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.] GenPhySE, INRAE, INP, ENVT, Université de Toulouse, 24 chemin de Borde-Rouge, Auzeville-Tolosane, 31326 Castanet Tolosan, 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

How different factors including demography, recombination or genome duplication may impact selection efficacy and the burden of deleterious mutations in different populations is a central question of evolutionary biology and genetics. 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 genetic load in Coho salmon (*Oncorhynchus kisutch*), a widely distributed salmonid species on the west coast of North America. Using whole genome resequencing data from 14 populations at different travelled migratory distances from their southern glacial refugium, we found support for postglacial gene surfing, with reduced N_e at the range recolonization front, thus inducing both a reduction in the efficacy of selection and a surf of deleterious alleles in populations evolving at low N_e. This inference was robust to various proxies of the load. 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 genetic drift effects, that is random change in allele frequencies. Due to the low number of colonizers, effective population size is reduced at the wave front [2,3] which induce an increase in levels of linkage disequilibrium [4], a loss of genetic diversity and increased

levels of genetic differentiation [2,3,5]. Also, the enhanced effect of genetic drift will impact the segregation of genetic variation. This process favours allele surfing, *i.e.* the fixation of alleles, including neutral, deleterious or advantageous mutations [5–7]. Allele surfing is expected to have two main negative consequences at the wave front *i*) an increase of the deleterious load (called expansion load [8,9]) and *ii*) a loss of adaptive potential [10]. In some cases genetic drift will overwhelm the effect of weak purifying selection[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 ratios of genetic diversity at non-synonymous over synonymous mutations (π_N/π_S)[8,9].

Moreover, 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 [12]. 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 [13]. Theory predicts increased deleterious load with long term effect on population fitness [8] and eventually population extinctions [14]. Evidence for an increased genetic load has been reported in humans which has been associated with the Out-of-Africa bottleneck [15,16]. Evidence has also been gathered in expanding populations in plants [17–21]. For instance in *Arabidopsis lyrata*, range expansion has been associated with an increased mutational load [17] and increased linkage disequilibrium [22]. Similar results were observed in *Escherichia coli* [23]. In contrast, a recent study in *A. lyrata* found no evidence of increased load following range expansion [24]. The reduction in N_e 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 *[25]*. Unless compensatory mutations counteract this effect in low N_e populations [26], one expects a lower proportion of substitutions driven by positive selection [27]. Support for a decreased rate of adaptation associated with range expansion was documented, particularly in plants [18,21]. For instance, such a pattern was reported in *A. lyrata* in North America [18] as well as an increased load in *A. alpina [28]* but not in the European populations of *A. lyrata* despite a slight increase in the genomic burden [24]. Similar to this latter observation, an increased load is observed in human populations and associated with the out-of-Africa scenario [29] but with limited evidence of a decrease in selection efficacy [30]. Beyond application to plants, humans and bacteria, there is a lack of empirical studies documenting the consequences of range expansion on deleterious load and the efficacy of natural selection.

In addition to the geographic scale at which the load unfolds, the genomic scale is also important. Both deleterious load and selection efficacy will vary along the genome as a function of recombination rate [31,32]. Indeed, non recombining regions are expected to more freely accumulate deleterious mutations, a process called Muller's Ratchet [33]. Moreover, Hill-Robertson interference, a process whereby competing alleles interfere with each other to become fixed, 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 [31,34].

Last, variation in tetrasomic inheritance and genomic architecture may affect the spatial distribution of the load along the genome [35]. For instance, whole genome duplication, by doubling chromosomal number, creates several evolutionary challenges immediately following whole genome duplication (WGD) such as altered gene expression and polyploid dysfunction (reviewed in [36]). Most specifically, this process has been associated with a decreased efficacy of purifying selection or transposable elements accumulation, which may then facilitate or inhibit local adaptation as observed in plants[37,38].

The Coho salmon (*Oncorhynchus kisutch*), is a Pacific salmon of culturally and economically major importance which has been severely declining over the last decades (reviewed in [39]). It has undergone further recent decline in multiple parts of its native range [40]. In addition to human induced perturbations, the species 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 [41–43]. Accordingly, these expansions may result in an increased expansion load, raising 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. The Coho salmon offers a unique opportunity to test the role of demography and recombination on the deleterious load and on the efficacy of selection.

Salmonids species, including the Coho salmon, are also a particularly relevant group of species to investigate the effect of whole genome duplication. A salient feature of all salmonids is that their common ancestor has undergone a whole genome duplication ~80-100MyA [44] and is still on its path to rediploidization. For instance, approximately 8% of the Coho salmon genome displays residual tetraploidy [41]. Residual tetraploidy offers a unique opportunity to investigate the direct effect of the genomic variation of N_e on the efficacy of selection. Theoretically, these regions should display an increased effective population size as compared to the rest of the genome (4 N_e instead of 2 N_e) and therefore a higher efficacy of selection and a reduced genetic load. We thus predict that these regions should display a lower genetic load compared to diploid regions of the genome.

To address these questions, we aimed at testing the following general hypotheses: i) whether demographic expansion and bottlenecks have led to an increased load and decreased efficacy of selection, ii) whether heterogeneous recombination levels induce genome-wide variation in the variation of load as a result of Hill-Robertson effects and iii) whether residual tetraploidy results in an increased efficacy of selection and reduced deleterious load through increased recombination and N_e.

Results and Discussion

Strong postglacial population expansion revealed through whole genome sequences

We generated 30X coverage whole genome resequencing data for 71 Coho salmon representing 14 populations distributed from California to some of the most upstream populations of the Porcupine River in Yukon, Canada (**Fig 1A**). We also included several outgroup species namely the Sockeye salmon (*O. nerka*), Chinook salmon (*O. tshawytscha*), Pink salmon *(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 S1-S3**).

To explore the amount of clustering a PCA was used (**Fig S1A, B**). The PCA illustrates that populations are spatially structured following both latitudinal and longitudinal gradients, with the first axis being strongly correlated with both latitude (Pearson r = 0.97, $p < 2e^{-16}$) and longitude (Pearson r = -0.82, $p < 2e^{-16}$) as expected when space drives population structure [45]. Moreover, admixture analyses (**Fig S1B**) revealed that Coho populations were generally structured at the river/watershed level. These observations corroborate our previous inferences based on lower resolution genomic dataset [42,43].

Under a single refugial expansion scenario, we expect a linear decrease in genetic diversity as a function of the distance from the source. We plotted the distribution of observed heterozygosity (H_o) as a function of the distance to the southernmost sample and observed a negative correlation with the distance to the southernmost site ($R^2 = 0.77$, $p<2e^{-16}$, **Fig S2B**). Next, we used the β_{ST} coefficient to identify likely ancestral populations from which the expansion could have started. β_{ST} can account for the non-independence among populations and negative values are indicative of ancestral populations [46]. This metric was also positively correlated with distance to the southernmost site ($R^2 = 0.75$, $p < 2e^{-16}$, **Fig S2A**).

The Salmon River from the Thompson area (dark green point **Fig 1A** and **Fig S2A**) exhibit lower diversity levels, as compared to all non-Alaskan populations. Excluding this outlier river increased the correlation with β_{ST} and H_O 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 [41].

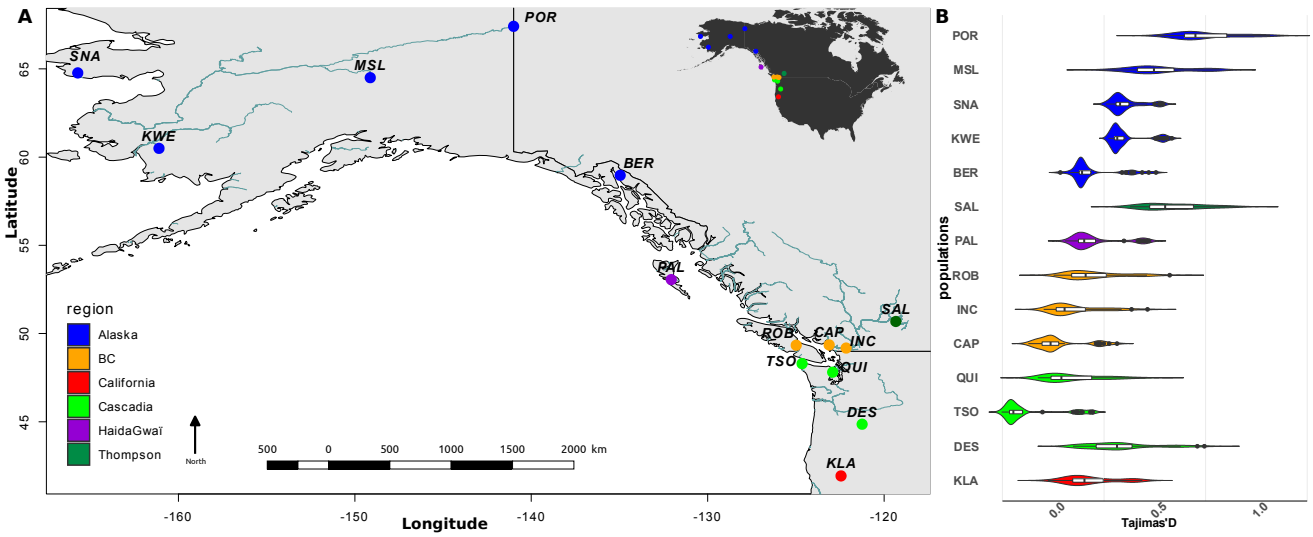

Figure 1: Sampling design and population size change inferred through whole genome sequences. A) Sample site along the Pacific coast of North America. Map produced in R using the World data from Maps package. B) Average Tajima's D values computed along the genome and summarized for each sample locality. 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 S2. 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.

Another indicator of changes in population sizes is an increase in Tajima's D values at the genome-wide level. We tested the hypothesis of increased Tajima's D values as a function of the traveled distance from the southernmost site to the sampling site (reflecting postglacial northern directed expansion) and as a function of the distance to the sea only (reflecting upstream directed founder events). Results are consistent with a signal of population expansion along the colonization axis from the south to the north based on Tajima's D values (R^2 =0.28, p< $2e^{-16}$, **Fig 1C, Fig S2**). Distance 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 e.g. Porcupine (POR), Mile Slough (MSL), Thompson (SAL) and Deschutes (DES).

Another hallmark of population bottlenecks and genetic drift associated with founder events is an increase in linkage disequilibrium across the genome[4]. We tested this hypothesis by measuring linkage disequilibrium (r^2) following [47]. In line with the above observations, linkage disequilibrium (LD) extends to long distance in remote Alaskan/Yukon populations (Porcupine and Mile Slough Rivers) and in the Salmon R. population (Thompson R. Watershed, **Fig S2C**). Accordingly, the LD decay, through a r^2 reaching half of its maximum value, is observed at 25 kb, 12kb and 6kb in these three populations, contrasting with values in the 0.5kbp range observed in populations with higher N_e (**Fig S2C**). These results show that there is a large variance in effective population sizes, likely associated with the strength of the founder events.

To complement the above analyses, we constructed a population phylogeny based on the matrix of pairwise F_{ST} values computed between each populations (**Fig 2A-B**) and on the matrix of variance-covariance in allele frequencies implemented in Treemix [48] (**Fig S3**). Both analyses confirmed that populations located in the

north underwent more genetic drift in line with the postglacial expansion from the south to the north and the subsequent founder effects associated with upstream colonisation from the west to the east.

To consolidate the above observations, we investigated historical changes in effective population size using the Sequentially Markovian Coalescent in SMC++ [49] which takes advantage of both the SMC and of the site frequency spectrum of each population. based on the 71 whole genome sequences (**Fig S2D)** and both a median mutation rate of $8e^{-9}$ mutations/bp/generation and a mean of $1.25e^{-8}$ mutations/bp/generation as estimated for the Atlantic salmon (*salmo* salar) a closely related salmonid (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 demographic 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 S4**) compatible with the onset of postglacial population expansion and colonisation of different rivers following glacial retreat [41,42]. Using the mean mutation rate yielded more recent estimates of split times (median = 9,6 KyA, **Fig S4**) (min = 5 KyA – max = 18 KyA). Overall, SMC++ results indicate that all populations shared a similar demographic history until they began to diverge following the end of the last glacial cycle (see **Note S1** and associated Supplementary Tables).

Range expansion explained variation in the mutation load

We estimated the mutation load of the populations using different metrics : *i)* π_N/π_S, *ii)* count of derived homozygous mutation (including both missense and Loss-of-function (LoF) mutation), and *iii)* total count of derived mutations (including both missense and LoF mutation). We observed a significant positive correlation of the π_N/π_S ratio of each local population and the distance to the southernmost site which was computed following the coastline and the downstream to upstream in-river distance (see methods) corresponding to the most-likely refugia as discussed in our previous work [42] ($R^2 = 0.73$, $p<0.0001$, **Fig 2D**). Similarly, using the tree branch length to the root of our tree (**Fig 2B**) as a proxy of expansion route revealed a significant correlation with the π_N/π_S ratio ($R^2 = 0.66$, $p<0.0001$, F**ig 2C**). This result was also supported by an analysis considering tree branch lengths inferred with Treemix (**Table S4**). Accordingly, there was a significant correlation between the π_N/π_S and two different estimates of the effective population size, namely π_S and SMC++ N_e, which we expected to be reduced both at the colonization front and in populations having undergone a bottleneck (π_S in **Fig S5A** and SMC++ N_e in **Fig S5B***)*. These results support expectations pertaining to variation in mutation load under a model of expansion load (e.g. [29]).

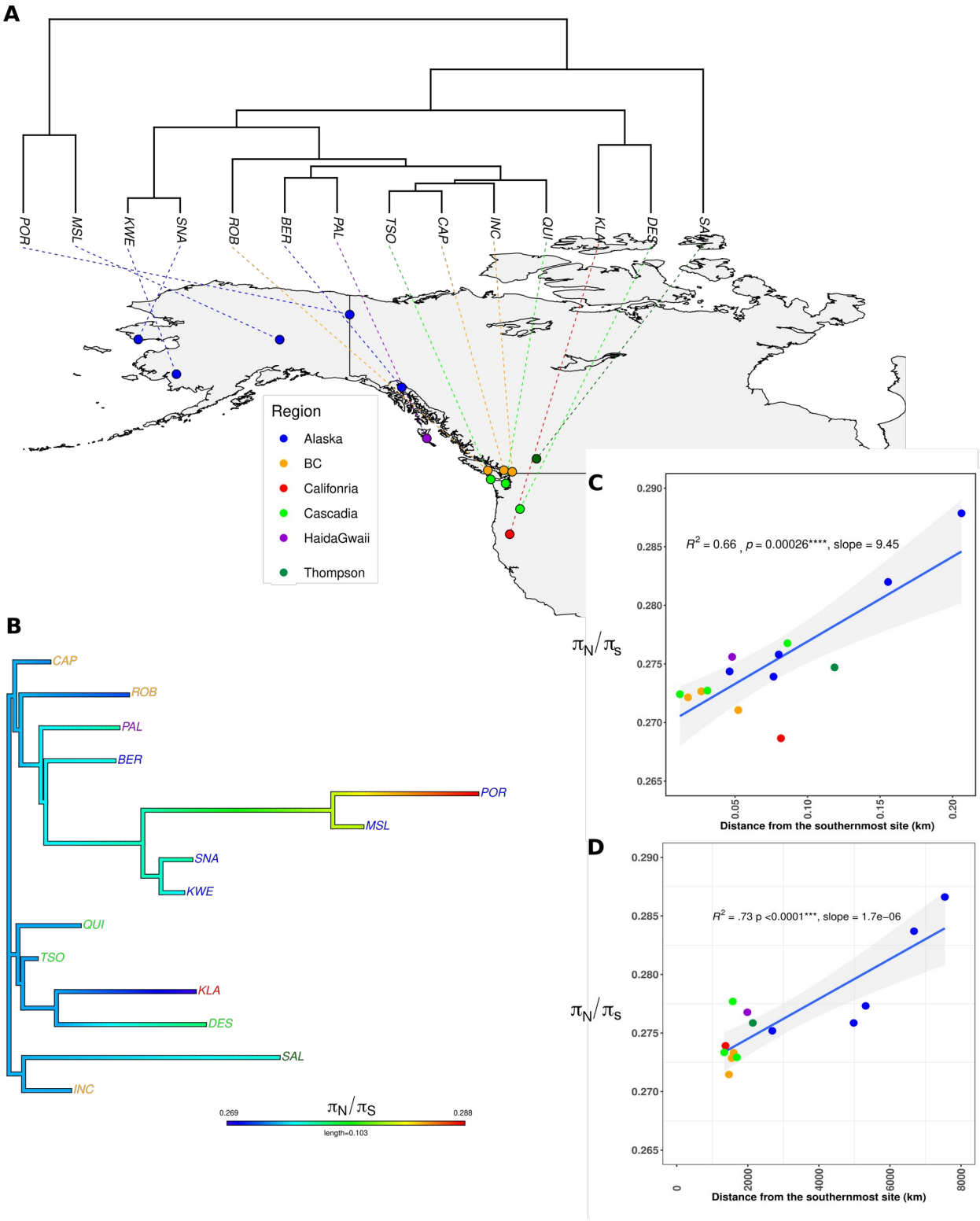

Figure 2: Expansion routes inferred from phylogenetic analyses revealed the extent of the mutation load. A) population phylogenetic tree inferred from F_{ST} values (see also Figure S3) superimposed on the map displaying sample location. Map produced with R using the worldHires data. B) Population specificf branc length of the genetic tree are coloured according to the value of the π_N/π_S ratio, used as a proxy of the genetic load. C) Correlation between tree branch lengths to the root and the π_N/π_S ratio. D) Correlation between the travelled distance to the southernmost site and the π_N/π_S ratio. The blue line represents the slope of the linear model between the two variables. The grey area in panel C and D represents the 95% confidence interval levels around the regression lines.

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/π_S will reach an equilibrium value faster than π_S because selected alleles will be subject to higher negative selection and undergo a faster turnover [15,50,51]. Thus, the π_N/π_S may not be the best predictor of the total burden of deleterious mutations as it is potentially affected by several factors including demography [15]. To circumvent this limitation, we counted and plotted the distribution of non-synonymous mutations classified according to i) their impact and expected consequences on fitness (*i.e.* missense and LoF mutations) and; ii) segregation patterns (*i.e.* total load composed of heterozygous and derived homozygous genotypes or fixed load composed of homozygous derived genotype) [52]. As for the π_N/π_S we found a 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**). Similarly, we observed a significant association between the tree branch length to the root and the derived load of missense and LoF mutations ($R^2 = 0.39$, $p = 0.0096$ and $R^2 = 0.40$, $p = 0.009$, respectively, **Table S4**).

The most deleterious mutations are efficiently purged across the range

The total number of missense mutations was significantly more abundant in the more southern populations (*i.e* California, Cascadia, **Table S5** for all Tukey-HSD *p-values*) than in northern ones (**Fig 3A**) while it was nearly constant from the south to the north regarding LoF mutations (**Fig 3B)** with no significant differences among populations **(Table S5**). The southernmost populations displayed a higher load of mutations in heterozygous state (**Fig 3**), as expected due to their higher historical effective population size, favouring the segregation of recessive mutations hidden in a heterozygous state [42,53].

The fixed load increases with population expansion

The fixed load (i.e. count of derived homozygous sites) increased from south to north for both missense and LoF with the most extremes samples, including Mile Slough, Porcupine and Salmon rivers being the most significantly loaded (**Table S5**), as expected due to founder events and allele surfing for Alaskan/Yukon samples and due to bottlenecks for the Salmon River (**Fig 3** right panel).

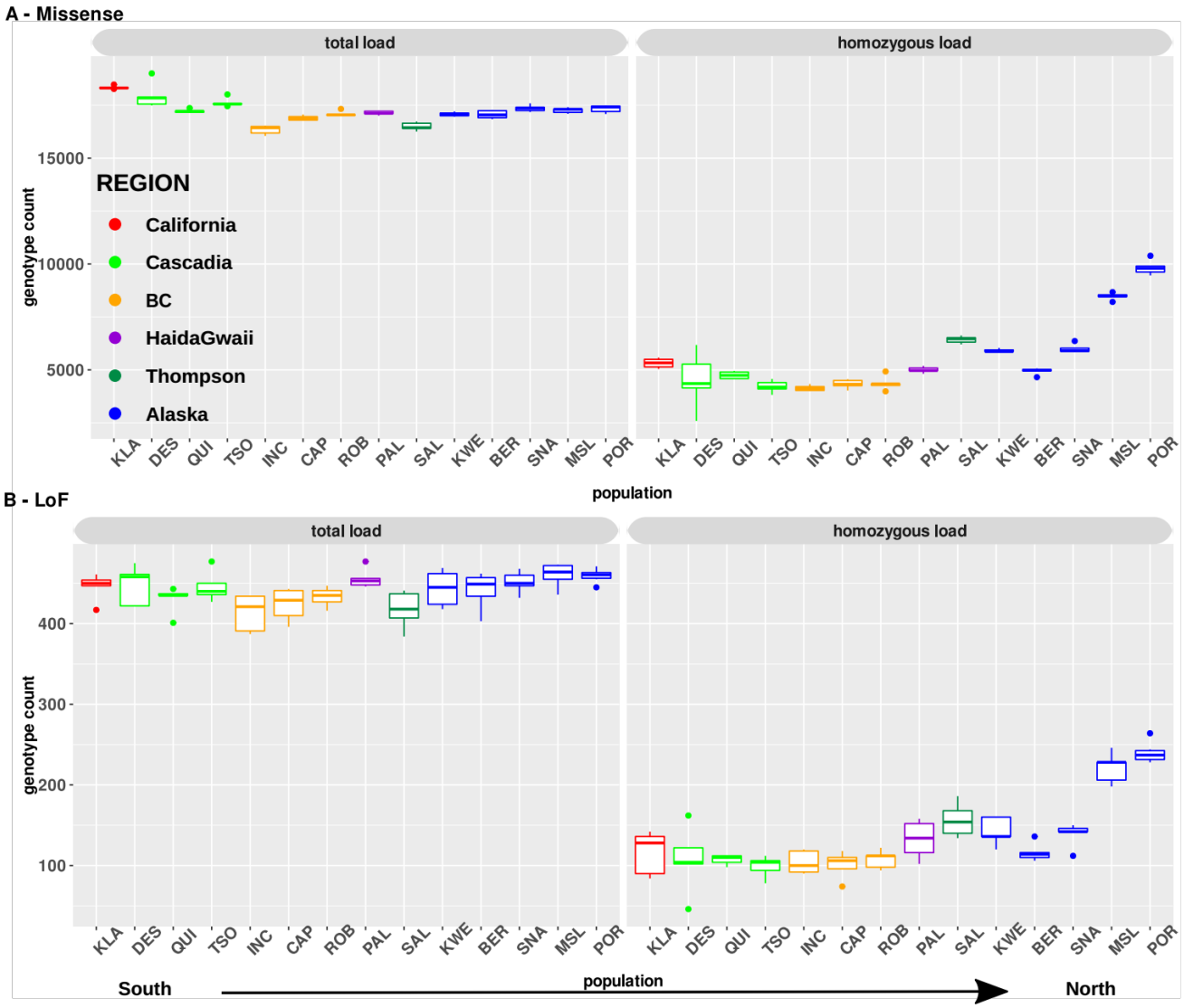

Figure 3: Number of deleterious alleles per river sorted from the South to the North showing the total load (left panel), and fixed derived homozygous (recessive) load (right panel) for A) missense mutation and B) Loss-Of-Function (LoF) mutations. No strong differences are observed in the total load 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.

Range expansion have reduced the efficacy of selection at range margin

We predicted that selection efficacy will be reduced at the expansion front and in populations with a lower N_e. To test this, we measured three parameters, namely the rate of non-adaptive and adaptive amino-acid substitutions relative to neutral divergence (ω_{NA} and ω_A, respectively; with $\omega_A = d_N/d_S - \omega_{NA}$) and the proportion of non-synonymous amino-acid substitution that result from positive selection ($\alpha = \omega_A/(d_N/d_S)$) using the software Grapes [54] using maximum likelihood estimation under a population genetic model. A general expectation is that populations at the expansion margin will exhibit a higher rate of non-adaptive substitutions (ω_{NA}), a lower rate of adaptive substitutions (ω_A) and proportion of amino-acid substitutions (α).

We observed a reduced efficacy of purifying selection likely associated with the multiple founder events and bottlenecks which resulted in a decrease rate of adaptive substitutions (ω_A, R^2 = 0.37, p = 0.013) and increased ω_{NA} as a function of the distance to the South (R^2 = 0.57, p = 0.0011, **Fig 4A**). We also found a significant positive correlation between ω_{NA} and the distance to the ocean (R^2 = 0.52, p = 0.002, **Fig S6A**) and conversely a negative correlation between ω_A and the distance to the ocean (R^2 = 0.40, p = 0.00895, **Fig S6B**), suggesting that upstream populations displays a lower adaptive potential.

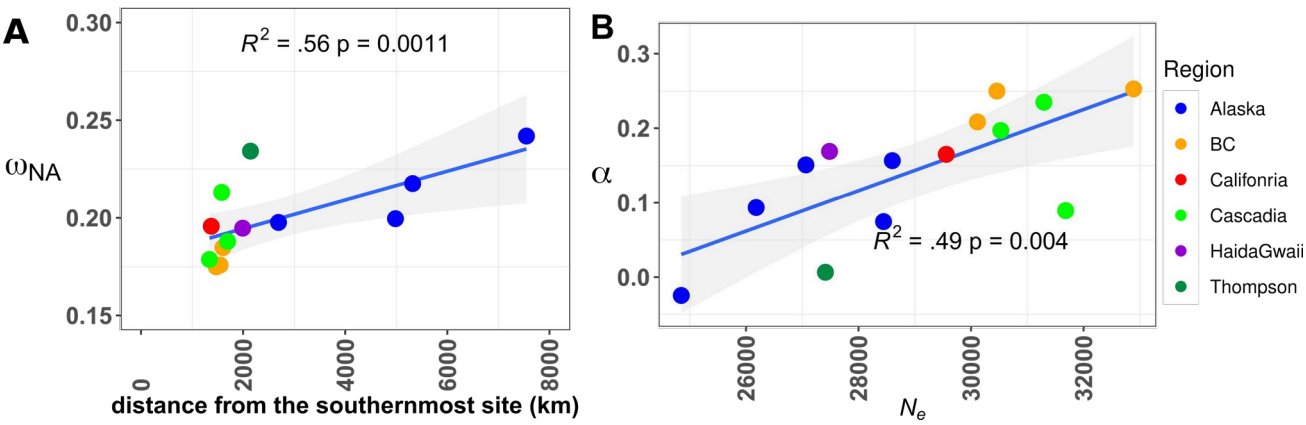

Figure 4: Efficacy of selection decrease at the expansion front and as a function of long-term effective population size. A) increased rate of nonsynonymous non-adaptive substitutions (ω_{NA}) in distant population of Coho salmon and in the bottlenecked Thompson River (dark green). B) Correlation between α and historical variation in the coalescent effective population size estimated from SMC++. See also **Fig S7-S8** for correlations based on tree branch lengths. 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.

Similarly strong correlation were observed when considering the tree branch length to the root (details in **Fig S7-S8, Table S4**). The Salmon R. population (Thompson R. watershed), which sufferent a recent decline in abundance, supposedly due to the sustained release of hatchery fish derived from few individuals and the Ryman-Laikre effect [55] (see discussion below), also displayed a high ω_{NA}.

We also predicted a higher α value in populations with higher N_e and, conversely, a decreased value of α in populations with lower N_e. 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) N_e (R^2 = 0.63, p = 0.0004, **Fig S9**). To more directly test the association with N_e, we correlated SMC++ based estimates of N_e (averaged over a 200KyA window) with α and recovered significant correlations (R^2 = 0.49, p = 0.004, **Fig 4B**). 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, results not shown). Altogether, the above results provide empirical support pertaining to the evolutionary consequences of allele surfing at expanding range margins, in particular regarding the loss of adaptive potential and the mutation burden.

Recombination rates shape the deleterious mutational landscape

In addition to the spatial structure associated with the postglacial recolonization, we investigated the genome-wide variation in mutation load. 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 [56]. To test this hypothesis, we used the GC content at third codon position (GC3 hereafter) as a proxy of the rate of recombination [57,58] (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; p = 0.0038) between levels of GC3 and π_N/π_S ratio among all Coho salmon populations (**Fig 5A**). An analysis focused on GC-conservative sites, which are not affected by GC-biased Gene Conversion (gBGC), revealed similarly strong patterns across all Coho salmon populations (**Fig 5B**, R^2 range = 0.665 - 0.876, p < 0.01). π_N/π_S ratios estimated based on all sites or GC-conservatives-sites only are highly correlated (Pearson r = 0.909, p <0.001). We tested the generality of this relationship using four other closely related Pacific 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 (R^2 range = 0.81 - 0.96, p <0.001), as well as Rainbow trout populations (R^2 range = 0.95 - 0.96, p <0.001, **Fig 5C**).

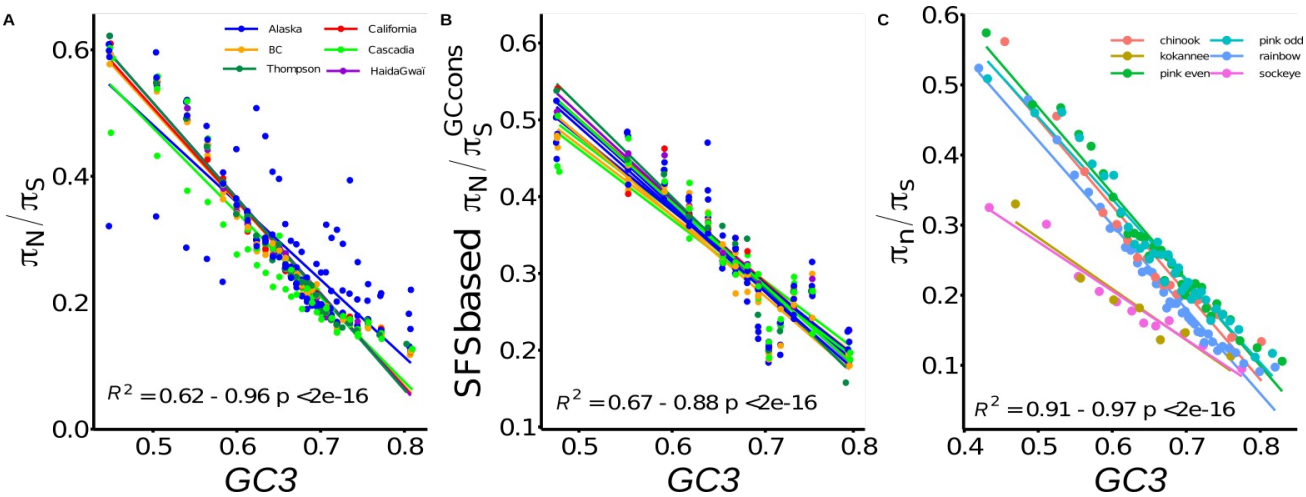

Figure 5) The deleterious load is determined by variation in GC3 content in multiple salmonids

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 colored by major regional group (See **Fig S10A** 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. 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 level, see **Fig S10B-C** for detail by population.

We then analyzed the difference in the correlations between levels of GC3 and π_N/π_S ratio among all Coho salmon populations. Indeed, the intensity of this correlation highlights the effect of recombination rate variation on the efficiency of purifying selection [59]. The populations at the expansion front (Mile Slough R.,

Porcupine R., Snake R.) in Alaska exhibit the lowest correlations 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). If both demographic factors (*i.e.* distance to the source here) and genomic factors (*i.e.* recombination) interact to shape the load, we may predict a positive relationship between these two factors. A linear model between the slope of the regression between π_N/π_S and GC3 and the distance to the ancestral source populations did not reveal such a relationship (p = 0.133, **Fig S11**).

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 N_e 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 6A**, 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 S12,** p < 0.0001, W Mann-Whitney = $3.04e^7$, see also **Fig S13** and **table S6** for differences in recombination among populations), our results suggest that this higher π_N/π_S could be mostly due to lower recombination rates. Another expected genomic consequence of this lower recombination rate is a higher load of transposable elements [60]. When computing the relative length of TE, *i.e.* 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 6B**, p <0.0001, WMann-Whitney = $1.36e^{10}$). 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. 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 S7**) 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.

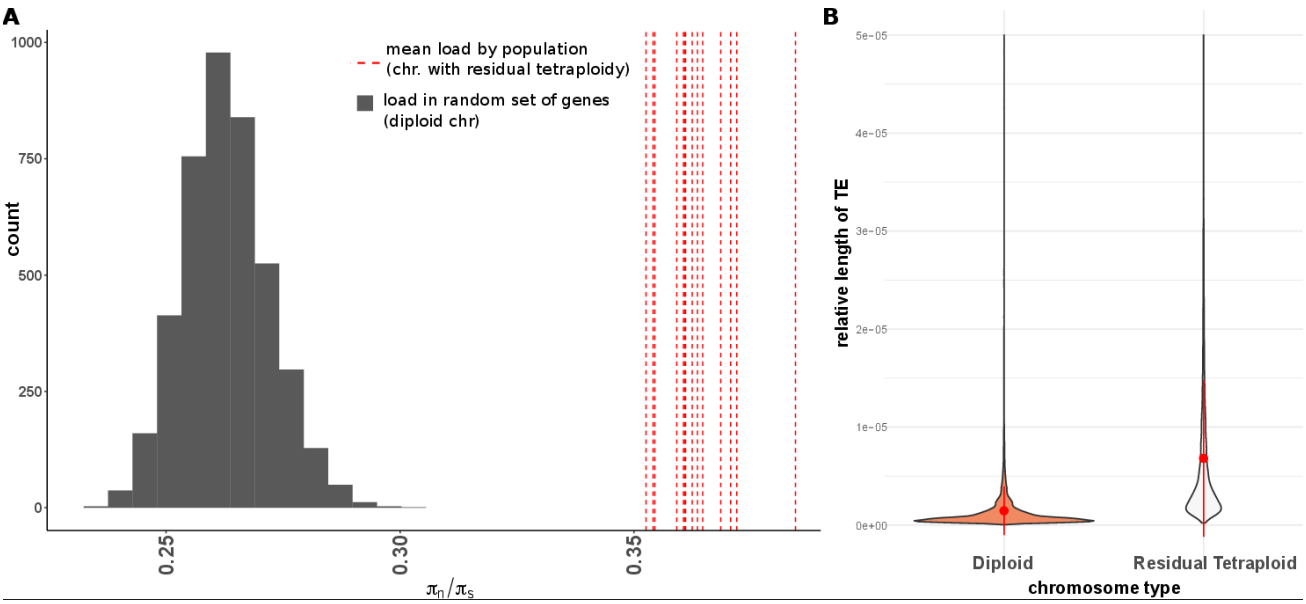

Figure 6: Increased load in region of residual tetraploidy. A) Distribution of π_N/π_S ratio when considering all genes in regions of residual tetraploidy (red lines) for each population compared to a set of 200 randomly generated samples of 4,000 genes in the rest of the genome (gray histogram). **B)** Violin plot 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, recombination and variation in inheritance in affecting the load and efficacy of natural selection is a central question, both with fundamental and applied consequences for biodiversity.

Using population genomics analyses, we investigated the evolutionary consequences of recent demographic events, particularly regarding N_e in Coho salmon. Our results supported gene surfing across North America [42]. Results also supported lower N_e 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 as expected under the nearly neutral hypothesis of molecular evolution [12]. 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 [42,43,61] and displayed an increased rate of long runs of homozygosity [41]. It has been subjected to extensive hatchery enhancement (from a single population) to circumvent the decline of Coho salmon in the Thompson drainage [62]. From a theoretical standpoint, if a few captive individuals were used as parents for subsequent release, then an increase in inbreeding and reduction in effective population size is expected, a phenomenon called the Ryman-Laikre effect [55]. This process is expected to generate a strong decrease in the effective population size of supplemented populations [63]. Consequently it is 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 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.

It is noteworthy that the southernmost population (Klamath R, California) was not charaterized by the highest α value. Several non-exclusive hypotheses may explain this pattern: first, some Californian populations are known to have undergone strong recent (human driven) decrease in abundance due to both habitat degradation and climate change, which may leave a detectable footprint in the genome [64] and the recent demography could have an impact on the accuracy of the fit of the synonymous and nonsynonymous SFS under Grapes. Second, the Klamath R. population is located in the far upstream part of that river, and may have undergone a founder event when reaching this upper part, a hypothesis that is supported by the correlations we observed between several metrics and distance to the river mouth. Finally, we did not include the southernmost populations from California, which are more likely to be more ancestral, as we previously reported [42].

Since the use of π_N/π_S ratio can be criticized [15,65,66] especially when used to quantify the load within species, we computed additional metrics to [54] more directly estimate mutation load, as advocated by others [15,67]. 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 and as a function of the tree branch length, both are used as proxies of the expansion route.

We demonstrated that the π_N/π_S ratio was negatively correlated to the GC content at third codon position, which represents a good proxy of the local recombination rate(but see **note S2**). The negative 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 by gBGC) [58]. These results indicated that recombination plays a key role in the efficacy of selection along the genome in salmonids and gave empirical support to theoretical studies about the accumulation of slightly deleterious mutations in non-recombining regions [31,33]. Similarly, increased prevalence of deleterious mutations in low recombining regions have been reported in plants [68] and in human populations [69–71]. In particular, recent work in human populations have shown that both variation in demographic history (i.e. change in effective population size) and recombination rates are affecting allele specific expression of harmful mutations [69]. The authors showed that allele specific expression causes under expression of harmful mutations more efficiently in normally and highly recombining regions compared to low recombining regions. They further documented variation of this process among populations with varying 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 [72]. Recent studies in polyploid plant species have also documented various evolutionary consequences of such duplication on local variation in the load and efficacy of selection [36]. For instance in *A. arenosa*, variation in ploidy levels was associated with a reduced efficacy of purifying selection [38]. Similarly, recent work in the allopolyploid cotton *(Gossypium)* demonstrated that this species accumulates more deleterious mutations than the diploid species [73]. This observation supports a theory proposed by Haldane that recessive deleterious mutations accumulate faster in allopolyploids because of the masking effect of duplicated genes [74]. To sum up, our results concur with predictions from the nearly neutral theory of molecular evolution [12] in which slightly deleterious mutations are effectively neutral and purged effectively in regions of higher recombination except perhaps in populations at the extreme of the expansion front.

Finally, our results have 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 on several plant and animal models [75–78]. However, results 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. 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. With this caveat in mind, our results for the Salmon R could 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 [53]. This strategy could reduce the occurrence of deleterious alleles and counteract the Ryman-Laikre effect described above. Moreover, populations (e.g. most upstream Alaskan/Yukon 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 most rapid at higher latitudes [79]. 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 [80]. More generally, 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 [53,81–83]. 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 [84].

Methods

Sampling design for Coho salmon

We sampled and sequenced 71 individuals representing 14 populations distributed from California to Alaska (**Table S01**). A set of 55 individuals was sequenced on an Illumina HiSeq2500 platform [38] and the other 16 individuals were sequenced on a NovaSeq6000 S4 platform using paired-end 150 bp reads. Reads were processed using fastp for trimming [85], and mapped to the most recent Coho reference genome (https://www.ncbi.nlm.nih.gov/assembly/GCF_002021735.2/) using bwa mem v2 [86]. 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 [87] 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 using a principal component analysis (PCA) performed using Ade4 [88] on a set of LD-pruned SNPs without missing data (1,739,037 SNPs) identified stringently with plink1.9 [89] (command indep-pairwise 100 50 0.1).

Outgroup dataset.

In order to test the generality of the relationship observed between π_N/π_S and GC3 we took advantage of the newly assembled reference genomes and resequencing data from other closely related Pacific salmonid species with similar demographic history[90–92]. Sockeye salmon published by [90] 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**). Three 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 [91]. These were indeed clearly separated based on a PCA. Finally, a number 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 quantified the load using the π_N/π_s ratio using the procedure described below for Coho salmon. For the Sockeye/Kokanee ecotypes, the Sockeye is a fully migratory ecotype, whereas the Kokanee is a resident (non-migratory) ecotype that typically comprises more isolated populations. The two alternative ecotypes are sometimes found in similar locations, with three rivers from our sampling design including both ecotypes (Table S03). For all other species we tested the relationship between π_N/π_S and GC3 (see below)..

Ancestral and derived alleles identification

To accurately recover the ancestral allelic states 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 [93]. This was used for demographic reconstruction detailed in **Note S1.**

Demographic reconstruction

We first tested our prediction that genetic diversity (observed heterozygosity H_o) decreases towards the North following our «out of Cascadia» model previously inferred [42,43]. Conversely, we verified that the β_{ST} coefficient, a measure of both genetic differentiation and ancestrality [46] increases northwards, as expected due to isolation by distance. β_{ST} and observed heterozygosity (H_o) were measured using the hierfstat R package [94] Details about β_{ST} computation are provided in Weir & Goudet [46] and H_o was computed following Nei (1987)[95] as 1 - $\sum_k \sum_i Pkii/np$ with *Pkii* the proportion of homozygotes *i* in sample *k* and *np* the number of samples. Oceanic coastal distances were computed using the marmap package [96] and waterway distance was computed using ArcGIS. Then we summed the 2 distances to obtain the distance to the most likely refugia that we identified in our previous studies [42,43]. A shapefile of the rivers used here is avaiable at https://github.com/QuentinRougemont/selection_efficacy/tree/main/00.raw_data.

Under postglacial expansion from a single refugium, the general hypothesis would be that all sampled populations should follow a common temporal trajectory of a population decline (bottleneck due to founder events by few individuals) followed by a (strong) increase in N_e corresponding to the expansion phase. To test this hypothesis, we inferred temporal changes in N_e using SMC++ [49]. 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 changes in population size 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, well documented for Coho, and a mutation rate of $8e^{-9}$ mutation/bp/generation were applied (corresponding to the median substitution rate inferred in Atlantic salmon, J. 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 [97]. 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 [98] package in R.

We observed slight discrepancies between our SMC++ estimates of divergence time and our previous work based on the site frequency spectrum [42]. To investigate this, we performed a new set of inference based on the unfolded joint site frequency spectrum (jSFS) using ∂a∂i [99] and a new set of refined models as detailed in **supp. Note S1, Table S8-S9**.

To reconstruct the population expansion routes we computed Weir and Cockerham [100] F_{ST} values among populations using vcftools [101] and constructed a phylogenetic population tree using the R package ape [102]. We then computed the tree branch lengths to the root using the function distroot from adephylo [103]. We also explored broad relationships among populations with treemix [104]. 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. Similarly, we computed the tree branch length to the root using the function adephylo and compared our results to those from F_{ST}-based values. The resulting tree branch length values were used as a proxy of the expansions routes and we tested their correlation using linear models in R, with all our metrics of deleterious load (π_N/π_S, total number of putative homozygous derived deleterious alleles (recessive load) and total number of deleterious alleles (both in homozygous and heterozygous states, additive load) for derived missense and LoF mutations) and metrics of

selection efficacy (ω_{NA}, ω_A, α detailed below). Similar inference was performed but using the reconstructed distances to the southernmost site included in our previous studies [42,43]. This distance was computed summing the oceanographic distance and river distances describe above.

Population-scaled recombination rate

Statistical phasing of the Coho whole genome sequences was performed using the shapeit software [105] considering all individuals at once. We then estimated effective recombination rates ($\rho=4.N_e.r$ where r represents the recombination rate per generation and N_e is the effective population size) along the genome using LDHat [106]. Phased genotypes were converted into LDHat format using vcftools after excluding SNPs with MAF < 10% since rare variants are not informative for such inferences. Following the guidelines of the LDHat manual), the genome was split in fragments of 2,000 SNPs with overlapping windows of 500 SNPs. We measured recombination rates independently for each population Difference in distribution of population scaled recombination was visualized using violin plot (Fig S13) and statistically tested using Tuckey HSD tests in R (Table S5) .

Genetic load estimation

π_N/π_S estimates

The approach developed in [57] was used to reconstruct fasta sequences for each individual. For each species (Coho and the other salmonids), coding sequences (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.

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: $N_{total} = 2 \times N_{homo} + N_{hetero}$ [29]. These individual values were then averaged per population. We tested for differences in the distribution of these counts among populations using an ANOVA followed by a TukeyHSD test. The p-values 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 applied the commonly used Provean [107] software based on a random set of non-synonymous mutations (but see **Note S4** for a brief discussion regarding the limitations). Provean measures the effect of a variation (*v*) on a protein A by measuring the similarity (using blast) of the protein A to a protein B before and after the introduction of the variation. The assumption is that a variation that reduces the similarity of protein A to the functional homolog protein B is more likely to cause a damaging effect. To this end, a change in the "alignment score" (delta score) can be used as a measure of change in "similarity" caused by a variation. A low delta score is interpreted as a deleterious effect of the variation, with a score below -2.5 being the recommended value to provide a tradeoff between specificity (i.e. a high confidence in the variants being identified as deleterious) and sensitivity (i.e. identifying a high number of deleterious variants). A high delta score is interpreted as neutral. The top 30 clusters of closely related proteins are used to compute a delta score, which are then averaged across clusters [107]. These results were then compared with results from non-synonymous mutations **(Fig S15**).

Correlation between π_N/π_S and recombination

We 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 GC-biased Gene Conversion (gBGC).

Finally, we measured the correlation between GC3 and π_N/π_S using linear models. 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 potential resulting from positive selection ($\alpha = \omega_A/(d_N/d_S)$, with d_N the rate of non-synonymous substitutions, d_S the rate of synonymous substitutions). To do so, we used the method implemented in Grapes v1.0 which builds upon the approach of [108]. 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). We used the following parameters in Grapes v1.0: we assumed a negative Gamma distribution to the synonymous and non-synoymous SFS (parameters GammaExpo in Grapes and parameter "unfolded" site frequency spectrum). To obtain suitable data we convert the quality filtered whole genome vcf file into a fasta file for each population containing sequence

information for each individual in the population (using vcf2fasta available at https://github.com/QuentinRougemont/selection_efficacy/tree/main/08.piNpiS/). This file was then converted into a site-frequency spectrum using bppml [109]. We required at least 10 sites without gap to keep a given site (-gapN_site), with a maximum proportion of gap of 0.5 (-gapN_seq), a minimum number of complete codon of 6 (-min_nb_codon), we did not allow frameshift (remove_frameshift parameter), we also require a sample size of 8 (-sample_size), the number of initial/terminal position in which Sop codons or FrameShift codons are tolerated was set to 20 (-tolerance_zone), we did not allow Stop or FrameShift codons between the two tolerance zones (-allow_internal_bc parameter), the kappa value (i.e Ts/Tv ratio) was set to 1.6 and we used an unfolded SFS using the three species as outgroup. Fitted parameters of the DFE were used to compute the expected d_N/d_S under near neutrality, which was compared to the observed d_N/d_S to estimate α, ω_{NA}, ωa.

Differences in load for region of residual tetraploidy in Coho salmon

π_N/π_S comparison

In the Coho reference genome, Chromosomes 1 to 30 are considered to represent generally diploid chromosomes, whereas chromosomes 31 to 38 represent those with a clear signal of residual tetraploid [41]. 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 [110] (made available on NCBI for the reference genome [41]) 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 are currently being deposited on NCBI SRA under project PRJNA982773. The coho vcf is available online at 10.5281/zenodo.8027705. All scripts to reproduce our analyses are currently being 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 the 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. This research was carried out in conjunction with EPIC4 (Enhanced Production in Coho: Culture, Community, Catch), a project supported in part by University Laval, Ressources Aquatiques Québec (RAQ), and the Government of Canada through Genome Canada, Genome British Columbia, and Genome Québec. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

3. 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

4. Ohta T. Linkage disequilibrium due to random genetic drift in finite subdivided populations. Proc Natl Acad Sci U S A. 1982;79: 1940–1944. doi:10.1073/pnas.79.6.1940

5. 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

6. 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

7. 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

8. 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

9. 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

10. 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

11. Haag CR, Roze D. Genetic Load in Sexual and Asexual Diploids: Segregation, Dominance and Genetic Drift. Genetics. 2007;176: 1663–1678. doi:10.1534/genetics.107.073080

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

13. 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

14. Lynch M, Conery J, Burger R. Mutation Accumulation and the Extinction of Small Populations. The American Naturalist. 1995;146: 489–518.

15. 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

16. 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

17. 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

18. Willi Y, Fracassetti M, Bachmann O, Van Buskirk J. Demographic Processes Linked to Genetic Diversity and Positive Selection across a Species' Range. Plant Communications. 2020;1: 100111. doi:10.1016/j.xplc.2020.100111

19. Perrier A, Sánchez-Castro D, Willi Y. Expressed mutational load increases toward the edge of a species' geographic range. Evolution. 2020;74: 1711–1723. doi:10.1111/evo.14042

20. de Pedro M, Riba M, González-Martínez SC, Seoane P, Bautista R, Claros MG, et al. Demography, genetic diversity and expansion load in the colonizing species Leontodon longirostris (Asteraceae) throughout its native range. Molecular Ecology. 2021;30: 1190–1205. doi:10.1111/mec.15802

21. 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

22. Lucek K, Willi Y. Drivers of linkage disequilibrium across a species' geographic range. PLOS Genetics. 2021;17: e1009477. doi:10.1371/journal.pgen.1009477

23. 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

24. 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

25. 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

26. 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

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

28. Laenen B, Tedder A, Nowak MD, Toräng P, Wunder J, Wötzel S, et al. Demography and mating system shape the genome-wide impact of purifying selection in Arabis alpina. PNAS. 2018;115: 816–821. doi:10.1073/pnas.1707492115

29. 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

30. 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

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

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

33. 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

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

35. Hill RR. Selection in autotetraploids. Theor Appl Genet. 1970;41: 181–186. doi:10.1007/BF00277621

36. Baduel P, Bray S, Vallejo-Marin M, Kolář F, Yant L. The "Polyploid Hop": Shifting Challenges and Opportunities Over the Evolutionary Lifespan of Genome Duplications. Frontiers in Ecology and Evolution. 2018;6. Available: https://www.frontiersin.org/articles/10.3389/fevo.2018.00117

37. Baduel P, Quadrana L, Hunter B, Bomblies K, Colot V. Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. Nat Commun. 2019;10: 5818. doi:10.1038/s41467-019-13730-0

38. Monnahan P, Kolář F, Baduel P, Sailer C, Koch J, Horvath R, et al. Pervasive population genomic consequences of genome duplication in Arabidopsis arenosa. Nat Ecol Evol. 2019;3: 457–468. doi:10.1038/s41559-019-0807-4

39. 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

40. 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

41. Rondeau EB, Christensen KA, Minkley DR, Leong JS, Chan MTT, Despins CA, et al. Population-size history inferences from the coho salmon (Oncorhynchus kisutch) genome. G3 Genes|Genomes|Genetics. 2023;13: jkad033. doi:10.1093/g3journal/jkad033

42. 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

43. 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. 2023;32: 542–559.

44. 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

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

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

47. Hill WG, Robertson A. Linkage disequilibrium in finite populations. Theor Appl Genet. 1968;38: 226–231. doi:10.1007/BF01245622

48. 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

49. 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

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

51. 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

52. 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

53. 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

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

55. 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

56. 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

57. 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

58. 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

59. 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

60. 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

61. 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. 2022;15: 1925–1944. doi:10.1111/eva.13489

62. 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

63. 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

64. 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

65. 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

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

67. 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

68. 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

69. 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

70. 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

71. 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

72. 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

73. Conover JL, Wendel JF. Deleterious Mutations Accumulate Faster in Allopolyploid Than Diploid Cotton (Gossypium) and Unequally between Subgenomes. Molecular Biology and Evolution. 2022;39: msac024. doi:10.1093/molbev/msac024

74. Haldane J. The Causes of Evolution. 1932 [cited 9 Jun 2023]. Available: https://jbshaldane.org/books/1932-Causes-of-Evolution/index.html

75. 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

76. 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

77. 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

78. 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

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

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

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

82. 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

83. 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

84. 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

85. 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

86. 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

87. 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

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

89. 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.

90. 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

91. 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

92. 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

93. 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

94. 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

95. Nei M. Molecular Evolutionary Genetics. Molecular Evolutionary Genetics. Columbia University Press; 2019. doi:10.7312/nei-92038

96. 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

97. 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

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

99. 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

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

101. 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

102. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019;35: 526–528. doi:10.1093/bioinformatics/bty633

103. Jombart T, Balloux F, Dray S. adephylo: new tools for investigating the phylogenetic signal in biological traits. Bioinformatics. 2010;26: 1907–1909. doi:10.1093/bioinformatics/btq292

104. 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

105. 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

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

107. 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

108. 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

109. Guéguen L, Gaillard S, Boussau B, Gouy M, Groussin M, Rochette NC, et al. Bio++: Efficient Extensible Libraries and Tools for Computational Molecular Evolution. Molecular Biology and Evolution. 2013;30: 1745–1750. doi:10.1093/molbev/mst097

110. 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