Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Development of a novel microhaplotype panel for steelhead/rainbow trout (Oncorhynchus mykiss) and application for phylogenetic analysis in California

Ronan Le Gall
Université Brest, CNRS, IRD, Ifremer, LEMAR, F-29280 Plouzane
Noé Barthelemy
Université Brest, CNRS, IRD, Ifremer, LEMAR, F-29280 Plouzane
Anthony J. Clemento
Institute of Marine Sciences, University of California, Santa Cruz
Cassondra D. Columbus
Institute of Marine Sciences, University of California, Santa Cruz
Ellen Campbell
Institute of Marine Sciences, University of California, Santa Cruz
Elenea C. A. Correa
Institute of Marine Sciences, University of California, Santa Cruz
Jeff A. Rodzen
California Department of Fish and Wildlife, Genetics Research Laboratory
John Carlos Garza
Southwest Fisheries Science Center, National Marine Fisheries Service
Devon E. Pearse (devon.pearse@noaa.gov)
Southwest Fisheries Science Center, National Marine Fisheries Service

Method Article

Keywords: genomics, microhaplotypes, genotyping, Oncorhynchus mykiss, phylogeography

Posted Date: June 6th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2949400/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

The rapid advance of high-throughput sequencing has prompted a transition in wildlife and fisheries genetics from using microsatellites toward markers that are more amenable to genotyping by sequencing. Microhaplotypes are novel multi-allelic genetic markers that utilize a high-throughput genomic amplicon sequencing approach to genotype large numbers of individuals for parentage and kinship analysis and population genetic studies, including applications in monitoring and fisheries management. We describe the development of a panel of microhaplotypes for *Oncorhynchus mykiss*, a species of high cultural and economic importance both in its native range in the North American and the Kamchatka Peninsula of northeast Asia, and globally through introductions for aquaculture and due to its reputation as a prized sport fish among recreational fishers. The panel includes 124 loci presumed to be neutral, a marker for the sex determination locus (*SdY)*, and 10 loci targeting previously identified adaptive genomic variants associated with important life-history traits in this species. We demonstrate that this panel provides high resolution for phylogeographic and other genetic analysis and provide an initial standardized reference population genetic baseline of California *O. mykiss*.

Introduction

The distribution of genetic variation within and among populations and the patterns of gene flow between populations have been studied with a variety of genetic markers through the years. Historically, the nuclear genetic markers most commonly used in population genetics and parentage analysis were allozymes and microsatellites, and more recently single-nucleotide polymorphism (SNPs), the most abundant form of variation in the genome for most species (Brumfield et al., 2003; Allendorf et al., 2010). When genotyped using single-locus assays, SNPs are characterized by low error rates, easy and fast genotyping with no calibration problems between genotyping platforms and laboratories, and simulations demonstrate their utility for parentage analysis (Anderson & Garza, 2006; Seeb et al., 2009). Rapid advances in technology have contributed to the growth of a new approach called Genotyping by Sequencing (GBS; Davey et al., 2011) that leverages the high-throughput DNA sequencing that is now dominant in molecular biology. Together with the associated changes in data handling and bioinformatics, GBS has dramatically increased the amount of data obtained for a lower cost with less time-consuming techniques. However, GBS generally surveys SNP variation, and because SNPs are typically bi-allelic, they do not provide the same power per locus as microsatellites (Narum et al., 2008; Hauser et al., 2011).

As addressing most questions in population biology does not require entire genome sequences, but rather a modest number of loci genotyped in a larger number of individuals, population geneticists tend to focus on a small number of genetic markers that enable more individuals to be genotyped. This has led to an emerging class of markers, microhaplotypes, which are characterized by two or more closely linked SNPs that can be genotyped together in a single marker (Baetscher et al., 2018). Because the multiple linked SNPs can appear in different allelic combinations, their combined sequences produce multiple haplotypes of tightly linked SNPs (Kidd et al., 2014; Oldoni et al., 2019). With their higher per locus

statistical power, microhaplotypes are particularly useful for pedigree reconstruction and categorical assignment (McKinney et al., 2017; Baetscher et al., 2018) as well as forensic applications (Pang et al., 2020). Microhaplotypes are amenable to highly reproducible data processing pipelines and provide much greater power per nucleotide of sequence data than approaches that focus on a single bi-allelic SNP. Microhaplotypes are also abundant in the genome, with a low genotyping error rates, and are becoming a popular molecular marker in genomics (Baetscher et al., 2018; Hendricks et al. 2018).

Study System

Rainbow Trout *Oncorhynchus mykiss* is a species of salmonid fish whose natural distribution ranges from northern Mexico to the Kamchatka Peninsula, Russia. The species includes a wide range of morphological differences, leading early researchers to describe more than 50 species in North America that have now been synonymized into ~ 14 subspecies (Busby et al., 1996; Pearse et al., 2011). Within the species *O. mykiss* a wide range of migratory behavior can be observed both within and among populations. This includes resident and anadromous life histories, as well as substantial variation in the timing and frequency of juvenile and adult migration. The anadromous form is termed “steelhead” trout, while the freshwater residents, that remain in freshwater throughout their lives, are referred to as “rainbow trout” (Behnke, 1992; Bagley & Gall, 1998). Steelhead usually spend one or two years in freshwater before migrating to the sea. After one to three years of growth, steelhead return to freshwater to spawn, typically returning to their natal stream. Unlike other Pacific salmon, *O. mykiss* can migrate and spawn multiple times (iteroparous; Shapovalov & Taft, 1954; Behnke, 1992; McPhee et al., 2007), despite a high mortality rate after the first spawn. Adaptive genomic variation associated with migratory life-history traits has been documented in this species, including a chromosomal inversion associated with expression of the resident and anadromous forms as well as variation in disease susceptibility (Pearse et al., 2019; Calboli et al. 2022). Similarly, other genetic loci have been linked to variation in migration timing (Waples et al., 2022) and age-at-return (Waters et al. 2021), further highlighting the need for simple yet efficient genotyping methods to target adaptive variation. In addition, the anadromous form has suffered major declines, especially in the southern part of its range (Swift et al., 1993; Clemento et al., 2009; Abadia-Cardoso *et al.*, 2016), and many populations are listed as threatened or endangered under the US Endangered Species Act (ESA; NOAA 2006). Thus, there is a continuing need for improved methods to understand the genetic diversity and gene flow among *O. mykiss* populations, monitor populations, and guide conservation and management to improve species resilience.

Materials & Methods

Microhaplotype Panel Development

SNP discovery and amplicon design

Double-digest RAD sequencing:

To discover suitable genomic targets for development of microhaplotype markers, we used a modified double-digest restriction site-associated DNA sequencing approach (ddRAD-seq; Peterson et al., 2012) on 32 individuals from 10 populations of *O. mykiss* and one population sample of coastal cutthroat trout, *O. clarkii clarkii* (Ascertainment samples; Supp. Table 1). Following double restriction enzyme digest using EcoR1 and Sph1, a five base pair barcode was ligated to each sample before pooling for size selection that targeted fragments in the range of 300–400bp using a Pippin Prep (Sage Science). The samples were sequenced in one run on a Miseq instrument (Illumina inc. Shen et al., 2005) using a 600-cycle paired-end sequencing kit.

ddRAD demultiplexing:

Raw reads obtained from the Illumina sequencing run were pre-filtered based on their average Phred-scaled base quality score (≥ 33). Using the *process_radtags* component of Stacks v1.48 (Catchen et al., 2013), reads were truncated to 325bp and then demultiplexed based on the unique 5bp individual barcode to assign reads to their corresponding individuals. Finally, Stacks was used to assemble reads into loci within and across individuals and populations with a minimum depth of coverage (-m) of four, the distance allowed between stacks of two (M), and distance between catalog loci of two (*n*).

Loci filtering:

Among the 29,024 potential loci, a total of 5,959 had more than two SNPs that were observed in at least 10 or more of the 32 individuals sequenced. Since our target amplicon insert length in the final panel was 100–105 bases, we selected only the loci that had two or more SNPs within 100–105 bases of each other and had enough non-variable bases on either side those variants to attempt primer design. This left us with 3,049 potential loci. Given the short lengths of the targeted sequences and the whole genome duplication event in the common ancestor of salmonid fishes (Berthelot et al., 2014; Lien et al., 2016), the risk of obtaining amplicons containing fully or even partially paralogous genomic regions was very high (Pearse et al., 2019). Consensus sequences from the 3,049 candidate loci were mapped to themselves using BLAT (Kent 2002). A strict filtering was applied using R (R core team 2022), which removed 2,218 'duplicate' targets that fully or partially matched another target locus. These 'duplicates' reflect mainly bioinformatics errors (Stacks errantly splitting reads into separate loci) and/or repetitive elements or paralogous regions in the genome. With only unique genomic regions represented in our filtered dataset, BLAT was used again to map our 831 remaining targets to an existing chromosome-scale genome assembly (Pearse et al., 2019) and using the same strict filtering, produced a list of 385 potential targets for the design of microhaplotype markers. Finally, we used the graphical interface of Stacks to assess overall variability, the potential for primer design, and the number of potential population-specific alleles of our loci in the 32 sequenced individuals and used these criteria to select loci for primer design.

Primer design:

We designed primers for 192 variable loci with the software Primer 3 (Untergasser et al., 2012) implemented in Geneious v.R11 (Kearse et al., 2012), using the Santa Lucia (1998) melting temperature

(Tm) calculation and salt correction method. The length range of primers was 18–27 bp (target length of 20 bp) and contained between 25 and 50% GC bases (optimal content of 50%), allowing a max Tm difference of 2°C between primers and otherwise using Primer 3 default parameters. We targeted an optimal product size of 130 bp (in the range of 90–145 bp), because short and uniform lengths of target sequences are important factors for PCR success (*i.e.*, uniform amplification among loci). Following initial testing and evaluation, loci with poor amplification or low polymorphism were removed from the panel resulting in the final list of 124 presumably neutral loci (Supp. Table 2).

Adaptive genetic variation and sex informative loci:

In addition to the newly discovered loci from the ddRAD data, markers for several previously identified functional gene regions were added to the panel. First, five microhaplotype markers were designed within the chromosomal inversion complex present on chromosome *Omy0*5, known to be strongly associated with expression of anadromous or resident migratory life-history phenotypes in some *O. mykiss* populations (Pearse et al., 2014, 2019). Second, we designed primers for microhaplotype loci targeted on five SNPs associated with run-timing in the *Greb1L* gene region on *Omy*28 (Waples et al. 2022). Finally, the 'Omy-Y1-2Sexy' locus was included by using the primers from Brunelli, Steele, and Thorgaard (2010). This marker amplifies only when the Y chromosome is present (*i.e.*, in males) and has been shown to be highly accurate in identifying males and females in coastal California steelhead (Rundio et al., 2012; Pearse et al., 2019).

Genotyping-by-sequencing

To be able to conduct 'genotyping in thousands' (GT-seq), Campbell, Harmon, and Narum (2015) developed a genotyping by sequencing (GBS) method to optimize the sequencing capacity of NGS technologies for population genetics and parentage studies. We used GT-seq to sequence up to 384 individuals with 135 microhaplotype loci in a single Illumina MiSeq® run, using a 150-cycle paired-end approach. All other details of the thermal cycling and library preparation are as in Baetscher et al. (2018).

Bioinformatics processing and panel finalization

Reads were de-multiplexed by the MiSeq Analysis Software (Illumina inc. Shen et al., 2005). Paired-end reads were combined using the Fast Length Adjustment of SHort reads (FLASH Magoč and Salzberg 2011) and mapped to the Stacks consensus sequences for the target loci using the Burrows-Wheeler Aligner (BWA-MEM, Li and Durbin 2009). Mapped reads were converted from Sequence Alignment/Map (SAM) files to Binary Alignment/Map (BAM) files with SAMtools (Li et al., 2009). We identified variable sites using FreeBayes (Garrison and Marth 2012); the positions of all SNPs for each locus were recorded in a variant call format (VCF) file. This VCF was then passed to *microhaplot* (Ng *et al.*, DOI: 10.5281/zenodo.820110), a microhaplotype dedicated software implemented as an R (R core team 2022) package and associated Shiny app (http://shiny.rstudio.com/), which assembled the SNPs for each amplicon into a microhaplotype using the SAM files. After filtering loci for a minimum read depth

coverage of 10 per individual and an allelic balance ratio of 0.3, the software was then used to export the microhaplotypes for downstream analyses.

Because the inference of genetic sex with the 'Omy-Y1-2Sexy' locus is based on non-amplification in females, the expected read depth for females is zero. However, due to individual barcode misidentification and other genotyping errors, some reads could possibly be incorrectly assigned to females. Thus, based on the observed distribution of read counts the threshold for the inference of female sex was set to a maximum of 5 reads.

Four sequencing runs were performed, with from one to four plates of 96 individuals at the same time. After each run, the variability across loci was assessed, and loci were filtered according to the following criteria: read depth, inconsistent allelic balance across individuals, deviation from Hardy-Weinberg equilibrium (HWE), and the presence of more than two haplotypes per individual, likely due to paralog loci or index sequencing errors (Larsson et al., 2017). Finally, primers associated with extremely high read-depth loci were diluted or removed, in order to limit their over-representation in the sequencing pools. The final panel was composed of 124 loci for parentage and population genetic analysis, 10 adaptive loci in the Omy05 and *greb1L* regions, and one locus to identify genetic sex in our individuals.

Phylogeographic Utility

In addition to the individuals genotyped for the panel development, samples from 28 additional populations were genotyped along with additional samples for some populations that were already included. This study is based on 124 loci genotyped in a total of 1,831 tissue samples from 58 populations (Supp. Table 1).

Genetic diversity estimates (*i.e.*, He, Ho, Na, Fis, pairwise Fst) were calculated for each population and across all populations using the MStoolkit (Park, 2008) as well as the R package 'diveRsity' (R core team 2022; Keenan et al., 2013), and Allelic Richness (A_R) was calculated using the software HPrare v1.1 (Kalinowski 2005). Genetic distances between populations were also analyzed with phylogenetic trees from the software PHYLIP (Felsenstein, 1993). A neighbour-joining tree (Saitou & Nei, 1987), representing genetic distance between populations, was calculated with PHYLIP using pairwise chord distances (Cavalli-Sforza & Edwards, 1967). The stability of the tree topology was examined using the Seqboot program, with 1,000 bootstrap replicates. Discriminant analysis of principal components (DAPC) was performed with the R package 'adegenet' (Jombart, Devillard & Balloux, 2010). The number of axes that maximize the results of the DAPC was defined by a "DAPC Cross-Validation" test (Jombart & Collins, 2015), and 150 PCA components were kept for the population panel, for each DAPC. In order to observe core patterns more clearly, each of the multivariate analyses had clearly differentiated populations removed for the next DAPC. Finally, individual-based ancestry evaluations were also implemented using the model-based clustering program STRUCTURE (Pritchard et al., 2000). Values of K were evaluated as follows: K = 2–10, 15, 20, 25, and 30. STRUCTURE output was then analyzed with CLUMPAK (Kopelman et al., 2015) and DISTRUCT (Rosenberg, 2004).

Results

Panel Validation

Overall, 754 individuals were successfully genotyped for the microhaplotype panel validation, and 96.8% of the 124 presumably neutral loci successfully genotyped for more than 90% of all individuals in each population. No consistent deviations from HWE were observed across populations. Mean global heterozygosity across all loci was high (0.42), with a total of 847 microhaplotype alleles distributed across all populations.

Population Genetics Statistics

Estimates of genetic diversity were calculated using the final dataset consisting of 124 loci and 1,831 genotyped individuals from 58 populations. No consistent deviations from HWE observed across populations. Average allelic richness among populations was 1.79, mean values of both expected and observed heterozygosity were 0.26, and the average number of alleles per locus was 2.18. The Sheepheaven Creek samples displayed the lowest genetic diversity estimates for all metrics (Supp. Table 1), with expected heterozygosity (0.06), observed heterozygosity (0.06), allelic richness (1.15), and number of alleles per loci (1.18). On the other hand, steelhead from Feather River Hatchery and Mokelumne River Hatchery showed the highest expected heterozygosity (0.40) and observed heterozygosity (0.40) values, and Nimbus Hatchery had a similarly high observed heterozygosity (0.40). Coleman National Fish Hatchery of Battle Creek displayed the highest allelic richness (2.36). The mean values of all genetic diversity estimates were higher for coastal samples (He = 0.33, Ho = 0.33, Ar = 2.06, N_A = 2.60) than for inland rainbow trout populations (He = 0.18, Ho = 0.19, Ar = 1.51, N_A = 1.76), but no significant differences were found.

Pairwise Fst estimates (Supp. Table 3) were highest between the outgroup, Cutthroat trout (1), and *O. mykiss* populations, with a mean Fst of 0.660 between those two species. Among the *O. mykiss* samples the mean Fst was 0.354, with Golden Trout complex samples (38–49) showing the greatest differentiation from the rest (mean Fst = 0.486; Supp. Table 3). Within the Golden Trout complex, the 'Wyoming' California Golden Trout hatchery strain (55) were similar to Golden Trout Creek populations of wild California Golden Trout, suggesting an absence of strong divergence despite multiple generations in captivity. Finally, the Pit River Hatchery (52), Spring Creek (6) and Coldwater Canyon Creek (19) populations stood out from other populations with consistently high pairwise Fst estimates in comparison to the rest of the populations. In contrast, the coastal and Central Valley populations showed relatively low genetic differentiation.

Neighbor-Joining Tree

The neighbor-joining tree highlighted previously known phylogenetic patterns among *O. mykiss* populations, with many nodes having strong bootstrap support (Fig. 1). For example, all Golden Trout complex samples were tightly clustered together, including the Golden Trout hatchery strain of Wyoming,

with a bootstrap value of 97% in the Neighbor-Joining tree (Fig. 1). Similarly, the Redband Trout of Deep, Shields, and Buck Creeks strongly clustered together, along with Spring Creek, an upper Klamath tributary (Fig. 1). Central Valley below-dam populations showed mixed ancestry, the majority of them clustered together along with the domesticated rainbow trout strains, with the exception of steelhead at Nimbus Hatchery that grouped with the coastal cluster, as expected given their lineage (Pearse & Garza, 2015).

DAPC: Discriminant Analysis of Principal Components

DAPC was used to visualize the differentiation and relationships of all the *O. mykiss* populations (Fig. 2). For each DAPC, populations that exhibited clear differentiation in the previous DAPC were removed to allow more detailed relationships among the remaining populations to resolve. The First DAPC (Fig. 2A) showed the cutthroat trout outgroup (1) isolated from all *O. mykiss* populations, as expected. With cutthroat trout removed, DAPC only including *O. mykiss* populations showed a significant separation between golden trout complex (38–49), the Wyoming strain of hatchery golden trout (55), and all other populations (Fig. 2B). Pit River Hatchery (52) and Coldwater Canyon Creek (19) populations also showed clear differentiation from the other *O. mykiss* populations. Interestingly, one individual from the Kamloops Hatchery (57) clustered with the Golden Trout complex. These results were also shown in the third DAPC (Fig. 2C). On this multivariate analysis, the redband trout subspecies (21–25) were differentiated from the rest. Northern coastal populations (2–8) also slightly displayed a break from the central group; this split is most visible on the last DAPC (Fig. 2D), in which Central Valley steelhead populations below dams (30–33, 35–36) also showed strong links with coastal populations.

STRUCTURE results (Supp. Figure 1) were clear and consistent among the 10 replicates made at all values of K. Since the presence of a highly divergent outgroup would not have yielded informative results on the genetic structure of Californian *O. mykiss*, the Cutthroat trout were excluded from this analysis. As in the DAPC analysis, the first split observed with STRUCTURE was between the Golden Trout complex and other *O. mykiss* populations. Then at K = 10 additional patterns are clear, including separation between coastal and inland populations, and an interior Redband Trout group consisting of Buck Creek, North Fork Shields Creek, South Fork Shields Creek, North Fork Deep Creek, and South Fork Deep Creek. At higher K-values, these patterns remain and most populations are clearly distinct (Supp. Figure 1).

Adaptive genetic variation and sex informative loci:

In addition to the microhaplotype loci used for population genetic analyses, we developed 10 microhaplotype loci in regions of adaptive genetic variation (*Omy*5 and *Greb1L*). The five microhaplotype markers designed within the inversion complex present on the chromosome 5 (*Omy0*5) produced a total of eight variable SNPs. Similarly, the five microhaplotype loci targeted in the region of the *Greb1L* gene on *Omy*28 contained six variable SNPs. Summary genotype frequencies for a single key SNP for each of these regions provide information on the distribution of known adaptive genetic variants in these regions (Supp. Table 1; G*reb1L*: mhap8_71 = 11667915 *Omy05*: omy5_9_54854574-19).

The ‘Omy-Y1-2Sexy’ marker and the G*reb1L* and *Omy05* microhaplotypes all amplified successfully in *O. clarkii* as well as *O. mykiss*. However, while these results demonstrate that these loci amplify in this species, they do not necessarily indicate that the same associations between these variants and specific life-history traits exist.

Finally, the ‘Omy-Y1-2Sexy’ marker, which amplifies only when the Y chromosome is present (*i.e.*, in males) provided highly accurate information on sex for the subset of individuals for which morphological sex information was available. All individuals identified as female had 0, 1, or 2 reads except for one individual with 44 reads, likely indicating a misidentified male. Similarly, with the exception of one male that had zero reads, indicating a possible metadata error, all other known males had > 8 and most had > 50 reads, clearly differentiating males and females.

Discussion

Variability of the microhaplotype panel

The novel microhaplotype panel described here contains much more variation than the 96 SNP panel used previously with *O. mykiss* (*e.g.* Abadia-Cardoso *et al.*, 2013; Pearse and Campbell 2018), providing improved resolution for genetic analysis comparable to microsatellite studies in this species (*e.g.*, Pearse et al., 2009; Garza et al. 2014; Pearse and Garza 2015). However, unlike microsatellites, microhaplotypes benefit from the same low error rates and ease of genotyping as SNPs, making them amenable to use with high-throughput genotype-by-sequencing pipelines. Thus, this panel of markers provides a valuable new tool for researchers, combining the main advantages of microsatellites and SNPs into a high-throughput genotyping approach (for an empirical comparison of SNPs and microsatellites see Glaubitz et al., 2003; Hauser et al., 2011). These markers will be especially useful for studies of kinship and parentage, where multi-allelic loci provide significant advantages over bi-allelic markers (Glaubitz et al., 2003; Baetscher et al., 2018).

Genetic sex determination

As in previous studies (Rundio et al., 2012; Pearse et al., 2019; Kelson et al., 2020), the ‘Omy-Y1-2Sexy’ genetic sex marker used here was very accurate in determining sex of adults from multiple populations, providing high confidence information about the sex of the genotyped individuals. However, the presence/absence detection mechanism of this marker makes it sensitive to technical errors such as failure to amplify in a male (*i.e* resulting in a ‘false-negative’ female) or index misidentification. In addition, ‘Omy-Y1-2Sexy’ was designed on the *sdY* gene, which is the master-sex determining gene in *O. mykiss* (Yano et al., 2013). However, the *sdY* gene is often transposed to a different location of the genome in salmonid species (Phillips 2013), so caution should be used in interpreting the information provided by this marker in novel populations.

Adaptive Genetic Variation

In the populations studied here, the proportions of anadromous-associated alleles (A) in the Omy05 region and alternative alleles in the *Greb1L* region are concordant with both previous estimations in some of the same populations (e.g. Pearse et al., 2014, 2019; Waples et al. 2022) and with expectations based on the habitats in which they were sampled. In addition to the information provided by single SNPs in both of these regions, together these SNPs provide information on the linkage disequilibrium and frequency of discordant genotypes, extending our understanding of the structure of adaptive genomic variation in these regions. Furthermore, the flexibility of the amplicon sequencing approach will also allow additional loci to be added to the panel in the future to further assess variation in the associations of specific SNPs, or markers associated with other adaptive phenotypic variation, such as loci recently identified as important for additional life-history traits (e.g. Six6 and vgll3; Waters et al., 2021).

Phylogeography of *O. mykiss*

The relationships recovered using the microhaplotype panel in this study clearly resolved the known population genetic structure of the study populations and concord with the broad-scale patterns known from previous studies. Cutthroat trout were strongly differentiated from *O. mykiss* in all the analyses, and the patterns of differentiation among the coastal, Central Valley, and inland *O. mykiss* populations (Buchanan et al., 1994; Pearse et al., 2011; Pearse and Garza 2015; Leitwein et al. 2017) were clearly visible in the DAPC and STRUCTURE results, as well as in the phylogeographic trees. Similarly, previous studies have shown that coastal *O. mykiss* populations above and below the dams or waterfall barriers are often genetically more differentiated between basins than populations within the same basin (Clemento et al., 2009; Pearse et al., 2009), demonstrating that they still share recent common ancestry. Furthermore, results from STRUCTURE and phylogeographic trees were consistent with previous studies that showed a pattern of isolation-by-distance among below-barrier populations (Pearse et al., 2007; Garza et al., 2014; Pearse and Garza 2015).

Among inland rainbow trout populations, the first group that showed strong genetic differentiation within all the results was the Golden Trout complex, a group of inland populations in the southern Sierra Nevada mountain range. Golden Trout have long been isolated from the sea and have diverged from other *O. mykiss* populations. Concordant with previous work, our analyses revealed two major lineages within the golden trout complex, one representing the populations of the Kern River (Little Kern Golden Trout *O. m. whitei* and Kern River Rainbow Trout *O. m. gilberti*), and another containing California Golden Trout (*O. m. aguabonita*) from Golden Trout Creek. These two clusters diverged within the last 5,000–10,000 years, since Golden Trout Creek was isolated above a waterfall that acts as a complete barrier (Cordes et al. 2006). However, the Chagoopa Creek population clustered with the Kern River lineage despite being above the waterfall, a pattern that can be explained by the fact that Kern River Golden Trout were heavily introduced from the lower Kern River into Chagoopa Creek in the past (Stephens *et al.*, 2006). Like the Golden Trout, the Sheepheaven Creek and Moosehead Creek samples of McCloud River Redband Trout (*O. m. stonei*) were clearly differentiated from other *O. mykiss* and also displayed low genetic diversity estimates, likely because they are small isolated populations with low effective size and limited gene flow (Nielsen et al., 1999; Simmons et al., 2010). Similarly, some small, isolated, southern coastal populations

also had very low genetic diversity (Garza et al., 2014), as well as introgression or complete replacement of wild population by hatchery fish (Abadia-Cardoso *et al.*, 2016).

Conclusions

The suite of novel microhaplotype markers described here provides a high-throughput amplicon sequencing approach to genotype large numbers of individuals for applications in monitoring and fisheries management, including parentage and kinship analysis. These markers offer high resolution for phylogeographic and other genetic analysis in a genotyping by sequencing framework and provide a population genetic baseline of California *O. mykiss* on which future studies can expand.

Declarations

Acknowledgements

This work was supported by the NOAA Southwest Fisheries Science Center, Interagency Agreement (IA) R15PG0006 between the U.S Department of the Interior, Bureau of Reclamation and the National Marine Fisheries Service, ISblue project, Interdisciplinary graduate school for the blue planet (ANR-17-EURE-0015), and by a grant from the French government under the program "Investissements d'Avenir". The authors thank Eric Anderson for help with primer design and for reviewing the draft manuscript. The authors declare no competing interests.

Authors contributions

LeGall, Barthelemy, Clemento, Rodzen, Garza, and Pearse designed the study; Rodzen, Garza, and Pearse provided biological materials; Columbus, Campbell, Correa, Le Gall, and Barthelemy conducted laboratory work and analyzed the sequence and marker data; Le Gall, Barthelemy, and Clemento conducted the population genetics analyses and made figures; Le Gall, Barthelemy, and Pearse wrote the manuscript. All authors approved the final manuscript.

All samples were collected non-lethally and in accordance with Federal and State regulations and approved IACUC animal care procedures at the institutions at which they were handled.

References

1. Abadía-Cardoso A, Anderson EC, Pearse DE, Carlos Garza J (2013) Large-scale parentage analysis reveals reproductive patterns and heritability of spawn timing in a hatchery population of steelhead (Oncorhynchus mykiss). Molecular Ecology 22(18). https://doi.org/10.1111/mec.12426
2. Abadía-Cardoso A, Pearse DE, Jacobson S, Marshall J, Dalrymple D, Kawasaki F, Ruiz-Campos G, Garza JC (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(3). https://doi.org/10.1007/s10592-016-0814-9

3. Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. Nature Reviews Genetics 11
4. Anderson EC, Garza JC (2006) The power of single-nucleotide polymorphisms for large-scale parentage inference. Genetics 172(4). https://doi.org/10.1534/genetics.105.048074
5. Baetscher DS, Clemento AJ, Ng TC, Anderson EC, Garza JC (2018) Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. Molecular Ecology Resources 18(2). https://doi.org/10.1111/1755-0998.12737
6. Bagley MJ, Gall GAE (1998) Mitochondrial and nuclear DNA sequence variability among populations of rainbow trout (Oncorhynchus mykiss). Molecular Ecology 7(8). https://doi.org/10.1046/j.1365-294x.1998.00413.x
7. Behnke RJ (1992) Native trout of western North America. Amer Fisheries Society
8. Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B, Bento P, da Silva C, Labadie K, Alberti A, Aury JM, Louis A, Dehais P, Bardou P, Montfort J, Klopp C, Cabau C, Gaspin C, Thorgaard GH, Boussaha M, Quillet E, Guyomard R, Galiana D, Bobe J, Volff JN, Genêt C, Wincker P, Jaillon O, Crollius HR, Guiguen Y (2014) The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nature Communications 5. https://doi.org/10.1038/ncomms4657
9. Brumfield RT, Beerli P, Nickerson DA, Edwards S v. (2003) The utility of single nucleotide polymorphisms in inferences of population history. Trends in Ecology and Evolution 18
10. Brunelli JP, Steele CA, Thorgaard GH (2010) Deep divergence and apparent sex-biased dispersal revealed by a Y-linked marker in rainbow trout. Molecular Phylogenetics and Evolution 56(3):983–990. https://doi.org/10.1016/j.ympev.2010.05.016
11. Buchanan D v, Hemmingsen AR, Currens KP (1994) Annual progress report. Native trout project. Oregon Department of Fish and Wildlife, Fish Research Project F-136-R-07, Annual Progress Report, Portland, OR.
12. Busby PJ, Wainwright TC, Bryant GJ, Lierheimer LJ, Waples RS, Waknitz FW, Lagomarsino IV (1996) Status review of west coast steelhead from Washington, Idaho, Oregon, and California. NOAA Technical Memorandum NMFS-NWFSC
13. Calboli FCF, Koskinen H, Nousianen A, Fraslin C, Houston RD, Kause A (2022) Conserved QTL and chromosomal inversion affect resistance to columnaris disease in 2 rainbow trout (*Oncorhyncus mykiss*) populations. G3 Genes|Genomes|Genetics. https://doi.org/10.1093/g3journal/jkac137
14. Campbell NR, Harmon SA, Narum SR (2015) Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. Molecular Ecology Resources 15(4). https://doi.org/10.1111/1755-0998.12357
15. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: An analysis tool set for population genomics. Molecular Ecology 22(11). https://doi.org/10.1111/mec.12354
16. Cavalli-Sforza LL, Edwards AW (1967) Phylogenetic analysis. Models and estimation procedures. American Journal of Human Genetics 19(3). https://doi.org/10.2307/2406616

17. Clemento AJ, Anderson EC, Boughton D, Girman D, Garza JC (2009) Population genetic structure and ancestry of Oncorhynchus mykiss populations above and below dams in south-central California. Conservation Genetics 10(5). https://doi.org/10.1007/s10592-008-9712-0
18. Cordes JF, Stephens MR, Blumberg MA, May B (2006) Identifying Introgressive Hybridization in Native Populations of California Golden Trout Based on Molecular Markers. Trans Am Fish Soc 135(1). https://doi.org/10.1577/t05-120.1
19. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics 12
20. Felsenstein J (1989) PHYLIP–Phylogenetic Inference Package. Cladistics 5
21. Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing – Free bayes – Variant Calling – Longranger. arXiv preprint arXiv:12073907
22. Garza JC, Gilbert-Horvath EA, Spence BC, Williams TH, Fish H, Gough SA, Anderson JH, Hamm D, Anderson EC (2014) Population Structure of Steelhead in Coastal California. Trans Am Fish Soc 143(1). https://doi.org/10.1080/00028487.2013.822420
23. Glaubitz JC, Rhodes OE, Dewoody JA (2003) Prospects for inferring pairwise relationships with single nucleotide polymorphisms. Molecular Ecology 12(4). https://doi.org/10.1046/j.1365-294X.2003.01790.x
24. Hauser L, Baird M, Hilborn R, Seeb LW, Seeb JE (2011) An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (Oncorhynchus nerka) population. Molecular Ecology Resources 11(SUPPL. 1). https://doi.org/10.1111/j.1755-0998.2010.02961.x
25. Hendricks S, Anderson EC, Antao T, Bernatchez L, Forester BR, Garner B, Hand BK, Hohenlohe PA, Kardos M, Koop B, Sethuraman A, Waples RS, Luikart G (2018) Recent advances in conservation and population genomics data analysis. Evolutionary Applications 11(8). https://doi.org/10.1111/eva.12659
26. Jombart T, Collins C (2015) A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet. Rvignette
27. 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. https://doi.org/10.1186/1471-2156-11-94
28. Kalinowski ST (2005) HP-RARE 1.0: A computer program for performing rarefaction on measures of allelic richness. Molecular Ecology Notes 5(1). https://doi.org/10.1111/j.1471-8286.2004.00845.x
29. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12). https://doi.org/10.1093/bioinformatics/bts199
30. Keenan K, Mcginnity P, Cross TF, Crozier WW, Prodöhl PA (2013) DiveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. Methods

in Ecology and Evolution 4(8). https://doi.org/10.1111/2041-210X.12067

31. Kelson SJ, Carlson SM, Miller MR (2020) Indirect genetic control of migration in a salmonid fish. Biology Letters 16(8). https://doi.org/10.1098/rsbl.2020.0299
32. Kent WJ (2002) BLAT-The BLAST-Like Alignment Tool Resource 656 Genome Research. Genome Res 12(4)
33. Kidd KK, Pakstis AJ, Speed WC, Lagacé R, Chang J, Wootton S, Haigh E, Kidd JR (2014) Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics. Forensic Science International: Genetics 12. https://doi.org/10.1016/j.fsigen.2014.06.014
34. Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources 15(5). https://doi.org/10.1111/1755-0998.12387
35. Larsson A, Stanley G, Sinha R, Weissman I, Sandberg R (2017) Computational correction of cross-contamination due to exclusion amplification barcode spreading. bioRxiv
36. Leitwein M, Garza JC, Pearse DE (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(1). https://doi.org/10.1111/eva.12416
37. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14). https://doi.org/10.1093/bioinformatics/btp324
38. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16). https://doi.org/10.1093/bioinformatics/btp352
39. Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, Leong JS, Minkley DR, Zimin A, Grammes F, Grove H, Gjuvsland A, Walenz B, Hermansen RA, von Schalburg K, Rondeau EB, di Genova A, Samy JKA, Olav Vik J, Vigeland MD, Caler L, Grimholt U, Jentoft S, Inge Våge D, de Jong P, Moen T, Baranski M, Palti Y, Smith DR, Yorke JA, Nederbragt AJ, Tooming-Klunderud A, Jakobsen KS, Jiang X, Fan D, Hu Y, Liberles DA, Vidal R, Iturra P, Jones SJM, Jonassen I, Maass A, Omholt SW, Davidson WS (2016) The Atlantic salmon genome provides insights into rediploidization. Nature 533. https://doi.org/10.1038/nature17164
40. Magoč T, Salzberg SL (2011) FLASH: Fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27(21). https://doi.org/10.1093/bioinformatics/btr507
41. McKinney GJ, Seeb JE, Seeb LW (2017) Managing mixed-stock fisheries: Genotyping multi-SNP haplotypes increases power for genetic stock identification. Canadian Journal of Fisheries and Aquatic Sciences 74(4). https://doi.org/10.1139/cjfas-2016-0443
42. McPhee M v., Utter F, Stanford JA, Kuzishchin K v., Savvaitova KA, Pavlov DS, Allendorf FW (2007) Population structure and partial anadromy in Oncorhynchus mykiss from Kamchatka: Relevance for conservation strategies around the Pacific Rim. Ecology of Freshwater Fish 16(4). https://doi.org/10.1111/j.1600-0633.2007.00248.x

43. Narum SR, Banks M, Beacham TD, Bellinger MR, Campbell MR, Dekoning J, Elz A, Guthrie CM, Kozfkay C, Miller KM, Moran P, Phillips R, Seeb LW, Smith CT, Warheit K, Young SF, Garza JC (2008) Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. Molecular Ecology 17(15). https://doi.org/10.1111/j.1365-294X.2008.03851.x
44. Nielsen JL, Crow KD, Fountain MC (1999) Microsatellite diversity and conservation of a relic trout population: McCloud River redband trout. In: Molecular Ecology
45. NOAA (2006) Department of Commerce National Endangered and threatened species: final listing determinations for 10 distinct population segments of west coast steelhead
46. Oldoni F, Kidd KK, Podini D (2019) Microhaplotypes in forensic genetics. Forensic Science International: Genetics 38
47. Pang JB, Rao M, Chen QF, Ji AQ, Zhang C, Kang KL, Wu H, Ye J, Nie SJ, Wang L (2020) A 124-plex Microhaplotype Panel Based on Next-generation Sequencing Developed for Forensic Applications. Scientific Reports 10(1). https://doi.org/10.1038/s41598-020-58980-x
48. Park S (2008) Excel Microsatellite Toolkit. Version 3.1. 1. Animal Genomics Lab website,(University College, Dublin, Ireland)
49. Pearse DE, Donohoe CJ, Garza JC (2007) Population genetics of steelhead (Oncorhynchus mykiss) in the Klamath River. Environmental Biology of Fishes 80(4). https://doi.org/10.1007/s10641-006-9135-z
50. Pearse DE, Hayes SA, Bond MH, Hanson C v., Anderson EC, MacFarlane RB, Garza JC (2009) Over the falls? Rapid evolution of ecotypic differentiation in steelhead/rainbow trout (Oncorhynchus mykiss). Journal of Heredity 100(5). https://doi.org/10.1093/jhered/esp040
51. Pearse DE, Gunckel SL, Jacobs SE (2011) Population structure and genetic divergence of coastal rainbow and redband trout in the Upper Klamath Basin. Trans Am Fish Soc 140(3). https://doi.org/10.1080/00028487.2011.583538
52. Pearse DE, Miller MR, Abadía-Cardoso A, Garza JC (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(1783). https://doi.org/10.1098/rspb.2014.0012
53. Pearse DE, Barson NJ, Nome T, Gao G, Campbell MA, Abadía-Cardoso A, Anderson EC, Rundio DE, Williams TH, Naish KA, Moen T, Liu S, Kent M, Moser M, Minkley DR, Rondeau EB, Brieuc MSO, Sandve SR, Miller MR, Cedillo L, Baruch K, Hernandez AG, Ben-Zvi G, Shem-Tov D, Barad O, Kuzishchin K, Garza JC, Lindley ST, Koop BF, Thorgaard GH, Palti Y, Lien S (2019) Sex-dependent dominance maintains migration supergene in rainbow trout. Nature Ecology and Evolution 3(12). https://doi.org/10.1038/s41559-019-1044-6
54. Pearse DE, Campbell MA (2018) Ancestry and Adaptation of Rainbow Trout in Yosemite National Park. Fisheries (Bethesda) 43(10). https://doi.org/10.1002/fsh.10136

55. Pearse DE, Garza JC (2015) You can't unscramble an egg: Population genetic structure of oncorhynchus mykiss in the California central valley inferred from combined microsatellite and single nucleotide polymorphism data. San Francisco Estuary and Watershed Science 13(4). https://doi.org/10.15447/sfews.2015v13iss4art1
56. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7(5). https://doi.org/10.1371/journal.pone.0037135
57. Phillips RB (2013) Evolution of the sex chromosomes in salmonid fishes. Cytogenetic and Genome Research 141(2–3). https://doi.org/10.1159/000355149
58. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2). https://doi.org/10.1093/genetics/155.2.945
59. R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria URL http://www R-project org
60. Rinne JN, Behnke RL (1994) Native Trout of Western North America. Copeia 1994(1). https://doi.org/10.2307/1446698
61. Rosenberg NA (2004) DISTRUCT: A program for the graphical display of population structure. Molecular Ecology Notes 4(1). https://doi.org/10.1046/j.1471-8286.2003.00566.x
62. Rundio DE, Williams TH, Pearse DE, Lindley ST (2012) Male-biased sex ratio of nonanadromous Oncorhynchus mykiss in a partially migratory population in California. Ecology of Freshwater Fish 21(2). https://doi.org/10.1111/j.1600-0633.2011.00547.x
63. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4). https://doi.org/10.1093/oxfordjournals.molbev.a040454
64. Seeb JE, Pascal CE, Ramakrishnan R, Seeb LW (2009) SNP genotyping by the 5'-nuclease reaction: advances in high-throughput genotyping with nonmodel organisms. Methods Mol Biol 578. https://doi.org/10.1007/978-1-60327-411-1_18
65. Shapovalov L, Taft AC (1954) The Life Histories of the Steelhead Rainbow Trout (Salmo gairdneri gairdneri) and Silver Salmon (Oncorhynchus kisutch) with Special Reference to Waddell Creek, California, and Recommendations Regarding Their Management. UC San Diego: Library – Scripps Digital Collection No. 98(Fish Bulletin)
66. Shen R, Fan JB, Campbell D, Chang W, Chen J, Doucet D, Yeakley J, Bibikova M, Garcia EW, McBride C, Steemers F, Garcia F, Kermani BG, Gunderson K, Oliphant A (2005) High-throughput SNP genotyping on universal bead arrays. Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis 573
67. Simmons RE, Lavretsky P, May B (2010) Introgressive Hybridization of Redband Trout in the Upper McCloud River Watershed. Trans Am Fish Soc 139(1). https://doi.org/10.1577/t08-245.1
68. Slatkin M (1987) Gene flow and the geographic structure of natural populations. Science (1979) 236(4803). https://doi.org/10.1126/science.3576198

69. Stephens SJ, McGuire C, Sims L (2004) Conservation Assessment and Strategy for the California Golden Trout (Oncorhynchus mykiss aguabonita) Tulare County, California. US Fish and Wildlife Service, Sacramento Office, Sacramento, California, USA
70. Swift CC, Haglund TR, Ruiz M, Fisher RN (1993) The Status and Distribution of the Freshwater Fishes of Southern California. Bulletin of the Southern California Academy of Sciences 92(3).
71. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3-new capabilities and interfaces. Nucleic Acids Research 40(15). https://doi.org/10.1093/nar/gks596
72. Waples RS, Ford MJ, Nichols K, Kardos M, Myers J, Thompson TQ, Anderson EC, Koch IJ, McKinney G, Miller MR, Naish K, Narum SR, O'Malley KG, Pearse DE, Pess GR, Quinn TP, Seamons TR, Spidle A, Warheit KI, Willis SC (2022) Implications of Large-Effect Loci for Conservation: A Review and Case Study with Pacific Salmon. J. Heredity 113(2): 121-144.
73. Waters, C.D., Clemento, A., Aykanat, T., Garza, J.C., Naish, K.A., Narum, S. and Primmer, C.R. (2021), Heterogeneous genetic basis of age at maturity in salmonid fishes. Mol Ecol, 30: 1435-1456. https://doi.org/10.1111/mec.15822
74. Yano A, Nicol B, Jouanno E, Quillet E, Fostier A, Guyomard R, Guiguen Y (2013) The sexually dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. Evolutionary Applications 6(3). https://doi.org/10.1111/eva.12032

Supplementary Figure

Supplementary Figure 1 is not available with this version

Figures

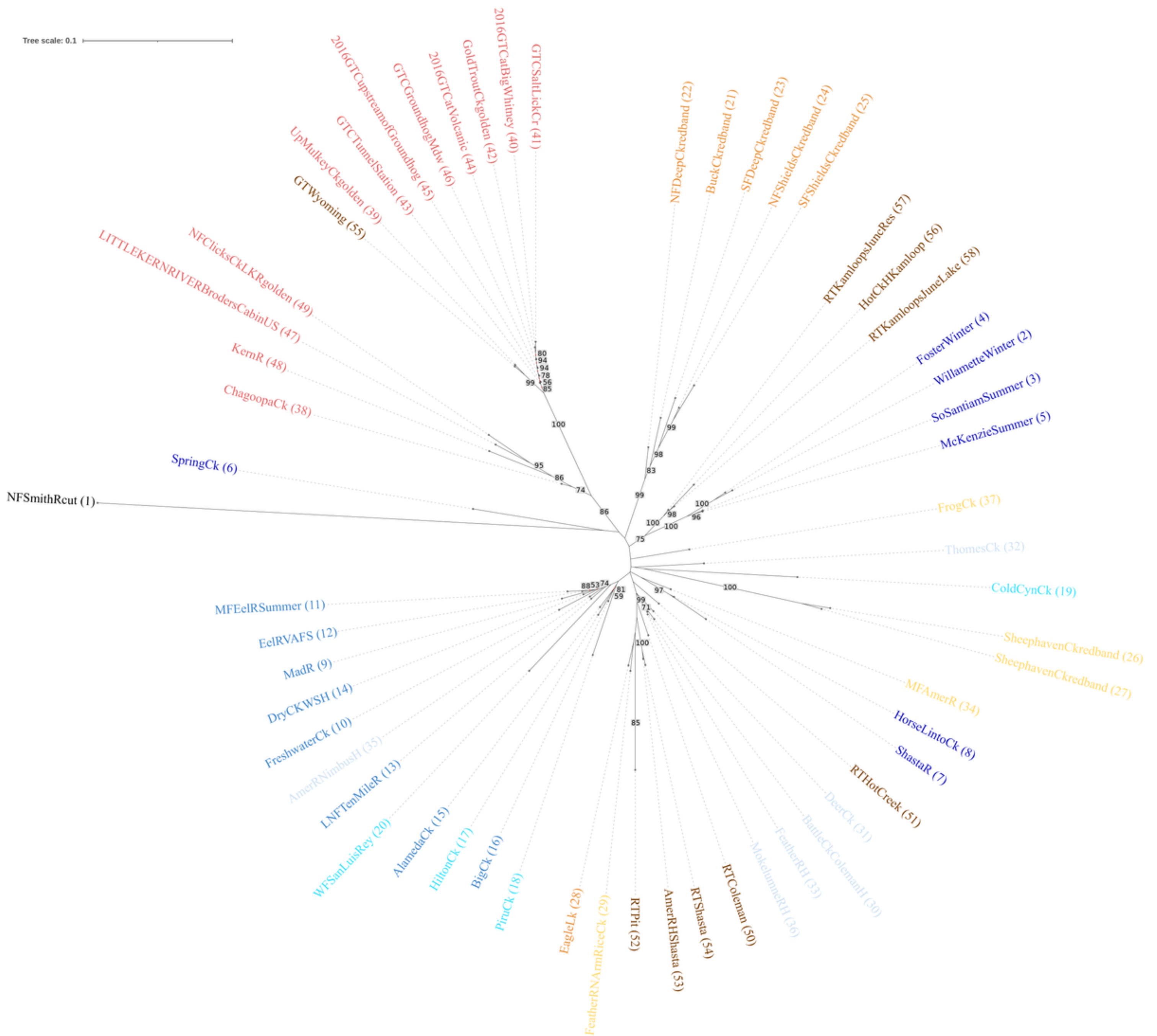

Figure 1

Neighbor joining tree of *O. mykiss* populations, showing distance between each nod. Constructed with chord distance, this tree displays all the population present in this study. The color code is the same as on Supp. Table 1, i.e. Black: Cutthroat trout, Blue: Coastal populations, Navy Blue: Northern Coastal, Turquoise: Southern Coastal, Light Blue: Central Valley Below dams, Yellow: Central Valley Above dams, Orange: Inland, Red: Golden Trout, Brown: Hatchery strain. Bootstrap consensus values from 1000 bootstrap replicates shown. Only bootstrap values above 50% are reported.

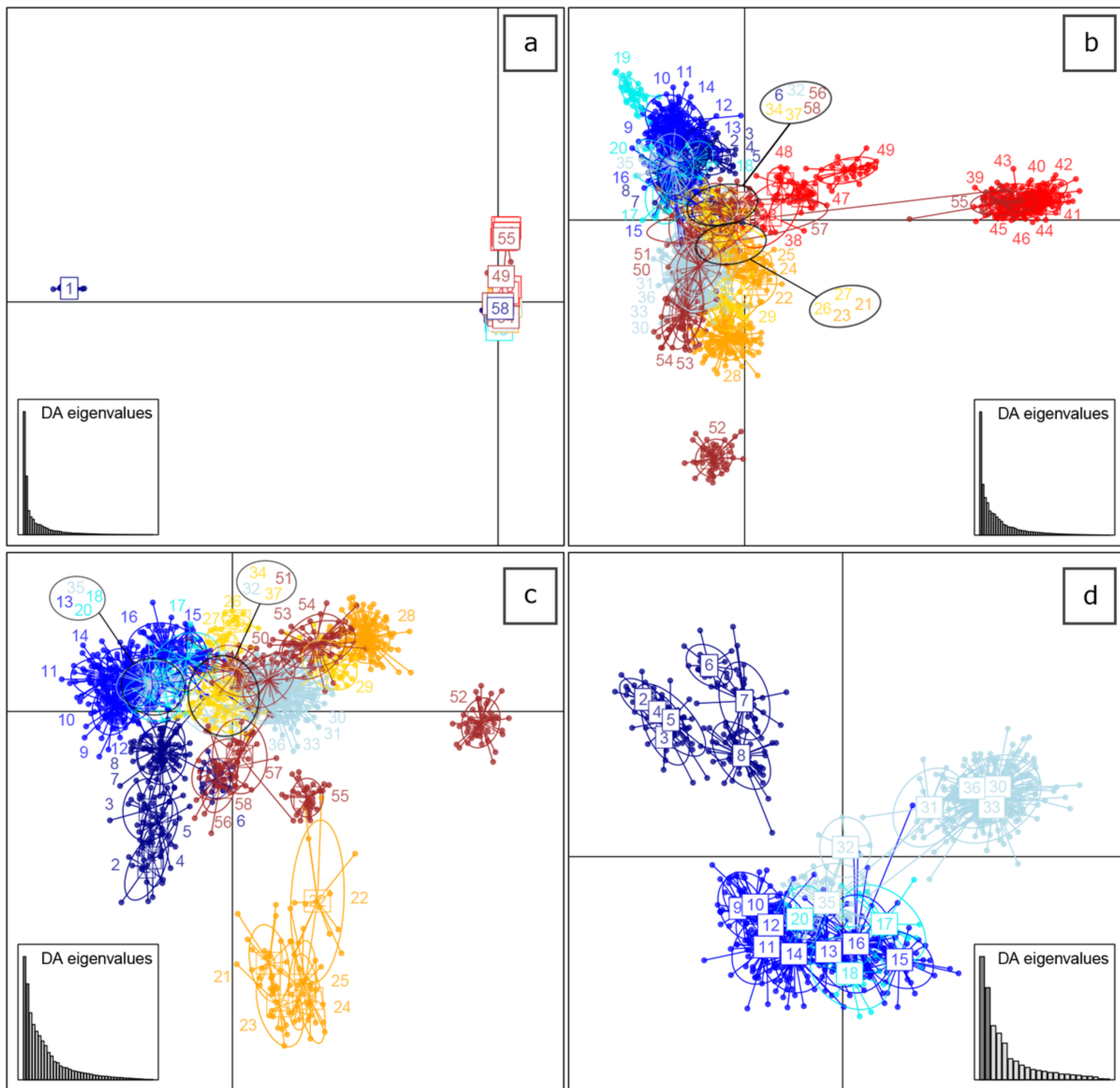

Figure 2

Four DAPCs showing population relationships on their first two axes that explained most of the variance. A = All the populations (1-58), B = *O. mykiss*populations without Cutthroat trout (2-58), C = *O. mykiss* populations except the golden trout and Coldwater Canyon Creek populations (2-18, 20-37, 50-58), D = Steelhead populations only (1-18, 20, 30-33, 35-36). The color code is the same as on Supp. Table 1, i.e. Black: Cutthroat trout, Blue: Coastal populations, Navy Blue: Northern Coastal, Turquoise: Southern Coastal, Light Blue: Central Valley Below dams, Yellow: Central Valley Above dams, Orange: Inland, Red: Golden Trout, Brown: Hatchery strain.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- LeGalletalGenotypeData.xlsx
- SuppTable1Submit.xlsx
- SuppTable2Primers.xlsx
- SuppTable3Fst.xlsx