

## RESEARCH ARTICLE OPEN ACCESS

# Effects of Different SNP Calling and Sequence Mapping Choices on the Inference of Genetic Architecture Underlying Migration Tendency

Giovanna Mottola<sup>1,2</sup>  | Frank Panitz<sup>3</sup> | Tuomas Leinonen<sup>4</sup> | Alexandre Lemopoulos<sup>1</sup> | Anssi Vainikka<sup>1</sup>

<sup>1</sup>Department of Environmental and Biological Sciences, University of Eastern Finland, Joensuu, Finland | <sup>2</sup>Genomics and Breeding Unit (GEJA), Natural Resources Institute Finland (LUKE), Turku, Finland | <sup>3</sup>Natural Resources Unit (LUVA), Natural Resources Institute Finland (LUKE), Turku, Finland | <sup>4</sup>Genomics and Breeding Unit (GEJA), Natural Resources Institute Finland (LUKE), Helsinki, Finland

**Correspondence:** Giovanna Mottola ([giovanna.mottola@uef.fi](mailto:giovanna.mottola@uef.fi))

**Received:** 6 January 2026 | **Revised:** 14 May 2026 | **Accepted:** 19 May 2026

**Keywords:** brown trout | dDocent | life-history | outlier | reference genome | *Salmo trutta* | stacks

## ABSTRACT

Genome-wide association studies with identification of biologically relevant genes rely on correct mapping of sequence variation. Here, we re-analysed RADseq data from two migration types of brown trout (*Salmo trutta*) from Koutajoki and Oulujoki watersheds by replacing the originally applied Atlantic salmon (*Salmo salar*) reference genome with the later published brown trout reference genome and by testing three alternative bioinformatic pipelines for identifying single nucleotide polymorphisms (SNPs). As expected, the results from population genomics and outlier analyses largely confirmed the original patterns of population structure and divergence, although the number of called SNPs varied between the used bioinformatic pipelines and reference genomes and was surprisingly lower with the conspecific reference genome. While only two SNP outliers were found by all the alternative methods, several other outlier SNPs related to migration differences among the populations were identified. These findings confirm that the choice of the reference genome is not critical for the inference of population structures but can improve the reliability of candidate gene identification in brown trout.

## 1 | Introduction

Conservation of genetic diversity as the ultimate level of biodiversity has become increasingly important in the recent decades as human-driven habitat fragmentation, environmental degradation and overharvesting have led to significant population declines of many iconic species globally (Fischer and Lindenmayer 2007; Frankham 2010; Heinrichs et al. 2016; Hohenlohe et al. 2021; IPBES 2019). Genetic diversity is hierarchically distributed between individuals, subpopulations, metapopulations and species, with important structuring effects from key life-history differences such as those relating to migratory behaviour in brown trout (*Salmo trutta*) (Lemopoulos, Uusi-Heikkilä, Huusko, et al. 2018). While understanding the patterns of genetic variation in the wild is essential for species

conservation, it is likewise important to understand the genetic architecture of ecologically significant life-history variation for the conservation of ecosystem functionality. Salmonids, like *Salmo* spp., *Oncorhynchus* spp. and *Salvelinus* spp., are economically and culturally significant fishes that often form distinctively resident or migratory populations with significant associated differences in several other life-history traits (Vera et al. 2023). Major efforts have been put into the assembly of their reference genomes and identification of genetic regions that could facilitate genomic selection, for example for growth and flesh quality in salmon aquaculture (Houston and Macqueen 2019). Among the first assembled salmonid reference genomes was that of the Atlantic salmon (*Salmo salar*) (Davidson et al. 2010; Lien et al. 2016), which has until recently been used as a reference for the rest of Salmonidae (Davidson 2013; Elias

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Ecology and Evolution* published by British Ecological Society and John Wiley & Sons Ltd.

et al. 2018; Lemopoulos, Uusi-Heikkilä, Huusko, et al. 2018; Lemopoulos, Uusi-Heikkilä, Vasemägi, et al. 2018; Lemopoulos, Uusi-Heikkilä, et al. 2019; Lemopoulos, Prokkola, et al. 2019; Robinson et al. 2017). Since then, reference genomes have been published for brown trout (*Salmo trutta*) (Hansen et al. 2021), rainbow trout (*Oncorhynchus mykiss*) (Gao et al. 2021), Coho salmon (*Oncorhynchus kisutch*) ([https://www.ncbi.nlm.nih.gov/assembly/GCF\\_002021735.2/](https://www.ncbi.nlm.nih.gov/assembly/GCF_002021735.2/)) and European whitefish (*Coregonus lavaretus*) (Gao et al. 2021; Pokharel et al. 2025) with rapid developments in understanding the genetic architecture of multiple traits (Clare et al. 2023; Gallardo-Hidalgo et al. 2021).

Migration tendency, as a trait separating resident populations from migratory (anadromous or potamodromous) populations, is partly driven by inherited variation (Ferguson et al. 2019). Multiple studies on brown trout have shown that migratory tendency is a threshold trait that is proximately affected also by a number of environmental factors, with significant genotype-by-environment ( $G \times E$ ) interactions (Olsson et al. 2006; Ferguson et al. 2016, 2019; Nevoux et al. 2019). Before the publication of the brown trout reference genome (Hansen et al. 2021), genomic studies of brown trout relied on *de novo* assembly or on hetero-specific alignment to the *S. salar* reference genome (Leitwein et al. 2016; Lemopoulos, Uusi-Heikkilä, Huusko, et al. 2018; Lemopoulos, Uusi-Heikkilä, Vasemägi, et al. 2018). The average nucleotide divergence observed between *S. salar* and *S. trutta* is 5.94% for mitochondrial DNA and the estimated average net nucleotide divergence ( $d_n$ ) is 0.0187 for genomic DNA (Bernatchez et al. 1992; Leitwein et al. 2016, 2017). The brown trout studies using *S. salar* reference assembly have identified variants that differentiate hatchery fish from wild ones (Leitwein et al. 2016), as well as genomic regions related to variation in smoltification and osmoregulation functions (Lemopoulos, Uusi-Heikkilä, Huusko, et al. 2018). However, no individual genes with large effects on migratory behaviour have been identified to date leaving it a question whether the trait itself is polygenic or whether some regulatory master genes with large effects have not yet been found (Lemopoulos, Uusi-Heikkilä, et al. 2019).

Reference genomes from phylogenetically close species (i.e., heterospecific mapping) have been valuable in early research of species-specific functional genomics, but conspecific mapping offers higher accuracy and reliability (Shafer et al. 2017) and limits the occurrence of false positives (Manel et al. 2016). The increased quality of the modern genome assemblies obtained through long-read sequencing will also increase the possibilities for genome-wide association studies in new species (Bringlee and Parent 2023). Lemopoulos, Uusi-Heikkilä, Huusko, et al. (2018), used the Atlantic salmon (*S. salar*) reference genome to study genomic associations of migratory behaviour in brown trout. While this study found some candidate SNPs with relevant functions in osmoregulation and immune defence, it left a technical need to re-examine the obtained sequence data with the now available brown trout reference genome (Hansen et al. 2021). Many variant calling pipelines have been developed over the last two decades for handling Restriction-site Associated DNA (RADseq) data (Catchen et al. 2013; Garrison and Marth 2012; Puritz et al. 2014; Rochette and Catchen 2017). Also, the choice of variant-calling pipeline introduces variability in genomic studies, as each pipeline contains unique bioinformatic algorithms and methods (Shafer et al. 2017; Wright et al. 2019).

Lemopoulos, Uusi-Heikkilä, Huusko, et al. (2018) used Stacks v1 (Catchen et al. 2013) that supports single end-reads, demultiplexing, assembling loci within and across individuals and conducting population genetic analyses. The updated version, Stacks v2 (Rochette et al. 2019), adds support for paired-end reads, processes loci across all individuals simultaneously within a meta-population, and uses a Bayesian genotype caller to detect SNPs. Another major software, dDocent (Puritz et al. 2014) also uses paired-end reads throughout and offers flexibility by integrating existing bioinformatics tools, which includes FreeBayes in the entire workflow (Garrison and Marth 2012). FreeBayes, like Stacks v2, is based on a Bayesian approach to variant calling. Both of these pipelines offer high quality SNP calling and have been extensively used for studying population genomics of fish (Brauer et al. 2016; Catchen et al. 2013).

In the present study, our aim was to examine the technical effects of alternative SNP identification methods and determine whether the use of the conspecific reference genome (Hansen et al. 2021) would change the list of candidate genes identified to comparatively explain the migratory differences of brown trout populations (Lemopoulos, Uusi-Heikkilä, Huusko, et al. 2018). To re-estimate the population structure for confirmation (Lemopoulos, Uusi-Heikkilä, Vasemägi, et al. 2018), and to compare migratory and resident populations of brown trout from two watersheds (Lemopoulos, Uusi-Heikkilä, Huusko, et al. 2018), we ran two different SNP-calling pipelines, dDocent and the newest version of Stacks (v2) on the publicly available sequence data. We predicted that the change of the reference genome would not change the inferred population structure as reported by Lemopoulos, Uusi-Heikkilä, Vasemägi, et al. (2018) but it could refine the candidate gene list for brown trout migratory tendency (Lemopoulos, Uusi-Heikkilä, Huusko, et al. 2018).

## 2 | Material and Methods

### 2.1 | Sequence Data

We used published RADseq data (GenBank accession numbers: Koutajoki: PRJNA431174; Oulujoki: SRP125540) from Lemopoulos, Uusi-Heikkilä, Huusko, et al. (2018) and Lemopoulos, Uusi-Heikkilä, Vasemägi, et al. (2018). The RAD-sequenced brown trout samples represented both migratory and resident populations sampled between 2010 and 2015 in two watersheds in North-Eastern Finland (for maps and description of these systems, see Lemopoulos, Uusi-Heikkilä, Huusko, et al. 2018; Lemopoulos, Prokkola, et al. 2019): Koutajoki (resident: MAA, PES, JUU; migratory: OUL, KIT, KUJ) and Oulujoki (resident: VAA, POH, TUH; migratory: OUV; Table 1). The resident individuals were sampled in small tributaries, while the migratory fish represented main stems and a hatchery population in Oulujoki (Figure 1; Table 1).

### 2.2 | Alignment and SNP Calling

To separate the effects of using different reference genomes from the effects caused by the variant-calling pipelines, we applied two distinct SNP calling methods, dDocent v. 2.9.4

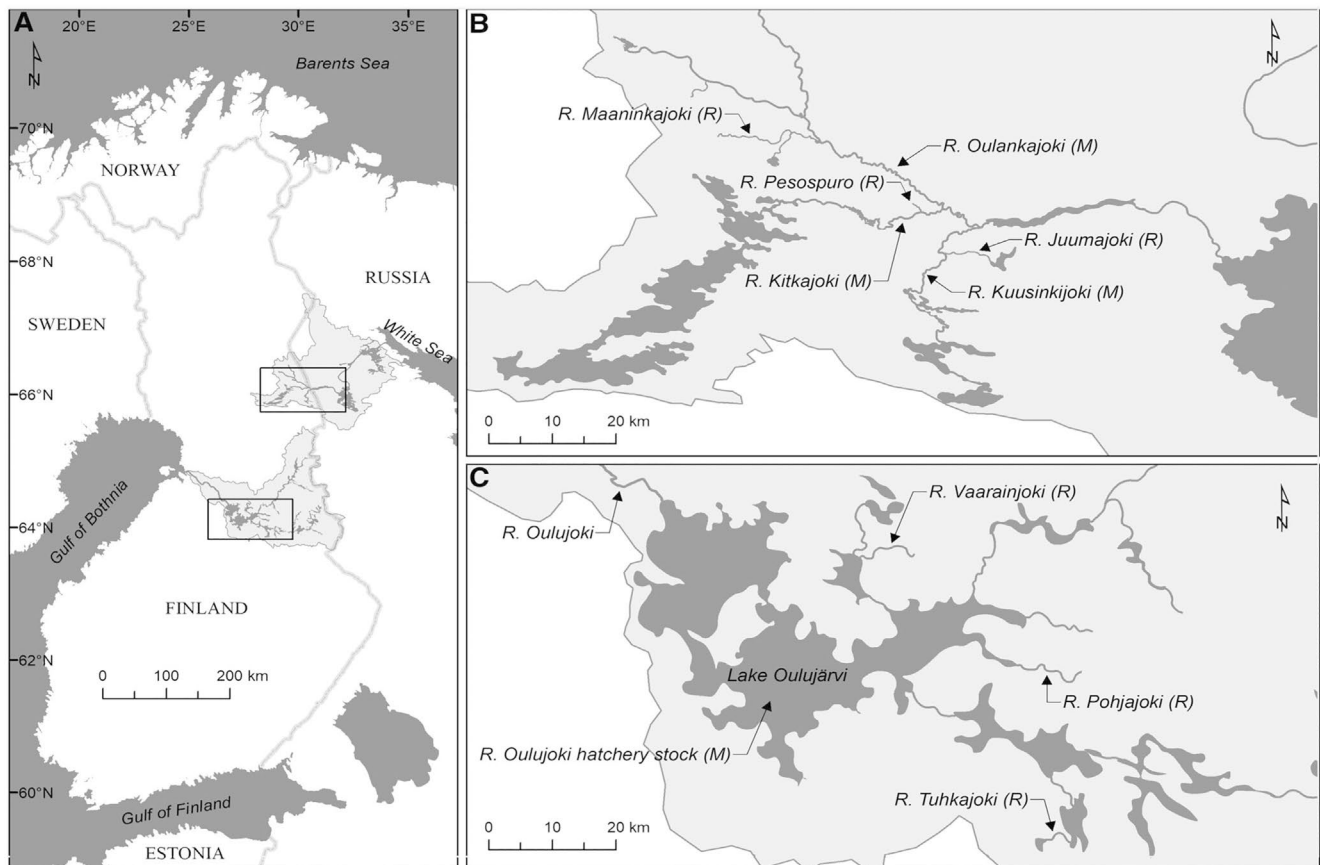
(Puritz et al. 2014) and Stacks v.2.65 (Rochette et al. 2019; Rochette and Catchen 2017), for the both sequence datasets. Since dDocent includes only *bwa* as alignment software and Lemopoulos, Uusi-Heikkilä, Huusko, et al. (2018) used *bowtie2*, we ran Stacks v.2.65 using both *bowtie2* and *bwa* as alignment software to account for variability arising from

this difference. The SNPs obtained using these pipelines were examined for overlap using the command *vcftoolz compare* in *vcftoolz* v. 1.2.3 (Davis 2019).

The Koutajoki library was demultiplexed using *process\_radtags* command present in Stacks v.2.65, while the Oulujoki library was already demultiplexed. Following the protocol in Lemopoulos, Uusi-Heikkilä, Huusko, et al. (2018), for *process\_radtags*, we added the *-inline\_null* argument to specify that the barcodes were inline with sequence. We also added the *-renz\_1* *pstI* and *-renz\_2* *bamHI*, which were the restriction enzymes used for the library preparation. Argument *-q* was added to discard reads with low quality scores while the *-i* argument was added to specify the input file type. Also, the *-c* argument was added to clean data and remove any reads with an uncalled base and *-r* argument to rescue barcodes and RAD-Tag cut sites. Since the barcodes length spanned from 9 to 13bp, we truncated the reads to 85bp by using the argument *-t* 85. Successively, the raw data were fed in dDocent workflow and Stacks. In dDocent, the raw data was aligned to the *S. trutta* reference genome (fSalTru1.1, GCA\_901001165.1) using *bwa* v0.7.17 and the successive indexing was done using *samtools* v1.17. The .bam files obtained were then used to call variants, both SNPs and INDELs, using FreeBayes v1.3.6 (Garrison and Marth 2012). The bioinformatic protocol with Stacks 2.65 followed Lemopoulos, Uusi-Heikkilä, Huusko, et al. (2018). The RADseq raw data were first aligned to the *S. trutta* reference genome (fSalTru1.1, GCA\_901001165.1)

**TABLE 1** | List showing the *N*-value of all the sequenced individuals, their behaviour (migratory or resident) and from which rivers and watershed they were sampled.

| Watershed | River                   | Code | Behaviour | <i>N</i> |
|-----------|-------------------------|------|-----------|----------|
| Koutajoki | Maaninkajoki            | MAA  | R         | 30       |
|           | Pesospuro               | PES  | R         | 30       |
|           | Juumajoki               | JUU  | R         | 30       |
|           | Kitkajoki               | KIT  | M         | 29       |
|           | Oulankajoki             | OUL  | M         | 30       |
|           | Kuusinkijoki            | KUU  | M         | 30       |
| Oulujoki  | Vaarainijoki            | VAA  | R         | 29       |
|           | Pohjajoki               | POH  | R         | 12       |
|           | Tuhkajoki               | TUH  | R         | 11       |
|           | Oulujoki hatchery stock | OUV  | M         | 28       |



**FIGURE 1** | Map showing the study areas in Finland (A) and specifically the Koutajoki (B) and Oulujoki (C) watersheds where fish were collected. The current image is an adaptation of a previous map published in Lemopoulos, Uusi-Heikkilä, Huusko, et al. (2018).

using *bowtie2* and *bwa*. For *bowtie2* the arguments used were *--end-to-end* (involving every character of the reads) and *--very-sensitive* as set previously by Lemopoulos, Uusi-Heikkilä, Huusko, et al. (2018), while the default settings were used when aligning using *bwa mem*. The .sam files obtained from the alignment were indexed and sorted using *samtools* v. 1.21 (Danecek et al. 2021). The resulting .bam files were then fed in the *gstacks*, including the population information as argument, and *population* modules, which were used to call variants. For the population module we followed Lemopoulos, Uusi-Heikkilä, Huusko, et al. (2018), i.e., the variants were called using the following filtering options: *-r 0.6* (*--min-samples-per-pop 0.6*) which indicates that a minimum of 60% individuals in a population should be required to process a locus. *-p 6/4* (*--min-populations 6/4*), which is the minimum number of populations a locus must be present in to process that locus. This parameter was set to 6 populations for Koutajoki and 4 populations for Oulujoki. Additionally, SNPs with a minor allele frequency of at least 0.05 (*--min-maf 0.05*) and maximum heterozygosity of 0.5 (*--max-obs-het 0.5*) were retained.

### 2.3 | Genome Scan, F-Statistics and Structure Analyses

All the datasets were filtered for Hardy–Weinberg equilibrium (HWE) population-wise to remove the SNPs (2868 for Koutajoki *bwa+dDocent*, 2195 for Oulujoki *bwa+dDocent*, 1826 for Koutajoki *bowtie2+Stacks* and 664 for Oulujoki *bowtie2+Stacks*, 1660 for Koutajoki *bwa+Stacks*, 646 for Oulujoki *bwa+Stacks*) departing from HWE. Population-wise HWE filtering was applied to remove loci subjected to genotyping errors without reducing the potential to infer population structures (Pearman et al. 2022).

This filtering was performed using the function *gl.filter.hwe* present in *dartR* (Mijangos et al. 2022) package for R v. 4.2.2 (R Core Team 2023). This function filters out loci showing significant departure from HW proportions based on observed frequencies of reference homozygotes, heterozygotes and alternate homozygotes (Mijangos et al. 2022). Moreover, by adding the *seppop* option available in the *adegenet* v. 2.1.10 (Jombart and Ahmed 2011) package for R v. 4.2.2 (R Core Team 2023) we performed a population-wise filtering that considered SNPs that departed from HWE only within each population. The datasets from the two distinct watersheds (Koutajoki and Oulujoki) were analysed separately for both population structure and outlier loci.

Genome scans were done with two widely used approaches: PCAdapt (v. 4.3.3) (Luu et al. 2017) package for R v. 4.2.2 (R Core Team 2023) and Bayescan (v. 2.1) (Foll and Gaggiotti 2008). An environmental association analysis was also performed by using BayeScEnv (v. 1.1) (de Villemereuil and Gaggiotti 2015). PCAdapt detects outlier loci based on population clustering and has increased statistical power compared to Bayesian models (Bekkevold et al. 2020; Luu et al. 2017). The outliers were detected based on q-values using a cutoff value of  $\alpha = 0.1$  in *qvalue* (v. 2.30.0) (Storey et al. 2015) package for R v. 4.2.2 (R Core Team 2023) and applying a Bonferroni correction for

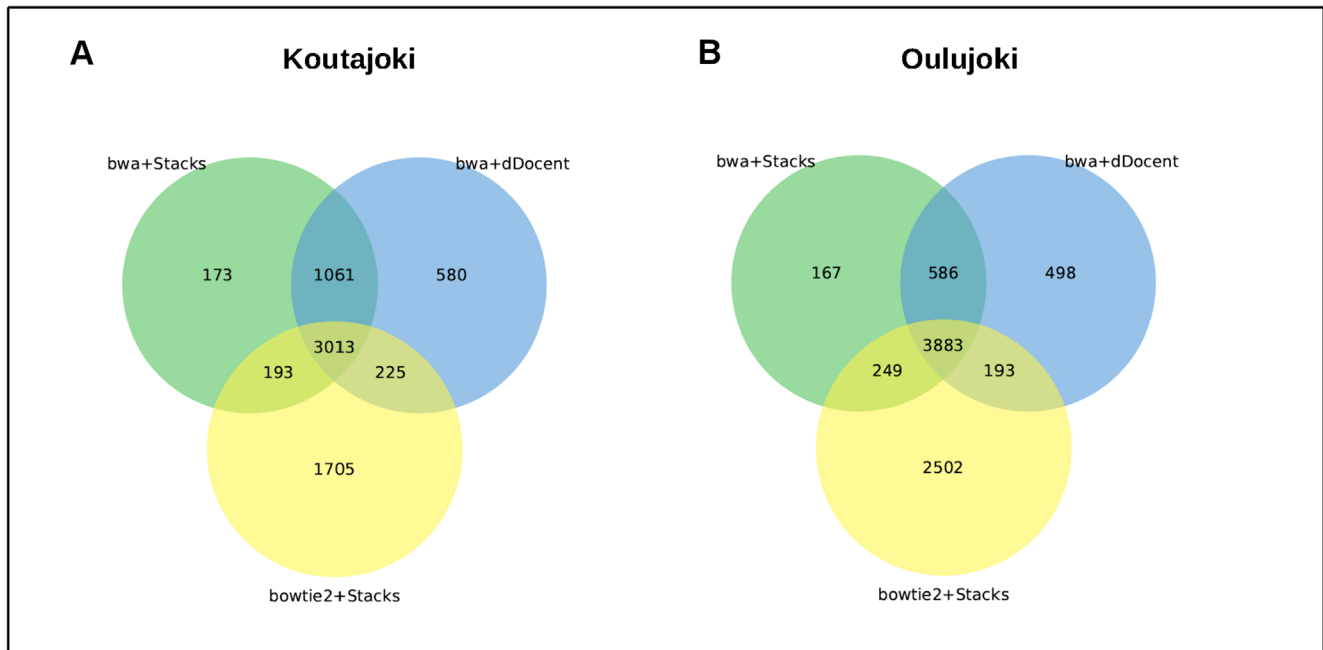
multiple testing. BayeScan identifies candidate loci under natural selection using differences in allele frequencies between populations (Foll and Gaggiotti 2008). The geste file was fed to Bayescan with a burn-in of 50,000 and 50,000 iterations. BayeScEnv aims at detecting local adaptation linked to a given environmental variable using medium- to high-density genotypic data. It considers the population structure, based on the F-model, which is the model behind the BayeScan software. BayeScEnv was also run with 2000 iterations and a burn-in phase of 50,000 iterations. In this case, the environmental variables were relative to the migratory distance of each population assigned as 0.5 for migratory populations and  $-0.5$  for resident populations (Lemopoulos, Uusi-Heikkilä, Huusko, et al. 2018). In the current study, only 1 out of 3 approaches (PCAdapt) identified loci under selection (see Results). Therefore, to have an additional validation we used PCAdapt only on SNPs shared by the SNP calling pipelines. For the variants detected as outliers, we located them in the DNA, determined its effect, and distance from genes using the annotated genome on the Ensembl Variant Effect Predictor (VEP). In VEP, to identify the genes and transcript close or that overlapped each outlier SNP, we consider a Upstream/Downstream distance from the variant of 5 kb. Therefore, we considered only genes present within a total 10 kb window.

Pairwise and global  $F_{st}$ —values (Weir and Cockerham 1984) and their 95%-confidence intervals were calculated using *hierfstat* v. 0.5.11 (Goudet and Jombart 2015) package for R v. 4.2.2 (R Core Team 2023) on the SNP datasets excluding the outlier loci that were detected during the previous genome scan step, as it is likely that the genetic population structure differs between neutral SNPs and those possibly under selection. The Structure analyses were done using the software STRUCTURE v2.3.4 (Pritchard et al. 2000), with a burn-in of 200,000 and 200,000 repetitions according to Lemopoulos, Uusi-Heikkilä, Vasemägi, et al. (2018). We set the maximum population number to six for Koutajoki and to four for Oulujoki. The results obtained from the STRUCTURE software were successively fed into CLUMPAK (Kopelman et al. 2015) with CLUMPP LargeGreedy algorithm and 2000 repeats, and dynamic MCL threshold for similarity scores. The best  $K$  was evaluated using Evanno's Delta  $K$  method (Evanno et al. 2005) (Supporting Information S1–S3). Ancestry proportion was plotted using *pophelper* server (Francis 2017).

## 3 | Results

### 3.1 | SNP Detection and Comparison Between Pipelines

The pipelines identified 4881 and 5164 SNPs for Koutajoki and Oulujoki, respectively using *bwa+dDocent*, 5236 and 6827 for Koutajoki and Oulujoki, respectively using *bowtie2+Stacks* and 4441 and 4890 for Koutajoki and Oulujoki, respectively, using *bwa+Stacks*. After the HWE-filtering step, 2268 and 4632 SNPs were retained for Koutajoki and Oulujoki, respectively using *bwa+dDocent*, 3055 and 4500 SNPs using *bowtie2+Stacks* and 2781 and 4244 using *bwa+Stacks*. In comparison to Stacks 1.40 used in Lemopoulos, Uusi-Heikkilä, Huusko, et al. (2018),



**FIGURE 2** | Venn diagrams obtained using vcftoolz and showing the overlaps among the numbers of SNPs obtained after alignment to *Salmo trutta* reference genome and identified using bwa+dDocent (green), bowtie2+Stacks (light blue) and bwa+Stacks (in orange) for Koutajoki (A) and Oulujoki (B) watersheds.

the number of remaining SNPs was significantly lower (originally 5519 SNPs for Koutajoki and 5670 for Oulujoki). Vcftoolz detected 43% SNPs overlap between the three pipelines for Koutajoki and 48% SNPs overlap between the three pipelines for Oulujoki (Figure 2).

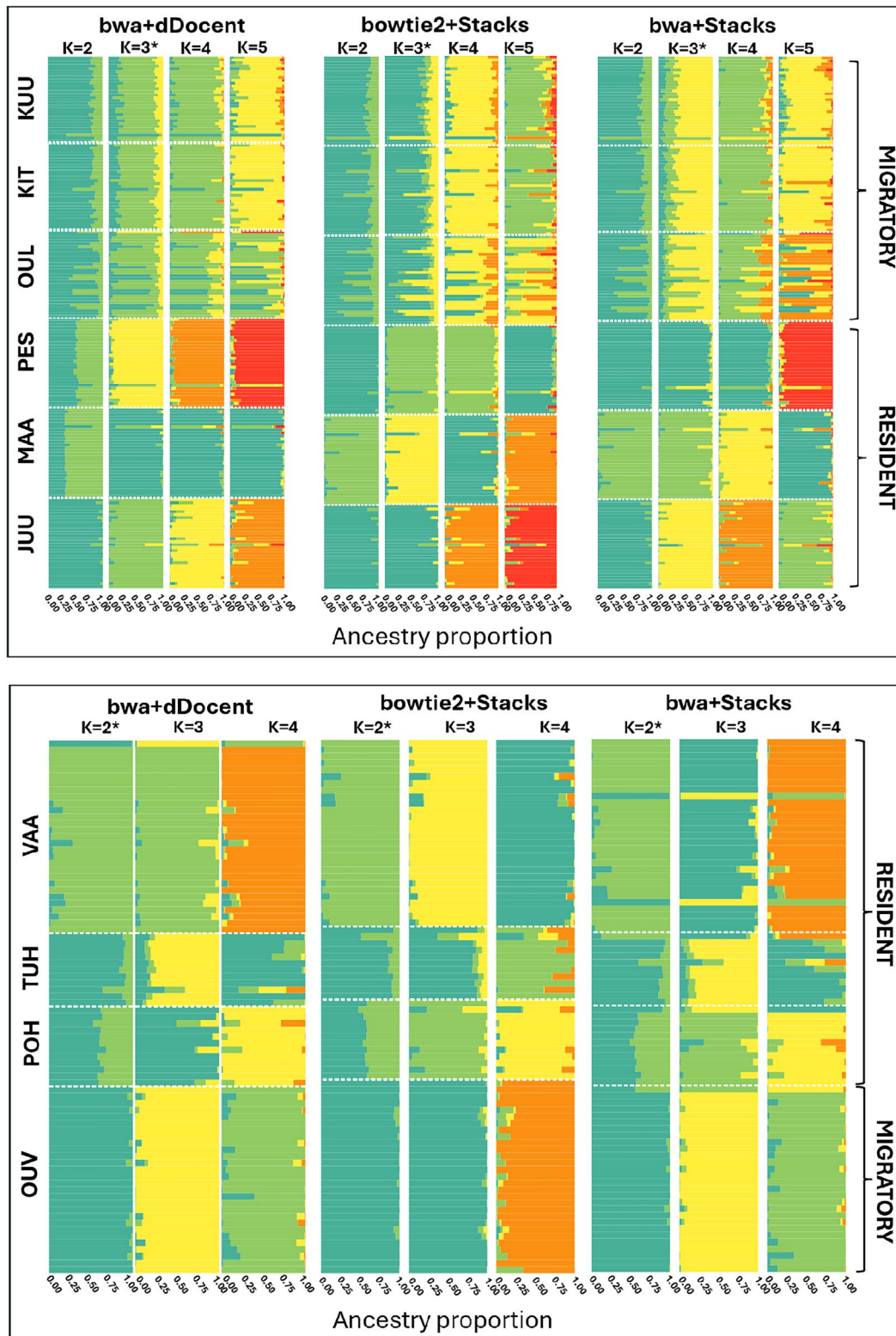
### 3.2 | Genetic Structure and Population Differentiation

Structure analyses confirmed the original distinction between resident and migratory populations in Koutajoki using both dDocent and Stacks identified SNPs (Supporting Information S1–S3). Different ancestry proportions and relative clusters were initially evident with  $K=2$  for the resident population MAA, that showed a clear divergence from the other resident and migratory populations (Figure 3), especially in the dDocent SNP set. Moreover, with  $K=3$  (the most likely  $K$  statistically), it was confirmed that the migratory populations were more similar to each other than to the resident populations (Figure 3).  $K=5$  with dDocent further suggested that KIT and KUU cannot be separated, pointing towards common genetic ancestry of those populations (Figure 3). Likewise, for Oulujoki, both datasets obtained from dDocent and Stacks ( $K=2$ ) (Supporting Information S1–S3) confirmed the documented distinction between the migratory and the resident populations (Figure 3), except for OUV and TUH that seem to share ancestry compared to the other two resident populations.

The divergence of the resident populations from each other and from migratory populations, as well as the lack of divergence among the migratory populations was confirmed by the  $F_{st}$ —values. PES and MAA were the population pair

showing the highest divergence (bwa+dDocent- $F_{st}$ : 0.294, 95% confidence interval = 0.280–0.308, bowtie2+Stacks- $F_{st}$ : 0.282, 95% confidence interval = 0.282–0.307, bwa+Stacks- $F_{st}$ : 0.310, 95% confidence interval = 0.323–0.296; Tables 2–4 and Supporting Information S4A–C), followed by MAA and JUU (bwa+dDocent- $F_{st}$ : 0.251, 95% confidence interval = 0.237–0.264, bowtie2+Stacks- $F_{st}$ : 0.245, 95% confidence interval = 0.234–0.255, bwa+Stacks- $F_{st}$ : 0.259, 95% confidence interval = 0.247–0.270, Tables 2–4 and Supporting Information S4A–C) and by PES and JUU (bwa+dDocent- $F_{st}$ : 0.238, 95% confidence interval = 0.224–0.250, bowtie2+Stacks- $F_{st}$ : 0.238, 95% confidence interval = 0.228–0.249, bwa+Stacks- $F_{st}$ : 0.246, 95% confidence interval = 0.235–0.258; Tables 2–4 and Supporting Information S4A–C). The migratory populations, on the other hand, had in general one order of magnitude smaller  $F_{st}$ -values, with the average  $F_{st}$  among KIT, OUL and KUU being 0.039 for the three bwa+dDocent, bowtie2+Stacks and bwa+Stacks (Tables 2–4 and Supporting Information S4A–C).

Similar trends were shown by the  $F_{st}$ -values calculated on both dDocent and Stacks datasets for Oulujoki, with POH and TUH having the highest  $F_{st}$  (bwa+dDocent- $F_{st}$ : 0.202, 95% confidence interval = 0.193–0.210, bowtie2+Stacks- $F_{st}$ : 0.237, 95% confidence interval = 0.227–0.247, bwa+Stacks- $F_{st}$ : 0.243, 95% confidence interval = 0.234–0.253; Tables 2–4 and Supporting Information S4D–F) and OUV and VAA having a lower  $F_{st}$  in two pipelines (dDocent- $F_{st}$ : 0.155, 95% confidence interval = 0.149–0.161, Stacks- $F_{st}$ : 0.186, 95% confidence interval = 0.180–0.193, bwa+Stacks- $F_{st}$ : 0.178, 95% confidence interval = 0.171–0.185; Tables 2–4 and Supporting Information S4D–F). The lowest  $F_{st}$ -value was found between the migratory population OUV and the resident population TUH (dDocent- $F_{st}$ : 0.100, 95% confidence



**FIGURE 3** | Genetic structure of the studied brown trout populations inhabiting the Koutajoki and Oulujoki watersheds and obtained by running STRUCTURE on bwa+dDocent, bowtie2+Stacks and bwa+Stacks datasets, excluding the outliers. Every column represents a single individual ( $N=179$  and  $N=80$  for Koutajoki and Oulujoki, respectively) and its probability of being assigned to its presumed population. The most probable  $K$  according to Evanno's delta  $K$  method is shown with an asterix.

interval = 0.094–0.105; Stacks- $F_{st}$ : 0.110, 95% confidence interval = 0.104–0.115, bwa+Stacks- $F_{st}$ : 0.115, 95% confidence interval = 0.110–0.121; Tables 2–4 and Supporting Information S4D–F).

### 3.3 | Loci Under Selection

Only PCAdapt identified loci under selection by suggesting 110, 95 and 77 SNPs loci under selection in bwa+dDocent,

**TABLE 2** | Pairwise  $F_{st}$  tables among the Koutajoki (above) and Oulujoki (below) watershed populations and obtained using bwa+dDocent.

| <b>Koutajoki<br/>bwa+dDocent</b> | <b>JUU</b> | <b>KIT</b> | <b>KUU</b> | <b>MAA</b> | <b>OUL</b> |
|----------------------------------|------------|------------|------------|------------|------------|
| KIT                              | 0.138      |            |            |            |            |
| KUU                              | 0.113      | 0.016      |            |            |            |
| MAA                              | 0.251      | 0.188      | 0.169      |            |            |
| OUL                              | 0.117      | 0.058      | 0.043      | 0.153      |            |
| PES                              | 0.238      | 0.175      | 0.163      | 0.294      | 0.165      |
| <b>Oulujoki bwa+dDocent</b>      | <b>OUV</b> | <b>POH</b> | <b>TUH</b> |            |            |
| POH                              | 0.175      |            |            |            |            |
| TUH                              | 0.100      | 0.202      |            |            |            |
| VAA                              | 0.155      | 0.190      | 0.176      |            |            |

Note: The 95%-confidence intervals can be found in the Supporting Information [S4A,D](#).

**TABLE 3** | Pairwise  $F_{st}$  tables among the Koutajoki (above) and Oulujoki (below) watershed populations and obtained using bowtie2+Stacks.

| <b>Koutajoki<br/>bowtie2+Stacks</b> | <b>JUU</b> | <b>KIT</b> | <b>KUU</b> | <b>MAA</b> | <b>OUL</b> |
|-------------------------------------|------------|------------|------------|------------|------------|
| KIT                                 | 0.132      |            |            |            |            |
| KUU                                 | 0.107      | 0.017      |            |            |            |
| MAA                                 | 0.245      | 0.180      | 0.160      |            |            |
| OUL                                 | 0.115      | 0.057      | 0.043      | 0.147      |            |
| PES                                 | 0.238      | 0.179      | 0.162      | 0.294      | 0.171      |
| <b>Oulujoki bowtie2+Stacks</b>      | <b>OUV</b> | <b>POH</b> | <b>TUH</b> |            |            |
| POH                                 | 0.198      |            |            |            |            |
| TUH                                 | 0.110      | 0.237      |            |            |            |
| VAA                                 | 0.186      | 0.225      | 0.211      |            |            |

Note: The 95%-confidence intervals can be found in the Supporting Information [S4B,E](#).

bowtie2+Stacks and bwa+Stacks SNP sets, respectively, in Koutajoki. In Oulujoki, PCAdapt identified 282, 189 and 177 loci under selection in bwa+dDocent, bowtie2+Stacks and bwa+Stacks SNP sets, respectively (Supplementary [S5](#) and [S6](#)). Two of the loci showed general importance by being shared with all the SNP subsets (Figure [4](#)), while other two loci were shared with five out of six datasets and four out of six datasets (Figure [4](#)). Hundred loci, on the other hand, were shared by three SNP subsets (Figure [4](#)). After inspection using VEP, we found that eight of the loci shared at least in 3 datasets were within or very close to candidate genes found to be associated with migratory tendency in previous studies (Table [5](#)). All the rest of the outliers are presented in Supporting Information [S7](#).

**TABLE 4** | Pairwise  $F_{st}$  tables among the Koutajoki (above) and Oulujoki (below) watershed populations and obtained using bwa+Stacks.

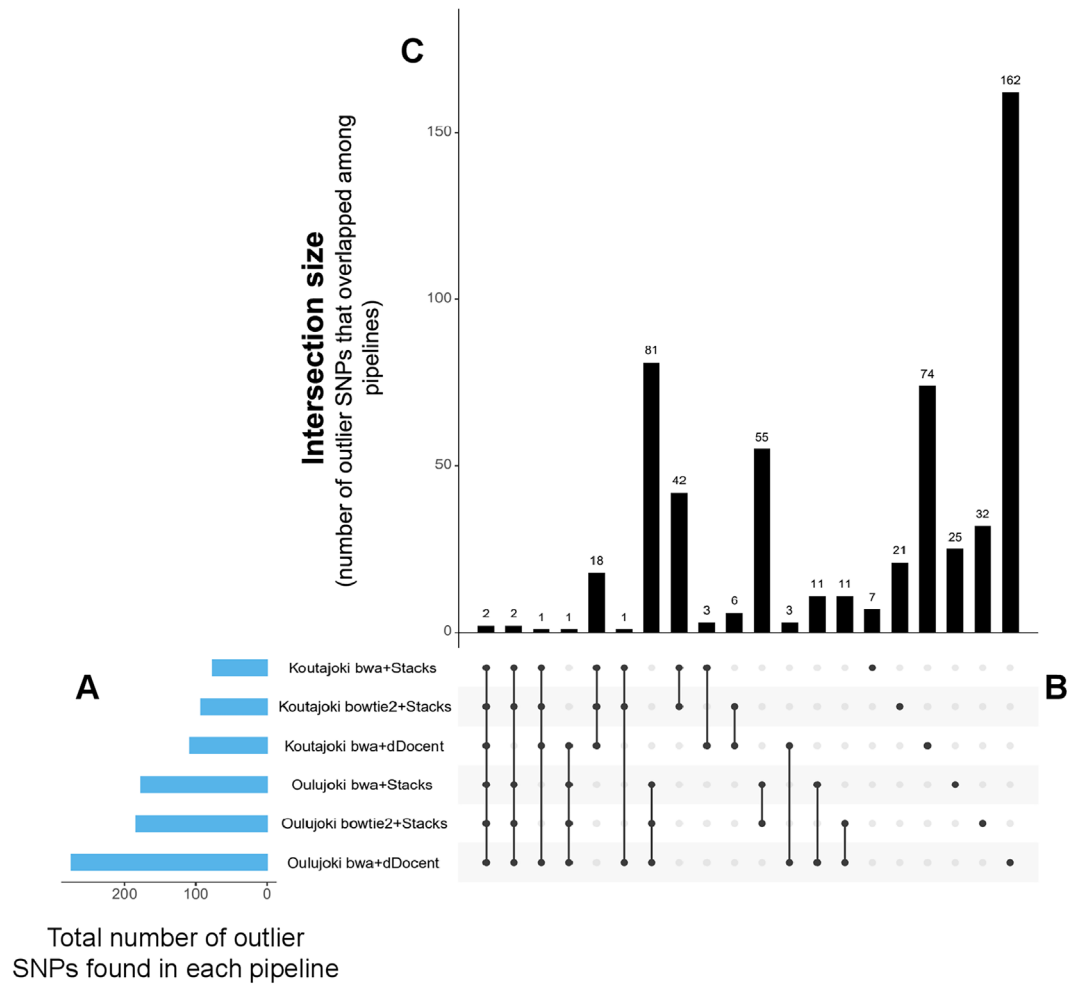
| <b>Koutajoki<br/>bwa+Stacks</b> | <b>JUU</b> | <b>KIT</b> | <b>KUU</b> | <b>MAA</b> | <b>OUL</b> |
|---------------------------------|------------|------------|------------|------------|------------|
| KIT                             | 0.138      |            |            |            |            |
| KUU                             | 0.113      | 0.018      |            |            |            |
| MAA                             | 0.259      | 0.192      | 0.172      |            |            |
| OUL                             | 0.118      | 0.059      | 0.045      | 0.160      |            |
| PES                             | 0.246      | 0.180      | 0.168      | 0.310      | 0.173      |
| <b>Oulujoki bwa+Stacks</b>      | <b>OUV</b> | <b>POH</b> | <b>TUH</b> |            |            |
| POH                             | 0.203      |            |            |            |            |
| TUH                             | 0.115      | 0.243      |            |            |            |
| VAA                             | 0.178      | 0.219      | 0.204      |            |            |

Note: The 95%-confidence intervals can be found in the Supporting Information [S4C,F](#).

## 4 | Discussion

The choice of software for SNP calling had a strong effect on both the number and identity of the detected SNPs, and respectively on the outlier detection results. The new conspecific sequence alignment did not affect the original inferences made on the population structure of the sampled fish (Lemopoulos, Uusi-Heikkilä, Vasemägi, et al. [2018](#)). However, compared to Lemopoulos, Uusi-Heikkilä, Huusko, et al. ([2018](#)), the alignment of sequences on the brown trout reference genome helped to identify new loci that might associate with migratory dichotomy in brown trout. Two of the candidate SNPs were located within a candidate gene region as identified using the salmon reference genome (Lemopoulos, Uusi-Heikkilä, Huusko, et al. [2018](#)). Therefore, our results overall are confirmatory and show that technical choices do matter for the outlier detection (see also Manel et al. [2016](#); Shafer et al. [2017](#)) while for basic population genetics, a set of some thousands of neutral SNPs is enough independently of the technical choices.

The two alternative SNP identification methods yielded different numbers of SNPs, with Stacks generally identifying more SNPs than dDocent when used in combination with bowtie2, for both study systems. This is not surprising due to the inherent differences in the algorithms of these pipelines (Liu et al. [2022](#); Torkamaneh et al. [2016](#)). dDocent uses Burrows-Wheeler Alignment algorithm (*bwa*) (Li and Durbin [2009](#)) for the reference genome alignment. It also detects structural variants (indels) other than SNPs and is therefore particularly suitable for long reads (our average read length was 85 bp after barcode trimming). Stacks offers more flexibility when choosing alignment tools, allowing both *bwa* or *bowtie2*. In our study it seems that the combination bowtie2+Stacks resulted in the highest number of SNPs, while using *bwa* seems a bit more conservative for both the pipelines. Our results showing a different SNP calling success between these pipelines are in line with previous studies (Puritz et al. [2014](#)). However, only about 45% of SNPs were shared between the pipelines, which



**FIGURE 4** | Plot showing the frequency of outlier SNPs that overlapped among different pipelines and watersheds: (A) the horizontal blue bars on the left-hand side represent the number of outlier SNPs found in each dataset using PCAdapt. (B) The vertical connection filled with dark circles represents the pipeline intersection sets. (C) The top vertical bar charts represent the intersection size, i.e., how many outlier SNPs are shared among different pipelines.

further highlights the differences between them. In the present study, downstream analyses of population structure were done using the complete filtered SNP dataset from each pipeline, rather than only the SNPs detected by both, which were used for outlier detection for robustness. Nonetheless, the inferred population structure and divergence in both watersheds were consistent across the datasets from both pipelines. This suggests that, although the pipelines identified different sets of SNPs, both produced enough SNPs to infer population structure and divergence, and as such there may not be a need to compare the analysis pipelines for basic population or landscape genetics (c.f. Shafer et al. 2017). However, the higher number of SNPs obtained with Stacks is reflected in higher statistical power in downstream analyses.

Heterospecific reference genomes should reduce alignment success. For example, by aligning RADseq data from the coho salmon (*Oncorhynchus kisutch*) to six different salmonid reference genomes, including the one of coho salmon itself, Bohling (2020) observed that increasing phylogenetic distance between the target species and the reference genome reduced read alignment success and in turn the number of SNPs detected. Although the Atlantic salmon has been considered

phylogenetically extremely close to the brown trout (Leitwein et al. 2017), these species show differences also in chromosomal arrangements (Leitwein et al. 2017). Against expectations, the use of the brown trout reference genome here resulted in a lower number of SNPs, after filtering, compared to the original study (Lemopoulos, Uusi-Heikkilä, Huusko, et al. 2018) with the Atlantic salmon reference genome. However, this result is confounded by the different SNP calling procedures, as Lemopoulos, Uusi-Heikkilä, Huusko, et al. (2018) used the older Stacks v1. The latest version implements a Bayesian genotype caller (BGC), meaning that SNP genotyping relies on a Bayesian prior before being defined. This procedure might explain the discrepancy. The *S. trutta* reference genome should provide significantly better contiguity ( $N50 = 52.21$  Mb in *S. trutta* vs. 17 Mb in *S. salar*) and structural quality (fewer scaffolds in *S. trutta* compared to *S. salar*), with fewer gaps compared to *S. salar* according to the assembly statistics for each genome. Consistent with our results, Shafer et al. (2017), used ddRADseq data from the Galápagos sea lion (*Zalophus wollebaeki*), and showed that the number of SNPs per locus generally increased as the reference genome became more distantly related. In their study, Shafer et al. (2017) observed that more distant genomes led to low Ts/Tv (transition/transversion) ratio, which suggests errors in variant

**TABLE 5** | List of candidate genes that have been found in 6 (bwa+dDocent Koutajoki, bwa+dDocent Koutajoki, bowtie2+Stacks Koutajoki, bwa+Stacks Koutajoki, bwa+Stacks Koutajoki) and bwa+Stacks Oulujoki), 5 datasets (bwa+dDocent River Koutajoki, bwa+dDocent River Koutajoki, bowtie2+Stacks River Koutajoki, bowtie2+Stacks River Oulujoki) or 4 (bwa+dDocent Oulujoki, bowtie2+Stacks Koutajoki, bowtie2+Stacks Oulujoki) and/or had previously observed to be related to migratory or residency behaviour, and their distances to the outlier SNPs that were detected by both dDocent and Stacks 2.65 pipelines.

| SNP ID            | Location<br>(CHR:POS) | Dataset<br>occurrence | Distance     |  | Gene symbol | Gene name                                                    | Function                                             | Previous studies                                                        | Species                         |
|-------------------|-----------------------|-----------------------|--------------|--|-------------|--------------------------------------------------------------|------------------------------------------------------|-------------------------------------------------------------------------|---------------------------------|
|                   |                       |                       | from<br>gene |  |             |                                                              |                                                      |                                                                         |                                 |
| 1_57,386,539_A/C  | 1:57386539            | 4                     | 2956         |  | ST6GALNAC2  | ST6 N-acetylgalactosaminide<br>alpha-2,6-sialyltransferase 2 | Mucus production in anadromous brown trout           | Malachowicz et al. 2017; Lemopoulos, Uusi-Heikkilä, Huusko, et al. 2018 | Salmo trutta m. trutta          |
| 3_9,902,677_C/T   | 3:9902677             | 6                     | 0            |  | PRDM9-like  | Histone-lysine N-methyltransferase PRDM9-like                | Driver of meiotic recombination hotspot in salmonids | Raynaud et al. 2025                                                     | Onchoryncus mykiss, Salmo salar |
| 3_9,902,717_T/C   | 3:9902717             | 6                     | 0            |  | PRDM9-like  | Histone-lysine N-methyltransferase PRDM9-like                | Driver of meiotic recombination hotspot in salmonids | Raynaud et al. 2025                                                     | Onchoryncus mykiss, Salmo salar |
| 9_16,268,531_C/T  | 9:16268531            | 4                     | 0            |  | TMEM132D    | Transmembrane protein 132D-like                              | Migratory behaviour                                  | Tigano and Russello 2022                                                | Onchoryncus nerka               |
| 13_17,025,604_C/T | 13:17025604           | 4                     | 0            |  | fbxw11a     | F-box/WD repeat-containing protein 7-like                    | Migratory behaviour                                  | Tigano and Russello 2022                                                | Onchoryncus nerka               |
| 16_20,774,328_G/A | 16:20774328           | 4                     | 0            |  | arl8a       | ADP-ribosylation factor-like protein 15                      | Migratory behaviour                                  | Hale et al. 2013                                                        | Onchoryncus mykiss              |
| 16_20,774,328_G/A | 16:20774328           | 4                     | 0            |  | znf76       | Zinc finger protein 76                                       | Smoltification                                       | Lemopoulos, Uusi-Heikkilä, Huusko, et al. 2018                          | Salmo trutta, Salmo salar       |
| 16_59,376,134_G/A | 16:59376134           | 5                     | 0            |  | PTPRT       | Protein tyrosine phosphatase receptor type T                 | Unknown in salmonids or other fish                   | —                                                                       | —                               |
| 21_282,471_G/A    | 21:282471             | 6                     | 0            |  | asap1a      | ArfGAP with SH3 domain, ankyrin repeat and PH domain 1a      | Role in processes relevant to membrane/cytoskeleton  | —                                                                       | —                               |

(Continues)

TABLE 5 | (Continued)

| SNP ID            | Location (CHR:POS) | Dataset occurrence | Distance from gene | Gene symbol | Gene name                                                  | Function                                                      | Previous studies                                                | Species                                                                             |
|-------------------|--------------------|--------------------|--------------------|-------------|------------------------------------------------------------|---------------------------------------------------------------|-----------------------------------------------------------------|-------------------------------------------------------------------------------------|
| 25_33,689,731_C/A | 25:33689731        | 4                  | 689                | CLDN20      | Claudin-20-like                                            | Freshwater vs marine adaptation                               | Ferchaud et al. 2014; Wang et al. 2020                          | <i>Gasterosteus aculeatus</i> ,<br><i>Pungitius pungitius</i>                       |
| 30_27,536,599_C/A | 30:27536599        | 4                  | 0                  | nup210      | Nuclear pore membrane glycoprotein 210-like                | Migratory behaviour                                           | Tigano and Russello 2022                                        | <i>Onchoryncus nerka</i>                                                            |
| 38_6,437,497_C/T  | 38:6437497         | 4                  | 7296               | KCNJ4       | Potassium inwardly rectifying channel subfamily J member 4 | Migratory behaviour and adaptation to saltmarsh microhabitats | Tigano and Russello 2022; Wagner et al. 2017; Moran et al. 2024 | <i>Onchoryncus nerka</i> ,<br><i>Fundulus heteroclitus</i> ,<br><i>Salmo trutta</i> |

Note: Since the SNP “16\_20,774,328\_G/A” was found to be really close to two genes, we listed it twice in the table.

calling, further recommending avoiding alignment on distantly related genomes, but instead to use phylogenetically as close references as possible. Thus, the discrepancy in the number of SNPs obtained here and in Lemopoulos, Uusi-Heikkilä, Huusko, et al. (2018) is likely explained partially by erroneous original fragment alignment and partially by improved Bayesian SNP calling algorithm in the new version of Stacks. Filtering steps between the previous study with alignment on *S. salar* reference and the present study were the same for Stacks v.2 and followed similar filtering criteria despite the usage of different filtering software for dDocent. This suggests that the SNP yield could be influenced by technical choices, and close phylogenetic match should in theory always produce more SNPs and less artefacts. In the absence of a conspecific genome, Atlantic salmon genome was the best possible reference for brown trout in Lemopoulos, Uusi-Heikkilä, Huusko, et al. (2018), and our results here confirm that the choice of the reference genome did not have any relevant influence on the inferred population structure among the studied fish.

In both River Koutajoki and Oulujoki, resident populations were genetically distinct from each other and from the migratory main stem populations, supporting the conclusion of reproductive isolation driven by geographic and behavioural factors (Lemopoulos, Uusi-Heikkilä, Vasemägi, et al. 2018). In contrast, fish from the main stems, i.e., the migratory stocks, showed higher admixture, as seen in the Structure analyses (Figure 3). The F-statistics confirmed these patterns across pipelines, with SNPs identified by Stacks producing generally higher  $F_{st}$ -values than SNPs from dDocent. The clearest difference between the current analyses and those of Lemopoulos, Uusi-Heikkilä, Vasemägi, et al. (2018) was for the River Tuhkajoki (TUH) population that here had similar ancestry proportion as the hatchery stock suggesting potential stocking-origin that was not detected in the original analyses. Similar cases of landscape-driven, as well as a potential behavioural or human induced, population isolation have been commonly observed for brown trout systems (Carlsson and Nilsson 2001; Lagunas et al. 2023). Non-migratory behaviour may reflect local adaptation and drift following reproductive isolation of formerly migratory founders that have colonised the current ranges after glacial recession (García Marín et al. 1999; Huusko et al. 2017; Hynes et al. 1996). Such a process could occur rather independently in each headwater population and explain why universal migratory genes have not been found. Yet, parallel evolution affecting common genetic regions is also possible (Miller et al. 2012; Pearse et al. 2014; Waples et al. 2004). On the other hand, we did not have individual movement data for all the sampled fish, and some tributaries may have included hatchery-produced migratory individuals with confounding effects on the analyses.

Aligning RAD tags to the *S. trutta* reference genome yielded novel outlier SNPs compared to the original study with alignment to the *S. salar* genome (Lemopoulos, Uusi-Heikkilä, Huusko, et al. 2018). Nonetheless, only one approach out of three yielded outliers in both study systems. Why only PCAdapt produced statistically significant outliers may relate to the more conservative approaches used in Bayescan and BayeScEnv (Luu et al. 2017). Moreover, PCAdapt has been shown to be among the most powerful approaches to detect trait-associated genetic variation in the presence of highly divergent populations like in

this study, while Bayescan has been shown to perform poorly when population divergence is high. Bayescan performance has been shown to be highly context-dependent (de Villemereuil and Gaggiotti 2015), especially in strongly hierarchical scenarios with confounding effects between demography and environmental variables. Therefore, its power could be weak, and the error rate could be high. This might be the reason why this approach was not successful in our study context and further highlight the importance of using more than one approach when searching for signals of adaptive divergence. It is also possible that evolutionary processes such as hitchhiking cause differences in detection of loci under selection with different approaches (e.g., Stephan 2019). Bekkevold et al. (2020) similarly showed contrasting results among genome scan approaches in brown trout. Since only one approach (PCAdapt) produced significant outlier loci, we examined only the outliers shared by the three pipelines and the two watersheds for robustness. By locating the outliers on the reference genome, they were found to be spread throughout the whole genome without evidence of blocks of highly divergent loci within few or small genomic regions (Supporting Information S6 and S7). Notably, among these outliers, 8 were found close to genes that have been identified also in previous studies of migration propensity. One SNP on chromosome 1 was found to be 2956 bp from the gene *ST6GALNAC2* (*ST6* N-acetylgalactosaminide alpha-2,6-sialyltransferase 2) that is involved in mucus production in brown trout (Malachowicz et al. 2017) and that was identified also in the original study of Lemopoulos, Uusi-Heikkilä, Huusko, et al. (2018). As fish often encounter a wide variety of microbiota, especially when migrating from one place to another, their immune system needs to differentiate between commensal and pathogenic microbiota to avoid excessive inflammatory responses (McMurtrie et al. 2025). Mucus on the skin and gills might host different microorganisms that contribute to immune defence and healthy skin microbiota, an important function when habitat changes expose fish to potentially new pathogens. Thus, this candidate gene suggests that mucus production plays a role in adaptation to migratory lifestyle, warranting further investigation. Another locus was found within the zinc finger protein 76 gene (*znf76*) confirming results of Lemopoulos, Uusi-Heikkilä, Huusko, et al. (2018). Zinc-finger proteins belong to the wide group of proteins involved in several molecular functions (Cassandri et al. 2017). In salmonids, these class of proteins include immune functions during smoltification by being expressed in liver (Harvey et al. 2024) and gills (Seear et al. 2010) of anadromous Atlantic salmon and in coho salmon (Gallagher et al. 2008), suggesting a similar function in brown trout. Another SNP was found about ~7 Kbp from *KCNJ4* (potassium inwardly rectifying channel subfamily J member 4), also suggested to be involved in migratory behaviour in brown trout (Moran et al. 2024) and other salmonids (Tigano and Russello 2022; Wagner et al. 2017) in relation to processes of acclimatisation to different salinity gradients in many other fishes including salmonids (Gideon et al. 2025; Hale et al. 2013). In our study, migratory brown trout were adfluvial without the need to tolerate saltwater. However, genetic variation tied to osmoregulation indicates that adfluvial fish inherit linked traits that support migration, including their ancestral ability to tolerate salinity gradients (Arostegui et al. 2019; Piironen et al. 2013; Vainikka et al. 2023). Other outliers were found nearby or within genes that have been previously considered as candidate genes

involved in freshwater vs. marine adaptation, again supporting the hypothesis that ancestral capacity to acclimate to different environmental conditions typical of the anadromy have been conserved also in adfluvial populations. Among those we identified *CLDN20*, detected in several studies (Chen and Narum 2021; Ferchaud et al. 2014; Tian et al. 2023; Wang et al. 2020). Some SNPs were found near gene regions previously observed to be involved in migratory behaviour in salmonids (Hale et al. 2013; Tigano and Russello 2022); among those genes we identified *arl8a*, *fbxw11a*, *nup210* and *TMEM132D*. Interestingly, two outliers were found in all the SNP subsets and both within the same gene *PRDM9*-like (histone-lysine N-methyltransferase *PRDM9*-like), which is involved in meiotic recombination in many species and which was recently discovered in salmonids (Raynaud et al. 2025). Other outliers found in all four datasets were within *PTPRT* (protein tyrosine phosphatase receptor type T) and *asap1a* (ArfGAP with SH3 domain, ankyrin repeat and PH domain 1a), whose role in salmonids or other fish is not known. Overall, the fact that several SNPs were in proximity to genes previously found to be involved in migratory/residence behaviour, osmoregulation and growth, suggests that these genes have a role in explaining also the migration dichotomy in the study populations addressed here. In *Onchorhynchus mykiss*, a large chromosomal inversion has been linked to migratory behaviour (Arostegui et al. 2019; Clare et al. 2023), which suggests that a similar pattern in *S. trutta* could not be totally excluded, yet most studies for now generally indicate that variation in migration tendency is shaped by many genes with small effects and that no specific genomic region seems to explain the phenotypic variation.

In conclusion, population structure and divergence patterns remained consistent with the original study (Lemopoulos, Uusi-Heikkilä, Vasemägi, et al. 2018) yet the re-analysis yielded fewer selectively neutral SNPs. SNP detection rate varied between pipelines and reference genomes, highlighting the impact of different bioinformatic methods and reference genomes on SNP detection and further downstream inferences. Different outlier analyses revealed a limited overlap among the genome scan approaches, compared to Lemopoulos, Uusi-Heikkilä, Huusko, et al. (2018), stressing a need for caution when inferring targets of selection. However, we identified several SNPs with known associations to migration, growth, and other biologically relevant traits, suggesting that migratory tendency in brown trout eventually has a multigenic architecture. Since the re-analysis performed here did not improve the original coverage of the RAD-sequencing approach, the genomic architecture of migration still awaits new whole-genome approaches.

#### Author Contributions

**Giovanna Mottola:** data curation (equal), formal analysis (equal), investigation (equal), software (equal), visualization (equal), writing – original draft (equal). **Frank Panitz:** formal analysis (equal), writing – review and editing (equal). **Tuomas Leinonen:** funding acquisition (equal), writing – review and editing (equal). **Alexandre Lemopoulos:** writing – review and editing (equal). **Anssi Vainikka:** conceptualization (equal), funding acquisition (equal), project administration (equal), supervision (equal), writing – review and editing (equal).

## Acknowledgements

We thank Prof. Anti Vasemägi for valuable discussions regarding the statistical analyses presented in this manuscript. We are also grateful to Dr. Luca Pettinau for his advice on data visualisation. Open access publishing facilitated by Ita-Suomen yliopisto, as part of the Wiley - FinELib agreement.

## Funding

This work was funded by Research Council of Finland grant #347367.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Data Availability Statement

All the data and codes used in the current manuscript are stored in Dryad and are publicly available at <https://doi.org/10.5061/dryad.v6wwpzh9f>.

## References

- Arostegui, M. C., T. P. Quinn, L. W. Seeb, J. E. Seeb, and G. J. McKinney. 2019. "Retention of a Chromosomal Inversion From an Anadromous Ancestor Provides the Genetic Basis for Alternative Freshwater Ecotypes in Rainbow Trout." *Molecular Ecology* 28: 1412–1427. <https://doi.org/10.1111/mec.15037>.
- Bekkevold, D., J. Höjesjö, E. E. Nielsen, et al. 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>.
- Bernatchez, L., R. Guyomard, and F. Bonhomme. 1992. "DNA Sequence Variation of the Mitochondrial Control Region Among Geographically and Morphologically Remote European Brown Trout *Salmo trutta* Populations." *Molecular Ecology* 1: 161–173. <https://doi.org/10.1111/j.1365-294X.1992.tb00172.x>.
- Bohling, J. 2020. "Evaluating the Effect of Reference Genome Divergence on the Analysis of Empirical RADseq Datasets." *Ecology and Evolution* 10: 7585–7601. <https://doi.org/10.1002/ece3.6483>.
- Brauer, C. J., M. P. Hammer, and L. B. Beheregaray. 2016. "Riverscape Genomics of a Threatened Fish Across a Hydroclimatically Heterogeneous River Basin." *Molecular Ecology* 25: 5093–5113. <https://doi.org/10.1111/mec.13830>.
- Bringloe, T. T., and G. J. Parent. 2023. "Contrasting New and Available Reference Genomes to Highlight Uncertainties in Assemblies and Areas for Future Improvement: An Example With Monodontid Species." *BMC Genomics* 24: 693. <https://doi.org/10.1186/s12864-023-09779-3>.
- Carlsson, J., and J. Nilsson. 2001. "Effects of Geomorphological Structures on Genetic Differentiation Among Brown Trout Populations in a Northern Boreal River Drainage." *Transactions of the American Fisheries Society* 130: 36–45. [https://doi.org/10.1577/1548-8659\(2001\)130<0036:EOGSOG>2.0.CO;2](https://doi.org/10.1577/1548-8659(2001)130<0036:EOGSOG>2.0.CO;2).
- Cassandri, M., A. Smirnov, F. Novelli, et al. 2017. "Zinc-Finger Proteins in Health and Disease." *Cell Death Discovery* 3: 17071. <https://doi.org/10.1038/cddiscovery.2017.71>.
- Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. "Stacks: An Analysis Tool Set for Population Genomics." *Molecular Ecology* 22: 3124–3140. <https://doi.org/10.1111/mec.12354>.
- Chen, Z., and S. R. Narum. 2021. "Whole Genome Resequencing Reveals Genomic Regions Associated With Thermal Adaptation in Redband Trout." *Molecular Ecology* 30: 162–174. <https://doi.org/10.1111/mec.15717>.
- Clare, C. I., K. M. Nichols, F. P. Thrower, E. A. Berntson, and M. C. Hale. 2023. "Comparative Genomics of Rainbow Trout (*Oncorhynchus mykiss*): Is the Genetic Architecture of Migratory Behavior Conserved Among Populations?" *Ecology and Evolution* 13: e10241. <https://doi.org/10.1002/ece3.10241>.
- Danecek, P., J. K. Bonfield, J. Liddle, et al. 2021. "Twelve Years of SAMtools and BCFtools." *GigaScience* 10: giab008. <https://doi.org/10.1093/gigascience/giab008>.
- Davidson, W. S. 2013. "Understanding Salmonid Biology From the Atlantic Salmon Genome." *Genome* 56: 548–550. <https://doi.org/10.1139/gen-2013-0163>.
- Davidson, W. S., B. F. Koop, S. J. Jones, et al. 2010. "Sequencing the Genome of the Atlantic Salmon (*Salmo salar*)." *Genome Biology* 11: 403. <https://doi.org/10.1186/gb-2010-11-9-403>.
- Davis, S. 2019. "VcfToolz: A Python Package for Comparing and Evaluating Variant Call Format Files." *Journal of Open Source Software* 4: 1144. <https://doi.org/10.21105/joss.01144>.
- de Villemereuil, P., and O. E. Gaggiotti. 2015. "A New FST-Based Method to Uncover Local Adaptation Using Environmental Variables." *Methods in Ecology and Evolution* 6: 1248–1258. <https://doi.org/10.1111/2041-210X.12418>.
- Elias, A., R. McLaughlin, R. Mackereth, C. Wilson, and K. M. Nichols. 2018. "Population Structure and Genomic Variation of Ecological Life History Diversity in Wild-Caught Lake Superior Brook Trout, *Salvelinus fontinalis*." *Journal of Great Lakes Research* 44: 1373–1382. <https://doi.org/10.1016/j.jglr.2018.08.006>.
- Evanno, G., S. Regnaut, and J. Goudet. 2005. "Detecting the Number of Clusters of Individuals Using the Software Structure: A Simulation Study." *Molecular Ecology* 14: 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>.
- Ferchaud, A.-L., S. H. Pedersen, D. Bekkevold, J. Jian, Y. Niu, and M. M. Hansen. 2014. "A Low-Density SNP Array for Analyzing Differential Selection in Freshwater and Marine Populations of Threespine Stickleback (*Gasterosteus aculeatus*)." *BMC Genomics* 15: 867. <https://doi.org/10.1186/1471-2164-15-867>.
- Ferguson, A., T. E. Reed, T. F. Cross, P. McGinnity, and P. A. Prodöhl. 2019. "Anadromy, Potamodromy and Residency in Brown Trout *Salmo trutta*: The Role of Genes and the Environment." *Journal of Fish Biology* 95: 692–718. <https://doi.org/10.1111/jfb.14005>.
- Ferguson, A., T. E. Reed, P. McGinnity, and P. Prodöhl. 2016. "Anadromy in Brown Trout (*Salmo trutta*): A Review of the Relative Roles of Genes and Environmental Factors and the Implications for Management and Conservation." In *Sea Trout: From Science to Management, Proceedings of the 2nd International Sea Trout Symposium*, edited by G. Harris, 1–40. Dundalk, Ireland, 20–22 October. isbn: 9781788035354.
- Fischer, J., and D. B. Lindenmayer. 2007. "Landscape Modification and Habitat Fragmentation: A Synthesis." *Global Ecology and Biogeography* 16: 265–280. <https://doi.org/10.1111/j.1466-8238.2007.00287.x>.
- Foll, M., and O. Gaggiotti. 2008. "A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective." *Genetics* 180: 977–993. <https://doi.org/10.1534/genetics.108.092221>.
- Francis, R. M. 2017. "Pophelper: An R Package and Web App to Analyse and Visualize Population Structure." *Molecular Ecology Resources* 17, no. 1: 27–32. <https://doi.org/10.1111/1755-0998.12509>.
- Frankham, R. 2010. "Challenges and Opportunities of Genetic Approaches to Biological Conservation." *Biological Conservation* 143: 1919–1927. <https://doi.org/10.1016/j.biocon.2010.05.011>.
- Gallagher, E. P., H. M. LaVire, T. K. Bammler, P. L. Stapleton, R. P. Beyer, and F. M. Farin. 2008. "Hepatic Expression Profiling in Smolting and Adult Coho Salmon (*Oncorhynchus kisutch*)." *Environmental Research* 106: 365–378. <https://doi.org/10.1016/j.envres.2007.10.001>.

- Gallardo-Hidalgo, J., A. Barriá, G. M. Yoshida, and J. M. Yáñez. 2021. "Genetics of Growth and Survival Under Chronic Heat Stress and Trade-Offs With Growth- and Robustness-Related Traits in Rainbow Trout." *Aquaculture* 531: 735685. <https://doi.org/10.1016/j.aquaculture.2020.735685>.
- Gao, G., S. Magadan, G. C. Waldbieser, et al. 2021. "A Long Reads-Based de-Novo Assembly of the Genome of the Arlee Homozygous Line Reveals Chromosomal Rearrangements in Rainbow Trout." *G3: Genes, Genomes, Genetics* 11: jkab052. <https://doi.org/10.1093/g3journal/jkab052>.
- García Marín, J.-L., F. M. Utter, and C. Pla. 1999. "Postglacial Colonization of Brown Trout in Europe Based on Distribution of Allozyme Variants." *Heredity* 82: 46–56. <https://doi.org/10.1046/j.1365-2540.1999.00447.x>.
- Garrison, E., and G. Marth. 2012. "Haplotype-Based Variant Detection From Short-Read Sequencing." *arXiv preprint arXiv:1207.3907* [q-bio.GN] 2012. <https://doi.org/10.48550/arXiv.1207.3907>.
- Gideon, S., B. Boyd, B. Ramirez Sierra, et al. 2025. "Voltage-Gated Ion Channels in Cultured Gill Epithelia of Rainbow Trout, *Oncorhynchus mykiss*, Change in Transcript Abundance With Exposure to Freshwater." *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology* 304: 111835. <https://doi.org/10.1016/j.cbpa.2025.111835>.
- Goudet, J., and T. Jombart. 2015. *Estimation and Tests of Hierarchical F-Statistics*. R Core Team 58.
- Hale, M. C., F. P. Thrower, E. A. Berntson, M. R. Miller, and K. M. Nichols. 2013. "Evaluating Adaptive Divergence Between Migratory and Nonmigratory Ecotypes of a Salmonid Fish, *Oncorhynchus mykiss*." *G3: Genes, Genomes, Genetics* 3: 1273–1285. <https://doi.org/10.1534/g3.113.006817>.
- Hansen, T., P. G. Fjelldal, S. Lien, et al. 2021. "The Genome Sequence of the Brown Trout, *Salmo trutta* Linnaeus 1758." *Wellcome Open Research* 6: 108. <https://doi.org/10.12688/wellcomeopenres.16838.1>.
- Harvey, T. N., G. B. Gillard, L. L. Røsaeg, et al. 2024. "The Genome Regulatory Landscape of Atlantic Salmon Liver Through Smoltification." *PLoS One* 19: e0302388. <https://doi.org/10.1371/journal.pone.0302388>.
- Heinrichs, J. A., D. J. Bender, and N. H. Schumaker. 2016. "Habitat Degradation and Loss as Key Drivers of Regional Population Extinction." *Ecological Modelling* 335: 64–73. <https://doi.org/10.1016/j.ecolmodel.2016.05.009>.
- Hohenlohe, P. A., W. C. Funk, and O. P. Rajora. 2021. "Population Genomics for Wildlife Conservation and Management." *Molecular Ecology* 30: 62–82. <https://doi.org/10.1111/mec.15720>.
- Houston, R. D., and D. J. Macqueen. 2019. "Atlantic Salmon (*Salmo salar* L.) Genetics in the 21st Century: Taking Leaps Forward in Aquaculture and Biological Understanding." *Animal Genetics* 50: 3–14. <https://doi.org/10.1111/age.12748>.
- Huusko, A., A. Vainikka, J. t. Syrjänen, P. Orell, P. Louhi, and T. Vehanen. 2017. "Life-History of the Adfluvial Brown Trout (*Salmo trutta* L.) in Eastern Fennoscandia." In *Brown Trout*, 267–295. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119268352.ch12>.
- Hynes, R. A., A. Ferguson, and M. A. McCann. 1996. "Variation in Mitochondrial DNA and Post-Glacial Colonization of North Western Europe by Brown Trout." *Journal of Fish Biology* 48: 54–67. <https://doi.org/10.1111/j.1095-8649.1996.tb01418.x>.
- IPBES. 2019. *Summary for Policymakers of the Global Assessment Report on Biodiversity and Ecosystem Services*. Zenodo. <https://doi.org/10.5281/zenodo.3553579>.
- Jombart, T., and I. Ahmed. 2011. "Adegenet 1.3-1: New Tools for the Analysis of Genome-Wide SNP Data." *Bioinformatics* 27: 3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>.
- Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg, and I. Mayrose. 2015. "Clumpak: A Program for Identifying Clustering Modes and Packaging Population Structure Inferences Across K." *Molecular Ecology Resources* 15: 1179–1191. <https://doi.org/10.1111/1755-0998.12387>.
- Lagunas, M., A. Pálsson, B. Jónsson, M. Jóhannsson, Z. O. Jónsson, and S. S. Snorrason. 2023. "Genetic Structure and Relatedness of Brown Trout (*Salmo trutta*) Populations in the Drainage Basin of the Ölfusá River, South-Western Iceland." *PeerJ* 11: e15985. <https://doi.org/10.7717/peerj.15985>.
- Leitwein, M., P.-A. Gagnaire, E. Desmarais, et al. 2016. "Genome-Wide Nucleotide Diversity of Hatchery-Reared Atlantic and Mediterranean Strains of Brown Trout *Salmo trutta* Compared to Wild Mediterranean Populations." *Journal of Fish Biology* 89: 2717–2734. <https://doi.org/10.1111/jfb.13131>.
- Leitwein, M., B. Guinand, J. Pouzadoux, E. Desmarais, P. Berrebi, and P.-A. Gagnaire. 2017. "A Dense Brown Trout (*Salmo trutta*) Linkage Map Reveals Recent Chromosomal Rearrangements in the *Salmo* Genus and the Impact of Selection on Linked Neutral Diversity." *G3: Genes, Genomes, Genetics* 7: 1365–1376. <https://doi.org/10.1534/g3.116.038497>.
- Lemopoulos, A., J. M. Prokko, S. Uusi-Heikkilä, et al. 2019. "Comparing RADseq and Microsatellites for Estimating Genetic Diversity and Relatedness — Implications for Brown Trout Conservation." *Ecology and Evolution* 9: 2106–2120. <https://doi.org/10.1002/ece3.4905>.
- Lemopoulos, A., S. Uusi-Heikkilä, A. Huusko, A. Vasemägi, and A. Vainikka. 2018. "Comparison of Migratory and Resident Populations of Brown Trout Reveals Candidate Genes for Migration Tendency." *Genome Biology and Evolution* 10: 1493–1503. <https://doi.org/10.1093/gbe/evyl02>.
- Lemopoulos, A., S. Uusi-Heikkilä, P. Hyvärinen, et al. 2019. "Association Mapping Based on a Common-Garden Migration Experiment Reveals Candidate Genes for Migration Tendency in Brown Trout." *G3: Genes, Genomes, Genetics* 9: 2887–2896. <https://doi.org/10.1534/g3.119.400369>.
- Lemopoulos, A., S. Uusi-Heikkilä, A. Vasemägi, A. Huusko, H. Kokko, and A. Vainikka. 2018. "Genome-Wide Divergence Patterns Support Fine-Scaled Genetic Structuring Associated With Migration Tendency in Brown Trout." *Canadian Journal of Fisheries and Aquatic Sciences* 75: 1680–1692. <https://doi.org/10.1139/cjfas-2017-0014>.
- Li, H., and R. Durbin. 2009. "Fast and Accurate Short Read Alignment With Burrows–Wheeler Transform." *Bioinformatics* 25: 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Lien, S., B. F. Koop, S. R. Sandve, et al. 2016. "The Atlantic Salmon Genome Provides Insights Into Rediploidization." *Nature* 533: 200–205. <https://doi.org/10.1038/nature17164>.
- Liu, J., Q. Shen, and H. Bao. 2022. "Comparison of Seven SNP Calling Pipelines for the Next-Generation Sequencing Data of Chickens." *PLoS One* 17: e0262574. <https://doi.org/10.1371/journal.pone.0262574>.
- Luu, K., E. Bazin, and M. G. B. Blum. 2017. "Pcadapt: An R Package to Perform Genome Scans for Selection Based on Principal Component Analysis." *Molecular Ecology Resources* 17: 67–77. <https://doi.org/10.1111/1755-0998.12592>.
- Malachowicz, M., R. Wenne, and A. Burzynski. 2017. "De Novo Assembly of the Sea Trout (*Salmo trutta* m. *Trutta*) Skin Transcriptome to Identify Putative Genes Involved in the Immune Response and Epidermal Mucus Secretion." *PLoS One* 12: e0172282. <https://doi.org/10.1371/journal.pone.0172282>.
- Manel, S., C. Perrier, M. Prati-long, et al. 2016. "Genomic Resources and Their Influence on the Detection of the Signal of Positive Selection in Genome Scans." *Molecular Ecology* 25: 170–184. <https://doi.org/10.1111/mec.13468>.
- McMurtrie, J., A. G. Bell, J. Cable, B. Temperton, and C. R. Tyler. 2025. "The Ecology and Plasticity of Fish Skin and Gill Microbiomes: Seeking

- What Matters in Health and Disease." *FEMS Microbiology Reviews* 49: fuaf027. <https://doi.org/10.1093/femsre/fuaf027>.
- Mijangos, J. L., B. Gruber, O. Berry, C. Pacioni, and A. Georges. 2022. "dartR v2: An Accessible Genetic Analysis Platform for Conservation, Ecology and Agriculture." *Methods in Ecology and Evolution* 13: 2150–2158. <https://doi.org/10.1111/2041-210X.13918>.
- Miller, M. R., J. P. Brunelli, P. A. Wheeler, et al. 2012. "A Conserved Haplotype Controls Parallel Adaptation in Geographically Distant Salmonid Populations." *Molecular Ecology* 21: 237–249. <https://doi.org/10.1111/j.1365-294X.2011.05305.x>.
- Moran, P. A., T. J. Colgan, K. P. Phillips, J. Coughlan, P. McGinnity, and T. E. Reed. 2024. "Whole-Genome Resequencing Reveals Polygenic Signatures of Directional and Balancing Selection on Alternative Migratory Life Histories." *Molecular Ecology* 33: e17538.
- Nevoux, M., B. Finstad, J. G. Davidsen, et al. 2019. "Environmental Influences of Life History Strategies in Partial Anadromous Brown Trout (*Salmo trutta*, Salmonidae)." *Fish and Fisheries* 20: 1–32. <https://doi.org/10.1111/faf.12396>.
- Olsson, C. I., L. A. Greenberg, E. Bergman, and K. Wysujack. 2006. "Environmentally Induced Migration: The Importance of Food." *Ecology Letters* 9: 645–651. <https://doi.org/10.1111/j.1461-0248.2006.00909.x>.
- Pearman, W. S., L. Urban, and A. Alexander. 2022. "Commonly Used Hardy–Weinberg Equilibrium Filtering Schemes Impact Population Structure Inferences Using RADseq Data." *Molecular Ecology Resources* 22: 2599–2613. <https://doi.org/10.1111/1755-0998.13646>.
- Pearse, D. E., M. R. Miller, A. Abadía-Cardoso, and J. C. Garza. 2014. "Rapid Parallel Evolution of Standing Variation in a Single, Complex, Genomic Region Is Associated With Life History in Steelhead/Rainbow Trout." *Proceedings of the Royal Society B: Biological Sciences* 281: 20140012. <https://doi.org/10.1098/rspb.2014.0012>.
- Piironen, J., P. Kiiskinen, H. Huuskonen, M. Heikura-Ovaskainen, and M. Vornanen. 2013. "Comparison of Smoltification in Atlantic Salmon (*Salmo salar*) From Anadromous and Landlocked Populations Under Common Garden Conditions." *Annales Zoologici Fennici* 50, no. 1/2: 1–15. <http://www.jstor.org/stable/23737172>.
- Pokharel, K., D. Fischer, T. Iso-Touru, et al. 2025. "The Genome Assembly of the Farmed European Whitefish *Coregonus lavaretus* L. From the Finnish Selective Breeding Programme." *BMC Genomic Data* 26: 22. <https://doi.org/10.1186/s12863-025-01313-6>.
- Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. "Inference of Population Structure Using Multilocus Genotype Data." *Genetics* 155: 945–959. <https://doi.org/10.1093/genetics/155.2.945>.
- Puritz, J. B., C. M. Hollenbeck, and J. R. Gold. 2014. "dDocent: A RADseq, Variant-Calling Pipeline Designed for Population Genomics of Non-Model Organisms." *PeerJ* 2: e431. <https://doi.org/10.7717/peerj.431>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Raynaud, M., P. Sanna, J. Joseph, et al. 2025. "PRDM9 Drives the Location and Rapid Evolution of Recombination Hotspots in Salmonid Fish." *PLoS Biology* 23: e3002950. <https://doi.org/10.1371/journal.pbio.3002950>.
- Robinson, N. A., G. Timmerhaus, M. Baranski, Ø. Andersen, H. Takle, and A. Krasnov. 2017. "Training the Salmon's Genes: Influence of Aerobic Exercise, Swimming Performance and Selection on Gene Expression in Atlantic Salmon." *BMC Genomics* 18: 971. <https://doi.org/10.1186/s12864-017-4361-7>.
- Rochette, N. C., and J. M. Catchen. 2017. "Deriving Genotypes From RAD-Seq Short-Read Data Using Stacks." *Nature Protocols* 12: 2640–2659. <https://doi.org/10.1038/nprot.2017.123>.
- Rochette, N. C., A. G. Rivera-Colón, and J. M. Catchen. 2019. "Stacks 2: Analytical Methods for Paired-End Sequencing Improve RADseq-Based Population Genomics." *Molecular Ecology* 28: 4737–4754. <https://doi.org/10.1111/mec.15253>.
- Seear, P. J., S. N. Carmichael, R. Talbot, J. B. Taggart, J. E. Bron, and G. E. Sweeney. 2010. "Differential Gene Expression During Smoltification of Atlantic Salmon (*Salmo salar* L.): A First Large-Scale Microarray Study." *Marine Biotechnology* 12: 126–140. <https://doi.org/10.1007/s10126-009-9218-x>.
- Shafer, A. B. A., C. R. Peart, S. Tusso, et al. 2017. "Bioinformatic Processing of RAD-Seq Data Dramatically Impacts Downstream Population Genetic Inference." *Methods in Ecology and Evolution* 8: 907–917. <https://doi.org/10.1111/2041-210X.12700>.
- Stephan, W. 2019. "Selective Sweeps." *Genetics* 211, no. 1: 5–13.
- Storey, J., A. Bass, A. Dabney, and D. Robinson. 2015. "Package 'qvalue'."
- Tian, F., B. Zhou, X. Li, et al. 2023. "Population Genomics Analysis to Identify Ion and Water Transporter Genes Involved in the Adaptation of Tibetan Naked Carps to Brackish Water." *International Journal of Biological Macromolecules* 247: 125605. <https://doi.org/10.1016/j.ijbio.2023.125605>.
- Tigano, A., and M. A. Russello. 2022. "The Genomic Basis of Reproductive and Migratory Behaviour in a Polymorphic Salmonid." *Molecular Ecology* 31: 6588–6604. <https://doi.org/10.1111/mec.16724>.
- Torkamaneh, D., J. Laroche, and F. Belzile. 2016. "Genome-Wide SNP Calling From Genotyping by Sequencing (GBS) Data: A Comparison of Seven Pipelines and Two Sequencing Technologies." *PLoS One* 11: e0161333. <https://doi.org/10.1371/journal.pone.0161333>.
- Vainikka, A., C. Elvidge, J. M. Prokkola, et al. 2023. "Two-Generation Common-Garden Experiment Reveals a Strong Genetic Contribution to Migration Tendency in Brown Trout (*Salmo trutta*)." *Canadian Journal of Fisheries and Aquatic Sciences* 80: 1394–1409. <https://doi.org/10.1139/cjfas-2022-0304>.
- Vera, M., E. Aparicio, S. Heras, et al. 2023. "Regional Environmental and Climatic Concerns on Preserving Native Gene Pools of a Least Concern Species: Brown Trout Lineages in Mediterranean Streams." *Science of the Total Environment* 862: 160739. <https://doi.org/10.1016/j.scitotenv.2022.160739>.
- Wagner, D. N., T. Z. Baris, D. I. Dayan, X. Du, M. F. Oleksiak, and D. L. Crawford. 2017. "Fine-Scale Genetic Structure due to Adaptive Divergence Among Microhabitats." *Heredity* 118: 594–604. <https://doi.org/10.1038/hdy.2017.6>.
- Wang, Y., Y. Zhao, Y. Wang, Z. Li, B. Guo, and J. Merilä. 2020. "Population Transcriptomics Reveals Weak Parallel Genetic Basis in Repeated Marine and Freshwater Divergence in Nine-Spined Sticklebacks." *Molecular Ecology* 29: 1642–1656. <https://doi.org/10.1111/mec.15435>.
- Waples, R. S., D. J. Teel, J. M. Myers, and A. R. Marshall. 2004. "Life-History Divergence in Chinook Salmon: Historic Contingency and Parallel Evolution." *Evolution* 58: 386–403. <https://doi.org/10.1111/j.0014-3820.2004.tb01654.x>.
- Weir, B. S., and C. C. Cockerham. 1984. "Estimating F-Statistics for the Analysis of Population Structure." *Evolution* 38: 1358–1370. <https://doi.org/10.2307/2408641>.
- Wright, B., K. A. Farquharson, E. A. McLennan, K. Belov, C. J. Hogg, and C. E. Grueber. 2019. "From Reference Genomes to Population Genomics: Comparing Three Reference-Aligned Reduced-Representation Sequencing Pipelines in Two Wildlife Species." *BMC Genomics* 20: 453. <https://doi.org/10.1186/s12864-019-5806-y>.

## Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Supporting Information: S1.**  $\Delta K$  statistics, calculated using Evanno's method, applied to determine the most likely

number of genetic clusters ( $K$ ) within River Koutajoki (graphs above) and Oulujoki (graphs below), using SNPs obtained by running bwa+dDocent pipeline after analyzing them using STRUCTURE software.

**Supporting Information: S2**  $\Delta K$  statistics, calculated using Evanno's method, applied to determine the most likely number of genetic clusters ( $K$ ) within the River Koutajoki and Oulujoki, using SNPs obtained by running bowtie2+Stacks pipeline after analyzing them using STRUCTURE software.

**Supporting Information: S3**  $\Delta K$  statistics, calculated using Evanno's method, applied to determine the most likely number of genetic clusters ( $K$ ) within the River Koutajoki and Oulujoki, using SNPs obtained by running bwa+Stacks pipeline after analyzing them using STRUCTURE software.

**Supporting Information: S4**  $F_{st}$ , 95%-confidence intervals for the pairwise  $F_{st}$  (lower 95% CIs on the lower diagonal, upper 95% CIs on the upper diagonal).

**Supporting Information: S5** Venn Diagrams showing the intersection among three different genome scan approaches (PCAdapt, Bayescan, BayeScEnv) to detect loci under selection using SNPs obtained from bwa+dDocent (in green), bowtie2+Stacks (in blue) and bwa+Stacks (in orange).

**Supporting Information: S6** Manhattan plot showing the distribution of the SNPs found in each dataset and watershed by chromosome (x-axis) and the log-transformation of the  $p$ -values ( $-\log_{10}(p)$ ) on the y-axis.

**Supporting Information: S7** List of the all the outliers overlapping between pipelines (bwa+dDocent, bowtie2+Stacks and bwa+Stacks) and watersheds (Koutajoki and Oulujoki), that did not show any known association with migration/residency.