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 Keehnen[1], Verena Esther Kutschera[2], Diana Ekman[2], Sten Karlsson[3], Marty Kardos[4], Ola Hössjer[1], Gunnar Ståhl[5], Fred Allendorf[4], Nils Ryman[1], and Linda Laikre[1]

[1]Stockholm University
[2]Science for Life Laboratory
[3]Norwegian Institute for Nature Research (NINA)
[4]University of Montana
[5]Biolab

January 11, 2021

Abstract

The sympatric existence of genetically distinct 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 only certain parts of the genome. We return to the first reported case of salmonid sympatry, initially detected through contrasting homozygosity at a single allozyme locus (lactate dehydrogenase, LDH-A1) in brown trout in the small Lakes Bunnersjöarna, central Sweden. We use DNA from samples collected in the 1970s and a 96 SNP fluidigm array to verify the existence of the coexisting demes. We then apply whole-genome resequencing of pooled DNA to explore genome-wide diversity within and between these demes; strong genetic divergence is observed with genome-wide FST=0.13. Nucleotide diversity is estimated to 0.0013 in Deme I but only 0.0005 in Deme II. Individual whole-genome resequencing of two individuals per deme suggests considerably higher inbreeding in Deme II vs. Deme I. Comparing with similar data from other lakes we find that the genome-wide divergence between the demes is similar to that between reproductively isolated populations. We located two genes for LDH-A and found divergence between the demes in a regulatory section of one of the genes, but we could not find a perfect fit between allozyme and sequence data. Our data demonstrate genome-wide divergence governed by genetic drift and diversifying selection, confirming reproductive isolation between the sympatric demes.

1 INTRODUCTION

Populations of the same species that co-exist spatially over at least a part of their life-cycle (Futuyma & Mayer, 1980; Mallet, Meyer, Nosil, & Feder, 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, Barton, & Coyne, 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, Armstrong, & Reed, 2015; Schindler et al., 2010) and to genetic diversity recognized in international conservation policy, e.g. 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, Beheregaray, & Möller, 2016; Guo et al., 2018; Knutsen et al., 2018; Orlov et al., 2012; Ravinet et al., 2016; Schönswetter et al., 2007; Verspoor, Smith, Mannion, Hurst, & Price, 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, Andersson, Ryman, & Laikre, 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 (Allendorf, Ryman, Stennek, & Ståhl, 1976; Ryman, Allendorf, & Ståhl, 1979). An additional 53 allozyme loci were screened and five of those supported the division while the remaining 48 were fixed (or nearly fixed) for the same allele in both demes resulting in the conclusion of the existence of reproductively isolated, sympatric populations with little genetic divergence (Ryman et al., 1979). Statistically significant body size differences between the two populations (denoted demes) 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). Further, the allozymes indicated greater amounts of genetic variation in Deme I than in Deme II.

Salmonid fishes represent one of the organism groups that has been most extensively studied with respect to natural population genetic structure (Utter, 2004), and a recent review found over 130 cases of sympatric populations to have been identified world-wide in salmonid fishes; 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 paper, we reanalyze samples from Lakes Bunnersjöarna and apply single nucleotide polymorphism (SNP) array analyses as well as whole-genome resequencing analyses based on a recent genome assembly for brown trout (https://www.ncbi.nlm.nih.gov/assembly/GCF_901001165.1) 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 MATERIAL AND METHODS

We used tissues from the material collected in 1975 (Allendorf et al., 1976; Ryman et al., 1979) that have been stored at -30°C. We first applied a 96 SNP fluidigm array to verify that the existence of the sympatric populations can be detected by markers other than allozymes. Next, we sequenced whole genomes from several pooled individuals (Pool-seq; Kofler, Langmüller, Nouhaud, Otte, & Schlötterer, 2016) as well as from two separate individuals of each deme.

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; Appendix S1, Figure S1). Both lakes are shallow, the southern lake only 0.5 m with a deeper middle (a few meters) 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); 68 and 72 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, Ståhl, & Ryman, 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).

We included data from eight additional brown trout populations that are part of other projects (Andersson et al., in prep; Kurland et al., in prep) to put the diversity and divergence patterns observed in Lakes Bunnersjöarna into perspective (Appendix S1, Figure S1).

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, MA, USA) according to the manufacturer´s instructions and eluted in 100 µl elution buffer. DNA quality was assessed by electrophoresis through a 1% agarose gel and subjectively assessing the proportion of high-molecular weight DNA relative to degraded DNA. The extraction procedure was the same for DNA used for genotyping using a 96 SNP array as for individual whole-genome sequencing (WGS). Double-stranded DNA was quantified using a Qubit fluorometer (ThermoScientific, MA, USA) and normalised to 30–50 ng/µl.

We genotyped all 140 fish using an EP1^{TM} 96.96 Dynamic array IFCs genotyping platform (Fluidigm, San Francisco, CA) comprising 96 SNPs shown to be variable in Danish brown trout and evenly distributed on 40 linkage groups, selected from 3,782 SNPs identified in brown trout (Bekkevold et al., 2020; their Table S8). Using TBLASTN (E. values <0.0001 & bitscore 80; Altschul, Gish, Miller, Myers, & Lipman, 1990) we identified the location of these 96 SNPs on the brown trout reference genome (https://vgp.github.io/genomeark/Salmo_trutta/). We located 95 of the SNPs on chromosomes, 1 SNP was located on an unplaced scaffold. Two chromosomes contain 1 SNP each, while all others carry 2-3 SNPs. Using bedtool's 'intersectBed' (v2.27.1; Quinlan & Hall, 2010), we investigated if SNPs were located within coding regions. We visualised SNPs with karyplotR in R (Gel & Serra, 2017; Supporting Information Figure S3). The results from the SNP analyses supported the existence of the two demes (see Results), and random samples of n =50 individuals per deme were used for Pool-seq. Also, we randomly selected n =2 individuals per deme for WGS.

DNA extraction for resequencing followed the same extraction protocol as above but with an additional RNase A treatment. DNA quality was assessed by visual inspection of DNA fragmentation on agarose gels and absorbance at 260/280. DNA with high molecular weight from each of 50 individuals per population was quantified using fluorometry (Qubit;

Thermo Scientific) and pooled at equal concentrations to achieve 3 µg pooled genomic DNA in a volume in the range of 65–120 µl; samples were pooled at equal concentrations per deme. Pool-seq and individual WGS samples were sent to the National Genomics Infrastructure (NGI) at the Science of Life Laboratory (SciLifeLab), Stockholm, Sweden. NGI conducted the construction of PCR-free paired-end libraries with an average insert size of 350 bp followed by Illumina HiSeq 2000 sequencing using read length 150 bp.

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 (version 4.3; Raymond, 1995; Rousset, 2008). Holm's (1979) sequential Bonferroni approach was applied to adjust significance levels when evaluating the results from multiple testing. F_{ST}(Weir & Cockerham, 1984)

was estimated using FSTAT v2.9.4 (Goudet, 2003). CHIFISH v5.0 (Ryman & Palm, 2006; available at http://www.zoologi.su.se/˜ryman/) was used for F_{ST} significance testing.

We also computed Nei´s (1973) parametricF_{ST}(F_{ST}=(H_T-H_S)/H_T) using GenAlEx v6.5 (Peakall & Smouse, 2012) to allow direct comparison with F_{ST}calculated from Pool-seq data where only Nei´sF_{ST} is possible to obtain. Confidence intervals for Nei's F_{ST} were calculated using the following equation: $F_{ST} \pm t_{df}$[?]s^2/n (s^2 is the variance ofF_{ST} among loci), and for Weir & Cockerham'sF_{ST} using FSTAT (Goudet, 2003). We note that Nei's pairwise F_{ST} is typically around half that of Weir & Cockerham's. To avoid confusion, we consistently try to indicate the type of F_{ST} we refer to.

We assessed the most likely number of populations (K) using STRUCTURE v2.3.4 (Pritchard, Stephens, & Donnelly, 2000) using the default model allowing population admixture and correlated allele frequencies. We used a burn-in of 250,000 steps and 500,000 Markov chain (MCMC) replicates to estimate Q(assignment probability for each individual to each cluster) and likelihoods for different K (= 1–15). Estimation of the most likely K was repeated over ten runs and the output was analyzed using STRUCTURE HARVESTER v0.6.94 (Earl & vonHoldt, 2012). Mean individual Q values to each deme over runs were derived from CLUMPP (Jakobsson & Rosenberg, 2007). We based our estimation of the most likely value of$K$ on the mean likelihood value from STRUCTURE, ΔK(Evanno, Regnaut, & Goudet, 2005) from STRUCTURE HARVESTER, and on results from KFinder v1.0 (Wang, 2019).

We also explored population structure using BAPS v6.0 (Corander, Marttinen, & Mantyniemi, 2006) and the details from this analysis are provided in Appendix S2. We constructed an individual-based neighbor-joining phylogenetic tree based on Nei's D_A distance estimates (Nei, Tajima, & Tateno, 1983) from the 96 SNP array using POPTREE2 (Takezaki, Nei, & Tamura, 2009), and MEGAX 10.0.5 (Kumar, Stecher, Li, Knyaz, & Tamura, 2018). We used the default number of bootstrap replications (1,000); the tree was condensed to only include branches with bootstrap support of at least 70%.

2.4 Pool-seq data processing and variant calling

We assessed the quality of the raw sequence reads of each pool using FastQC v0.11.5 (Leggett, Ramirez-Gonzalez, Clavijo, Waite, & Davey, 2013), and the results from different pools were jointly evaluated using MultiQC v1.5 (Ewels, Magnusson, Lundin, & Käller, 2016). Low quality bases (phred score <20) and Illumina adapters were trimmed off the reads using BBDuk as implemented in BBTools v38.08 (http://sourceforge.net/projects/bbmap/). The trimmed reads were mapped against the brown trout reference assembly (comprising 2,371,863,509 bp; https://vgp.github.io/genomeark/Salmo_trutta/) using the Burrows–Wheeler Aligner v0.7.17 (BWA, using bwa mem algorithm; Li & Durbin, 2009). Resulting bam files were sorted, merged per pool and filtered for paired reads using SAMtools v1.8 (Li et al., 2009). The quality of the obtained bam files per pool were evaluated with Qualimap v2.2.1 (García-Alcalde et al., 2012) and summarized in MultiQC v1.5. Read depth histograms obtained from Qualimap were assessed to define minimum and maximum depth thresholds for subsequent population genomic analyses. SAMtools was applied for variant calling using minimum mapping quality and base quality scores of 20 and the parameter "base alignment quality" (BAQ; "-B") to reduce false SNPs caused by misalignments, resulting in one mpileup file for the two pools. We used the 'identify-genomic-indel-regions.pl' script of PoPoolation2 v1201 (Kofler, Pandey, & Schlötterer, 2011) to remove any indels from the mpileup file. No SNPs were kept from the error-prone 5 bp upstream and downstream of indels. A synchronized file was created for downstream analyses using the 'mpileup2sync.jar' script of PoPoolation2.

2.4.1 Population genomic analyses of Pool-seq data

PoPoolation v1.2.2 (Kofler, Orozco-terWengel, et al., 2011) was used to estimate nucleotide diversity (π · Tajima, 1983), Watterson's theta (; Watterson, 1975) and Tajima's D (T_D; Tajima, 1989) separately for each pool. Estimates of π andT_D from Pool-seq data are sensitive to sequencing errors and variation in coverage (Kofler, Orozco-terWengel, et al., 2011). Therefore, we subsampled mpileup files per pool to uniform depths that were chosen based on the mode of the read depth histogram for each pool (using a

minimum depth of 0.5*mode and a maximum depth of mode+0.5*mode for subsampling; see Kurland et al., 2019). Subsampling was done without replacement using the 'subsample-pileup.pl' script implemented in PoPoolation v1.2.2 (Kofler, Orozco-terWengel, et al., 2011). The 'variance-sliding.pl' script of PoPoolation v1.2.2 (Kofler, Orozco-terWengel, et al., 2011) was used to estimate π(including invariant sites) and T_{D}. Estimations were done for non-overlapping 5 kilo base pair (kb) windows across the assembly, with a minor allele count of 2 for a SNP to be called and applying the same depth thresholds as for the subsampling. Finally, only windows covered to [?]80% with data within the depth thresholds were used. We estimated π and T_{D} using only the sites retained in both pools. All retained sites with 20X to 150X depth of coverage were used to estimate with the same script and parameters as described for $\pi$and T_{D} but without subsampling.

PoPoolation2 v1201 (Kofler, Pandey, et al., 2011) was used to calculate F_{ST}(Nei, 1973) between the two pools using the 'fst-sliding.pl' script. A minor allele count of 3 was applied for SNP calling. We only retained variant positions with 20X to 150X depth of coverage, which is supposed to remove paralogous regions (inflated coverage) and regions represented by a small number of individuals (Kofler, Pandey, et al., 2011; their supporting information 1). We estimated F_{ST}between the demes per 1 bp site, as well as for larger, non-overlapping windows of 100 bp, 1, 5, 10, and 20 kb sizes to minimize stochastic errors linked to small window sizes (Kofler, Pandey, et al., 2011). Windows were only included if the fraction of the window covered with data was [?]0.8. In addition to these genome-wide F_{ST}calculations, the gene annotation was transferred onto the synchronized file using 'create-genewise-sync.pl'. F_{ST} was calculated per gene using the 'fst-sliding.pl' script with the same settings mentioned above, with a larger window size (1,000,000 bp) than the length of the largest gene present in the genome, following the PoPoolation authors' recommendation. For subsequent analyses we focused on windows of 1 bp and 5 kb for estimating diversity, divergence, and for performing outlier analyses. Genome-wide divergence patterns were visualized with a Manhattan plot created from F_{ST}estimates using the R package qqman (Turner, 2014). Confidence intervals (95%) of statistical parameters were calculated in R-studio v3.5.1 (R Core Team, 2018).

2.4.2 Identifying potential outliers

A modified version of POWSIM (Ryman & Palm, 2006) was used to investigate whether the distribution ofF_{ST} values in our Pool-seq data was consistent with the expectation for selectively neutral loci, or if there was deviation from such a distribution which could reflect directional selection (cf. Lamichhaney et al., 2012). We compared the observed genome-wide distribution ofF_{ST} with a simulated "expected" distribution under drift only. The observed distribution was based on per SNP estimates (1 bp; n =12,177,462; meanF_{ST}=0.083). We used a conservative approach defining outliers as SNPs with an observed F_{ST}above the largest value of the expected distribution (F_{ST}>0.864; details in Appendix S3). We also defined outliers as those SNPs with anF_{ST} above a thresholds corresponding to the 99.5$^{\mathrm{th}}$ percentile (e.g. Fabian et al., 2012; Keehnen, Hill, Nylin, & Wheat, 2018) of the genome-wide F_{ST} distribution. A threshold of the 97.5$^{\mathrm{th}}$ percentile was used forF_{ST} based on 5 kb windows (e.g. Gagnaire et al., 2018; Keehnen et al., 2019).

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 the outlier SNPs and outlier 5 kb windows. This analysis tests if multiple genes can be linked with certain GO functions. First, 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 from Ensembl (https://www.ensembl.org/Salmo_trutta/Info/Index). Second, we converted the NCBI *S. trutta* annotation release 100 from GFF to bed format using bedtools v2.27.1 (Quinlan & Hall, 2010), including only CDS regions, and extracted all genes that overlapped with the identified outlier SNPs and windows using bedtool's (Quinlan & Hall, 2010) 'intersect' function. Third, the R package topGO (Alexa & Rahnenfuhrer, 2020) was used to test for over-representation of GO biological processes using a node size of 5 and the 'classic' algorithm to account for structural relationships among GO terms. Finally, GO terms with p-values [?]0.01 (i.e., enriched GO terms) were retained and then filtered in REVIGO (Supek, Bošnjak, Škunca, & Šmuc, 2011) to remove redundant GO terms. Treemaps were drawn in R (Supek et al., 2011) to visualize enriched and non-redundant GO superclusters.

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), and we identified the DNA sequences relating to this enzyme. First, all genes related to LDH (EC number 1.1.1.27, GO:0004457) were identified using the GO term protein annotations from EggNOG (section 2.4.3.). Second, to ensure no gene copy was missed, we used TBLASTN (Altschul et al., 1990) with the FASTA sequence for LDH extracted from the Uniprot database to search the genome assembly, extracting hits with E- values<0.0001 and bitscore>60. Finally, each *LDH-A* copy was manually inspected for exon boundaries, reading frame and mapping performance by loading the brown trout reference assembly and the bam files generated with the Pool-seq and individual whole-genome sequencing pipelines (see below for the latter) into the Integrative Genomics Viewer (IGV) v2.4.2 (Robinson et al., 2011). In both cases, the sequencing reads had only been filtered for minimum base quality 20 and for being mapped as pairs, i.e. without depth or additional quality filtering, to ensure that reads mapping to the *LDH-A* loci had not been removed. The amino acid compositions for all LDH genes were extracted using an in-house script (Appendix S4).

We calculated allele frequencies for the *LDH-A* gene copies in both demes from the Pool-seq data (gene-based sync files; section 2.4.1.) using 'snp-frequency-diff.pl' from Popoolation2 (Kofler, Pandey, et al., 2011). Sequence variants were called for these genes on 14 of the individual WGS data using bcftools call (Danecek, McCarthy, & Li, 2015), inspected using IGV, and visualized with Gviz (Hahne & Ivanek, 2016) and ggplot2 in R (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 (Li et al., 2009). No data was filtered before mapping sequences against the reference genome (unlike the pipeline for Pool-seq data). Alignment generated two bam files for each individual (one per lane), which were merged per individual using PICARD v2.10.3 MarkDuplicates (https://broadinstitute.github.io/picard/). PCR duplicates were also marked using PICARD v2.10.3. The quality of the obtained bam files 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 2.6) was performed with GATK GenotypeGVCFs. To filter out low quality variants, we applied a hard filter approach, using GATK's VariantFiltration tool, separately for SNPs (QD < 2.0, MQ < 40.0, FS > 10.0, MQRankSum < -5.0, ReadPosRankSum < -5.0, SOR > 5.0) and indels (QD < 2.0, FS > 10.0, ReadPosRankSum < -5.0, SOR > 5.0). VCFtools v0.1.15 (Danecek et al., 2011) was used to retain only bi-allelic SNPs with a minor allele frequency [?] 0.01. Further, variants from the un-assigned scaffolds were removed. We used BCFtools v1.8 (Li et al., 2009) to annotate the ID field of the joint VCF file.

2.5.1 Estimation of inbreeding

Inbreeding can be estimated from individual WGS as the fraction of the genome covered by 'runs of homozygosity' (ROH), and their length (LnROH; Gomez-Raya, Rodriguez, Barragan, & Silio, 2015; Kardos, Qvarnstrom, & Ellegren, 2017; Magi et al., 2014). We identified ROHs in each individual using the PLINK '–homozyg' method v1.90b4.9 (Purcell et al., 2007), since that approach provides lower false-positives than other methods (Kancheva et al., 2015). Homozygosity was determined for 1,000 kb overlapping windows. As suggested for low coverage data like ours (Ceballos, Hazelhurst, & Ramsay, 2018), the minimum length of a ROH was set to 300 kb which must contain at least 50 SNPs, and a maximum of 3 heterozygous genotypes per window were allowed. For measuring LnROH, we categorized lengths of ROHs into three groups: between 300 and 500 kb, >500 kb but [?]1,000 kb, and >1,000 kb. The fraction of ROH (F_{ROH}) for each

individual was calculated as the total length of ROHs (including all three length groups) divided by the total length of the autosomal genome sequenced.

3 RESULTS

3.1 Population divergence and diversity using SNP array data

A total of 77 of the 96 SNPs were polymorphic in Lakes Bunnersjoarna. 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). When grouping the individuals into demes based on the*LDH-1* genotype, we found a strong genetic divergence between the populations; Weir & Cockerham's (1984)F_{ST}=0.24 (CI: 0.19-0.29), and Nei's (1973) F_{ST}=0.12 (CI: 0.10-0.14). The correlation between these two approaches for measuring F_{ST} is high (r =0.99, p< 0.001). We found pronounced differences in the amount of genetic variation within the demes, consistent with data from the eight polymorphic allozyme loci (Ryman et al., 1979); 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).

When analyzing the SNP array data without prior grouping of individuals, STRUCTURE suggested K =2 clusters, and these two clusters are almost completely consistent with the *LDH-1* groupings (Figure 2a). Only one individual of Deme I (from *LDH-1* genotype) was assigned to the other deme, but the assignment probability was very close to 0.5 ($Q = 0.502$). Kfinder suggested K =3 as the most likely number of clusters, but the STRUCTURE barplot for K =3 did not indicate a third separate cluster (Figure 2b). Further, the log likelihood approach from STRUCTURE suggests K =11, where four of these clusters occur primarily within allozyme *LDH-1* Deme II and the remaining seven in the *LDH-1* Deme I (Figure 2c, also see Figure S3). BAPS suggested a total of six clusters and five of these clusters were located within *LDH-1* Deme I whereas the sixth cluster coincided completely with *LDH-1* Deme II (Appendix S2).

An individual-based neighbor-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 and showed additional sub-structuring within Deme I (Figure 3). From these observations, we conclude that the two demes identified by allozyme *LDH-1* are supported by SNP array data.

3.2 Pool-seq data and population genomics

In total, 213 giga base pairs (Gb; 1,527 million reads) and 197 Gb (1,302 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 1,366 and 1,169 million reads mapped as pairs, remained (Table S1). Read depth was larger in data from Deme I than from Deme II (mode of read depth=104 vs. 90). The average mapping quality (c. 33) and edit distance between the reads and reference (ca 0.6%) were similar in data from both pools (Table S1).

3.2.1 Genome-wide diversity and divergence

Genome-wide diversity, measured as nucleotide diversity (π) and Watterson's theta ( ), was considerably larger in Deme I than in Deme II (Table 2; Deme I π =0.0013 vs. Deme II π =0.00046). The average values for 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 2).

The genome-wide divergence between the two demes was high;F_{ST}=0.13 using window size of 5 kb (Table 3); this value was largely consistent for other window sizes (Table S2) and agrees with estimates from the 96 SNP array (Table 3). 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. There was a strong correlation between Nei´s (1973)F_{ST} from Pool-seq versusF_{ST} from the 96 SNP array (r =0.78, p<0.001; Figure S4).

3.2.2. F_{ST} outliers

High genomic divergence between the demes was found all across the genome (Figure 4). The POWSIM results showed that the distributions of the observed and simulated (expected) F_{ST} values differed from each other (Kolmogorov-Smirnov, $p<0.05$). The observed distribution showed a higher frequency of large F_{ST} values than the expected distribution (Figure 5). 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 observed distribution, of which 19 had $F_{ST}=1$. Of these 194 outliers, one was associated with a gene significantly enriched ($p<0.05$) 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 99.5^{th} percentile of the F_{ST} distribution (F_{ST} [?] 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[?]0.01) 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 sulfate 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 S5).

For 5 kb windows, 8,508 out of 340,297 windows were considered as outliers with F_{ST}-values above the 97.5^{th} percentile (F_{ST} [?] 0.35). 2,148 outlier windows were within genes. TopGO identified 1,494 feasible genes, associated to 113 significant GO terms (p[?]0.01). 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 S6).

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 per Lakes Bunnersjoarna individual (Table S5). This corresponds to approximately 98% of the raw data per individual. An average mapping quality of 32 and depth of coverage c. 10X was observed (Tables 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 bi-allelic 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. Similar results were obtained from individual WGS from other lakes (Table S6).

3.3.1 Inbreeding

Both individuals from Deme II showed considerably higher inbreeding measured as length of runs of homozygosity (LnROHs; Figure 6a, Table S7) and F_{ROH} (Figure 6b) than those from Deme I. Individual 1 from Deme II was the most homozygous; a total of 950,545 kb involving 1,742 ROH segments were estimated as identically homozygous in this individual. This translates into an F_{ROH} from PLINK of 0.405 (95% CI: 0.405–0.405; Figure 6b). The second individual from Deme II showed 1,520 ROH segments covering 782,973 kb (Figure 6a) and $F_{ROH}=0.334$ (95% CI: 0.334–0.334; Figure 6b). Much lower inbreeding levels were observed in the individuals from Deme I. The first had 175 ROH segments expanded over 75,092 kb with an $F_{ROH}=0.032$ (95% CI: 0.032–0.032) and the second 276 ROH segments spanning 127,296 kb and $F_{ROH}=0.055$ (95% CI: 0.055–0.055; Figure 6a, b). Individuals from Deme II showed some ROH segments longer than 1,000 kb; individual 1 had 114 such segments and individual 2 had 73.

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 S8). 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 (or 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-wise 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 theF_{ST} analyses of coding regions. However, the 3' UTR 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 S7, Table S9). In contrast, no divergence of patterns between the demes was observed for *LDH-A*on chromosome 7 (Figure S8).

The 3' UTR region on chromosome 17 was also analyzed 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 S9). We also looked at the 3' UTR region of chromosome 17 in 11 individuals from other lakes. All of these individuals had been scored for allozyme locus*LDH-1* and had all been scored as 100/100. Only one of these individuals (from Lake Saxvattnet) showed a similar IGV profile as those from the two demes in Lakes Bunnersjoarna, whereas all the others did not (Figure S9).

3.5 Comparisons with other lakes

The genetic divergence between the two demes of Lakes Bunnersjoarna was similar to the lower estimates of divergence between fish from lakes with no migratory contact (Tables S10a, b-S11). Lake Annsjon is the closest lake, 8.5 km downstream of Lakes Bunnersjorna (Figure 1), and the most parsimonious source for the colonization of Lakes Bunnersjorna. Weir & Cockerham's F_{ST} for the 96 SNP array between Lake Annsjon and Deme I (F_{ST}=0.12; Table S10a) was around half that of the divergence between the two demes (F_{ST}=0.24) while F_{ST}between Lake Annsjon and Deme II was 0.36.

There are strong correlations for different types ofF_{ST} values between population pairs from Pool-seq data and those from the SNP array data (Figure S10a, b). The TreeMix dendrogram from Pool-seq data (Figure 7; c.f. Appendix S1 for methods) and the neighbor-joining tree from SNP array data (Figure S11) illustrate how Deme II stands out as isolated and less connected to both Deme I and Lake Annsjon.

The nucleotide diversity estimated in brown trout from Bunnersjoarna Deme I (π =0.0013; Table 2) was of the same order of magnitude as estimates from other lakes where π ranged between 0.00104 to 0.00151 (Table S12), while π for Deme II was almost an order of magnitude lower (Table 2). The two individuals from Deme II showed higher average inbreeding levels measured as length of ROH (Figure 6a, Figure S12) and F_{ROH} (Figure 6b, Figure S13) than brown trout in any of the other lakes. In contrast, Deme I individuals showed the lowest values. Long runs of homozygosity were also more frequent in Deme II than in any other lake (Table S7).

4 DISCUSSION

4.1 Evidence for reproductive isolation

The present findings confirm the occurrence of two reproductively isolated populations of brown trout in lakes Bunnersjöarna as previously reported by Ryman et al. (1979).

We find indications that Deme II is reproductively isolated from Deme I, i.e. that gene flow from Deme I to Deme II is rare or non-existing. This conclusion is supported by the following observations i) a strikingly lower level of genetic diversity within Deme II as compared to Deme I for all variability measures used

(Tables 1, 2; Figure 6a, b), including that 53 loci of the 96 SNP array are monomorphic in Deme II but polymorphic in Deme I (Table 1), and *ii*) the highF_{ST} between the two demes which is in the same order as between populations from reproductively isolated lakes. In contrast, we cannot rule out a small amount of gene flow from Deme II to Deme I based on the present genomic and SNP array data. Deme II does not show any private alleles with the SNP array, and the STRUCTURE analysis suggests some minor admixture of Deme II into the Deme I cluster (Figure 2c).

The amount of divergence between the demes in Bunnersjöarna is confirmed to be large through-out the genome and it appears high in comparison to observations from other cases of sympatry in salmonids (Jorde et al., 2018). The observedF_{ST} values between the demes from the Pool-seq data (Nei´s F_{ST}=0.13), as well as from the 96 SNP array (Nei´s F_{ST}=0.14; Weir & Cockerham´sF_{ST}= 0.24), are considerably higher than those observed in other cryptic, sympatric salmonid populations including Arctic charr (Adams, Wilson, & Ferguson, 2008; Wilson et al., 2004), Atlantic salmon (Aykanat et al., 2015), and brown trout (Andersson et al., 2017; Palmé, Laikre, & Ryman, 2013). Rather, the amount of divergence between the demes appears more in line with what is found between sympatric salmonids showing substantial phenotypic divergence such as different morphotypes of Arctic charr (Conejeros et al., 2014; Gordeeva, Osinov, Alekseyev, Matveev, & Samusenok, 2010) or brown trout (Ferguson & Taggart, 1991).

Our outlier analyses indicate that most differentiation is caused by genetic drift but a number of genes may be under diversifying selection in the Lakes Bunnersjöarna demes (Table S4, Figures S5, S6). We found genes with high F_{ST} involved in growth process. It is interesting to note that a striking difference with respect to growth was observed between the two demes with fish from Deme II being markedly smaller than fish from Deme I (Ryman et al., 1979). Further, several of the genes putatively under selection were associated with reproductive functions. This, along with the strong genetic differentiation observed between the demes, indicates differences in reproductive characteristics between the two demes. Spatial separation due to separate spawning grounds in streams resulting from a strong homing behavior is a typical population separator in brown trout (Ferguson, Reed, Cross, McGinnity, & Prodöhl, 2019). In Bunnersjöarna such a mechanism has been suggested where 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).

A large number of putatively selected genes were associated with metabolic processes including genes associated with "NADH oxidation". Interestingly, LDH functions as the catalyst to all these processes (Parra-Bonilla, Alvarez, Al-Mehdi, Alexeyev, & Stevens, 2010; also see https://www.genome.jp/dbget-bin/www_bget?ko:K00016). Putatively selected genes were also associated with immunological processes, the roles of which are currently unknown in brown trout. However, adaptive divergence of immune related genes has been identified in Atlantic salmon (e.g. Kjærner-Semb et al., 2016) and it has been suggested that such adaptive genes have largely contributed to genetic divergence between Atlantic salmon populations in Norway. Observations of signatures of selection in particularly duplicated regions of the salmon genome have led researchers to hypothesize that genome duplications might increase the opportunity for evolutionary adaptations (Kjærner-Semb et al., 2016). Genome-wide association studies still remain to be carried out in brown trout but might clarify this issue further. However, it is also important to note that we observe no clear genome regions of selection – rather the signature (elevated differentiation) is diffuse across the genome with a weak signal with underlying loci being difficult to identify. This may be due to either 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 demes

We find a strikingly lower level of genetic variation within Deme II than Deme I for all variability measures (Tables 1, 2), and genomic data from individual sequencing estimated much higher inbreeding levels in Deme II than in Deme I (Figure 6a, b). 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 in our populations. Deme II individuals have the highest number with a total of 14 ROHs > 2Mb (Table S7). This might imply that recent inbreeding is not pronounced in any of the populations studied here 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 in Lakes Bunnersjöarna to 0.405 in individual 1 in Deme II (Figure 6b). The samples from other lakes show values between these extremes but all values for other lakes are aboveF_{ROH}>0.09 with an average of 0.18 (Table S7). This appears high, and is above what has been estimated 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´Ambrosia 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 genotype is recommended (e.g. Ceballos et al., 2018).

4.3 Comparison to other lakes

The comparison with eight other lakes showed that Deme II in Bunnersjöarna stands out as less genetically variable, more inbred, and more isolated than any other population. For the other populations, nucleotide diversity was around 0.001 (Table 2; Table S12). This is in line with what we have observed previously from brown trout of the same geographic area using the *Salmo salar* reference genome for mapping (Kurland et al., 2019). Our estimate is considerably lower than that observed using dd-RADseq for brown trout populations of the Atlantic lineage (same lineage as our study system) held in a hatchery in south-western France withπ =0.004 and for wild Mediterranean lineage populations withπ =0.005 (Leitwein, Gagnaire, Desmarais, Berrebi, & Guinand, 2018). Again, Deme II stands out as an extreme in comparison with other brown trout populations with genome wide nucleotide diversity of below 0.0005 (Table 2).

4.4 Linking the LDH enzymes to genes

Contrasting homozygosity for the allozyme locus *LDH-1* detected the sympatric populations in Lake Bunnersjöarna (Allendorf et al., 1976; Ryman et al., 1979), and we tried to identify the *LDH-1* gene in the genome and find the sequence differences that result in the allozyme divergence. We were not fully successful in this, however. We located two *LDH-A* loci as expected from the genome duplication event. One locus was located on chromosome 7 and the other on chromosome 17. The amino acid composition of the protein products from these loci suggested the LDH locus on chromosome 7 to be *LDH-1* due to a slightly more positive charge of the product at this locus as compared to the one on chromosome 17, as expected from allozyme patterns (Allendorf et al., 1984). However, our SNP- and gene-wise analyses revealed that neither copy of *LDH-A* that we identified in the reference genome shows a fixed or even strong differentiation between the demes.

When investigating the Pool-seq reads mapping to the two gene copies visually in the Integrative Genomics Viewer (IGV), we identified a region in the 3' UTR of the *LDH-A* on chromosome 17 that showed strong divergence, albeit not fixed differences, between the two demes, which was confirmed by allele frequency calculations (Figures S7, S9).*LDH-A* is highly regulated at the transcriptional and posttranscriptional levels (Jungmann, Huang, & Tian, 1998), and the 3' UTR region is known to be a critical determinant for the mRNA stabilizing activity for this gene (Tian, Huang, Short, Short, & Jungmann, 1998). We hypothesized that the divergence in this region could result in the null allele for Deme II observed in allozyme data, but this would imply that *LDH-1* is located on chromosome 17 (and not on 7 as indicated by the amino acid charges). A previous study suggests that regulatory regions may primarily be involved in divergence between the two sympatric forms of whitefish – dwarf and normal – in Cliff Lake in St John River drainage in

Maine, USA, since outliers were predominantly found in non-coding regions within genes (Hebert, Renaut, & Bernatchez, 2013).

When we examined the 3´ UTR of the *LDH-A* on chromosome 17 in the whole-genome sequences from a total of 16 individuals (from 7 lakes, including 2 individuals from each of the demes) scored for allozyme genotype in LDH-1, we found that the majority of the individuals (n =11) showed similar patterns as Deme I in Pool-seq data. However, one individual from Lake Saxvattnet, and all four individuals from Lakes Bunnersjöarna (both demes) showed the Deme II Pool-seq genotype at the 3' UTR region of chromosome 17 (Figure S9). These results indicate that *LDH-1* expression is a complex feature, and we were not able to resolve the genomic background to the contrasting allozyme expressions in the demes. Rather, further studies need to identify the exact genetic control behind the LDH-1 enzyme in brown trout. Our results suggest that contrasting homozygosity of an allozyme (F_{ST}=1) may not result directly from contrasting homozygosity in a DNA sequence. Identifying the genes and the regulatory mechanisms underlying these enzymes is particularly challenging in duplicated genomes.

4.5 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 results suggest that the divergence between the two demes results primarily from genetic drift, but that 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 also in relation to brown trout of other lakes. Deme I shows the highest degree of nucleotide diversity and lowest measures of runs of homozygosity compared to all other samples studied here, whereas Deme II shows radically lower levels of nucleotide diversity than any of the other samples, and the highest measures of ROH (cf. Table 2, Figure 6 vs. Tables S7, S12, Figures S12, S13). Further, we find that the genomic background behind the *LDH-1* allozyme expressions is complex and likely 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 human induced 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 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 important implications for management and conservation.

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 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, SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. This research was supported by the Swedish Research Council Formas (L.L.), the Swedish Research Council (L.L.) the Swedish Agency for Marine and Water Management (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.

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. doi: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., & Stahl, G. (1976). Genetic variation in Scandinavian brown trout (*Salmo trutta L .*): evidence of distinct sympatric populations.*Hereditas, 83* (1), 73-82. doi:10.1111/j.1601-5223.1976.tb01572.x

Allendorf, F. W., Stahl, 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. doi: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. doi:https://doi.org/10.1016/S0022-2836(05)80360-2

Andersson, A., Jansson, E., Wennerstrom, L., Chiriboga, F., Arnyasi, M., Kent, M. P., . . . 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. doi:10.1007/s10592-017-0972-4

Attard, C. R. M., Beheregaray, L. B., & Moller, 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. doi:10.1038/srep22291

Aykanat, T., Johnston, S. E., Orell, P., Niemela, 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. doi:10.1111/mec.13383

Bekkevold, D., Hojesjo, J., Nielsen, E. E., Aldven, 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. doi:10.1111/eva.12877

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., . . . Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution, 22* (3), 148-155. doi: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. doi:10.1186/s12864-018-4489-0

Conejeros, P., Phan, A., Power, M., O'Connell, M., Alekseyev, S., Salinas, I., & Dixon, B. (2014). Differentiation of Sympatric Arctic Char Morphotypes Using Major Histocompatibility Class II Genes. *Transactions of the American Fisheries Society, 143* (3), 586-594. doi:10.1080/00028487.2014.880734

Corander, J., Marttinen, P., & Mantyniemi, S. (2006). Bayesian identification of stock mixtures from molecular marker data. *Fishery Bulletin, 104* , 550-558.

D'Ambrosio, J., Phocas, F., Haffray, P., Bestin, A., Brard-Fudulea, S., Poncet, C., . . . 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. doi:10.1186/s12711-019-0468-4

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Group, G. P. A. (2011). The variant call format and VCFtools. *Bioinformatics, 27* (15), 2156-2158. doi:10.1093/bioinformatics/btr330

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. doi: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. doi:10.1111/j.1365-294X.2005.02553.x

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

Fabian, D. K., Kapun, M., Nolte, V., Kofler, R., Schmidt, P. S., Schlotterer, C., & Flatt, T. (2012). Genome-wide patterns of latitudinal differentiation among populations of*Drosophila melanogaster* from North America. *Molecular Ecology, 21* (19), 4748-4769. doi:10.1111/j.1365-294X.2012.05731.x

Ferguson, A., Reed, T. E., Cross, T. F., McGinnity, P., & Prodohl, 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. doi:10.1111/jfb.14005

Ferguson, A., & Taggart, J. B. (1991). Genetic differentiation among the sympatric brown trout (*Salmo trutta*) populations of Lough Melvin, Ireland.*Biological Journal of the Linnean Society, 43* (3), 221-237. doi:10.1111/j.1095-8312.1991.tb00595.x

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

Gagnaire, P.-A., Lamy, J.-B., Cornette, F., Heurtebise, S., Degremont, L., Flahauw, E., . . . Lapegue, S. (2018). Analysis of genome-wide differentiation between native and introduced populations of the cupped oysters *Crassostrea gigas*and *Crassostrea angulata* . *Genome Biology and Evolution, 10* (9), 2518-2534. doi:10.1093/gbe/evy194

Garcia-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L. M., Gotz, S., Tarazona, S., . . . Conesa, A. (2012). Qualimap: evaluating next-generation sequencing alignment data.*Bioinformatics, 28* (20), 2678-2679. doi:10.1093/bioinformatics/bts503

Gel, B., & Serra, E. (2017). karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics (Oxford, England), 33* (19), 3088-3090. doi:10.1093/bioinformatics/btx346

Gomez-Raya, L., Rodriguez, C., Barragan, C., & Silio, 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. doi:10.1186/s12711-015-0153-1

Gordeeva, N. V., Osinov, A. G., Alekseyev, S. S., Matveev, A. N., & Samusenok, V. P. (2010). Genetic differentiation of Arctic charr (*Salvelinus alpinus*) complex from Transbaikalia revealed by microsatellite markers. *Journal of Ichthyology, 50* (5), 351-361. doi:10.1134/S0032945210050012

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.htm

Guo, Y., Song, Z., Luo, L., Wang, Q., Zhou, G., Yang, 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. doi:10.1186/s13071-018-2814-8

Hahne, F., & Ivanek, R. (2016). Visualizing Genomic Data Using Gviz and Bioconductor. In E. Mathe & S. Davis (Eds.), *Statistical Genomics: Methods and Protocols* (pp. 335-351). New York, NY: Springer New York.

Hebert, F. O., Renaut, S., & Bernatchez, L. (2013). Targeted sequence capture and resequencing implies a predominant role of regulatory regions in the divergence of a sympatric lake whitefish species pair (*Coregonus clupeaformis*).*Mol Ecol, 22* (19), 4896-4914. doi:10.1111/mec.12447

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6* (2), 65-70.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernandez-Plaza, A., Forslund, S. K., Cook, H., . . . 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. doi: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. doi: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. doi:10.1111/mec.14846

Jungmann, R. A., Huang, D., & Tian, D. (1998). Regulation of LDH-A gene expression by transcriptional and posttranscriptional signal transduction mechanisms. *Journal of Experimental Zoology, 282* (1-2), 188-195. doi:10.1002/(sici)1097-010x(199809/10)282:1/2<188::aid-jez21>3.0.co;2-p

Kancheva, D., Atkinson, D., De Rijk, P., Zimon, M., Chamova, T., Mitev, V., . . . Jordanova, A. (2015). Novel mutations in genes causing hereditary spastic paraplegia and Charcot-Marie-Tooth neuropathy identified by an optimized protocol for homozygosity mapping based on whole-exome sequencing. *Genetics In Medicine, 18* , 600. doi:10.1038/gim.2015.139

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

Kardos, M., Qvarnstrom, 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. doi:10.1534/genetics.116.198861

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

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

Keehnen, N. L. P., Fors, L., Jarver, P., Spetz, A.-L., Nylin, S., Theopold, U., & Wheat, C. W. (2019). Geographic variation in hemocyte diversity and phagocytic propensity shows a diffuse genomic signature in the green veined white butterfly.*bioRxiv* , 790782. doi:10.1101/790782

Keehnen, N. L. P., Hill, J., Nylin, S., & Wheat, C. W. (2018). Microevolutionary selection dynamics acting on immune genes of the green-veined white butterfly, *Pieris napi* .*Molecular Ecology, 27* (13), 2807-2822. doi:10.1111/mec.14722

Kjaerner-Semb, E., Ayllon, F., Furmanek, T., Wennevik, V., Dahle, G., Niemela, E., . . . Edvardsen, R. B. (2016). Atlantic salmon populations reveal adaptive divergence of immune related genes - a duplicated genome under selection. *BMC Genomics, 17* (1), 610-610. doi:10.1186/s12864-016-2867-z

Knutsen, H., Jorde, P. E., Hutchings, J. A., Hemmer-Hansen, J., Gronkjaer, P., Jorgensen, K.-E. M., . . . Olsen, E. M. (2018). Stable coexistence of genetically divergent Atlantic cod ecotypes at multiple spatial scales. *Evolutionary Applications, 11* (9), 1527-1539. doi:10.1111/eva.12640

Kofler, R., Langmuller, A. M., Nouhaud, P., Otte, K. A., & Schlotterer, C. (2016). Suitability of different mapping algorithms for genome-wide polymorphism scans with Pool-seq data. *G3: Genes|Genomes|Genetics, 6* (11), 3507-3515. doi:10.1534/g3.116.034488

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

Kofler, R., Pandey, R. V., & Schlotterer, C. (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq).*Bioinformatics, 27* (24), 3435-3436. doi: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. doi:10.1093/molbev/msy096

Kurland, S., Wheat, C. W., de la Paz Celorio Mancera, M., Kutschera, V. E., Hill, J., Andersson, A., . . . Laikre, L. (2019). Exploring a Pool-seq-only approach for gaining population genomic insights in nonmodel species. *Ecology and Evolution, 0* (0). doi:10.1002/ece3.5646

Lamichhaney, S., Barrio, A. M., Rafati, N., Sundstrom, G., Rubin, C.-J., Gilbert, E. R., . . . Andersson, L. (2012). Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring.*Proceedings of the National Academy of Sciences, 109* (47), 19345-19350. doi: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* , 288-288. doi: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. doi:10.1111/mec.14816

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Subgroup, G. P. D. P. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics, 25* (16), 2078-2079. doi: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. doi:10.1111/j.1558-5646.1999.tb05413.x

Magi, A., Tattini, L., Palombo, F., Benelli, M., Gialluisi, A., Giusti, B., . . . Pippucci, T. (2014). H3M2 : detection of runs of homozygosity from whole-exome sequencing data.*Bioinformatics, 30* (20), 2852-2859. doi: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. doi:10.1111/j.1420-9101.2009.01816.x

Maynard Smith, J. (1966). Sympatric speciation. *The American Naturalist, 100* (916), 637-650.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., . . . DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res, 20* (9), 1297-1303. doi:10.1101/gr.107524.110

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences, 70* (12), 3321-3323. doi: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. doi: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. doi:10.1134/S1022795412010152

Palkopoulou, E., Mallick, S., Skoglund, P., Enk, J., Rohland, N., Li, H., . . . Dalen, L. (2015). Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Curr Biol, 25* (10), 1395-1400. doi:10.1016/j.cub.2015.04.007

Palme, 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. doi:10.1007/s10592-013-0475-x

Parra-Bonilla, G., Alvarez, D. F., Al-Mehdi, A.-B., Alexeyev, M., & Stevens, T. (2010). Critical role for lactate dehydrogenase A in aerobic glycolysis that sustains pulmonary microvascular endothelial cell proliferation. *American journal of physiology. Lung cellular and molecular physiology, 299* (4), L513-L522. doi:10.1152/ajplung.00274.2009

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

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155* (2), 945-959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., . . . 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. doi:https://doi.org/10.1086/519795

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics, 26* (6), 841-842. doi:10.1093/bioinformatics/btq033

R Core Team. (2018). R: A language and environment for statistical computing. Retrieved from https://www.R-project.org/

Ravinet, M., Westram, A., Johannesson, K., Butlin, R., Andre, 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. doi:10.1111/mec.13332

Raymond, M. (1995). GENEPOP (version 1.2) : population genetics software for exact tests and ecumenicism.*J. Hered., 86* , 248-249.

Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology, 29* (1), 24-26. doi: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. doi:10.1111/j.1471-8286.2007.01931.x

Roux, C., Fraisse, 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. doi:10.1371/journal.pbio.2000234

Ryman, N., Allendorf, F. W., & Stahl, 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. doi: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. doi: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. doi:10.1038/nature09060

Schonswetter, P., Lachmayer, M., Lettner, C., Prehsler, D., Rechnitzer, S., Reich, D. S., . . . 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. doi: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. doi:10.1371/journal.pone.0021800

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics, 105* (2), 437-460.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.*Genetics, 123* (3), 585-595.

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. doi: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. doi:10.1023/A:1008955229420

Tian, D., Huang, D., Short, S., Short, M. L., & Jungmann, R. A. (1998). Protein kinase A-regulated instability site in the 3'-untranslated region of lactate dehydrogenase-A subunit mRNA. *The Journal of biological chemistry, 273* (38), 24861-24866. doi:10.1074/jbc.273.38.24861

Turelli, M., Barton, N. H., & Coyne, J. A. (2001). Theory and speciation. *Trends in Ecology & Evolution, 16* (7), 330-343. doi: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* , 005165. doi:10.1101/005165

Utter, F. (2004). Population genetics, conservation and evolution in salmonids and other widely cultured fishes: some perspectives over six decades. *Reviews in Fish Biology and Fisheries, 14* (1), 125-144. doi:10.1007/s11160-004-3768-9

van der Valk, T. (2019).*Genomics of population decline.* (1822 Doctoral thesis, comprehensive summary), Acta Universitatis Upsaliensis, Uppsala. 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. doi:10.1101/248237

Via, S. (2001). Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution, 16* (7), 381-390. doi: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. doi: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. doi: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. doi:10.1111/j.1558-5646.1984.tb05657.x

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* : Springer-Verlag New York.

Wilson, A. J., Gíslason, D., Skúlason, S., Snorrason, S. S., Adams, C. E., Alexander, 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. doi:10.1111/j.1365-294X.2004.02149.x

DATA ACCESSIBILITY

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 xxxx (to be completed after manuscript is accepted for publication).

AUTHOR CONTRIBUTION

L.L., A.A., Sa.Ku., N.R. designed the study; N.R., F.W.A., G.S., L.L. provided the material and allozyme data, St.Ka. provided 96 SNP array genotypes, A.A., A.S., N.R., L.L. analyzed SNP array data, A.S. analyzed genomics data initially instructed by Sa.Ku. and further supervised by V.E.K., D.E., and M.K. O.H and N.R. performed POWSIM theoretical evaluation and supervised simulations performed by A.S. N.L.P.K guided and located the 96 SNPs in the reference genome and performed the*LDH-A* analysis with allozyme guidance from F.W.A. and G.S. A.S. and L.L. led the writing with contribution from all authors. L.L. funded the study.

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 available material from these lakes. These polymorphic loci consist of 77 SNPs from the 96 SNP fluidigm assay and 8 allozyme loci. All loci are bi-allelic 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. n = number of fish,N_{A} = mean number of alleles per locus,H_{E} = expected heterozygosity. Statistically significant F_{IS} values are in bold (p<0.05).

Marker	**Deme**	n	$N_{\mathbf{A}}$	$F_{\mathbf{IS}}$	$H_{\mathbf{E}}$	**Number of polymorphic loci**
SNPs	I	68	1.80	**-0.168**	0.269	77
	II	72	1.25	**0.071**	0.079	24
Allozymes*	I	217	1.88	**0.096**	0.219	7
	II	177	1.63	0.069	0.123	5

*From Ryman et al. 1979 and unpublished data

Table 2 Descriptive statistics from genome-wide Pool-seq data (window size = 5,000 bp and fraction depth

covered [?] 0.8 i.e. a window was only retained if at least 80% of its SNPs had a read depth between 20X and 150X) for each of the two demes in Lakes Bunnersjöarna.π = nucleotide diversity (Tajima, 1983),T_D = Tajima's D (Tajima, 1989), = Watterson's theta (Watterson, 1975).

Statistic	**Statistic**	**Deme**	**Deme**	**Mean (95% CI)**	**No. windows**
π	I	I	0.00130 (0.00128; 0.0013)	0.00130 (0.00128; 0.0013)	0.00130 (0.00128; 0.0013)
	II	II	0.00046 (0.00046; 0.00046)	0.00046 (0.00046; 0.00046)	0.00046 (0.00046; 0.00046)
$T_{\mathbf{D}}$	I	I	-0.22665 (-0.2309; -0.2224)	-0.22665 (-0.2309; -0.2224)	-0.22665 (-0.2309; -0.2224)
	II	II	-1.0148 (-1.0194; -1.0102)	-1.0148 (-1.0194; -1.0102)	-1.0148 (-1.0194; -1.0102)
	I	I	0.001670 (0.00167; 0.00168)	0.001670 (0.00167; 0.00168)	0.001670 (0.00167; 0.00168)
	II	II	0.000789 (0.00079; 0.00079)	0.000789 (0.00079; 0.00079)	0.000789 (0.00079; 0.00079)

Table 3 F_{ST} estimates between demes I and II of Lakes Bunnersjöarna. For Pool-seq data Nei´s (1973)F_{ST} was used with window size = 5,000 bp, read depth = 20X – 150X and fraction depth covered (the fraction to which windows were covered with data) [?] 0.8. For the 96 SNP fluidigm assay, 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). $F_{ST}$was computed using Nei's (1973) and Weir and Cockerham's (1984) approaches.

			Nei's F_{ST}	**Nei's F_{ST}**	**Weir & Cocker-ham's F_{ST}**	**Weir & Cocker-ham's F_{ST}**
Data	**Number of windows**	**Number of loci**	**Mean F_{ST} (95% CI)**	**Min.;max. F_{ST}**	**Mean F_{ST} (95% CI)**	**Min.;max. F_{ST}**
Pool-seq	340,297	8,495,563	0.127 (0.127-0.127)	0.001;0.849	-	-
SNPs	-	77	0.116 (0.097-0.136)	0.001;0.502	0.242 (0.194-0.292)	0.001;0.657
Allozymes	-	7*	0.053 (0.036-0.070)	0.017;0.080	0.091 (0.055-0.113)	0.022;0.126
	-	8**	0.296 (0.069-0.523)	0.017;1.000	0.440 (0.069-0.801)	0.022;1.000

*excluding diagnostic locus *LDH-1* for whichF_{ST} = 1.0 **including diagnostic locus *LDH-1* for whichF_{ST} = 1.0

Figure legends

Figure 1 Map showing the 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 =meters). Our samples were collected in the northern Bunnersjön (N. Bunnersjön;n = 68) and in the southern Bunnersjön (S. Bunnersjön;n =72). For comparison, we also used material from Lake Ånnsjön (shown here) as well as from several other lakes in other areas (see Figure S1a-d).

Figure 2 Membership coefficient (Q) plots showing the assignment probability of individual fish (total n =140) to clusters using 96 SNPs. Membership coefficients (Q) were obtained from STRUCTURE and the most likely number of clusters (K) for the data was estimated using three different *ad hoc* methods: log probability of data (ln$\Pr X$ —K), ΔK , and parsimony index (PI). Panel a) shows K =2, panel b) K =3, and panel c) K =11 as the most likely numbers of K suggested by ΔK , PI, and ln$\Pr X$ —K , respectively.

Each fish is represented by a vertical bar and the order of individuals is the same in all panels, starting with 68 individuals classified to Deme I, followed by 72 individuals classified to Deme II based on genotype in the allozyme locus *LDH-1* .

Figure 3 Individual-based neighbor-joining tree illustrating the genetic relationship among brown trout from Lakes Bunnersjöarna using 70 SNP loci. Individuals marked I vs. II (purple vs. blue color) are classified to belong to Deme I vs. Deme II based on their allozyme*LDH-I* genotype. 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.

Figure 4 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=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 5: 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 analyzed 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.

Figure 6 a) Estimates of total lengths of runs of homozygosity (Ln*ROH*) of three categories and 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.

Figure 7 : The genetic relationships among brown trout populations from the two demes of Lakes Bunnersjöarna and from other lakes in the same geographic region and TreeMix analysis assuming potential migration between lakes in the same geographic area where creek waterways between lakes exist (cf. Figure 1, Figure S1, Appendix S1). The scale indicates the proportion of genetic divergence per unit length of the branch (indicated by the scale length).