

ORIGINAL ARTICLE

The role of recombination on genome-wide patterns of local ancestry exemplified by supplemented brook charr populations

Maeva Leitwein¹  | Hugo Cayuela¹  | Anne-Laure Ferchaud¹  |
 Éric Normandeau¹  | Pierre-Alexandre Gagnaire²  | Louis Bernatchez¹ 

¹Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, QC, Canada

²ISEM, CNRS, IRD, EPHE, Univ. Montpellier, Montpellier, France

Correspondence

Maeva Leitwein, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, QC, Canada.
 Email: maeva.leitwein.pro@gmail.com

Funding information

MFFP; Canadian Research Chair in Genomics and Conservation of Aquatic Resources; Ressources Aquatiques Québec

Abstract

Assessing the immediate and long-term evolutionary consequences of human-mediated hybridization is of major concern for conservation biology. Several studies have documented how selection in interaction with recombination modulates introgression at a genome-wide scale, but few have considered the dynamics of this process within and among chromosomes. Here, we used an exploited freshwater fish, the brook charr (*Salvelinus fontinalis*), for which decades of stocking practices have resulted in admixture between wild populations and an introduced domestic strain, to assess both the temporal dynamics and local chromosomal variation in domestic ancestry. We provide a detailed picture of the domestic ancestry patterns across the genome using about 33,000 mapped single nucleotide polymorphisms genotyped in 611 individuals from 24 supplemented populations. For each lake, we distinguished early- and late-generation hybrids using information regarding admixture tracts. To assess the selective outcomes following admixture we then evaluated the relationship between recombination and admixture proportions at three different scales: the whole genome, chromosomes and within 2-Mb windows. This allowed us to detect a wide range of evolutionary mechanisms varying along the genome, as reflected by the finding of favoured or disfavoured introgression of domestic haplotypes. Among these, the main factor modulating local ancestry was probably the presence of deleterious recessive mutations in the wild populations, which can be efficiently hidden to selection in the presence of long admixture tracts. Overall, our results emphasize the relevance of taking into consideration local ancestry information to assess both the temporal and the chromosomal variation in local admixture ancestry toward better understanding post-hybridization evolutionary outcomes.

KEYWORDS

evolutionary mechanisms, genomic landscape, hybridization, introgression, salmonid, stocking

1 | INTRODUCTION

Understanding the evolutionary consequences of voluntary or involuntary anthropogenic hybridization is of major concern for conservation biology (Allendorf, 2017; McFarlane & Pemberton, 2019;

Waples, 1991). Indeed, anthropogenic hybridization may affect fitness components (survival, growth and reproduction), which may in turn impact population dynamics, genetic diversity and long-term viability (e.g., Allendorf, Hohenlohe, & Luikart, 2010; McFarlane & Pemberton, 2019). However, the consequences of induced gene

flow between foreign and local populations are not well understood and have been considered as potentially either beneficial or harmful to the local populations, depending on the context (e.g., McFarlane & Pemberton, 2019; Todesco et al., 2016). On the positive side, the introduction of foreign individuals may be used to rescue endangered, inbred populations (i.e., genetic rescue) with the goal of increasing the mean fitness of individuals in the local population (Frankham, 2015; Harris, Zhang, & Nielsen, 2019). On the negative side, outbreeding depression may occur when the extent of genetic divergence between populations or species is sufficiently important to cause genetic incompatibilities (i.e., Dobzhansky-Muller Incompatibilities; Orr, 1995; Turelli & Orr, 2000), which may lead to a loss of local adaptation through the disruption of co-adapted genes (Verhoeven, Macel, Wolfe, & Biere, 2011; Waples, 1991). Additionally, while positive effects may be observed in the first hybrid generations by masking the effect of accumulated recessive deleterious alleles (i.e., associative overdominance) (Chen, 2010; Harris et al., 2019; Kim, Huber, & Lohmueller, 2018; Lippman & Zamir, 2007), negative effects may arise in later generations of admixture when maladapted recessive alleles are exposed to selection (Harris & Nielsen, 2016; Harris et al., 2019; Racimo, Sankararaman, Nielsen, & Huerta-Sánchez, 2015).

Therefore, considering both the time since hybridization has occurred and the variation in recombination rate along the genome is critical to distinguish between the immediate and long-term consequences of admixture (Harris & Nielsen, 2016; Harris et al., 2019). Because recombination is expected to progressively reduce the length of introgressed haplotypes across generations following initial admixture (Racimo et al., 2015), the length of introgressed haplotypes can be used as a proxy to estimate the time since hybridization (Gravel, 2012; Racimo et al., 2015). Thus, longer admixture tracts are expected in early hybrids while later hybrid generations tend to display shorter tracts (Leitwein, Gagnaire, Desmarais, Berrebi, & Guinand, 2018; Racimo et al., 2015). Additionally, the local variation in the recombination rate is also expected to affect the introgressed haplotype length with longer and shorter haplotypes expected in lower and higher recombining regions, respectively (Martin & Jiggins, 2017; Racimo et al., 2015). As a consequence, the effects of selection in interaction with recombination should vary along the genome between low- and high-recombination regions. Thus, in genomic regions of low recombination, long introgressed haplotypes may combine the individual effects of multiple selected mutations acting collectively at a block scale (*sensus* Anderson & Stebbins, 1954), as in early-generation hybrids (Leitwein et al., 2018). One could expect such a block effect to generate a lower introgression rate in low recombining regions due to genetic barriers to introgression (Barton, 1983; Duranton et al., 2018; Martin, Davey, Salazar, & Jiggins, 2019; Schumer et al., 2018). Such pattern was observed in swordtail fish hybrid populations for which the introgressed ancestry was more persistent in high-recombining regions where incompatibility alleles uncoupled more quickly. Inversely, a higher introgression rate can also be observed in low-recombining regions due to the presence of recessive deleterious mutations (i.e., associative overdominance;

Harris et al., 2019; Kim et al., 2018). This is because longer haplotypes will be more efficient in masking the effect of multiple recessive deleterious alleles (Harris & Nielsen, 2016). In genomic regions of high recombination rate but also in later hybrid generations, introgressed haplotypes should be shorter and thus selective effects would be more likely to be revealed locally, that is at the locus scale. Therefore, both highly recombining regions and anciently introgressed haplotypes could display either low or high introgression rates depending on the adaptive or maladaptive nature of introgressed alleles at specific loci and the mutation load of the recipient populations (e.g., Harris & Nielsen, 2016; Harris et al., 2019; Racimo et al., 2015).

Variable patterns of genome-wide admixture and introgression may result from the interplay of multiple evolutionary processes (e.g., drift, positive or negative selection for the introgressed alleles) rather than a single, general mechanism. Moreover, antagonistic evolutionary mechanisms (e.g., positive and negative selection) may act differentially across the genome, especially if several pulses of hybridization have occurred, resulting in both historical and contemporary gene flow within a population from an exogenous source (Gravel, 2012; McFarlane & Pemberton, 2019). A few recent studies have investigated how the interaction between recombination rate and selection may modulate the genome-wide temporal dynamics of introgression (e.g., Duranton et al., 2018; Harris et al., 2019; Kim et al., 2018; Martin et al., 2019; Martin & Jiggins, 2017; Schumer et al., 2018).

The general goal of this study was to investigate the level of heterogeneity of admixture along the genome, as well as the underlying evolutionary mechanisms in a freshwater fish, the brook charr. In Québec, Canada, this socio-economically important species has undergone intense stocking from a domestic strain for many decades (detailed in Létourneau et al., 2018). The history of each stocking event with the captive bred fish has been recorded in provincial wildlife reserves, which allows an assessment of the temporal dynamics of the domestic introgression in wild populations (Lamaze, Sauvage, Marie, Garant, & Bernatchez, 2012; Létourneau et al., 2018). In a recent study, Létourneau et al. (2018) documented a negative relationship between the proportion of domestic ancestry and the mean number of years since the most important stocking event. However, they did not investigate the selective consequences (positive or negative) of the introgressed domestic ancestry within wild populations, which thus remains poorly understood. The recent availability of a high-density linkage map developed for *Salvelinus fontinalis* (Sutherland et al., 2016) and a reference genome for the sister species the Arctic Charr (*Salvelinus alpinus*) (Christensen et al., 2018) open new opportunities to investigate how selection in interaction with recombination modulates introgression at a genome-wide scale, as well as within and among chromosomes following human-mediated hybridization events.

More specifically, we used a restriction site-associated DNA sequencing (RADseq) data set collected from 24 brook charr populations that have been stocked with the aforementioned domestic strain to assess genome-wide patterns of variation in local domestic

ancestry for both early- and late-generation hybrids. Moreover, we considered the local recombination rate to investigate which selective effects (positive, negative or neutral) may drive the genome-wide domestic ancestry pattern at three different scales: whole genome, chromosomes and sliding windows of 2 Mb. Finally, we examined how the presence of putative deleterious mutations may modulate genome-wide domestic ancestry, also taking recombination rate into account.

2 | MATERIALS AND METHODS

2.1 | Study system

As with many salmonids, the brook charr is a socio-economically important species that is highly valued for recreational fishing. As a consequence, intensive stocking programmes have been developed to support this industry. The domestic brook charr strain has been reproduced and maintained in captivity for more than 100 years (i.e., around 33 generations considering an approximate mean age at maturity of 3 years) to sustain supplementation programmes (Ministère du Développement Durable, de l'Environnement, de la Faune et des Parcs, 2013). On average, in the province of Québec, Canada, more than 650 tons of brook charr are released annually into the wild (Létourneau et al., 2018; Ministère du Développement Durable, de l'Environnement, de la Faune et des Parcs, 2013), resulting in frequent hybridization between wild and domestic populations (Lamaze et al., 2012; Létourneau et al., 2018; Marie, Bernatchez, & Garant, 2010).

2.2 | Sampling, sequencing and genotyping

The brook charr populations analysed on this study were sampled in 2014 and 2015 (Létourneau et al., 2018) and consist of 611 individuals from 24 lakes located in two wildlife reserves (Mastigouche and St-Maurice) in Québec, Canada (location map is given in Figure S1 adapted from Létourneau et al., 2018). Additionally, 37 domestic fish originating from the *Truite de la Mauricie* Aquaculture Center broodstock were used as reference for domestic samples. Stocking intensity was variable among lakes, as can be seen from the data on the history of stocking, including the number of years since the mean year of stocking (mean_year), the total number of stocking events (nb_stock_ev), the mean number of fish stocked per stocking event (mean_stock_fish) and the total number of fish stocked per hectare (total_ha) as identified by Létourneau et al. (2018) and reported in Table 1. Genotyping by sequencing (GBS) library preparation was performed as in Létourneau et al. (2018), after the extraction of genomic DNA from fin clips. Libraries were amplified by PCR and sequenced on the Ion Torrent Proton P1v2 chip. Raw reads were checked for quality and the presence of adapters with FASTQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), and reads were then demultiplexed with STACKS version 1.40 (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013) with the option *process_radtags* as described by Létourneau et al. (2018). Demultiplexed

reads were aligned to the Arctic charr (*S. alpinus*) reference genome (Christensen et al., 2018) with BWA_MEM version 0.7.9 (Li & Durbin, 2010) before calling of individual single nucleotide polymorphisms (SNPs) with the *pstacks* module (using a minimum depth of coverage $m = 3$ and the bounded error model with the chi-square significance level $\alpha = .05$). To build the catalogue in *cstacks*, we randomly used 10 individuals per population with a coverage depth of at least $10\times$ and a minimum of 500,000 reads. Each individual was then matched against the catalogue with *sstacks*. The *population* module was then run separately for each of the 24 lakes and the domestic strain, in order to generate one VCF file per population with loci passing the following filters: (a) a minimum depth of four reads per locus, (b) a genotype call rate of at least 60% per population, (c) a minimum allele frequency of 2% and (d) a maximum observed heterozygosity of 80%. Individuals with a high percentage of missing data (>20%) were removed, resulting in a final data set of 33 domestic individuals and 603 wild caught individuals (Table 1). Finally, to avoid merging paralogues, we removed for each individual loci with more than two alleles with the R package *STACKR* (Gosselin & Bernatchez, 2016).

2.3 | Inference of local ancestry

Local ancestry inference was performed following the same methodology developed by Leitwein et al. (2018). First, we used the program ELAI version 1.01 (Guan, 2014) based on a two-layer hidden Markov model to detect individual ancestry dosage from the domestic strain along each individual linkage group (LG). The program was run 20 times for each 42 brook charr LGs to assess convergence. Prior to running ELAI, we retrieved the relative mapping positions of our markers along each chromosome after controlling for synteny and collinearity between the Arctic charr and the brook charr genomes. To identify blocks of conserved synteny between the two species, we anchored both the brook charr and the Arctic charr linkage maps (Nugent, Easton, Norman, Ferguson, & Danzmann, 2017; Sutherland et al., 2016, respectively) to the Arctic charr reference genome using MAPCOMP (Sutherland et al., 2016). Results were visualized with the web-based VGSC (Vector Graph toolkit of genome Synteny and Collinearity: <http://bio.njfu.edu.cn/vgsc-web/>). We were thus able to order RAD loci (that were assembled against the Arctic charr reference genome) with respect to their relative positions along each of the 42 brook charr LGs before running ELAI. We then performed local ancestry inference separately for each population, using the 33 domestic individuals as a source population and the wild caught individuals as the admixed population. For each LG in each of the 24 populations, 20 replicate runs of ELAI were performed with the number of upper clusters (-C) set to 2 (i.e., assuming that each fish was a mixture of domestic and wild populations), the number of lower clusters (-c) to 15 and the number of expectation-maximization steps (-s) to 20. Finally, the number of admixed generations (-mg) was estimated using the mean year of stocking and the approximate mean age at maturity of 3 years and ranged from three to 16 (Table 1). We then generated individual domestic ancestry profiles by plotting

TABLE 1 Description of the 24 sampled brook charr lakes in Québec, Canada, along with the stocking variables

Reserve	Lake	Label	N_samples	N_samples_filters	N_SNPs	N_SNPs_mapped	mean_year	n_generation	nb_stock_ev	mean_stock_fish	total_ha	
Mastigouche	Abénakis	ABE	21	21	46,269	31,529	24	8	10	957	1,915	
	Arbout	ARB	28	28	50,376	31,713	38	13	6	317	380	
	Chamberlain	CHA	20	20	42,490	33,290	28	9	9	1,917	958	
	Cougouar	COU	29	29	40,612	33,455	30	10	14	954	1,669	
	Deux-Etapes	DET	28	28	40,806	33,580	22	7	22	2,039	3,647	
	Gélinotte	GEL	25	25	38,716	31,377	10	3	8	1,033	1,652	
	Grignon	GRI	24	23	41,053	33,298	15	5	11	2,086	846	
	Jones	JON	26	26	37,724	30,553	32	11	8	1,638	468	
	Ledoux	LED	27	26	38,262	30,941	33	11	4	13,550	3,956	
	Lemay	LEM	28	28	37,358	30,253	37	12	6	2,909	914	
	Saint-Maurice	Brown	BRO	28	28	70,074	29,716	45	15	4	7,813	114
		Brulôt	BRU	25	25	58,368	35,463	44	15	3	667	247
		Corbeil	COR	23	23	67,425	28,071	45	15	2	2,500	526
		Gaspard	GAS	31	28	61,935	34,844	45	15	2	1,500	259
Maringouins		MAR	26	26	57,489	31,102	35	12	16	416	1,073	
Melchior		MEL	27	26	60,256	30,720	45	15	2	1,000	455	
Milord		MIL	25	25	77,145	41,035	26	9	16	3,428	1,175	
Perdu		PER	24	24	72,604	32,484	36	12	16	3,167	2,293	
Porc-Epic		POE	26	26	68,923	32,100	41	14	6	742	1,648	
Portage		POR	27	27	64,727	34,415	30	10	11	4,451	1,044	
Vierge	Soucis	SOU	22	22	73,117	30,350	39	13	1	750,000	2,804	
	Tempête	TEM	18	17	65,097	32,852	28	9	23	2,057	3,784	
	À la truite	TRU	30	26	62,661	31,741	41	14	7	1,685	1,814	
	Vierge	VIE	29	26	52,919	33,747	48	16	2	594	208	

Note: N_samples: number of individuals per lake; N_samples_filters: number of individuals after filtering; N_SNPs: the number of SNPs after filtering; N_SNPs_mapped: the number of mapped SNPs; mean_year: the number of years since the mean year of stocking; nb_stock_ev: the total number of stocking events; mean_stock_fish: the mean number of fish stocked per stocking event; and total_ha: the total number of fish stocked per stocking event.

the estimated number of domestic allele copies for each replicate run and its median along each LG with R (R Core Team, 2015). The pipeline we used is available on GitHub (https://github.com/mleitwein/local_ancestry_inference_with_ELAI). We then compared the percentage of domestic ancestry computed here with ELAI to the previous study from Létourneau et al. (2018) using a Spearman's correlation.

2.4 | Estimation of length and number of domestic ancestry tracts

The number and length of domestic tracts were determined based on the positions of junctions retrieved from ELAI ancestry dosage output, following the method used and detailed by Leitwein et al. (2018). When the ELAI domestic ancestry dosage median value was between [0.9 to 1.1] and [1.9 to 2], we considered the presence of one domestic tract (i.e., in heterozygous state) and two domestic tracts (i.e., in homozygous state), respectively. Junction positions within "uncertainty areas" (when the domestic ancestry dosage was either between 0.1 and 0.9 or between 1.1 and 1.9) were determined as the position where the domestic ancestry dosage crossed the 0.5 or 1.5 value (for details see Leitwein et al., 2018). Junction positions were used to estimate the number and length of introgressed domestic tracts for each of the 42 LGs in each population.

2.5 | Hybrid class determination

To define hybrid categories with respect to the number of generations of crossing in nature, we computed the chromosomal ancestry imbalance (CAI) developed by Leitwein et al. (2018). Briefly, the CAI represents the cumulated length differences of the domestic ancestry between the two parental chromosome copies divided by the respective LG length. Thus, pure domestic or pure wild individuals have a CAI of 0 whereas F_1 hybrids between wild and domestic parents have a CAI of 1. Due to uncertainty concerning individual haplotype structure (i.e., local ancestry profiles were inferred from unphased domestic ancestry dosage), we could not precisely determine the CAI of admixed genotypes resulting from several generations of admixture. Therefore, we conservatively classified individuals as early-generation hybrids when their CAI was ≥ 0.25 , and as late-generation hybrids when their CAI was ≤ 0.125 . Early-generation hybrids therefore correspond to F_1 , F_2 , first-generation backcrosses and other types of crosses generated among hybrids and parental pedigrees during the first generations of admixture. By contrast, late-generation hybrids comprise genotypes that mostly comprise wild-type ancestry, while being introgressed by varied proportions of domestic alleles. In both hybrid categories, we removed individuals with a genome-wide percentage of domestic ancestry equal to 0% (pure wild individual) and higher than 60% (see Figure S2) in order to exclude individuals with a mostly domestic ancestry. This should not impact the following analysis as these individuals are probably pure domestic

individuals or hybrids between F_1 and domestic parents; this concerns only six individuals (Figure S1).

2.6 | Impact of stocking on domestic haplotype number and length

We first assessed the relationships between domestic tract characteristics and the four variables reflecting stocking history (i.e., mean_year, nb_stock_ev, mean_stock_fish and total_ha; Table 1) using linear mixed models. The mean percentage of domestic ancestry, and the number and length of introgressed domestic tracts were treated as dependent variables whereas the stocking variables were introduced in the model as explanatory terms in an additive way, with no interaction to avoid model over-parameterization. The dependent variables were log-transformed and the explanatory variables were scaled (i.e., centred and reduced). Population (24 populations) and region (two regions, Mastigouche and St-Maurice) were introduced as random effects in the model. Normality of the residuals was examined graphically using a quantile-quantile plot. We used a likelihood ratio test to assess the significance of the tested relationship by comparing the models with and without the explanatory term. We calculated marginal R^2 to quantify the proportion of variance explained by the explanatory variable only. All analyses were performed in R using the package "LME4" (Bates, Mächler, Bolker, & Walker, 2015).

2.7 | Domestic ancestry profiles as a function of hybrid classes

The fraction of domestic ancestry rate was estimated separately for both hybrid classes, with the early-generation hybrids comprising 47 individuals from 17 populations and the late-generation hybrids comprising 358 individuals from 24 populations. Local ancestry rate from the domestic strain was estimated at 11,803 SNP positions distributed along the 42 LGs and common between all individuals originating from different lakes and found in the two hybrid categories. Estimates of genome-wide domestic ancestry rate for each hybrid category were plotted with R (R Core Team, 2015) (https://github.com/mleitwein/local_ancestry_inference_with_ELAI). The mean domestic ancestry rate and its 95% confidence interval (CI) were reported along the 42 LGs. Then, genomic regions exceeding the 95% CI were considered as displaying an excess or deficit of domestic ancestry.

Linear mixed models were used to investigate how domestic ancestry profiles (i.e., mean number and length of domestic tracts) differ between early and late hybrid classes. The domestic tract characteristics were log-transformed and treated as dependent variables whereas the hybrid class (discrete variable with two modalities) was introduced in the model as an explanatory term. The population (24 populations) and region (two regions, Mastigouche and St-Maurice) terms were introduced as random effects in the model. All analyses were performed in R using the package "LME4" (Bates et al., 2015).

2.8 | Local recombination rate estimation

To estimate genome-wide variation in recombination rate, we anchored the brook charr mapped RAD loci from Sutherland et al. (2016) to the Arctic charr reference genome. The relative positions of these loci were extracted from the `BWA_MEM` alignment for each collinearity block identified with `MAPCOMP` (Sutherland et al., 2016) allowing the reconstruction of a brook charr collinear reference genome. The local variation in recombination rate was estimated across the collinear reference genome by comparing the physical (bp) and genetic position (cM) of each marker using `MAREYMAP` (Rezvoy, Charif, Guéguen, & Marais, 2007). The polynomial Loess regression method was used to assess the recombination rate with a degree of smoothing (*span*) set to 0.9. Finally, to estimate the recombination rate of markers that were not included in the linkage map, we computed the weighted mean recombination rate using the two closest markers based on their relative physical positions.

2.9 | Relationships between domestic ancestry and recombination rate

To investigate the selective forces shaping the domestic ancestry pattern along the genome, we took in consideration the recombination rate. Indeed, a positive correlation between recombination rate and domestic ancestry rate reflects selective forces against domestic ancestry in low-recombining regions. Inversely, a negative correlation will reflect selective forces favouring domestic ancestry in low-recombining regions. As the recombination rate is highly variable along the genome, we investigated the relationship between the local domestic ancestry rate and the recombination rate at three different levels: (a) at the whole genome level, (b) at the LG level and (c) within 2-Mb windows. For the three levels, we used regression models in which the local domestic ancestry rate was treated as a dependent variable and the recombination rate was incorporated as an explanatory term. In all models, the domestic ancestry rate was log-transformed and the recombination rate was centred-reduced. At the whole genome level, we used a linear mixed model in which the LGs and the 2-Mb windows were introduced as random effects. By doing so, we considered the nonindependency of the domestic ancestry rate estimates by specifying that the estimates belong to a given window within a given LG. The hybrid class (early and late) was also added as an explanatory term in an interactive way (recombination \times class). The marginal R^2 describing the proportion of variance explained by the fixed factors (i.e., the recombination rate and the hybrids class) was calculated and we assessed the significance of the hybrid class effect using a likelihood ratio test. At the LG level, to avoid model over-parameterization, the analyses were performed separately for the early and late hybrids. A linear mixed model was built for each of the 42 LGs and the 2-Mb sliding window was introduced as a random effect in all models. The slope coefficient (and the 95% CI) of the relationship between domestic ancestry rate and recombination rate was reported for the 42 LGs. To correct for multiple

testing, the conservative Bonferroni correction was applied such that only values with $p < .001$ were considered significant (i.e., $\alpha_{\text{corrected}} = .05/42$ LGs) (Henry, 2015). At the 2-Mb window level, generalized linear models were built for each window containing at least five positions with an estimated recombination and introgression rate (resulting in 689 models) because a regression analysis could not be performed with fewer values. The slope coefficient and its 95% CI were reported for each sliding window. In all models, the significance of the effect of the explanatory terms was assessed with a likelihood ratio test. All analyses were performed in R using the package “`LME4`” (R Core Team, 2015). To control for serial correlation potentially that could affect the within-LG regression analyses we performed a Mantel test in R using the package “`ADE4`” (Dray & Dufour, 2007).

2.10 | Relationships between potentially deleterious mutations and ancestry rate

All SNPs present in all populations were used for the identification of potentially deleterious mutations. First, read sequences associated with the SNPs were blasted against the Arctic charr proteome available on NCBI (https://www.ncbi.nlm.nih.gov/assembly/GCF_002910315.2). All hits with a minimum amino acid sequence length alignment of 25 and a similarity in protein sequence of $\geq 70\%$ between the sequences of interest and the reference proteome were retained (those thresholds have been optimally chosen after running several tests with different parameter combinations on the observed data). Then, `PROVEAN` (Protein Variation Effect Analyzer; Choi, Sims, Murphy, Miller, & Chan, 2012) was used to predict the deleterious effect of nonsynonymous mutations. As in previous studies (e.g., Ferchaud, Laporte, Perrier, & Bernatchez, 2018; Renaut & Rieseberg, 2015) a threshold of -2.5 in Provean score was applied to distinguish between nonsynonymous mutations potentially deleterious (≤ -2.5) and neutral (> -2.5). In `PROVEAN`, the deleteriousness of a variant can be predicted based on its effect on gene functioning (such as protein changing, stop-gain, stop-loss), for example by assessing the degree of conservation of an amino acid residue across species. The pipeline used for the entire process is available on github (`gbs_synonymy_genome`: https://github.com/QuentinRougemont/gbs_synonymy_with_genome).

To evaluate the relationship between the presence of potentially deleterious mutations and the domestic ancestry rate as a function of recombination rate, a mean domestic ancestry and recombination rate was estimated in a window size of 400 kb surrounding the position of the potentially deleterious mutation (i.e., 200 kb before and after). The analysis was performed at the genome-wide level and LG level. For both levels, we first applied a regression model in which the mean 400-kb window domestic ancestry rate was treated as a dependent variable and the associated recombination rate window was incorporated as an explanatory term. At the genome-wide scale, a linear mixed model was performed and LGs were incorporated as a random effect and the recombination rate as a control co-variable. We then performed a linear model in which the residuals of the model (domestic ancestry rate \sim recombination rate) were included as a dependent variable and

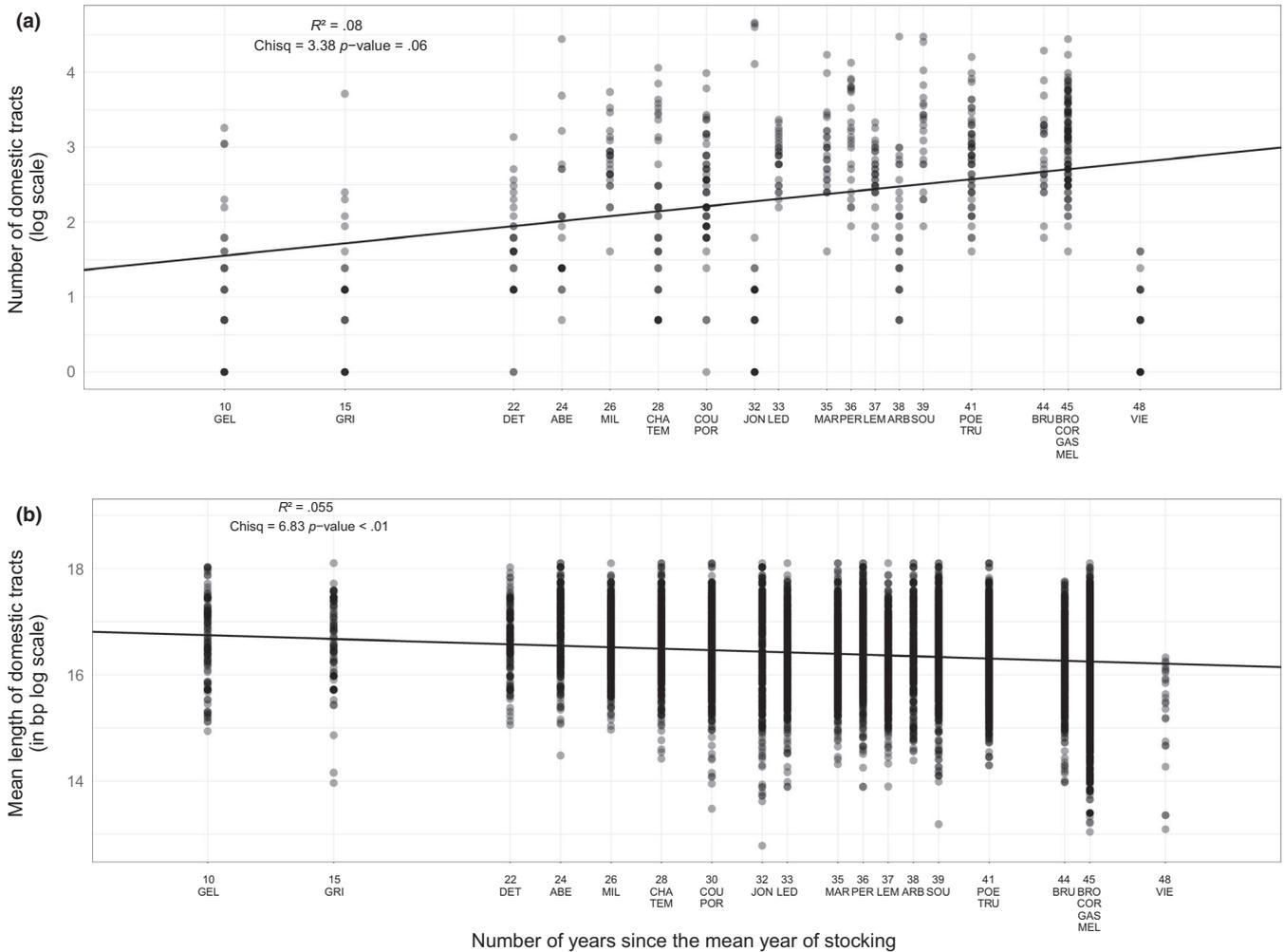


FIGURE 1 Relationships between domestic tract features (number and the mean length) and stocking variables. (a) Positive correlation between the number of domestic tracts (log scale) per individual as a function of the number of years since the mean year of stocking events for each sampled lake (lakes labels are described in Table 1; $R^2 = .08$, $p = .06$). (b) Negative correlation between the mean length of domestic tracts in base pair log scale per individual as a function of the number of years since the mean year of stocking events for each sampled lake (lakes labels are described in Table 1; $R^2 = .055$, $p < .01$)

the presence of potentially deleterious alleles (Provean score < -2.5) or neutral (Provean score > -2.5) as an explanatory term (i.e., discrete variable with two modalities, “deleterious” vs. “neutral”). At the LG level, similar linear models were built for each LG.

3 | RESULTS

3.1 | SNP calling

Demultiplexed and cleaned raw reads resulted in an average of 2.38 million reads per individual. After filtering, 636 individuals were kept for further analyses. On average, $1,225,704 \pm 285,583$ reads were properly mapped to the Arctic charr reference genome with an average depth of $11.5 \pm 2.4 \times$ per individual. After applying population filters, an average of $55,267 \pm 11,779$ SNPs were kept per population for subsequent analysis (Table 1).

3.2 | Domestic ancestry tracts detection and hybrid classes

We performed local ancestry inference with ELAI with an average of $32,442 \pm 1,799$ SNPs per population that were mapped to the reconstructed brook charr reference genome (see Methods and Table 1). Individual local ancestry was summarized across LGs to determine the number and mean length of domestic ancestry tracts at the individual level. Both domestic ancestry tract length and abundance were highly variable among individuals, ranging from 356 kb to 72 Mb in length and from one to 103 for the number of domestic ancestry tracts per individual (Tables S1 and S2). A total of 47 individuals were assigned to the early hybrid-generation class ($CAI \geq 0.25$) and 358 to the late hybrid-generation class ($CAI \leq 0.125$; Table 1). A total of 195 remaining fish were unclassified and therefore considered as intermediate between

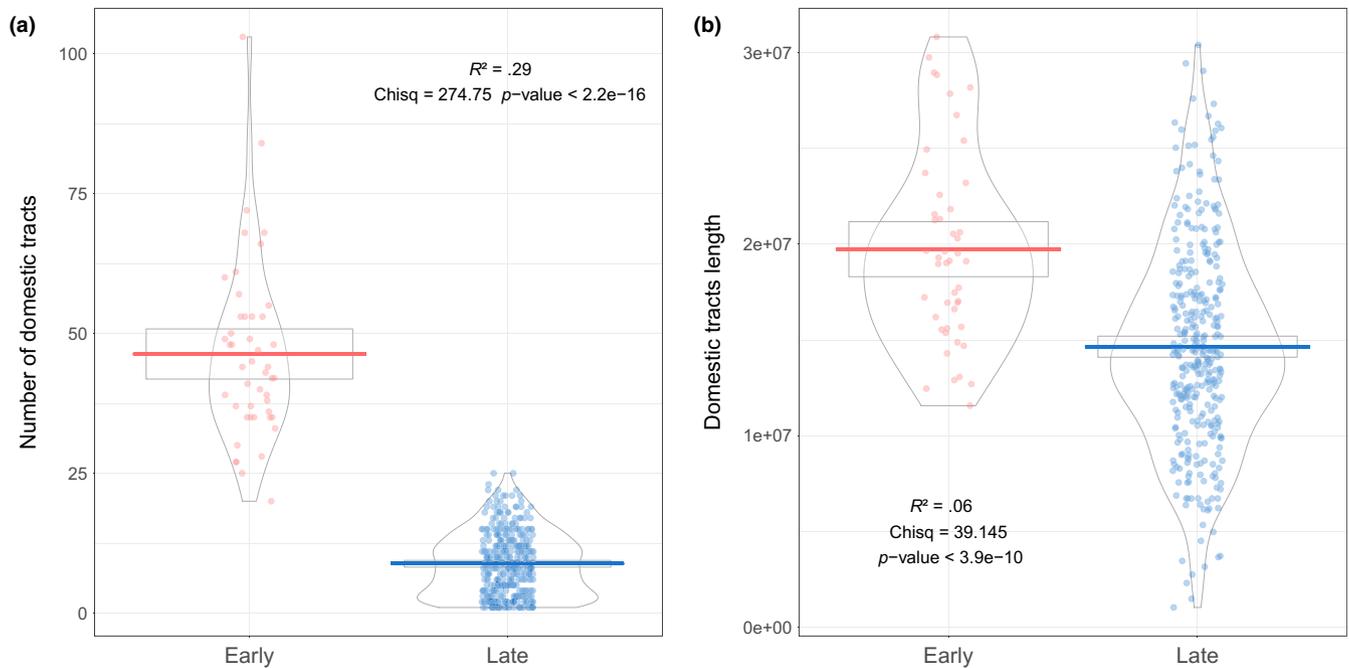


FIGURE 2 Domestic tract features and hybrid generations (early and late). The number (a) and the length (in bp) (b) of domestic tracts as a function of the hybrid categories; early (in pink) and late (in blue). Boxes indicate 95% CI, horizontal line represents the mean and violins indicate the density

early- and late-generation hybrid categories based on these empirical thresholds. The mean individual percentage of domestic ancestry was significantly positively correlated (Spearman's $\rho = 0.68$, $p < 10^{-10}$, Figure S3) with the percentage of domestic ancestry computed with *ADMIXTURE* by Létourneau et al. (2018) although that was based on a much smaller number of markers (~4,579 SNPs per population).

3.3 | Influence of stocking history on domestic tract number and length

No significant relationship was found between the mean number of years since stocking or the intensity of stocking and the mean individual percentage of domestic ancestry within each population (Table S1). However, the number of domestic tracts tended to be higher for populations for which supplementation began more than 30 years ago. Indeed, a marginally significant positive relationship was detected between the mean number of domestic tracts and the number of years since the mean year of stocking ($R^2 = .08$, $p = .06$, Figure 1a). Also, a significant negative correlation was observed between the mean length of individual domestic tracts and the number of years since the mean year of stocking ($R^2 = .055$, $p < .01$, Figure 1b). Therefore, the length of the domestic ancestry tracts tended to decrease over time, whereas it tended to increase with the number of stocking events (Figure S4, $R^2 = .055$, $p = .049$). The number of stocking events was also negatively correlated with the mean year of stocking practices (Figure S5, Spearman's $\rho = -0.75$, $p < .001$).

3.4 | Domestic tract characteristics as a function of hybrid classes

The mean number of domestic ancestry tracts was significantly lower for the late hybrid individuals ($\text{mean}_{\text{number}} = 8.01$) compared to the early hybrid individuals ($\text{mean}_{\text{number}} = 46.31$; Figure 2a, $R^2 = .22$, $p < 2.2e-16$). Moreover, the mean length of the introgressed domestic tracts was shorter for the late hybrid ($\text{mean}_{\text{length}} = 14.6$ Mb) compared to early hybrid individuals ($\text{mean}_{\text{length}} = 19.7$ Mb; Figure 2b, $R^2 = .06$, $p < 3.9e-10$).

3.5 | Domestic ancestry and recombination rates

The genome-wide level of domestic ancestry rate was estimated for both early and late hybrid generations using 11,803 SNPs (i.e., corresponding to common polymorphic sites shared among lakes and found in the two hybrid categories) distributed along the 42 brook charr LGs. The mean genome-wide domestic ancestry rate in early hybrid generations was 0.307 (95% CI 0.18–0.43; Figure S6a) and was highly variable within and among LGs with several genomic regions displaying a local excess (LGs 1, 9, 11 and 12, Figure S6a) or deficit of domestic ancestry (LGs 14, 28, 30 and 40; Figure S6a). The mean domestic ancestry rate of late hybrid generations was significantly lower with a mean of 0.042 (95% CI 0.012–0.081; Figure S6a), and genomic regions displaying excess (LGs 9, 12, 20 and 41) or deficit (LG14) of domestic ancestry were found (Figure S6b). Thus, two LGs displaying an excess (LGs 9 and 12) and one displaying a local deficit of domestic ancestry (LG14) were found both in early and in late hybrid generations.

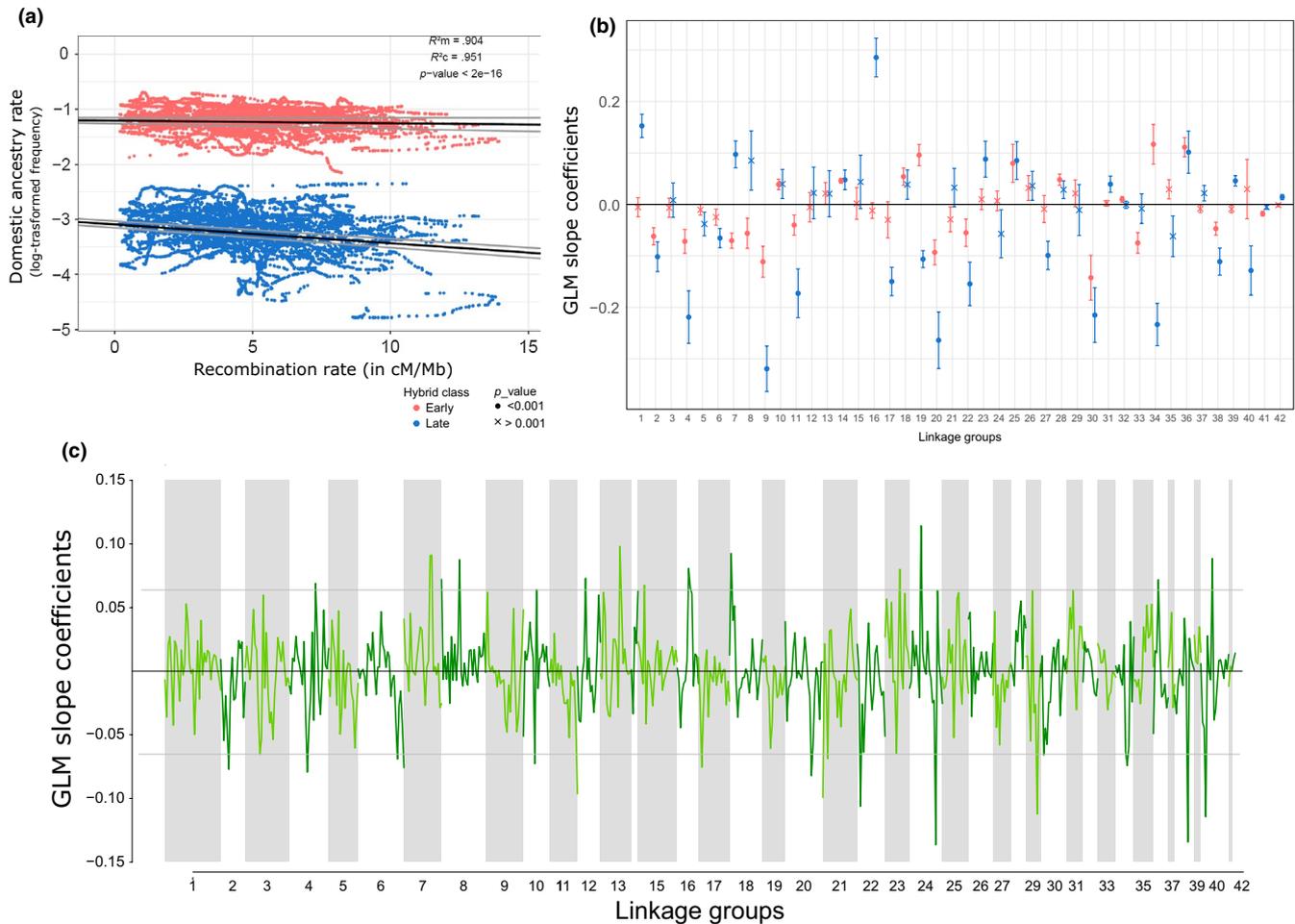


FIGURE 3 Relationship between the domestic introgression rate and the recombination rate (cM/Mb) at different genomic scales. (a) The whole genome scale; negative correlation between the log-transformed introgression rate (frequency among individuals) and the recombination rate (cM/Mb) for both early (in pink) and late (in blue) hybrid categories assessed with a generalized linear mixed model (GLM); $\beta_{\text{early}} = -0.07$, $\beta_{\text{Late}} = -0.09$; $R^2_m = .90$, $R^2_c = .95$; LRtest $\chi^2_{\text{Recombination}} = 441.7$, $\chi^2_{\text{Hybrids}} = 60,420$, $\chi^2_{\text{Hybrids_interaction}} = 177.57$; $p < 2.2e-16$). (b) The linkage group level; slope coefficient of the GLM of the introgression rate and recombination rate models for each 42 Brook charr linkage group for both early (in pink) and late (in blue) hybrid categories. Dots represent the significance of each model with the dark circle and cross representing $p < .001$ and $p > .001$, respectively. (c) The 2-Mb sliding windows level; slope coefficient of the GLM of the introgression rate and recombination rate models of the late hybrids for each 2-Mb sliding window along the 42 brook charr linkage groups. Grey lines represent the 95% confidence interval

The local recombination rate (r) was estimated at 10,740 SNPs out of the 11,803 SNPs and was found to be highly variable within and among LGs, with a mean value of 4.48 ± 1.83 cM/Mb (Figure S6c) over the 42 LGs. At the genome-wide scale, for both early- and late-generation hybrids, a negative correlation was found between the proportion of domestic ancestry and recombination rate (Figure 3a; slope coefficient $\beta_{\text{early}} = -0.07$, $\beta_{\text{Late}} = -0.09$; $p < 2.2e-16$). The marginal R^2 of the full model was equal to 0.90, the likelihood ratio test was significant for recombination rate ($\chi^2_{\text{Recombination}} = 441.7$, $p < 2.2e-16$), hybrid class, which displayed the strongest effect ($\chi^2_{\text{Hybrids}} = 60,420$, $p < 2.2e-16$), and the interaction between recombination and hybrid class ($\chi^2_{\text{Hybrids_interaction}} = 177.57$; $p < 2.2e-16$). Moreover, the interaction between recombination rate and hybrid class was significant ($\beta_{\text{early:Late}} = 2.8$; $p < 2.2e-16$), indicating a stronger negative correlation between introgression

and recombination rate in late- compared to early-generation hybrids.

The results of the linear mixed model fitted at the LG level are presented in Figure 3b (mean and 95% CI) for both hybrid classes. For the late-generation hybrids, 10 LGs showed a significant positive correlation and 14 LGs showed a significant negative correlation between the domestic ancestry rate and recombination rate (Figure 3b). For the early-generation hybrids, nine and 12 LGs respectively showed a significant positive and a significant negative correlation between the domestic ancestry rate and recombination rate (Figure 3b). Eight LGs displayed significant negative correlations between domestic ancestry and recombination for both early- and late-generation hybrid classes, whereas only three LG consistently showed positive correlations. Conversely, three LGs presented opposite correlations between the early- and the late-generation hybrid classes (LGs 7, 19 and 34).

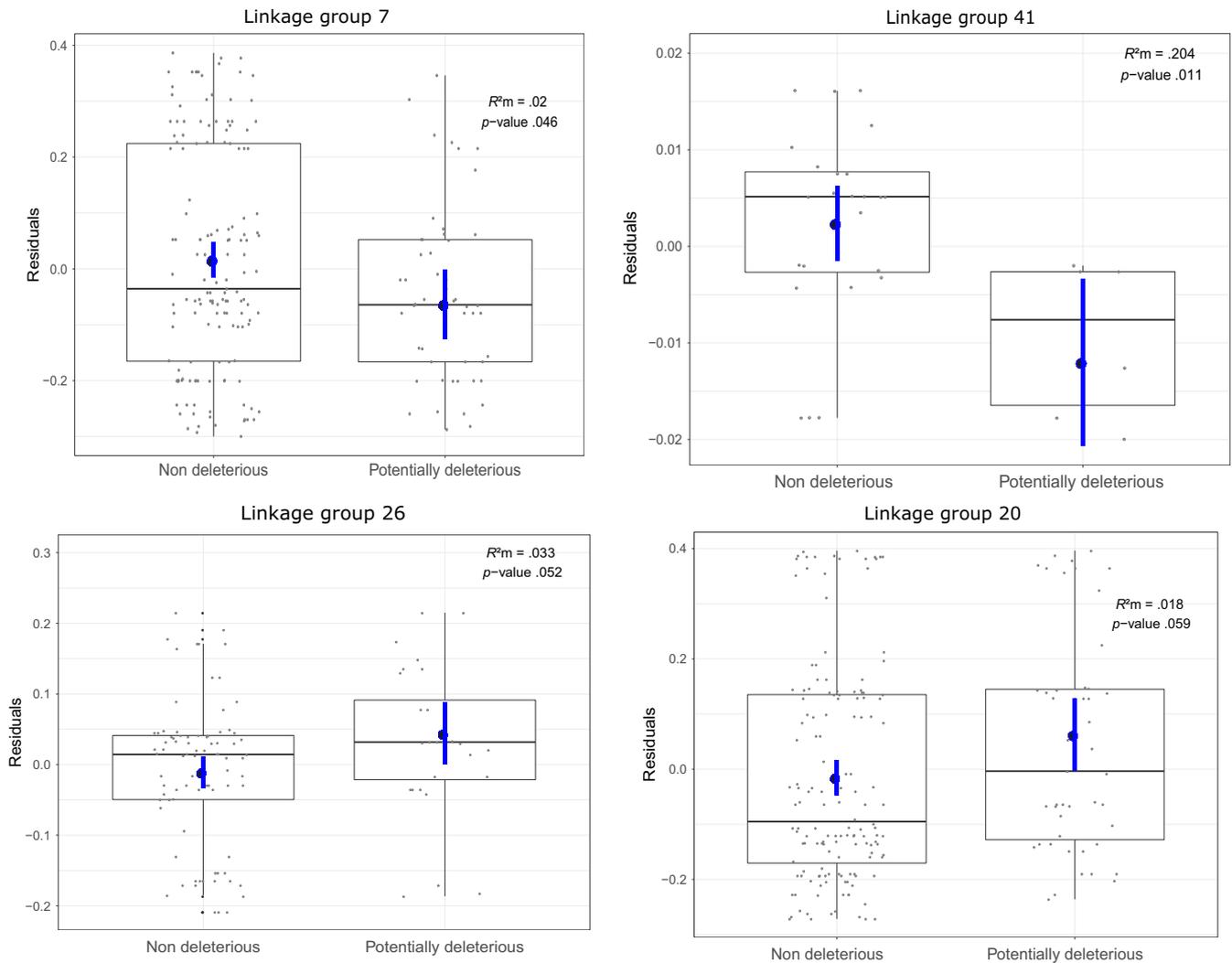


FIGURE 4 Domestic ancestry rate excess or deficit and presence of potentially deleterious alleles. Residuals of the regression between the domestic ancestry and recombination rates models as a function of the presence of nondeleterious or potentially deleterious alleles for four linkage groups in late hybrids generations. LG7 and LG41 displayed a lower introgression rate of domestic ancestry compared to model prediction in genomic regions surrounding potentially deleterious alleles. LG26 and LG20 displayed an excess of introgressed domestic ancestry compared to the model expectation around potentially deleterious alleles. Boxes indicate 95% confidence intervals, and horizontal line represents the median. Blue dots and blue line represent the estimates and the confidence intervals of the linear model estimates. Marginal R^2 (R^2m) and significance are displayed at the upper right of each plot

We then focused the window-scale analyses on late-generation hybrids only because the effect of selection needs more than a few generations after hybridization to leave detectable footprints at such a local scale. The estimated slopes of linear models performed for each 2-Mb sliding window in late-generation hybrids were highly variable within and among LGs, and several genomic regions exceeded the 95% CI of the estimated distribution (grey lines; Figure 3c). For example, 12 genomic regions showed a strong positive correlation (e.g., LGs 7, 8, 13, 16, 24, 40) and 15 genomic regions showed a strong negative correlation (e.g., LGs 2, 12, 22, 24, 29, 38, 40; Figure 3c) between local domestic ancestry rate and recombination rate. No spatial autocorrelations were observed between our 2-Mb windows within each LG (Table S3).

3.6 | Potentially deleterious mutations

Among a total of 141,332 sequences that were blasted against the Arctic charr reference proteome (Christensen et al., 2018), 36,743 loci had significant hits and were retrieved. Among these, 21,288 had more than 70% similarity between the sequences of interest and the reference proteome and a mean length greater than 25 nucleotides. Finally, 7,799 were nonsynonymous and among them 1,932 were potentially deleterious (Provean score < -2.5).

At the genome scale, we did not detect any significant correlation between the residuals of the linear relationship between local ancestry and recombination rate and the presence of deleterious mutations, in both early- ($\beta_{\text{early}} = -0.0015$; $p = .6137$) and late-generation hybrids ($\beta_{\text{late}} = -0.0114$; $p = .0713$; Figure S7). At the LG

scale, two LGs (LGs 4 and 15) showed a negative relationship between the residuals and the presence of potentially deleterious alleles for the early hybrids (Figure S7a,b; LG4: $R^2 = 0.018$, $p = .04$; LG15: $R^2 = 0.02$, $p = .028$). For the late hybrids, two LGs showed a significant negative correlation, LG7 ($R^2 = 0.02$, $p = .045$) and LG41 ($R^2 = 0.20$, $p = .01$), and two LGs showed a marginally significant positive correlation, LG20 ($R^2 = 0.01$, $p = .059$) and LG26 ($R^2 = 0.033$, $p = .051$), between the presence of potentially deleterious alleles and residuals of the linear mixed model of domestic ancestry rate and the recombination rate (Figure 4).

4 | DISCUSSION

The main goal of this study was to document the genetic outcomes of more than four decades of stocking domestic brook charr into wild populations. To this end, we assessed genome-wide patterns of domestic ancestry within 24 wild populations with different histories of stocking. We more specifically considered the relationship between recombination rate and domestic ancestry rate at different scales: (a) the whole genome, (b) individual LGs and (c) within 2-Mb windows. This allowed us to detect a wide range of patterns that can be attributed to different evolutionary processes acting across the genome. Among these, the main factor modulating local domestic ancestry was probably associative overdominance, leading to an increased frequency of domestic ancestry within low-recombining regions. This pattern might be explained by the presence of mildly deleterious recessive mutations in the wild populations. At the regional scale (i.e., chromosomes and 2-Mb windows), inverted correlations were found, which suggested local selection against domestic ancestry. Our results highlight the importance of taking into consideration genome-wide variation in both recombination and ancestry rates to understand the evolutionary outcomes of human-induced admixture.

4.1 | History of supplementation impacts domestic ancestry

No significant relationship was observed between the mean individual proportion of domestic ancestry and the main stocking variables previously identified as having an impact on the levels of domestic introgression (i.e., mean year since stocking, number of stocking event, number of fish stocked per hectare and the number of stocking events; L  tourneau et al., 2018). This discrepancy could be explained by different sensitivities of the methods. Indeed, although the correlation between individual domestic ancestry proportions estimated here and in L  tourneau et al. (2018) was good (Spearman's $\rho = 0.68$, $p < 10^{-10}$), low proportions of domestic ancestry were not detected in the previous study, which resulted in considering weakly introgressed individuals as pure wild genotypes (Figure S2). Moreover, thanks to the higher number of markers and their positions we were able to retrieve the approximate length and number of domestic tracts within each population. Consistent with theoretical

predictions (Racimo et al., 2015), we observed an increase in the number of domestic ancestry haplotypes and a decrease in the mean domestic haplotype length as a function of the number of years since the main stocking event. This corroborates findings by Leitwein et al. (2018) who suggested that the mean length of foreign ancestry could be used as a proxy to retrieve the history of stocking practice. We indeed observed a higher number of smaller introgressed domestic haplotypes for the lakes where stocking has stopped earlier in the past. We also observed longer domestic haplotypes for the lakes displaying a higher number of stocking events, as well as more recent stocking events (Figure S4). Additionally, the length and the number of domestic tracts also allowed us to distinguish among early- and late-generation-hybrids within populations, which is important as the lakes have undergone several successive events of supplementation. The presence of early hybrids (i.e., F_1 , F_2 and backcrosses of first hybrid generations) is not surprising considering the relatively recent events of stocking practices and the presence of pure domestic individuals remaining in some lakes (Figure S2, Table 1 and see L  tourneau et al., 2018 for more details). The time since hybridization, by between distinguishing early- and late-generation hybrids (i.e., individuals with a small number of short domestic haplotypes), is also important to assess the potential evolutionary outcomes of supplementation. Indeed, selection is expected to be more efficient at a locus scale in later hybrid generations when introduced domestic haplotypes have been sufficiently shortened by recombination. We note that because more individuals were available for the late hybrids than for the early hybrids, this could result in slight differences in analysis power.

4.2 | General tendency of selective effects and putative deleterious mutations

Both the time since hybridization and the genomic scales (i.e., global vs. local) were important to assess the selective outcomes of gene flow between wild and domestic individuals. At the genome-wide scale, we observed a negative correlation between domestic ancestry and recombination rate for both early- and late-generation hybrids. While we detected highly variable selective effects along the genome, this negative correlation was also predominant at the LG and at the local scales (i.e., more genomic regions exceeding the 95% CI for negative slope at the 2-Mb window scale) for the late hybrid generations. Such a negative correlation could be the result of different evolutionary mechanisms. First, the presence of beneficial mutations under positive selection may locally increase the introgression of foreign alleles in low recombining regions where hitchhiking of neutral foreign alleles could occur (Charlesworth, 2009; Fay & Wu, 2000; Felsenstein, 1974). For example, adaptive introgression of Neanderthal ancestry has been reported in modern humans and interpreted as an adaptation to high altitude in Tibetan populations (Huerta-S  nchez et al., 2014) and the immune response (Dannemann & Racimo, 2018; Racimo et al., 2015). However, this mechanism is more likely to explain local-scale correlations than genome-wide patterns. Second, given the generally small effective population size

of these brook charr lacustrine populations (Gossieux, Bernatchez, Sirois, & Garant, 2019), random drift may also be responsible for variable ancestry locally along the genome (Martin & Jiggins, 2017) and is not expected to produce consistent patterns across lakes. Third, the most likely evolutionary mechanism that could explain the general negative correlation would be a dominant effect of associative overdominance (Kim et al., 2018). Indeed, favoured domestic ancestry within regions of low recombination rate is likely to reflect the action of associative-overdominance, especially if the recipient population tends to accumulate recessive deleterious alleles preferentially in low-recombining regions (Charlesworth & Willis, 2009). Consequently, long introgressed foreign haplotypes of domestic origin would be favoured by masking the effect of linked recessive deleterious mutations present in the small local populations (i.e., local heterosis effect; Charlesworth & Willis, 2009; Kim et al., 2018). Kim et al. (2018) suggested that the presence of slightly deleterious recessive mutations might modulate the genome-wide introgression rate in hybrid populations. More specifically, if the mutation load of receiving (wild) populations is higher than in the introduced domestic strain, a negative correlation between domestic ancestry rate and recombination rate is expected (Schumer et al., 2018). Here, when considering the genome-wide scale, we did not observe any relationship between the presence of putative deleterious mutations and the residuals of the linear model relating introgression to recombination rate. Although the domestic strain is genetically more diversified than the wild populations (Létourneau et al., 2018; Martin, Savaria, Audet, & Bernatchez, 1997), we could not determine if the potential deleterious mutations were originating from the domestic strain or the wild populations. However, we were able to detect an increase of domestic ancestry in the presence of deleterious mutations for some of the chromosomes (LGs 26 and 20). This may reflect a positive effect (i.e., associative over-dominance) of the presence of domestic ancestry in these regions. This interpretation is congruent with the hypothesis of temporarily reduced genetic load, caused by the masking of deleterious mutations, and resulting in an increase of introgressed ancestry as observed in a simulation study (Kim et al., 2018). This is also consistent with the expected higher accumulation of deleterious mutations in small lacustrine brook charr populations as observed by Ferchaud et al. (2019) as well as with the higher allelic richness observed in the domestic strain compared to these small lacustrine populations (Martin et al., 1997). Finally, this is also consistent with the fact that the domestic strain and the wild populations are not highly divergent as the domestic strain has been created only ~33 generations ago (Létourneau et al., 2018). Therefore, genetic incompatibilities are, if any, not expected to be frequent between domestic and wild brook char.

4.3 | Variability of selective effects revealed at local scales

At local genomic scales, molecular signatures suggesting the action of variable selective effects were observed along the genome. In contrast to the general tendency, some LGs displayed strong positive

associations between the recombination rate and the domestic ancestry rate, which might be explained by several mechanisms. (a) Local variation could reflect the stochastic outcomes of genetic drift (Martin & Jiggins, 2017). (b) The presence of hybrid incompatibilities (i.e., selection against introgressed alleles) may also result in lower foreign ancestry in low recombining regions (Schumer et al., 2018). Such a pattern has been observed in swordtail fish species where the retention of minor parent ancestry was more pronounced in highly recombining regions (Schumer et al., 2018), as well as in European sea bass (Duranton et al., 2018). Similarly, Martin et al. (2019) observed stronger barriers to introgression within longer chromosomes displaying lower average recombination rate on average. However, the occurrence of multiple hybrid incompatibilities seems unlikely here as the domestic strain is not highly divergent from the wild populations. Finally, (c) in the situation of selection against domestic haplotypes due to unconditionally deleterious introgression (or hybridization load), foreign loci of small individual effect are expected to be removed more quickly within low-recombining regions because selection is more effective in removing linked deleterious mutations (Kim et al., 2018; Martin & Jiggins, 2017; Schumer et al., 2018). We observed such a pattern for two LGs (7 and 41) displaying a deficit of domestic introgression in the presence of potentially deleterious mutations. Moreover, LG7 broadly displayed lower domestic introgression within low-recombining regions. Together, these combined pieces of information suggest that selection has been acting against the domestic haplotypes potentially carrying deleterious alleles.

4.4 | Importance of the time since hybridization

The time since hybridization seems to be an important parameter that determines the fate of admixture tracts, as different ancestry patterns were observed as a function of the hybrids category considered. Late-generation hybrids displayed a stronger negative correlation between domestic ancestry and recombination rate and even reversed pattern for three LGs (7, 19 and 34) compared to the early hybrid generations. These discrepancies between the early- and late-generation hybrids could be caused by a time-specific effect of selection mediated by recombination. Indeed, the outcomes of hybridization are expected to vary with the number of generations elapsed since the introduction of foreign alleles into the population. If the hybridization event is recent, long introgressed haplotypes are expected because it takes time for recombination to break down domestic tracts across generations (Racimo et al., 2015). Thus, the selective effects in recent hybrid generations are expected to act at the scale of large haplotype blocks, as in low-recombining regions. Inversely, smaller introgressed haplotypes are expected for older hybridization events (Racimo et al., 2015) and thus selective effects would act at a more localized (i.e., small blocks or locus) scale, as for high-recombining regions (see above) (McFarlane & Pemberton, 2019). Here, we suspect that for the early hybrid individuals, the admixture events were too recent for selection to efficiently occur and be detectable. Continuously, in our late hybrid individuals, although the introgressed haplotypes are smaller than the early hybrid individuals they are still long (~15 Mb) and selection would tend to act at the block scale, explaining the predominance

of the associative overdominance effects. As a consequence, the hybridization events studied here are still too recent to fully assess the long-term selective outcomes of supplementation. Indeed, while positive effects of associative overdominance are predominant in our study, it is possible that later on maladaptative introgressed domestic alleles reveal their individual effects, which could become detrimental for the supplemented populations over the long term (Harris & Nielsen, 2016).

4.5 | Implications for conservation biology

Understanding the evolutionary outcomes of anthropogenic hybridization is of major concern for conservation biology and management (Allendorf, 2017; McFarlane & Pemberton, 2019). Indeed, supplementation with a foreign population could either be beneficial for the recipient population (i.e., genetic rescue; Aitken & Whitlock, 2013; Allendorf et al., 2010; Goedbloed et al., 2013; Uller & Leimu, 2011), or could induce outbreeding depression because of maladaptation, loss of local adaptation or hybrid incompatibilities (Orr, 1995; Randi, 2008; Verhoeven et al., 2011; Waples, 1991). This is especially true when stocking is done with highly differentiated captive breeds, as is widely the case in salmonid species (Araki, Berejikian, Ford, & Blouin, 2008; Hansen, Fraser, Meier, & Mensberg, 2009; Naish et al., 2007; Waples, 1991). Here, quantifying the relationship between introgression and recombination rates allowed us to reveal complex patterns of selective effects acting along the genome. Such patterns showed that interpreting the consequences of anthropogenic hybridization is not straightforward, because several antagonistic mechanisms may cause both positive and negative effects of domestic ancestry to co-occur along the genome. From a conservation point of view, it is important to weigh the pros and cons of those different outcomes and evaluate the main conservation focus (i.e., the species, the population or the local genetic inheritance). Our method may help to consider the potential evolutionary consequences of hybridization in the short and possibly long terms in order to take sound management decisions. Moreover, evaluating the length and distribution of introgressed haplotypes allows determination of the approximate time since hybridization (Duranton, Bonhomme, & Gagnaire, 2019; Leitwein et al., 2018) and hybrid categories, which can influence management practices and conservation decisions (Allendorf, Leary, Spruell, & Wenburg, 2001). Indeed, a population carrying a small proportion of domestic ancestry might be of more interest from a conservation standpoint than a population carrying a stronger one. This seems particularly important when the history of supplementation is unknown, for instance due to illicit stocking (Johnson, Arlinghaus, & Martinez, 2009). Moreover, the selective forces modulating the introgression rate along the genome will be influenced by the length of introgressed haplotypes, which is dependent on both the number of generations since hybridization and the recombination rate. As a result, both the time and the recombination rate variation are important to understand the consequences (positive or negative) of hybridization with a foreign population and to better orientate management decisions. Finally, applying such methods to nonmodel species is becoming more and more accessible and

thus may be helpful in conservation for a wide range of natural hybridization contexts.

5 | CONCLUSIONS

It has recently been recognized that consideration of the recombination rate is of prime importance to interpret how natural selection shapes the genomic landscape of introgression in admixed populations (Duranton et al., 2018; Martin et al., 2019; Martin & Jiggins, 2017; Schumer et al., 2018). In our study we highlight the importance of the temporal dynamics of hybridization and the genome-wide variation of the recombination rates, both of which results in a complex interplay of multiple evolutionary processes occurring along the genome. By assessing the pattern of introgression and recombination at three different scales (i.e., global, LG and 2-Mb windows size) we were able to provide a detailed picture of these antagonistic evolutionary mechanisms (e.g., positive and negative selection) occurring along the genome. In particular, our results show that the interplay between the recombination rate and the presence of potentially deleterious recessive mutations may be responsible for these variable selective patterns along the genome. Such variability of both selective forces and recombination rate along the genome reflect the complexity of genome evolution and need to be considered before drawing conclusions regarding the beneficial and/or negative effects of hybridization. In particular, an apparent beneficial outcome of hybridization during early generations could be detrimental in later hybrid generations when the individual effects of maladaptative loci are being exposed (Harris & Nielsen, 2016). In the same way, the whole genome could display a general beneficial outcome of hybridization while analyses at the local scale could reveal strong selection against maladaptative introgressed alleles.

ACKNOWLEDGEMENTS

We thank biologists and technicians of the Société d'Établissement de Plain Air du Québec (SEPAQ) and the Ministère des Forêts, de la Faune et des Parcs du Québec (MFFP), in particular Amélie Gilbert and Isabel Thibault for their help with the project and/or their field assistance and sampling. We thank M. Laporte for statistical advice and constructive discussion. We also thank Associate editor J. A. H. Benzie and three anonymous reviewers for their constructive comments that helped to improve the manuscript. This research was funded by the MFFP, the Canadian Research Chair in Genomics and Conservation of Aquatic Resources, Ressources Aquatiques Québec (RAQ) as well as by a Strategic Project Grant from the Natural Science and Engineering Research Council of Canada (NSERC) to L.B., D. Garant and P. Sirois.

AUTHOR CONTRIBUTIONS

M.L. and L.B. conceived the study. M.L. and H.C. designed the analyses and performed the analyses. E.N., H.C. and A-L.F. contributed to the bioinformatics and interpretation of the results. M.L. wrote the manuscript. P-A.G. helped in structuring the manuscript and all

co-authors critically revised the manuscript and approved the final version to be published.

DATA AVAILABILITY STATEMENT

Table S1 provides the individual chromosomal ancestry imbalance, number of domestic haplotypes, total percentage of domestic ancestry per individuals, number of generations since the mean years of stocking practice and hybrid generation. Table S2 describes the individual domestic ancestry tract position and length. Raw data are available from Létourneau et al. (2018) at the Dryad Digital Repository: <https://doi.org/10.5061/dryad.s5qt3>.

ORCID

Maeva Leitwein  <https://orcid.org/0000-0001-5172-0970>

Hugo Cayuela  <https://orcid.org/0000-0002-3529-0736>

Anne-Laure Ferchaud  <https://orcid.org/0000-0002-9577-5508>

Éric Normandeau  <https://orcid.org/0000-0003-2841-9391>

Pierre-Alexandre Gagnaire  <https://orcid.org/0000-0002-1908-3235>

Louis Bernatchez  <https://orcid.org/0000-0002-8085-9709>

REFERENCES

- Aitken, S. N., & Whitlock, M. C. (2013). Assisted gene flow to facilitate local adaptation to climate change. *Annual Review of Ecology, Evolution, and Systematics*, 44(1), 367–388. <https://doi.org/10.1146/annurev-ecolsys-110512-135747>
- Allendorf, F. W. (2017). Genetics and the conservation of natural populations: Allozymes to genomes. *Molecular Ecology*, 26(2), 420–430. <https://doi.org/10.1111/mec.13948>
- Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, 11(10), 697–709. <https://doi.org/10.1038/nrg2844>
- Allendorf, F. W., Leary, R. F., Spruell, P., & Wenburg, J. K. (2001). The problems with hybrids: Setting conservation guidelines. *Trends in Ecology and Evolution*, 16(11), 613–622. [https://doi.org/10.1016/S0169-5347\(01\)02290-X](https://doi.org/10.1016/S0169-5347(01)02290-X)
- Anderson, E., & Stebbins, G. L. (1954). Hybridization as an evolutionary stimulus. *Evolution*, 8(4), 378–388. <https://doi.org/10.1111/j.1558-5646.1954.tb01504.x>
- Araki, H., Berejikian, B. A., Ford, M. J., & Blouin, M. S. (2008). Fitness of hatchery-reared salmonids in the wild. *Evolutionary Applications*, 1(2), 342–355. <https://doi.org/10.1111/j.1752-4571.2008.00026.x>
- Barton, N. H. (1983). Multilocus clines. *Evolution*, 37(3), 454–471. <https://doi.org/10.1111/j.1558-5646.1983.tb05563.x>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124–3140. <https://doi.org/10.1111/mec.12354>
- Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10(3), 195–205. <https://doi.org/10.1038/nrg2526>
- Charlesworth, D., & Willis, J. H. (2009). The genetics of inbreeding depression. *Nature Reviews Genetics*, 10(11), 783–796. <https://doi.org/10.1038/nrg2664>
- Chen, Z. J. (2010). Molecular mechanisms of polyploidy and hybrid vigor. *Trends in Plant Science*, 15(2), 57–71. <https://doi.org/10.1016/j.tplants.2009.12.003>
- Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE*, 7(10), e46688. <https://doi.org/10.1371/journal.pone.0046688>
- Christensen, K. A., Rondeau, E. B., Minkley, D. R., Leong, J. S., Nugent, C. M., Danzmann, R. G., ... Koop, B. F. (2018). The Arctic charr (*Salvelinus alpinus*) genome and transcriptome assembly. *PLoS ONE*, 13(9), e0204076. <https://doi.org/10.1371/journal.pone.0204076>
- Dannemann, M., & Racimo, F. (2018). Something old, something borrowed: Admixture and adaptation in human evolution. *Current Opinion in Genetics and Development*, 53, 1–8. <https://doi.org/10.7287/peerj.preprints.26782v2>
- Dray, S., & Dufour, A.-B. (2007). The ADE4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(1), 1–20. <https://doi.org/10.18637/jss.v022.i04>
- Duranton, M., Allal, F., Fraïsse, C., Bierne, N., Bonhomme, F., & Gagnaire, P.-A. (2018). The origin and remodeling of genomic islands of differentiation in the European sea bass. *Nature Communications*, 9(1), 2518. <https://doi.org/10.1038/s41467-018-04963-6>
- Duranton, M., Bonhomme, F., & Gagnaire, P.-A. (2019). The spatial scale of dispersal revealed by admixture tracts. *Evolutionary Applications*, 12(9), 1743–1756. <https://doi.org/10.1111/eva.12829>
- Fay, J. C., & Wu, C.-I. (2000). Hitchhiking under positive Darwinian selection. *Genetics*, 155(3), 1405–1413.
- Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics*, 78(2), 737–756.
- Ferchaud, A.-L., Laporte, M., Perrier, C., & Bernatchez, L. (2018). Impact of supplementation on deleterious mutation distribution in an exploited salmonid. *Evolutionary Applications*, 11(7), 1053–1065. <https://doi.org/10.1111/eva.12660>
- Ferchaud, A.-L., Leitwein, M., Laporte, M., Boivin-Delisle, D., Bougas, B., Hernandez, C., Bernatchez, L. (2019). Adaptive and maladaptive genetic diversity in small populations: insights from the Brook Charr (*Salvelinus fontinalis*) case study. *bioRxiv*. Retrieved from <https://www.biorxiv.org/content/10.1101/660621v1.abstract>
- Frankham, R. (2015). Genetic rescue of small inbred populations: Meta-analysis reveals large and consistent benefits of gene flow. *Molecular Ecology*, 24(11), 2610–2618. <https://doi.org/10.1111/mec.13139>
- Goedbloed, D. J., van Hooft, P., Megens, H.-J., Langenbeck, K., Lutz, W., Crooijmans, R. P. M. A., ... Prins, H. H. T. (2013). Reintroductions and genetic introgression from domestic pigs have shaped the genetic population structure of Northwest European wild boar. *BMC Genetics*, 14, 43. <https://doi.org/10.1186/1471-2156-14-43>
- Gosselin, T., & Bernatchez, L. (2016). STACKR: GBS/RAD data exploration, manipulation and visualization using R. Retrieved from <https://github.com/thierrygosselin/stackr>
- Gossieux, P., Bernatchez, L., Sirois, P., & Garant, D. (2019). Impacts of stocking and its intensity on effective population size in Brook Charr (*Salvelinus fontinalis*) populations. *Conservation Genetics*, 20(4), 729–742. <https://doi.org/10.1007/s10592-019-01168-2>
- Gravel, S. (2012). Population genetics models of local ancestry. *Genetics*, 191(2), 607–619. <https://doi.org/10.1534/genetics.112.139808>
- Guan, Y. (2014). Detecting structure of haplotypes and local ancestry. *Genetics*, 196(3), 625–642. <https://doi.org/10.1534/genetics.113.160697>
- Hansen, M. M., Fraser, D. J., Meier, K., & Mensberg, K.-L.-D. (2009). Sixty years of anthropogenic pressure: A spatio-temporal genetic analysis of brown trout populations subject to stocking and population declines. *Molecular Ecology*, 18(12), 2549–2562. <https://doi.org/10.1111/j.1365-294X.2009.04198.x>

- Harris, K., & Nielsen, R. (2016). The genetic cost of Neanderthal introgression. *Genetics*, 203(2), 881–891. <https://doi.org/10.1534/genetics.116.186890>
- Harris, K., Zhang, Y., & Nielsen, R. (2019). Genetic rescue and the maintenance of native ancestry. *Conservation Genetics*, 20(1), 59–64. <https://doi.org/10.1007/s10592-018-1132-1>
- Henry, R. (2015). Etymologia: Bonferroni correction. *Emerging Infectious Diseases*, 21(2), 289. <https://doi.org/10.3201/eid2102.ET2102>
- Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B. M., Vinckenbosch, N., ... Nielsen, R. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, 512(7513), 194–197. <https://doi.org/10.1038/nature13408>
- Johnson, B. M., Arlinghaus, R., & Martinez, P. J. (2009). Are we doing all we can to stem the tide of illegal fish stocking? *Fisheries*, 34(8), 389–394. <https://doi.org/10.1577/1548-8446-34.8.389>
- Kim, B. Y., Huber, C. D., & Lohmueller, K. E. (2018). Deleterious variation shapes the genomic landscape of introgression. *PLoS Genetics*, 14(10), e1007741. <https://doi.org/10.1371/journal.pgen.1007741>
- Lamaze, F. C., Sauvage, C., Marie, A., Garant, D., & Bernatchez, L. (2012). Dynamics of introgressive hybridization assessed by SNP population genomics of coding genes in stocked brook charr (*Salvelinus fontinalis*). *Molecular Ecology*, 21(12), 2877–2895.
- Leitwein, M., Gagnaire, P.-A., Desmarais, E., Berrebi, P., & Guinand, B. (2018). Genomic consequences of a recent three-way admixture in supplemented wild brown trout populations revealed by local ancestry tracts. *Molecular Ecology*, 27(17), 3466–3483. <https://doi.org/10.1111/mec.14816>
- Létourneau, J., Ferchaud, A.-L., Luyer, J. L., Laporte, M., Garant, D., & Bernatchez, L. (2018). Predicting the genetic impact of stocking in Brook Charr (*Salvelinus fontinalis*) by combining RAD sequencing and modeling of explanatory variables. *Evolutionary Applications*, 11(5), 577–592. <https://doi.org/10.1111/eva.12566>
- Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26(5), 589–595. <https://doi.org/10.1093/bioinformatics/btp698>
- Lippman, Z. B., & Zamir, D. (2007). Heterosis: Revisiting the magic. *Trends in Genetics*, 23(2), 60–66. <https://doi.org/10.1016/j.tig.2006.12.006>
- Marie, A. D., Bernatchez, L., & Garant, D. (2010). Loss of genetic integrity correlates with stocking intensity in Brook Charr (*Salvelinus fontinalis*). *Molecular Ecology*, 19, 2025–2037. <https://doi.org/10.1111/mec.2010.19.issue-10>
- Martin, S. H., Davey, J., Salazar, C., & Jiggins, C. D. (2019). Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biology*, 17(2), e2006288. <https://doi.org/10.1371/journal.pbio.2006288>
- Martin, S., & Jiggins, C. D. (2017). Interpreting the genomic landscape of introgression. *Current Opinion in Genetics and Development*, 47, 69–74. <https://doi.org/10.1016/j.gde.2017.08.007>
- Martin, S., Savaria, J.-Y., Audet, C., & Bernatchez, L. (1997). Microsatellites reveal no evidence for inbreeding effects but low inter-stock genetic diversity among brook charr stocks used for production in Québec. *Aquaculture Association of Canada*, 97(2), 21–23.
- McFarlane, E., & Pemberton, J. M. (2019). Detecting the true extent of introgression during anthropogenic hybridization. *Trends in Ecology and Evolution*, 34(4), 315–326.
- Ministère du Développement Durable, de l'Environnement, de la Faune et des Parcs. (2013). *Outil D'aide À L'ensemencement Des Plans D'eau - Omble de Fontaine* (*Salvelinus fontinalis*) (pp. 1–12). Direction de la faune aquatique, Québec: Direction générale de l'expertise sur la faune et ses habitats.
- Naish, K. A., Taylor, J. E., Levin, P. S., Quinn, T. P., Winton, J. R., Huppert, D., & Hilborn, R. (2007). An evaluation of the effects of conservation and fishery enhancement hatcheries on wild populations of salmon. *Advances in Marine Biology*, 53, 61–194. [https://doi.org/10.1016/S0065-2881\(07\)53002-6](https://doi.org/10.1016/S0065-2881(07)53002-6)
- Nugent, C. M., Easton, A. A., Norman, J. D., Ferguson, M. M., & Danzmann, R. G. (2017). A SNP based linkage map of the arctic charr (*salvelinus alpinus*) genome provides insights into the diploidization process after whole genome duplication. *G3: Genes, Genomes, Genetics*, 7(2), 543–556. <https://doi.org/10.1534/g3.116.038026>
- Orr, H. A. (1995). The population genetics of speciation: The evolution of hybrid incompatibilities. *Genetics*, 139(4), 1805–1813.
- R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Racimo, F., Sankararaman, S., Nielsen, R., & Huerta-Sánchez, E. (2015). Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics*, 16(6), 359–371. <https://doi.org/10.1038/nrg3936>
- Randi, E. (2008). Detecting hybridization between wild species and their domesticated relatives. *Molecular Ecology*, 17(1), 285–293. <https://doi.org/10.1111/j.1365-294X.2007.03417.x>
- Renaut, S., & Rieseberg, L. H. (2015). The accumulation of deleterious mutations as a consequence of domestication and improvement in sunflowers and other Compositae crops. *Molecular Biology and Evolution*, 32(9), 2273–2283. <https://doi.org/10.1093/molbev/msv106>
- Rezvoy, C., Charif, D., Guéguen, L., & Marais, G. A. (2007). MAREYMAP: An R-based tool with graphical interface for estimating recombination rates. *Bioinformatics*, 23(16), 2188–2189. <https://doi.org/10.1093/bioinformatics/btm315>
- Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., ... Przeworski, M. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, 360(6389), 656–660. <https://doi.org/10.1126/science.aar3684>
- Sutherland, B. J. G., Gosselin, T., Normandeau, E., Lamothe, M., Isabel, N., Audet, C., & Bernatchez, L. (2016). Salmonid chromosome evolution as revealed by a novel method for comparing RADseq linkage maps. *Genome Biology and Evolution*, 8(12), 3600–3617. <https://doi.org/10.1093/gbe/evw262>
- Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., ... Rieseberg, L. H. (2016). Hybridization and extinction. *Evolutionary Applications*, 9(7), 892–908. <https://doi.org/10.1111/eva.12367>
- Turelli, M., & Orr, H. A. (2000). Dominance, epistasis and the genetics of postzygotic isolation. *Genetics*, 154(4), 1663–1679.
- Uller, T., & Leimu, R. (2011). Founder events predict changes in genetic diversity during human-mediated range expansions. *Global Change Biology*, 17(11), 3478–3485. <https://doi.org/10.1111/j.1365-2486.2011.02509.x>
- Verhoeven, K. J. F., Macel, M., Wolfe, L. M., & Biere, A. (2011). Population admixture, biological invasions and the balance between local adaptation and inbreeding depression. *Proceedings of the Royal Society B: Biological Sciences*, 278(1702), 2–8. <https://doi.org/10.1098/rspb.2010.1272>
- Waples, R. S. (1991). Genetic interactions between hatchery and wild salmonids: Lessons from the Pacific Northwest. *Canadian Journal of Fisheries and Aquatic Sciences*, 48(S1), 124–133. <https://doi.org/10.1139/f91-311>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Leitwein M, Cayuela H, Ferchaud A-L, Normandeau É, Gagnaire P-A, Bernatchez L. The role of recombination on genome-wide patterns of local ancestry exemplified by supplemented brook charr populations. *Mol Ecol*. 2019;00:1–15. <https://doi.org/10.1111/mec.15256>