Received: 18 December 2020 | Revised: 16 October 2021 | Accepted: 18 October 2021

DOI: 10.1111/mec.16252

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Whole-genome resequencing confirms reproductive isolation between sympatric demes of brown trout (*Salmo trutta*) detected with allozymes

Atal Saha[1] | Anastasia Andersson[1] | Sara Kurland[1] | Naomi L. P. Keehnen[1] | Verena E. Kutschera[2] | Ola Hössjer[3] | Diana Ekman[2] | Sten Karlsson[4] | Marty Kardos[5,6] | Gunnar Ståhl[7] | Fred W. Allendorf[8] | Nils Ryman[1] | Linda Laikre[1]

[1]Division of Population Genetics, Department of Zoology, Stockholm University, Stockholm, Sweden

[2]Department of Biochemistry and Biophysics, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Stockholm University, Solna, Sweden

[3]Department of Mathematics, Stockholm University, Stockholm, Sweden

[4]Norwegian Institute for Nature Research (NINA), Trondheim, Norway

[5]Flathead Lake Biological Station, University of Montana, Montana, USA

[6]National Marine Fisheries Service, Northwest Fisheries Science Center, National Oceanic and Atmospheric Administration, Seattle, Washington, USA

[7]Biolab, Skärholmen, Sweden

[8]Division of Biological Sciences, University of Montana, Missoula, Montana, USA

Correspondence
Linda Laikre and Atal Saha, Division of Population Genetics, Department of Zoology, Stockholm University, Stockholm, Sweden.
Emails: linda.laikre@popgen.su.se; atal.saha@zoologi.su.se

Funding information
Svenska Forskningsrådet Formas, Grant/Award Number: FR-2020/0008 and FR-2016/0005; Vetenskapsrådet, Grant/Award Number: 2019-05503; Carl Tryggers Stiftelse för Vetenskaplig Forskning; Erik Philip-Sörensen Foundation

Abstract

The sympatric existence of genetically distinguishable populations of the same species remains a puzzle in ecology. Coexisting salmonid fish populations are known from over 100 freshwater lakes. Most studies of sympatric populations have used limited numbers of genetic markers making it unclear if genetic divergence involves certain parts of the genome. We returned to the first reported case of salmonid sympatry, initially detected through contrasting homozygosity at a single allozyme locus (coding for lactate dehydrogenase A) in brown trout in the small Lakes Bunnersjöarna, Sweden. First, we verified the existence of the two coexisting demes using a 96-SNP fluidigm array. We then applied whole-genome resequencing of pooled DNA to explore genome-wide diversity within and between these demes; nucleotide diversity was higher in deme I than in deme II. Strong genetic divergence is observed with genome-wide $F_{ST} \approx 0.2$. Compared with data from populations of similar small lakes, this divergence is of similar magnitude as that between reproductively isolated populations. Individual whole-genome resequencing of two individuals per deme suggests higher inbreeding in deme II versus deme I, indicating different degree of isolation. We located two gene-copies for LDH-A and found divergence between demes in a regulatory section of one of these genes. However, we did not find a perfect fit between the sequence data and previous allozyme results, and this will require further research. Our data demonstrates genome-wide divergence governed mostly by genetic drift but also by diversifying selection in coexisting populations. This type of hidden biodiversity needs consideration in conservation management.

KEYWORDS
coexisting populations, conservation genetics, conservation genomics, hidden biodiversity, population genetic structure, salmonid

1 | INTRODUCTION

Populations of the same species that co-exist spatially over at least a part of their life-cycle (Futuyma & Mayer, 1980; Mallet et al., 2009) have interested evolutionary ecologists for decades since they may represent the first steps of speciation (Maynard Smith, 1966; Via, 2001). Reproductive isolation between sympatric populations may arise from adaptations to ecological niches, even in the absence of migration barriers (Kawecki, 1996, 1997; Turelli et al., 2001). In biodiversity research and conservation management sympatric populations represent genetic diversity below the species level that is important to identify and monitor. Such populations contribute to the portfolio effect in ecosystem stability (Schindler et al., 2010, 2015) and to genetic diversity recognized in international conservation policy, for example, the Convention on Biological Diversity (www.cbd.int).

Sympatric populations have been documented in a wide range of taxa from insects to large mammals, in both terrestrial and aquatic environments, as well as in plants (Attard et al., 2016; Guo et al., 2018; Knutsen et al., 2018; Orlov et al., 2012; Ravinet et al., 2016; Schönswetter et al., 2007; Verspoor et al., 2018). Theoretically, they can represent a continuum of genomic divergence dependent on their evolutionary history with respect to degree of isolation over time (Roux et al., 2016). Empirically, different degrees of genetic divergence between sympatric populations have been reported, indicating different evolutionary backgrounds and degree of isolation (Lu & Bernatchez, 1999; Taylor, 1999).

In the vast majority of cases, sympatric populations have been detected because the populations differ phenotypically (Jorde et al., 2018). Sympatric populations can be referred to as "cryptic" when no obvious morphological divergence has been detected between the populations (Bickford et al., 2007), and where their detection has been based exclusively on genetic data (Andersson et al., 2017). The first case of cryptic sympatry in salmonids was reported for brown trout (*Salmo trutta*) in 1976 in the small twin mountain Lakes Bunnersjöarna in central Sweden where contrasting homozygosity at an allozyme locus (a lactate dehydrogenase locus denoted *LDH-1*) indicated the existence of two coexisting, genetically distinct groupings (denoted demes; Allendorf et al., 1976; Ryman et al., 1979). Further, the allozymes indicated greater amounts of genetic variation in deme I than in deme II. Statistically significant body size differences between the two populations were detected (deme II fish smaller than those in deme I) but it was not possible to classify fish to deme based on visual inspection (Ryman et al., 1979), thus the Lakes Bunnersjöarna sympatry can be referred to as cryptic. Since this early finding over 130 cases of sympatric populations have been identified world-wide in salmonid fishes and less than 10 of those cases were cryptic (Jorde et al., 2018). Most of these studies used few genetic markers and it remains unclear if genetic divergence in sympatry evolves primarily through reproductive isolation and genetic drift, or if divergent selection acting on a restricted part of the genome is the primary evolutionary mechanism for such structures.

In the present study, we reanalyse samples from Lakes Bunnersjöarna and apply whole-genome resequencing and single nucleotide polymorphism (SNP) array data to test if (i) the existence of two reproductively isolated demes in these tiny lakes is supported, and if so, (ii) what the genome-wide divergence between the demes is, and (iii) whether the previously observed differences in amount of genetic variation in a few allozyme loci is a genome-wide phenomenon or limited to a small number of loci.

2 | MATERIALS AND METHODS

2.1 | Samples

Lakes Bunnersjöarna are closely connected oligotrophic twin lakes (total area of 0.67 km^2) located at an elevation of 955 m near the Norwegian border in the County of Jämtland, Sweden (Figure 1). Both lakes are shallow, the southern lake only 0.5 m with a deeper middle (a few metres) and the northern lake c. 2 m deep. The brown trout in these lakes were sampled in 1975 as part of some of the first population genetic screenings of natural populations (Allendorf et al., 1976; Ryman et al., 1979). Material from that collection has been stored in a frozen tissue bank at the Department of Zoology, Stockholm University, Sweden. Here, we used 140 samples that were still available (out of 151 reported in Allendorf et al., 1976; Ryman et al., 1979); 62 and 78 from the northern and southern lake, respectively.

The allozyme studies showed contrasting homozygosity at one locus coding for lactate dehydrogenase (locus *LDH-1*). About half of the fish were homozygous for the 100 allele most common in brown trout in the study area, and the others homozygous for a rare null allele with no active enzyme product (Allendorf et al., 1984).

We classified the 140 fish into deme I or II based on the *LDH-1* genotype, resulting in 68 individuals from deme I (100/100 homozygous for *LDH-1*) and 72 individuals from deme II (homozygous for the null allele). The 68 deme I fish were from both lakes (northern: $n = 39$, southern: $n = 29$), as were the 72 deme II fish (northern: $n = 23$, southern: $n = 49$).

2.2 | Genotyping and sequencing

Genomic DNA was extracted from c. 50 mg muscle tissue from each of the 140 individuals using the KingFisher cell and tissue DNA kit (ThermoScientific) according to the manufacturer's instructions and normalised to 30–50 ng/μl.

We genotyped all 140 fish using an EP1 96.96 dynamic array IFCs genotyping platform (Fluidigm) comprising 96 SNPs shown to be variable in Danish brown trout (Bekkevold et al., 2020; their table S8). Using TBLASTN (E. values <0.0001 & bitscore 80; Altschul et al., 1990) we identified the location of these 96 SNPs on the brown trout reference genome (GenBank accession number GCA_901001165.1;

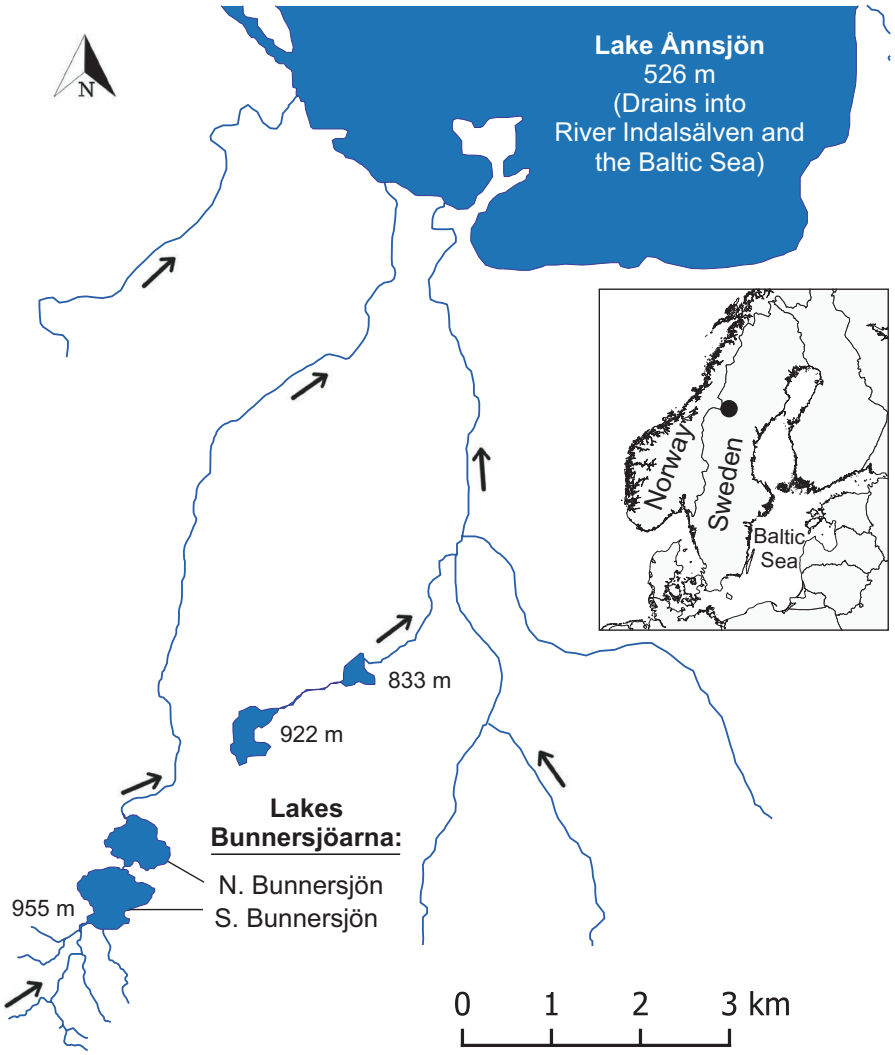

FIGURE 1 Location of Lakes Bunnersjöarna in the mountainous area of central Sweden. The lakes are found in the uppermost part of the water system (elevation indicated, *m*, metres). Our samples were collected in the northern Bunnersjön (N. Bunnersjön; $n = 62$) and in the southern Bunnersjön (S. Bunnersjön; $n = 78$)

Figure S1). The results from the SNP analyses supported the existence of the two demes (see Section 3).

Random samples of $n = 50$ individuals per deme were used for Pool-seq; similarly, $n = 2$ individuals per deme were chosen for individual whole-genome sequencing (WGS). DNA extraction for Pool-seq and WGS followed the same extraction protocol as above but with an additional RNase A treatment. For Pool-seq, DNA with high molecular weight from each of 50 individuals per population was pooled at equal concentrations per deme. The National Genomics Infrastructure (NGI) at the Science of Life Laboratory (SciLifeLab), Stockholm, Sweden conducted the construction of PCR-free paired-end libraries followed by Illumina HiSeq 2000 sequencing.

Additional details about DNA extractions, genotyping and sequencing are provided in Appendix S1. To put the diversity and divergence patterns observed in Lakes Bunnersjöarna into perspective, we compared with similar data from the same time period (1970s) from brown trout of other mountain lakes. This data had been generated as described below for other projects that will be reported elsewhere (Andersson et al., unpublished data; Kurland et al., unpublished data). The samples are from eight additional mountain lakes; one of them (Lake Ånnsjön) is connected to the focal Bunnersjöarna twin lakes (Figure 1). The others represent small lakes of comparable size to the Bunnersjöarna lakes but from separate geographic areas without migration possibilities to them (Appendix S2).

2.3 | Population genetic analyses of 96 SNP array

Allele frequencies and deviations from Hardy–Weinberg proportions measured as F_{IS} and their associated significance levels for the 96 SNP fluidigm array were obtained from GENEPOP v4.3 (Raymond, 1995; Rousset, 2008). Holm's (1979) sequential Bonferroni approach was applied to adjust for multiple testing. F_{ST} (Weir & Cockerham, 1984) was estimated using FSTAT v2.9.4 (Goudet, 2003). We also computed Nei's (1973) parametric F_{ST} ($F_{ST} = (H_T - H_S)/H_T$) using GENALEX v6.5 (Peakall & Smouse, 2012) to allow direct comparisons with Pool-seq F_{ST} (cf. below). CHIFISH v5.0 (Ryman & Palm, 2006; available at http://www.zoologi.su.se/~ryman/) was used for F_{ST} significance testing. Details on simulations of F_{ST} distributions and relationships between various F_{ST} measures are provided in Appendix S3 and S4.

We assessed the most likely number of populations (K) using STRUCTURE v2.3.4 (Pritchard et al., 2000) using the default model

allowing population admixture and correlated allele frequencies with a burnin of 250,000 steps and 500,000 Markov chain (MCMC) replicates. The results of 20 runs were analysed using STRUCTURE HARVESTER v0.6.94 (Earl & vonHoldt, 2012) and CLUMPP (Jakobsson & Rosenberg, 2007). The most likely value of K was determined from the mean likelihood value from STRUCTURE, ΔK (Evanno et al., 2005) from STRUCTURE HARVESTER, and from KFINDER v1.0 (Wang, 2019). Appendix S1 includes additional details.

We constructed an individual-based neighbour-joining phylogenetic tree based on Nei's D_A distance estimate (Nei et al., 1983) from the 96 SNP array using POPTREE2 (Takezaki et al., 2009), and MEGAX 10.0.5 (Kumar et al., 2018) with 1000 bootstrap replications.

2.4 | Pool-seq data processing and variant calling

We assessed the quality of the raw sequence reads using FASTQC v0.11.5 (Leggett et al., 2013) and MULTIQC v1.5 (Ewels et al., 2016). BBDuk as implemented in BBTOOLS v38.08 (http://sourceforge.net/projects/bbmap/) was used for adapter and quality trimming. Trimmed reads were mapped against the brown trout reference assembly (comprising 2,371,863,509 bp; GenBank accession number GCA_901001165.1) using the BURROWS–WHEELER ALIGNER v0.7.17 (BWA, using bwa mem algorithm; Li, 2013; Li & Durbin, 2009) and resulting bam files were filtered for paired reads using SAMTOOLS v1.8 (Li et al., 2009). Bam file quality was evaluated with QUALIMAP v2.2.1 (García-Alcalde et al., 2012) and MULTIQC v1.5. SAMtools was applied for variant calling to produce mpileup files. Indels and SNPs 5 bp upstream and downstream of indels were removed using the "identify-genomic-indel-regions.pl" script of POPOOLATION2 v1201 (Kofler, Pandey, et al., 2011). Additional details, including quality filtering parameters, are provided in Appendix S1.

2.4.1 | Population genomic analyses of Pool-seq data

The script "variance-sliding.pl" as implemented in POPOOLATION v1.2.2 (Kofler, Orozco-terWengel, et al., 2011) was used to estimate nucleotide diversity (π; Tajima, 1983), Watterson's theta (Θ; Watterson, 1975) and Tajima's D (T_D; Tajima, 1989) separately for each pool. We subsampled mpileup files per pool to uniform depths without replacement using the "subsample-pileup.pl" script implemented in POPOOLATION v1.2.2 because estimates of π and T_D from Pool-seq data are sensitive to sequencing errors and variation in coverage (Kofler, Orozco-terWengel, et al., 2011). Θ was estimated with the same parameters as used for π and T_D estimates but without subsampling. Further details are provided in Appendix S1.

F_{ST} between pairs of pools was calculated with POPOOLATION2 v1201 using the default approach (Kofler, Orozco-terWengel, et al., 2011) and the "fst-sliding.pl" script for nonoverlapping windows of different sizes. Additionally, F_{ST} was calculated per gene with the same settings but with a larger window size (1,000,000 bp) than the length of the largest gene present in the genome. We also used the Karlsson et al. (2007) F_{ST} provided by PoPoolation2 which is expected to return F_{ST} more in line with the approach used for the 96 SNP array (i.e., Weir & Cockerham, 1984; see Appendix S4 for an elaboration on how different F_{ST} estimators relate to each other). A Manhattan plot was created using the R package QQMAN v0.1.4 (Turner, 2014) and confidence intervals (95%) were calculated from observed per window F_{ST} with R v3.6.3 (R Core Team, 2018). Additional details on window sizes, quality filtering and parameter settings are described in Appendix S1.

A maximum likelihood tree was estimated and admixture analyses were performed using TREEMIX v1.12 (Pickrell & Pritchard, 2012) from allele frequencies calculated in POPOOLATION2 v1201. Detailed methods are described in Appendix S2.

2.4.2 | Identifying potential outliers

A modified version of POWSIM (Ryman & Palm, 2006) was used to investigate whether the distribution of F_{ST} values in our Pool-seq data was consistent with the expectation for selectively neutral loci (cf. Lamichhaney et al., 2012). We compared the observed genome-wide distribution of F_{ST} with a simulated "expected" distribution under drift only and defined SNPs with an observed F_{ST} above the largest value of the expected distribution as potential outliers (theory and numerical examples provided in Appendix S3). Further, outliers were also defined from percentile thresholds of the observed genome-wide F_{ST} distribution (Appendix S1).

2.4.3 | Identification of biological function of outlier genes

A gene ontology (GO) set enrichment analysis (GSEA) was performed to associate biological and common gene functions to F_{ST} outliers. Functional annotation of the brown trout reference was performed on the EGGNOG v5.0 web-interface (http://eggnog-mapper.embl.de/; Huerta-Cepas et al., 2018) using brown trout protein FASTA files (GenBank accession GCA_901001165.1). We extracted all genes overlapping with outlier SNPs and windows from the NCBI *S. trutta* annotation release 100. The R package topGO (Alexa & Rahnenfuhrer, 2020) was used to test for overrepresentation of GO biological processes. GO terms with p-values ≤.01 were retained and then filtered for redundancy in REVIGO (Supek et al., 2011). For visualization of results, treemaps were drawn in R v3.6.3.

2.4.4 | Locating *LDH-A* genes in the brown trout reference assembly

The sympatric populations of Lakes Bunnersjöarna were identified with a single allozyme locus, *LDH-1*, that showed contrasting homozygosity in the demes. *LDH-1* is one of two loci (*LDH-1* and

LDH-2) in salmonids coding for LDH-A (Allendorf et al., 1984). We identified the DNA sequences relating to this enzyme using (1) GO term protein annotations from EggNOG and (2) TBLASTN (Altschul et al., 1990) with the FASTA sequence for LDH extracted from the Uniprot database to search the genome assembly. Each *LDH-A* copy was manually inspected in the Integrative Genomics Viewer (IGV) v2.4.2 (Robinson et al., 2011). The amino acid compositions for all LDH genes were extracted using an in-house script (Appendix S5). See Appendix S1 for more details.

We calculated allele frequencies for the *LDH-A* gene copies in both demes from the Pool-seq data using "snp-frequency-diff.pl" from POPOOLATION2 v1201. Variants were called for these genes in individual WGS data using BCFTOOLS call v1.8 (Danecek et al., 2015), inspected using IGV, and visualized with GVIZ (Hahne & Ivanek, 2016) and GGPLOT2 in R v3.6.3 (Wickham, 2016).

2.5 | Individual whole-genome sequencing data processing and variant calling

Sequenced reads from two individuals per deme were aligned against the brown trout reference assembly using BWA mem v0.7.17 (BWA; Li & Durbin, 2009) and sorted using SAMTOOLS v1.8. Resulting bam files were merged per individual using MarkDuplicates as implemented in PICARD v2.10.3 (https://broadinstitute.github.io/picard/), which was also used to mark PCR duplicates. Bam file quality was assessed with QUALIMAP v2.2.1 (García-Alcalde et al., 2012).

Individual genomic variant call format files (gVCFs) were generated with HaplotypeCaller from the Genome Analysis ToolKit (GATK) v3.8 (McKenna et al., 2010), and joint genotyping of all brown trout samples (including fish from other lakes, see Section 2.2) was performed with GATK GenotypeGVCFs. GATK's VariantFiltration tool was used to remove low-quality variants. Details are provided in Appendix S1.

2.5.1 | Estimation of inbreeding

Inbreeding was estimated from individual WGS as the fraction of the genome covered by "runs of homozygosity" (ROH), and their length (LnROH; Gomez-Raya et al., 2015; Kardos et al., 2017; Magi et al., 2014) using the PLINK v1.90b4.9's (Purcell et al., 2007) "—homozyg" method. Details on this analysis are described in Appendix S1.

3 | RESULTS

3.1 | Population divergence and diversity using SNP array data

A total of 77 of the 96 SNPs were polymorphic in Lakes Bunnersjöarna. The average call rate of these loci was 0.989 (range: 0.707–1.0). The average per-individual-call-rate for these loci was 0.989 (range: 0.935–1.0). Without prior grouping of individuals, STRUCTURE suggested $K = 2$ clusters, and these two clusters are almost completely consistent with the *LDH-1* groupings (Figures 2a and S2). An individual-based neighbour-joining tree including 127 individuals with full genotypes in 70 of the 77 polymorphic SNP loci further illustrated the clear partition of the two demes (Figure 2b). We found pronounced differences in the amount of genetic variation within the demes; only 24 of the 77 SNPs loci were variable in deme II and expected heterozygosity is 0.27 for deme I and 0.08 for deme II (Table 1). We observed a strong genetic divergence between the demes with $F_{ST} = 0.24$ (Table 2).

3.2 | Pool-seq data and population genomics

In total, 213 giga base pairs (Gb; 1527 million reads) and 197 Gb (1302 million reads) of raw Pool-seq data were generated for the samples from deme I and II, respectively. After quality filtering of raw reads and mapping, 184 and 134 Gb of data, corresponding to 1366 and 1169 million reads mapped as pairs, remained. Mode of read depth of coverage was 104 and 90, for demes I and II, respectively. Average mapping quality was c. 33 and edit distance between the reads and reference was c. 0.6% (Table S1).

3.2.1 | Genome-wide diversity and divergence

Genome-wide diversity, measured as nucleotide diversity (π) and Watterson's theta (Θ), was considerably larger in deme I than in deme II (Table 3; deme I $\pi = 0.0013$ vs. deme II $\pi = 0.00046$). The average values for Θ were 0.0017 (deme I) vs. 0.0008 (deme II). Tajima's *D* values were all below 0, with deme II having a larger negative estimate (Table 3).

The genome-wide divergence between the two demes was high across the genome; F_{ST}=0.13 using window size of 5 kb (Table 2; Figure 3); this value was largely consistent for other window sizes (Table S2) and agrees with other estimates (Table 2, Appendix S6), and the relationship among F_{ST} from different approaches are in line with expectations (Appendices S4 and S6). The fraction to which windows were covered with data after quality and depth filtering did not affect F_{ST} values ($F_{ST} = 0.13$ for all fraction depths; Table S3). To minimize stochastic errors linked with small window sizes, while not losing too much of data, we chose 5 kb windows with ≥80% fraction depth coverage.

3.2.2 | F_{ST} outliers

The POWSIM results showed that the distributions of the observed and simulated ("expected") F_{ST} values differed from each other (Kolmogorov-Smirnov, $p < .05$). The observed distribution showed a higher frequency of large F_{ST} values than the expected one (Figure 4). The highest value of the expected distribution was $F_{ST} = 0.864$, and we found 194 SNPs (out of 12,177,462) with a larger F_{ST} in the

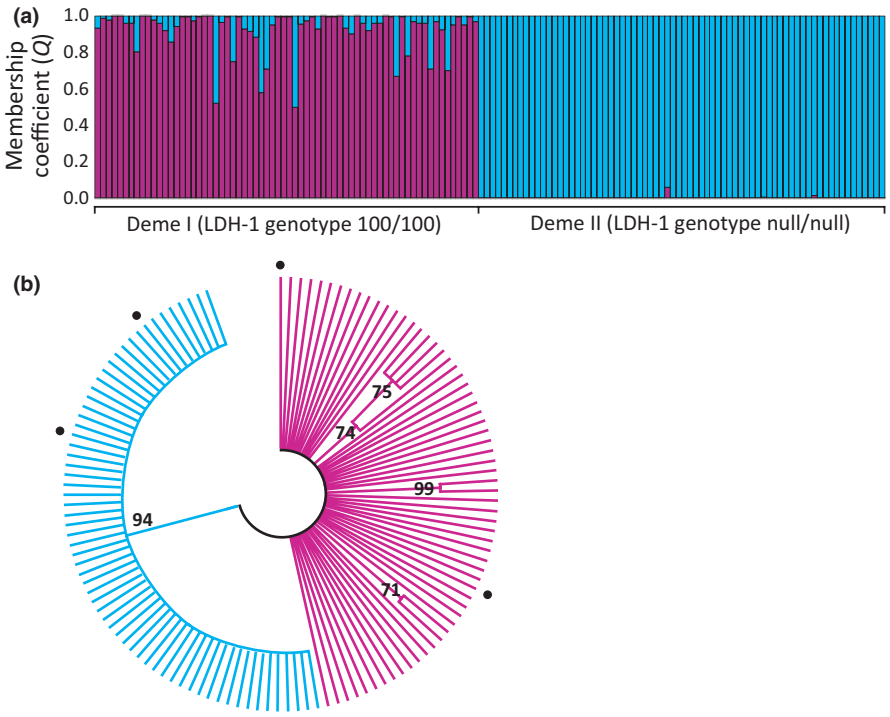

FIGURE 2 Membership coefficient (*Q*) plots showing the assignment probability of individual fish (total $n = 140$) to clusters using 96 SNPs (a), and individual-based neighbour-joining tree based on 70 SNPs (b). Membership coefficients (*Q*) were obtained from STRUCTURE and the most likely number of clusters (*K*) for the data was estimated using ΔK; each fish is represented by a vertical bar. The phylogenetic tree shows individuals classified to deme I and II based on their LDH-1 genotype in purple and blue, respectively. The tree was constructed based on Nei's D_A distance, and has been compressed to include branches with bootstrap values of at least 70%. Numbers along the branches indicate bootstrap values in percentages. The black dots mark the four individuals that were randomly selected for individual whole-genome sequencing

TABLE 1 Measures of genetic diversity for the two sympatric brown trout populations of Lakes Bunnersjöarna using individual genotyping of loci that show genetic variation in the total material from these lakes

Marker	Deme	*n*	N_A	F_{IS}	H_E	Number of polymorphic loci	Number of private alleles
SNPs	I	68	1.80	**−0.168**	0.269	77	53
	II	72	1.25	**0.071**	0.079	24	0
Allozymes[a]	I	217	1.88	**0.096**	0.219	7	3
	II	177	1.63	0.069	0.123	5	1

Note: These polymorphic loci consist of 77 SNPs from the 96 SNP fluidigm assay and eight allozyme loci. All loci are biallelic and each private allele corresponds to a locus that carries an allele that only occurs in that particular deme, while being monomorphic in the other deme. Statistically significant F_{IS} values are in bold ($p < .05$).

Abbreviations: H_E, expected heterozygosity;*N*, number of fish; N_A, mean number of alleles per locus.

[a]From Ryman et al. (1979) and unpublished data.

observed distribution, of which 19 had $F_{ST} = 1$. Of these 194 outliers, one was associated with a gene significantly enriched ($p < .05$) by two GO terms connected with the growth process (GO:0048590): "extracellular matrix and structure organization" (Table S4).

A total of 60,887 SNPs out of 12,177,462 occurred above the 99.5th percentile of the F_{ST} distribution ($F_{ST} \geq 0.53$). 698 of these outlier SNPs were within genes and 432 genes were identified as feasible for topGO analysis. Results showed that these 432 genes were significantly ($p \leq .01$) enriched with 69 GO terms: 21 genes were significantly linked with GO terms associated to "glycosaminoglycan biosynthesis" (Table S4). Seven genes were associated to "gonad morphogenesis" while two were involved in binding of sperm with eggs. The five top GO term superclusters (the ones associated with the smallest *p*-values) were: "chondroitin sulphate metabolism" (i.e., glycosaminoglycan biosynthesis), "response to ozone" (linked with animals' response to stimulus/stress, GO:0050896), "cell-cell adhesion involved in gastrulation" (linked with embryonic morphogenesis, GO:0048598), "fatty acid derivative metabolism" (energy storage), and "reactive oxygen species metabolism" (linked with phagocytosis and signal transduction; Figure S3).

TABLE 2 F_{ST} estimates between demes I and II of Lakes Bunnersjöarna

Data	Number of windows	Number of loci	Mean F_{ST} (95% CI)	Min.; max. F_{ST}
Pool-seq	340,297	8,495,563	0.127[a] (0.127–0.127)	0.001; 0.849
SNPs	–	77	0.242 (0.194–0.292)	0.001; 0.657
Allozymes	–	7[b]	0.091 (0.055–0.113)	0.022; 0.126
	–	8[c]	0.440 (0.069–0.801)	0.022; 1.000

Note: For Pool-seq data Popoolation2 default F_{ST} (Nei, 1973) was used with window size = 5000 bp, read depth = 20–150× and fraction depth covered (the fraction to which windows were covered with data) ≥0.8. For the 96 SNP fluidigm array, 77 polymorphic loci were used and for the allozymes 7 and 8 polymorphic loci were used (excluding and including *LDH-1*, respectively; data from Ryman et al., 1979 and unpublished data). F_{ST} for the SNP array and allozymes use Weir and Cockerham's (1984) approach. See Appendix S4 for a description of how various F_{ST} estimators relate to each other.

[a]Popoolation2 also provide an F_{ST} from an approach by Karlsson et al. (2007) which gives a mean F_{ST} = 0.199 (Appendix S6).

[b]Excluding diagnostic locus *LDH-1* for which F_{ST} = 1.0..

[c]Including diagnostic locus *LDH-1* for which F_{ST} = 1.0.

TABLE 3 Descriptive statistics from genome-wide Pool-seq data (window size = 5000 bp and fraction depth covered ≥0.8 that is, a window was only retained if at least 80% of its SNPs had a read depth between 20x and 150×) for each of the two demes in Lakes Bunnersjöarna

Statistic	Deme	Mean (95% CI)	No. windows	No. SNPs	Subsampling depth
π	I	0.00130 (0.00128; 0.0013)	278,410	6,741,183	52–135×
	II	0.00046 (0.00046; 0.00046)	278,410	2,884,642	52–135×
T_D	I	−0.22665 (−0.2309; −0.2224)	278,410	6,741,183	52–135×
	II	−1.0148 (−1.0194; −1.0102)	278,410	2,884,642	52–135×
Θ	I	0.001670 (0.00167; 0.00168)	354,917	10,480,433	20–150×
	II	0.000789 (0.00079; 0.00079)	346,678	4,627,177	20–150×

Note: π, nucleotide diversity (Tajima, 1983), T_D, Tajima's *D* (Tajima, 1989), Θ, Watterson's theta (Watterson, 1975).

For 5 kb windows, 8508 out of 340,297 windows were considered as outliers with F_{ST}-values above the 97.5th percentile (F_{ST} ≥ 0.35). 2148 outlier windows were within genes. TopGO identified 1494 feasible genes, associated to 113 significant GO terms (p ≤ .01). Seven genes were significantly enriched with the GO terms "glucose catabolic process to pyruvate" and "canonical glycolysis" (Table S4). Four genes were enriched with the GO term "NADH oxidation". Furthermore, six genes were significantly enriched with the GO term "sperm capacitation". Top five GO term superclusters identified by REVIGO were: "phagocytosis", "thrombin-activated receptor signaling pathway", ciliary body morphogenesis", "peptidyl-glutamic acid modification", and "fatty acid derivative biosynthesis" These GO terms are mainly linked with metabolic (GO.0008152) or immunological (GO:0006910) processes (Figure S4). Overall, a limited number of loci/windows were identified as outliers and only a small fraction of identified outliers were linked with functional parts of the genome.

3.3 | Individual whole-genome resequencing

Between 21 and 27 giga base pairs (Gb) of sequencing data, corresponding to over 150 million reads mapped as pairs to the *Salmo trutta* reference genome for each individual from Lakes Bunnersjöarna. This corresponds to approximately 98% of the raw data per individual. An average mapping quality of 32 and depth of coverage c. 10× was observed (Table S5). In total, 21 million variants were called of which c. 20 million variants were retained after hard filtering. After removing indels, while keeping only biallelic SNPs with minor allele frequency (MAF) ≥0.01 assigned to any of the 40 *S. trutta* chromosomes, c. 10 million SNPs were retained and were used for ROH estimation.

3.3.1 | Inbreeding

Both individuals from deme II showed considerably higher inbreeding measured as fraction and length of runs of homozygosity (F_{ROH} and LnROH) than those from deme I (Figure 5). Individual 1 from deme II was the most homozygous one; a total of 950,545 kb involving 1742 ROH segments were estimated as identically homozygous in this individual. This translates into an F_{ROH} of 0.405 (95% CI: 0.405–0.405; Table S6). Individuals from deme II showed some long ROH segments which could indicate recent inbreeding. Individual 1 had 122 segments longer than 1000 kb; eight of those were >2000 kb. Individual 2 had 79 such segments with six >2000 kb.

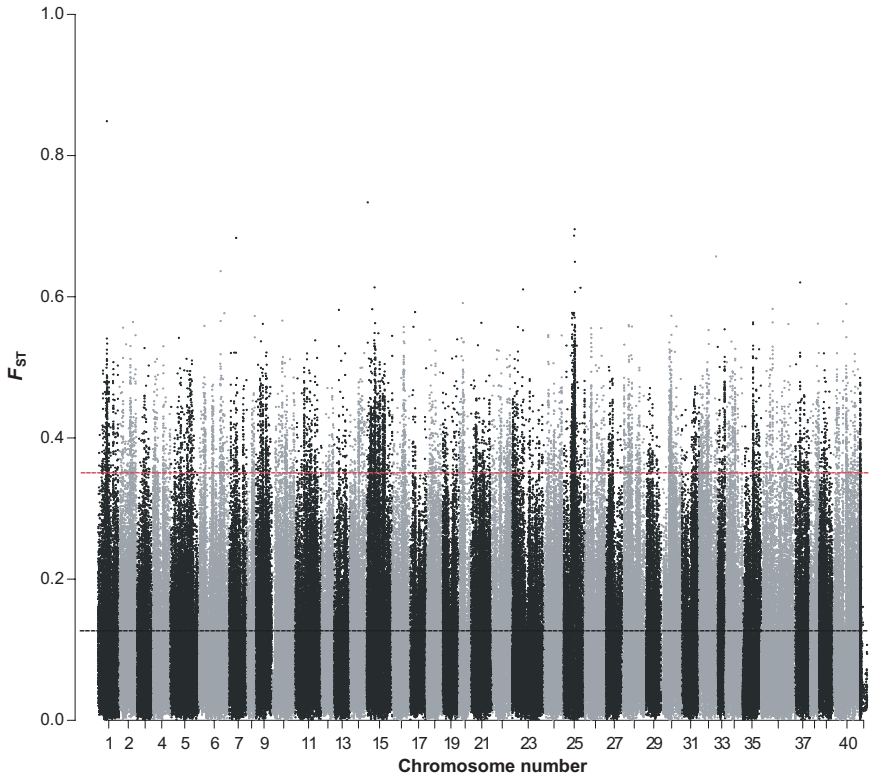

FIGURE 3 Pairwise F_{ST} values between deme I and II of Lakes Bunnersjöarna estimated from whole-genome Pool-seq data using 5 kb windows across 40 brown trout chromosomes. NA (to the right of chromosome 40) = F_{ST} values from scaffolds not possible to assign to a chromosome. The horizontal black dashed line shows the genome-wide mean F_{ST} = 0.13 while the red dashed line marks the 97.5% limiting F_{ST} = 0.35

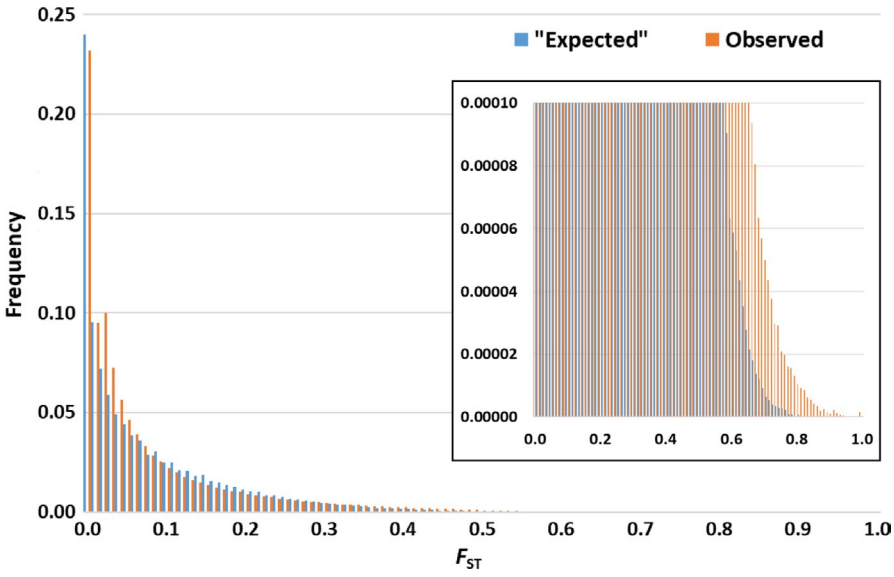

FIGURE 4 Distribution of observed (range: 0 to 1) and simulated ("expected"; range: 0 to 0.864) F_{ST} values based on the POWSIM method (Ryman & Palm, 2006). A total of 12,177,462 SNPs derived from Pool-seq data were analysed and the mean F_{ST} for these was 0.083 (Table S2). The largest simulated F_{ST} was 0.864 and we had 194 observed F_{ST} values above this value. The framed subfigure is a close-up to show the distribution of the extreme values in the right-hand tail (further details on this analysis are provided in Appendix S3)

Deme I individuals only had very few >1000 kb segments; none >2000 kb (Table S6).

3.4 | *LDH-A* genes in the genome

Two copies of the *LDH-A* gene were identified, one on chromosome 7 and one on chromosome 17. Their amino acid composition showed small differences in the number of residues of leucine, serine, valine, isoleucine, as well as in the charged amino acids histidine (positive), and aspartic acid (negative; Table S7). In allozyme studies, protein products from the *LDH-1* locus move closer to the negative (cathodal) pole while products from the *LDH-2* locus move more to the positive (anodal) pole (Allendorf et al., 1984) indicating that *LDH-1* products are more positively charged than those from *LDH-2*. The difference in number of aspartic acid and histidine indicate that the protein product from chromosome 7 is slightly more positively charged and thus expected to be more cathodal in electrophoresis than the product from the chromosome 17 locus. No large divergence between the two demes were detected in either of the gene copies; gene-wide F_{ST} for *LDH-A* from the Pool-seq data was 0.154 and 0.187 for the gene copy on chromosome 7 and chromosome 17, respectively. Furthermore, we found no outlier values for *LDH-A* in any of the F_{ST} analyses of coding regions. However, the 3' UTR

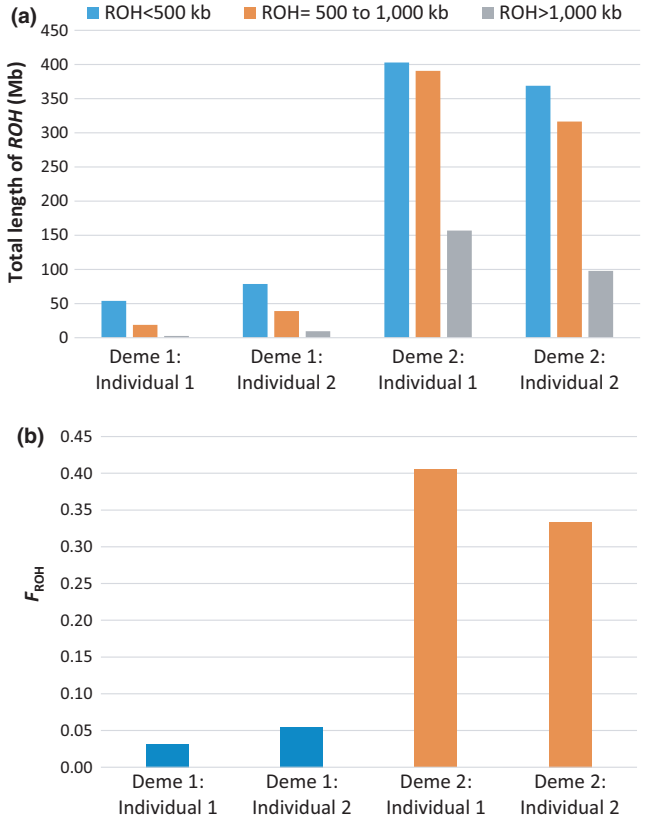

FIGURE 5 (a) Total genomic lengths (measured in number of megabases, MB) of ROH belonging to either of three size categories including small (<500 kb), medium (500 to 1000 kb) and large (>1000 kb) size categories for individuals from two demes. (b) F_{ROH} (fraction of runs of homozygosity, *ROH*, expanded over the genome) using individual whole-genome sequencing data from each of two individuals per deme from Lakes Bunnersjöarna

region of the *LDH-A* on chromosome 17 showed contrasting allele counts between the two pools, and IGV visualization indicated some genetic variation in this region in deme I, whereas deme II was fixed across the whole region (Figure S5, Table S8). In contrast, no divergence of patterns between the demes was observed for *LDH-A* on chromosome 7 (Figure S6).

The 3′ UTR region on chromosome 17 was also analysed using the individual WGS data. In contrast to the results from Pool-seq data the individual sequences did not show any difference between the demes. All four individuals showed lack of variation in the region (Figure S7).

3.5 | Comparisons with other lakes

The genetic divergence between the two demes of Lakes Bunnersjöarna was similar to the lower estimates of divergence between fish from lakes with no migratory contact (Appendix S2). Lake Ånnsjön is the closest lake, 8.5 km downstream of Lakes Bunnersjörna (Figure 1), and the most parsimonious source for the colonization of Lakes Bunnersjörna. F_{ST} for the 96 SNP array between Lake Ånnsjön and deme I (F_{ST} = 0.12) was around half that of the divergence between the two demes (F_{ST} = 0.24) while F_{ST} between Lake Ånnsjön and deme II was 0.36 (Appendix S2). Comparing each of the two demes and Lake Ånnsjön with respect to private alleles using the 96 SNP array data, deme I shows nine and Lake Ånnsjön seven while deme II does not have any private alleles compared to Lake Ånnsjön.

The TreeMix dendrogram from Pool-seq data illustrates how deme II stands out as isolated and less connected to both deme I and Lake Ånnsjön (Appendix S2, c.f. Appendix S1 for methods). Admixture analyses of Pool-seq data from deme I, deme II, Lake Ånnsjön, and a geographically separate lake used as outgroup (Lake Blanktjärnen; Appendix S2) showed ambiguous signals. TreeMix analyses suggested a migration event from deme II to deme I, but a three-population test returned nonsignificant *f*3-values for all possible population combinations, providing no evidence for admixture (Appendix S2, c.f. Appendix S1 for methods).

The nucleotide diversity estimated in brown trout from Bunnersjöarna deme I (π = 0.0013; Table 3) was of the same order of magnitude as estimates from other lakes where π ranged between 0.00104 to 0.00151 (Appendix S2), while π for deme II was almost an order of magnitude lower (Table 3). The two individuals from deme II showed higher average inbreeding levels measured LnROH and F_{ROH} than brown trout in any of the other lakes (Figures S8–S9) and long runs of homozygosity were also more frequent in deme II than in any other lake. In contrast, deme I individuals showed the lowest values (Table S6).

4 | DISCUSSION

4.1 | Evidence for reproductive isolation

The existence of two sympatric demes identified by a single allozyme locus in the small mountain Lakes Bunnersjöarna (Ryman et al., 1979) is supported by the present SNP array data. Deme II appears reproductively isolated from deme I. This conclusion is supported by (i) a strikingly lower level of genetic diversity within deme II as compared to deme I for all variability measures used, including that 53 loci of the 96 SNP array are monomorphic in deme II but polymorphic in deme I, and (ii) the high F_{ST} between the two demes which is of the same order as between populations from reproductively isolated lakes. However, we cannot rule out a small amount of gene flow from deme II to deme I; deme II does not show any private alleles with the SNP array, the STRUCTURE analysis suggested some minor admixture of deme II into the deme I cluster, and the TreeMix analysis provided some support for a migration event from deme II to deme I.

The amount of divergence between the demes is confirmed to be large throughout the genome and it appears high in comparison with observations from other cases of sympatry in salmonids (Jorde et al., 2018). The observed F_{ST} (Table 2), is higher than those reported in other cryptic, sympatric salmonid populations (Adams et al., 2008;

Andersson et al., 2017; Aykanat et al., 2015; Palmé et al., 2013; Wilson et al., 2004).

The high divergence over the entire genome and the fact that our outlier analyses indicate relatively few SNPs/windows to be outlier indicate that most differentiation is caused by genetic drift. However, some genes may be under diversifying selection in the demes. We found genes with high F_{ST} involved in growth process which appears to be in agreement with the differences in body length that have been found between the demes with deme II fish being significantly smaller than deme I fish (Ryman et al., 1979). Further, several of the putatively selected genes were associated with reproductive functions possibly supporting differences in reproductive characteristics between the two demes. Spatial separation of spawning grounds in streams due to a strong homing behaviour is a typical population separator in brown trout (Ferguson et al., 2019). In Bunnersjöarna such a mechanism has been suggested since deme II fish have been observed to primarily occur relatively close to the inlet stream in the southern lake, whereas deme I fish are found in the northern lake with the outlet stream towards Lake Ånnsjön (Ryman et al., 1979). We also find outlier SNPs linked with other genes (Figures S3 and S4), but note that we observe no clear genome regions of selection. Rather, the signature is diffuse across the genome with a weak signal with underlying loci being difficult to identify. This may be due to many false positives (no real underlying selection) or that selective differences are highly polygenic. Further work is needed to resolve this issue.

4.2 | Genetic diversity within and between populations

We find a strikingly lower level of genetic variation within deme II than deme I for all variability measures, and genomic data from individual sequencing estimated much higher inbreeding levels in deme II than in deme I. Similar degrees of difference in diversity appear rarely observed in sympatric salmonids (Jorde et al., 2018). Relatively few studies estimate inbreeding from whole-genome sequencing data in natural populations. Those that do typically focus on extinct or highly threatened populations such as woolly mammoth (Palkopoulou et al., 2015), Scandinavian wolves (Kardos et al., 2018), and gorillas (van der Valk, 2019). In these populations considerably longer (>2 Mb) runs of homozygosity than ours were observed. We find only a few ROHs above 2 Mb. Deme II individuals have the highest number with a total of 14 ROHs >2Mb (Table S6). This might imply that recent inbreeding is not pronounced and that inbreeding is mainly due to deep historical shared ancestors of parents. Similar observations with few long ROHs suggesting limited recent inbreeding have been observed in wild *Ficedula* flycatcher populations (Kardos et al., 2017).

Our estimates of F_{ROH} range from 0.032 in individual 1 in deme I to 0.405 in individual 1 in deme II. The samples from eight other lakes show values between these extremes but all values for other lakes are above F_{ROH} >0.09 with an average of 0.18 (range: 0.09-0.26; Table S6). This appears high, and is above estimates for hatchery strains of rainbow trout where the highest observations were typically around 0.1–0.2 (D'Ambrosio et al., 2019). A large extent of this difference might be due to differences in settings of ROH analyses: D'Ambrosio et al. (2019) allowed a maximum of one heterozygous genotype per ROH while we allowed three. For a low coverage data like ours, a maximum of three heterozygous genotypes is recommended (e.g., Ceballos et al., 2018).

In comparison with eight other lakes deme II in Lakes Bunnersjöarna stands out as less genetically variable, more inbred, and more isolated than any other population (Appendix S2; Table S6; Figures S8 and S9). For the other populations, nucleotide diversity was around 0.001. This is in line with what we have observed previously from brown trout of the same geographic area (Kurland et al., 2019), but considerably lower than that observed using ddRADseq for brown trout populations of the Atlantic lineage (same lineage as our study system) held in a hatchery in south-western France (π = 0.004) and for wild Mediterranean lineage populations (π = 0.005; Leitwein et al., 2018).

4.3 | Linking the LDH enzymes to genes

Contrasting homozygosity for the allozyme locus *LDH-1* coding for LDH-A detected the sympatric demes in Lake Bunnersjöarna (Allendorf et al., 1976; Ryman et al., 1979), and we tried to identify the *LDH-1* gene in the genome and to find the sequence differences that result in the allozyme divergence. However, we were not fully successful. We located two *LDH-A* loci as expected from the genome duplication, on chromosomes 7 and 17, respectively. The amino acid composition of the protein products suggested the locus on chromosome 7 to be *LDH-1* but our SNP- and gene-wise analyses revealed that neither copy of *LDH-A* shows a fixed or strong differentiation between the demes. The regulatory region of *LDH-A* on chromosome 17 did show strong divergence in Pool-seq data that seemed to agree well with the null allele characterizing deme II in *LDH-1* (Allendorf et al., 1984), but this pattern was not observed in the individual sequences. This indicates that *LDH-1* expression is complex, that contrasting homozygosity of an allozyme may not result directly from contrasting homozygosity in a DNA sequence, and that identifying the genes and the regulatory mechanisms underlying these enzymes may be particularly challenging in duplicated genomes.

4.4 | What additional information did genomic analysis provide?

The present results are consistent with patterns detected using only a few allozymes over 40 years ago. However, several new insights have been gained from the whole-genome analyses applied here. First, the divergence between the two demes appears to result primarily from genetic drift, but selection might also play

a role, particularly including polygenic traits of metabolic and growth processes. Second, the pattern of divergence and the allele occurrence over the SNP array indicate that deme II is reproductively isolated, whereas deme I appears to be in at least some contact with Lake Ånnsjön, possibly by downstream migration. Third, the level of genome-wide diversity and inbreeding is strikingly different between the two demes and appears large in relation to brown trout of other lakes. Also, we find that the genomic background behind the *LDH-1* allozyme expressions is complex and probably involves regulatory mechanisms and possibly interactions between several genes.

This study has not addressed what causes the reproductive isolation between these demes and if this structuring has evolved sympatrically or allopatrically, if it reflects natural evolution or if man mediated release has played a role. Preliminary analysis of the mitochondrial DNA sequences of these demes do not lend support for separate lineages reflecting colonization from different glaciation refugia (unpublished data). Further, we cannot exclude human translocation. For example, such translocation could imply that deme I was translocated from one place (e.g., Lake Ånnsjön) and deme II from another lake, and that the populations did not, or only to a very limited degree, hybridize in the new environment. There are lakes c 20 km away where the LDH-1 null allele has been observed (Allendorf et al., 1984). Further, we have not been able to address the potential temporal stability of these structures which appears highly warranted, but difficult in light of a current strict protection of these lakes.

5 | CONCLUSIONS

We revisited the first reported case of cryptic sympatry in brown trout that was detected by contrasting homozygosity at an allozyme locus in the tiny Lakes Bunnersjöarna, central Sweden using a 96 SNP array and various genomic tools. The present findings confirm reproductive isolation between the sympatric demes. Our genomic data show that divergence between these two demes is a genome-wide phenomenon governed by genetic drift but also by selective mechanisms. Our work demonstrates that populations from the same habitat may have large genome-wide divergence without obvious morphological distinction, which has implications for management and conservation. This type of hidden biodiversity needs to be mapped, monitored, and managed sustainably.

ACKNOWLEDGEMENTS

We thank Sigbjørn Lien and Matthew Peter Kent at CIGENE, Norwegian University of Life Sciences for providing SNP sequences and linkage map for development of the 96-SNP array on the EP1™ Fluidigm genotyping platform. We thank Daniel Ruzzante, one anonymous reviewer and the Subject Editor Prof. J. A. H. Benzie for valuable comments and suggestions. We acknowledge support from the National Genomics Infrastructure (NGI) in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council. The computations and data handling for genomic data were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at Uppsala Multidisciplinary Center for Advanced Computational Science partially funded by the Swedish Research Council through grant agreement no. 2018-05973. This research was supported by the Swedish Research Council Formas (L.L.), the Swedish Research Council (L.L.), the Carl Trygger and the Erik Philip-Sörensen Foundations (L.L.), the SciLifeLab Bioinformatics Long-term Support (L.L.) funded by the Knut and Alice Wallenberg foundation (grant no. 2014.0278). V.E.K. and D.E. are financially supported by the Knut and Alice Wallenberg Foundation as part of the National Bioinformatics Infrastructure Sweden at SciLifeLab.

AUTHOR CONTRIBUTIONS

Linda Laikre, Anastasia Andersson, Sara Kurland, and Nils Ryman designed the study; Nils Ryman, Fred W. Allendorf, Gunnar Ståhl, and Linda Laikre provided the material and allozyme data, Sten Karlsson provided 96 SNP array genotypes, Anastasia Andersson, Atal Saha, Nils Ryman, Linda Laikre analysed SNP array data, Atal Saha analysed genomics data in collaboration with Sara Kurland with supervision provided by Verena E. Kutschera, Diana Ekman, and Marty Kardos. Ola Hössjer and Nils Ryman performed POWSIM theoretical evaluation and supervised simulations performed by Atal Saha. Ola Hössjer evaluated F_{ST} estimators for degree of bias. Naomi L. P. Keehnen guided and located the 96 SNPs in the reference genome and performed the *LDH-A* analysis with allozyme guidance from Fred W. Allendorf and Gunnar Ståhl. Atal Saha and Linda Laikre led the writing with contribution from all authors. Linda Laikre funded the study.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY STATEMENT

Illumina raw sequences from this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB41224 (https://www.ebi.ac.uk/ena/browser/view/PRJEB41224). Processed data are available at Dryad: https://doi.org/10.5061/dryad.jm63xsjc3.

ORCID

Atal Saha https://orcid.org/0000-0003-1334-928X
Anastasia Andersson https://orcid.org/0000-0002-5698-4948
Sara Kurland https://orcid.org/0000-0002-5370-1236
Naomi L. P. Keehnen https://orcid.org/0000-0002-8782-3477
Verena E. Kutschera https://orcid.org/0000-0002-8930-534X
Ola Hössjer https://orcid.org/0000-0003-2767-8818
Fred W. Allendorf https://orcid.org/0000-0003-3663-6425
Nils Ryman https://orcid.org/0000-0003-3342-8479
Linda Laikre https://orcid.org/0000-0001-9286-3361

REFERENCES

Adams, C. E., Wilson, A. J., & Ferguson, M. M. (2008). Parallel divergence of sympatric genetic and body size forms of Arctic charr, *Salvelinus alpinus*, from two Scottish lakes. *Biological Journal of the Linnean Society*, *95*(4), 748–757. https://doi.org/10.1111/j.1095-8312.2008.01066.x

Alexa, A. & Rahnenfuhrer, J. (2020). topGO: Enrichment analysis for gene ontology. R package version 2.40.0.

Allendorf, F. W., Ryman, N., Stennek, A., & Ståhl, G. (1976). Genetic variation in Scandinavian brown trout (*Salmo trutta* L.): evidence of distinct sympatric populations. *Hereditas*, *83*(1), 73–82. https://doi.org/10.1111/j.1601-5223.1976.tb01572.x

Allendorf, F. W., Ståhl, G., & Ryman, N. (1984). Silencing of duplicate genes: a null allele polymorphism for lactate dehydrogenase in brown trout (*Salmo trutta*). *Molecular Biology and Evolution*, *1*(3), 238–248. https://doi.org/10.1093/oxfordjournals.molbev.a040315

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Andersson, A., Jansson, E., Wennerström, L., Chiriboga, F., Arnyasi, M., Kent, M. P., Ryman, N., & Laikre, L. (2017). Complex genetic diversity patterns of cryptic, sympatric brown trout (*Salmo trutta*) populations in tiny mountain lakes. *Conservation Genetics*, *18*(5), 1213–1227. https://doi.org/10.1007/s10592-017-0972-4

Attard, C. R. M., Beheregaray, L. B., & Möller, L. M. (2016). Towards population-level conservation in the critically endangered Antarctic blue whale: the number and distribution of their populations. *Scientific Reports*, *6*, 22291. https://doi.org/10.1038/srep22291

Aykanat, T., Johnston, S. E., Orell, P., Niemelä, E., Erkinaro, J., & Primmer, C. R. (2015). Low but significant genetic differentiation underlies biologically meaningful phenotypic divergence in a large Atlantic salmon population. *Molecular Ecology*, *24*(20), 5158–5174. https://doi.org/10.1111/mec.13383

Bekkevold, D., Höjesjö, J., Nielsen, E. E., Aldvén, D., Als, T. D., Sodeland, M., & Hansen, M. M. (2020). Northern European *Salmo trutta* (L.) populations are genetically divergent across geographical regions and environmental gradients. *Evolutionary Applications*, *13*, 400–416. https://doi.org/10.1111/eva.12877

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., Ingram, K. K., & Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, *22*(3), 148–155. https://doi.org/10.1016/j.tree.2006.11.004

Ceballos, F. C., Hazelhurst, S., & Ramsay, M. (2018). Assessing runs of homozygosity: A comparison of SNP Array and whole genome sequence low coverage data. *BMC Genomics*, *19*(1), 106. https://doi.org/10.1186/s12864-018-4489-0

D'Ambrosio, J., Phocas, F., Haffray, P., Bestin, A., Brard-Fudulea, S., Poncet, C., Quillet, E., Dechamp, N., Fraslin, C., Charles, M., & Dupont-Nivet, M. (2019). Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding. *Genetics Selection Evolution*, *51*(1), 26. https://doi.org/10.1186/s12711-019-0468-4

Danecek, P., McCarthy, S., & Li, H. (2015). bcftools—utilities for variant calling and manipulating vcfs and bcfs.

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Ferguson, A., Reed, T. E., Cross, T. F., McGinnity, P., & Prodöhl, P. A. (2019). Anadromy, potamodromy and residency in brown trout *Salmo trutta*: the role of genes and the environment. *Journal of Fish Biology*, *95*(3), 692–718. https://doi.org/10.1111/jfb.14005

Futuyma, D. J., & Mayer, G. C. (1980). Non-allopatric speciation in animals. *Systematic Zoology*, *29*(3), 254–271. https://doi.org/10.2307/2412661

García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L. M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T. F., & Conesa, A. (2012). Qualimap: Evaluating next-generation sequencing alignment data. *Bioinformatics*, *28*(20), 2678–2679. https://doi.org/10.1093/bioinformatics/bts503

Gomez-Raya, L., Rodríguez, C., Barragán, C., & Silió, L. (2015). Genomic inbreeding coefficients based on the distribution of the length of runs of homozygosity in a closed line of Iberian pigs. *Genetics Selection Evolution*, *47*(1), 81. https://doi.org/10.1186/s12711-015-0153-1

Goudet, J. (2003) FSTAT (ver. 2.9.4), a program to estimate and test population genetics parameters. Retrieved from https://www2.unil.ch/popgen/softwares/fstat.html

Guo, Y., Song, Z., Luo, L., Wang, Q., Zhou, G., Yang, D., Zhong, D., & Zheng, X. (2018). Molecular evidence for new sympatric cryptic species of *Aedes albopictus* (Diptera: Culicidae) in China: A new threat from *Aedes albopictus* subgroup? *Parasites & Vectors*, *11*(1), 228. https://doi.org/10.1186/s13071-018-2814-8

Hahne, F., & Ivanek, R. (2016). Visualizing genomic data using Gviz and bioconductor. In E. Mathé, & S. Davis (Eds.), *Statistical genomics: Methods and protocols* (pp. 335–351). Springer.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70. http://www.jstor.org/stable/4615733

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2018). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*(D1), D309–D314. https://doi.org/10.1093/nar/gky1085

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*(14), 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jorde, P. E., Andersson, A., Ryman, N., & Laikre, L. (2018). Are we underestimating the occurrence of sympatric populations? *Molecular Ecology*, *27*(20), 4011–4025. https://doi.org/10.1111/mec.14846

Kardos, M., Åkesson, M., Fountain, T., Flagstad, Ø., Liberg, O., Olason, P., Sand, H., Wabakken, P., Wikenros, C., & Ellegren, H. (2018). Genomic consequences of intensive inbreeding in an isolated wolf population. *Nature Ecology & Evolution*, *2*(1), 124–131. https://doi.org/10.1038/s41559-017-0375-4

Kardos, M., Qvarnström, A., & Ellegren, H. (2017). Inferring Individual Inbreeding and demographic history from segments of identity by descent in *Ficedula* Flycatcher genome sequences. *Genetics*, *205*(3), 1319–1334. https://doi.org/10.1534/genetics.116.198861

Karlsson, E. K., Baranowska, I., Wade, C. M., Salmon Hillbertz, N. H. C., Zody, M. C., Anderson, N., Biagi, T. M., Patterson, N., Pielberg, G. R., Kulbokas, E. J., Comstock, K. E., Keller, E. T., Mesirov, J. P., von Euler, H., Kämpe, O., Hedhammar, Å., Lander, E. S., Andersson, G., Andersson, L., & Lindblad-Toh, K. (2007). Efficient mapping of Mendelian traits in dogs through genome-wide association. *Nature Genetics*, *39*(11), 1321–1328. https://doi.org/10.1038/ng.2007.10

Kawecki, T. J. (1996). Sympatric speciation driven by beneficial mutations. *Proceedings of the Royal Society of London B, Biological Sciences*, *263*(1376), 1515–1520. https://doi.org/10.1098/rspb.1996.0221

Kawecki, T. J. (1997). Sympatric speciation via habitat specialization driven by deleterious mutations. *Evolution*, *51*(6), 1751–1763. https://doi.org/10.1111/j.1558-5646.1997.tb05099.x

Knutsen, H., Jorde, P. E., Hutchings, J. A., Hemmer-Hansen, J., Grønkjaer, P., Jørgensen, K.-E., André, C., Sodeland, M., Albretsen, J., & Olsen, E. M. (2018). Stable coexistence of genetically divergent Atlantic cod ecotypes at multiple spatial scales. *Evolutionary Applications*, *11*(9), 1527–1539. https://doi.org/10.1111/eva.12640

Kofler, R., Orozco-terWengel, P., De Maio, N., Pandey, R. V., Nolte, V., Futschik, A., Kosiol, C., & Schlötterer, C. (2011). PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One*, *6*(1), e15925. https://doi.org/10.1371/journal.pone.0015925

Kofler, R., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, *27*(24), 3435–3436. https://doi.org/10.1093/bioinformatics/btr589

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Kurland, S., Wheat, C. W., Paz Celorio Mancera, M., Kutschera, V. E., Hill, J., Andersson, A., Rubin, C.-J., Andersson, L., Ryman, N., & Laikre, L. (2019). Exploring a Pool-seq-only approach for gaining population genomic insights in nonmodel species. *Ecology and Evolution*, *9*(19), 11448–11463. https://doi.org/10.1002/ece3.5646.

Lamichhaney, S., Barrio, A. M., Rafati, N., Sundstrom, G., Rubin, C.-J., Gilbert, E. R., Berglund, J., Wetterbom, A., Laikre, L., Webster, M. T., Grabherr, M., Ryman, N., & Andersson, L. (2012). Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(47), 19345–19350. https://doi.org/10.1073/pnas.1216128109

Leggett, R. M., Ramirez-Gonzalez, R. H., Clavijo, B. J., Waite, D., & Davey, R. P. (2013). Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Frontiers in Genetics*, *4*, 288. https://doi.org/10.3389/fgene.2013.00288

Leitwein, M., Gagnaire, P.-A., Desmarais, E., Berrebi, P., & Guinand, B. (2018). Genomic consequences of a recent three-way admixture in supplemented wild brown trout populations revealed by local ancestry tracts. *Molecular Ecology*, *27*(17), 3466–3483. https://doi.org/10.1111/mec.14816

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM (Publication no. arXiv:1303.3997v2 [q-bio.GN]).

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

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*(5), 1491–1505. https://doi.org/10.1111/j.1558-5646.1999.tb05413.x

Magi, A., Tattini, L., Palombo, F., Benelli, M., Gialluisi, A., Giusti, B., Abbate, R., Seri, M., Gensini, G. F., Romeo, G., & Pippucci, T. (2014). H3M2: detection of runs of homozygosity from whole-exome sequencing data. *Bioinformatics*, *30*(20), 2852–2859. https://doi.org/10.1093/bioinformatics/btu401

Mallet, J., Meyer, A., Nosil, P., & Feder, J. L. (2009). Space, sympatry and speciation. *Journal of Evolutionary Biology*, *22*(11), 2332–2341. https://doi.org/10.1111/j.1420-9101.2009.01816.x

Maynard Smith, J. (1966). Sympatric speciation. *The American Naturalist*, *100*(916), 637–650. https://doi.org/10.1086/282457

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, *70*(12), 3321–3323. https://doi.org/10.1073/pnas.70.12.3321

Nei, M., Tajima, F., & Tateno, Y. (1983). Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution*, *19*(2), 153–170. https://doi.org/10.1007/BF02300753

Orlov, V. N., Borisov, Y. M., Cherepanova, E. V., Grigor'eva, O. O., Shestak, A. G., & Sycheva, V. B. (2012). Narrow hybrid zone between Moscow and Western Dvina chromosomal races and specific features of population isolation in common shrew *Sorex araneus* (*Mammalia*). *Russian Journal of Genetics*, *48*(1), 70–78. https://doi.org/10.1134/S1022795412010152

Palkopoulou, E., Mallick, S., Skoglund, P., Enk, J., Rohland, N., Li, H., Omrak, A., Vartanyan, S., Poinar, H., Götherström, A., Reich, D., & Dalén, L. (2015). Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Current Biology*, *25*(10), 1395–1400. https://doi.org/10.1016/j.cub.2015.04.007

Palmé, A., Laikre, L., & Ryman, N. (2013). Monitoring reveals two genetically distinct brown trout populations remaining in stable sympatry over 20 years in tiny mountain lakes. *Conservation Genetics*, *14*(4), 795–808. https://doi.org/10.1007/s10592-013-0475-x

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research–an update. *Bioinformatics*, *28*(19), 2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, *8*(11), e1002967. https://doi.org/10.1371/journal.pgen.1002967

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ravinet, M., Westram, A., Johannesson, K., Butlin, R., André, C., & Panova, M. (2016). Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Molecular Ecology*, *25*(1), 287–305. https://doi.org/10.1111/mec.13332

Raymond, M., & Rousset, F. (1995). GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, *86*, 248–249. https://doi.org/10.1093/oxfordjournals.jhered.a111573

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754

Rousset, F. (2008). Genepop'007: A complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, *8*(1), 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

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*(12), e2000234. https://doi.org/10.1371/journal.pbio.2000234

Ryman, N., Allendorf, F. W., & Ståhl, G. (1979). Reproductive isolation with little genetic divergence in sympatric populations of brown trout (*Salmo trutta*). *Genetics*, *92*(1), 247–262.

Ryman, N., & Palm, S. (2006). POWSIM: A computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology Notes*, *6*(3), 600–602. https://doi.org/10.1111/j.1471-8286.2006.01378.x

Schindler, D. E., Armstrong, J. B., & Reed, T. E. (2015). The portfolio concept in ecology and evolution. *Frontiers in Ecology and the Environment*, *13*(5), 257–263. https://doi.org/10.1890/140275

Schindler, D. E., Hilborn, R., Chasco, B., Boatright, C. P., Quinn, T. P., Rogers, L. A., & Webster, M. S. (2010). Population diversity and the portfolio effect in an exploited species. *Nature*, *465*(7298), 609–612. https://doi.org/10.1038/nature09060

Schönswetter, P., Lachmayer, M., Lettner, C., Prehsler, D., Rechnitzer, S., Reich, D. S., Sonnleitner, M., Wagner, I., Hülber, K., Schneeweiss, G. M., Trávníček, P., & Suda, J. (2007). Sympatric diploid and hexaploid cytotypes of *Senecio carniolicus* (Asteraceae) in the Eastern Alps are separated along an altitudinal gradient. *Journal of Plant Research*, *120*(6), 721–725. https://doi.org/10.1007/s10265-007-0108-x

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*, *6*(7), e21800. https://doi.org/10.1371/journal.pone.0021800

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*, *105*(2), 437–460. https://doi.org/10.1093/genetics/105.2.437

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595. https://doi.org/10.1093/genetics/123.3.585

Takezaki, N., Nei, M., & Tamura, K. (2009). POPTREE2: software for constructing population trees from allele frequency data and computing other population statistics with windows interface. *Molecular Biology and Evolution*, *27*(4), 747–752. https://doi.org/10.1093/molbev/msp312

Taylor, E. B. (1999). Species pairs of north temperate freshwater fishes: Evolution, taxonomy, and conservation. *Reviews in Fish Biology and Fisheries*, *9*(4), 299–324. https://doi.org/10.1023/A:1008955229420

Turelli, M., Barton, N. H., & Coyne, J. A. (2001). Theory and speciation. *Trends in Ecology & Evolution*, *16*(7), 330–343. https://doi.org/10.1016/S0169-5347(01)02177-2

Turner, S. D. (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv*, 5165. https://doi.org/10.1101/005165

van der Valk, T. (2019). *Genomics of population decline* (1822 Doctoral thesis, comprehensive summary)), Uppsala University, Uppsala, Sweden: Uppsala University, Disciplinary Domain of Science and Technology, Biology, Department of Ecology and Genetics, Animal ecology.Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-384346 DiVA database

Verspoor, R. L., Smith, J. M. L., Mannion, N. L. M., Hurst, G. D. D., & Price, T. A. R. (2018). Strong hybrid male incompatibilities impede the spread of a selfish chromosome between populations of a fly. *bioRxiv*, 248237. https://doi.org/10.1101/248237

Via, S. (2001). Sympatric speciation in animals: The ugly duckling grows up. *Trends in Ecology & Evolution*, *16*(7), 381–390. https://doi.org/10.1016/S0169-5347(01)02188-7

Wang, J. (2019). A parsimony estimator of the number of populations from a STRUCTURE-like analysis. *Molecular Ecology Resources*, *19*(4), 970–981. https://doi.org/10.1111/1755-0998.13000

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, *7*(2), 256–276. https://doi.org/10.1016/0040-5809(75)90020-9

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Wilson, A. J., Gíslason, D., Skúlason, S., Snorrason, S. S., Adams, C. E., Alexander, G., Danzmann, R. G., & Ferguson, M. M. (2004). Population genetic structure of Arctic Charr, *Salvelinus alpinus* from northwest Europe on large and small spatial scales. *Molecular Ecology*, *13*(5), 1129–1142. https://doi.org/10.1111/j.1365-294X.2004.02149.x

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Saha, A., Andersson, A., Kurland, S., Keehnen, N. L. P., Kutschera, V. E., Hössjer, O., Ekman, D., Karlsson, S., Kardos, M., Ståhl, G., Allendorf, F. W., Ryman, N., & Laikre, L. (2022). Whole-genome resequencing confirms reproductive isolation between sympatric demes of brown trout (*Salmo trutta*) detected with allozymes. *Molecular Ecology*, 31, 498–511. https://doi.org/10.1111/mec.16252