Conservation Genetics Resources
https://doi.org/10.1007/s12686-024-01374-2

METHODS AND RESOURCES

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

Ronan Le Gall[1] · Noé Barthelemy[1] · Anthony J. Clemento[2] · Cassondra D. Columbus[2] · Ellen Campbell[2] · Elena C. A. Correa[2] · Jeff A. Rodzen[3] · John Carlos Garza[2,4,5] · Devon E. Pearse[2,4,6]

Received: 17 May 2023 / Accepted: 9 November 2024

Abstract

Advances in high-throughput sequencing and bioinformatic data processing have prompted a transition in wildlife and fisheries genetics from the use of allozymes, mtDNA, or microsatellites towards markers that are more amenable to genotyping by sequencing, increasing the amount of data obtained for a lower cost with less time-consuming techniques. 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 throughout its native range 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 on initial standardized reference population genetic baseline of California *O. mykiss*.

Keywords Genomics · Microhaplotypes · Genotyping · *Oncorhynchus mykiss* · Phylogeography

✉ Devon E. Pearse
devon.pearse@noaa.gov

Ronan Le Gall
ronan.legall@univ-brest.fr

Noé Barthelemy
noe.barthelemy@hotmail.fr

Anthony J. Clemento
anthony.clemento@noaa.gov

Cassondra D. Columbus
ccolumbu@ucsc.edu

Ellen Campbell
emcampbe@ucsc.edu

Elena C. A. Correa
elena.correa@noaa.gov

Jeff A. Rodzen
Jeff.rodzen@wildlife.ca.gov

John Carlos Garza
carlos.garza@noaa.gov

[1] Université Brest, CNRS, IRD, Ifremer, LEMAR, 29280 Plouzane, France

[2] Institute of Marine Sciences, University of California, Santa Cruz, CA, USA

[3] Genetics Research Laboratory, California Department of Fish and Wildlife, Sacramento, CA, USA

[4] Southwest Fisheries Science Center, National Marine Fisheries Service, Santa Cruz, CA, USA

[5] Department of Ocean Sciences, University of California, Santa Cruz, CA, USA

[6] Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA, USA

Published online: 28 November 2024

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 reduced calibration problems between genotyping platforms and laboratories, and simulations demonstrate their utility for parentage analysis (Anderson and 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 often focus on a small number of genetic markers that enable more individuals to be genotyped on a limited budget. 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 rate similar to SNPs, and are becoming a popular molecular marker in genomics (Baetscher et al. 2018; Hendricks et al. 2018).

Rainbow Trout (*Oncorhynchus mykiss)* is a species of salmonid fish whose native distribution ranges from northern Mexico to the Kamchatka Peninsula, Russia. The species encompasses a wide range of morphological variation, 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 and Gall 1998). Steelhead usually spend 1 or 2 years in freshwater before migrating to the sea. After 1 to 3 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, although the rate of iteroparity is typically low (Shapovalov and Taft 1954; Behnke 1992; McPhee et al. 2007; Beulke et al. 2023; Goetz et al. 2024). 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; Abadía-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. Here we present the development of a panel of microhaplotype markers for *O. mykiss*, and explore their utility for elucidation of phylogeographic information among populations in California and beyond.

Materials and methods

Microhaplotype panel development

SNP discovery and amplicon design

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. Size selection using a Pippin Prep (Sage Science) targeted fragments in the range of 300–400 bp. The samples were then sequenced in one run on a Miseq instrument (Illumina inc. Shen et al. 2005) using a 600-cycle paired-end sequencing kit. Sequencing yielded 19,324,096 pairs of raw reads. Raw reads obtained from the Illumina sequencing run were filtered based on their average Phred-scaled base quality score (≥ 33) and, as a further quality-control measure, paired-end reads were combined using FLASH v1.2.11 (Fast Length Adjustment of SHort reads; Magoč and Salzberg 2011) with min overlap of 20 and max overlap of 300, retaining 15,499,545 extended reads. Using the *process_radtags* component of Stacks v1.48 (Catchen et al. 2013), reads were truncated to 325 bp and then demultiplexed based on the unique 5 bp individual barcode to assign reads to their corresponding individuals. Raw reads per individual ranged from 74,878 to 651,980 and averaged 449,200. 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*). As reported by Stacks, depth of coverage for processed samples ranged from 13.3× to 69.5× and averaged 47.4× across all individuals.

Among the 29,024 potential loci, a total of 5959 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 loci that had two or more SNPs within 100–105 bases of each other and had a minimum of 18 non-variable bases on either side of those variants to attempt primer design. This left 3049 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 paralogous genomic regions was very high (Pearse et al. 2019). Consensus sequences from the 3049 candidate loci were mapped to themselves using BLAT (Kent 2002). BLAT output was processed using R (R core team 2022), removing 2218 'duplicate' targets that fully or partially matched another target locus, regardless of the number of indels or mismatches. 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 the Omyk_1.0 chromosome-scale genome assembly (Pearse et al. 2019) and produce a list of 385 potential targets for the design of microhaplotype markers. Finally, we used the graphical interface of Stacks to assess overall variability (number of SNPs and observed haplotypes), the potential for successful primer design (variation more than 18 bp from either end of the fragment), and the proportion of population- or species-specific alleles in the 32 sequenced individuals, and used these criteria to select loci for 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 a product size of 130 bp (in the range of 90–143 bp), to optimize sequencing overlap and because short and uniform lengths of target sequences are important factors for uniform PCR amplification among loci. Following initial testing and PCR multiplex optimization, loci that continued to garner the largest proportion of reads (possibly due to a paralogous or repetitive sequence) and the lowest proportion of reads (due to inefficient amplification or variation in the primer sequence) were removed from the panel, as were loci where previously identified SNPs were absent or at much lower frequency than expected, resulting in the final list of 124 presumably neutral loci (Supp. Table 2). In addition to these newly discovered loci from the ddRAD data, markers targeting two previously identified functional gene regions and the Y-chromosome were added to the panel. First, five microhaplotype markers were designed within the chromosomal inversion complex present on chromosome *Omy05*, 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 in the *Greb1L* gene region on *Omy*28 that have been repeatedly and broadly associated with variation run-timing (Waples et al. 2022). Finally, the 'Omy-Y1-2Sexy' locus was included by using the primers from Brunelli et al. (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 et al. (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 found in Baetscher et al. (2018). Sequencing reads were de-multiplexed by the MiSeq Analysis Software (Illumina inc. Shen et al. 2005). Paired-end reads were combined using FLASH (min overlap of 4 and max overlap of 50; Magoč and Salzberg 2011); merge rates were generally over 90% although some individuals with lower quality DNA had rates as low as 60%. Merged reads were then mapped to the Stacks consensus sequences for the target loci using the BWA-MEM v0.7.17-r1188 (*-M -v 3 -t 10 -R*; Li and Durbin 2009). Mapped reads were converted from Sequence Alignment/Map (SAM) files to BAM files with SAMtools v1.13 (*view -bhS*; Li et al. 2009). We identified variable sites using FreeBayes v1.3.6 (*–haplotype-length 0 -kwVa –no-mnps –no-complex*; Garrison and Marth 2012); the positions of all SNPs for each locus were recorded in a VCF file. The *microhaplot* R package (Ng et al., https://doi.org/10.5281/zenodo.820110; R core team 2022) and associated Shiny app (http://shiny.rstudio.com/) were employed to assemble the SNPs for each amplicon into a microhaplotype using the SAM files and the positions specified in the VCF. After filtering loci for a minimum depth of coverage of 10 reads per individual and an allelic balance ratio greater than 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, each containing one to four plates of 96 individuals at the same time. After each run, 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, 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. Based on previous studies, we defined several phylogeographic groups; the Cutthroat trout outgroup, Northern Coastal (Oregon and Klamath system; Pearse et al. 2011), Southern Coastal (Southern California, Abadía-Cardoso et al. 2016), the remaining Coastal populations (Northern California to Central California, Garza et al. 2014), Central Valley Below dams (Pearse and Garza 2015), Central Valley Above dams (Pearse and Campbell 2018), Inland (Redband and Eagle Lake subspecies; Nielsen et al. 1999; Moyle 2002), Golden Trout (Subspecies; Cordes et al. 2006), and Hatchery strains. This study is based on 124 loci genotyped in a total of 1831 tissue samples from 58 populations (Fig. 1; Table 1) (Pritchard et al. 2012; Garza et al. 2014; Abadía-Cardoso et al. 2019).

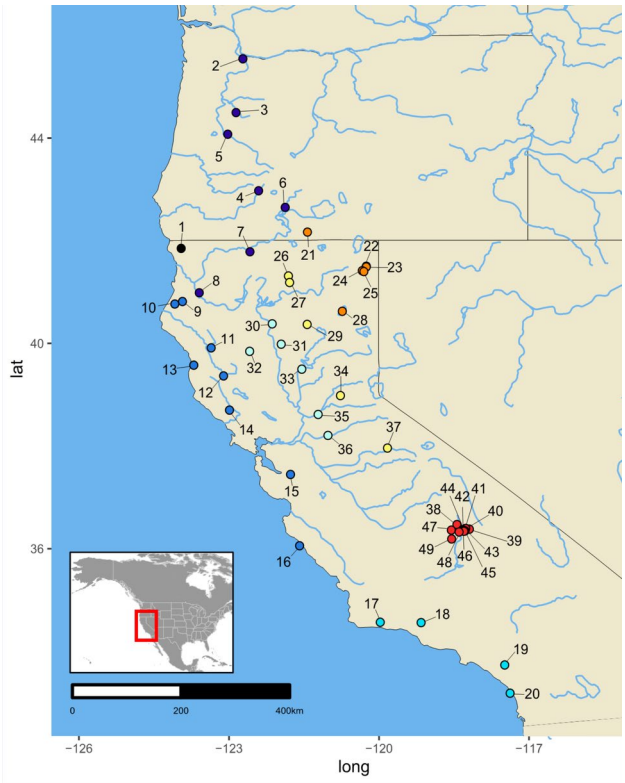

Fig. 1 Map of rivers and watersheds of California and Oregon. Each wild population sample used for this study is represented by a point, estimated with GPS coordinates (in Supp. Table 1). The color code is the same as on Table 1, i.e. Black: Cutthroat trout, Blue: Coastal populations, Dark Blue: Northern Coastal, Turquoise: Southern Coastal, Light Blue: Central Valley Below dams, Yellow: Central Valley Above dams, Orange: Inland, Red: Golden Trout. Hatchery strains are not represented on the map

Table 1 Population samples used for this study

#	Region	Country, States	Sample site	n	Reference
1	Coastal Cutthroat	United States, California	NF Smith R.	16	Pritchard et al. 2012
2	Coastal	United States, Oregon	Willamette R.	12	J. Rodzen, pers. comm.
3	Coastal	United States, Oregon	South Santiam R.	21	J. Rodzen, pers. comm.
4	Coastal	United States, Oregon	Foster Ck.	9	J. Rodzen, pers. comm.
5	Coastal	United States, Oregon	Mckenzie R.	25	J. Rodzen, pers. comm.
6	Coastal Klamath	United States, Oregon	Spring Ck.	16	Pearse et al. 2011
7	Coastal Klamath	United States, California	Shasta R.	36	Pearse et al. 2007
8	Coastal Klamath	United States, California	Horse Linto Ck.	46	Garza et al. 2013
9	Coastal	United States, California	Mad R.	16	Pearse & Garza 2015
10	Coastal	United States, California	Freshwater Ck.	42	Garza et al. 2013
11	Coastal	United States, California	MF Eel R.	46	Clemento 2006
12	Coastal	United States, California	Eel R. (Van Arsdale Fisheries Station)	46	This study
13	Coastal	United States, California	Little NF (Tenmile Ck.)	44	Garza et al. 2013
14	Coastal	United States, California	Dry Ck. (Russian R.)	78	Abadia-Cardoso et al. 2013
15	Coastal	United States, California	Alameda Ck.	40	Leitwein et al., 2017
16	Coastal	United States, California	Big Ck.	72	Barthelemy 2018
17	Coastal – SoCal	United States, California	Hilton Ck. (Santa Ynez R.)	16	Clemento et al 2009
18	Coastal – SoCal	United States, California	Piru Ck. (Santa Clara R.)	15	Clemento et al 2009
19	Coastal – SoCal	United States, California	Coldwater Canyon Ck.	35	Abadia-Cardoso et al. 2016
20	Coastal – SoCal	United States, California	WF San Luis Rey R.	9	J. Rodzen, pers. comm.
21	Inland	United States, California	Buck Ck. (Goose Lake)	24	Pearse et al., 2011
22	Inland	United States, California	NF Deep Ck.	15	J. Rodzen, pers. comm.
23	Inland	United States, California	SF Deep Ck.	15	J. Rodzen, pers. comm.
24	Inland	United States, California	NF Shields Ck. (Pit R.)	15	J. Rodzen, pers. comm.
25	Inland	United States, California	SF Shields Ck. (Pit R.)	13	J. Rodzen, pers. comm.
26	Central Valley-A	United States, California	Sheepheaven Ck. (McCloud R.)	13	J. Rodzen, pers. comm.
27	Central Valley-A	United States, California	Moosehead Ck. (McCloud R.)	14	J. Rodzen, pers. comm.
28	Inland - Eagle Lake	United States, California	Eagle Lake	123	J. Rodzen, pers. comm.
29	Central Valley-A	United States, California	North Arm Rice Ck. (Feather R.)	29	J. Rodzen, pers. comm.
30	Central Valley-B	United States, California	Coleman Hatchery (Battle Ck.)	76	Pearse & Garza 2015
31	Central Valley-B	United States, California	Deer Ck.	18	Pearse & Garza 2015
32	Central Valley-B	United States, California	Thomes Ck.	6	Pearse & Garza 2015
33	Central Valley-B	United States, California	Feather R. Hatchery	62	Pearse & Garza 2015
34	Central Valley-A	United States, California	MF American R.	42	Abadia-Cardoso et al. 2019
35	Central Valley-B	United States, California	Nimbus Hatchery (Amercian R.)	47	This study
36	Central Valley-B	United States, California	Mokelumne R. Hatchery	95	Goetz et al. In Prep.
37	Central Valley-A	United States, California	Frog Ck. (Tuolumne R.)	20	Pearse & Campbell 2018
38	Inland - Golden Trout	United States, California	Chagoopa Ck.	13	J. Rodzen, pers. comm.
39	Inland - Golden Trout	United States, California	Upper Mulkey Ck.	30	J. Rodzen, pers. comm.
40	Inland - Golden Trout	United States, California	Big Whitney (Golden Trout Ck.)	15	J. Rodzen, pers. comm.
41	Inland - Golden Trout	United States, California	Salt Lick Ck. (Golden Trout Ck.)	26	J. Rodzen, pers. comm.
42	Inland - Golden Trout	United States, California	Golden Trout Ck.	29	J. Rodzen, pers. comm.
43	Inland - Golden Trout	United States, California	Tunnel Station (Golden Trout Ck.)	22	J. Rodzen, pers. comm.
44	Inland - Golden Trout	United States, California	Volcanic Ck. (Golden Trout Ck.)	37	J. Rodzen, pers. comm.
45	Inland - Golden Trout	United States, California	Upper Groundhog meadow (Golden Trout Ck.)	25	J. Rodzen, pers. comm.
46	Inland - Golden Trout	United States, California	Groundhog meadow (Golden Trout Ck.)	26	J. Rodzen, pers. comm.
47	Inland - Golden Trout	United States, California	Broders Cabin (Little Kern R.)	28	J. Rodzen, pers. comm.
48	Inland - Golden Trout	United States, California	Kern R.	14	J. Rodzen, pers. comm.
49	Inland - Golden Trout	United States, California	NF Clicks Ck. (Little Kern R.)	30	J. Rodzen, pers. comm.
50	Hatchery strain	United States, California	Coleman Hatchery	35	J. Rodzen, pers. comm.
51	Hatchery strain	United States, California	Hot Ck. Hatchery	36	J. Rodzen, pers. comm.
52	Hatchery strain	United States, California	Pit R. Hatchery	47	J. Rodzen, pers. comm.
53	Hatchery strain	United States, California	Mt. Shasta Hatchery (American R.)	15	Pearse & Garza 2015
54	Hatchery strain	United States, California	Shasta R. Hatchery	46	J. Rodzen, pers. comm.
55	Hatchery strain	United States, California	Wyoming Hatchery	32	J. Rodzen, pers. comm.
56	Hatchery strain	United States, California	Hot Ck. Hatchery (Kamloops Strain)	16	J. Rodzen, pers. comm.
57	Hatchery strain	Canada, British Columbia	Kamloops Hatchery JuncRes	22	J. Rodzen, pers. comm.
58	Hatchery strain	Canada, British Columbia	Kamloops Hatchery June Lake	20	J. Rodzen, pers. comm.

Each color refers to a subgroup of populations used in all figures. Black: Cutthroat trout, Blue: Coastal populations, Dark 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. The ID corresponds to the identification name used in our dataset, and n is the number of individuals per population. See Supp. Table 1 for complete sample details and population genetic results

Genetic diversity estimates [*i.e.*, expected heterozygosity (He), observed heterozygosity (Ho), number of alleles (Na), inbreeding coefficient (Fis), and pairwise fixation index (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 1989). A neighbour-joining tree (Saitou and Nei 1987), representing genetic distance between populations, was calculated with PHYLIP using pairwise chord distances (Cavalli-Sforza and Edwards 1967). The stability of the tree topology was examined using the Seqboot program, with 1000 bootstrap replicates. Discriminant analysis of principal components (DAPC) was performed with the R package 'adegenet' (Jombart et al. 2010). The number of axes that maximize the results of the DAPC was defined by a "DAPC Cross-Validation" test (Jombart and 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, 729 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 1831 genotyped individuals from 58 populations. No consistent deviations from HWE were 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; He = 0.06, Ho = 0.06, Ar = 1.15, Na = 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. 2). 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. 2). Similarly, the Redband Trout of Deep, Shields, and Buck Creeks strongly clustered together, along with Spring Creek, an upper Klamath tributary (Fig. 2). 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 and Garza 2015).

DAPC: discriminant analysis of principal components

DAPC was used to visualize the differentiation and relationships of all the *O. mykiss* populations (Fig. 3). For

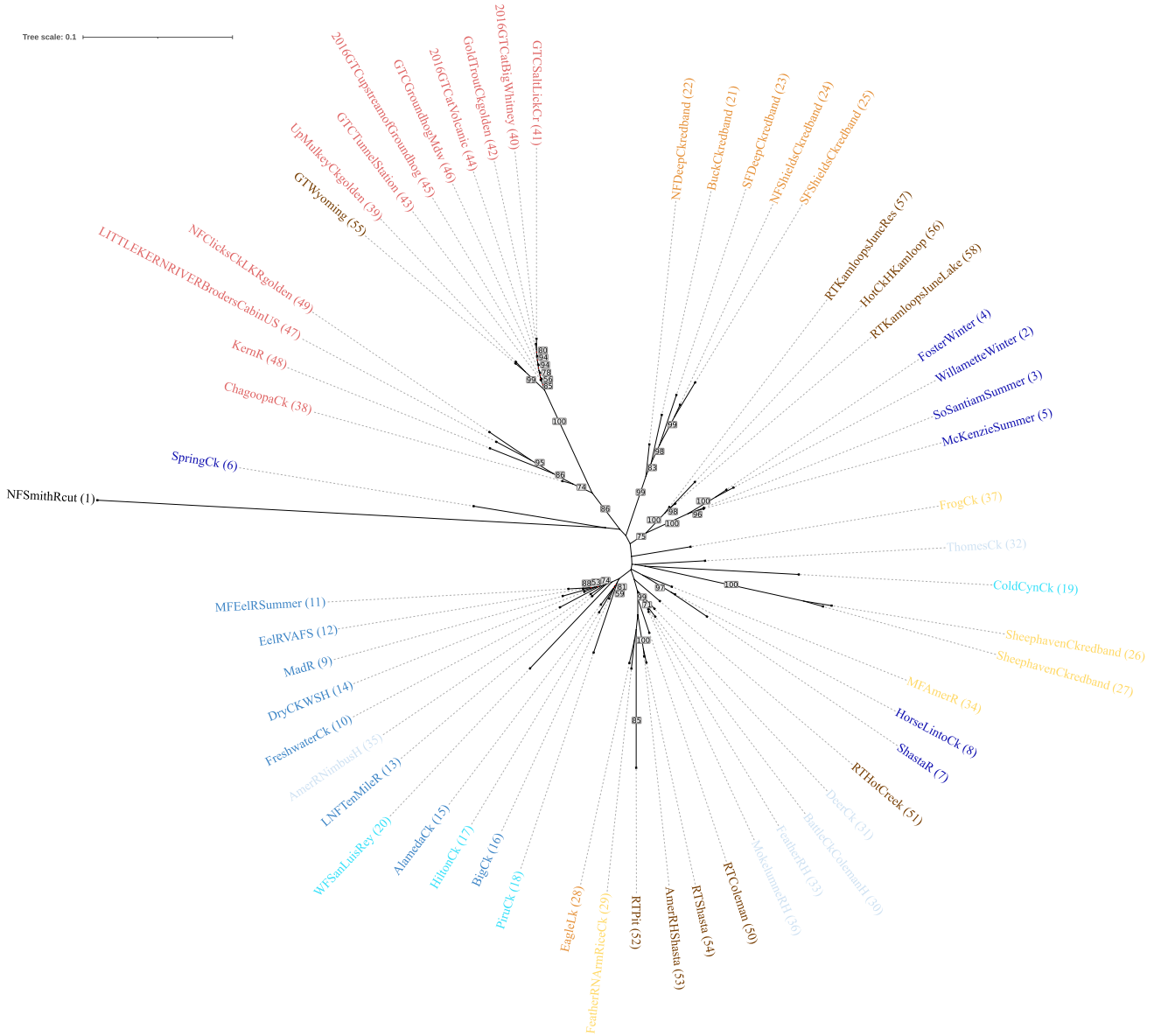

Fig. 2 Neighbor joining tree of *O. mykiss* populations, showing distance between each nod. Constructed with chord distance, this tree displays all the population present in this study. The color code is the same as on Table 1, i.e. Black: Cutthroat trout, Blue: Coastal populations, Dark Blue: Northern 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

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. 3A) 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 some members of the golden trout complex (38–49), the Wyoming strain of hatchery golden trout (55), and all other populations (Fig. 3B). Pit River Hatchery (52) and Coldwater Canyon Creek (19) populations also showed clear differentiation from the other *O. mykiss* populations. These results were also shown with golden trout complex populations removed (Fig. 3C), in which case the redband trout subspecies (21–25) were differentiated from the rest. Northern coastal populations (2–8) also displayed a break from the central group; this split is most visible on the last DAPC (Fig. 3D), in which Central Valley steelhead populations below dams (30–33, 35–36) also showed strong links with coastal populations.

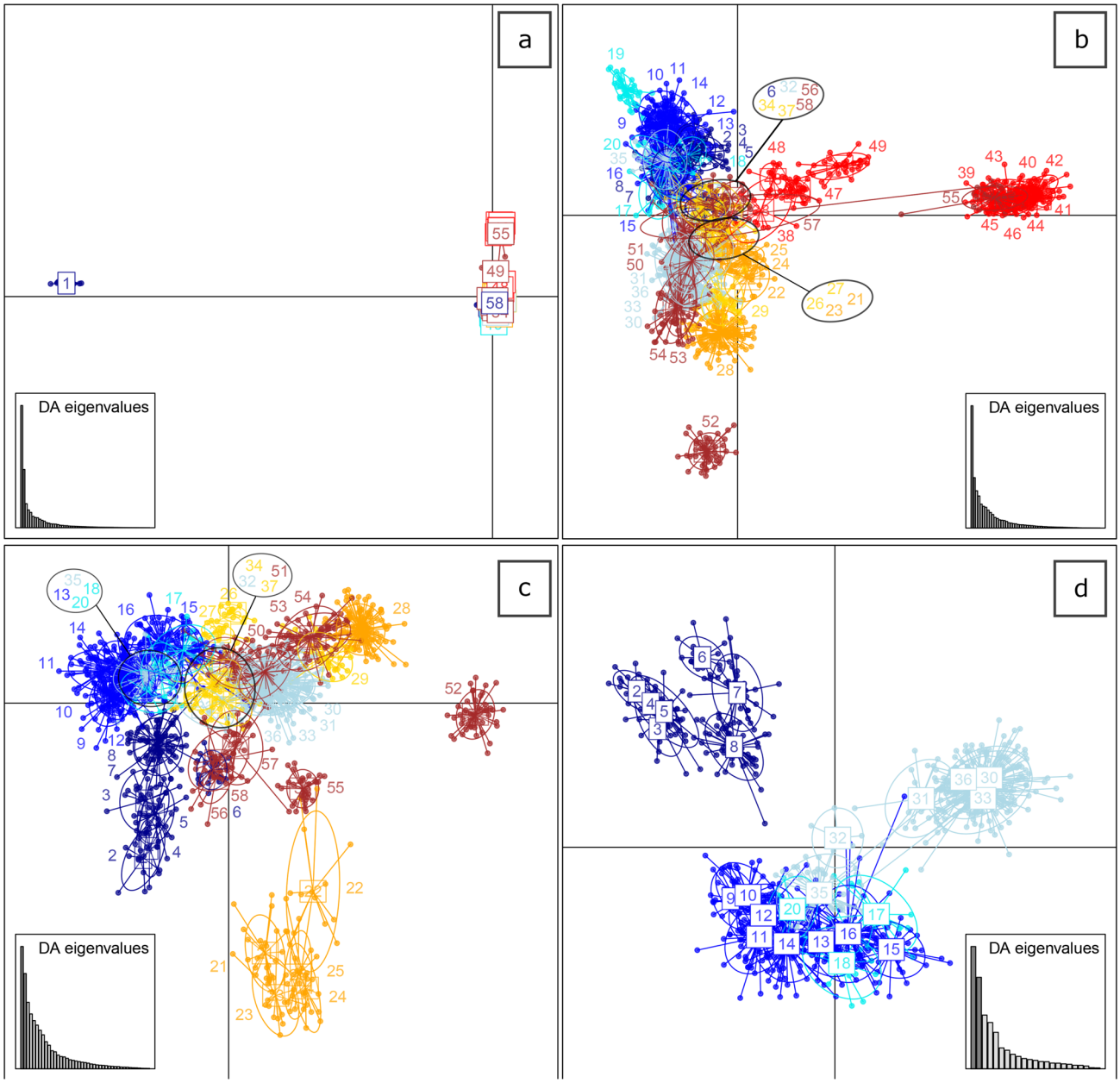

Fig. 3 Four DAPCs showing population relationships on their first two axes that explained most of the variance. **A** all the populations (1–58), **B** *O. mykiss* populations without Cutthroat trout (2–58), **C** *O. mykiss* populations except the golden trout and Coldwater Canyon Creek populations (2–18, 20–37, 50–58), **D** steelhead populations only (1–18, 20, 30–33, 35–36). The color code correspond to: Black: Cutthroat trout, Blue: Coastal populations, Dark 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

STRUCTURE

STRUCTURE results (Supp. Fig. 1) were clear and consistent among the 10 replicates 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. Fig. 1).

Adaptive genetic variation and sex informative loci

In addition to the microhaplotype loci designed for population genetic analyses, we developed 10 microhaplotype loci specifically to target regions of adaptive genetic variation (*Omy*5 and *Greb1L*). 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 this demonstrates that these loci amplify in this species, it does not necessarily indicate that the same associations between these variants and specific life-history traits exist.

The five microhaplotype markers designed within the inversion complex present on chromosome 5 (*Omy05*) 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. Based on previous studies and examination of the genotype data, a single key SNP for each of these regions was used to 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). For example, as has been previously shown (Pearse et al. 2019), the Omy05 allele associated with residency is fixed in inland redband trout populations, but the alternate allele appears at high frequencies in populations known to support anadromous migration (Supp. Table 1). Similarly, the population sample from the Middle Fork of the Eel River is fixed for the 'early' or 'premature' allele at Greb1L loci, as has been previously shown in this summer run steelhead population, while other coastal steelhead populations are fixed for the alternate allele (Waples et al. 2022).

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, no other known males failed to amplify, suggesting that that individual was likely a female and indicating a possible metadata error. All other known males had a minimum of eight reads, and most had > 50 reads, clearly differentiating males and females.

Discussion

Effectiveness of the microhaplotype panel

The novel microhaplotype panel described here provides sufficient variation to resolve population genetic groups previously identified by microsatellite and SNP-based studies in this species in California (*e.g.*, Pearse et al. 2009; Garza et al. 2014; Pearse and Garza 2015; Pearse and Campbell 2018). 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 and offering a multi-allelic alternative to existing bi-allelic SNP genotyping panels (e.g. Campbell et al. 2015). Thus, this panel of markers provides a valuable new tool for researchers, combining the main advantages of microsatellites and SNPs into a single high-throughput genotyping approach (for an empirical comparison of SNPs and microsatellites see Glaubitz et al. 2003; Hauser et al. 2011). These markers should 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 were 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, further investigations of the linkage disequilibrium and frequency of discordant genotypes among the existing microhaplotypes, as well as the potential to add microhaplotypes targeting additional adaptive variants, will extend 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 expected 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 pairwise Fst, 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 pairwise Fst, 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), as were overall patterns of within-population variation as reflected in the low estimates of heterozygosity and allelic richness found in isolated inland populations and populations above dams, and relatively high values of these parameters found in coastal steelhead populations below dams.

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 5000–10,000 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. 2004). 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, high pairwise Fst, 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 (Abadía-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, as well as having potential utility for 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.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12686-024-01374-2.

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.

Author 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.

Declarations

Conflict of interest The authors declare no competing interests.

References

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*). Mol Ecol. https://doi.org/10.1111/mec.12426

Abadía-Cardoso A, Brodsky A, Cavallo B, Arciniega M, Garza JC, Hannon J, Pearse DE (2019) Anadromy Redux? Genetic analysis to inform development of an indigenous American steelhead broodstock. J Fish Wildl Manag 10(1):137–147. https://doi.org/10.3996/072018-JFWM-063

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. Conserv Genet. https://doi.org/10.1007/s10592-016-0814-9

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. Nat Rev Genetics 11:697–709

Anderson EC, Garza JC (2006) The power of single-nucleotide polymorphisms for large-scale parentage inference. Genetics. https://doi.org/10.1534/genetics.105.048074

Baetscher DS, Clemento AJ, Ng TC, Anderson EC, Garza JC (2018) Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. Mol Ecol Resour. https://doi.org/10.1111/1755-0998.12737

Bagley MJ, Gall GAE (1998) Mitochondrial and nuclear DNA sequence variability among populations of rainbow trout (*Oncorhynchus mykiss*). Mol Ecol. https://doi.org/10.1046/j.1365-294x.1998.00413.x

Behnke RJ (1992) Native trout of western North America. Amer Fisheries Society, Bethesda

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. Nat Commun. https://doi.org/10.1038/ncomms4657

Beulke AK, Abadía-Cardoso A, Pearse DE, Goetz LC, Thompson NF, Anderson EC, Garza JC (2023) Distinct patterns of inheritance shape life-history traits in steelhead trout. Mol Ecol 32:6896–6912. https://doi.org/10.1111/mec.17182

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. Trends Ecol Evol 18:249–256

Brunelli JP, Steele CA, Thorgaard GH (2010) Deep divergence and apparent sex-biased dispersal revealed by a Y-linked marker in rainbow trout. Mol Phylogenet Evol 56(3):983–990. https://doi.org/10.1016/j.ympev.2010.05.016

Buchanan DV, 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

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

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 Geneslgenomeslgenetics. https://doi.org/10.1093/g3journal/jkac137

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. Mol Ecol Resour. https://doi.org/10.1111/1755-0998.12357

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. Mol Ecol. https://doi.org/10.1111/mec.12354

Cavalli-Sforza LL, Edwards AW (1967) Phylogenetic analysis. Models and estimation procedures. Am J Hum Genet. https://doi.org/10.2307/2406616

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. Conserv Genet. https://doi.org/10.1007/s10592-008-9712-0

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. https://doi.org/10.1577/t05-120.1

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499–510

Felsenstein J (1989) PHYLIP--phylogenetic inference package. Cla 5:164–166

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing---Free bayes---Variant Calling---Longranger. arXiv preprint arXiv:1207.3907

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. https://doi.org/10.1080/00028487.2013.822420

Glaubitz JC, Rhodes OE, Dewoody JA (2003) Prospects for inferring pairwise relationships with single nucleotide polymorphisms. Mol Ecol. https://doi.org/10.1046/j.1365-294X.2003.01790.x

Goetz LC, Nuetzel H, Vendrami DLJ, Beulke AK, Anderson EC, Garza JC, Pearse DE (2024) Genetic parentage reveals the (un) natural history of Central Valley hatchery steelhead. Evol Appl 17:e13681. https://doi.org/10.1111/eva.13681

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. Mol Ecol Resour. https://doi.org/10.1111/j.1755-0998.2010.02961.x

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. Evolut Appl. https://doi.org/10.1111/eva.12659

Jombart T, Collins C (2015) A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0.0. 2015, pp 1–43. Available online: https://adegenet.r-forge.r-project.org/files/tutorial-dapc.pdf (Accessed 25 Jan 2022)

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. https://doi.org/10.1186/1471-2156-11-94

Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. Mol Ecol Notes. https://doi.org/10.1111/j.1471-8286.2004.00845.x

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. https://doi.org/10.1093/bioinformatics/bts199

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 Ecol Evol. https://doi.org/10.1111/2041-210X.12067
Kelson SJ, Carlson SM, Miller MR (2020) Indirect genetic control of migration in a salmonid fish. Biol Lett. https://doi.org/10.1098/rsbl.2020.0299
Kent WJ (2002) BLAT-the BLAST-like alignment tool resource 656 genome research. Genome Res 12(4):656–664
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 Sci Int: Genet. https://doi.org/10.1016/j.fsigen.2014.06.014
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. Mol Ecol Resour. https://doi.org/10.1111/1755-0998.12387
Larsson A, Stanley G, Sinha R, Weissman I, Sandberg R (2017) Computational correction of cross-contamination due to exclusion amplification barcode spreading. bioRxiv 48:1430
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. Evolut Appl. https://doi.org/10.1111/eva.12416
Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. https://doi.org/10.1093/bioinformatics/btp324
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics. https://doi.org/10.1093/bioinformatics/btp352
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. https://doi.org/10.1038/nature17164
Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. https://doi.org/10.1093/bioinformatics/btr507
McKinney GJ, Seeb JE, Seeb LW (2017) Managing mixed-stock fisheries: Genotyping multi-SNP haplotypes increases power for genetic stock identification. Can J Fish Aquat Sci. https://doi.org/10.1139/cjfas-2016-0443
McPhee MV, Utter F, Stanford JA, Kuzishchin KV, 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. Ecol Freshw Fish. https://doi.org/10.1111/j.1600-0633.2007.00248.x
Moyle PB (2002) Inland fishes of California: revised and expanded. Univ of California Press, Berkeley
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. Mol Ecol. https://doi.org/10.1111/j.1365-294X.2008.03851.x
Nielsen JL, Crow KD, Fountain MC (1999) Microsatellite diversity and conservation of a relic trout population: McCloud River redband trout. Mol Ecol. https://doi.org/10.1046/j.1365-294X.1999.00817.x
NOAA (2006) Department of Commerce National Endangered and threatened species: final listing determinations for 10 distinct population segments of west coast steelhead. Fed Regist 71:843–862
Oldoni F, Kidd KK, Podini D (2019) Microhaplotypes in forensic genetics. Forensic Sci Int: Genet 38:54–69
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. Sci Rep. https://doi.org/10.1038/s41598-020-58980-x
Park S (2008) Excel microsatellite toolkit. version 3.1. 1. Animal Genomics Lab website (University College, Dublin, Ireland)
Pearse DE, Campbell MA (2018) Ancestry and adaptation of rainbow trout in Yosemite National Park. Fisheries (Bethesda). https://doi.org/10.1002/fsh.10136
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 Franc Estuary Watershed Sci. https://doi.org/10.15447/sfews.2015v13iss4art1
Pearse DE, Donohoe CJ, Garza JC (2007) Population genetics of steelhead (*Oncorhynchus mykiss*) in the Klamath River. Environ Biol Fishes. https://doi.org/10.1007/s10641-006-9135-z
Pearse DE, Hayes SA, Bond MH, Hanson CV, Anderson EC, MacFarlane RB, Garza JC (2009) Over the falls? Rapid evolution of ecotypic differentiation in steelhead/rainbow trout (*Oncorhynchus mykiss*). J Hered. https://doi.org/10.1093/jhered/esp040
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. https://doi.org/10.1080/00028487.2011.583538
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. Proc R Soc B. https://doi.org/10.1098/rspb.2014.0012
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. Nat Ecol Evol. https://doi.org/10.1038/s41559-019-1044-6
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. https://doi.org/10.1371/journal.pone.0037135
Phillips RB (2013) Evolution of the sex chromosomes in salmonid fishes. Cytogenet Genome Res. https://doi.org/10.1159/000355149
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics. https://doi.org/10.1093/genetics/155.2.945
Pritchard VL, Abadía-Cardoso A, Garza JC (2012) Discovery and characterization of a large number of diagnostic markers to discriminate Oncorhynchus mykiss and O. clarkii. Mol Ecol Resour 12:918–931. https://doi.org/10.1111/j.1755-0998.2012.03149.x
R Core Team (2022) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://wwwR-project.org. (Accessed 22 Jan 2022)
Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Rundio DE, Williams TH, Pearse DE, Lindley ST (2012) Male-biased sex ratio of nonanadromous *Oncorhynchus mykiss* in a partially migratory population in California. Ecol Freshw Fish. https://doi.org/10.1111/j.1600-0633.2011.00547.x

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. https://doi.org/10.1093/oxfordjournals.molbev.a040454

SantaLucia J, Jr (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc Natl Acad Sci USA 95:1460–1465

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. https://doi.org/10.1007/978-1-60327-411-1_18

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)

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. Mutat Res—Fundam Mol Mech Mutagenes 573:70–82

Simmons RE, Lavretsky P, May B (2010) Introgressive hybridization of redband trout in the upper McCloud river watershed. Trans Am Fish Soc. https://doi.org/10.1577/t08-245.1

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

Swift CC, Haglund TR, Ruiz M, Fisher RN (1993) The status and distribution of the freshwater fishes of Southern California. Bull South Calif Acad Sci 92(3):101–167

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3-new capabilities and interfaces. Nucleic Acids Res. https://doi.org/10.1093/nar/gks596

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 Hered 113(2):121–144

Waters CD, Clemento A, Aykanat T, Garza JC, Naish KA, Narum S, Primmer CR (2021) Heterogeneous genetic basis of age at maturity in salmonid fishes. Mol Ecol 30:1435–1456. https://doi.org/10.1111/mec.15822

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. Evolut Appl. https://doi.org/10.1111/eva.12032