (i.e. from the least recombining region, or lower GC3 value) and the distance to the source (p = 0.79) whereas the relationship was significant when considering the minimum π_N/π_s value (p = 0.0006, R^2 = 0.60). Finally, a linear model between the slope of the regression between π_N/π_s and GC3 and the distance to the ancestral source populations indicated a lack of relationship between these two variables (p=0.133, **Fig S09**). A significant association may have been predicted under a linear coupling between recombination efficiency and demographic expansions.

Regions of residual tetraploidy revealed drivers of the load in Coho salmon

We tested the hypothesis that regions with residual tetraploidy, which in theory should display a two-fold *Ne* also exhibit a reduced load due to more efficient purifying selection. Contrary to this expectation, increased deleterious load was observed in regions of residual tetraploidy (3,700 genes) as compared to 200 randomly sampled set of 4,000 genes from diploid regions (**Fig 5A**, red dotted line mean π_N/π_S > 0.35, diploid region π_N/π_S < 0.30). Given that we also observed lower levels of recombination in regions with residual tetraploidy compared to re-diploidized genomic regions (**Fig S10,** p < 0.0001, W Mann-Whitney = 3.04e7), our results suggest that this higher π_N/π_S could be mostly due to lower recombination rates. Another genomic consequence of this lower recombination rate is a higher load of transposable elements[51]. When computing the relative length of TE, that is the length of TE corrected by the chromosome length (see methods), we found a significant enrichment of TE in the regions of residual tetraploidy as compared to diploid chromosomes (**Fig 5B**, p <0.0001, WMann-Whitney = 1.36e10). This tendency was also observed across the different TE categories (**Fig S14)**. To more directly test the N_e-effect hypothesis, we eliminated the effect of the recombination rate by comparing the load across similar bins of GC in diploid vs. tetraploid regions. Interestingly, the π_N/π_S was systematically higher in diploid regions than in the tetraploid ones after excluding the class with the lowest GC content (**Table S05**) indicating that the load was significantly higher in diploid compared to tetraploid regions, with the notable exception of regions with extremely reduced recombination. Therefore, it is still possible that increased efficacy of selection is at play in recombining regions of residual tetraploidy, therefore following the general hypothesis of higher effective size in these regions.

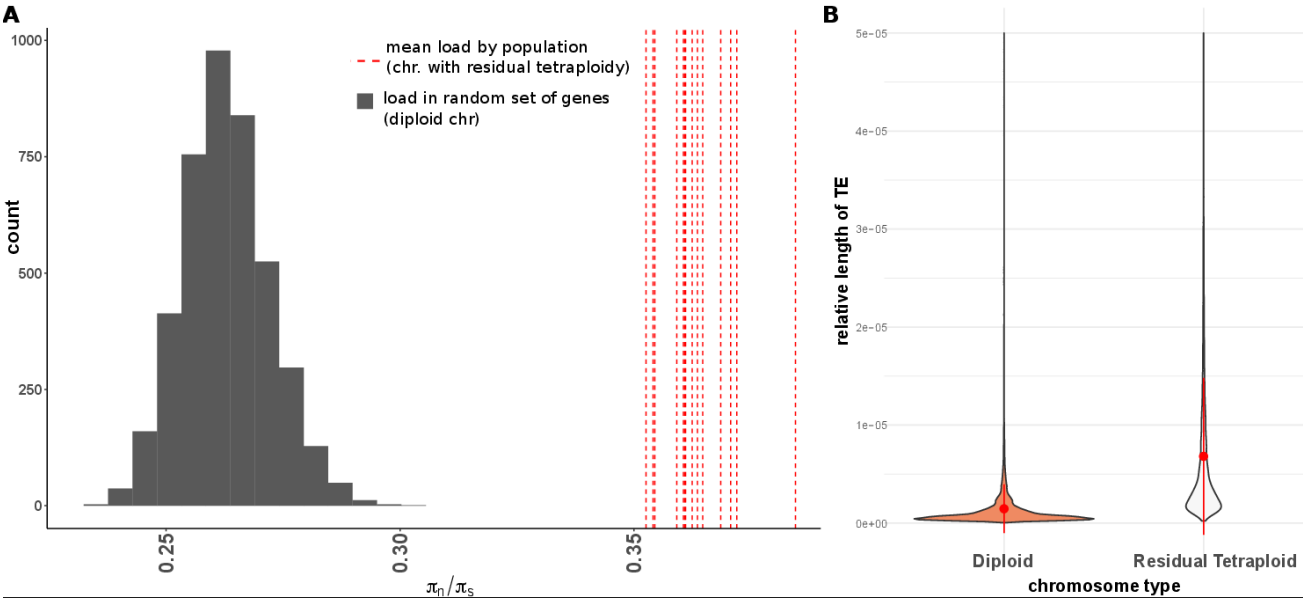

Figure 5: Increased load in region of residual tetraploidy. *A) Distribution of π_N/π_S ratio when considering all genes in regions of residual tetraploidy (red lines) for each population compared to a set of 200 randomly generated samples of 4,000 genes in the rest of the genome (gray histogram).* ***B)*** *Boxplot showing that regions of residual tetraploidy displayed significantly longer transposable elements when compared to diploid regions. Orange = diploid chromosome. Gray = chromosome with residual tetraploidy. Red point = mean +/- 1*sd.*

Conclusion

The role of evolutionary processes including genetic drift and recombination in affecting the efficacy of natural selection and therefore evolutionary potential is a central question, both with fundamental and applied consequences for biodiversity. Using whole genome sequences from Coho salmon and several closely related salmonid species, we investigated the determinants of the deleterious load across the geographic and genomic space.

Using population genomics analyses, we investigated the evolutionary consequences of recent demographic events, particularly regarding N_e in Coho salmon. First, our results provided remarkable empirical support for gene surfing across North America. This result confirms previous early findings regarding allele surfing in this system and more specifically expansion load in Alaskan populations based on a lower resolution approach (Genotyping By Sequencing)[35].Performing whole-genome sequencing analyses further allowed providing strong empirical support for lower Ne at the northern range expansion front, which induced two main evolutionary consequences: a surf of slightly deleterious mutations and a putative reduction of the adaptive potential. Overall, this result provides remarkable support for the nearly neutral hypothesis of molecular evolution [13]. We further demonstrated that one population, the Salmon River (Thompson R. watershed) displayed an increased fixed load and decreased selection efficacy compared to southern populations. Previous studies showed that this population is highly divergent from any others [35,36,52] and displayed an increased rate of long runs of homozygosity [38]. It has been subjected to extensive hatchery enhancement (from a single population) to circumvent the decline of Coho salmon in the Thompson drainage [53]. From a theoretical point of view, if only a few captive individuals were used as parents for subsequent release, then an increase in inbreeding and reduction in effective population size is a process called the Ryman-Laikre effect [45]. This process is expected to generate a strong decrease in the effective population size of supplemented populations [54]. As a consequence it is entirely possible that the decreased efficacy of selection, increased segregation of non-synonymous non-adaptive substitutions and increased fixed load could be explained by both long-term evolution at low N_e and by the Ryman-Laikre effect associated with hatchery enhancements. Regardless, this suggests that careful enhancement needs to be performed with a diversity of parents to maximize genetic diversity of supplemented populations. This result has therefore implications for fisheries enhancement, as well as genetic rescue programs aiming at reducing the inbreeding load of declining populations and restoring the fitness of these populations.

Thanks to the increasing availability of salmonids dataset, we also reported that the magnitude of the load in Coho salmon was modest as compared to the observed variation among species of Pacific salmon (**note S2**). This could question the fitness consequences of these slightly deleterious loads and call for further investigations of the relationship between effective population size, census size, fitness effect and the conservation of populations (see discussion in [27]).

Since the use of π_N/π_S ratio can be criticized [12,55,56] especially when quantifying the load within species, we used additional metrics to quantify the efficacy of selection (i.e. α and related metrics [26]) as well as more direct count of putatively deleterious mutations, as advocated by other [12,18]. We further look at the effect of predicted Loss-of-Function mutations, whose fitness effects are expected to be strong. In this case, we found small differences in the total load among Coho salmon populations, but a linear increase in the fixed load as a function of the distance to the Southernmost sites. As argued by Simons and Sella [12] however, this increased load is also to be expected due to founder events, reduced effective populations, and fixation of neutral mutations due to increased genetic drift. Yet, the same pattern is also observed when considering LoF mutations, therefore our results are consistent with an increase of the deleterious burden due to allele surfing, especially in the most isolated Alaskan samples, as previously suggested [35]. Using these data, we also demonstrated that the π_N/π_S ratio was strongly correlated to the GC content at third codon position, a good proxy for the load (but see **note S3**). The strong correlation between π_N/π_S and GC3 was repeatedly observed across different salmonid species using both a sequence based estimate of π_N/π_S and an estimate based on GC-conservative site (non-affected to gBGC) [49]. Therefore, these results indicated that recombination plays a key role in the efficacy of selection along the genome in salmonids. More generally, these results give empirical support to many theoretical studies about the accumulation of slightly deleterious mutations in non-recombining regions [28,30]. Similarly, increased prevalence of deleterious mutations in low recombining regions have been reported in plants [57] and in human populations [32,58,59]. In particular, recent work in human populations have shown that both variation in demographic history (i.e. change in effective population size) and recombination rate variation are affecting allele specific expression of harmful mutations [32]. These authors showed that allele specific expression more efficiently cause under expression of harmful mutations in normally and highly recombining regions compared to low recombining regions and that this varies among human populations with different demographic histories.

In line with the key role of recombination rate, we observed that only regions of residual tetraploidy with extremely low recombination rate (lower GC content) displayed an increased load in Coho salmon, accumulated more transposable elements suggesting efficient selection in more “normally” recombining region of residual tetraploidy. The detailed consequences of ongoing rediploidization have been extensively studied at the regulatory levels elsewhere, with support for an increased load (higher d_N/d_S and TE load) in duplicated genes undergoing lower expression [60]. To sum up, our results concur with predictions from the nearly neutral theory of molecular evolution [13] in which slightly deleterious mutations are effectively neutral and purged effectively by recombination except perhaps in populations at the extreme of the expansion front.

Finally, our results have important implications for conservation practices. We showed that the additive load

is approximately constant, indicative of efficient purging across populations for both missense and LoF, similar to some recent studies [61–64]. However, evidence from π_N/π_S, rate of adaptation, and recessive load indicate that population at higher latitude from the Yukon watershed (Mile Slough R., Porcupine R.) or from the bottlenecked Salmon R. have not entirely purged the most deleterious mutations, including missense and LoF mutations, which may impose a fitness cost to these populations and to the Salmon River in particular. Further empirical evidence for a causal link between the putative fitness cost of LoF mutations and adaptive phenotypic variation will be necessary to validate our observations. Indeed, there is no easy translation between our population genetic inferences and their consequences at the phenotypic level and fitness consequences. With this caveat in mind, our results could still guide practices in supplementation programs. For instance, choosing a diversity of parents from moderately differentiated populations of modest size may help increase the levels of heterozygosity and mask the expression of recessive deleterious mutations [39]. This strategy could reduce the occurrence of deleterious alleles and counteract the Ryman-Laikre effect described above. Moreover, populations (e.g. most upstream Alaskan populations) for which our results suggest a reduced adaptive potential are also the most strongly exposed to rapid climate change, as the rate of temperature increase is the most rapid at higher latitudes [65]. In all cases, maximizing the connectivity among populations and limiting habitat degradation appears as fundamental strategies to maintain high effective population size and increase the adaptive potential of Coho salmon to the multiple ongoing anthropogenic pressures [66]. More generally, this question of how the genomic load translates into a reduced fitness and impacts population persistence is an unsolved issue that deserves more attention in future research. As well, how best to manage declining populations and guide conservation policies in these conditions to minimize the load and/or maximize genetic diversity is another debated issue [39–42]. In the meantime, while conservation genomics undoubtedly have a major role for the short-term preservation of endangered species, this should not override the crucial need for reducing human impacts on natural ecosystems to preserve biodiversity over long time scales [67].

Methods

Sampling design for Coho salmon

We sampled 71 individuals representing 14 populations from California to Alaska (**Table S01**). A set of 55 individuals was sequenced on an Illumina HiSeq2500 platform [38] and a set of 16 additional individuals was sequenced on a NovaSeq6000 S4 platform using paired-end 150 bp reads. Reads were processed using fastp for trimming [68], and mapped to the latest coho reference genome (https://www.ncbi.nlm.nih.gov/assembly/GCF_002021735.2/) using bwa mem v2 [69]. Reads with a minimum quality below 20 were discarded with samtools v1.7. Duplicates were removed with picard (http://broadinstitute.github.io/picard/). SNP calling was performed using GATK v4.2 [70] using our pipeline

available at github.com/quentinrougemont/gatk_haplotype/. We generated a Haplotype gVCF for each sample individually, combined all gVCF and then performed a joint genotyping. We checked the variants quality score of our data and filtered our genotypes according to their quality following GATK best practices and based on quantiles distributions of quality metrics. We excluded all sites that did not match the following criterion: MQ < 30, QD < 2, FS > 60, MQRankSum < -20, ReadPosRankSum < 10, ReadPosRankSum > 10. We excluded multiallelic SNPs, as well as indels. Genotypes with a depth lower than 6 or higher than 100 reads were also excluded to remove low confidence genotypes potentially associated to paralogs. Finally, we also generated a separate vcf file using the emit-all-site option to obtain a full vcf to reconstruct sequence data (see the **Genetic load estimation in Coho salmon and related species** section).
A total of 14,701,439 SNPs were identified without missing data. Population structure was evaluated usinga principal component analysis (PCA) was performed using Ade4 [71] on a set of LD-pruned SNPs without missing data (1,739,037 SNPs) identified stringently with plink1.9 [72] (command indep-pairwise 100 50 0.1).

Outgroup dataset. Sockeye salmon published by [73] were retrieved from NCBI PRJNA530256. Samples with a high number of individuals per ecotype (Sockeye and Kokanee) were chosen from the NCBI table. We retained a total of 5 Kokanee populations and 5 Sockeye populations (from Fraser & Columbia watershed described in T**able S01 and Table S03**). 3 Chinook salmon samples were provided by B. Koop (also available at NCBI PRJNA694998. 11 samples of "even" and 10 samples "odd" pink salmon (*O. gorbuscha*) were also downloaded from PRJNA 556728 and included. Here "even" and "odd) refers to Salmon returning to their natal rivers in "even" and "odd" years to spawn, leading to a temporal isolation of these ecotypes [74]. These were indeed clearly separated based on a PCA. Finally, a set of rainbow trout available from NCBI PRJNA386519 were used (n = 19 from 3 random populations showing genomic differentiation based on a PCA). Each sample was downloaded and mapped onto its species' reference genome downloaded from NCBI and using the exact same procedure as described above relying on fastp, bwa-mem2, picard and GATK 4.2.5.0 to generate a final vcf filter based on usual GATK quality criteria and variance in sequencing depth. For each species we then quantify the load using the π_N/π_S ratio using the procedure described below for Coho salmon. For the Sockeye-Kokanee ecotype, we specifically tested the hypothesis that the non-resident ecotype displayed a higher load due to a lower effective population size (**note S2, table S03**). For all other species we tested the relationship between π_N/π_S and GC3 (see below). For Sockeye-Koanee ecotype we further tested for differences in effective population sizes with smc++ using the procedure described below for Coho salmon.

Ancestral and derived alleles identification

To accurately recover the ancestral state we used three outgroup species including the chinook salmon and rainbow trout sample (see above, n = 5 for rainbow trout) plus data from Atlantic salmon (n = 5, *Salmo salar*

SRP059652). Each individual was aligned against the Coho salmon V2 genome (GCF_002021745.2) using GATK with the same procedure as above and calling every SNP using the EMIT_ALL_SITES modes. We then determined the ancestral state of each SNPs if 1) the SNP was homozygous in at least 90% of the individuals from our three outgroups, and 2) match one of the two alleles identified in Coho salmon. Otherwise, the site was inferred as missing and was not used in subsequent analyses of the load. In addition, we reconstructed a consensus ancestral fasta sequence using the doFasta option from angsd [75]. This was used for demographic reconstruction detailed in **Note S1.**

Demographic reconstruction

We first tested our prediction that genetic diversity (H_e) decreases towards the North following our «out of Cascadia» model previously inferred [35,36]. Conversely, we verified that the β_{ST} coefficient, a measure of genetic differentiation [76] increases towards the North, as expected due to isolation by distance. β_{ST} and gene diversity (H_S) were measured using the hierfstat R package [77] and oceanographic distances were computed using the marmap package [78]. Waterway distance was computed using ArcGIS.

Under postglacial expansion from a single refugia, the general hypothesis would be that all sampled populations should follow a common temporal trajectory of a population decline (bottleneck due to fonder event by few individuals) followed by a (strong) increase in *Ne* corresponding to the expansion phase. To test this hypothesis, we inferred temporal changes in *Ne* used SMC++ [44]. SMC++ works similarly to the PSMC model but takes into account the Site Frequency Spectrum (SFS) and is better suited to large sample sizes. Estimates of population size changes were performed for all populations. To validate the fact that expansions are indeed postglacial, splitting time was estimated between all pairs of samples from different geographic areas based on the joint SFS (n = 75 pairwise comparisons). A generation time of 3.5 years and a mutation rate of $9e^{-9}$ mutation/bp/generation were applied (corresponding to the median substitution rate inferred in Atlantic salmon, Wang, Personal communication). We also compared the results to the mean substitution rate of $1.25e^{-8}$ mutation/bp/generation also inferred by Wang.

In addition, pairwise linkage disequilibrium provides valuable information regarding population size and inbreeding. We computed the squared correlation coefficient between genotypes in each sample and all populations separately using popLDdecay [79]. We used a MAF of 5% in each population, keeping between 3.7 and 6.4 million SNPs with populations having undergone stronger bottleneck/founding events displaying the lowest amount of variation. We estimated LD decay by plotting LD (R_2) against physical distance measured in base pairs in ggplot2 [80] package in R.

We observed slight discrepancies between our SMC++ estimates of divergence time and our previous work based on the site frequency spectrum [35]. To investigate this, we performed a new set of inference based on the unfolded joint site frequency spectrum (jSFS) using ∂a∂i [81] and a new set of refined as detailed in **supp. Note S1**. We also explored broad relationship among populations with treemix [82]. We fitted a model

with an increasing number of migration edges. We chose a model with K = 3 migration edges as all edges displayed significant p-value and captured a high proportion of explained variance. Fitting more edges decreased the p-value without really improving the fit to our data.

Population-scaled recombination rate

Statistical phasing of the Coho whole genome sequences was performed using the shapeit software [83] considering all individuals at once. We then estimated effective recombination rates (ρ=4.*Ne*.r where r represents the recombination rate per generation and *Ne* is the effective population size) along the genome using LDHat [84]. Phased genotypes were converted into LDHat format using vcftools after excluding SNPs with MAF < 10% since rare variant are not informative for such inferences. Following the guidelines of the LDHat manual), the genome was split in chunks of 2,000 SNPs with overlapping windows of 500 SNPs. We measured recombination rates independently for each river, as well as among all populations. For this latter analysis, we excluded the population from the Thompson R. watershed, which is too divergent from the rest of the samples. Difference in distribution of population scaled recombination was visualized using violin plot and statistically tested using a Mann-Whitney test.

Genetic load estimation

The approach developed in [27] was used to reconstruct fasta sequences for each individual. For each species, CDS from each reconstructed sequence were extracted using the gff files available with the reference genome to estimate the nucleotide diversity (π). We also concatenated the CDS sequences into different classes according to their length and computed π_N and π_S over 4-Mb concatenated gene space windows. Such large windows reduce the stochasticity due to the low π_s values in Coho salmon. We also computed the GC content at third-codon positions of protein coding genes (GC3) which has been shown to be an accurate proxy of local recombination rate in other species and are generally silent [85,86]. To compute π_N/π_S values we sorted genes by ascending GC3 values, which enabled us to obtain a ratio based on genes with similar GC3 values. Moreover, we also used the site frequency-based approach proposed by Rousselle et al. to estimate the π_N/π_S ratios. This approach enabled us to compute SFS separately for GC conservative sites (A<->T and C<->G mutations), that is, not affected by gBGC.

We then tested the correlation between population-scaled recombination rate (ρ), measured over 250-kb windows, and GC3. To do so, we assigned a ρ value for each gene according to its position into each 250-kb window. Then we averaged ρ values across the genes falling in the same window of 4 Mb concatenated gene space (see above) in order to compare the median population recombination rate estimates and GC3. Finally, we measured the correlation between GC3 and π_N/π_S using linear model. We replicated these analyses considering only genes (n = 3,500) and SNPs in regions of residual tetraploidy (8% of the genome).

DFE estimation and rate of adaptation in Coho salmon

We estimated the rate of non-adaptive and adaptive synonymous substitutions (ω_{NA} and ω_A, respectively; with $\omega_A = d_N/d_S - \omega_{NA}$) and the proportion of amino-acid substitution that result from positive selection ($\alpha = \omega_A/(d_N/d_S)$). To do so, we used the method implemented in Grapes v1.0 [12] which builds upon the approach of [87]. Grapes models the effect of favourable mutations on non-synonymous site frequency spectrum (SFS) while accounting for various confounding factors distorting the SFS (demographic change, linked selection, genotyping errors, SNP misorientation). More precisely, we model the distribution of fitness effect (DFE) using a negative Gamma distribution to the synonymous and non-synonymous SFS (model GammaExpo in Grapes). Fitted parameters of the DFE were used to compute the expected dN/dS under near neutrality, which was compared to the observed dN/dS to estimate α, $\omega_{NA,}$ ωa.

Identifying potential deleterious non-synonymous alleles

We tested the difference in count of non-synonymous mutations in each local population of Coho salmon across non-synonymous missense mutations (putatively deleterious) and Loss of Function (LoF) mutations (likely to be strongly deleterious) identified with SNPeff. We analyzed data in two ways: first we counted the total number of putative homozygous derived deleterious alleles (recessive load) per individual as well as the total number of deleterious alleles (both in homozygous and heterozygous states, additive load) using: $Ntotal = 2 \times N_{homo} + N_{hetero}$ [34]. These individual values were then averaged per population (Figure 3). We tested for differences in the distribution of these counts among populations using an anova followed by a TukeyHSD test. Obtained p-value were corrected using a Bonferroni correction. We also computed mean derived allele frequencies (DAF) in all sampling locations and across mutation categories (synonymous, non-synonymous missense and LoF). We also used the popular Provean [88] software based on a random set of non-synonymous mutations (but see **Note S4** for a brief discussion regarding the limitations). Mutations with a score of -2.5 or lower were considered as potentially deleterious. These Provean results were then compared with results from non-synonymous mutations **(Fig S15**).

Differences in load for region of residual tetraploidy in Coho salmon

π_N/π_S comparison

In the reference Coho genome, chromosomes from 1 to 30 correspond to regions that were rediploidized, while chromosomes from 31 to 38 correspond to regions of tetraploidy. We took advantage of this specificity regarding chromosome evolution to contrast the load for the 3700 genes in regions of residual tetraploidy (average across all genes for each population). for diploid regions, we generated 200 datasets of 4,000 genes randomly sampled and then estimated the load for each of these datasets.

TE annotations

We used the TEs annotation file from repeatmasker [89] (made available on NCBI for the reference genome [38]) and tested for difference in the length of TEs between diploid and tetraploid regions, after correcting for the difference in chromosome length.

Data availability

New sequencing data will be deposited on NCBI SRA. All vcf will be deposited online. All scripts to reproduce our analyses will be deposited online at https://github.com/QuentinRougemont/selection_efficacy.

Acknowledgements

We are grateful to Alysse Perreault-Payette and Bérénice Bougas for help in DNA extraction of additional individuals for whole genome sequencing. We thank Benoit Nabholz for initial development of the π_N/π_S C++ app we used here. We are grateful to the genomic platform at Compute Canada (Canada), IBIS (Canada), and MBB (Montpellier, France) for providing computational capacity as well as Bird (Nantes, France) for data storage.

References:

1. 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

2. Hallatschek O, Hersen P, Ramanathan S, Nelson DR. Genetic drift at expanding frontiers promotes gene segregation. Proc Natl Acad Sci U S A. 2007;104: 19926–19930. doi:10.1073/pnas.0710150104

3. Excoffier L, Foll M, Petit RJ. Genetic Consequences of Range Expansions. Annual Review of Ecology, Evolution, and Systematics. 2009;40: 481.

4. Slatkin M, Excoffier L. Serial Founder Effects During Range Expansion: A Spatial Analog of Genetic Drift. Genetics. 2012;191: 171–181. doi:10.1534/genetics.112.139022

5. Klopfstein S, Currat M, Excoffier L. The Fate of Mutations Surfing on the Wave of a Range Expansion. Molecular Biology and Evolution. 2006;23: 482–490. doi:10.1093/molbev/msj057

6. Travis JMJ, Münkemüller T, Burton OJ, Best A, Dytham C, Johst K. Deleterious Mutations Can Surf to High Densities on the Wave Front of an Expanding Population. Molecular Biology and Evolution. 2007;24: 2334–2343. doi:10.1093/molbev/msm167

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

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

9. 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

10. Bertorelle G, Raffini F, Bosse M, Bortoluzzi C, Iannucci A, Trucchi E, et al. Genetic load: genomic estimates and applications in non-model animals. Nat Rev Genet. 2022;23: 492–503. doi:10.1038/s41576-022-00448-x

11. Chen J, Glémin S, Lascoux M. Genetic Diversity and the Efficacy of Purifying Selection across Plant and Animal Species. Molecular Biology and Evolution. 2017;34: 1417–1428. doi:10.1093/molbev/msx088

12. 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

13. Ohta T. Slightly Deleterious Mutant Substitutions in Evolution. Nature. 1973;246: 96–98. doi:10.1038/246096a0

14. Hedrick PW, Garcia-Dorado A. Understanding Inbreeding Depression, Purging, and Genetic Rescue. Trends in Ecology & Evolution. 2016;31: 940–952. doi:10.1016/j.tree.2016.09.005

15. Lohmueller KE. The Impact of Population Demography and Selection on the Genetic Architecture of Complex Traits. Williams SM, editor. PLoS Genet. 2014;10: e1004379. doi:10.1371/journal.pgen.1004379

16. Lohmueller KE, Indap AR, Schmidt S, Boyko AR, Hernandez RD, Hubisz MJ, et al. Proportionally more deleterious genetic variation in European than in African populations. Nature. 2008;451: 994–997. doi:10.1038/nature06611

17. 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

18. 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

19. Do R, Balick D, Li H, Adzhubei I, Sunyaev S, Reich D. No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. Nat Genet. 2015;47: 126–131. doi:10.1038/ng.3186

20. 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

21. 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

22. Takou M, Hämälä T, Koch EM, Steige KA, Dittberner H, Yant L, et al. Maintenance of Adaptive Dynamics and No Detectable Load in a Range-Edge Outcrossing Plant Population. Molecular Biology and Evolution. 2021;38: 1820–1836. doi:10.1093/molbev/msaa322

23. Rousselle M, Simion P, Tilak M-K, Figuet E, Nabholz B, Galtier N. Is adaptation limited by mutation? A timescale-dependent effect of genetic diversity on the adaptive substitution rate in animals. PLoS Genet. 2020;16: e1008668. doi:10.1371/journal.pgen.1008668

24. Lanfear R, Kokko H, Eyre-Walker A. Population size and the rate of evolution. Trends Ecol Evol. 2014;29: 33–41. doi:10.1016/j.tree.2013.09.009

25. McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila. Nature. 1991;351: 652–654. doi:10.1038/351652a0

26. Galtier N. Adaptive Protein Evolution in Animals and the Effective Population Size Hypothesis. PLOS Genetics. 2016;12: e1005774. doi:10.1371/journal.pgen.1005774

27. Leroy T, Rousselle M, Tilak M-K, Caizergues AE, Scornavacca C, Recuerda M, et al. Island songbirds as windows into evolution in small populations. Current Biology. 2021 [cited 3 Mar 2021]. doi:10.1016/j.cub.2020.12.040

28. Felsenstein J. The evolutionary advantage of recombination. Genetics. 1974;78: 737–756. doi:10.1093/genetics/78.2.737

29. Nordborg M, Charlesworth B, Charlesworth D. The effect of recombination on background selection. Genet Res. 1996;67: 159–174.

30. Muller HJ. The relation of recombination to mutational advance. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis. 1964;1: 2–9. doi:10.1016/0027-5107(64)90047-8

31. Hill WG, Robertson A. The effect of linkage on limits to artificial selection. Genetics Research. 1966;8: 269–294. doi:10.1017/S0016672300010156

32. Harwood MP, Alves I, Edgington H, Agbessi M, Bruat V, Soave D, et al. Recombination affects allele-specific expression of deleterious variants in human populations. Sci Adv. 8: eabl3819. doi:10.1126/sciadv.abl3819

33. Sergeant CJ, Sexton EK, Moore JW, Westwood AR, Nagorski SA, Ebersole JL, et al. Risks of mining to salmonid-bearing watersheds. Science Advances. 2022;8: eabn0929. doi:10.1126/sciadv.abn0929

34. 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

35. Rougemont Q, Moore J-S, Leroy T, Normandeau E, Rondeau EB, Withler RE, et al. Demographic history shaped geographical patterns of deleterious mutation load in a broadly distributed Pacific Salmon. PLOS Genetics. 2020;16: e1008348. doi:10.1371/journal.pgen.1008348

36. Rougemont Q, Xuereb A, Dallaire X, Moore J-S, Normandeau E, Perreault-Payette A, et al. Long-distance migration is a major factor driving local adaptation at continental scale in Coho salmon. Molecular Ecology. n/a. doi:10.1111/mec.16339

37. Macqueen DJ, Johnston IA. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proc Biol Sci. 2014;281. doi:10.1098/rspb.2013.2881

38. Rondeau EB, Christensen KA, Minkley DR, Leong JS, Chan MTT, Despins CA, et al. Isolation-by-distance and population-size history inferences from the coho salmon (Oncorhynchus kisutch) genome. bioRxiv; 2022. p. 2022.06.14.496192. doi:10.1101/2022.06.14.496192

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

40. Teixeira JC, Huber CD. The inflated significance of neutral genetic diversity in conservation genetics. PNAS. 2021;118. doi:10.1073/pnas.2015096118

41. García-Dorado A, Caballero A. Neutral genetic diversity as a useful tool for conservation biology. Conserv Genet. 2021;22: 541–545. doi:10.1007/s10592-021-01384-9

42. Kardos M, Armstrong EE, Fitzpatrick SW, Hauser S, Hedrick PW, Miller JM, et al. The crucial role of genome-wide genetic variation in conservation. Proceedings of the National Academy of Sciences. 2021;118: e2104642118. doi:10.1073/pnas.2104642118

43. 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

44. Terhorst J, Kamm JA, Song YS. Robust and scalable inference of population history from hundreds of unphased whole genomes. Nature Genetics. 2017;49: 303–309. doi:10.1038/ng.3748

45. Ryman N, Laikre L. Effects of Supportive Breeding on the Genetically Effective Population Size. Conservation Biology. 1991;5: 325–329. doi:https://doi.org/10.1111/j.1523-1739.1991.tb00144.x

46. Gravel S. When Is Selection Effective? Genetics. 2016;203: 451–462. doi:10.1534/genetics.115.184630

47. Brandvain Y, Wright SI. The Limits of Natural Selection in a Nonequilibrium World. Trends in Genetics. 2016;32: 201–210. doi:10.1016/j.tig.2016.01.004

48. Berdan EL, Blanckaert A, Butlin RK, Bank C. Deleterious mutation accumulation and the long-term fate of chromosomal inversions. PLOS Genetics. 2021;17: e1009411. doi:10.1371/journal.pgen.1009411

49. Rousselle M, Laverré A, Figuet E, Nabholz B, Galtier N. Influence of Recombination and GC-biased Gene Conversion on the Adaptive and Nonadaptive Substitution Rate in Mammals versus Birds. Mol Biol Evol. 2019;36: 458–471. doi:10.1093/molbev/msy243

50. MacPherson B, Scott R, Gras R. Sex and recombination purge the genome of deleterious alleles: An Individual Based Modeling Approach. Ecological Complexity. 2021;45: 100910. doi:10.1016/j.ecocom.2021.100910

51. Kent TV, Uzunović J, Wright SI. Coevolution between transposable elements and recombination. Philosophical Transactions of the Royal Society B: Biological Sciences. 2017;372: 20160458. doi:10.1098/rstb.2016.0458

52. Xuereb A, Rougemont Q, Dallaire X, Moore J-S, Normandeau E, Bougas B, et al. Re-evaluating Coho salmon (Oncorhynchus kisutch) conservation units in Canada using genomic data. Evolutionary Applications. n/a. doi:10.1111/eva.13489

53. Government of Canada PS and PC. Stock assessment of Thompson River/upper Fraser River coho salmon / J.R. Irvine ... [et al.].: Fs70-1/1999-28E-PDF - Government of Canada Publications - Canada.ca. 1 Jul 2002 [cited 11 Aug 2022]. Available: https://publications.gc.ca/site/eng/9.804890/publication.html

54. Waples RS, Hindar K, Karlsson S, Hard JJ. Evaluating the Ryman–Laikre effect for marine stock enhancement and aquaculture. Curr Zool. 2016;62: 617–627. doi:10.1093/cz/zow060

55. Kratochvíl L, Rovatsos M. Ratios can be misleading for detecting selection. Curr Biol. 2022;32: R28–R30. doi:10.1016/j.cub.2021.11.066

56. Leroy T, Nabholz B. Response to Kratochvíl and Rovatsos. Curr Biol. 2022;32: R30–R31. doi:10.1016/j.cub.2021.11.067

57. Ramu P, Esuma W, Kawuki R, Rabbi IY, Egesi C, Bredeson JV, et al. Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. Nat Genet. 2017;49: 959–963. doi:10.1038/ng.3845

58. Hussin JG, Hodgkinson A, Idaghdour Y, Grenier J-C, Goulet J-P, Gbeha E, et al. Recombination affects accumulation of damaging and disease-associated mutations in human populations. Nat Genet. 2015;47: 400–404. doi:10.1038/ng.3216

59. Alves I, Houle AA, Hussin JG, Awadalla P. The impact of recombination on human mutation load and disease. Philos Trans R Soc Lond B Biol Sci. 2017;372: 20160465. doi:10.1098/rstb.2016.0465

60. Gillard GB, Grønvold L, Røsæg LL, Holen MM, Monsen Ø, Koop BF, et al. Comparative regulomics supports pervasive selection on gene dosage following whole genome duplication. Genome Biology. 2021;22: 103. doi:10.1186/s13059-021-02323-0

61. Khan A, Patel K, Shukla H, Viswanathan A, van der Valk T, Borthakur U, et al. Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers. Proceedings of the National Academy of Sciences. 2021;118: e2023018118. doi:10.1073/pnas.2023018118

62. 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

63. Grossen C, Guillaume F, Keller LF, Croll D. Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. Nat Commun. 2020;11: 1001. doi:10.1038/s41467-020-14803-1

64. Kleinman-Ruiz D, Lucena-Perez M, Villanueva B, Fernández J, Saveljev AP, Ratkiewicz M, et al. Purging of deleterious burden in the endangered Iberian lynx. Proceedings of the National Academy of Sciences. 2022;119: e2110614119. doi:10.1073/pnas.2110614119

65. Parmesan C, Yohe G. A globally coherent fingerprint of climate change impacts across natural systems. Nature. 2003;421: 37–42. doi:10.1038/nature01286

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

67. Ripple WJ, Wolf C, Newsome TM, Gregg JW, Lenton TM, Palomo I, et al. World Scientists' Warning of a Climate Emergency 2021. BioScience. 2021;71: 894–898. doi:10.1093/biosci/biab079

68. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34: i884–i890. doi:10.1093/bioinformatics/bty560

69. 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

70. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics. 2011;43: 491. doi:10.1038/ng.806

71. Dray S, Dufour A-B. The ade4 Package: Implementing the Duality Diagram for Ecologists. Journal of Statistical Software. 2007;22: 1–20.

72. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet. 2007;81: 559–575.

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

74. Christensen KA, Rondeau EB, Sakhrani D, Biagi CA, Johnson H, Joshi J, et al. The pink salmon genome: Uncovering the genomic consequences of a two-year life cycle. PLoS One. 2021;16: e0255752. doi:10.1371/journal.pone.0255752

75. 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

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

77. 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

78. Pante E, Simon-Bouhet B. marmap: A Package for Importing, Plotting and Analyzing Bathymetric and Topographic Data in R. PLOS ONE. 2013;8: e73051. doi:10.1371/journal.pone.0073051

79. Zhang C, Dong S-S, Xu J-Y, He W-M, Yang T-L. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics. 2019;35: 1786–1788. doi:10.1093/bioinformatics/bty875

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

81. 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

82. 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

83. Delaneau O, Zagury J-F, Robinson MR, Marchini JL, Dermitzakis ET. Accurate, scalable and integrative haplotype estimation. Nat Commun. 2019;10: 5436. doi:10.1038/s41467-019-13225-y

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

85. Singhal S, Leffler EM, Sannareddy K, Turner I, Venn O, Hooper DM, et al. Stable recombination hotspots in birds. Science. 2015;350: 928–932. doi:10.1126/science.aad0843

86. Bolívar P, Mugal CF, Nater A, Ellegren H. Recombination Rate Variation Modulates Gene Sequence Evolution Mainly via GC-Biased Gene Conversion, Not Hill-Robertson Interference, in an Avian System. Mol Biol Evol. 2016;33: 216–227. doi:10.1093/molbev/msv214

87. Eyre-Walker A, Keightley PD. Estimating the Rate of Adaptive Molecular Evolution in the Presence of Slightly Deleterious Mutations and Population Size Change. Molecular Biology and Evolution. 2009;26: 2097–2108. doi:10.1093/molbev/msp119

88. 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

89. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics. 2009;Chapter 4: Unit 4.10. doi:10.1002/0471250953.bi0410s25