Received: 9 May 2019 | Revised: 6 August 2019 | Accepted: 8 August 2019

DOI: 10.1111/mec.15226

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Polygenic selection drives the evolution of convergent transcriptomic landscapes across continents within a Nearctic sister species complex

Clément Rougeux[1] | Pierre-Alexandre Gagnaire[2] | Kim Praebel[3] | Ole Seehausen[4] | Louis Bernatchez[1]

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

[2]ISEM, CNRS, IRD, EPHE, University of Montpellier, Montpellier, France

[3]Norwegian College of Fishery Science, UiT The Arctic University of Norway, Tromsø, Norway

[4]Aquatic Ecology and Evolution, Institute of Ecology & Evolution, University of Bern, Bern, Switzerland

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

Funding information
This research was supported by a Discovery Research grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to L.B. and strategic funding from UiT to K.P. L.B. also holds the Canadian Research Chair in genomics and conservation of aquatic resources, which funded the research infrastructure for this project.

Abstract

In contrast to the plethora of studies focusing on the genomic basis of adaptive phenotypic divergence, the role of gene expression during speciation has been much less investigated and consequently less understood. Yet, the convergence of differential gene expression patterns between closely related species-pairs might reflect the role of natural selection during the process of ecological speciation. Here, we test for intercontinental convergence in differential transcriptional signatures between limnetic and benthic sympatric species-pairs of Lake Whitefish (*Coregonus clupeaformis*) and its sister lineage, the European Whitefish (*Coregonus lavaretus*), using six replicated sympatric species-pairs (two in North America, two in Norway and two in Switzerland). We characterized both sequence variation in transcribed regions and differential gene expression between sympatric limnetic and benthic species across regions and continents. Our first finding was that differentially expressed genes (DEG) between limnetic and benthic whitefish tend to be enriched in shared polymorphism among sister lineages. We then used both genotypes and covariation in expression in order to infer polygenic selection at the gene level. We identified parallel outliers and DEG involving genes primarily overexpressed in limnetic species relative to the benthic species. Our analysis finally revealed the existence of shared genomic bases underlying parallel differential expression across replicated species-pairs from both continents, such as a *cis*-eQTL affecting the pyruvate kinase expression level involved in glycolysis. Our results are consistent with a long-standing role of natural selection in maintaining trans-continental diversity at phenotypic traits involved in ecological speciation between limnetic and benthic whitefishes.

KEYWORDS

convergence, *Coregonus*, ecological speciation, polygenic selection, population genetics, RNAseq

1 | INTRODUCTION

Deciphering the genomic basis of differential adaptations between divergent populations, ultimately leading to ecological speciation, has been of foremost interest over the last decade. Adaptive divergence implies different population phenotypic responses to constraints associated with selective pressures stemming from different environments. This is particularly well illustrated by the occurrence of independent parallel phenotypic evolution among closely related and locally adapted nascent species (Endler, 1986; Losos, 2011; Orr, 2005). Parallel phenotypic evolution can emerge from repeated divergence of the same genomic regions (Conte, Arnegard, Peichel, & Schluter, 2012) or from different genes involved in similar or different biological pathways (Cohan & Hoffmann, 1989; Losos, 2011) and has been associated with changes in gene expression during adaptive divergence (Harrison, Wright, & Mank, 2012; Manceau, Domingues, Mallarino, & Hoekstra, 2011; Pavey, Collin, Nosil, & Rogers, 2010). Genetic variation underlying parallel phenotypic changes may originate from parallelism at the molecular level that has arisen from de novo mutations affecting the same genes (Manceau, Domingues, Linnen, Rosenblum, & Hoekstra, 2010; Rockman, 2012). However, such mutations are generally associated with loci of large effect controlling the expression of a given phenotypic trait with a mono-/oligogenic architecture (Manceau et al., 2010). This contrasts with the polygenic architecture of most complex traits, including those thought to be involved in ecological speciation, and more generally in adaptation (Gagnaire & Gaggiotti, 2016; Savolainen, Lascoux, & Merilä, 2013; Yeaman, 2015). For such traits, standing genetic variation is usually seen as an important source of adaptive mutations (Welch & Jiggins, 2014). Several recent studies have showed that standing variation may originate from past admixture events (Martin et al., 2015; Meier et al., 2017; Roesti, Gavrilets, Hendry, Salzburger, & Berner, 2014; Rougeux, Bernatchez, & Gagnaire, 2017), suggesting an important role of anciently diverged variants in the process of ecological speciation (Marques, Meier, & Seehausen, 2019). Despite the increasing number of studies underlining the fundamental role of standing variation as the main fuel for adaptation (Barrett & Schluter, 2008; Schrider & Kern, 2017), relatively few have focused on the possible consequences of different levels of standing genetic variation (e.g., because of different historical contingencies) across populations on the fate of parallel phenotypic evolution (Nelson & Cresko, 2018).

In this study, we compare the genomic basis of limnetic–benthic divergence among sympatric species-pairs from two different evolutionary lineages: the North American lake whitefish (*Coregonus clupeaformis* species complex) and the European whitefish (*Coregonus lavaretus* species complex). The two sister lineages *C. clupeaformis* (from North America) and *C. lavaretus* (from Europe) became geographically isolated ~500,000 years ago and have evolved independently since then (Bernatchez & Dodson, 1991, 1994; Jacobsen et al., 2012). In both lineages, several isolated lakes harbour partially reproductively isolated sympatric benthic (normal) and limnetic (dwarf) species-pairs. Limnetic and benthic sympatric species-pairs display sufficient levels of reproductive isolation to be considered as valid biological species (Kottelat & Freyhof, 2007). Therefore, we will use the term "species" throughout the manuscript to refer strictly to limnetic and benthic whitefish, independently of their continent of origin. This system thus offers a valuable model to study the genomic and transcriptomic underpinnings of parallel differential adaptations leading to ecological speciation in independent lineages. The European whitefish species-pairs appear to be the result of a secondary contact between glacial sublineages (Rougeux, Gagnaire, & Bernatchez, 2019), resulting in intralacustrine evolution of benthic and limnetic species across Scandinavian and Alpine lakes (Douglas, Brunner, & Bernatchez, 1999; Østbye et al., 2006; Østbye, Bernatchez, Naesje, Himberg, & Hindar, 2005). The North American lake whitefish sympatric species-pairs are also the result of a postglacial secondary contact between two glacial sublineages during the late Pleistocene. In both lineages, the allopatric phase that likely lasted about 60,000–100,000 years has allowed the accumulation of genomic incompatibilities between sublineages, while secondary contact around 12,000 years ago has provoked character displacement in sympatry, leading to the current phenotypic and ecological divergence (Bernatchez & Dodson, 1990, 1991; Pigeon, Chouinard, & Bernatchez, 1997; Rougeux et al., 2017).

Lake whitefish have been the subject of numerous studies pertaining to the ecological and genomic basis of adaptive divergence between limnetic and benthic species. The limnetic species differ from the benthic species in their use of habitat and trophic resources, with a higher metabolic rate and more active swimming behaviour for foraging and predator avoidance (Bernatchez, Chouinard, & Lu, 1999; Rogers, Gagnon, & Bernatchez, 2002; Trudel, Tremblay, Schetagne, & Rasmussen, 2001), reduced energy allocated to growth relative to benthic whitefish (Rogers & Bernatchez, 2004; Trudel et al., 2001) and differences in morphology, life history and physiological traits (Dalziel, Laporte, Rougeux, Guderley, & Bernatchez, 2017; Dalziel, Martin, Laporte, Guderley, & Bernatchez, 2015; Rogers & Bernatchez, 2007), which are under polygenic control (Gagnaire, Pavey, Normandeau, & Bernatchez, 2013; Laporte et al., 2015). Given the recent divergence between American and European sister lineages, an important part of the divergence between limnetic and benthic would have predated the divergence between continents, especially for genes associated with phenotypic diversification. Thus, the divergence between species may have stemmed from divergent selection acting on standing genetic variation. A possible outcome could be that divergent selection between species has been actively maintaining shared polymorphism at selected variants across continents, protecting them from being lost by drift.

The main goal of this study was to investigate the role of ancestral polymorphism on differential transcriptional signatures between limnetic and benthic species. Our aim was to test the general hypothesis that an overlapping polygenic basis underlies the parallel phenotypic divergence observed between sympatric species-pairs from sister lineages living on two different continents. Specifically, (a) we first documented the amount and functional role of shared ancestral genetic polymorphism within coding genes among populations

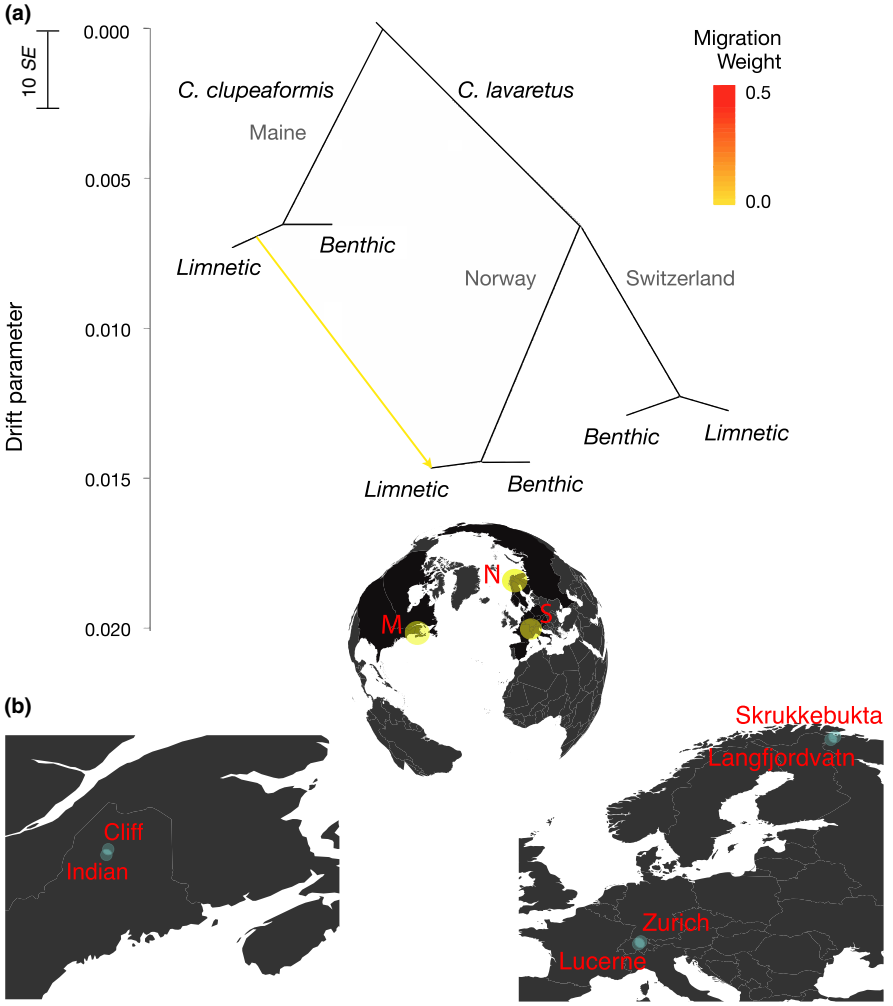

FIGURE 1 Details about the whitefish study system. (a) Treemix analysis illustrating independent differentiation between sympatric Benthic and Limnetic species, from the closely related sister lineages *C. clupeaformis* in North America and *C. lavaretus* in Europe. The lake level was removed by merging populations from the same region for more clarity. Vertical branch lengths are proportional to the amount of genetic drift in each branch; the scale bar indicates 10 times the average standard error (*SE*) of the entries in the covariance matrix between pairs of populations. The colour scale indicates the weight of inferred migration events or shared ancestral polymorphism in absence of possible gene flow across continents, represented by the arrow between tree tips. (b) Locations of the three sampled regions (yellow circles) M: Maine for *Coregonus clupeaformis*, N: Norway and S: Switzerland for *Coregonus lavaretus*. Two lakes (blue circles) containing sympatric species-pairs were sampled per region

of the entire system, (b) we then quantified the extent of differential gene expression between benthic and limnetic species at the local (lake), regional and intercontinental scales, (c) we tested whether genes differentially expressed display an excess of shared polymorphism between sister lineages and finally (d) explored associations between polymorphism and variation in expression at the gene level.

2 | MATERIALS AND METHODS

2.1 | Sample collection, library preparation and sequencing

Coregonus clupeaformis samples were collected from Indian Lake and Cliff Lake, Maine (USA; Figure 1), in 2010. These lakes are part of a well-studied lake whitefish system (Bernatchez et al., 2010) and comprise the most divergent species-pairs along the divergence continuum described in previous studies (Gagnaire, Pavey, et al., 2013; Renaut et al., 2012; Rougeux et al., 2017). In parallel, *C. lavaretus* individuals were sampled in two Scandinavian lakes in Norway (2014): Skrukkebukta, Langfjordvatn and two alpine lakes in Switzerland (2012): Zurich and Lucerne (Figure 1). We chose these European lakes as they each contained only two sympatric limnetic–benthic populations (i.e., excluding potential genetic interactions with other sympatric whitefish forms that occur in other lakes) consistent with our sampling for *C. clupeaformis*. For each species-pair, six benthic and six limnetic individuals were sampled (72 samples in total). Fresh liver biopsies were taken, flash-frozen and stored at −80°C for Lake whitefish, while European whitefish livers were stored directly in RNAlater. All individuals were sampled during summer for each locality, and only mature males were selected for this study in order to reduce sex-specific and life-stage gene expression variability.

Prior to extraction, all samples were assigned randomly to different extraction groups (i.e., groups of samples extracted at the same time) to minimize batch effect of any specific group of samples. Total RNA was extracted from liver tissue pieces of equal size using the RNAeasy Mini Kit following the manufacturer's instructions (Qiagen). RNA quantification was done with a NanoDrop2000 spectrophotometer (Thermo Scientific), and quality was assessed using the 2100 Bioanalyser (Agilent). Only high-quality samples with a RIN value greater than or equal to eight (intact rRNA and no detectable trace of gDNA) were kept

for subsequent steps. Final RNA concentration was measured with Quant-iT RiboGreen RNA Assay Kit (Invitrogen, Life Technologies) 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 HiSeq2000 platform for the 72 libraries randomly distributed on a total of nine lanes (eight libraries/lane) at the McGill University and Genome Quebec Innovation Centre (Montreal, Canada).

2.2 | De novo transcriptome assembly and annotation

Raw sequencing reads were cleaned to remove adaptor and individual tag sequences and trimmed using TRIMMOMATIC v0.36 (Bolger, Lohse, & Usadel, 2014). We applied a quality score threshold of 30 across a 10-bp sliding window and removed all reads <60 nucleotides in length after quality processing. Reads were merged among all individuals using FLASH v1.2.11 with default parameters (Magoc & Salzberg, 2011) and used to assemble a de novo reference transcriptome using TRINITY v2.2.0 suite (Haas et al., 2013). We aimed to build a combined orthologous gene composite reference (hereafter: common reference) for both the North American and European lineages in order to identify and compare orthologous genes involved in the divergence process between limnetic and benthic species across lineages. Successive filtering steps were applied to the de novo common reference transcriptome; contigs lacking an ORF longer than 200 bp were discarded as well as redundant transcripts per ORF in favour of a unique ORF per transcript using TRANSDECODER v3.0.1 pipeline (-Transdecoder.LongOrfs; Haas et al., 2013). In the absence of a reference genome and for comparative purposes with other salmonids transcriptomes, only the longest isoform per transcript was kept (Carruthers et al., 2018; Pasquier et al., 2016). Finally, a scaling factor of one transcript per million (TPM) was applied to normalize the raw reads count per gene for the gene expression analysis. We finally used a BlastX approach against the Swissprot database (http://www.uniprot.org) and the Ensembl *Danio rerio* database (Zv9) to annotate the filtered common reference. In parallel, we also assembled two separate lineage-specific transcriptomes for Lake and European whitefish following the same procedure as detailed above. Using normalized transcripts, we considered the reciprocal best hits within each transcriptome, for both *C. clupeaformis* and *C. lavaretus* to identify and discard paralogous genes (Carruthers et al., 2018). We then blasted each lineage-specific transcriptome to the common reference and discarded unmapped contigs, 98.7% and 98.2% of such orthologous hits for *C. lavaretus* and *C. clupeaformis*, respectively, thus avoiding imbalanced mapping between limnetic and benthic species of both lineages. We kept only common transcripts (i.e., found in both lineages) that we refer to as orthologous genes in the common reference.

2.3 | Differential gene expression analysis

Individual reads were mapped to the orthologous gene common reference with BOWTIE2 v2.1.0 (Langmead & Salzberg, 2012) using the *-end-to-end* mode, and reported multiple alignments were discarded. Then, resulting Bam files were parsed to estimate individual reads counts with EXPRESS v1.5.1 (Roberts & Pachter, 2012). Differential expression analysis was conducted with the R packages DESEQ2 v1.14.1 (Love, Huber, & Anders, 2014). In order to take into account the hierarchical structure of the studied populations, generalized linear models (glm) were built to allow for comparisons between benthic and limnetic species while integrating progressively lakes, regions and continents (hereafter called "phylogeographic") effects, as covariates on gene expression. The final model for limnetic and benthic comparisons across continents was composed of: "~Species*Lake*Continent" in order to integrate interactions between "Continents" and "Species" in "Lake," and interactions between "Lake" and "Species" within both "Continents" successively. We then controlled for the presence of DEGs associated with phylogeographic effects by removing DEGs associated with "Lake" and "Continent" factors (i.e., including interaction and intersection) from the list of identified DEGs, in order to keep only DEGs associated with the "Species" factor. Inference of differentially expressed genes (DEGs) relied on normalized counts matrix by the integration of the size factor per library to correct for heterogeneity in sample sequencing depth. DEGs were determined by controlling for false discovery rate (FDR) as implemented in DESEQ2 (Benjamini–Hochberg correction; Benjamini & Hochberg, 1995), with a threshold of a FDR < 0.05. Then, GO enrichment analysis was performed with GOATOOLS (Klopfenstein et al., 2018), based on Fisher's exact test. For all tested lists of genes, GO enrichment was associated with FDR < 0.05 (Benjamini & Hochberg, 1995) and a minimum of three genes represented per category.

2.4 | SNP genotyping and sequence divergence

In order to document the extent of polymorphism within and divergence between *C. clupeaformis and C. lavaretus* and among divergent sympatric benthic and limnetic species-pairs, individual reads were mapped (71.41% overall alignment mean success rate) to the common reference transcriptome using BOWTIE2 v2.1.0 "end-to-end" alignment (Langmead & Salzberg, 2012). Resulting SAM files were converted to BAM files and sorted using SAMTOOLS v1.3 (Li et al., 2009) and duplicates were removed with the PICARD-TOOLS program v1.119 (http://broadinstitute.github.io/picard/). The physical mapping information of reads to the reference was used for calling SNPs with FREEBAYES v0.9.10-3-g47a713e (Garrison & Marth, 2012). Variable sites were considered for a minimum coverage of three reads per individual in order to process a site and for a minimum of two reads per individual to consider an alternative allele. We used the *vcffilter* program from *vcflib* (https://github.com/ekg/vcflib) to process the variant call format (VCF) file obtained from FREEBAYES, in order to specifically retain biallelic SNPs with a phred scaled quality

score above 30, a genotype quality with a phred score higher than 20, and an allele depth balance between 0.30 and 0.70. Following these quality control steps, we filtered the resulting VCF file using VCFTOOLS v0.1.12b (Danecek et al., 2011), in order to remove miscalled and low-quality SNPs for subsequent population genomics analyses. For each of the 12 populations, we kept loci with less than 10% of missing genotypes and filtered for Hardy–Weinberg disequilibrium using a *p*-value exclusion threshold of 0.01. Finally, we merged the VCF files from all the 12 populations, resulting in a unique VCF file containing 20,911 SNPs passing all the filters in each population. Since we did not apply any minor allele frequency threshold within populations, the final VCF represents a nonascertained dataset of genetic variation. Intrapopulation nucleotide diversity (π) was estimated within nonoverlapping 100-bp windows (due to transcriptome data specificities, i.e., N50 of 1,672 bp) with VCFTOOLS v0.1.12b (Danecek et al., 2011) on the VCF file of shared variant and invariant sites. We then reported the mean π per gene as the a posteriori mean of all windows per gene. Finally, we estimated the between-species nucleotide diversity (*Dxy*) with a custom *perl* script, using nonoverlapping 100-bp windows. Absolute divergence (*Dxy*) was calculated as the fraction of nucleotide differences between two sequences taken from two different species (Cruickshank & Hahn, 2014; Nei, 1987). As for the nucleotide diversity, we estimated the mean *Dxy* per gene from all windows contained in each gene.

2.5 | Shared polymorphism and historical relationships among lineages and species

Historical relationships among all populations were inferred using TREEMIX v1.12 (Pickrell & Pritchard, 2012) applied to the VCF file containing 20,911 polymorphic SNPs. This program uses the covariance structure of allele frequencies between all tested populations and a Gaussian approximation for genetic drift to build a maximum-likelihood graph relating populations with an ancestral genetic pool. The number of migration edges was determined empirically to improve the fit to the inferred tree. Migration edges among and between species may either reflect gene flow, or the retention of shared ancestral polymorphism among geographically isolated populations.

We then quantified and compared the amount of shared polymorphism retained at the lake, region and continental hierarchical levels. More precisely, we defined the number of SNPs that were shared and polymorphic among all populations for each hierarchical level (i.e., between sympatric species at the lake level, among all populations from Norway, Switzerland and Maine at region level, and among all populations across both continents at continent level). We also tested for the increased probability of limnetic–benthic DEGs relative to non-DEGs to display shared polymorphism, which could hint to a possible role for selection in maintaining variation at these genes. We defined the proportions of DEGs with shared polymorphism (DEG_{SP}) relative to the total number of DEGs (DEG_T) and of non-DEGs with shared polymorphism ($NDEG_{SP}$) relative to the total number of non-DEGs ($NDEG_T$). From these two proportions, we realized a ratio test to compare the relative proportion of shared polymorphism (SPRT: shared polymorphism ratio test) in each category of genes (i.e., DEGs and NDEGs):

$$SPRT = \frac{(DEG_{SP}/DEG_T)}{(NDEG_{SP}/NDEG_T)}$$

Confident intervals (CIs) of 95% were determined using 1,000 bootstrapping iterations per comparison on the empirical data set. Obtained ratios and associated CIs were compared to the expected ratio of one (i.e., no difference in the amount of shared polymorphism between DEGs and NDEGs) to test for enrichment of shared polymorphism in DEGs (i.e., SPRT > 1).

2.6 | Detection of adaptive variation

The detection of adaptive variation using F_{ST}-based approaches is challenging in study systems with complex population structures, which can be accounted for by multivariate outlier detection methods (Duforet-Frebourg, Luu, Laval, Bazin, & Blum, 2016; Luu, Bazin, & Blum, 2017). Redundancy analysis (RDA; Legendre & Legendre, 1998) is an efficient constrained ordination method to detect (adaptive) variation under the effect of divergent selection, especially when the selection gradient is weakly correlated with population structure (Capblancq, Luu, Blum, & Bazin, 2018; Forester, Lasky, Wagner, & Urban, 2018). In order to account for the hierarchical population structure, we used a conditioned (partial) redundancy analysis (cRDA), as implemented in the VEGAN v2.4-3 (Jari Oksanen et al., 2018) R package, to identify genes that diverge the most between limnetic and benthic species, independently of the population structure and hierarchical (region and continent) levels. Consequently, the first axis of the cRDA corresponds to the variance explained by the constrained "Species" variable and the second axis corresponds to the PC1 of the PCA (nested into the RDA) and represents the main axis of unconstrained variance. Because RDA can be sensitive to missing data, we only kept loci that contained no missing genotypes across populations, representing a total of 9,093 SNPs. Briefly, a RDA allows evaluating the variation that can be explained by the applied constraints. We conditioned the RDA to remove the effects of continents, regions and lakes to control for the hierarchical genetic structure. We tested the significance level of the cRDA with an analysis of variance (ANOVA), performed with 1,000 permutations. From the conditioned ordination, each SNP was assigned a locus score that corresponds to the coordinates used to ordinate points and vectors. Then, we identified outlier SNPs by putting significance thresholds at ±2.6 and 3.0 standard deviations from the mean score of the constrained axis, corresponding to *p*-value thresholds of 0.01 and 0.001, respectively (Forester et al., 2018).

We also applied cRDA to gene expression data (i.e., on the 32,725 orthologous genes from the common reference) and tested the significance of the constrained ordination model with an ANOVA using 1,000 permutations. We aimed at identifying covarying DEGs between limnetic and benthic species after correcting for hierarchical population structure and applying a significance threshold

on expression scores (*p*-value of .01 and of .001). This approach is expected to limit the local effect of highly expressed genes in the identification of outliers/DEGs if they were not associated with the constrained independent variables. Such covarying DEGs could reflect the effect of polygenic selection acting in parallel between benthic and limnetic whitefish.

2.7 | Gene subnetwork analysis

In order to test for selection acting on subnetworks of genes involved in common biological pathways, we performed a gene network analysis designed specifically to detect polygenic selection (Gouy, Daub, & Excoffier, 2017). The level of differential expression between limnetic and benthic whitefish captured by individual locus scores in the expression cRDA was scaled to a *z*-score, such that individual locus scores have a mean of 0 and a standard deviation of 1. We obtained KEGG Ontology (KO) for each transcript of the common reference with an Entrez gene ID annotation from the KASS (KEGG Automatic Annotation Server, http://www.genome.jp/tools/kaas/). Polygenic selection was tested using the R package *signet* (Gouy et al., 2017) on the *Danio rerio* and *Homo sapiens* KEGG databases (we present only the results obtained from the human database because of lack of power using the smaller *D. rerio* database). This package defines subnetworks of genes that interact with each other and present similar patterns attributed to selection, such as covariation in expression level for genes involved in the same biological pathway. A null distribution of subnetwork scores was generated by random sampling to create 10,000 subnetworks of variable sizes. Each pathway of the KEGG database was parsed to identify gene subnetworks with a high score using 10,000 iterations of simulated annealing. Finally, the *p*-value of the subnetworks showing parallelism (i.e., for which we found evidence of differential expression between most to all limnetic and benthic populations) in gene expression was inferred based on the distribution of 10,000 permuted scores from the randomly generated subnetworks. We then tested for similarities among limnetic species for differential gene expression against benthic species across continents (subnetworks $p < .05$).

2.8 | eQTL analysis

We related differential gene expression with sequence divergence to identify eQTLs. We thus generated a new VCF file containing shared loci among all the populations that showed polymorphism across continents (i.e., trans-continental polymorphisms shared between *C. clupeaformis* and *C. lavaretus*), which corresponded to 2,240 SNPs. We extracted the 1,272 associated annotated genes and their expression level and tested for correlations between genotype and expression level (eQTL), using different models. We applied a linear model testing for gene expression variation in response to genotype variation by controlling for the lake and continent covariates as environmental effects to correct for population genetic structure (expression ~ genotype+covariates). We compared the tested linear model to a theoretical simulated data set, as suggested by Shabalin (2012). This analysis was run using the R package MATRIXEQTL v2.1.1 (Shabalin, 2012). We identified significant (FDR < 0.05) *cis*-eQTL by focusing on SNPs affecting the expression level of the gene to which they were physically linked, from the output of the linear model including all covariates.

3 | RESULTS

3.1 | Reference transcriptome assembly

A total of 1.74×10^9 100 bp single-end reads (average of 24.22×10^6 reads per individual) were generated from 72 individuals (Table S1). Filtered libraries (1.69×10^9 reads) were used to de novo assemble a composite reference transcriptome. Starting from a raw assembly of 277,194 contigs (mean length = 768 bp; N50 = 1,347 bp), we kept only the longest ORF per transcript, which reduced the initial assembly by 66.9% (91,715 contigs remaining, see Table S2). Keeping the longest isoform per transcript and normalizing the reads distribution resulted in a composite reference transcriptome composed of 54,514 contigs (mean length = 1,121 bp; N50 = 1,672 bp) with 79% of uniquely annotated transcripts (43,501 contigs). This liver-specific assembly is consistent with the number of transcripts assembled using several organs separately in one *C. clupeaformis* male and one *C. lavaretus* female (range: 66,996–74,701, respectively; Pasquier et al., 2016). Comparing transcriptomes separately assembled in *C. clupeaformis* (55,104 contigs, with 83.8% uniquely annotated) and *C. lavaretus* (58,321 contigs, with 73.9% uniquely annotated) to identify orthologous genes and filtering out paralogous genes (i.e., self-mapped transcripts hits, 15%), we ended up with a common reference transcriptome of 32,725 annotated contigs (comparable to Carruthers et al., 2018) that were used for downstream analyses of gene expression and sequence divergence. The common reference transcriptome N50 was 1,797 bp with a contig size distribution ranging from 297 bp to 13,427 bp and a mean contig size of 1,185 bp.

3.2 | Genetic relationships among populations

Characterizing the genetic relationships among the studied populations with TreeMix indicated the presence of shared polymorphisms maintained across the entire hierarchical genetic structure (Figures 1 and S1). The different hierarchical levels were composed by limnetic–benthic species-pairs (hereafter, Species-pair level) in both North America and Europe (the lake level was removed for more clarity in this analysis). Similarities in branch length at the Species-pair level reflected the similar degrees of genetic differentiation among species-pairs from different regions consistent with a relatively similar timing of divergence and postglacial admixture among species-pairs from both continents. The second hierarchical level was composed by intracontinent regional divergence (hereafter, Region level), which was represented by the two European regions: Central alpine (Switzerland) and Fennoscandinavia (Norway). The highest hierarchical level of divergence was between the two sister lineages *C. clupeaformis* and *C. lavaretus*. The sharing of ancestral

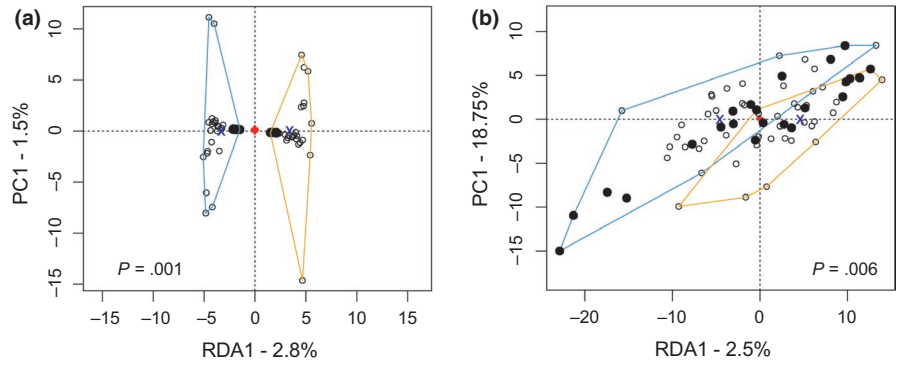

FIGURE 2 Conditioned redundancy analysis (cRDA) clustering individuals per species (all limnetic vs. all benthic from both continents). The cRDA was used to capture divergence between limnetic and benthic species (along RDA axis 1) while correcting for three hierarchical levels of population structure (lake, region and continent) based on (a) genotypes and (b) gene expression levels. Orange and blue clusters correspond to benthic and limnetic species, respectively. Circles represent individuals. Open and filled circles represent the *Coregonus lavaretus* and *Coregonus clupeaformis* lineages, respectively

polymorphism between geographically isolated taxa across continents was captured by the inferred migration link connecting two limnetic populations from Maine and Norway (Figure 1). This link indicates an excess of shared ancestral polymorphism after accounting for drift along the population tree, which could indicate the presence of balanced polymorphisms across continents or past admixture events.

3.3 | Trans-continental polymorphism quantification

Given the evidence for shared polymorphisms maintained among species from different continents, we documented the overall extent of trans-continental polymorphism. Trans-continental polymorphism corresponds to ancestral variation shared among all populations of limnetic and benthic species from North America (*C. clupeaformis*) and Europe (*C. lavaretus*), that is, loci that are polymorphic in all populations on both continents. Among the 20,911 SNPs initially obtained after genotyping and filtering steps, we identified 2,241 SNPs (10.7%) distributed among 1,251 genes (3.8%) that met our criteria of trans-continental shared polymorphic loci. The genes containing trans-continental polymorphisms showed a significantly higher mean level of nucleotide diversity (π) per gene within species (Wilcoxon signed-rank test, $p < .001$, Figure S2), and a higher mean level of absolute sequence divergence (*Dxy*, Nei, 1987) per gene between limnetic and benthic species (Wilcoxon signed-rank test, $p < .001$, Figure S3) compared to genes with no trans-continental SNPs. These results point to the influence of evolutionary processes (after controlling for artefact from the data, Figure S4) acting differentially between these two categories of genes.

3.4 | Parallel genetic differentiation

We used conditioned ordination to test whether divergence at the limnetic/benthic species-pair level involves parallel changes in allele frequency across sister lineages from different continents (*C. clupeaformis-C. lavaretus*). We conditioned the ordination to account for the hierarchical genetic structure among populations (lake, region and continent). The cRDA thus allowed the identification of variants associated with limnetic–benthic species divergence, explaining 2.8% of the total genetic variance (ANOVA, $F = 1.259$, $p = .001$), after controlling for the variance explained by regional and continental population structure (Figures 2a and S5). The distribution of individual locus scores on the first cRDA axis discriminating all limnetic and benthic samples allowed identifying 348 outlier markers ($p < .001$, 3.0*SD*) showing parallel allele frequency differences between limnetic and benthic species across both continents (Figure S6). These 348 SNPs, which represent 15% of the 2,241 of the trans-continental polymorphic loci, may be interpreted as being enriched for shared genetic bases of limnetic–benthic species divergence across continents. Gene ontology (GO) analysis of transcripts associated with parallel outlier SNPs revealed significant enrichment ($p < .001$) in metabolic process (i.e., catabolism), immune system process and developmental process, among others (Table S1).

3.5 | Differential gene expression between Limnetic and Benthic Species

Using the benthic whitefish populations as the reference level, we quantified differentially expressed genes (DEG) between limnetic and benthic whitefish at the Lake and Region levels using multivariate glm. Despite different sampling locations, storage methods and sequencing runs, we did not observe any sign of experimental bias in our transcriptomic data (Figure S7). Our expectation was a decreasing number of shared DEGs in higher comparisons levels, due to a reduced fraction of shared regulatory variants. In North America, Cliff and Indian Lakes showed 3,175 (9.7%) and 238 (0.7%) significantly differentially expressed genes (false discovery rate; FDR < 0.05) between sympatric limnetic and benthic species, respectively (Figure 3). In both lakes, approximately twice as many genes showed higher expression in the limnetic species compared to the benthic species (Cliff Lake: 2,001 vs. 1,174, χ^2 test, $p < .001$; Indian Lake: 159 vs. 79, χ^2 test, $p < .001$; Figure S8). The lower level of DEGs

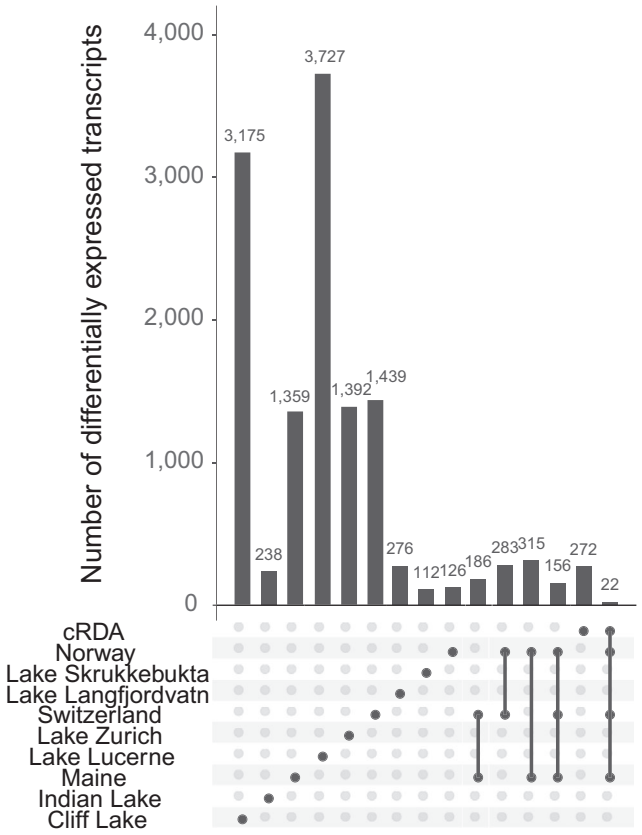

FIGURE 3 Frequency of shared differentially expressed transcripts between species across hierarchical levels. Limnetic/Benthic comparisons are indicated by the dots (intralake and intraregion). Linked dots represent the combinations of pooled populations per species for interspecies comparison levels (inter-regions and intercontinents). The number of significant differentially expressed genes (DEGs) associated with species divergence (FDR < 0.05) per comparison is indicated on top of each bar, as determined from the multivariate tests, from the cRDA for all limnetic to all benthic comparison, as well as the overlap of DEGs between cRDA and DESeq2 (last column)

identified in Indian Lake was likely associated with lower level of differentiation between species-pairs, but also to a much higher interindividual variance in this lake (Figure S8). While we found 44 common DEGs, using a less stringent significance threshold (*q*-value < 0.1 instead of 0.05) led to the identification of 1,926 DEGs in Indian Lake, 318 of which were shared with the 3,175 DEGs from Cliff Lake. In Langfjordvatn and Skukkebukta lakes from Norway, 276 and 112 significant DEGs were identified between limnetic and benthic whitefish, respectively. Contrary to North America, more genes showed lower expression in limnetic populations (45 vs. 231, χ^2 test, p < .001 in Langfjordvatn; 44 vs. 68, χ^2 test, p = .023, in Skrukkebukta; see also Figure S8). In Swiss lakes, 3,727 and 1,392 genes showed a significantly different expression level between limnetic and benthic species in Lake Lucerne and Lake Zurich, respectively. In contrast with North America and Norway however, a similar number of genes showed lower and higher expression in the limnetic species compared to benthic species in both lakes (1,870 vs. 1,857, χ^2 test, p = .831 in Lake Lucerne and 691 vs. 701, χ^2 test, p = .787, in Lake Zurich).

To document the extent of parallelism at the Region level (i.e., common DEGs between species across lakes within a given region), we considered the proportion of DEGs between limnetic and benthic whitefish while integrating the lakes as covariates in a multivariate glm. Thus, the degree of parallelism in DEG was most pronounced in Swiss lakes whereby 1,439 parallel DEGs were found between limnetic and benthic species among lakes. These parallel genes showed a higher proportion of downregulated genes in limnetic species (625 vs. 814, χ^2 test, p < .001; Figure S9). Similarly in Maine, 126 parallel DEGs were identified between limnetic and benthic species from Indian and Cliff lakes. These genes showed a higher gene expression in the limnetic species than in the benthic species (818 vs. 541, χ^2 test, p < .001; Figure S9). Finally, 126 parallel DEGs were identified between limnetic and benthic whitefish from the two Norwegian lakes. Here, however, parallel DEGs comprised a significantly higher proportion of downregulated genes in the limnetic species (53 vs. 239, χ^2 test, p < .001; Figure S10).

GO enrichment analysis at the Lake and Region levels provided evidence for parallelism in biological functions. Indeed, DEGs were significantly enriched (FDR < 0.05, see Table S3) in limnetic species for immune system response, detoxification and antioxidant activity in both North America and Europe. Moreover, we found enrichment in genes associated with growth and development at both Lake (Indian, Lucerne, Skrukkebukta and Langfjordvatn lakes) and Region levels (Maine) in benthic whitefish. DEGs were also enriched in limnetic species for metabolic processes, electron carrier activity and catabolic processes, which are associated with differences in the metabolic rate between limnetic and benthic species (Dalziel, Laporte, Guderley, & Bernatchez, 2017; Dalziel, Laporte, Rougeux, et al., 2017; Laporte, Dalziel, Martin, & Bernatchez, 2016).

At the Continent level (integrating Region and Continent as covariates), we found 156 parallel DEGs between limnetic and benthic species between both continents. Again, these 156 genes showed similar proportions of up- and downregulated genes in all limnetic whitefish compared to all benthic whitefish (72 vs. 84, χ^2 test, p < .299; Figure S10). From enriched biological functions included metabolic process (p = .016) and antigen binding (p = .021) associated with immune response (e.g., immunoglobulin domain) and cellular metabolic process (p = .002; e.g., SPRY-associated domain). DEGs were also enriched for oxido-reductase activity (e.g., *TSTA3*, a gene able to activate fructose and mannose metabolism via an oxydo-reductase step, involved in glycolysis; *Hsp90*, a gene responding to environmental stress with effects on growth).

3.6 | Enrichment in trans-continental polymorphism in DEGs

The excess of ancestral polymorphism shared among limnetic and benthic species across both continents suggests the existence of a mechanism responsible for the maintenance of balanced ancestral variation, against the stochastic effect of drift in each lineage.

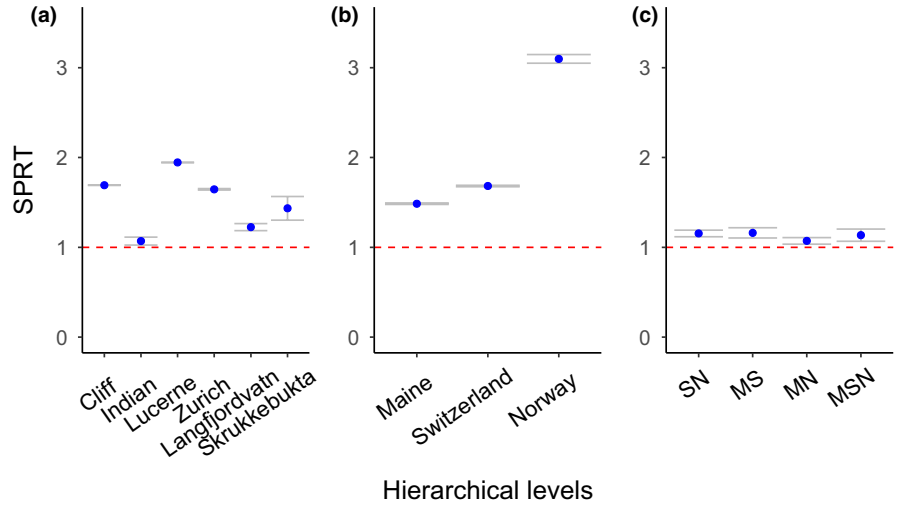

FIGURE 4 Shared polymorphism enrichment in DEGs between species compared to Non-DEGs. Test ratio (SPRT) of the proportion of shared trans-continental polymorphism in DEGs compared to non-DEGs, for (a) intralake, (b) interlake within regions and (c) inter-regions (SN: Switzerland/Norway, MS: Maine/Switzerland, MN: Maine/Norway and MSN: Maine/Switzerland/Norway) comparisons. All ratios are above the red-dashed line ($y = 1$), which illustrates the general enrichment of shared polymorphism in DEGs compared to non-DEGs. Grey bars indicate the 95% confidence interval associated with the observed value obtained using 1,000 bootstrap resampling of the data set

In order to further test if the retention of ancestral polymorphism could be linked to differential selection on adaptive traits between limnetic and benthic species, we tested if *cis*-regulating regions (i.e., regions physically linked to transcripts) of DEGs show an increased probability of having shared polymorphisms. The shared polymorphism ratio test (SPRT), which compares the proportions of shared polymorphism in DEGs to nondifferentially expressed genes (NDEGs), revealed an enrichment of shared polymorphism in DEGs at the three hierarchical levels (Figure 4), and a parallel analysis suggested that DEGs do not have a higher level of expression (i.e., a higher sequencing depth) than NDEGs (Figure S11).

3.7 | Identification of gene subnetworks showing patterns of parallel gene expression

A cRDA performed on the expression data of the 32,725 orthologous genes revealed that 2.5% of total variance in expression was explained by net differences between limnetic and benthic species across continents (ANOVA, $F = 2.516$, $p = .006$, Figure 2b), while accounting for the hierarchical population structure. The z-scored distribution of the gene expression on the first RDA axis constrained for divergence between limnetic and benthic whitefish ranged from [−4.14; 3.99] (Figure S12). Applying two different significance thresholds ($p < .01$ and $p < .001$) allowed identifying 272 ($p < .01$) and 66 ($p < .001$) putative outliers DEGs. These were significantly enriched for biological regulation ($p < .001$) and metabolic process ($p < .001$) in both sets of genes, and growth ($p = .026$) for the subset of 272 genes (Table S3). Twenty-two out of the 156 (14%) parallel DEGs at the Continent level (which is more than expected by chance, hypergeometric test, $p < .001$) identified with the *glm* analysis overlapped with the 272 DEGs from the cRDA on gene expression, including the previously mentioned genes *TSTA3* and *Hsp90*.

We then investigated the polygenic basis of transcriptomic differences using genes expression scores defined by the ordination analysis. The *z*-scored transformation of cRDA's gene expression scores was used as a quantitative measure for assessing the extent of parallel expression between limnetic and benthic whitefish across continents. A total of 22,188 out of the 32,725 orthologous genes (from the common reference) that were successfully annotated with an Entrez gene ID were analysed with *signet* based on information from KEGG databases. In *signet*, gene subnetworks (i.e., genes showing patterns of convergence within limnetic populations) were identified for each pathway and we considered the significance of subnetworks ($p < .05$) in the analysis against the *Homo sapiens* KEGG database.

Ten metabolic pathways with significant subnetworks of genes were identified (Table S4). Five of these pathways shared genes and were therefore merged together to identify genes showing convergent patterns among the significant subnetworks (Figures 5 and S13), mainly for peripheral genes (Figure S14). From the ten identified metabolic pathways composed of 73 parallel DEGs between species, three categories of metabolic functions were represented. The first category corresponded to energetic metabolism (e.g., pentose phosphate pathway, glycerolipid metabolism, nicotinate and nicotinamide metabolism, FoxO signalling pathways) which is involved in regulation of glycolysis and energy production (from ATP to NADH). The second category was the detoxification metabolism and immune system (e.g., CYP450 and glutathione metabolism), which is mainly associated with detoxification and oxidative stress, maintaining cell integrity by preventing damage due to reactive oxygen species (ROS). The third category was the cell cycle metabolism and control (e.g., FoxO signalling pathways, purine and pyrimidine metabolism, cAMP signalling pathway, PI3K-Akt signalling pathway). These pathways

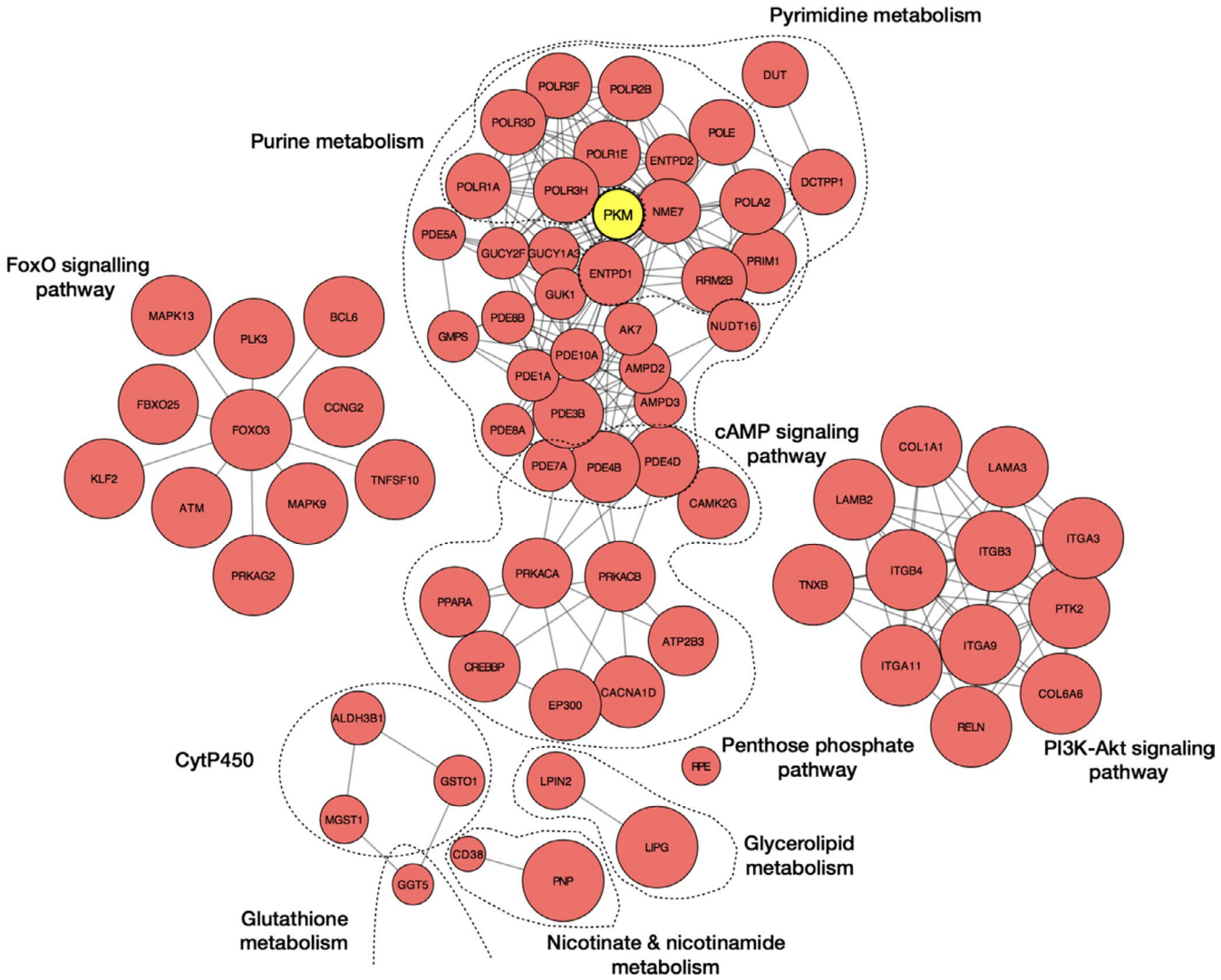

FIGURE 5 Merged significant subnetworks for Limnetic/Benthic species comparisons. Detailed subset of differentially expressed genes between species showing convergence across the studied system, separately displayed for each metabolic pathway. Pyruvate kinase (PKM) gene expression (in yellow) is associated with a *cis*-eQTL in its 3′UTR. Gene annotation is based on the Ensembl nomenclature. For each node represented by a gene, the relative size is proportional to the contribution score of the associated gene to the significance of the metabolic pathway. The score for each gene corresponds to a probability of convergent adaptation between individuals of the same species

play critical roles in regulating diverse cellular functions including metabolism, growth, proliferation, survival, transcription and protein synthesis.

3.8 | cis-eQTL markers associated with expression

We finally tested whether the level of expression was associated with SNPs located within a given gene. We found 451 SNPs significantly ($p < .05$) associated with the level of expression of the gene (i.e., UTR and coding sequence) to which they belong (*cis*-eQTL). That is, the level of expression significantly varied among the three possible genotypes at a given SNP. Controlling for multiple tests and testing for deviation in the distribution of *p*-values due to unaccounted covariates (Figure S15), we retained 134 significant (FDR < 0.05) *cis*-eQTL across continents associated with differences between limnetic and benthic species.

We identified SNPs and genes showing overlap between the different analyses. Thus, two significant *cis*-eQTL overlapped the 20 outliers SNPs from the cRDA based on genetic variation (hypergeometric test, $p < .001$). They were physically linked to the complement factor H (*CFH*, Entrez 3075; Figure 6b) and protein kinase AMP-activated noncatalytic subunit beta 1 (*Prkab1*, Entrez 19079; Figure 6c). These genes are, respectively, involved in immune response and in regulation of the cellular energy metabolism. Moreover, among the 134 significant *cis*-eQTL, 19 genes were shared with genes identified in the polygenic subnetwork analysis (hypergeometric test, $p = .006$). However, only the pyruvate kinase gene (*PKM*, isoform M2; Entrez 5315) remained significant in both (subnetworks and eQTL) analyses (Figure 4). This gene encodes a protein involved in glycolysis, which generates ATP and pyruvate. The level of expression of this gene was higher in heterozygous and homozygous individuals for the minor allele (Figures 6a and S16, linear model, $p = .001$). Finally,

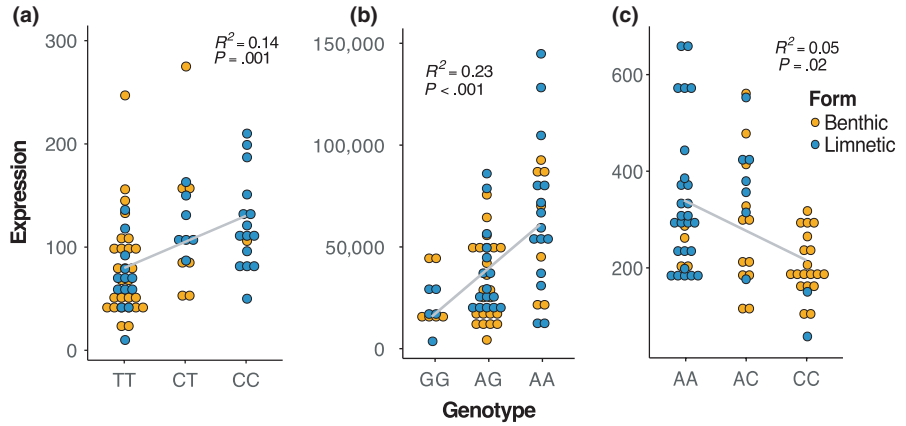

FIGURE 6 Associations between significant *cis*-eQTL genotypes and the level of expression (relative counts) of three genes in limnetic and benthic whitefish species, independently of their geographic origin. Three examples of transcripts abundance per individual (circles) varying with genotypes for a *cis*-eQTL located in 3′UTR of (a) pyruvate kinase (PKM), (b) complement factor H (CFH) and (c) protein kinase AMP-activated noncatalytic subunit beta 1 (Prkab1) genes. The grey line corresponds to the linear model fitted to the data and associated statistics (coefficient of determination: R^2 and *p*-value: *p*) detailed in each panel. Individuals from benthic and limnetic sister species are represented in orange and blue, respectively

we inferred the gene structure (i.e., identification of 5′ and 3′ UTR, exonic and intronic regions) of our de novo assembled transcriptome and more particularly the *PKM* gene. We localized the variant affecting the level of expression of the *PKM* gene in its 3′UTR region, which could impact the regulation of the transcription of this gene.

4 | DISCUSSION

Coregonus clupeaformis (North America) and *C. lavaretus* (Eurasia) sister lineages have been geographically isolated for the past 500,000 years (Bernatchez & Dodson, 1991, 1994; Jacobsen et al., 2012). Yet, they maintained similar habitat preferences in cold freshwater lakes (Bernatchez & Dodson, 1991; Douglas et al., 1999; Østbye et al., 2006, 2005), with a frequent occurrence of sympatric species-pairs being, respectively, associated with benthic and limnetic ecological and trophic niches (Amundsen, Bøhn, & Våga, 2004; Häkli, Ostbye, Kahilainen, Amundsen, & Praebel, 2018; Kahilainen & Østbye, 2006; Landry, Vincent, & Bernatchez, 2007; Lu & Bernatchez, 1999). In both *C. clupeaformis* and *C. lavaretus*, historical demographic events and selective processes initiated species diversification (Rougeux et al., 2017, 2019) and resulted in a repeated ecological specialization to limnetic and benthic habitats in each region. Here, the analysis of gene sequence divergence and differential expression in limnetic–benthic species has the potential to provide new insights into the genomic bases of parallel adaptation and parallel ecological speciation.

A salient result from our analysis is that pairs of limnetic and benthic species from independent divergence events exhibit parallelism in DEGs associated with repeated divergent adaptation to different ecological niches (i.e., limnetic and benthic niches). The identification of 156 significant parallel DEGs involved in energetic metabolism, immune response, cell cycle and growth is congruent with previous transcriptomic analysis conducted in *C. clupeaformis* on the same organ tissue but with low-resolution methods (microarrays with a reduced representation of transcripts), highlighting life history tradeoffs between growth and energetic costs associated with occupying the limnetic niche (Jeukens, Bittner, Knudsen, & Bernatchez, 2009; St-Cyr, Derome, & Bernatchez, 2008). Those congruent results also highlight the fact that despite differential storage methods, our data set was not affected by any of the controlled batch effects and that such putative source of stochasticity (increased variance) did not significantly affect our inference of differential expression and parallelism between species. Moreover, our results show that seeking to detect parallel DEGs based on a single gene approach may lack the power to detect polygenic changes in expression levels, as it could be expected for the complex phenotypic traits involved in the divergence of these species-pairs. Indeed, a gene-by-gene approach may be too conservative and not well adapted to capture subtle parallel expression differences at genes involved in the same biological pathways under selection. The statistical approach employed to detect polygenic selection based on transcript abundance covariation allowed us to integrate this level of information. Consistent with results from the negative binomial glm, the RDA allowed identifying a parallel genetic basis of phenotypic and ecological divergence by revealing parallel DEGs between limnetic and benthic species. Moreover, we found that these DEGs are involved in several metabolic pathways belonging to energetic, growth, cell cycle metabolisms and transcription factor, regulating genes associated with energetic metabolism, as *observed* in Jacobs et al. (2018). This approach also allowed detecting congruent expression signals at the integrated pathway scale, where the same effect on the selected phenotype can be achieved via regulation of different genes, because of the complexity and redundancy of the multigenic regulatory systems (Yeaman, 2015). The accumulated results coupled with previous analyses on this system thus highlight the repeated

action of natural selection on gene expression patterns (Jeukens & Bernatchez, 2011; Jeukens et al., 2009; St-Cyr et al., 2008), as well as on partially shared polygenic bases of phenotypic traits (Gagnaire, Normandeau, Pavey, & Bernatchez, 2013; Gagnaire, Pavey, et al., 2013).

In addition to trans-continental parallelism in interspecific divergence of transcript abundance, we found parallel differentiation between limnetic and benthic species also at the gene sequence level. Indeed, from identified outliers between replicate species-pairs, 15% were defined as parallel outliers (i.e., shared among all species comparisons), consistent with what has been found in other systems (ranging from 6% to 28%; Deagle et al., 2012; Le Moan, Gagnaire, & Bonhomme, 2016; Meier, Marques, Wagner, Excoffier, & Seehausen, 2018; Ravinet et al., 2015; Westram et al., 2014). We also identified 1,251 genes exhibiting shared polymorphism across the intercontinental complex of *Coregonus* lineages. Similar patterns were observed in the tree species *Populus tremula*, in which genes in the networks' peripheries (i.e., with lower level of connectivity) were more likely to be enriched in polymorphism due to reduced evolutionary constraints (Mähler et al., 2017). For these genes, patterns of genetic diversity and DEGs enrichment between limnetic and benthic species suggest the action of divergent selection in the presence of gene flow (Charlesworth, Nordborg, & Charlesworth, 1997) globally maintaining alleles associated with different expression levels between sympatric species. The fact that absolute genetic divergence between sympatric limnetic and benthic species in genes with trans-continental shared polymorphism was elevated compared to other genes may reflect the active maintenance of these alleles over the long term (i.e., since the lineage divergence). Such long-term maintenance could result from the interaction of divergent selection and admixture between sympatric species-pairs (Han et al., 2017; Ma et al., 2017). The action of gene flow between sympatric limnetic and benthic whitefish may indeed further contribute to maintain the alleles favoured in each species in a balanced state within each lake, thus protecting polymorphism from being lost at a global scale even across continents. This could be eased by the apparently highly polygenic nature of the traits under divergent selection, meaning that the intensity of selection acting on each underlying locus could be weak (Le Corre & Kremer, 2012). Therefore, even a modest amount of gene flow could possibly maintain a balanced polymorphism within each species-pair. This, however, remains to be investigated more formally, since strong divergent selection without migration within lakes would also help the maintenance of polymorphism across continents. Moreover, such genomic patterns could be directly caused by the sorting of sieved ancestral balanced polymorphism, which genomic signatures would mimic patterns built by other evolutionary process (e.g., gene flow, as discussed above; Guerrero & Hahn, 2017).

The identification of orthologous genes with trans-continental polymorphism associated with differential expression between benthic and limnetic species supports the existence of *cis*-acting SNPs on transcripts abundance. Moreover, the characterization of DEGs enriched in shared polymorphism across continents suggests the long-term action of some form of balancing selection, maintaining ancestral polymorphisms that predate the onset of regional and continental divergence of the different limnetic–benthic species-pairs. Consistent with theory and empirical studies (Zheng, Gianoulis, Karczewski, Zhao, & Snyder, 2011), our analysis of orthologous genes supports a role of polymorphism originating from standing genetic variation both in protein coding sequences (CDS) and in regulatory motives (e.g., untranslated regions UTRs) in the process of adaptive divergence between limnetic–benthic whitefish sister species (Zheng et al., 2011). For instance, we found a parallel *cis*-eQTL in the 3′UTR of the pyruvate kinase gene (*PKM*), affecting the relative expression level of this gene between species. The *PKM* isoform M2 corresponds to a glycolytic enzyme (isoenzyme) that is expressed in liver tissue. Given the importance of 3′UTRs in regulating the transcription process and transcripts abundance (Merritt, Rasoloson, Ko, & Seydoux, 2008; Wittkopp & Kalay, 2011), this 3′UTR SNP could be under divergent selection between limnetic and benthic species and therefore protected from being lost by drift within populations over the long term within any given limnetic–benthic pair as hypothesized above. Consequently, it is likely that such a *cis*-eQTL could have been recruited from standing genetic variation by natural selection, increasing in frequency in limnetic whitefish on both continents, while modifying the level of expression of a central gene in energetic metabolism.

The inferred module of gene co-expression from liver tissue allowed quantifying a partial view of the gene co-expression network associated with species phenotypic differentiation. Interactions between nearby genes within the module could result in *cis*- regulation affecting the level of expression of other genes and ultimately, affect the activity of genes farther in the genome (Boyle, Li, & Pritchard, 2017). It is noteworthy that genes interactions between metabolic pathways are conserved among sister lineages. Indeed, it has been shown that co-expression modules are maintained through evolutionary times despite variation in the set of regulatory genes that activate them (Tanay, Regev, & Shamir, 2005). Moreover, those patterns of gene expression changes, maintained across the system, could be associated with genes involved in local adaptation by which co-evolving traits are integrated into the same module (Wagner, Pavlicev, & Cheverud, 2007). Despite the partial portrait of the polygenic basis of phenotypic differentiation between species, we stress that no gene of main effect (i.e., hub gene) was identified in the module. Such patterns would suggest that modularity in gene expression and genes interaction into a module can recruit less constrained genes without affecting highly constrained central genes (Wagner, 1996), while peripheral genes could be directly and indirectly involved in the co-expression network via a “hub gene” effect in the initial metabolic pathways of the recruited gene. However, further investigations on quantifying the gene–gene interactions or protein–protein interactions (PPIs), in order to infer individual gene constraint levels or position in the module, should be realized in a more formal framework on several tissues. Thus, this could allow identifying a most complete picture (qualitatively and

quantitatively) of the gene co-expression network associated with phenotypic differentiation between limnetic and benthic species.

In conclusion, our study provides a quantitative assessment of DEGs and gene sequence divergence based on an extensive transcriptomic data set, enabling to infer the effects of polygenic divergent selection acting on complex traits that diverge between sympatric benthic and limnetic species, within both the *C. clupeaformis* and *C. lavaretus* species radiations (Gagnaire, Pavey, et al., 2013; Laporte et al., 2015). Our results also extend previous findings by revealing patterns of parallelism between species on two continents, derived from two evolutionary lineages that diverged at least half a million years ago. Furthermore, they show the effects of polygenic selection on genes associated with fundamental and constrained metabolic pathways, such as functions associated with energetic metabolism (Dalziel, Laporte, Guderley, et al., 2017). Due to the additive effects of multiple genes in controlling the expression of polygenic phenotypic traits, the probability of identifying a shared genetic basis from standing genetic variation (likely to increase with the number of genes involved) is higher compared to the alternative de novo mutation to generate local polymorphism. This suggests an important contribution of ancestral polymorphism in the repeated evolution of sympatric species-pairs. This was illustrated by the identification of a genetic variant in the UTR gene region associated with phenotypic differences between species, as previously reported in other taxa (Jones et al., 2012; Schluter, Clifford, Nemethy, & McKinnon, 2004; Uebbing et al., 2016; Verta, Landry, & MacKay, 2016; Wang, Uebbing, & Ellegren, 2017; Wittkopp, Haerum, & Clark, 2004). The resolution of future studies could be enhanced using a comparative whole-genome resequencing approach to provide a more detailed understanding of the genomic architecture of phenotypic differences between species, and the role of old standing variants in ecological speciation.

ACKNOWLEDGEMENTS

We thank Jérémy Le Luyer and Ben J. G. Sutherland for providing valuable advice and discussions on transcriptomic analysis, and Martin Laporte for inspiring and stimulating discussions about the lake whitefish system. We thank Alexandre Gouy for support with *signet* in the early times of the package, as well as Kyle W. Wellband for commenting on an earlier version of the manuscript. We thank Shripathi Bhat and the Freshwater ecology group at UiT for participating with the sampling in Norway. Finally, we are grateful to AE Shawn Narum and two anonymous referees for their constructive and helpful comments on an earlier version of the manuscript.

CONFLICT OF INTERESTS

There are no competing interests.

AUTHOR CONTRIBUTIONS

C.R. and L.B. designed the project. K.P. and O.S. shared samples from Norway and Switzerland, respectively. C.R. produced and analysed the data. C.R. drafted the manuscript and all authors contributed to the writing and approved the final draft of the manuscript.

DATA AVAILABILITY STATEMENT

Raw sequence data are available through the NCBI sequence read archive (SRA) database under accession SRP136771. Scripts for analysis are available under the Clément Rougeux github repository (https://github.com/crougeux).

ORCID

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

REFERENCES

Amundsen, P.-A., Bøhn, T., & Våga, G. H. (2004). Gill raker morphology and feeding ecology of two sympatric morphs of European whitefish (*Coregonus lavaretus*). *Annales Zoologici Fennici*, *41*, 291–300.

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, *23*, 38–44. https://doi.org/10.1016/j.tree.2007.09.008

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *1*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bernatchez, L., Chouinard, A., & Lu, G. (1999). Integrating molecular genetics and ecology in studies of adaptive radiation: Whitefish, *Coregonus* sp., as a case study. *Biological Journal of the Linnean Society*, *68*, 173–194. https://doi.org/10.1111/j.1095-8312.1999.tb01165.x

Bernatchez, L., & Dodson, J. J. (1990). Allopatric origin of sympatric populations of lake whitefish (*Coregonus clupeaformis*) as revealed by mitochondrial-DNA restriction analysis. *Evolution*, *24*, 890–908.

Bernatchez, L., & Dodson, J. J. (1991). Phylogeographic structure in mitochondrial DNA of the lake whitefish (*Coregonus clupeaformis*) and its relation to Pleistocene glaciations. *Evolution*, *45*, 1016–1035.

Bernatchez, L., & Dodson, J. J. (1994). Phylogenetic relationships among Palearctic and Nearctic whitefish (*Coregonus* sp.) populations as revealed by mitochondrial DNA variation. *Canadian Journal of Fisheries and Aquaculture*, *51*, 240–251.

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

Bolger, A. M., Lohse, M., & Usadel, B. (2014). TRIMMOMATIC: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: From polygenic to omnigenic. *Cell*, *169*, 1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

Capblancq, T., Luu, K., Blum, M. G. B., & Bazin, É. (2018). Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources*, *18*, 1223–1233. https://doi.org/10.1111/1755-0998.12906

Carruthers, M., Yurchenko, A. A., Augley, J. J., Adams, C. E., Herzyk, P., & Elmer, K. R. (2018). De novo transcriptome assembly, annotation and comparison of four ecological and evolutionary model salmonid fish species. *BMC Genomics*, *19*, 1–17.

Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, *70*, 155–174.

Cohan, F. M., & Hoffmann, A. A. (1989). Uniform selection as a diversifying force in evolution: Evidence from Drosophila. *The American Naturalist*, *134*, 613–637. https://doi.org/10.1086/285000

Conte, G. L., Arnegard, M. E., Peichel, C. L., & Schluter, D. (2012). The probability of genetic parallelism and convergence in natural populations. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 5039–5047. https://doi.org/10.1098/rspb.2012.2146

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*, 3133–3157. https://doi.org/10.1111/mec.12796

Dalziel, A. C., Laporte, M., Guderley, H., & Bernatchez, L. (2017). Do differences in the activities of carbohydrate metabolism enzymes between Lake Whitefish ecotypes match predictions from transcriptomic studies? *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *224*, 138–149.

Dalziel, A. C., Laporte, M., Rougeux, C., Guderley, H., & Bernatchez, L. (2017). Convergence in organ size but not energy metabolism enzyme activities among wild Lake Whitefish (*Coregonus clupeaformis*) species pairs. *Molecular Ecology*, *26*, 225–244.

Dalziel, A. C., Martin, N., Laporte, M., Guderley, H., & Bernatchez, L. (2015). Adaptation and acclimation of aerobic exercise physiology in Lake Whitefish ecotypes (*Coregonus clupeaformis*). *Evolution*, *69*, 2167–2186.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Deagle, B. E., Jones, F. C., Chan, Y. F., Absher, D. M., Kingsley, D. M., & Reimchen, T. E. (2012). Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 1277–1286. https://doi.org/10.1098/rspb.2011.1552

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.

Duforet-Frebourg, N., Luu, K., Laval, G., Bazin, É., & Blum, M. G. B. (2016). Detecting genomic signatures of natural selection with principal component analysis: Application to the 1000 genomes data. *Molecular Biology and Evolution*, *33*, 1082–1093. https://doi.org/10.1093/molbev/msv334

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

Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Molecular Ecology*, *27*, 2215–2233. https://doi.org/10.1111/mec.14584

Gagnaire, P.-A., & Gaggiotti, O. E. (2016). Detecting polygenic selection in marine populations by combining population genomics and quantitative genetics approaches. *Current Zoology*, *62*, 603–616. https://doi.org/10.1093/cz/zow088

Gagnaire, P.-A., Normandeau, E., Pavey, S. A., & Bernatchez, L. (2013). Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the Lake Whitefish (*Coregonus clupeaformis*). *Molecular Ecology*, *22*, 3036–3048.

Gagnaire, P.-A., Pavey, S. A., Normandeau, E., & Bernatchez, L. (2013). The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution*, *67*, 2483–2497. https://doi.org/10.1111/evo.12075

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

Gouy, A., Daub, J. T., & Excoffier, L. (2017). Detecting gene subnetworks under selection in biological pathways. *Nucleic Acids Research*, *45*(16), e149. https://doi.org/10.1093/nar/gkx626

Guerrero, R. F., & Hahn, M. W. (2017). Speciation as a sieve for ancestral polymorphism. *Molecular Ecology*, *26*, 5362–5368. https://doi.org/10.1111/mec.14290

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*, 1494–1512. https://doi.org/10.1038/nprot.2013.084

Häkli, K., Ostbye, K., Kahilainen, K. K., Amundsen, P.-A., & Praebel, K. (2018). Diversifying selection drives parallel evolution of gill raker number and body size along the speciation continuum of European whitefish. *Ecology and Evolution*, *8*, 2617–2631. https://doi.org/10.1002/ece3.3876

Han, F., Lamichhaney, S., Grant, B. R., Grant, P. R., Andersson, L., & Webster, M. T. (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Research*, *27*, 1004–1015. https://doi.org/10.1101/gr.212522.116

Harrison, P. W., Wright, A. E., & Mank, J. E. (2012). The evolution of gene expression and the transcriptome–phenotype relationship. *Seminars in Cell and Developmental Biology*, *23*, 222–229. https://doi.org/10.1016/j.semcdb.2011.12.004

Jacobs, A., Carruthers, M., Yurchenko, A., Gordeeva, N., Alekseyev, S., Hooker, O., … Elmer, K. R. (2018). Convergence in form and function overcomes non-parallel evolutionary histories in Arctic Charr. *bioRxiv*, 1–84. https://doi.org/10.1101/265272

Jacobsen, M. W., Hansen, M. M., Orlando, L., Bekkevold, D., Bernatchez, L., Willerslev, E., & Gilbert, M. T. (2012). Mitogenome sequencing reveals shallow evolutionary histories and recent divergence time between morphologically and ecologically distinct European whitefish (*Coregonus* spp.). *Molecular Ecology*, *21*, 2727–2742.

Jari Oksanen, F., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner, H. (2018). VEGAN: *Community ecology package*. R package version 2.5-3. Retrieved from https://CRAN.R-project.org/package=vegan

Jeukens, J., & Bernatchez, L. (2011). Regulatory versus coding signatures of natural selection in a candidate gene involved in the adaptive divergence of whitefish species pairs (*Coregonus* spp.). *Ecology and Evolution*, *2*, 258–271.

Jeukens, J., Bittner, D., Knudsen, R., & Bernatchez, L. (2009). Candidate genes and adaptive radiation: Insights from transcriptional adaptation to the limnetic niche among coregonine fishes (*Coregonus* spp., Salmonidae). *Molecular Biology and Evolution*, *26*, 155–166. https://doi.org/10.1093/molbev/msn235

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., … Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*, 55–61. https://doi.org/10.1038/nature10944

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

Klopfenstein, D. V., Zhang, L., Pedersen, B. S., Ramírez, F., Vesztrocy, A. W., Naldi, A., … Dampier, W. (2018). GOATOOLS: A Python library for Gene Ontology analyses. *Scientific Reports*, *8*, 10872. https://doi.org/10.1038/s41598-018-28948-z

Kottelat, M., & Freyhof, J. (2007). Handbook of European freshwater fishes. *Copeia*, *2008*(3), 725–727. https://doi.org/10.1643/OT-08-098a.1

Landry, L., Vincent, W. F., & Bernatchez, L. (2007). Parallel evolution of lake whitefish dwarf ecotypes in association with limnological features of their adaptive landscape. *Journal of Evolutionary Biology*, *20*, 971–984. https://doi.org/10.1111/j.1420-9101.2007.01304.x

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with BOWTIE 2. *Nature Methods*, *9*, 357–359. https://doi.org/10.1038/nmeth.1923

Laporte, M., Dalziel, A. C., Martin, N., & Bernatchez, L. (2016). Adaptation and acclimation of traits associated with swimming capacity in Lake Whitefish (*Coregonus clupeaformis*) ecotypes. *BMC Evolutionary Biology*, *16*, 1–13. https://doi.org/10.1186/s12862-016-0732-y

Laporte, M., Rogers, S. M., Dion-Côté, A.-M., Normandeau, E., Gagnaire, P. A., Dalziel, A. C., … Bernatchez, L. (2015). RAD-QTL mapping reveals both genome-level parallelism and different genetic architecture underlying the evolution of body shape in Lake Whitefish (*Coregonus clupeaformis*) species pairs. *G3: Genes*, *Genomes*, *Genetics*, *5*, 1481–1491.

Le Corre, V., & Kremer, A. (2012). The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, *21*, 1548–1566. https://doi.org/10.1111/j.1365-294X.2012.05479.x

Le Moan, A., Gagnaire, P. A., & Bonhomme, F. (2016). Parallel genetic divergence among coastal-marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. *Molecular Ecology*, *25*, 3187–3202. https://doi.org/10.1111/mec.13627

Legendre, L., & Legendre, L. (1998). *Numerical ecology*. Amsterdam, The Netherlands: Elsevier.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

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

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*, 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, G., & Bernatchez, L. (1999). Correlated trophic specialization and genetic divergence in sympatric lake whitefish ecotypes (*Coregonus clupeaformis*): Support for the ecological speciation hypothesis. *Evolution*, *53*, 1491–1505.

Luu, K., Bazin, É., & Blum, M. G. B. (2017). PCADAPT: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*, 67–77.

Ma, T., Wang, K., Hu, Q., Xi, Z., Wan, D., Wang, Q., … Liu, J. (2017). Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. *Proceedings of the National Academy of Sciences of the United States of America*, *5*, 201713288–201713298.

Magoc, T., & Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*, 2957–2963. https://doi.org/10.1093/bioinformatics/btr507

Mähler, N., Wang, J., Terebieniec, B. K., Ingvarsson, P. K., Street, N. R., & Hvidsten, T. R. (2017). Gene co-expression network connectivity is an important determinant of selective constraint. *PLOS Genetics*, *13*, e1006402–e1006433. https://doi.org/10.1371/journal.pgen.1006402

Manceau, M., Domingues, V. S., Linnen, C. R., Rosenblum, E. B., & Hoekstra, H. E. (2010). Convergence in pigmentation at multiple levels: Mutations, genes and function. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *365*, 2439–2450. https://doi.org/10.1098/rstb.2010.0104

Manceau, M., Domingues, V. S., Mallarino, R., & Hoekstra, H. E. (2011). The developmental role of Agouti in color pattern evolution. *Science*, *331*, 1062–1065. https://doi.org/10.1126/science.1200684

Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A combinatorial view on speciation and adaptive radiation. *Trends in Ecology & Evolution*, *34*(6), 531–544.

Martin, C. H., Cutler, J. S., Friel, J. P., Dening Touokong, C., Coop, G., & Wainwright, P. C. (2015). Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. *Evolution*, *69*, 1406–1422. https://doi.org/10.1111/evo.12674

Meier, J. I., Marques, D. A., Wagner, C. E., Excoffier, L., & Seehausen, O. (2018). Genomics of parallel ecological speciation in Lake Victoria cichlids. *Molecular Biology and Evolution*, *35*(6), 1489–1506.

Meier, J. I., Sousa, V. C., Marques, D. A., Selz, O. M., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Demographic modelling with whole-genome data reveals parallel origin of similar *Pundamilia* cichlid species after hybridization. *Molecular Ecology*, *26*, 123–141.

Merritt, C., Rasoloson, D., Ko, D., & Seydoux, G. (2008). 3′ UTRs are the primary regulators of gene expression in the *C. elegans* germline. *Current Biology*, *18*, 1476–1482. https://doi.org/10.1016/j.cub.2008.08.013

Nei, M. (1987). *Molecular evolutionary genetics*. New York, NY: Columbia University Press.

Nelson, T. C., & Cresko, W. A. (2018). Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evolution Letters*, *114*, 7061–7113.

Orr, H. A. (2005). The probability of parallel evolution. *Evolution*, *59*, 216–220. https://doi.org/10.1111/j.0014-3820.2005.tb00907.x

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. https://doi.org/10.1111/j.1365-294X.2005.02737.x

Pasquier, J., Cabau, C., Nguyen, T., Jouanno, E., Severac, D., Braasch, I., … Bobe, J. (2016). Gene evolution and gene expression after whole genome duplication in fish: The PhyloFish database. *BMC Genomics*, *17*, 368. https://doi.org/10.1186/s12864-016-2709-z

Pavey, S. A., Collin, H., Nosil, P., & Rogers, S. M. (2010). The role of gene expression in ecological speciation. *Annals of the New York Academy of Sciences*, *1206*, 110–129. https://doi.org/10.1111/j.1749-6632.2010.05765.x

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genetics*, *8*, e1002967.

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.

Ravinet, M., Westram, A., Johannesson, K., Butlin, R., André, C., & Panova, M. (2015). Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Molecular Ecology*, *25*, 287–305.

Renaut, S., Maillet, N., Normandeau, E., Sauvage, C., Derome, N., Rogers, S. M., & Bernatchez, L. (2012). Genome-wide patterns of divergence during speciation: The lake whitefish case study. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 354–363. https://doi.org/10.1098/rstb.2011.0197

Roberts, A., & Pachter, L. (2012). Streaming fragment assignment for real-time analysis of sequencing experiments. *Nature Methods*, *10*, 71–73. https://doi.org/10.1038/nmeth.2251

Rockman, M. V. (2012). The QTN program and the alleles that matter for evolution: All that's gold does not glitter. *Evolution*, *66*, 1–17. https://doi.org/10.1111/j.1558-5646.2011.01486.x

Roesti, M., Gavrilets, S., Hendry, A. P., Salzburger, W., & Berner, D. (2014). The genomic signature of parallel adaptation from shared genetic variation. *Molecular Ecology*, *23*, 3944–3956. https://doi.org/10.1111/mec.12720

Rogers, S. M., & Bernatchez, L. (2004). FAST-TRACK: Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (*Coregonus clupeaformis*). *Molecular Ecology*, *14*, 351–361. https://doi.org/10.1111/j.1365-294X.2004.02396.x

Rogers, S., & Bernatchez, L. (2007). The genetic architecture of ecological speciation and the association with signatures of selection in Natural Lake Whitefish (*Coregonus* sp. Salmonidae) species pairs. *Molecular Biology and Evolution*, *24*, 1423–1438. https://doi.org/10.1093/molbev/msm066

Rogers, S. M., Gagnon, V., & Bernatchez, L. (2002). Genetically based phenotype-environment association for swimming behavior in lake whitefish ecotypes (*Coregonus clupeaformis* Mitchill). *Evolution*, *56*, 2322–2329. https://doi.org/10.1111/j.0014-3820.2002.tb00155.x

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., & Bernatchez, L. (2019). Model-based demographic inference of introgression history in European whitefish species pairs'. *Journal of Evolutionary Biology*, *32*(8), 806–817. https://doi.org/10.1111/jeb.13482

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, *14*, 807–820. https://doi.org/10.1038/nrg3522

Schluter, D., Clifford, E. A., Nemethy, M., & McKinnon, J. S. (2004). Parallel evolution and inheritance of quantitative traits. *The American Naturalist*, *163*, 809–822. https://doi.org/10.1086/383621

Schrider, D. R., & Kern, A. D. (2017). Soft Sweeps are the dominant mode of adaptation in the human genome. *Molecular Biology and Evolution*, *34*, 1863–1877. https://doi.org/10.1093/molbev/msx154

Shabalin, A. A. (2012). Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. *Bioinformatics*, *28*, 1353–1358. https://doi.org/10.1093/bioinformatics/bts163

St-Cyr, J., Derome, N., & Bernatchez, L. (2008). The transcriptomics of life-history trade-offs in whitefish species pairs (*Coregonus* sp.). *Molecular Ecology*, *17*, 1850–1870.

Tanay, A., Regev, A., & Shamir, R. (2005). Conservation and evolvability in regulatory networks: The evolution of ribosomal regulation in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 7203–7208. https://doi.org/10.1073/pnas.0502521102

Trudel, M., Tremblay, A., Schetagne, R., & Rasmussen, J. B. (2001). Why are dwarf fish so small? An energetic analysis of polymorphism in lake whitefish (*Coregonus clupeaformis*). *Canadian Journal of Fisheries and Aquatic Sciences*, *58*, 394–405.

Uebbing, S., Künstner, A., Mäkinen, H., Backström, N., Bolivar, P., Burri, R., … Ellegren, H. (2016). Divergence in gene expression within and between two closely related flycatcher species. *Molecular Ecology*, *25*, 2015–2028. https://doi.org/10.1111/mec.13596

Verta, J.-P., Landry, C. R., & MacKay, J. (2016). Dissection of expression-quantitative trait locus and allele specificity using a haploid/diploid plant system – Insights into compensatory evolution of transcriptional regulation within populations. *New Phytologist*, *211*, 159–171. https://doi.org/10.1111/nph.13888

Wagner, G. P. (1996). Homologues, natural kinds and the evolution of modularity. *The American Zoologist*, *36*, 36–43. https://doi.org/10.1093/icb/36.1.36

Wagner, G. P., Pavlicev, M., & Cheverud, J. M. (2007). The road to modularity. *Nature Reviews Genetics*, *8*, 921–931. https://doi.org/10.1038/nrg2267

Wang, M., Uebbing, S., & Ellegren, H. (2017). Bayesian inference of allele-specific gene expression indicates abundant cis-regulatory variation in natural flycatcher populations. *Genome Biology and Evolution*, *9*, 1266–1279. https://doi.org/10.1093/gbe/evx080

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

Westram, A. M., Galindo, J., Alm Rosenblad, M., Grahame, J. W., Panova, M., & Butlin, R. K. (2014). Do the same genes underlie parallel phenotypic divergence in different *Littorina saxatilis* populations? *Molecular Ecology*, *23*, 4603–4616.

Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2004). Evolutionary changes in cis and trans gene regulation. *Nature*, *430*, 85–88. https://doi.org/10.1038/nature02698

Wittkopp, P. J., & Kalay, G. (2011). Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, *13*, 59–69.

Yeaman, S. (2015). Local adaptation by alleles of small effect. *The American Naturalist*, *186*(S1), S74–S89. https://doi.org/10.1086/682405

Zheng, W., Gianoulis, T. A., Karczewski, K. J., Zhao, H., & Snyder, M. (2011). Regulatory variation within and between species. *Annual Review of Genomics and Human Genetics*, *12*, 327–346. https://doi.org/10.1146/annurev-genom-082908-150139

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, Praebel K, Seehausen O, Bernatchez L. Polygenic selection drives the evolution of convergent transcriptomic landscapes across continents within a Nearctic sister species complex. *Mol Ecol*. 2019;00:1–16. https://doi.org/10.1111/mec.15226