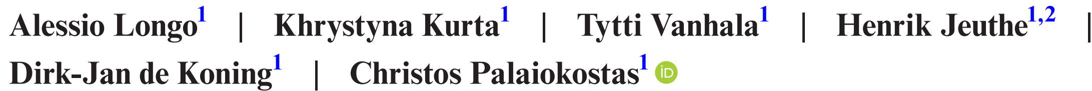
Received: 10 February 2023 | Accepted: 1 November 2023

DOI: 10.1111/age.13378

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

Genetic diversity patterns in farmed rainbow trout (*Oncorhynchus mykiss*) populations using genome-wide SNP and haplotype data



Alessio Longo[1] | Khrystyna Kurta[1] | Tytti Vanhala[1] | Henrik Jeuthe[1,2] | Dirk-Jan de Koning[1] | Christos Palaiokostas[1]

[1]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden

[2]Aquaculture Center North, Kälarne, Sweden

Correspondence
Christos Palaiokostas, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 7090, 750 07 Uppsala, Sweden.
Email: christos.palaiokostas@slu.se

Funding information
Sveriges Lantbruksuniversitet

Abstract
Rainbow trout is one of the most popular aquaculture species worldwide, with a long history of domestication. However, limited information exists about the genetic diversity of farmed rainbow trout populations globally, with most available reports relying on low-throughput genotyping technologies. Notably, no information exists about the genetic diversity status of farmed rainbow trout in Sweden. Double-digest restriction-site-associated DNA sequencing was performed on more than 500 broodfish from two leading producers in Sweden and from the country's national breeding program. Following the detection of single nucleotide polymorphisms (SNPs), genetic diversity was studied by using either individual SNPs ($n = 8680$; one SNP retained per 300 bp sequence reads) or through SNP haplotypes ($n = 20\,558$; all SNPs retained in 300 bp sequence reads). Similar amounts of genetic diversity were found amongst the three populations when individual SNPs were used. Furthermore, principal component analysis and discriminant analysis of principal components suggested two genetic clusters with the two industry populations grouped together. Genetic differentiation based on the F_{ST} fixation index was ~0.01 between the industry populations and ~0.05 when those were compared with the breeding program. Preliminary estimates of effective population size (N_e) and inbreeding (based on runs of homozygosity; F_{ROH}) were similar amongst the three populations ($N_e \approx 50$–80; median $F_{ROH} \approx 0.11$). Finally, the haplotype-based analysis suggested that animals from the breeding program had higher shared coancestry levels than those from the other two populations. Overall, our study provides novel insights into the genetic diversity and structure of Sweden's three main farmed rainbow trout populations, which could guide their future management.

KEYWORDS
ddRAD, genetic diversity, rainbow trout

INTRODUCTION

The salmonid *Oncorhynchus mykiss*, originally native to the Pacific basin of North America and the Kamchatka Peninsula in north-eastern Eurasia, has been introduced over the years in at least 99 countries (Stanković et al., 2015). *Oncorhynchus mykiss*, commonly known as steelhead trout (anadromous) and rainbow trout (spends its entire lifetime in fresh water) is one of the most domesticated and farmed aquaculture species worldwide (Teletchea & Fontaine, 2014). The latter, with a production volume of approximately 740 000 t, is the most

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2023 The Authors. *Animal Genetics* published by John Wiley & Sons Ltd on behalf of Stichting International Foundation for Animal Genetics.

farmed salmonid globally in freshwater (FAO, 2022). Rainbow trout is expected to play a key role toward scaling up the total farmed fish production volume required to cover the needs of the expanding human population. In line with the above lies the fact that between 2015 and 2019, the worldwide production of farmed trout increased by more than 20% (D'Agaro et al., 2022).

Farmed in both the south and north of Europe, rainbow trout is a key aquaculture species for the continent's industry. Swedish aquaculture, in particular, largely relies on rainbow trout, which, with an annual production volume close to 10 000 t, representing approximately 85% of the country's farmed finfish production (Jordbruksverket, 2022). Notably, finfish consumption in Sweden is still largely dependent on imports rather than on the domestic aquaculture industry, which is still considered embryonic compared with other neighbouring countries. Aiming to boost production efficiency, a rainbow trout breeding program was initiated in Sweden during the early 1980s. The founding population originated from a Norwegian one that had previously undergone four generations of selection for increased growth rate and delayed sexual maturation (Sylvén & Elvingson, 1992). However, this attempt was short-lived, primarily owing to disease and funding issues, with the program being discontinued during the mid-1990s.

Nevertheless, a new breeding program was formed in 2011 using the remnants of the original strain in combination with broodstock from the Swedish rainbow trout industry. Relying on traditional pedigree recordings, the first generation was established in 2016, and there are currently two generations of selection. Compared with other salmonid breeding programs, the Swedish one is relatively small, with approximately 70 full-sib families produced during the last year class (Kurta et al. 2023).

Genetic variance is a critical component of every breeding program, as selection acts directly upon it. In particular, in the case of aquaculture species that are commonly characterised by high fecundity, genetic variance can be depleted rapidly as short-term production goals, in many cases, can be accomplished by using a small number of broodfish (Saura et al., 2021). Pedigree recordings can offer solutions allowing for relationships between the breeding candidates to be estimated using 'classic' quantitative genetics theory. As such, the magnitude of inbreeding accumulation per generation can be controlled to a certain extent (Meuwissen, 1997; Nielsen et al., 2011). Nevertheless, the above assumes that the founding animals are considered unrelated (Aguilar & Misztal, 2008; Meuwissen & Luo, 1992). However, this assumption conflicts with the background knowledge of the Swedish rainbow trout breeding program. Therefore, it is questionable whether the current gene pool of the farmed population contains sufficient variation that guarantees the program's long-term sustainability.

High-throughput genotyping allows for in-depth genetic diversity studies regardless of the availability of genealogy-related information. Genotyping by sequencing platforms, like double digest restriction-site associated DNA (ddRAD-seq) (Peterson et al., 2012), have proved valuable in a wide range of aquaculture studies (Robledo et al., 2018). Notably, several ddRAD-seq studies have focused on genetic diversity in farmed fish (Hosoya et al., 2018; Nedoluzhko et al., 2021; Nyinondi et al., 2020; Palaiokostas et al., 2022; Torati et al., 2019). At the same time, many studies applied high-throughput genotyping technologies in farmed rainbow trout (Lhorente et al., 2019). However, the focus has been primarily on production-related traits, such as body weight and metabolism (Reis Neto et al., 2019), parasitic disease resistance (Barria et al., 2018), chronic heat stress (Yoshida & Yáñez, 2022) and fillet yield (Gonzalez-Pena et al., 2016), amongst others.

On the other hand, prior research studying genetic diversity in rainbow trout has mainly focused on introgression (Hohenlohe et al., 2013) and on quantifying the risk of outbreeding depression of wild populations owing to the introduction of domesticated animals from hatcheries (Abadía-Cardoso et al., 2016; Consuegra et al., 2011; Leitwein et al., 2017). Surprisingly, even though rainbow trout is one of the most popular aquaculture species with a long history of domestication, limited information exists about the diversity status of farmed populations, with only a handful of studies using modern genotyping platforms (D'Ambrosio et al., 2019; Liu et al., 2017). To the best of our knowledge, no prior study has attempted to evaluate the genetic diversity of farmed rainbow trout in Sweden using any genotyping technology.

In the current study, we attempted to gain the first insights into the genetic diversity levels of Swedish farmed rainbow. More than 500 broodfish from two main producers and from the national breeding program were genotyped using ddRAD-seq. Genetic diversity metrics were computed both within and across populations. Moreover, we investigated the existence of genetic clusters and estimated population-level coancestry coefficients which could be of value for the management of the breeding populations. Additionally, possible directions regarding the future of the national breeding program are discussed.

MATERIALS AND METHODS

Animal ethics

The current study was performed in accordance with the Swedish legislation described in the Animal Welfare Act 2018:1192 (ethics permit: 5.2.18-09859/2019).

Background of sampled animals

Fin clips were collected from the 2016 year-class broodfish ($n=177$) of the national Swedish rainbow trout

13652052, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/age.13378 by Swedish University Of Agricultural Sciences, Wiley Online Library on [15/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

breeding program located at facilities of Aquaculture Centre North (Kälarne, Sweden denoted in the following as Breeding Pop). A nested breeding design was used where eggs from two females were fertilised separately by milt from one male, while matings amongst close relatives (e.g. full or half sibs) were avoided. Since the core activities of the breeding program require pedigree information, the fish were marked with passive integrated transponder tags. For the needs of this study, fin clips were also collected from broodfish of two main rainbow trout producers in Sweden. Those two populations were denoted as Industry Pop1 ($n = 192$) and Pop2 ($n = 192$). No pedigree records were kept in those populations with mass selection being used to identify suitable broodfish.

DNA extraction and ddRAD library preparation

Genomic DNA was extracted using a salt-based precipitation method previously described in Palaiokostas et al. (2022). In short, fin tissue was digested at 55°C for 4h using 200μL of SSTNE (50mM Tris base, 300mM NaCl, 0.2mM each of EGTA and EDTA, 0.15mM of spermine tetrahydrochloride and 0.28mM of spermidine trihydrochloride; pH 9; Sigma-Aldrich, Darmstadt, Germany), 10% sodium dodecylsulfate (Bio-Rad, Hercules, USA) and 100μg proteinase K. RNaseA (Thermo Fisher, Vilnius, Lithuania; 2mg/mL) was added following the digestion (5μL), and the samples were incubated at 37°C for 60min. Next, proteins were precipitated by adding 0.7 volume of 5M NaCl (Sigma-Aldrich). After that DNA was pelleted by adding 0.7 vols of isopropanol and centrifuging (Pico 21; Thermo Fisher, Waltham, MA, USA) at 14000*g* for 5min and incubated overnight with 75% ethanol. The DNA pellet was dissolved in 30μL of 5mM Tris (pH 8.0; Sigma-Aldrich), and its content and quality were assessed using a NanoDrop 8000 (Thermo Scientific) spectrophotometer, agarose gel electrophoresis and Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA). Finally, the DNA samples were diluted to 15ng/μL using 5mM Tris (pH 8.0) and stored at 4°C before library preparation.

The ddRAD library was prepared following a modified version of the original protocol (Peterson et al., 2012) described in detail by Palaiokostas et al. (2015). In short, three ddRAD libraries were prepared from 550 samples. A 15ng aliquot from each individual DNA sample was digested at 37°C for 60min with the high-fidelity enzyme *SbfI* recognising the CCTGCA|GG motif and the *NlaIII* recognising the CATG motif (New England Biolabs, Ipswich, UK). Individual-specific P1 and P2 adapters with unique 5 or 7bp barcodes were ligated with the samples incubated at room temperature for 120min. The ligation reaction was stopped after adding 2.5 vols of PB buffer (Qiagen, Hilden, Germany), after which the samples were combined in a multiplex pool and purified with a MinElute PCR Purification kit (Qiagen).

The libraries were size-selected (400–600bp) by electrophoresis on a 1.1% TAE agarose gel, followed by gel purification. The gel was run at constant voltages of 45V for 3min, 60V for 3min and 90V for around 70min. PCR amplification was performed on a thermal cycler T100 (Bio-Rad, Redmond, WA, USA) using the following cycling conditions: 98°C for 30s, 13–14 PCR cycles of 98°C for 10s, 65°C for 30s and 72°C for 30s, then a final step of 72°C for 5min. Each amplified library was purified using an equal volume of AMPure beads (Beckman Coulter, Brea, CA, USA), and eluted at 20μL with EB buffer (MinElute Gel Purification Kit; Qiagen). Finally, the libraries were quality controlled using the TapeStation system and Genomic DNA ScreenTape assay (Agilent Technologies, Waldbronn, Germany) and sequenced in an Illumina NovaSeq 6000 using three lanes of SP flow cells (150 bp end reads) at the National Genomics Infrastructure centre (Uppsala, Sweden).

Sequenced data filtering and SNP detection

Trimming of adapter-oligomeric sequences and the filtering out of reads with a Phred quality score below 30 were performed using FASTP v0.22.0 (Chen et al., 2018). The remaining reads were then processed with the *process-radtags* module of STACKS v2.5 (Rochette et al., 2019) for demultiplexing and checking for the existence of the expected restriction site. Thereafter retained reads were aligned to the reference genome (Assembly:USDA_OmykA_1.1; Genbank accession number GCA_013265735.3) using BOWTIE2 (Langmead & Salzberg, 2012). Following genotyping calling using STACKS, two approaches were followed using the *populations* module. For the first approach only a single SNP per ddRAD-tag was retained, followed by filtering for those found in at least two of the populations, having observed heterozygosity below 0.6, minor allele frequency above 0.05 and calling rate above 80%. Finally, the vcf file was further filtered with VCFTOOLS v0.1.16 (Danecek et al., 2011), removing animals with more than 30% missing data. The second approach involved retaining all SNPs located on the same ddRAD-tag and applying the previous filters haplotype-wise (Figure 1).

Genetic diversity metrics and population structure based on individual SNPs

Generic diversity metrics like mean observed heterozygosity (H_o), expected heterozygosity (H_E) and Wright's F statistics, such as the individual F_{IS} coefficient and the fixation index F_{ST}, were estimated using the *populations* module of STACKS. In the latter case, a kernel-smoothed average estimate was used, while the default sliding

13652052, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/age.13378 by Swedish University Of Agricultural Sciences, Wiley Online Library on [15/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

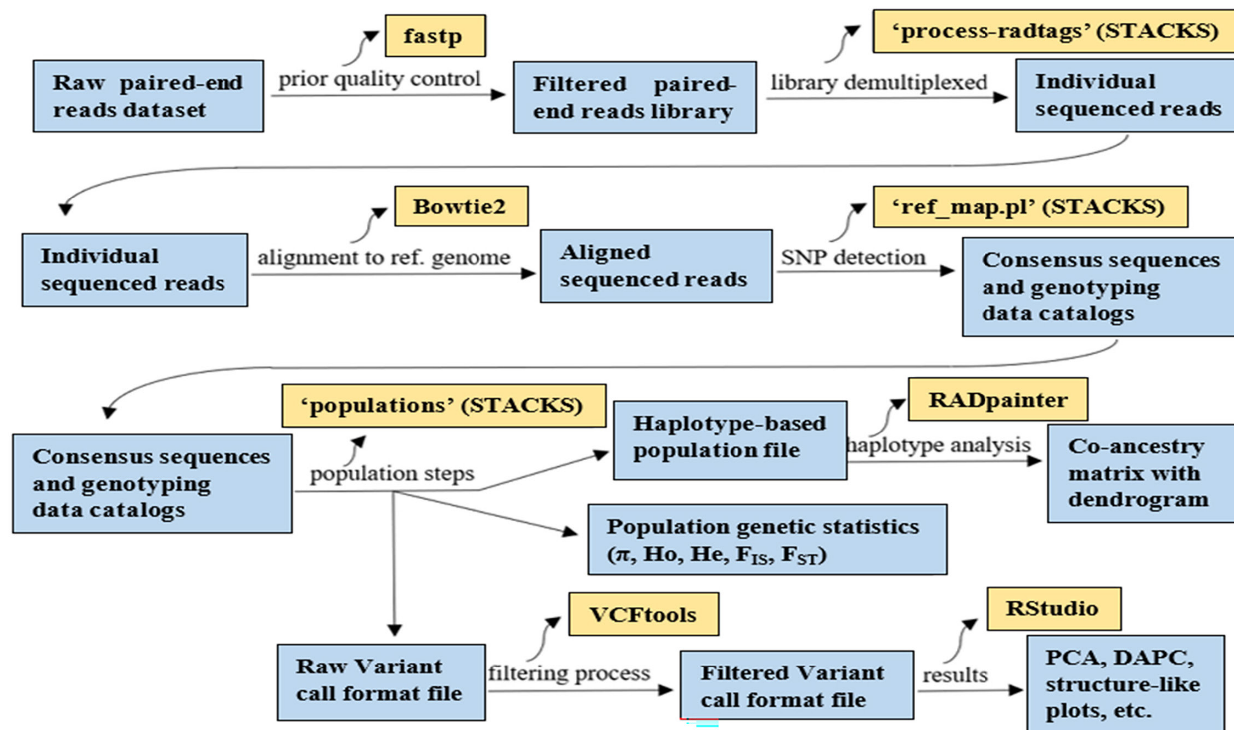


FIGURE 1 Bioinformatic workflow for detecting SNPs and estimating genetic diversity metrics.

window size was equal to 450 kbp. Moreover, Manhattan plots of the kernel-smoothed F_{ST} values were built with the R package CMPLOT v4.2.0 (Yin et al., 2021).

Principal component analysis (PCA) was conducted using the R package adegenet v2.1.5 (Jombart, 2008) for deciphering the underlying genetic structure of the studied populations. In addition, a discriminant analysis of principal components (DAPC) (Jombart et al., 2010) was performed to detect genetic clusters using the same software. Following PCA, a cross-validation step was performed using the xvalDapc function to identify the optimal number of principal components (PCs), followed by a discriminant analysis step. Finally, the selection of the optimal number of clusters (K) relied on the elbow method depicting the corresponding Bayesian information criterion (BIC) values of each tested K (Jombart et al., 2010).

Effective population size and inbreeding estimation

Effective population size (N_e) estimates were obtained using the software GONE (Santiago et al., 2020). As no genetic map was available, physical distances were used, assuming that 10 cM corresponded to 6 Mb. The above was observed in a genetic map of French rainbow trout (Fraslin et al., 2018). In addition, the Haldane function was used to adjust the genetic distances. Furthermore, runs of homozygosity (ROH) with a sliding window approach were estimated with the R package detectRUNS v0.9.6 (Biscarini et al., 2018). The window length was set to 15 SNPs, the maximum gap at 1 Mb and the minimum ROH length at 250 kb. Simultaneously, the minimum number of SNPs per run was set to 20, and the maximum number of missing genotypes per window to 1. Finally, inbreeding coefficients (F_{ROH}) were estimated from the obtained ROH for each individual using the *Froh_inbreeding* function of detectRUNS.

Haplotype-derived population structure

A haplotype-based analysis of the studied populations was performed with the RADPAINTER and FINERADSTRUCTURE software (Malinsky et al., 2018). More specifically, a coancestry matrix was estimated from all individuals using RADPAINTER followed by clustering using the Markov chain Monte Carlo algorithm of FINESTRUCTURE (Lawson et al., 2012). Finally, a heatmap depicting the clustered coancestry matrix was constructed with FINERADSTRUCTURE.

RESULTS

Sequencing output and SNP detection

Approximately 2.9 billion 150 bp paired-end reads were produced. Around 10% of these were removed owing to their missing the expected RAD cut site and 0.2% owing to a low-quality Phred score. Overall, approximately 82%

13652052, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/age.13378 by Swedish University Of Agricultural Sciences, Wiley Online Library on [15/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

13652052, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/age.13378 by Swedish University Of Agricultural Sciences, Wiley Online Library on [15/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of the initially obtained reads were retained. From those reads, 27219 polymorphic loci were found in at least 80% of the genotyped animals, of which 20558 SNPs passed quality control. Those SNPs were used for the haplotype-based analysis. In the case of individual SNP-based analysis, only one marker per locus was retained, resulting in a dataset of 8680 SNPs. Moreover, 22 animals with more than 30% missing genotypes were removed. Overall, 529 animals from the three populations were retained for downstream analysis. The mean sequencing coverage for the retained loci in those animals was 32× (SD 19×). Moreover, a similar minor allele frequency distribution was observed amongst the three populations (Figure S1).

Genetic diversity metrics

Of the 529 animals that passed quality control, 172 were from the Breeding Pop, 184 from Industry Pop1 and 173 from Industry Pop2. The values of H_o and H_E were nearly identical between the three populations, with the former metric having, on average, a value of 0.25, while the latter was 0.27. Regarding the F_{IS} coefficient, the obtained values ranged between 0.06 and 0.095, with the industry Pop1 having the highest value (Table 1).

Population structure – differentiation based on individual SNPs

The PCA suggested a clear separation between the breeding program and the two industry populations across the first PC. The latter two populations were grouped together with variation though observed across the second PC. In terms of explained variance the first two PCs accounted for 8 and 2% respectively (Figure 2).

A DAPC was used to further decipher the genetic structure of the three populations. Cross-validation suggested that the optimal number of principal components for clustering was 50, with DAPC indicating that the optimal number of clusters (*K*) was two (Figure S2).

In line with the PCA and DAPC results, the lowest estimated genetic distance according to the F_{ST} metric was between the two industry populations (0.015). Nevertheless, substantial variation of the F_{ST} metric was found across the rainbow trout genome, with values exceeding the 99% quantile cutoff of the empirical distribution found in 10 chromosomes (Figure 3a). Furthermore, a larger genetic distance was found between the Breeding Pop and the ones from the industry. More specifically, F_{ST} was 0.047 between the Breeding Pop and Industry Pop1. Notably, during this pairwise comparison (Breeding Pop vs Industry Pop1), the kernel-smoothed F_{ST} values exceeded the 99% quantile cutoff of the empirical distribution in genomic regions across 25 chromosomes, the largest of which ($F_{ST}=0.41$) was found in chromosome 30 (Figure 3b). Similarly, the F_{ST} based genetic distance between the Breeding Pop and Industry Pop2 was 0.051. As for genomic regions where F_{ST} exceeded the 99% quantile cutoff of the empirical distribution, those were found in 26 different chromosomes, with the most significant value ($F_{ST}=0.44$) located on chromosome 5 (Figure 3c). Overall, genomic regions exceeding the *a priori* 99% quantile threshold in both cases were found in 20 chromosomes.

TABLE 1 General metrics of genetic diversity.

Population	H_o (SE)	H_E (SE)	F_{IS} (SE)
Breeding Pop	0.26 (0.03)	0.28 (0.03)	0.070 (0.03)
Industry Pop1	0.25 (0.02)	0.27 (0.02)	0.095 (0.03)
Industry Pop2	0.25 (0.03)	0.27 (0.02)	0.060 (0.02)

Abbreviations: F_{IS}, inbreeding coefficient; H_E, expected heterozygosity; H_o, observed heterozygosity.

Effective population size and inbreeding estimates

All N_e estimates for the most recent generation were below the suggested threshold of 100. Amongst the three populations, the one with the lowest N_e was Industry Pop2, with a value of 43, while Industry Pop1 had the highest value ($N_e=85$; Figure 4).

Inbreeding coefficients based on ROH for each studied animal ranged between 0.002 and 0.30. The median values of the inbreeding coefficients of each of the three populations were similar and approximately equal to 0.11 (Figure 5).

Population structure – differentiation based on haplotypes

The haplotype-derived coancestry values provided additional insights regarding the underlying genetic structure of the three populations (Figure 6). As expected, individuals within the Breeding Pop showed higher coancestry levels between them than with the two industry populations forming a distinct cluster. Moreover, from the produced heatmap, it was apparent that the coancestry levels between the animals from the industry populations were lower than those from the breeding program. Finally, indications of further population substructuring were shown compared with the previous DAPC results, where animals from the industry populations show higher shared coancestry.

DISCUSSION

Surprisingly, little information exists about the genetic diversity status of farmed rainbow trout despite it being a

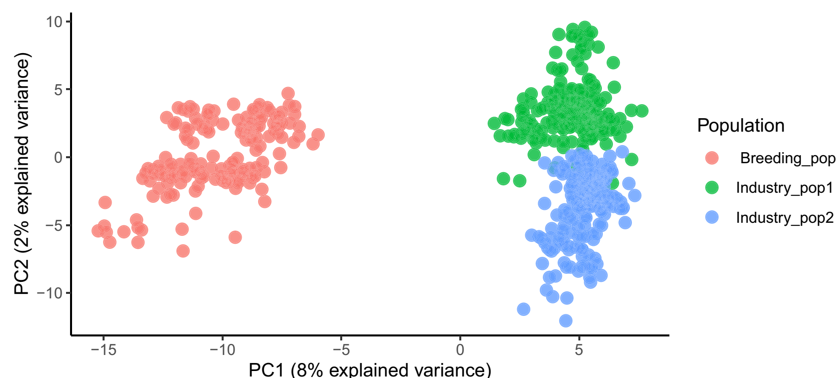


FIGURE 2 Principal component analysis (PCA) of farmed rainbow trout populations in Sweden.

critical determining factor for the long-term success of any animal farming industry (Kristensen & Sørensen, 2005). With the exception of the recent study by D'Ambrosio et al. (2019), where a detailed screening of approximately 300 rainbow trout individuals from five French populations was performed using a 57k SNP array, all previous studies focusing on genetic diversity relied on low-throughput genotyping technologies. In the case of Sweden, no prior information exists about the genetic diversity of any of the farmed populations. Therefore, we conducted a large-scale screening using ddRAD involving more than 500 broodfish from two major Swedish producers and from the national breeding program.

Genetic diversity within populations

Based on the estimated generic diversity metrics like the heterozygosity levels, minimal and indistinguishable differences were found between the three populations. More specifically, the mean H_o was approximately 0.25, with a small heterozygosity deficit as the mean H_E was 0.27. Generally, the above values were within the reported range in the literature ($H_E \approx 0.1$–0.4) on similar scale studies in wild or farmed fish using ddRAD (Drinan et al., 2018; Lemopoulos et al., 2019; Nyinondi et al., 2020). On the other hand, slightly higher differences were found between the studied populations in terms of the F_{IS} coefficient, with the first industry population having a higher value (0.095). Positive F_{IS} values, besides indicating a potential loss of heterozygosity, could also mean the existence of non-random mating or population subdivision (Allendorf & Luikart, 2007). Considering common aquaculture practices, e.g. a closed nucleus in the case of the breeding program and selection based on observed phenotypes (taking place in all studied populations), all the above could be the driving forces toward higher F_{IS} values. Notably, an average F_{IS} of 0.39 was recently reported in Italian populations of wild brown trout using ddRAD (Magris et al., 2022).

Even though direct comparisons with low-throughput genotyping studies on rainbow trout are probably not meaningful, cases of both slight excess and a deficit of heterozygosity have been previously reported (Abadía-Cardoso et al., 2016; Leitwein et al., 2017). Furthermore, high values (>0.3) of F_{IS} have been reported for Chilean farmed rainbow trout (Consuegra et al., 2011). In terms of previous studies, probably the most relevant to ours is the one by D'Ambrosio et al. (2019), where a slight excess of heterozygosity was reported. Nevertheless, this does not necessarily imply a diverse gene pool as in the same study, high levels of inbreeding were reported reaching up to 19.5%, but instead could indicate a recent (<100 generations) bottleneck (Cornuet & Luikart, 1996).

Effective population size estimates and inbreeding

Overall, the N_e estimates of our study are within the reported range in previous aquaculture studies (Saura et al., 2021; Villanueva et al., 2022). More specifically, N_e estimates from farmed fish populations appear to be lower than 100 and, in some cases (Barria et al., 2018; D'Ambrosio et al., 2019; Garcia et al., 2018), as in our study (Industry Pop2), lower than 50, which is considered an empirical threshold below which the probability of inbreeding depression increases. Taking into consideration the high fecundity of fish and the fact that breeding programs are relatively new in aquaculture, with crosses often performed without taking into consideration the relationship levels of the mating pair, the above results are probably not surprising. At the same time, it is important to stress that N_e estimates largely rely on genotyping density and the underlying demographic-genealogy model each software uses. In terms of the former, it is important to stress that our study lies in the lower range of genotyping density for conducting such estimations, so the reported values should be treated with caution. Probably, the most closely comparable study is the one of Saura et al. (2021), where both a similar genotyping density (~10 000–15 000 SNPs) and the same software were used to estimate N_e in farmed carp, seabass, seabream and turbot. In all those species, the reported N_e was lower than 50 (31–46) when animals from the latest available generation were used.

13652052, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/age.13378 by Swedish University Of Agricultural Sciences, Wiley Online Library on [15/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

13652052, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/age.13378 by Swedish University Of Agricultural Sciences, Wiley Online Library on [15/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

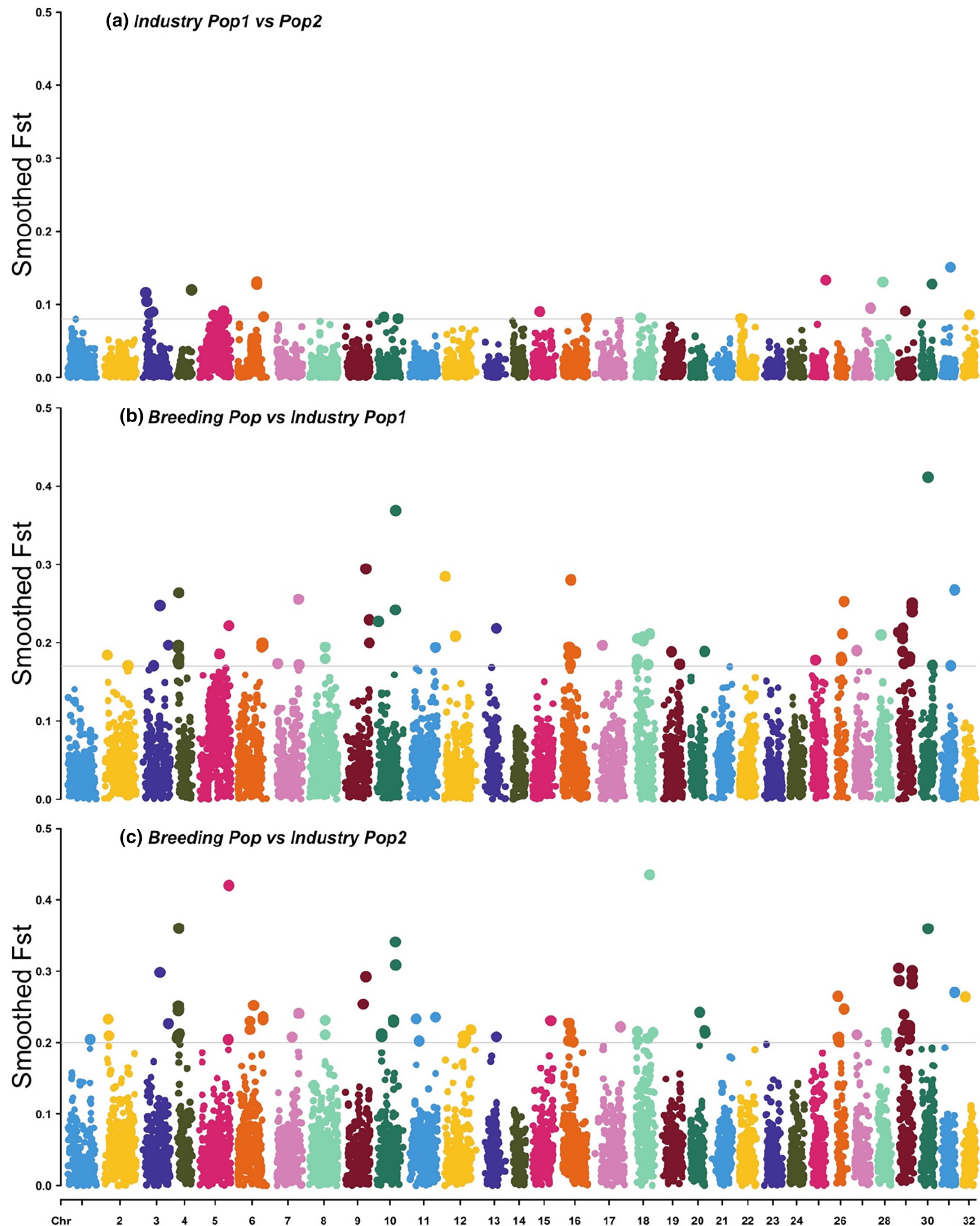


FIGURE 3 Manhattan plot showing F_{ST} metric values for each pairwise population comparison. (a) Comparison between industry populations. (b) Comparison between the breeding population and the first industry population. (c) Comparison between the breeding population. Points represent the 450 kb genomic windows for which kernel-smoothed F_{ST} values were estimated and plotted along the horizontal axis based on physical position. The horizontal grey line corresponds to the 99% percentile cutoff of each empirical F_{ST} distribution.

Similarly, the estimated inbreeding coefficients (based on ROH) were consistent with the N_e estimates and in line with reported values from the literature. In particular, similar F_{ROH} values were obtained amongst the three populations with a median of approximately 0.11. In the study of D'Ambrosio et al. (2019) for similar

N_e values, F_{ROH} ranged between 0.11 and 0.20. It should be stressed that neither the reported N_e estimates nor the inbreeding coefficients suffice on their own to conclude whether the populations under study are experiencing inbreeding depression. Nevertheless, it appears prudent to schedule appropriate management actions before inbreeding depression becomes evident, as its appearance could lead to a non-reversible situation in terms of production. Therefore in the case of the studied rainbow trout populations, it would be interesting to investigate the possibility of a cross-breeding scheme between the industry populations and the one from the breeding program. Taking into account the farming history of the latter, the above should be seriously considered. Simultaneously, since inbreeding accumulation is not uniform across the genome (Howard et al., 2017), the inclusion of genotyping information is expected to be beneficial in the management of all three populations.

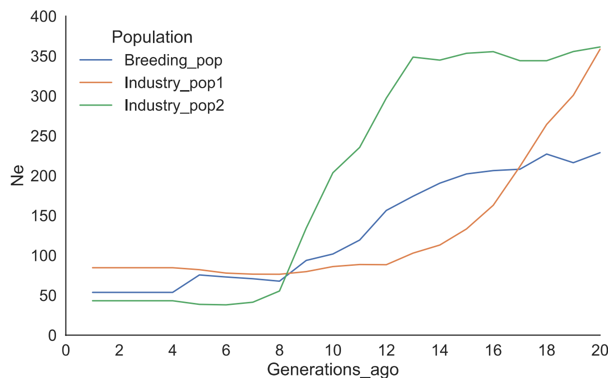


FIGURE 4 N_e estimates across the last 20 generations for each studied population. Different colors are used to denote each population.

Population differentiation and underlying genetic structure

Overall, low genetic differentiation was observed between the populations (F_{ST}<0.05). In particular, in the case of those from the industry, the obtained genetic distance between them was minimal ($F_{ST}\approx0.01$), suggesting that those two populations might have a common origin. In connection with that, two distinct genetic clusters were suggested, with the two industry populations grouped together and the breeding program population appearing as a separate group. Unfortunately, we could not obtain background information about the origin of the two industry populations, so we cannot confirm this hypothesis. In general, moderate to high levels of genetic differentiation between rainbow trout populations have been reported in several cases. On the other hand, previous studies based on low-throughput genotyping suggested that most of the genetic variation in farmed rainbow trout is to be found within a population (Hershberger, 1992; Martsikalis et al., 2014). However, as previously mentioned, limited information is currently available regarding the levels of genetic differentiation between farmed rainbow trout populations.

One of the advantages of high-throughput genotyping platforms like ddRAD is that instead of relying on an average F_{ST} value, they allow screening at high resolution of its variation across the genome. In our case, even though, based on the average F_{ST} values, low genetic differentiation was suggested overall, it was evident by inspecting the Manhattan plots that a high degree of variation existed across the genome. Furthermore, even in the case of the industry populations that formed a distinct genetic cluster, genomic regions suggesting high differentiation (F_{ST}>0.15) were found. Those regions could represent signatures of selection owing to domestication and differing farming practices (López et al., 2014; Paul et al., 2022). However,

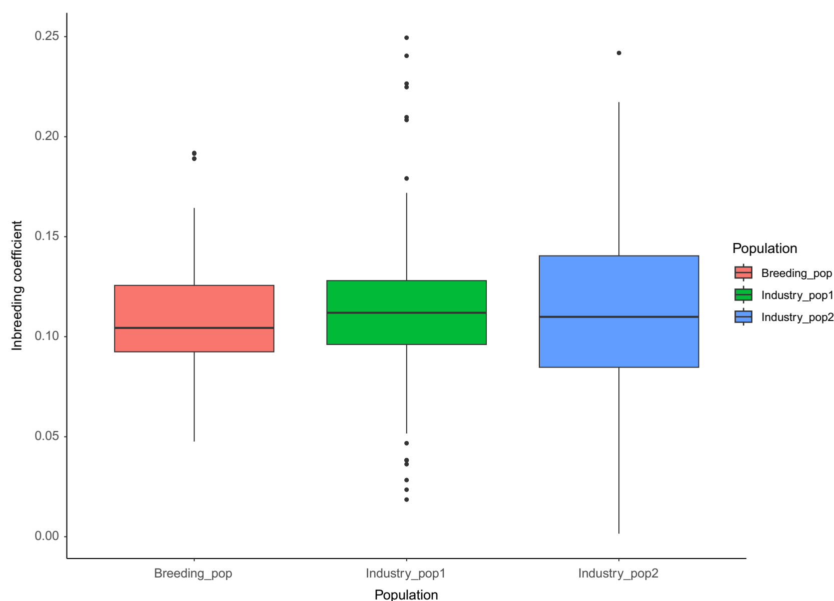


FIGURE 5 Boxplots depicting the range of inbreeding coefficients in terms of F_{ROH} of the studied populations. The solid horizontal line within each boxplot depicts the median value.

13652052, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/age.13378 by Swedish University Of Agricultural Sciences, Wiley Online Library on [15/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

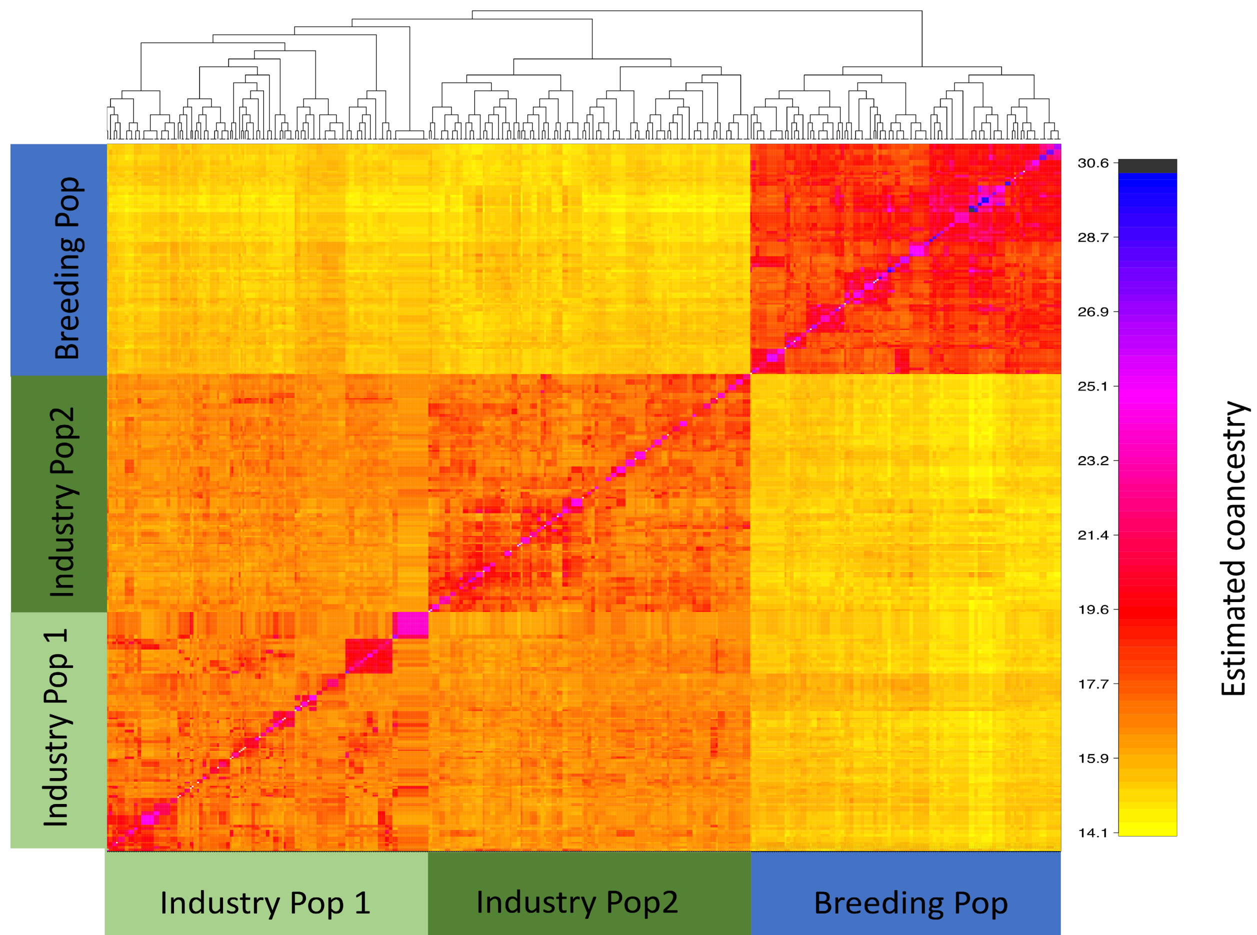


FIGURE 6 Heatmap depicting the haplotype-derived coancestry matrix. The color gradient shows the magnitude of shared coancestry between each pair of individuals.

since we do not have information about the origin of the industry populations, those regions could also reflect a distinct genetic background before domestication.

Even though genotyping platforms like ddRAD have already been applied to a wide range of aquaculture species (Li & Wang, 2017), mainstream data processing revolves almost exclusively around individual SNPs. Nevertheless, considering that nowadays, 300–600 bp long sequence reads are produced in genotyping platforms like ddRAD, additional information could be extracted from the data. Since individual SNPs at such short distances are, in the vast majority of cases, in complete linkage disequilibrium, it is common during data processing for only a single SNP in each sequenced read to be kept for downstream analysis. However, treating the SNPs within the same sequence read as phased haplotypes can allow for additional insights, especially regarding relationships amongst the studied animals (Malinsky et al., 2018). For instance, haplotype-based coancestry analysis using ddRAD suggested that different ecotypes of Arctic charr (*Salvelinus alpinus*) evolved independently in parallel (Jacobs et al., 2020). In our study, the haplotype-based analysis shed additional light on the underlying population structure of the three populations. More specifically, it became clear that shared coancestry was higher amongst the animals from the breeding program than the corresponding one in the two industry populations. Moreover, the haplotype analysis revealed an underlying genetic structure in the case of the two industry populations compared with when DAPC was performed based on individual SNPs. Overall, complementing the typical single SNP-based genetic diversity analysis with haplotype-based information could offer additional insights, revealing previously undetected population structure.

CONCLUSIONS

Our study represents the first large-scale effort to evaluate the genetic diversity status of farmed rainbow trout in Sweden using high-throughput genotyping. Based on generic metrics, all three populations had comparable levels of genetic diversity. However, the two industry populations formed one cluster in terms of the underlying genetic structure, while the breeding program population appeared to be distinct. Haplotype-based analysis suggested that the

13652052, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/age.13378 by Swedish University Of Agricultural Sciences, Wiley Online Library on [15/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

13652052, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/age.13378 by Swedish University Of Agricultural Sciences, Wiley Online Library on [15/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

animals of the latter population displayed higher levels of shared coancestry compared with the populations from the industry. Even though no evidence of inbreeding depression is currently apparent in the breeding program, our results, together with the background information about their origin, suggest that expanding the gene pool of the breeding nucleus through crossing with other populations should be taken into serious consideration. As the two industry populations studied here appear to be of common genetic origin, even using broodfish from only one of them could suffice to expand the breeding program's gene pool. On the other hand, as neither of the two companies keeps pedigree recordings, using broodfish from both would reduce the chances of crossing close relatives. Nevertheless, carefully considering industry-valued phenotypic traits will be required before embarking on such an endeavour.

AUTHOR CONTRIBUTIONS

AL contributed toward the preparation of the sequencing libraries, the initial data analysis and the compilation of a first draft. KK and TV performed DNA extraction and prepared the sequencing libraries. HJ contributed toward the original study design and in sample collection. DJK contributed toward the original study design and the manuscript's writing and reviewing. CP finalised the study design, reviewed all the data analysis steps and contributed to the manuscript's writing.

ACKNOWLEDGEMENTS

The authors would like to thank the VBCN personnel for assisting with the sampling.

FUNDING INFORMATION

The authors were supported by funding from the Swedish government under grant agreement no. N2017/02366/SUN and from the SLU aquaculture platform. Sequencing was performed by the SNP & SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure Sweden and Science for Life Laboratory. The SNP & SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Sequence information of the utilised samples in the form of fastq files was deposited in the National Centre for Biotechnology Information repository in bam format under project ID PRJNA933589.

ORCID

Christos Palaiokostas https://orcid.org/0000-0002-4480-4612

REFERENCES

Abadía-Cardoso, A., Pearse, D.E., Jacobson, S., Marshall, J., Dalrymple, D., Kawasaki, F. et al. (2016) Population genetic structure and ancestry of steelhead/rainbow trout (*Oncorhynchus mykiss*) at the extreme southern edge of their range in North America. *Conservation Genetics*, 17, 675–689.

Aguilar, I. & Misztal, I. (2008) Technical note: recursive algorithm for inbreeding coefficients assuming nonzero inbreeding of unknown parents. *Journal of Dairy Science*, 91, 1669–1672.

Allendorf, F. & Luikart, G. (2007) *Conservation and the genetics of populations*. Blackwell Oxfors.

Barria, A., López, M.E., Yoshida, G., Carvalheiro, R., Lhorente, J.P. & Yáñez, J.M. (2018) Population genomic structure and genome-wide linkage disequilibrium in farmed Atlantic salmon (*Salmo salar* L.) using dense SNP genotypes. *Frontiers in Genetics*, 9, 649.

Biscarini, F., Cozzi, P., Gaspa, G. & Marras, G. (2018) detectRUNS: An R Package to detect runs of homozygosity and heterozygosity in diploid genomes. CRAN (The Compr. R Arch. Network). Available at https://rdrr.io/cran/detectRUNS/f/vignettes/detectRUNS.vignette.Rmd [Accessed September 2023].

Chen, S., Zhou, Y., Chen, Y. & Gu, J. (2018) Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34, i884–i890.

Consuegra, S., Phillips, N., Gajardo, G. & de Leaniz, C.G. (2011) Winning the invasion roulette: escapes from fish farms increase admixture and facilitate establishment of non-native rainbow trout. *Evolutionary Applications*, 4, 660–671.

Cornuet, J.M. & Luikart, G. (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, 144, 2001–2014.

D'Agaro, E., Gibertoni, P. & Esposito, S. (2022) Recent trends and economic aspects in the rainbow trout (*Oncorhynchus mykiss*) sector. *Applied Sciences*, 12, 8773.

D'Ambrosio, J., Phocas, F., Haffray, P., Bestin, A., Brard-Fudulea, S. et al. (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, 26.

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

Drinan, D.P., Gruenthal, K.M., Canino, M.F., Lowry, D., Fisher, M.C. & Hauser, L. (2018) Population assignment and local adaptation along an isolation-by-distance gradient in Pacific cod (*Gadus macrocephalus*). *Evolutionary Applications*, 11, 1448–1464.

FAO. (2022) *Fishery and aquaculture statistics 2022*. FAO. https://www.fao.org/3/cc0461en/cc0461en.pdf

Fraslin, C., Dechamp, N., Bernard, M., Krieg, F., Hervet, C., Guyomard, R. et al. (2018) Quantitative trait loci for resistance to *Flavobacterium psychrophilum* in rainbow trout: effect of the mode of infection and evidence of epistatic interactions. *Genetics Selection Evolution*, 50, 60.

Garcia, A.L.S., Bosworth, B., Waldbieser, G., Misztal, I., Tsuruta, S. & Lourenco, D.A.L. (2018) Development of genomic predictions for harvest and carcass weight in channel catfish. *Genetics Selection Evolution*, 50, 66.

Gonzalez-Pena, D., Gao, G., Baranski, M., Moen, T., Cleveland, B.M. et al. (2016) Genome-wide association study for identifying loci that affect fillet yield, carcass, and body weight traits in rainbow trout (*Oncorhynchus mykiss*). *Front Genet*, 7, 203.

Hershberger, W.K. (1992) Genetic variability in rainbow trout populations. *Aquaculture*, 100, 51–71.

Hohenlohe, P.A., Day, M.D., Amish, S.J., Miller, M.R., Kamps-Hughes, N., Boyer, M.C. et al. (2013) Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Molecular Ecology*, 22, 3002–3013.

Hosoya, S., Kikuchi, K., Nagashima, H., Onodera, J., Sugimoto, K., Satoh, K. et al. (2018) Assessment of genetic diversity in coho

salmon (*Oncorhynchus kisutch*) populations with no family records using ddRAD-seq. *BMC Research Notes*, 11, 548.

Howard, J.T., Pryce, J.E., Baes, C. & Maltecca, C. (2017) Invited review: inbreeding in the genomics era: inbreeding, inbreeding depression, and management of genomic variability. *Journal of Dairy Science*, 100, 6009–6024.

Jacobs, A., Carruthers, M., Yurchenko, A., Gordeeva, N.V., Alekseyev, S.S., Hooker, O. et al. (2020) Parallelism in eco-morphology and gene expression despite variable evolutionary and genomic backgrounds in a Holarctic fish. *PLoS Genetics*, 16, e1008658.

Jombart, T. (2008) Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405.

Jombart, T., Devillard, S. & Balloux, F. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, 11, 94.

Jordbruksverket, V. (2022) https://jordbruksverket.se/om-jordbruksverket/jordbruksverkets-officiella-statistik/jordbruksverkets-statistikrapporter/statistik/2023-08-31-vattenbruk-2022#h-SummaryinEnglish

Kristensen, T.N. & Sørensen, A.C. (2005) Inbreeding – lessons from animal breeding, evolutionary biology and conservation genetics. *Animal Science*, 80, 121–133.

Kurta, K.H., Jeuthe, D.-J., de Koning & Palaiokostas, C. (2023) Insights from the early generations of the Swedish rainbow trout (Oncorhynchus mykiss) breeding program. *Acta Agriculturae Scandinavica, Section A — Animal Science*, 1–12.

Langmead, B. & Salzberg, S.L. (2012) Fast gapped-read alignment with bowtie 2. *Nature Methods*, 9, 357–359.

Lawson, D.J., Hellenthal, G., Myers, S. & Falush, D. (2012) Inference of population structure using dense haplotype data. *PLoS Genetics*, 8, e1002453.

Leitwein, M., Garza, J.C. & Pearse, D.E. (2017) Ancestry and adaptive evolution of anadromous, resident, and adfluvial rainbow trout (*Oncorhynchus mykiss*) in the San Francisco bay area: application of adaptive genomic variation to conservation in a highly impacted landscape. *Evolutionary Applications*, 10, 56–67.

Lemopoulos, A., Prokkola, J.M., Uusi-Heikkilä, S., Vasemägi, A., Huusko, A., Hyvärinen, P. et al. (2019) Comparing RADseq and microsatellites for estimating genetic diversity and relatedness – implications for brown trout conservation. *Ecology and Evolution*, 9, 2106–2120.

Lhorente, J.P., Araneda, M., Neira, R. & Yáñez, J.M. (2019) Advances in genetic improvement for salmon and trout aquaculture: the Chilean situation and prospects. *Reviews in Aquaculture*, 11, 340–353.

Li, Y.-H. & Wang, H.-P. (2017) Advances of genotyping-by-sequencing in fisheries and aquaculture. *Reviews in Fish Biology and Fisheries*, 27, 535–559.

Liu, S., Palti, Y., Martin, K.E., Parsons, J.E. & Rexroad, C.E. (2017) Assessment of genetic differentiation and genetic assignment of commercial rainbow trout strains using a SNP panel. *Aquaculture*, 468, 120–125.

López, M.E., Neira, R. & Yáñez, J.M. (2014) Applications in the search for genomic selection signatures in fish. *Frontiers in Genetics*, 5, 458.

Magris, G., Marroni, F., D'Agaro, E., Vischi, M., Chiabà, C. et al. (2022) ddRAD-seq reveals the genetic structure and detects signals of selection in Italian brown trout. *Genetics Selection Evolution*, 54, 8.

Malinsky, M., Trucchi, E., Lawson, D.J. & Falush, D. (2018) RADpainter and fineRADstructure: population inference from RADseq data. *Molecular Biology and Evolution*, 35, 1284–1290.

Martsikalis, P., Gkafas, G.A., Apostolidis, A.P. & Exadactylos, A. (2014) Genetic structure profile of rainbow trout (*Oncorhynchus mykiss*) farmed strains in Greece. *Turkish Journal of Fisheries and Aquatic Sciences*, 14, 749–757.

Meuwissen, T. & Luo, Z. (1992) Computing inbreeding coefficients in large populations. *Genetics Selection Evolution*, 24, 305.

Meuwissen, T.H.E. (1997) Maximizing the response of selection with a predefined rate of inbreeding. *Journal of Animal Science*, 75, 934–940.

Nedoluzhko, A.V., Gladysheva-Azgari, M.V., Shalgimbayeva, G.M., Volkov, A.A., Slobodova, N.V., Tsygankova, S.V. et al. (2021) Genetic contribution of domestic European common carp (*Cyprinus carpio carpio*) and Amur carp (*Cyprinus carpio haematopterus*) to the wild Vietnamese carp population as revealed by ddRAD sequencing. *Aquaculture*, 544, 737049.

Nielsen, H.M., Sonesson, A.K. & Meuwissen, T.H.E. (2011) Optimum contribution selection using traditional best linear unbiased prediction and genomic breeding values in aquaculture breeding schemes. *Journal of Animal Science*, 89, 630–638.

Nyinondi, C.S., Mtolera, M.S.P., Mmochi, A.J., Lopes Pinto, F.A., Houston, R.D., de Koning, D.J. et al. (2020) Assessing the genetic diversity of farmed and wild Rufiji tilapia (*Oreochromis urolepis urolepis*) populations using ddRAD sequencing. *Ecology and Evolution*, 10(18), 10044–10056.

Palaiokostas, C., Anjum, A., Jeuthe, H., Kurta, K., Lopes Pinto, F. & de Koning, D.J. (2022) A genomic-based vision on the genetic diversity and key performance traits in selectively bred Arctic charr (*Salvelinus alpinus*). *Evolutionary Applications*, 15, 565–577.

Palaiokostas, C., Bekaert, M., Khan, M.G.Q., Taggart, J.B., Gharbi, K., McAndrew, B.J. et al. (2015) A novel sex-determining QTL in Nile tilapia (*Oreochromis niloticus*). *BMC Genomics*, 16, 1–10.

Paul, K., Restoux, G. & Phocas, F. (2022) Genome-wide detection of positive and balancing selection signatures shared by four domesticated rainbow trout populations (*Oncorhynchus mykiss*). *bioArxiv*. https://doi.org/10.1101/2022.12.08.519621

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. & Hoekstra, H.E. (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, 7, e37135.

Reis Neto, R.V., Yoshida, G.M., Lhorente, J.P. & Yáñez, J.M. (2019) Genome-wide association analysis for body weight identifies candidate genes related to development and metabolism in rainbow trout (*Oncorhynchus mykiss*). *Molecular Genetics and Genomics*, 294, 563–571.

Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P. & Houston, R. (2018) Applications of genotyping by sequencing in aquaculture breeding and genetics. *Reviews in Aquaculture*, 10, 670–682.

Rochette, N.C., Rivera-Colón, A.G. & Catchen, J.M. (2019) Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, 28, 4737–4754.

Santiago, E., Novo, I., Pardiñas, A.F., Saura, M., Wang, J. & Caballero, A. (2020) Recent demographic history inferred by high-resolution analysis of linkage disequilibrium. *Molecular Biology and Evolution*, 37, 3642–3653.

Saura, M., Caballero, A., Santiago, E., Fernández, A., Morales-González, E., Fernández, J. et al. (2021) Estimates of recent and historical effective population size in turbot, seabream, seabass and carp selective breeding programmes. *Genetics Selection Evolution*, 53, 85.

Stanković, D., Crivelli, A.J. & Snoj, A. (2015) Rainbow trout in Europe: introduction, naturalization, and impacts. *Reviews in Fisheries Science & Aquaculture*, 23, 39–71.

Sylvén, S. & Elvingson, P. (1992) Comparison of rainbow trout (*Oncorhynchus mykiss*) strains for body weight, length and age at maturity in different Swedish production systems. *Aquaculture*, 104, 37–50.

Teletchea, F. & Fontaine, P. (2014) Levels of domestication in fish: implications for the sustainable future of aquaculture. *Fish and Fisheries*, 15, 181–195.

Torati, L.S., Taggart, J.B., Varela, E.S., Araripe, J., Wehner, S. & Migaud, H. (2019) Genetic diversity and structure in *Arapaima gigas* populations from Amazon and Araguaia-Tocantins river basins. *BMC Genetics*, 20, 13.

Villanueva, B., Fernández, A., Peiró-Pastor, R., Peñaloza, C., Houston, R.D., Sonesson, A.K. et al. (2022) Population structure and genetic variability in wild and farmed Mediterranean populations

13652052, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/age.13378 by Swedish University Of Agricultural Sciences, Wiley Online Library on [15/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of gilthead seabream and European seabass inferred from a 60K combined species SNP array. *Aquaculture Reports*, 24, 101145.

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z. et al. (2021) rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics, Proteomics & Bioinformatics*, 19, 619–628.

Yoshida, G.M. & Yáñez, J.M. (2022) Increased accuracy of genomic predictions for growth under chronic thermal stress in rainbow trout by prioritizing variants from GWAS using imputed sequence data. *Evolutionary Applications*, 15, 537–552.

SUPPORTING INFORMATION

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

How to cite this article: Longo, A., Kurta, K., Vanhala, T., Jeuthe, H., de Koning, D.-J. & Palaiokostas, C. (2024) Genetic diversity patterns in farmed rainbow trout (*Oncorhynchus mykiss*) populations using genome-wide SNP and haplotype data. *Animal Genetics*, 55, 87–98. Available from: https://doi.org/10.1111/age.13378

13652052, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/age.13378 by Swedish University Of Agricultural Sciences, Wiley Online Library on [15/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License