UC Santa Cruz

Title

Characterization and distribution of a 14-Mb chromosomal inversion in native populations of rainbow trout (Oncorhynchus mykiss).

Permalink

https://escholarship.org/uc/item/2tr889sd

Journal

Authors

Hale, Matthew

Pearse, Devon

Campbell, Matthew

Publication Date

2024-07-08

DOI

Peer reviewed

Characterization and distribution of a 14-Mb chromosomal inversion in native populations of rainbow trout (*Oncorhynchus mykiss*)

Matthew C. Hale (ID) ,[1,]* Devon E. Pearse (ID) ,[2] Matthew A. Campbell (ID) [3,4]

[1]Department of Biology, Texas Christian University, Fort Worth, TX 76109, USA
[2]Southwest Fisheries Science Center, National Oceanic and Atmospheric Administration, Santa Cruz, CA 95064, USA
[3]Department of Animal Science, University of California Davis, Davis, CA 95616, USA
[4]Present address: Centre for Carbon, Water and Food, University of Sydney, Brownlow Hill 2570, Australia

*Corresponding author: Department of Biology, Texas Christian University, Fort Worth, TX 76109, USA. Email: m.c.hale@tcu.edu

Multiple studies in a range of taxa have found links between structural variants and the development of ecologically important traits. Such variants are becoming easier to find due, in large part, to the increase in the amount of genome-wide sequence data in nonmodel organisms. The salmonids (salmon, trout, and charr) are a taxonomic group with abundant genome-wide datasets due to their importance in aquaculture, fisheries, and variation in multiple ecologically important life-history traits. Previous research on rainbow trout (*Oncorhynchus mykiss*) has documented a large pericentric (~55 Mb) chromosomal inversion (CI) on chromosome 5 (Omy05) and a second smaller (~14 Mb) chromosome inversion on Omy20. While the Omy05 inversion appears to be associated with multiple adaptive traits, the inversion on Omy20 has received far less attention. In this study, we re-analyze RAD-seq and amplicon data from several populations of rainbow trout (*O. mykiss*) to better document the structure and geographic distribution of variation in the Omy20 CI. Moreover, we utilize phylogenomic techniques to characterize both the age- and the protein-coding gene content of the Omy20 CI. We find that the age of the Omy20 inversion dates to the early stages of *O. mykiss* speciation and predates the Omy05 inversion by ~450,000 years. The 2 CIs differ further in terms of the frequency of the homokaryotypes. While both forms of the Omy05 CI are found across the eastern Pacific, the ancestral version of the Omy20 CI is restricted to the southern portion of the species range in California. Furthermore, the Omy20 inverted haplotype is comparable in genetic diversity to the ancestral form, whereas derived CIs typically show substantially reduced genetic diversity. These data contribute to our understanding of the age and distribution of a large CI in rainbow trout and provide a framework for researchers looking to document CIs in other nonmodel species.

Keywords: chromosomal inversions; homokaryotypes; salmonids; anadromy; structural variation

Introduction

With the advent of next-generation sequencing and high-power computational analyses, it has become easier to generate genetic maps and identify segregating structural variants (SVs), improving their detection (Flagel *et al.* 2019). A lot of interest has recently been generated in identifying and documenting chromosomal inversions (CIs) in natural populations and linking different homokaryotypes (often referred to as "supergenes") to ecologically important traits. This interest is due to recombination suppression between alternative homokaryotypes through nonviability of recombinant gametes, which results in the protection of locally adapted alleles (Kirkpatrick 2010; Schwander *et al.* 2014; reviewed in Wellenreuther and Bernatchez 2018; Huang and Rieseberg 2020). Although many forms of SVs can suppress recombination, CIs are often large—in the order of several megabases—and therefore are more easily found, described, and documented than smaller SVs such as deletions, insertions, and copy number variants (but see Huang *et al.* 2020; Gao *et al.* 2021 for examples of smaller CIs).

The salmonids (salmon, trout, and charrs) are an exemplary group to study genomics in wild organisms, as many species are economically of interest due to human consumption, and are major recreational fishing, aquaculture and commercial fisheries species. Additionally, numerous salmonids are important to native peoples for cultural and subsistence reasons. These factors have led to near-complete genome sequences, as well as significant resequencing efforts, for several species of salmonid that are crucial resources in understanding variation in genome architecture and the genomic basis to the development of ecologically important traits (Lien *et al.* 2016; Pearse *et al.* 2019; Bertolotti *et al.* 2020). For example, previous investigations have found CIs in chum salmon (*Oncorhynchus keta*) from western Alaska (McKinney *et al.* 2020), in Arctic charr (*Salvelinus alpinus*) from Nunavut, Canada (Hale *et al.* 2021), in Alaskan sockeye salmon (*Oncorhynchus nerka*; Euclide *et al.* 2023), and in Atlantic salmon (*Salmo salar*, e.g. Lehnert *et al* 2019; Stenløkk *et al.* 2022). In rainbow trout (*Oncorhynchus mykiss*), much research has focused on a 55-Mb pericentric inversion on chromosome 5 (Omy05). This inversion contains ~1,100 protein-coding genes and has been found

to be associated with life-history development (i.e. residency vs anadromy) as well as other adaptive traits such as embryonic development rate, sexual maturation, phototransduction, and behavior (Sundin *et al.* 2005; Nichols *et al.* 2007; Hale *et al.* 2014; Pearse *et al.* 2019). However, an additional 14-Mb paracentric inversion on the p-arm of Omy20 with ~300 protein-coding genes was described by Pearse *et al.* (2019) and Gao *et al.* (2021) in 2 homozygous inbred lines of *O. mykiss*, but has received far less attention. Indeed, only one other publication—to the best of our knowledge—has noted patterns of variation in the Omy20 inversion in wild populations (Campbell *et al.* 2021), and that study was limited to trout from 4 coastal populations of rainbow trout (*O. m. irideus*) located in central to southern California.

The study of Campbell *et al.* (2021) sampled not only a restrictive set of populations, but also a small proportion of rainbow trout phylogenetic diversity. Coastal rainbow trout (*O. m. irideus*) are widely distributed across the North Pacific and represent a more recent postglacial expansion (McCusker *et al.* 2000). Long-term occupancy of California is shown through the diverse lineages present there, such as the McCloud River Trout (*O. m. calisulat*) and the Golden Trout Complex, which may be recognized as a separate species (*Oncorhynchus aguabonita*; Campbell *et al.* 2023). Archaic "redband" trout are present not only as a geographic isolate in the form of McCloud River Trout, but also as more recent dispersers from the Great Basin/Upper Columbia River into California (*O. m. newberrii* and *O. m. stonei*), further increasing the phylogenetic diversity of *O. mykiss* in California (Campbell *et al.* 2023). Given this phenotypic and genetic diversity, the scarcity of data documenting the different forms of the Omy20 inversion makes it difficult to determine the geographic and phylogenetic distribution of the 2 homokaryotypes, the frequency of heterozygotes, or the association between the different forms of the inversion and selected traits. Herein, we analyze preexisting datasets to describe and document the Omy20 CI and to determine the distribution of the 2 karyotypes over a broad geographic range from Alaska to California. In addition, we investigate the roles and functions of the protein-coding genes within the inversion and use sequence information from these genes to estimate the age of the inversion. These results provide information on this less studied CI in rainbow trout, as well as providing an evolutionary framework for estimating the age and recombination patterns within CIs.

Materials and methods

Two data sources were used to determine the Omy20 karyotype of rainbow trout from the west coast of the contiguous United States and Alaska. First, we re-analyzed previously published reduced representation (RAD-seq) sequencing data downloaded from the GenBank SRA for samples of *O. mykiss* (see Supplementary Table 1 for details of the sampling location, sample size, and the GenBank accession numbers of the raw data). Second, we used data from a newly published amplicon sequencing panel focused on Californian rainbow trout (Le Gall *et al.* 2024; details provided in Supplementary Table 2). These 2 datasets were analyzed separately (see below), and the results were combined to evaluate the inferred frequencies of the Omy20 CI across the sampled populations.

Analysis of RAD-seq data
RAD-seq data from 701 samples were aligned to Version 2 of the rainbow trout genome (Pearse *et al.* 2019) using bwa-mem v0.7.17 with default parameters (Li and Durbin 2009). SAMtools v1.18 was then used to convert sequence alignments into sorted

bam files, and genotypes were determined using ANGSD v0.93 (Korneliussen *et al.* 2014) with the following parameters (snp_pval 1e-6, postCutoff 0.95, minQ 20, minMapQ 20, minMaf 0.05, and minInd 500). The SAMtools model was used for determining genotypes (Li 2011), and the results of ANGSD were outputted in plink format. PLINK1.9 (Purcell *et al.* 2007; Chang *et al.* 2015) was then used to calculate Linkage Disequilibrium (LD) for Omy20 to determine the presence and location of the CI. LD was calculated between all pairs of markers on Omy20 using the following parameters: -ld-window-r2 0, -r2 inter-chr. Estimates of LD were then filtered to include only loci that produced r^2 values >0.8. All samples were then recalled from the existing sequence data using 75 loci that mapped to the CI region of Omy20 and were indicative of the CI karyotypes (i.e. homozygous for the ancestral form, heterozygous, or homozygous for the inverted form). The ancestral karyotype was identified as the inversion orientation that was shared with other *Oncorhynchus* and the inverted homokaryotype as the form of the CI specific to rainbow trout (see Campbell *et al.* 2021, for more details). Heterozygosity was calculated for all loci within the inversion for the 3 karyotypes using PLINK1.9 (Chang *et al.* 2015) with default parameters. The same methods as above were also used to genotype the same samples for the Omy05 inversion in this dataset to compare neutrality statistics and heterozygosity.

Population genetic analyses of the RAD-seq dataset
Site Frequency Spectrum (SFS) files were generated in ANGSD v0.93 (Korneliussen *et al.* 2014) using datasets for each Omy20 and Omy05 homokaryotype in fish from populations that were variable for the inversion (i.e. those south of the Siletz River). Briefly, 4 bam lists were created for homozygous individuals (1 for the ancestral karyotype and 1 for the inverted karyotype for each CI), and SFS files were created in ANGSD using the following specified parameters (snp_pval 1e-6, minQ 20, minMapQ 20, and minMaf 0.05). The SFS files were then processed to determine both Tajima's *D* and Watterson's theta using a sliding-window analysis with a window size of 50,000 bp and a slide size of 10,000 bp. All windows that were not in the inversion regions and those within the inversions with fewer than 100 segregating sites were removed to increase confidence in the estimates of neutrality statistics. The same statistics were also calculated for rainbow trout from the same populations that were homokaryotypes for the Omy05 inversion and compared with the Omy20 inversion. Significant differences between homokaryotypes for both CIs were determined using a 2-tailed *t*-test assuming unequal variance. All statistics were run in R using an alpha value of 0.05. Heterozygosity was also calculated for each of the homokaryotypes and compared both between genotypes and between CIs. Inversion-wide heterozygosity was calculated in PLINK1.9 by calculating the per sample number of heterozygous genotypes for all polymorphic positions within the CI region. We also estimated Weir and Cockerham's F_{ST} for each SNP located along Omy20 between homokaryotypes. These calculations were performed from the SFS files generated in ANGSD, as described above.

Amplicon sequencing–based Omy20 inversion data
To expand the geographic sampling beyond that in publicly available RAD-seq data, we examined high-throughput amplicon sequencing data available for rainbow trout, focusing on representative California populations (Le Gall *et al.* 2024). The dataset of Le Gall *et al.* (2023) includes a range of ocean accessible coastal

rainbow trout and isolated populations from ~33°N to ~45°N latitude. We first determined whether any of the 135 amplicons were located within the Omy20 inversion region by using BLASTN optimized for short sequences (blastn-task blastn-short) against Version 2 of the rainbow trout genome. We filtered alignments based on >95% similarity to the rainbow trout genome and occurring on Omy20 within the CI region. Amplicons within this region were then compared with the RAD-seq data (detailed above) to identify alleles associated with each of the 2 homokaryotypes of the CI. These combined genotype data were then imported into R and filtered to geographic sampling locations; frequencies of the inversion types were plotted geographically with latitude on the x-axis and the frequency of the ancestral Omy20 CI were plotted with latitude on the y-axis.

Dating of Omy20 and Omy05 inversions

To estimate the age of both Omy20 and Omy05 inversions, we identified 2 individuals from the Pearse et al. (2019) dataset from coastal California that were homozygous for alternative forms of the inversions. That is, individual M075219 was homozygous ancestral for both Omy20 and Omy05 inversions, and individual M075289 was homozygous for the derived forms of both Omy20 and Omy05. Sequences of all protein-coding loci from the genomes of both M075219 and M075289 were generated from the genome-wide variants identified by Pearse et al. (2019) with vcf2fasta (https://github.com/santiagosnchez/vcf2fasta). Coding regions (CDs) from annotated protein-coding genes were combined to generate 1 FASTA-formatted file for both samples.

Additional sequences for phylogenetic analyses were obtained by downloading protein-coding genes from northern pike (*Esox lucius*), eastern mudminnow (*Umbra pygmaea*), Atlantic salmon, and European grayling (*Thymallus thymallus*) from SalmoBase (Samy et al. 2017). The lake whitefish (*Coregonus clupeaformis*) protein-coding genes were obtained from the NCBI (GCF_020615455.1). For all species, orthologs were determined by using OrthoFinder (Emms and Kelly 2019; program options: -d -t 5 -a 5) and filtered to groups of putative orthologs that are present as single copies in northern pike and eastern mudminnow but as duplicates in all salmonids to increase confidence in ortholog groups. Subsequently, ortholog groups with a gene in the Omy20 or Omy05 CIs were identified for additional analyses.

For each high-confidence ortholog group with a gene in the Omy20 or Omy05 CI, a multiple sequence alignment (MSA) was generated with MAFFT (Katoh et al. 2002; Katoh and Toh 2008) and then trimmed with BMGE (Criscuolo and Gribaldo 2010). A maximum-likelihood) gene tree for each of the resulting MSAs was generated with IQTREE with support for nodes assessed with ultrafast bootstrap replicates (Nguyen et al. 2015; Hoang et al. 2018; -m GTR + G -bb 1000). The topology of each gene tree was evaluated to verify that rooting with northern pike and mudminnow sequences produced a topology with 2 clades of salmonid sequences originating from the salmonid-specific whole-genome duplication, that Atlantic salmon and rainbow trout sequences were most closely related in each salmonid sequence clade, and the inversion-region sequences from rainbow trout exhibited a sister relationship and were not identical.

Ortholog groups meeting these additional topological criteria were analyzed separately with MrBayes 3.2 (Ronquist et al. 2012). For each ortholog, a time-calibrated phylogeny was generated by providing a starting tree with esociform sequences as outgroups and the 2 clades of salmonid orthologs constrained with lake whitefish and European grayling as sister taxa and salmonine taxa monophyletic. A relaxed clock was used with 2-state nucleotide substitution model with gamma-distributed rate variation. Three exponentially distributed calibrations were used that are based off fossils, dating the Time to The Most Recent Common Ancestor (TMRCA) of northern pike and eastern mudminnow, offsetexp (85, 110.5; Wilson et al. 1992; Campbell et al. 2013) and TMRCA of Salmonidae in the 2 homeolog clades, offsetexp (51.8, 67.34; Wilson and Williams 1992; Near et al. 2012). The TMRCA of *Salmo* and *Oncorhynchus* in the 2 homeolog clades, offsetexp (27.3, 35.49), is based on dates from molecular studies (Campbell et al. 2013; MacQueen and Johnston 2014). The root age of the tree was constrained with a uniform prior from 96.1 to 121.5 MYA based on a divergence time of 112 MYA Esociformes and Salmoniformes (Kumar et al. 2022).

Initially, MrBayes analyses of 3 runs with 3 chains each were conducted for increasing run lengths, with convergence assessed after a 25% burn in for a subset of 10 orthologs (sump command in MrBayes). Convergence and sufficient effective sample sizes ($\gg$200) of parameters were reached consistently across orthologs with 50 million generations sampled every 5,000 generations. Phylogenies for all orthologs generated with these parameters and the posterior median of the estimated age of divergence of ancestral and inverted forms of each CI were extracted from the consensus tree files created by MrBayes (sumt command). Following Pearse et al. (2019), rainbow trout genes that were invariant were removed for dating CIs. Consensus trees of the Omy20 and Omy05 ortholog time trees were created with the consensus.edges function of the phytools package in R (Revell 2012).

Functional characterization of the Omy20 inversion

We used a BLAST-based approach to determine the functions of protein-coding genes found within the Omy20 CI. Briefly, protein-coding sequences for all 311 genes (as described in Pearse et al. 2019) were downloaded from Version 2 of the rainbow trout genome and annotated against the zebrafish reference protein database using BLASTX with default parameters (apart from: maximum e-value = 1.0e−10, maximum number of blast hits saved per sequence = 15). BLAST hits were then uploaded into Blast2GO v6 (Conesa et al. 2005) to obtain Gene Ontology (GO) terms associated with the zebrafish protein sequences. A Fisher's exact test was used to test for enrichment of protein-coding genes within the Omy20 CI by comparing GO terms associated with the 311 genes within the Omy20 CI and the rest of the protein-coding genes within the rainbow trout genome. Significantly enriched GO terms were identified using a Benjamini–Hochberg False Discovery Rate (FDR)-corrected P-value (alpha = 0.05).

Results

Geographic distribution of the Omy20 CI karyotypes

LD analysis of 421 RAD-seq SNPs that mapped to Omy20 identified 75 SNPs that are indicative of the Omy20 CI karyotype (Pearse et al. 2019; Campbell et al. 2021). The Omy20 CI spans 13.7 Mb and is located between 5.6 and 19.2 Mb in the rainbow trout genome (Fig. 1). Our RAD-seq data analyses of 41 populations (Supplementary Table 1) revealed a large difference in the frequency of the homokaryotypes with 591 out of 701 samples (84.3%) being homozygous for the inverted form of the inversion, whereas 66 (9.4%) individuals were heterozygous, and 44 (6.3%) were homozygous for the ancestral karyotype. Homozygotes for
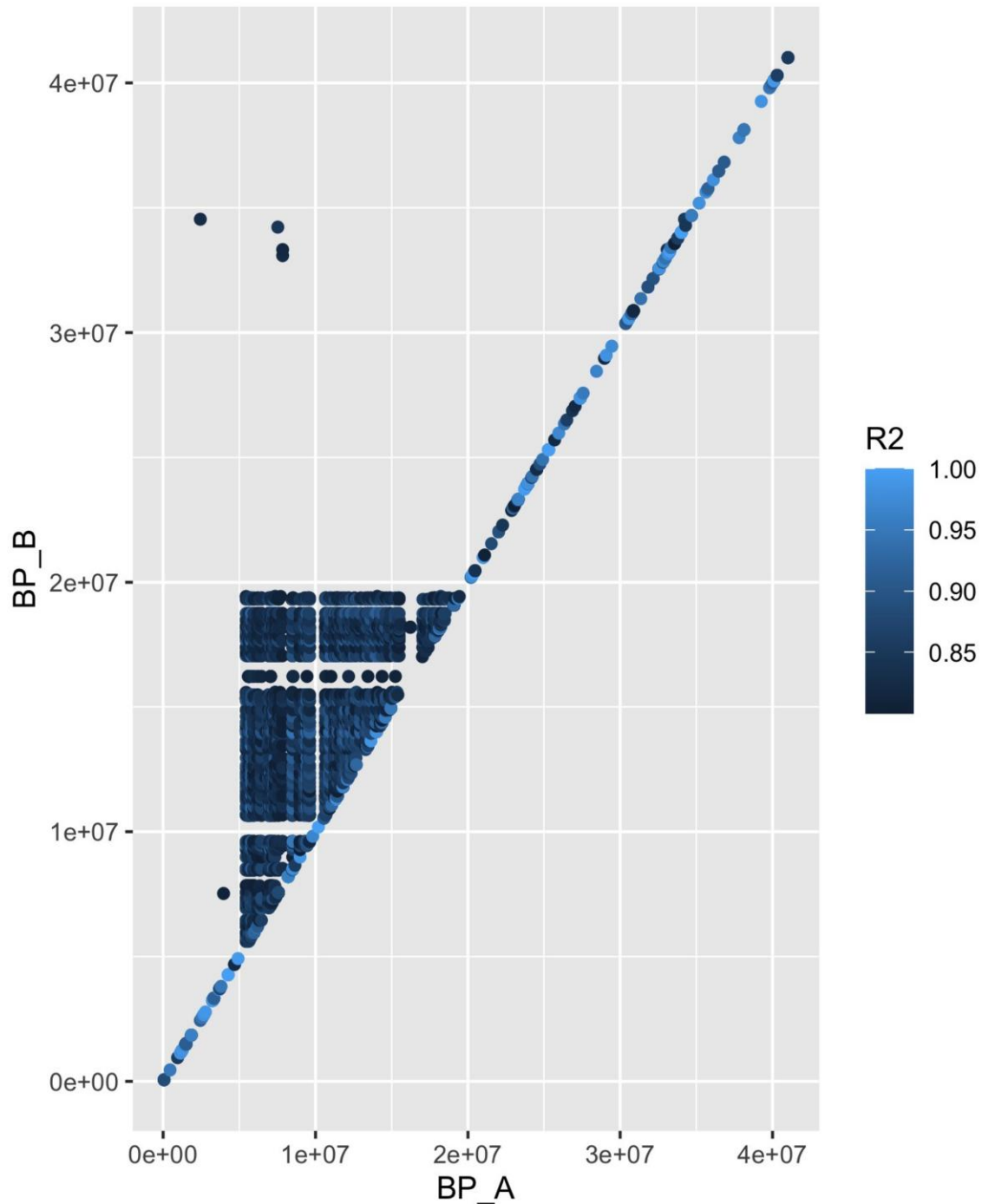
Fig. 1. Patterns of LD measured between pairs of loci on Omy20. Only estimates of LD ≥r^2 values of 0.8 are shown for simplicity. The location of the CI is determined to be between 5.6 and 19.2 Mb.

the ancestral form were restricted to California, with the karyotype being fixed in Scott Creek, and at moderate-to-high frequency in Matilija Creek, Big Creek, the San Gabriel River, and the Eel River (63.3, 86.8, 30.0, and 58.0%, respectively; Fig. 2). Examination of the amplicon sequencing dataset of Le Gall *et al.* (2024) identified 1 microhaplotype locus (Omy_1521) that mapped within the Omy20 inversion region, with alternate alleles that are diagnostic for the ancestral and inverted karyotypes. Population frequencies of the Omy20 ancestral form from Le Gall *et al.* (2024) were estimated based on this locus and merged with the

Omy20 haplotype frequencies inferred from the RAD-seq data. In the combined dataset, the ancestral form Omy20 CI form is abundant from ~33°N to ~41°N, but rapidly drops in frequency, being absent at ~44°N in anadromous populations except for the Elwha River, Washington (Fig. 2; Supplementary Tables 1 and 2), and no ancestral homozygotes were found in any population outside of California. There were populations outside of California with the heterozygous karyotype, but these were rare compared with the frequency of the inverted homokaryotype (e.g. 3.7% in the Elwha River, Washington). Interestingly, we did find evidence
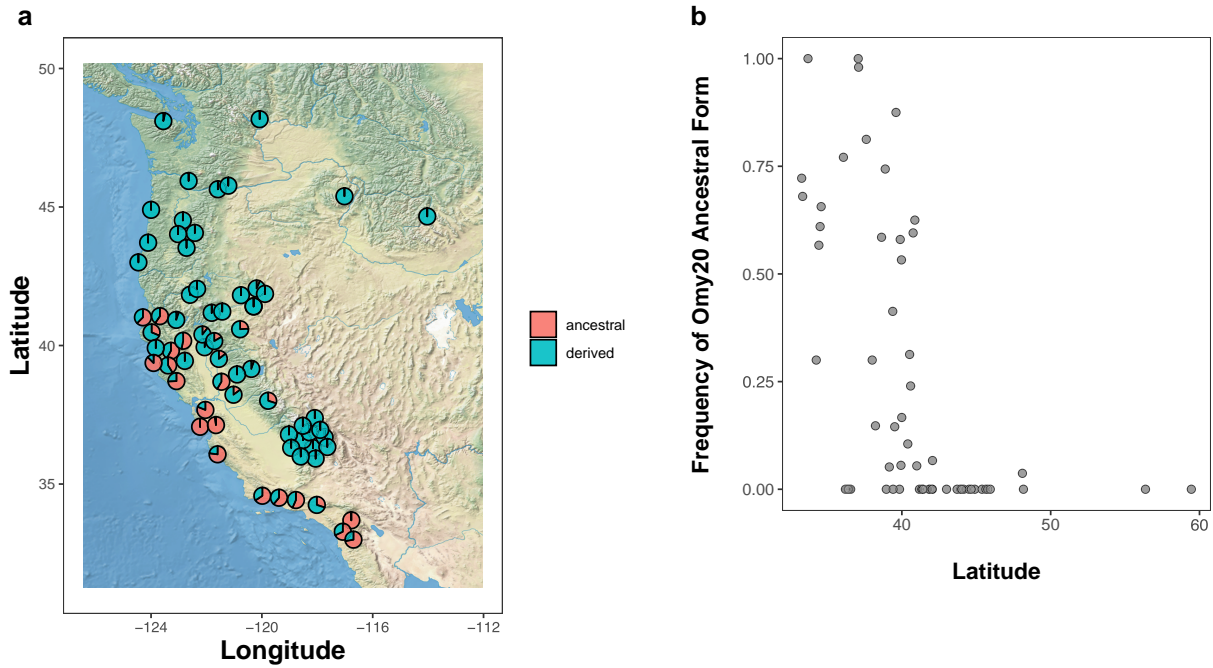
Fig. 2. Geographic distribution of Omy20 inversion types within Rainbow Trout from 69 populations (21 from RAD-seq data, 31 from amplicon data, and 17 from both datasets). Sample location, sample sizes, and frequencies of inversion types are reported in Supplementary Tables 1 and 2. Note that hatchery populations and populations with unclear ancestry are not shown. The frequency of ancestral and inverted types is plotted geographically in (a) with locations north of 50° latitude ($n = 2$) omitted, and (b) latitude on the *x*-axis is plotted vs frequency of the ancestral form the *y*-axis.
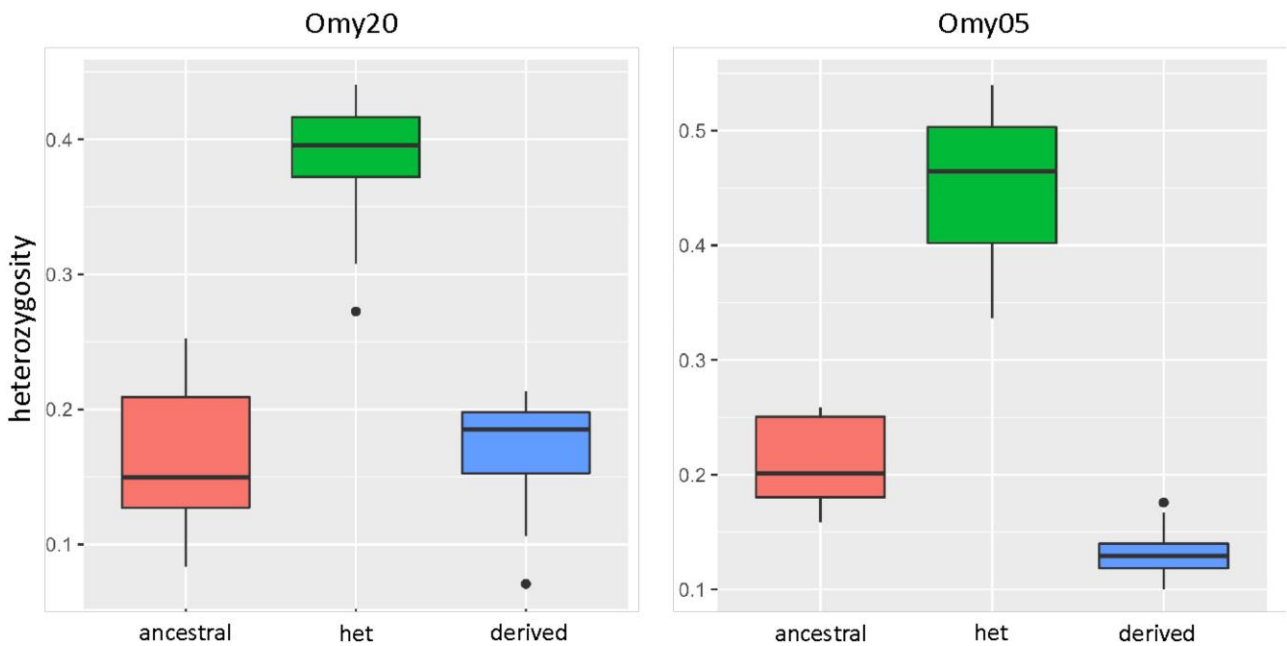

Fig. 3. Comparisons of heterozygosity for 76 fish from southern California. Each sample was genotyped at both the Omy20 and Omy05 CIs using 75 and 94 SNPs within these inversion regions, respectively. Heterozygosity was calculated in PLINK1.9. Differences in CI wide heterozygosity were tested using ANOVA: the differences between homokaryotypes were statistically significant for the Omy05 inversion and not for the Omy20 inversion.

for heterokaryotypes in 4 redband trout (*O. m. stonei*, Pit River and Surprise Valley regions, California), but note that this was a small percentage of the total number of redbands analyzed (1.4%; data not shown), suggesting the ancestral form is rare and is likely due to the movement of hatchery fish (that were presumably ancestral for the Omy20 CI) and subsequent hybridization with redband trout.

Population genetics statistics of the homokaryotypes

Differences in heterozygosity between the homokaryotypes are reduced for the Omy20 CI compared with the different forms of the Omy05 homokaryotypes (not significant between Omy20 homokaryotypes: $t = 0.105$, $P = 0.456$; significant between Omy05 homokaryotypes, $t = 7.427$, $P \leq 0.001$; Fig. 3). These results imply
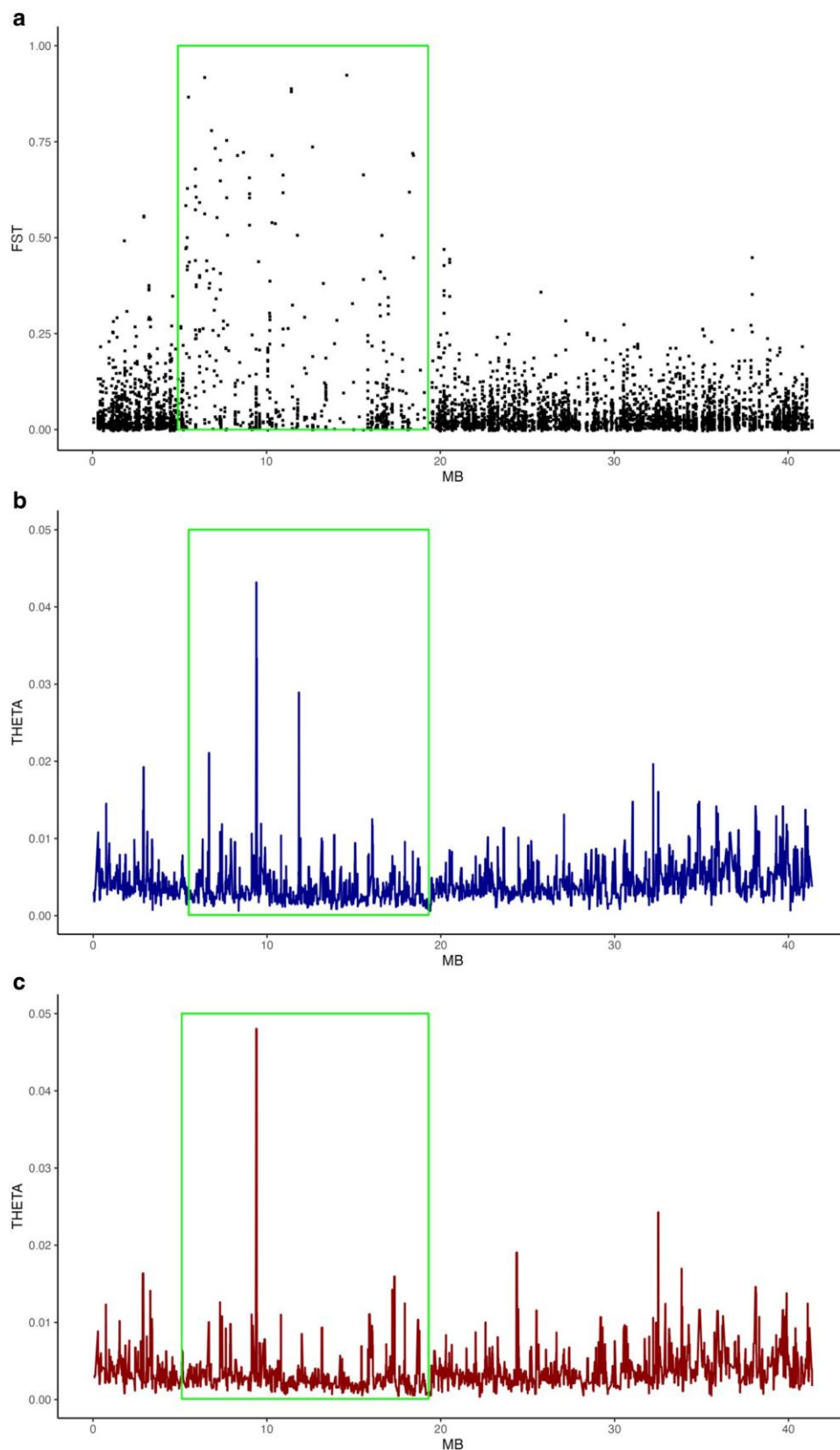
Fig. 4. Divergence among Omy20 CI homokaryotypes in *O. mykiss* from southern California. a) Nucleotide divergence (pointwise F_{ST}) between alternative homokaryotypes reveals the Omy20 CI as an area of elevated divergence. b) and c) show sequence diversity (π) among the ancestral and derived homokaryotypes. Estimates of sequence diversity are based on 50 kb windows. The location of the CI is indicated by a box.
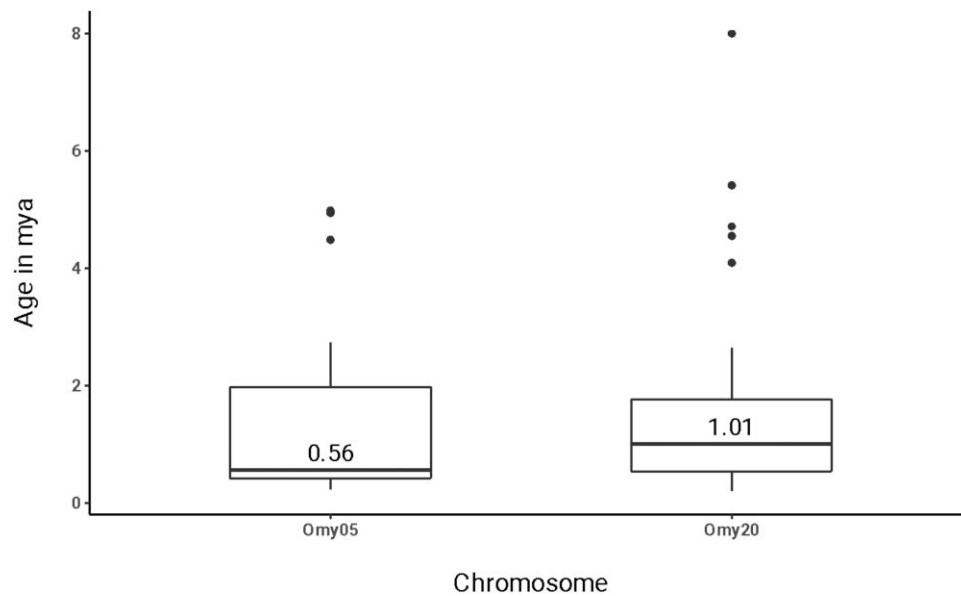
Fig. 5. Age of large inversions in rainbow trout. The median age of divergence times is indicated by the orthologs found across each inversion. These estimates are from the analysis of 38 protein-coding genes from the Omy20 CI and 44 genes from the Omy05 CI.

that the Omy20 inversion is older than the Omy05 inversion, as the inverted homokaryotype has had more time to accrue mutations than the inverted Omy05 homokaryotype, although the combined effects of drift and selection make it difficult to determine based on patterns of sequence variation (Stenløkk *et al.* 2022). We further investigated this difference in age between the 2 CIs by calculating F_{ST}, Watterson's theta, and Tajima's D for the homokaryotic individuals for Omy05 and Omy20. Our calculations of pairwise F_{ST} show the location of the Omy20 CI, with multiple loci producing high F_{ST} values compared with regions outside the CI (Fig. 4a). The Omy20 homokaryotypes showed similar patterns of Watterson's theta ($t = 2.234$, $P = 0.013$), with the ancestral homokaryotype showing marginally higher estimates of theta than the inverted form (although note that there do not appear to be differences in the distribution of theta values between the CI and non-CI regions of Omy20 (Fig. 4b and c). In contrast, the Omy05 homokaryotypes showed much greater differences in theta values between the different forms of the CI ($t = 12.303$, $P < 0.001$; data not given), suggesting the Omy05 inversion is more recent than the Omy20 CI. Tajima's D results were significant for both tests, with the inverted homokaryotype of the Omy05 CI showing significantly negative values compared with the ancestral ($t = 4.576$, $P < 0.001$; data not shown), suggesting the inverted form of the CI is accruing rare polymorphisms.

Dating of Omy20 and Omy05 inversions
The initial ortholog search identified a set of 34,294 genome-wide orthologs, of which 3,387 were single-copy orthologs in pike outgroups but duplicated in salmonids. Of these orthologs, 55 were located within the CI on Omy20, of which 44 met our requirements for topology (2 ohnolog clades present, sister relationships between rainbow trout and Atlantic salmon in each clade, inversion gene sequences from ancestral and derived forms most closely related). After requiring variation in rainbow trout genes to be present, 38 protein-coding genes were kept for estimating the age of the Omy20 CI inversion. Similarly, for the Omy05 inversion, a set of 266 high-confidence orthologs located within the Omy05

inversion was reduced to an inversion-wide set of 50 genes, with final filtering reducing the total number to 44 genes after orthologs with invariant rainbow trout genes were removed. The older age of the Omy20 CI compared with the Omy05 CI was supported by phylogenetic analyses (Fig. 5, Supplementary Fig. 1). The median age estimated for divergence of the Omy20 CI is ~1.00 MYA, with the median age of the Omy05 inversion of 0.56 MYA. The mean estimated ages are 1.61 and 1.20 MYA, respectively.

Functional characterization of the Omy20 inversion
Enrichment tests found 128 overrepresented GO terms in the protein-coding genes located within the Omy20 CI compared with the rest of the genome. Of these terms, 8 were associated with Cellular Component, 25 were with Molecular Functions, and 97 were with Biological Processes. Of the Biological Processes, 41 terms produced FDR-corrected P-values <0.00001 and are reported in Table 1. Although making inferences regarding the biological importance of enriched GO terms is difficult, it is striking that many GO terms connected to Wnt-signaling pathways, general development, and ion regulation were found to be over-present in the Omy20 CI. The Omy05 CI contains multiple genes connected to development, phototransduction, photoperiod recognition, and anadromy.

Discussion
The rarity and limited geographic range of the ancestral inversion suggest that selection for the inverted form of the Omy20 CI caused an increase in frequency during the colonization of the Pacific Northwest by rainbow trout. Presumably, the restriction of the ancestral version to (mostly) anadromous populations of coastal rainbow trout (*O. m. irideus*) in the southern portion of the species' range is because it encodes benefits to anadromous migrations or to local climate conditions, and that these benefits are lost in both land-locked resident populations and higher latitude anadromous populations. Pearse *et al.* (2019) proposed sex-dependent balancing selection as a mechanism behind the

Table 1. List of enriched GO terms within the Omy20 CI with a maximum FDR-corrected P-value of 0.00001.

GO ID	GO term	P-value	FDR P-value
0010456	Cell proliferation in dorsal spinal cord	1.59E−08	4.99E−12
0010226	Response to lithium ion	1.59E−08	5.82E−12
0060847	Endothelial cell fate specification	2.39E−08	1.05E−11
0060827	Regulation of canonical Wnt-signaling pathway involved in neural plate anterior/posterior pattern formation	5.94E−08	3.04E−11
0060829	Negative regulation of canonical Wnt-signaling pathway involved in neural plate anterior/posterior pattern formation	6.54E−08	3.86E−11
0060823	Canonical Wnt-signaling pathway involved in neural plate anterior/posterior pattern formation	6.54E−08	4.30E−11
0060784	Regulation of cell proliferation involved in tissue homeostasis	7.49E−08	5.48E−11
0048873	Homeostasis of number of cells within a tissue	7.63E−08	6.14E−11
0007004	Telomere maintenance via telomerase	8.87E−08	9.08E−11
0006278	RNA-templated DNA biosynthetic process	8.87E−08	9.08E−11
0060729	Intestinal epithelial structure maintenance	1.82E−07	2.13E−10
0010833	Telomere maintenance via telomere lengthening	2.24E−07	2.94E−10
0030277	Maintenance of gastrointestinal epithelium	3.14E−07	4.42E−10
0035545	Determination of left/right asymmetry in nervous system	3.14E−07	4.83E−10
0035462	Determination of left/right asymmetry in diencephalon	3.14E−07	4.83E−10
2000054	Negative regulation of Wnt-signaling pathway involved in dorsal/ventral axis specification	3.55E−07	5.71E−10
0019747	Regulation of isoprenoid metabolic process	4.52E−07	8.59E−10
1900052	Regulation of retinoic acid biosynthetic process	4.52E−07	8.59E−10
0032350	Regulation of hormone metabolic process	4.52E−07	8.59E−10
0030656	Regulation of vitamin metabolic process	5.09E−07	1.01E−09
0045995	Regulation of embryonic development	1.12E−06	2.30E−09
0002138	Retinoic acid biosynthetic process	1.14E−06	2.43E−09
0021979	Hypothalamus cell differentiation	1.16E−06	2.55E−09
0010669	Epithelial structure maintenance	1.19E−06	2.78E−09
0016102	Diterpenoid biosynthetic process	1.95E−06	4.70E−09
0042180	Cellular ketone metabolic process	2.35E−06	5.84E−09
0016114	Terpenoid biosynthetic process	3.09E−06	8.14E−09
0042573	Retinoic acid metabolic process	3.58E−06	9.69E−09
0050769	Positive regulation of neurogenesis	3.58E−06	9.94E−09
0050821	Protein stabilization	4.64E−06	1.32E−08
0044283	Small molecule biosynthetic process	5.39E−06	1.58E−08
0072148	Epithelial cell fate commitment	6.20E−06	1.86E−08
0021879	Forebrain neuron differentiation	6.67E−06	2.05E−08
0021872	Forebrain generation of neurons	7.36E−06	2.41E−08
0060839	Endothelial cell fate commitment	7.36E−06	2.42E−08
2000053	Regulation of Wnt-signaling pathway involved in dorsal/ventral axis specification	7.59E−06	2.68E−08
0000723	Telomere maintenance	7.59E−06	2.72E−08
0032200	Telomere organization	7.96E−06	2.91E−08
0072330	Monocarboxylic acid biosynthetic process	9.35E−06	3.49E−08
0060323	Head morphogenesis	9.45E−06	3.59E−08
0021513	Spinal cord dorsal/ventral patterning	9.54E−06	3.77E−08
0010456	Cell proliferation in dorsal spinal cord	1.59E−08	4.99E−12

Each term was enriched for protein-coding genes within the CI compared with protein-coding genes outside the CI.

maintenance of both forms of the Omy05 CI in rainbow trout (anadromous individuals benefit from the ancestral form and resident individuals from the inverted form), and it is tempting to suggest that the 2 forms of the Omy20 CI may also be maintained by some form of balancing selection. An alternative hypothesis is that dispersal and colonization of new environments after the last ice age were predominated by fish with the inverted version of the Omy20 CI, because those individuals were predisposed to dispersal, had a selective advantage, or simply because of the random effects of drift. These hypotheses are not mutually exclusive, especially when one considers that many anadromous steelhead in northern latitudes are homozygous for the inverted forms of both Omy05 and Omy20 CIs (e.g. Hale *et al.* 2013; Weinstein *et al.* 2019), demonstrating that anadromy can and does occur in fish without the ancestral form of either CI.

Given the association (albeit more restricted in geographic range) between the 2 Omy20 homokaryotypes and anadromy, it is tempting to suggest that selection for different developmental processes has maintained the 2 homokaryotypes—at least in California (Campbell *et al.* 2021). However, it is important to note

that anadromy is a widespread behavior for rainbow trout across the species' range, and there was no indication of the ancestral homokaryotype in steelhead from Oregon (e.g. Little Sheep Creek, Siletz, Hood, Willamette, and Umpqua Rivers), Washington (the Elwha, Methow, Klickitat, and Lewis Rivers), and Alaska (Sashin Creek and potadromy in Lake Iliamna). Ultimately, although many of the enriched GO terms are associated with development, the functions of the terms suggest that many of the protein-coding genes within the Omy20 CI are connected to different pathways. This is not surprising, as the large size of the CI encompasses 311 protein-coding genes. Further research into how the 2 homokaryotypes are connected to different phenotypes is necessary to more fully understand how selection has maintained the 2 forms of the CI.

Ultimately, the characterization of these alternative forms of the Omy20 CI, together with our population genetic analyses, increases the knowledge of this so far understudied CI in rainbow trout. Although Campbell *et al.* (2021) suggested a link between the Omy20 CI and migratory life history, we recommend caution for 2 reasons: (1) The ancestral form of the CI is nearly absent

north of 42°, at least in the populations sampled (reemphasizing the idea that anadromy and life-history development are controlled by both population-level-specific genetic mechanisms and the environment, Kendall *et al.* 2015), and (2) the lack of phenotypic data associated with the investigated samples makes it impossible to associate the effects of the 2 homokaryotypes on specific traits. This finding contrasts with many studies that note the association of the Omy05 CI and migratory life history in multiple populations of anadromous rainbow trout throughout the Pacific (e.g. Pearse *et al.* 2014, 2019; Arostegui *et al.* 2019; Fraik *et al.* 2021), as well as other related life-history traits (development rate and juvenile growth rate), despite a strong cline in the frequency of the inverted haplotype (Pearse *et al.* 2019). What is clear from our analyses is that the Omy20 CI predates the Omy05 CI, suggesting that the derived form of the CI increased in frequency after expanding from southern glacial refugia into the Pacific Northwest. Moreover, the lack of evidence for the ancestral form of the CI in interior forms of rainbow trout strongly points to the inverted form of the CI increasing in frequency during colonization.

Identification of CIs often relies on comparing the genetic diversity of homozygous and heterozygous individuals, and ancestral heterozygosity is expected to exceed inverted heterozygosity (e.g. Hale *et al.* 2021). While this pattern is true for the Omy05 CI in rainbow trout, it is not true for the Omy20 CI (Fig. 3). We find that the age of the Omy20 CI exceeds that of the Omy05 CI and that the ancestral Omy20 CI is geographically and phylogenetically restricted. In the case of the Omy20 CI, the age of the inversion has substantial impacts on the observed levels of diversity, a factor that should be considered by others as more and more studies document and characterize the presence of CIs in nonmodel organisms (Wellenreuther and Bernatchez 2018).

Data availability

All sequence data analyzed in this study are publicly available with repository information available in Supplementary Tables 1 and 2.

Supplemental material available at G3 online.

Acknowledgments

The authors would like to thank the members of the Hale lab, Krista Nichols, and 2 anonymous reviewers for their thoughtful comments on previous drafts of the manuscript.

Conflicts of interest

The author(s) declare no conflicts of interest.

Literature cited

Arostegui MC, Quinn TP, Seeb LW, Seeb JE, McKinney GJ. 2019. Retention of a chromosomal inversion from an anadromous ancestor provides the genetic basis for alternative freshwater ecotypes in rainbow trout. Mol Ecol. 28(6):1412–1427. doi:10.1111/mec.15037.

Bertolotti AC, Layer RM, Gundappa MK, Gallagher MD, Pehlivanoglu E, Nome T, Robledo D, Kent MP, Rosaeg LL, Holen MM, *et al.* 2020. The structural variation landscape in 492 Atlantic salmon genomes. Nat Commun. 11(1):5176. doi:10.1038/s41467-020-18972-x.

Campbell MA, Anderson EC, Garza JC, Pearse DE. 2021. Polygenic basis and the role of genome duplication in adaptation to similar selective environments. J Hered. 112(7):614–625. doi:10.1093/jhered/esab049.

Campbell MA, Habibi E, Auringer G, Stephens M, Rodzen J, Conway KW, Finger AJ. 2023. Molecular systematics of Redband Trout from genome-wide DNA sequencing substantiates the description of a new taxon (Salmonidae: *Oncorhynchus mykiss calisulat*) from the McCloud River. Zootaxa. 5254(1):1–29. doi:10.11646/zootaxa.5254.1.1.

Campbell MA, López JA, Sado T, Miya M. 2013. Pike and salmon as sister taxa: detailed intraclade resolution and divergence time estimation of Esociformes + Salmoniformes based on whole mitochondrial genome sequences. Gene. 530(1):57–65. doi:10.1016/j.gene.2013.07.068.

Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer databases. GigaScience. 4(1):7. doi:10.1186/s13742-015-0047-8.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization, and analysis in functional genomics research. Bioinformatics. 21(18):3674–3676. doi:10.1093/bioinformatics/bti610.

Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol. 10(1):210. doi:10.1186/1471-2148-10-210.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20(1):238. doi:10.1186/s13059-019-1832-y.

Euclide PT, Larson WA, Shi Y, Christensen KA, Seeb J, Seeb L. 2023. Conserved islands of divergence associated with adaptive variation in sockeye salmon are maintained by multiple mechanisms. Mol Ecol. 00:1–21. doi:10.1111/mec.17126.

Flagel LE, Blackman BK, Fishman L, Monnahan PJ, Sweigart A, Kelly JK. 2019. GOOGA: a platform to synthesize mapping experiments and identify genomic structural diversity. PLoS Comput Biol. 15(4):e1006949. doi:10.1371/journal.pcbi.1006949.

Fraik AK, Mcmillan JR, Liermann M, Bennett T, Mchenry ML, Mckinney GJ, Wells AH, Winans G, Kelley JL, Pess GR, *et al.* 2021. The impacts of dam construction and removal on the genetics of recovering steelhead (*Oncorhynchus mykiss*) populations across the Elwha River watershed. Genes. 12(1):89. doi:10.3390/genes12010089.

Gao G, Magadan S, Waldbieser GC, Youngblood RC, Wheeler PA, Scheffler BE, Thorgaard GH, Palti Y. 2021. A long-reads based de-novo assembly of the genome of the Arlee homozygous line reveals chromosomal rearrangements in rainbow trout. G3 (Bethesda). 11(4):jkab052. doi:10.1093/g3journal/jkab052.

Hale MC, Campbell MA, McKinney GJ. 2021. A candidate chromosome inversion in Arctic charr (*Salvelinus alpinus*) identified by population genetic analysis techniques. G3 (Bethesda). 11(10):jkab267. doi:10.1093/g3journal/jkab267.

Hale MC, Colletti JA, Gahr SA, Scardina J, Thrower FP, Harmon M, Carter M, Phillips RB, Thorgaard GH, Rexroad CA. 2014. Mapping and expression of candidate genes for development rate in rainbow trout (*Oncorhynchus mykiss*). J Heredity. 105(4):506–520. doi:10.1093/jhered/esu018.

Hale MC, Thrower FP, Berntson EA, Miller MR, Nichols KM. 2013. Evaluating adaptive divergence between migratory and nonmigratory ecotypes of a salmonid fish, *Oncorhynchus mykiss*. G3 (Bethesda). 3(8):1273–1285. doi:10.1534/g3.113.006817.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 35(2):518–522. doi:10.1093/molbev/msx281.

Huang K, Andrew RL, Owens GL, Ostevik KL, Rieseberg LH. 2020. Multiple chromosomal inversion contribute to adaptive divergence of a dune sunflower ecotype. Mol Ecol. 29(14):2535–2549. doi:10.1111/mec.15428.

Huang K, Rieseberg LH. 2020. Frequency, origins, and evolutionary role of chromosomal inversions in plants. Front Plant Sci. 11: 296. doi:10.3389/fpls.2020.00296.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30(14):3059–3066. doi:10.1093/nar/gkf436.

Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform. 9(4):286–298. doi:10.1093/bib/bbn013.

Kendall NW, McMillan JR, Sloat MR, Buehrens TW, Quinn TP, Pess GR, Kuzishchin KV, McClure MM, Zabel RW. 2015. Anadromy and residency in steelhead and Rainbow Trout (*Oncorhynchus mykiss*): a review of the processes and patterns. Can J Fish Aquat Sci. 342:319–342.

Kirkpatrick M. 2010. How and why chromosome inversions evolve. PLoS Biol. 8(9):e1000501. doi:10.1371/journal.pbio.1000501.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. BMC Bioinformatics. 15(1):356. doi:10.1186/s12859-014-0356-4.

Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB. 2022. TimeTree 5: an expanded resource for species divergence times. Mol Biol Evol. 39(8):msac174. doi:10.1093/molbev/msac174.

Le Gall R, Barthelemy N, Clemento AJ, Columbus CD, Campbell E, Correa ECA, Rodzen JA, Garza JC, Pearse DE. 2023. Development of a novel microhaplotype panel for steelhead/rainbow trout (*Oncorhynchus mykiss*) and application for phylogenetic analysis in California. Res Square. doi:https://doi.org/10.21203/rs.3.rs-2949400/v1.

Le Gall R, Barthelemy N, Clemento AJ, Columbus CD, Campbell E, Correa ECA, Rodzen JA, Garza JC, Pearse DE. 2024. Development of a novel microhaplotype panel for steelhead/rainbow trout (*Oncorhynchus mykiss*) and application for phylogenetic analysis in California. Conserv Genet Resour https://doi.org/10.21203/rs.3.rs-2949400/v1x, preprint: not peer reviewed.

Lehnert SJ, Bentzen P, Kess T, Lien S, Horne JB, Clement M, Bradbury IR. 2019. Chromosome polymorphisms track trans-Atlantic divergence and secondary contact in Alantic salmon. Mol Ecol. 28(8):2074–2087. doi:10.1111/mec.15065.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 27(21): 2987–2993. doi:10.1093/bioinformatics/btr509.

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

Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, Leong JS, Minkley DR, Zimin A, *et al.* 2016. The Atlantic salmon genome provides insights into rediploidization. Nature. 533(7602):200–205. doi:10.1038/nature17164.

Macqueen DJ, Johnston IA. 2014. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proc R Soc B. 281(1778): 20132881. doi:10.1098/rspb.2013.2881.

McCusker MR, Parkinson E, Taylor EB. 2000. Mitochondrial DNA variation in rainbow trout (*Oncorhynchus mykiss*) across its native range: testing biogeographical hypotheses and their relevance to conservation. Mol Ecol. 9(12):2089–2108. doi:10.1046/j.1365-294X.2000.01121.x.

McKinney G, McPhee MV, Pascal C, Seeb JE, Seeb LW. 2020. Network analysis of linkage disequilibrium reveals genome architecture in chum salmon. G3 (Bethesda). 10(5):1553–1561. doi:10.1534/g3.119.400972.

Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, Wainwright PC, Friedman M, Smith WL. 2012. Resolution of ray-finned fish phylogeny and timing of diversification. Proc Natl Acad Sci U S A. 109(34):13698–13703. doi:10.1073/pnas.1206625109.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol Biol Evol. 32(1):268–274. doi:10.1093/molbev/msu300.

Nichols KM, Broman KW, Sundin K, Young JM, Wheeler PA, Thorgaard GH. 2007. Quantitative trait loci x maternal cytoplasmic environment interaction for development rate in *Oncorhynchus mykiss*. Genetics. 175(1):335–347. doi:10.1534/genetics.106.064311.

Pearse DE, Barson NJ, Nome T, Gao G, Campbell MA, Abadía-Cardoso A, Anderson EC, Rundio DE, Williams TH, Naish KA, *et al.* 2019. Sex-dependent dominance maintains migration supergene in rainbow trout. Nat Ecol Evol. 3(12):1731–1742. doi:10.1038/s41559-019-1044-6.

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 Biol Sci. 281(1783):20140012. doi:10.1098/rspb.2014.0012

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreria M, Bender D, Maller J, Sklar P, de Bakker P, Daly MJ, *et al.* 2007. PLINK: a tool-set for whole-genome and population-based linkage analyses. Am J Hum Genet. 81(3):559–575. doi:10.1086/519795.

Revell LJ. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol. 3(2):217–223. doi:10.1111/j.2041-210X.2011.00169.x.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. Mrbayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 61(3):539–542. doi:10.1093/sysbio/sys029.

Samy JKA, Mulugeta TD, Nome T, Sandve SR, Grammes F, Kent MP, Lien S, Våge DI, *et al.* 2017. SalmoBase: an integrated molecular data resource for Salmonid species. BMC Genomics. 18(1):482. doi:10.1186/s12864-017-3877-1.

Schwander T, Libbrecht R, Keller L. 2014. Supergenes and complex phenotypes. Curr Biol. 24(7):R288–R294. doi:10.1016/j.cub.2014.01.056.

Stenløkk K, Saitou M, Rud-Johansen L, Nome T, Moser M, Arnyasi M, Kent M, Barson NJ, Lien S. 2022. The emergence of supergenes from inversions in Atlantic salmon. Philos Trans R Soc Lond B. 377(1856):20210195. doi:10.1098/rstb.2021.0195.

Sundin K, Brown KH, Drew RE, Nichols KM, Wheeler PA, *et al.* 2005. Genetic analysis of a development rate QTL in backcrosses of clonal rainbow trout, *Oncorhynchus mykiss*. Aquaculture. 247(1–4): 75–83. doi:10.1016/j.aquaculture.2005.02.054.

Weinstein SY, Thrower FP, Nichols KM, Hale MC. 2019. A large-scale chromosomal inversion is not associated with life history development in rainbow trout from Southeast Alaska. PLoS One. 14(9):e0223018. doi:10.1371/journal.pone.0223018.

Wellenreuther M, Bernatchez L. 2018. Eco-evolutionary genomics of chromosomal inversions. Trends Ecol Evol. 33(6):427–440. doi:10.1016/j.tree.2018.04.002.

Wilson MVH, Brinkman DB, Neuman AG. 1992. Cretaceous esocoidei (Teleostei): early radiation of the pikes in North American fresh waters. J Paleontol. 66(5):839–846. doi:10.1017/S0022336000020849.

Wilson MVH, Williams RRG. 1992. Phylogenetic, biogeographic, and ecological significance of early fossil records of North American freshwater teleostean fishes. In: Mayden RL, editor. Systematics, Historical Ecology, & North American Freshwater Fishes. Stanford (CA): Stanford University Press. p. 969.