Received: 6 February 2019 | Revised: 27 March 2019 | Accepted: 22 April 2019

DOI: 10.1111/jeb.13482

RESEARCH PAPER

WILEY Journal of Evolutionary Biology eseb

Model-based demographic inference of introgression history in European whitefish species pairs'

Clément Rougeux[1] | Pierre-Alexandre Gagnaire[2] | Louis Bernatchez[1]

[1]Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Quebec, Québec, Canada

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

Correspondence
Clément Rougeux, Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, G1V 0A6, Québec, Canada
Email: clement.rougeux@gmail.com

Abstract

Parallel phenotypic differentiation is generally attributed to parallel adaptive divergence as an evolutionary response to similar environmental contrasts. Such parallelism may actually originate from several evolutionary scenarios ranging from repeated parallel divergence caused by divergent selection to a unique divergence event followed by gene flow. Reconstructing the evolutionary history underlying parallel phenotypic differentiation is thus fundamental to understand the relative contribution of demography and selection on genomic divergence during speciation. In this study, we investigate the divergence history of replicate European whitefish (*Coregonus lavaretus*), limnetic and benthic species pairs from two lakes in Norway and two lakes in Switzerland. Demographic models accounting for semi-permeability and linked selection were fitted to the unfolded joint allele frequency spectrum built from genome-wide SNPs and compared to each other in each species pair. We found strong support for a model of asymmetrical post-glacial secondary contact between glacial lineages in all four lakes. Moreover, our results suggest that heterogeneous genomic differentiation has been shaped by the joint action of linked selection accelerating lineage sorting during allopatry, and heterogeneous migration eroding divergence at different rates along the genome following secondary contact. Our analyses reveal how the interplay between demography, selection and historical contingency has influenced the levels of diversity observed in previous whitefish phylogeographic studies. This study thus provides new insights into the historical demographic and selective processes that shaped the divergence associated with ecological speciation in European whitefish.

KEYWORDS

Coregonus, demographic inference, ecological speciation, JAFS, population genomics, secondary contact

1 | INTRODUCTION

Independent phenotypic divergence among closely related and locally adapted nascent species provides evidence for population phenotypic response to similar constraints stemming from environmental variation (Endler, 1986; Losos, 2011). Therefore, parallel phenotypic divergence among replicate species pairs provides a valuable framework to understand the molecular basis of species diversification by natural selection (Elmer & Meyer, 2011). Repeated phenotypic diversification can arise through similar or different genomic bases as a function of the

origin and effects of selected mutations. Different scenarios regarding the origin of adaptive variation are usually distinguished, depending on whether selection has been recruiting de novo mutations (Pearce et al., 2009), mutations already present in the ancestral genetic variation (Yeaman et al., 2016), or alternatively adaptive alleles transferred by gene flow between populations from similar environments (Schluter & Conte, 2009; Welch & Jiggins, 2014). These distinctions are important for understanding whether repeated phenotypic divergence is facilitated by the existence of divergent alleles that have evolved in isolation in the past, while being later reused by selection over shorter timescales (Van Belleghem et al., 2018; Welch & Jiggins, 2014).

Along with the question pertaining to the genomic basis and the origin of the mutations associated with parallel phenotypic differentiation, the chronology and the mode of divergence remain highly discussed (Welch & Jiggins, 2014). Indeed, parallelism does not necessarily imply independent evolution, since it could alternatively result from a unique divergence event that is shared among replicates (Bierne, Gagnaire, & David, 2013). Consequently, it is important to decipher whether phenotypic diversification has evolved in a context of primary or secondary divergence, which are two different evolutionary scenarios with respect to the timing of gene flow (Smadja & Butlin, 2011). Primary divergence is usually associated with truly parallel evolution among replicates, in response to similar selective pressures. By contrast, secondary divergence may involve a single divergence event between two lineages followed by their spatial redistribution and admixture. Another important aspect in both secondary and primary divergence scenarios is the joint influence of demographic and selective forces on the heterogeneity of divergence across the genome (Barton & Bengtsson, 1986; Harrison & Larson, 2016; Sousa & Hey, 2013). Therefore, understanding the different facets of the evolutionary history underlying phenotypic divergence can be quite challenging.

Recent developments in demographic inference methods provide a powerful framework to decipher the relative contributions of demographic and selective pressures accompanying parallel phenotypic differentiation (Sousa & Hey, 2013). In particular, methods that integrate varying introgression rates among loci allow capturing the barrier effect caused by speciation genes (Rougeux, Bernatchez, & Gagnaire, 2017; Roux et al., 2016; Sousa, Carneiro, Ferrand, & Hey, 2013; Tine et al., 2014), which in turn provides an estimation of the species barrier's permeability to gene flow. In addition, the effect of linked selection can be captured by allowing different sets of loci to experience varying rates of genetic drift (Rougeux et al., 2017; Roux et al., 2016; Sousa et al., 2013). This extended framework thus provides the opportunity to integrate different demographic and selective effects separately or simultaneously, within models of increasing complexity that help us improving the reconstruction of the history of speciation, while accounting for heterogeneous patterns of differentiation in genomic data sets (Rougeux et al., 2017).

The European whitefish (*Coregonus lavaretus*), hereafter whitefish, represents a valuable model to study the evolutionary history of repeated parallel phenotypic differentiation. Indeed, the *C. lavaretus* species complex is composed of several whitefish species believed to have originated following the last glaciation events and distributed across the European continent, notably in Alpine lakes from Switzerland (Hudson, Vonlanthen, & Seehausen, 2010) and in Fennoscandinavia lakes (Douglas, Brunner, & Bernatchez, 1999; Østbye, Bernatchez, Naesje, Himberg, & Hindar, 2005). Ecological divergence in this system has resulted in phenotypic differentiation associated with the use of limnetic and benthic ecological niches (Praebel et al., 2013; Vonlanthen et al., 2009). Phenotypic divergence is believed to have evolved repeatedly across different lakes (Douglas et al., 1999; Hudson et al., 2010). Whereas the sympatric species pairs of the Alpine whitefish radiation have been taxonomically distinguished (Kottelat & Freyhof, 2007), this is not the case for Norwegian whitefish. Nevertheless, we here use the terms limnetic and benthic species to refer to these similar morphotypes found in different European locations. The limnetic and benthic whitefish species differ in body size, body shape, habitat utilization and resources with a feeding-related morphology (Praebel et al., 2013; Vonlanthen et al., 2009). Differences between species in the use of habitat and ecological niches likely induce reproductive isolation as a by-product of ecological divergence (Praebel et al., 2013; Schluter, 2000; Woods, Müller, & Seehausen, 2009). However, it remains unclear whether primary or secondary divergence has accompanied whitefish diversification in Eurasia.

Mitochondrial DNA (mtDNA)-based phylogeographic studies have identified three clades in the whitefish distribution area: the North European, the Siberian and the South European clades (Østbye et al., 2005). Those three clades reflect the occupation of three different glacial refugia during the last glacial period (Østbye et al., 2005). However, haplotypes of all three clades are observed together in some areas, suggesting that lineage mixing may have resulted from habitat recolonization following glacial retreat (Østbye et al., 2005). Moreover, genetic analysis comparing limnetic and benthic species revealed heterogeneous genetic differentiation along the genome between sympatric species in Alpine lakes (Feulner & Seehausen, 2018), suggesting local reduction in gene flow for some genomic regions associated with adaptive divergence. Whereas such heterogeneous genome divergence has not been characterized in studies on Norwegian whitefish, intraspecies populations from different lakes showed more genetic similarities between them than between sympatric species, suggesting a common origin for the populations of a given species (Praebel et al., 2013). Although such results support the hypothesis that limnetic and benthic whitefish species share a common history of divergence across Europe, this remains to be rigorously tested using historical divergence inference methods in order to understand the evolutionary processes associated with their diversification.

Here, we use RNAseq single nucleotide polymorphism (SNP) (De Wit, Pespeni, & Palumbi, 2015) to document the joint allele frequency spectrum (JAFS) of each sympatric species pair and infer its demographic divergence history. We test for temporal and genomic variations in the rate of gene flow for each sympatric species pairs. We then characterize genetic relationships and perform historical gene flow analyses among lakes to determine a general parsimonious scenario for whitefish diversification. Finally, we document the relative contribution of selective and demographic parameters in shaping the genomic differentiation between sympatric species pairs.

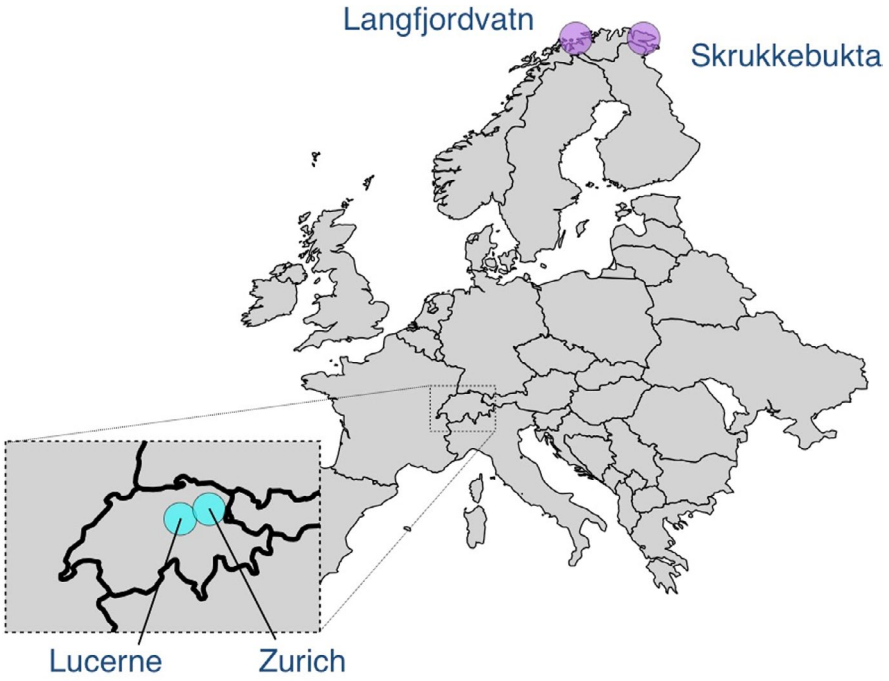

FIGURE 1 Geographic locations of the lakes where sympatric whitefish species pairs were sampled. Two lakes were sampled in Norway (purple): Langfjordvatn and Skrukkebukta lakes; and two lakes were sampled in Switzerland (blue, zoomed in the box): Zurich and Lucerne lakes

2 | MATERIAL AND METHODS

2.1 | Sampling, library preparation and sequencing

Four lakes harbouring sympatric limnetic and benthic species were sampled across the distribution range of *C. lavaretus* in Europe. Six individual per species (i.e. 12 individuals per lake) were collected in two Scandinavian lakes in Norway (Skrukkebukta and Langfjordvatn lakes) and two lakes from Switzerland (Lucerne and Zurich lakes) (Figure 1), for a total of 48 individuals. Liver tissue was taken on each individual and stored in RNAlater before mRNA extractions.

We used the same sequencing data as generated in a previous study (Rougeux, Gagnaire, Praebel, Seehausen, & Bernatchez, 2018). However, the analyses performed in that previous study addressed a totally different topic (comparing patterns of differential transcription vs. genomic differentiation) and did not include any model-based

TABLE 1 The demographic divergence history of whitefish species pairs inferred parameters

Lake	Model	MLE	AIC	Theta	w_{AIC}	N_{ref}	N_1	N_2	*hrf*	m_{e12}
Lf	SC2m	−1003.5	2027.1	2756.0	1.0	3608.5	708835.4	182275.6	—	0.00051
							[76535.9;1341021.3]	[0;375859.8]		[0.0002;0.0007]
Sk	SC2N2m	−803.2	1630.4	3990.2	1.0	6028.9	685784.1	244350.1	0.13	0.00102
							[481947.9;887088.1]	[50642.5;438479.7]	[0.09;0.17]	[0.0005;0.0015]
Lu	SC2N2m	−887.0	1797.9	3440.0	1.0	6057.4	295539	297235	0.62	0.00079
							[141621.2;449456.6]	[61482.2;532563.7]	[0.00;1.29]	[0.0003;0.0012]
Zu	SC2m	−1133.2	2286.3	3988.0	1.0	8874.7	174565.1	79517.2	—	0.00005
							[0;765974.1]	[11803.3;147231.0]		[0;0.0013]

Note: For all lakes, details of statistics and demographic parameter values for the fittest models determined under the threshold of ΔAIC < 10. The table contains in this order the maximum likelihood (MLE) for each model, the value of Akaike information criterion (AIC) and the weighted AIC (w_{AIC}). Then, the inferred demographic parameter values converted with the estimate of theta: the ancestral effective population size before population split (N_{ref}); the effective population size after split for limnetic (N_1) and benthic (N_2) populations; the Hill–Robertson factor (*hrf*) corresponds to the degree to which the effective population size of diverging populations is locally reduced linked selection effect. Migration parameters (in migrant per generation) include migration rates from benthic population to limnetic population (m_{e12}) and reciprocally (m_{e21}), and a second category of effective migration rates ($m_{e'12}$) and ($m_{e'21}$) applying to a second category of loci. Time parameters include the duration (in years) of the allopatric divergence period (T_s) and the duration of the migration period (T_{SC}). Finally, the table also contains proportion parameters such as the proportion (Q) of the genome with effective population sizes N_1 and N_2. A proportion (P) of the genome is occupied by loci with effective migration rates m_{e12} and m_{e21} (and a second category of loci occupying a fraction 1−P of the genome has effective population sizes m'_{e12} and m'_{e21}). The parameter (O) is the proportion of correct SNP orientation. Numbers into brackets denote 95% confidence intervals obtained using the MLE parameter values ± 2 *SE*.

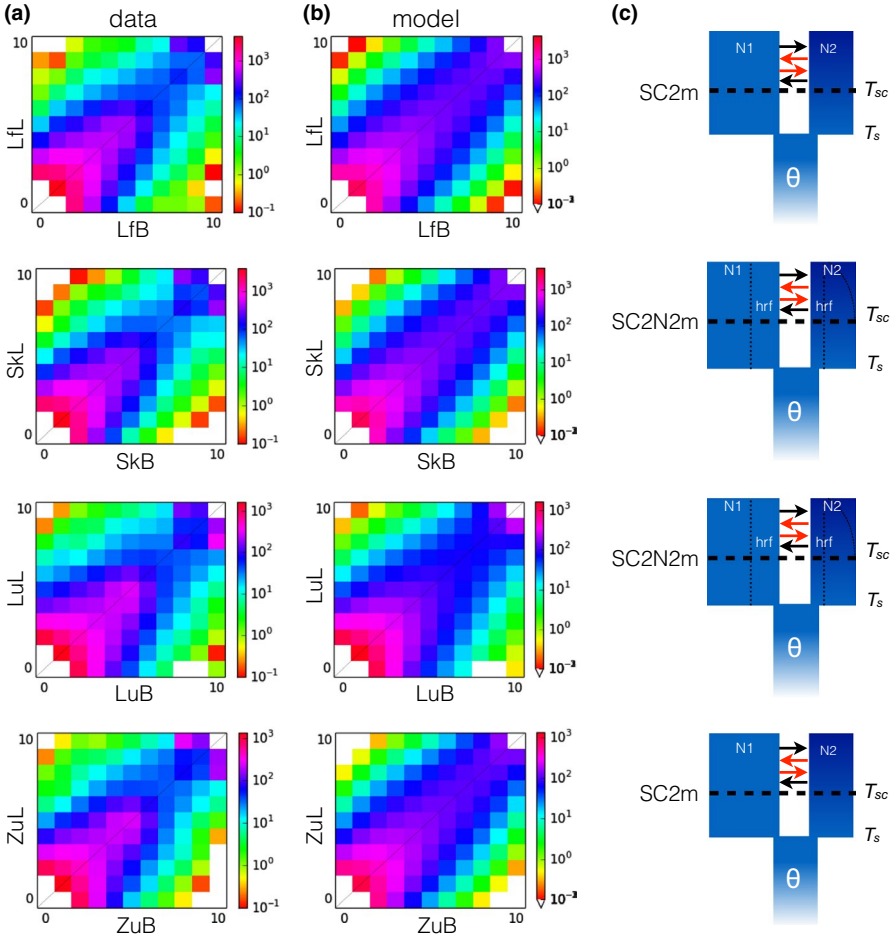

FIGURE 2 Historical demography of whitefish species pairs. (a) Observed joint allele frequency spectrum (JAFS) for benthic (-B; *x*-axis) and limnetic (-L; *y*-axis) species for each of the four lakes (Lf, Langfjordvatn Lake; Sk, Skrukkebukta Lake; Lu, Lucerne Lake; and Zu, Zurich Lake), obtained by projection of empirical data to 10 chromosomes. (b) Predicted JAFS of the fittest model for each lake. For each JAFS, the colour scale corresponds to the number of SNPs for each bin, defined by the counts of the number of derived alleles observed in limnetic and benthic populations. (c) Representation of the fittest model for each lake. All fittest models are based on a secondary contact (SC) model and integrate heterogeneous parameters, such as a heterogeneous rate of gene flow (−2 m) and linked selection (−2N). Each blue block corresponds to a population (with effective population size of N_1, N_2 and θ), and dashed lines indicate the duration (T_{sc}) of the secondary contact period after the duration (T_s) of the allopatric period: black arrows symbolize neutral gene flow, whereas red arrows symbolize reduced gene flow around locally adapted loci

demographic inference. Briefly, total RNA was extracted using the RNeasy Mini Kit following the manufacturer's instructions (Qiagen). We first quantified the amount of RNA via a NanoDrop 2000 spectrophotometer (Thermo Scientific) and assessed RNA quality using the 2100 Bioanalyzer (Agilent). We finally measured RNA concentration with Quant-iT RiboGreen RNA Assay Kit (Invitrogen, Life

m_{e21}	$m_{e'12}$	$m_{e'21}$	T_s	T_{sc}	P	Q	O
0.00072	0.00004	0.00034	121497.7	12377.1	0.95	–	0.98
[0.0007;0.0008]	[0;0.0001]	[0.0002;0.0003]	[53802.5;189192.8]	[4041.5;10608.9]	[0.36;1.53]		[0.98;0.99]
0.01149	0.00010	0.00045	68789.4	15614.8	0.64	0.07	0.98
[0.008;0.0149]	[0;0.0001]	[0;0.001]	[60348.9;77651.8]	[0;50642.5]	[0.21;1.07]	[0.00;0.19]	[0.96;1.00]
0.00069	0.00058	0.00001	154833.5	24168.9	0.50	0.02	0.98
[0;0.0013]	[0;0.001]	[0;0.00008]	[31801.1;277865.7]	[3392.1;44945.6]	[0.35;0.65]	[0.00;0.07]	[0.96;1.00]
0.02647	0.00000	0.00014	106851.2	28576.5	0.91	–	0.99
[0.0093;0.0436]	[0;0.0142]	[0;0.0004]	[59016.6;154685.7]	[1242.4;54668.0]	[0.33;1.48]		[0.99;0.99]

Technologies) on samples reaching our quality criteria (RIN value greater than or equal to eight) before library preparation.

Individual libraries were prepared from 2 μg of RNA using the TruSeq RNA sample preparation kit V2 (Illumina) following the manufacturer's instructions. Library size and concentration were evaluated using DNA High Sensitivity Chip on the 2100 Bioanalyzer (Agilent). Single read sequencing (100 bp) was performed on the Illumina HiSeq 2000 platform at the McGill University and Genome Quebec Innovation Centre (Montreal, Canada).

2.2 | Genotyping

In order to document the extent of polymorphism within and among limnetic and benthic species pairs in *C. lavaretus*, individuals reads of the 48 *C. lavaretus* individuals were mapped to a reference transcriptome composed by orthologous genes to *C. clupeaformis* and *C. lavaretus* (Rougeux et al., 2018) using *Bowtie2* v2.1.0 (Li & Durbin, 2010). Briefly, the composite assembled reference transcriptome was composed by transcripts with a N50 of 1,797 bp, and identified paralogous genes were discarded. Six benthic individuals and six limnetic of the American Lake Whitefish (*C. clupeaformis*) that were taken from (Rougeux et al., 2018) were also aligned to the same reference transcriptome to provide outgroup information for orienting mutations and identify the derived alleles within *C. lavaretus*.

BAM files were generated and sorted using Samtools v1.3 (Li et al., 2009), and filtered for duplicate reads that were removed with the Picard-tools program v1.119 (http://broadinstitute.github.io/picard/). SNP calling from mapped reads was realized with *Freebayes* v0.9.10-3-g47a713eb (Garrison & Marth, 2012), considering alleles with at least two reads per sample (default), and keeping good-confidence alignments with a minimum quality of five and a minimum coverage of five.

Variable sites with a minimum coverage lower than three reads per individuals were filtered out. We then used *vcffilter* program from *vcflib* (Garrison & Marth, 2012) to process the VCF file generated by *Freebayes*. We retained biallelic SNPs with a Phred-scaled quality score above 30 and individual genotypes with a Phred score higher than 20. Following quality control steps, we removed miscalled and low-quality SNPs for subsequent population genomics analyses using *VCFtools* (Danecek et al., 2011). SNPs with more than 10% of missing genotypes in at least one *C. lavaretus* population were removed. A lower exclusion threshold of 50% was applied for the *C. clupeaformis* outgroup population, in order to maintain an important number of SNPs for the orientation step (described below). We controlled for Hardy–Weinberg disequilibrium within each population using a *p*-value exclusion threshold of 0.05 and kept only one SNP per nonoverlapping window of 100 bp along assembled transcripts in order to avoid tightly linked SNPs. Finally, we merged the filtered data sets of limnetic and benthic populations within each lake together with the Lake Whitefish outgroup, resulting in four lake-outgroup VCF files containing 412,292, 408,999, 410,433 and 408,071 SNPs for Langfjordvatn, Skrukkebukta, Lucerne and Zurich lakes, respectively. Finally, only oriented loci (i.e. monomorphic loci in *C. clupeaformis* and polymorphic in *C. lavaretus*) were kept after removing loci that were polymorphic in the outgroup (Rougeux et al., 2017; Tine et al., 2014). This resulted in 157,516, 136,499, 117,125 and 92,678 oriented SNPs (corresponding to a mean number of 8.9 SNPs/transcript for each lake, SD 0.49) for Langfjordvatn, Skrukkebukta, Lucerne and Zurich lakes, respectively, that were used to generate the unfolded JAFS of each lake, by projecting the data to 10 chromosomes (i.e. 5 individual diploid copies) per population (Figure 2).

2.3 | Divergence history inferences

The demo-selective histories of the four species pairs were inferred using a custom version of the software *∂a∂i* v1.7 (Gutenkunst, Hernandez, Williamson, & Bustamante, 2009). Based on the four basic models representing alternative modes of divergence, Strict Isolation (SI), Isolation with Migration (IM), Ancient Migration (AM) and Secondary Contact (SC), we considered 13 model extensions that integrate additional demographic and selective parameters. In these models, an ancestral population of size N_{ref} splits into two populations of size N_1 (limnetic) and N_2 (benthic) during T_S (SI and IM), $T_{AM} + T_S$ (AM) or $T_S + T_{SC}$ (SC) generations. Migration events during periods of T_S (IM), T_{AM} (AM) and T_{SC} (SC) generations occurred at rates m_{e12} from population 2 (benthic) to population 1 (limnetic), and vice versa. Next, we incorporated extensions to the models with gene flow in order to capture the effect of selection inducing reduced gene flow around loci associated with adaptive divergence (heterogeneous gene flow: *−2m*). These semi-permeability models allowed quantifying the proportion *P* of loci freely exchanged at rate m_e, whereas a proportion 1−*P* corresponded to the fraction of the genome experiencing a reduced effective migration rate m'_{e12} from population 2 (benthic) to population 1 (limnetic), and vice versa (Tine et al., 2014). We further incorporated the effect of background selection by considering local genomic variation in N_e across the genome due to Hill–Robertson effects at linked neutral sites (Charlesworth, Morgan, & Charlesworth, 1993; Hill & Robertson, 1966). We used a scaling factor *hrf* < 1, multiplying the N_e of each population for a proportion 1−*Q* of loci that are affected by linked selection (Rougeux et al., 2017). In each model, an orientation parameter *O* was used to quantify the proportion of correct SNP ancestral orientation (Tine et al., 2014).

We generated 20 independent optimization runs for each model in each lake in order to check for model convergence and kept the best-fit run of each model in order to perform model comparisons for each lake on the basis of the Akaike information criterion (AIC). This allowed us to account for the number of parameters in each model. We retained models with the smallest AIC ($\Delta AIC_i < 10$) and estimated the weighted AIC (w_{AIC}) of each model to estimate its relative probability to be the best of all tested models (Rougeux et al., 2017). We then estimated parameter uncertainty for each retained model using the Godambe information matrix method from *∂a∂i* v1.7. A nonparametric procedure was used to estimate confidence intervals (CIs) and standard errors (SE) from 1,000 bootstrap resamplings. Finally, we converted estimated parameter values using *theta* (*θ*) (Table 1), by applying equations from Rougeux et al. (2017).

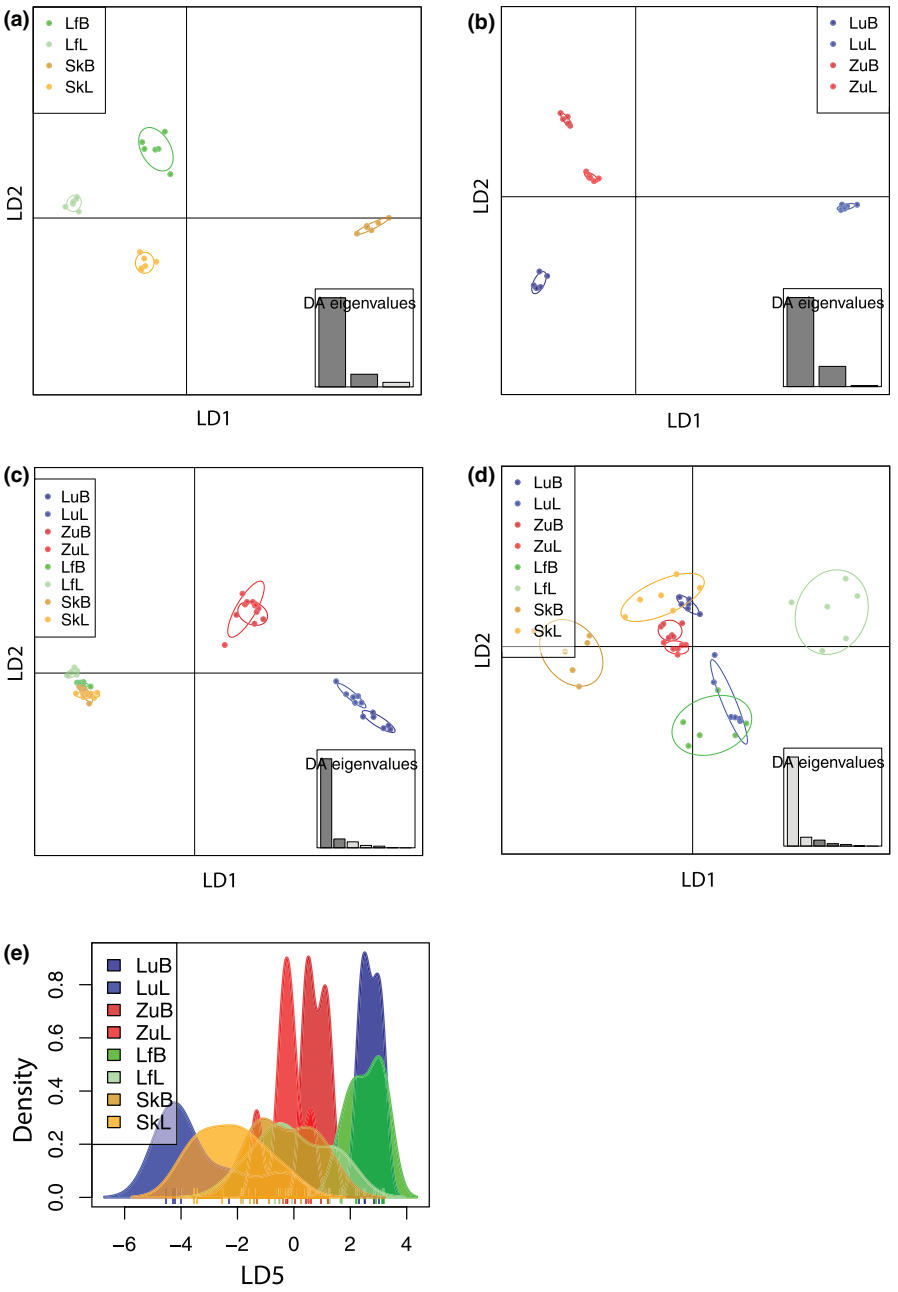

FIGURE 3 Genetic structure and relationships among lakes and species. (a) Discriminant analysis of principal components (dAPC) of the lakes in Norway and (b) for the lakes in Switzerland. For both regions, the first axis captures the signal of species differentiation and the second axis separates species pairs into their respective lakes. (c) dAPC of the different lakes of the study. The first axis separates lakes according to their region. The second axis discriminates mainly lakes from Switzerland. (d) dAPC including all lakes, with most of the variation discriminating populations from Norway, capturing the signal of lakes on the axis three and the signal of species differentiation on the axis four. (e) The fifth axis captures the residual genetic proximity of populations of the same species, allowing separating limnetic and benthic sympatric populations

2.4 | Patterns of shared ancestry and admixture

We generated a merged VCF file composed by only polymorphic loci that were retained after filtering in the four studied lakes to search for signals of shared ancestry and gene flow among all four species pairs. Using the 47,556 retained SNPs, we performed a discriminant analysis of principal components (dAPC), in *Adegenet* v2.0.0 (Jombart, Devillard, & Balloux, 2010), partitioning the variance into between-group variance and within-group variance in order to maximize discrimination between groups. The dAPC was used to characterize the genetic structure and relationships among lakes and species in the entire system (Figure 3a), and in Norway and Switzerland separately (Figure 3b,c).

Finally, we inferred historical relationships among populations using TreeMix v1.12 and *f*4-statistics (Pickrell & Pritchard, 2012). *f*4-Statistics assess the relationship among a group of four populations, based on an unrooted phylogenetic tree, and test consistency with SNP allele frequency data (Reich, Thangaraj, Patterson, Price, & Singh, 2009). Then, evaluating *f*4-statistics between populations A, B, C and D under the topology *f*4(A;B;C;D) corresponds to measure the frequency differences between A and B, C and D and test for the correlation of the differences of both comparisons. Significant deviation from 0 is associated with the effect of mixture between populations of the tree. More precisely, significantly positive value of *f*4-statistics means that A and C or B and D are more related than expected under null hypothesis (i.e. genetic drift effect

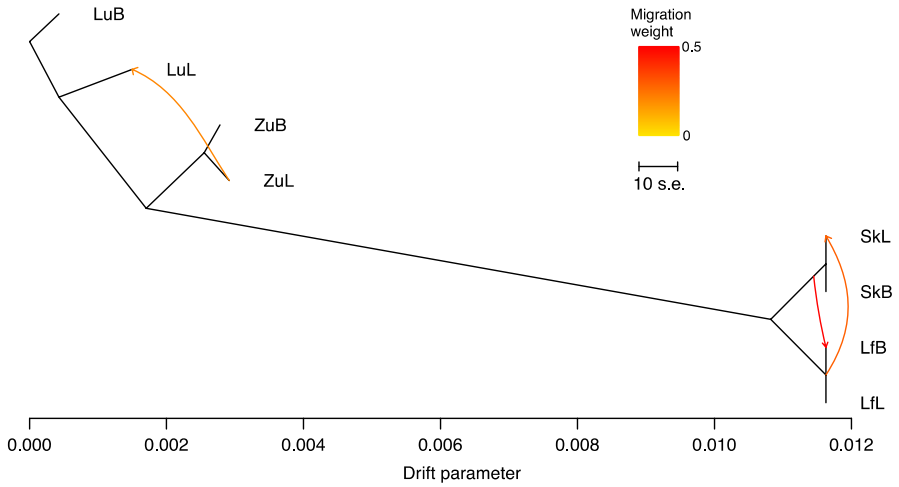

FIGURE 4 Shared ancestral genetic variation between allopatric and admixture events between sympatric species pairs. The TreeMix tree represented here allowed three migration events. Migration events link limnetic populations from different lakes in Switzerland, and ancestral populations from each lake in Norway with contemporary populations. Horizontal branch lengths are proportional to the amount of genetic drift in each branch, and the scale bar indicates 10 times the average standard error (*SE*) of the entries in the covariance matrix between pairs of populations. Colour scale indicates the weight of inferred migration events

alone). Alternatively, a significantly negative value of *f*4-statistics means that A and D or B and C are more related to each other than expected (Herman et al., 2018). TreeMix allows modelling migration events between populations by discrete mixture events that are added to a bifurcating population tree. Consequently, such events may reflect either gene flow between sympatric populations or excess of shared ancestral polymorphism among populations from different lakes that are geographically isolated. We allowed three migration events to be inferred among the branches of the whitefish population tree, in agreement with the resulting increase in likelihood (Figure S1).

3 | RESULTS

A total of 1.19×10^9 100-bp single raw reads were generated from the 48 individuals of *C. lavaretus* (1.5×10^9 when including the 12 *C. clupeaformis* individuals) (Table S1). A total of 1.44×10^9 (including both *C. lavaretus* and *C. clupeaformis*) trimmed and cleaned reads were aligned to the reference transcriptome generated in Rougeux et al. (2018), resulting in the mapping of 8.3×10^8 reads.

3.1 | Divergence models

The JAFS of each of the four species pairs showed a high density of SNPs with similar allele frequencies between limnetic and benthic populations (i.e. located around the diagonal) (Figure 2). However, the presence of markers located on the outer frame of the spectra showed that some regions of the genome display strong differences in allele frequencies between sympatric species in each lake.

In general, the model producing the best fit to the observed JAFS was an extension of the secondary contact (SC) model, in which gene flow occurred following an allopatric phase. The incorporation of selective parameters in the demographic models generally improved AIC values that penalize model likelihood by the number of parameters to avoid overfitting. For all the four lakes, heterogeneous gene flow along the genome was a necessary component of the secondary contact model (SC2 m), whereas for Skrukkebukta and Lucerne lakes, integrating local reductions in N_e to account for linked selection improved their respective best-fit model (SC2N2 m; Table 1 and Table S2). All the highest ranked model for each lake received a $w_{AIC} = 1.0$, thus showing maximal support for the SC2 m model in Langfjordvatn and Zurich lakes and for the SC2N2 m model in Skrukkebukta and Lucerne lakes.

3.2 | Inference of model parameters

The inferred proportion of correctly oriented markers in the unfolded JAFS (*O*) ranged from 98% to 99%, indicating that almost all of ancestral allelic states were correctly inferred using *C. clupeaformis* as an outgroup. The conversion of estimated demographic and selective parameters allowed intra- and interlake population comparisons. Considering only the best-fit model for each lake, the effective population sizes of the limnetic populations in the Norwegian lakes (*nu1*) were about three times greater than the effective population sizes of the benthic populations (*nu2*). Although wide confidence intervals overlapped between limnetic and benthic populations, effective population sizes were similar in Lucerne Lake and a larger effective population size was inferred for the limnetic population in Zurich Lake (Table 1). The migration rate followed the same pattern with a generally higher rate from the limnetic species to the benthic species (m_{e21}), but not in Lucerne Lake where migration rates were comparable for both directions. Finally, the fraction of the genome showing unconstrained migration (*P*) was estimated to be 95%, 64%, 50% and 91% for Langfjordvatn, Skrukkebukta, Lucerne and Zurich, respectively. These values suggest that limnetic and benthic

whitefish from Lucerne and Skrukkebukta lakes are more strongly reproductively isolated than those from Zurich and Langfjordvatn lakes. Consistent with these observations, the highest ranked model for Lucerne and Skrukkebukta lakes also included heterogeneous effective population size across the genome. The fraction (*Q*) of the genome with a locally reduced N_e was estimated to be 7% for Skrukkebukta Lake and 2% for Lucerne Lake, and the factor of reduction (*hrf*) in N_e in those fractions was 13% and 62% in Skrukkebukta and Lucerne lakes, respectively, although this was associated with a large uncertainty for Skrukkebukta (Table 1).

Finally, the estimated duration of the allopatric phase (T_s) was variable among lakes, although in the same order of magnitude, with T_s of 121,000, 69,000, 155,000 and 107,000 years for Langfjordvatn, Skrukkebukta, Lucerne and Zurich, respectively. Moreover, the inferred duration of secondary contact (T_{sc}) corresponds roughly with the last glacial retreat and corroborates the more recent deglaciation in northern Europe relative to the central Alpine region, with secondary gene flow being initiated 12,400, 15,600, 24,200 and 28,600 years ago for Langfjordvatn, Skrukkebukta, Lucerne and Zurich, respectively.

3.3 | A shared history of divergence

Partitioning the genetic variation using a dAPC, either within Norway and Switzerland separately or in the entire system, revealed distinct patterns. In Norway, populations clustered by species along the first axis (LD1 6.4% of the variance), whereas the second axis (LD2 5.3% of the variance) separated the populations according to their lake of origin (Figure 3a). A similar pattern emerged from the dAPC of lakes from Switzerland whereby the first axis (LD1 9.9% of the variance) mostly separated the species, whereas the second axis (LD2 7.4% of the variance) clustered the populations by lake (Figure 3b). When projecting the partitioned genetic variation of all four lakes in the same dAPC, the first axis discriminated Norwegian and Swiss populations, whereas the second axis discriminated lakes within each region, albeit to a lesser extent (Figure 3c). The third axis separated lakes in Norway, and the fourth axis discriminated limnetic and benthic species from Norway (Figure 3d). The fifth axis separated the species across the entire system (Figure 3e). We note that at the scale of the entire system, despite proximity between sympatric populations for Zurich and Skrukkebukta lakes, populations from the same species show more similarities between them than with sympatric populations of the other species.

The genetic relationships among populations analysed with TreeMix (Figure 4) revealed a clear separation between Norway and Switzerland, with Norwegian populations showing less divergence among them. Within regions, this analysis revealed two levels of signal. Firstly, sympatric populations were grouped by lake, reflecting the homogenizing effect of gene flow following secondary contact in each lake. Secondly, migration links inferred between allopatric populations of the same species (i.e. from different lakes between which gene flow is impossible due to the absence of physical connectivity) illustrated the excess of shared ancestral polymorphism (i.e. shared ancestry). This is particularly the case in Switzerland between the related populations of the limnetic species from Lucerne and Zurich lakes. We observed a more complex pattern in Norway with links between an ancestral population of the Skrukkebukta Lake (shared branch between species from Skrukkebukta Lake) and the benthic species of Langfjordvatn Lake, and reciprocally between an ancestral population of Langfjordvatn Lake and the limnetic species of Skrukkebukta. This suggests an excess of shared polymorphism between populations of the same species between lakes, considering the inferred population bifurcating tree. This pattern was confirmed by allowing only two migration events in TreeMix, which showed a link between Norwegian populations of the limnetic species (Figure S3). Finally, patterns of genetic similarity between allopatric populations of the same species were confirmed by *f*4-statistics, which identified significant excesses of shared polymorphism between populations of the same species between different lakes (*z*-scores, $P < 0.001$) (Table S3). Those results were observed within both Norway and Switzerland when analysed separately.

4 | DISCUSSION

Repeated evolution of phenotypic diversification (i.e. parallel phenotypic divergence) may stem from independent histories of divergence among diverging populations or have a common basis of divergence before independent evolution (Bierne et al., 2013; Welch & Jiggins, 2014). Therefore, determining the mode of divergence and, by extension, inferring the timing of divergence and gene flow between diverging population is required in order to understand the origin of the divergence leading to parallel phenotypic diversification (Herman et al., 2018; Rosenblum, Parent, & Brandt, 2014; Rougeux et al., 2017).

Limnetic and benthic species of the *C. lavaretus* complex offer a relevant system to study the origin of parallel phenotypic differentiation. Whereas previous phylogeographic studies aimed to understand and elucidate the evolutionary history of such diversification, they relied on mtDNA (Østbye et al., 2005) or a combination of mtDNA and reduced genomic representation polymorphisms such as AFLP (Hudson et al., 2010). On the other hand, a recent population genomics study based on RADseq data has mainly focused on the identification of outlier loci associated with divergence between limnetic and benthic species (Feulner & Seehausen, 2018), but without reconstructing the divergence history of the studied lakes. Here, our transcriptome-derived SNP data set allowed achieving an unprecedented coverage of genome-wide variation in the European whitefish. These data thus provide relevant material for inferring the historical demography and mode of divergence between diverging sympatric species pairs of European whitefish. More specifically, we addressed a series of questions pertaining to: (a) the origin of the phenotypic diversification, (b) the timing of divergence between sympatric species, (c) the amount of gene flow during divergence, and (d) the demographic and selective pressures shaping heterogeneous genomic differentiation between species pairs. This study

thus provides a complementary framework for understanding the repeated evolution of the limnetic and the benthic species in whitefish, as it was previously realized for the Lake Whitefish (*C. clupeaformis*) in North America (Rougeux et al., 2017).

4.1 | Multiple evidences of secondary contact

Firstly, we inferred the divergence history of each species pair separately, using the JAFS as a summary statistic of genome-wide differentiation. The best scenario explaining our data was a secondary contact (SC) for each species pair of the four studied lakes. It is noteworthy that these data allowed identifying the secondary contact despite the low degree of genetic divergence between some species pairs observed on the JAFS, and relatively modest sample size per population (but see Fraïsse et al., 2018). This corroborates our finding that the allopatric period was relatively long compared to the period of gene flow following secondary contact (Roux et al., 2016) and that the resulting JAFS showed in each lake an excess of shared intermediate frequency alleles, as expected following the erosion of divergence by gene flow (Alcala, Jensen, & Telenti, 2016).

The secondary contact scenario was clearly supported for each lake. Previous phylogeographic studies (Hudson et al., 2010; Østbye et al., 2005) have inferred secondary contacts between mitochondrial clades (i.e. glacial lineages). However, in some lakes, the origin of sympatric species was assumed to be the result of an adaptive radiation from a single clade due to the presence of a single mtDNA lineage (Østbye et al., 2005). Such a lack of mtDNA polymorphism has also been observed in the American Lake Whitefish system (*C. clupeaformis*), in which one of the five studied lakes where sympatric limnetic and benthic species co-occur (East Lake) is fixed for only one mitochondrial haplotype. Consequently, this observation has previously been interpreted in favour of sympatric divergence in this lake (Pigeon, Chouinard, & Bernatchez, 1997). However, recent historical demographic analysis based on genome-wide polymorphism data inferred a secondary contact between two glacial lineages also in this lake, where mitochondrial lineage fixation has probably been facilitated by asymmetrical gene flow (Rougeux et al., 2017). Here, estimated parameters from the JAFS analysis indicated an important asymmetrical effective population size between sympatric species for most lakes, associated with an asymmetrical migration rate mainly from the limnetic species to the benthic species (except for Lucerne Lake). Interestingly, in Lucerne Lake sympatric species display two different mitochondrial lineages, which is not the case in Zurich Lake where both species are fixed for the "limnetic haplotype" lineage (Østbye et al., 2005). A similar pattern emerges for Norwegian lakes where mainly only one mitochondrial haplotype is more frequent (Kahilainen & Østbye, 2006; Østbye et al., 2005, 2006; Praebel et al., 2013). Thus, the asymmetrical gene flow from the limnetic to the benthic species, which likely occurred due to an asymmetrical effective population size (from the larger to the smaller population) (Beysard, Perrin, Jaarola, Heckel, & Vogel, 2011), has probably resulted in the swamping of the benthic mitochondrial lineage in several lakes, but not in Lucerne Lake in which both clades were described (Østbye et al., 2005).

Our demographic inferences allowed quantifying the proportion of the genome affected by linkage to speciation genes (i.e. genes involved in local adaptation or reproductive isolation between diverging species), which locally reduce gene flow in their vicinity. We also tried to dissociate this effect from that of linkage to selected genes that do not play a role in reproductive isolation (i.e. linked selection). Heterogeneous introgression was found in all four lakes, reflecting the existence of partial reproductive isolation among benthic and limnetic species. Moreover, the most reproductively isolated species pairs have also retained the signal of heterogeneous genome divergence during allopatric isolations, as it has been found for the North American Lake Whitefish (Rougeux et al., 2017). These results confirm that historical demographic inference models accounting for selective effects may help disentangling the relative contribution of the different evolutionary forces (Rougeux et al., 2017; Roux et al., 2016; Sousa et al., 2013; Van Belleghem et al., 2018).

Temporal aspects are fundamental in speciation studies for our understanding of the rate of divergence, the duration of putative gene flow or periods of complete isolation. Our study brings new evidence supporting previous findings based on mtDNA that the timing of allopatric divergence matches with the last glacial period (Østbye et al., 2005). In Norway, we found that the divergence time (T_S) between sympatric species was about 121,000 (CI [53,800;189,000]) years before present (ybp) for Langfjordvatn Lake and about 69,000 (CI [60,300;77,600]) ybp for Skrukkebukta Lake. Similarly, T_S between sympatric species was about 155,000 (CI [31,800;277,800]) ybp for Lucerne Lake and about 107,000 (CI [59,000;154,000]) ybp for Zurich Lake. Such divergence times correspond to the late Würm (Alps) and Weichselian (North Europe) glacial episodes during the Pleistocene (110,000 to 10,000 ybp). This also matches with the glacial lineages identified in a previous study in which one clade was localized in northern Europe and a second clade in a refugia located south of the 53°N (Østbye et al., 2005). Consequently, multiple evidences support that two anciently allopatric glacial lineages have been admixed across Europe following the last glacial retreat. Furthermore, the time of secondary contact between glacial lineages was inferred around 12,400 (CI [4,000;10,600]), 15,600 (CI [0;50,642]), 24,000 (CI [3,300;44,900]) and 28,000 (CI [1,200;55,000]) ybp for Langfjordvatn, Skrukkebukta, Lucerne and Zurich lakes, respectively. Those inferences roughly match with the last glacial retreat allowing lake colonization, which started around 20,000 ybp in Switzerland (Andersen & Borns, 1994) but only later in Norway around 10,300 and 12,000 ybp (Björck, 1995). Of course, those figures must be interpreted cautiously given the broad confidence intervals, yet they all point to relatively recent secondary contact.

Repeated evolution of phenotypic diversification is likely to result from divergent selection acting following post-glacial lake colonization and minimizing competitive interactions with other glacial lineage (Bernatchez et al., 2010; Østbye et al., 2005). Moreover, the evolution of different morphs in sympatry resulting from a contact

between different lineages has been described in several freshwater fishes besides North American and European Lake Whitefish, for instance ciscoes of the *C. artedii* complex (Turgeon & Bernatchez, 2003), but also in anadromous fishes such as the rainbow smelt (*Osmerus mordax*) (Bernatchez, 1997), as well as between avian species such as the pied and the collared flycatcher (Nadachowska-Brzyska et al., 2013).

4.2 | A shared history of divergence

Although the JAFS analyses provided strong support for a secondary contact in each lake, they did not specifically address whether species pairs had a common history before these secondary contacts.

The genetic similarities revealed by TreeMix produced different levels of grouping in the population tree. These different hierarchical levels appeared in the tree as sympatric species pairs clustered by lake, and lakes clustered per region. This pattern likely reflects the relative importance of gene flow within lakes and genetic drift among lakes. However, inferring migration events among populations within the tree enabled us to detect links between populations of the same species but from different lakes. This kind of patterns is more likely to indicate an excess of shared genetic variation considering the inferred population tree supported by the majority of the loci and can be due to gene flow or co-ancestry (Rougeux et al., 2017). Namely, inferred links between limnetic populations from Zurich and Lucerne lakes support the view that these two populations of the same species were genetically similar before being isolated in their respective lakes. Slightly more complex relationships emerged among populations from Norway. Here, the ancestral population from Skrukkebukta Lake was linked to the benthic population of Langfjordvatn Lake, and reciprocally for the ancestral population of Skrukkebukta and Langfjordvatn lakes. This is consistent with an important genetic similarity between all populations from these two lakes due to the more recent divergence and extensive gene flow. It is however important to note that the second most likely tree allowing two migration events linked populations of the limnetic species from Skrukkebukta and Langfjordvatn lakes (Figure S3). Finally, the *f*4-statistics supported the results from TreeMix and revealed genetic similarities between benthic populations from Lucerne and Zurich lakes but also between populations of the same species across Norway and Switzerland (Table S3), suggesting a common origin of the populations of the same species.

The analysis of diversity patterns performed with the dAPC provided a complementary approach to disentangle signals of genetic differentiation between species from genetic differentiation among lakes. At the regional scale, the first axis separated the species (Figure 3a,b), and the second axis separated the lakes. Within the entire system, only the fifth axis illustrated parallel genetic differentiation among the four species pairs. This supports that the signal of past divergence between the two glacial lineages has been largely reshuffled since the onset of secondary contact and the demographic contingencies in each region. However, the projection of limnetic and benthic populations from Lucerne Lake on the fifth discriminant axis indicated the positions of the two least introgressed (more differentiated) populations of the data set, in accordance with results obtained from demographic inferences (Table 1) and with previous observations (Østbye et al., 2005). Therefore, these populations are currently probably the most representative of two glacial lineages in our data set. Interestingly, Zurich Lake populations as well as limnetic and benthic populations from Norway were separated by the same axis. However, Skrukkebukta and Langfjordvatn species pairs showed less differentiation than species pairs from Zurich Lake (Figure 3c). Thus, variable levels of separation along the 5th axis could mirror variable levels of proportion of the ancestry from initial glacial lineages due to variable levels of introgression during the secondary contact.

Complementarity of the approaches realized in this study suggests that the most parsimonious scenario corresponds to a secondary contact that occurred at different dates between two main glacial lineages during glacial retreat. These lineages that correspond to mitochondrial clades identified in Østbye et al. (2005) were isolated for roughly 100,000 years and have subsequently undergone different extents of introgression in different lakes, resulting from variable demographic contingencies following lake colonization some 10,000–20,000 years ago. Because roads of colonization were described in much detail by Østbye et al. (2005), we mostly focused our analysis on deciphering the timing of historical migration between divergent lineages, and the selective pressures that shaped the genomic landscape of divergence before and after secondary contact. However, it is interesting to leave an open question concerning the degree of parallelism in phenotypic diversification. Whereas the secondary contact between two clades has resulted in phenotypic diversification, it is difficult to distinguish whether the diversification occurred before the colonization of lakes, because of retention of ancestral polymorphism in populations of the same species across the system, or whether the diversification occurred in sympatry following colonization events of each lake. The latter would suggest that one clade could have been more prone to switch for the use of an alternative ecological niche, and phenotypically, probably as a consequence of the allopatric phase.

In summary, we propose that the genetic divergence in whitefish species pairs results from a complex interplay between the genetic divergence between glacial lineages during the allopatric phase (allowing for mutation accumulation and lineage sorting) and the reduced introgression around locally adapted loci, but also selective pressures on phenotypes and the demographic contingency of each species pairs. We also argue that integrating selective events in demographic models allows disentangling the relative contribution of deterministic factors (e.g. selection) from historical contingency in order to better understand the historical demography of the populations, and then address questions of biological diversification. This study adds to a growing list of findings that support a role of ancient allopatric divergence in speciation even when phenotypically divergent lineages seem to have emerged in sympatry.

ACKNOWLEDGMENTS

We thank Kim Praebel, Shripathi Bhat and the Freshwater ecology group at UiT for participating with the sampling in Norway, and Ole Seehausen for samples from Switzerland. We are also grateful to Associate Editor A. Gonzalez- Voyer as well as two anonymous reviewers for their constructive feedbacks on this manuscript. This research was supported by a Discovery Research grant from the Natural Sciences and Engineering Research Council of Canada(NSERC) to L.B. L.B also holds the Canadian Research Chair in genomics and conservation of aquatic resources, which funded the research infrastructure for this project.

DATA ACCESSIBILITY

Raw RNAseq sequence data are available through the NCBI Sequence Read Archive (SRA) database under accession SRP136771.

ORCID

Clément Rougeux https://orcid.org/0000-0002-1776-5187

REFERENCES

Alcala, N., Jensen, J. D., & Telenti, A. (2016). The genomic signature of population reconnection following isolation: From theory to HIV. *G3: Genes*, *6*, 107–120. https://doi.org/10.1534/g3.115.024208

Andersen, K. G., & Borns, H. J. (1994). *The Ice Age world: an introduction to Quaternary history andresearch with emphasis on North America and northern Europe during the last 2.5 million years*. (1st ed.). Oslo, Norway: Scandinavian University Press.

Barton, N. H., & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity*, *57*, 357–376. https://doi.org/10.1038/hdy.1986.135

Bernatchez, L. (1997). Mitochondrial DNA analysis confirms the existence of two glacial races of rainbow smelt Osmerus mordax and their reproductive isolation in the St Lawrence River estuary (Quebec, Canada). *Molecular Ecology*, *6*, 73–83. https://doi.org/10.1046/j.1365-294X.1997.00156.x

Bernatchez, L., Renaut, S., Whiteley, A. R., Derome, N., Jeukens, J., Landry, L., … St-Cyr, J. (2010). On the origin of species: Insights from the ecological genomics of lake whitefish. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*, 1783–1800. https://doi.org/10.1098/rstb.2009.0274

Beysard, M., Perrin, N., Jaarola, M., Heckel, G., & Vogel, P. (2011). Asymmetric and differential gene introgression at a contact zone between two highly divergent lineages of field voles (Microtus agrestis). *Journal of Evolutionary Biology*, *25*, 400–408.

Bierne, N., Gagnaire, P. A., & David, P. (2013). The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. *Current Zoology*, *59*, 72–86. https://doi.org/10.1093/czoolo/59.1.72

Björck, S. (1995). A review of the history of the Baltic Sea, 13.0-8.0 ka BP. *Quaternary International*, *27*, 19–40. https://doi.org/10.1016/1040-6182(94)00057-C

Charlesworth, B., Morgan, M. T., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*, *134*, 1289–1303.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … DePristo, M. A. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158.

De Wit, P., Pespeni, M. H., & Palumbi, S. R. (2015). SNP genotyping and population genomics from expressed sequences - current advances and future possibilities. *Molecular Ecology*, *24*, 2310–2323. https://doi.org/10.1111/mec.13165

Douglas, M. R., Brunner, P. C., & Bernatchez, L. (1999). Do assemblages of *Coregonus* (Teleostei: Salmoniformes) in the Central Alpine region of Europe represent species flocks? *Molecular Ecology*, *8*, 589–603. https://doi.org/10.1046/j.1365-294x.1999.00581.x

Elmer, K. R., & Meyer, A. (2011). Adaptation in the age of ecological genomics: Insights from parallelism and convergence. *Trends in Ecology & Evolution*, *26*, 298–306. https://doi.org/10.1016/j.tree.2011.02.008

Endler, J. A. (1986). *Natural selection in the wild*. Princeton, NJ: Princeton University Press.

Feulner, P. G. D., & Seehausen, O. (2019). Genomic insights into the vulnerability of sympatric whitefish species flocks. *Molecular Ecology*, *28*(3), 615–629. http://doi.org/10.1111/mec.14977

Fraïsse, C., Roux, C., Gagnaire, P.-A., Romiguier, J., Faivre, N., Welch, J. J., & Bierne, N. (2018). The divergence history of European blue mussel species reconstructed from Approximate Bayesian Computation: The effects of sequencing techniques and sampling strategies. *PeerJ*, *6*, e5198.

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv* arXiv:1207.3907.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, *5*, e1000695.

Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, *25*, 2454–2466. https://doi.org/10.1111/mec.13582

Herman, A., Brandvain, Y., Weagley, J., Jeffery, W. R., Keene, A. C., Kono, T. J. Y., … McGaugh, S. E. (2018). The role of gene flow in rapid and repeated evolution of cave-related traits in Mexican tetra, *Astyanax mexicanus*. *Molecular Ecology*, *27*, 4397–4416.

Hill, W. G., & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetical Research*, *8*, 269–294. https://doi.org/10.1017/S0016672300010156

Hudson, A. G., Vonlanthen, P., & Seehausen, O. (2010). Rapid parallel adaptive radiations from a single hybridogenic ancestral population. *Proceedings of the Royal Society B: Biological Sciences*, *278*, 58–66.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*, 94. https://doi.org/10.1186/1471-2156-11-94

Kahilainen, K., & Østbye, K. (2006). Morphological differentiation and resource polymorphism in three sympatric whitefish *Coregonus lavaretus* (L.) forms in a subarctic lake. *Journal of Fish Biology*, *68*, 63–79. https://doi.org/10.1111/j.0022-1112.2006.00876.x

Kottelat, M., & Freyhof, J. (2007). Handbook of European freshwater fishes. Publications Kottelat Cornol, Switzerland.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*, 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Homer, N. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*, 2078–2079.

Losos, J. B. (2011). Convergence, adaptation, and constraint. *Evolution*, *65*, 1827–1840. https://doi.org/10.1111/j.1558-5646.2011.01289.x

Nadachowska-Brzyska, K., Burri, R., Olason, P. I., Kawakami, T., Smeds, L. A., & Ellegren, H. (2013). Demographic divergence history of pied flycatcher and collared flycatcher inferred from whole-genome re-sequencing data. *PLoS Genetics*, *9*, e1003942. https://doi.org/10.1371/journal.pgen.1003942

Østbye, K., Amundsen, P.-A., Bernatchez, L., Klemetsen, A., Knudsen, R., Kristoffersen, R., … Hindar, K. (2006). Parallel evolution of ecomorphological traits in the European whitefish *Coregonus lavaretus* (L.) species complex during postglacial times. *Molecular Ecology*, *15*, 3983–4001.

Østbye, K., Bernatchez, L., Naesje, T. F., Himberg, K. J. M., & Hindar, K. (2005). Evolutionary history of the European whitefish *Coregonus lavaretus* (L.) species complex as inferred from mtDNA phylogeography and gill-raker numbers. *Molecular Ecology*, *14*, 4371–4387.

Pearce, R. J., Pota, H., Evehe, M.-S. B., Bâ, E.-H., Mombo-Ngoma, G., Malisa, A. L., … Roper, C. (2009). Multiple origins and regional dispersal of resistant DHPS in African plasmodium falciparum malaria. *PLoS Medicine*, *6*, e1000055–15.

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, *8*, e1002967. https://doi.org/10.1371/journal.pgen.1002967

Pigeon, D., Chouinard, A., & Bernatchez, L. (1997). Multiple modes of speciation involved in the parallel evolution of sympatric morphotypes of lake whitefish (*Coregonus clupeaformis*, Salmonidae). *Evolution*, *51*, 196.

Praebel, K., Knudsen, R., Siwertsson, A., Karhunen, M., Kahilainen, K. K., Ovaskainen, O., … Amundsen, P.-A. (2013). Ecological speciation in postglacial European whitefish: Rapid adaptive radiations into the littoral, pelagic, and profundal lake habitats. *Ecology and Evolution*, *3*, 4970–4986.

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, *461*, 489–494. https://doi.org/10.1038/nature08365

Rosenblum, E. B., Parent, C. E., & Brandt, E. E. (2014). The Molecular Basis of Phenotypic Convergence. *Annual Review of Ecology Evolution and Systematics*, *45*, 203–226. https://doi.org/10.1146/annurev-ecolsys-120213-091851

Rougeux, C., Bernatchez, L., & Gagnaire, P.-A. (2017). Modeling the multiple facets of speciation-with-gene-flow toward inferring the divergence history of lake whitefish species pairs (*Coregonus clupeaformis*). *Genome Biology and Evolution*, *9*, 2057–2074. https://doi.org/10.1093/gbe/evx150

Rougeux, C., Gagnaire, P.-A., Praebel, K., Seehausen, O., & Bernatchez, L. (2018). Convergent transcriptomic landscapes under polygenic selection accompany inter-continental parallel evolution within a Nearctic Coregonus (Salmonidae) sister-species complex. *bioRxiv* 1–27.

Roux, C., Fraïsse, C., Romiguier, J., Anciaux, Y., Galtier, N., & Bierne, N. (2016). Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biology*, *14*, e2000234–22.

Schluter, D. (2000). Ecological character displacement in adaptive radiation. *American Naturalist*, *156*, S4–S16. https://doi.org/10.1086/303412

Schluter, D., & Conte, G. L. (2009). Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(Suppl 1), 9955–9962. https://doi.org/10.1073/pnas.0901264106

Smadja, C. M., & Butlin, R. K. (2011). A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology*, *20*, 5123–5140. https://doi.org/10.1111/j.1365-294X.2011.05350.x

Sousa, V. M. C., Carneiro, M., Ferrand, N., & Hey, J. (2013). Identifying Loci Under Selection Against Gene Flow in Isolation-with-Migration Models. *Genetics*, *194*, 211–233. https://doi.org/10.1534/genetics.113.149211

Sousa, V., & Hey, J. (2013). Understanding the origin of species with genome-scale data: Modelling gene flow. *Nature Reviews Genetics*, *14*, 404–414.

Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R. S. T., … Reinhardt, R. (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications*, *5*, 1–10.

Turgeon, J., & Bernatchez, L. (2003). Reticulate evolution and phenotypic diversity in North American ciscoes, *Coregonus* ssp. (Teleostei: Salmonidae): Implications for the conservation of an evolutionary legacy. *Conservation Genetics*, *4*, 67–81. https://doi.org/10.1023/A:1021860910719

Van Belleghem, S. M., Vangestel, C., De Wolf, K., De Corte, Z., Möst, M., Rastas, P., … Hendrickx, F. (2018). Evolution at two time frames: Polymorphisms from an ancient singular divergence event fuel contemporary parallel evolution. *PLoS Genetics*, *14*, e1007796–26.

Vonlanthen, P., Roy, D., Hudson, A. G., Largiader, C. R., Bittner, D., & Seehausen, O. (2009). Divergence along a steep ecological gradient in lake whitefish (Coregonus sp.). *Journal of Evolutionary Biology*, *22*, 498–514. https://doi.org/10.1111/j.1420-9101.2008.01670.x

Welch, J. J., & Jiggins, C. D. (2014). Standing and flowing: The complex origins of adaptive variation. *Molecular Ecology*, *23*, 3935–3937. https://doi.org/10.1111/mec.12859

Woods, P. J., Müller, R., & Seehausen, O. (2009). Intergenomic epistasis causes asynchronous hatch times in whitefish hybrids, but only when parental ecotypes differ. *Journal of Evolutionary Biology*, *22*, 2305–2319. https://doi.org/10.1111/j.1420-9101.2009.01846.x

Yeaman, S., Hodgins, K. A., Lotterhos, K. E., Suren, H., Nadeau, S., Degner, J. C., … Aitken, S. N. (2016). Convergent local adaptation to climate in distantly related conifers. *Science*, *353*, 1431–1433.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Rougeux C, Gagnaire P-A, Bernatchez L. Model-based demographic inference of introgression history in European whitefish species pairs'. *J Evol Biol*. 2019;32:806–817. https://doi.org/10.1111/jeb.13482