Molecular Ecology Resources

WILEY

MOLECULAR ECOLOGY RESOURCES

RESOURCE ARTICLE OPEN ACCESS

De Novo Genome Assemblies of Four Rainbow Trout Genetic Lines Reveal Structural Variants in Pursuit of a Pangenome Reference

Ali Ali[1] | Geoffrey C. Waldbieser[2] | Ramey C. Youngblood[3] | Paul A. Wheeler[4] | Brian E. Scheffler[5] | Stuart Willis[6] | Shawn R. Narum[6] | Gary H. Thorgaard[4] | Mohamed Salem[1] | Yniv Palti[7]

[1]Department of Animal and Avian Sciences, University of Maryland, College Park, Maryland, USA | [2]Warmwater Aquaculture Research Unit, USDA-ARS, Stoneville, Mississippi, USA | [3]Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Mississippi State, Mississippi, USA | [4]School of Biological Sciences and Center for Reproductive Biology, Washington State University, Pullman, Washington, USA | [5]Genomics and Bioinformatics Research Unit, USDA-ARS, Stoneville, Mississippi, USA | [6]Fishery Science Department, Columbia River Inter-Tribal Fish Commission, Hagerman, Idaho, USA | [7]National Center for Cool and Cold Water Aquaculture, USDA-ARS, Kearneysville, West Virginia, USA

Correspondence: Mohamed Salem (mosalem@umd.edu) | Yniv Palti (yniv.palti@usda.gov)

Received: 10 October 2025 | **Revised:** 27 April 2026 | **Accepted:** 10 June 2026

Keywords: de-novo genome assembly | pangenome | rainbow trout | reference genome | structural variants

ABSTRACT

Rainbow trout (*Oncorhynchus mykiss*) exhibit extensive genomic diversity shaped by domestication, life history and geographic origin. To advance the development of a comprehensive pangenome reference, we present new de novo genome assemblies of two genetically and ecologically distinct lines: Whale Rock (WR; wild, landlocked, Central California) and Keithley Creek (KC; wild, resident, interior Columbia Basin), along with the previously published assemblies of the Arlee (domesticated, Northern California) and Swanson (semi-domesticated, resident, Alaska) lines. All assemblies provide nearly complete coverage of known genes (BUSCO 95.8%–99.7%) and are similar in genome size (~2.3 Gb), with scaffold N50 values between 3.4 Mb (KC) and 52.4 Mb (Swanson). Comparative whole-genome alignments revealed high sequence conservation (97%–98% identity) among assemblies, but also evidence of extensive structural variation of at least 50 bp in length. Structural variant (SV) profiling identified tens of thousands of deletions, insertions and complex rearrangements largely in noncoding sequences. In an initial assessment of the utility of having multiple de novo genome assemblies for rainbow trout, we found that two strains (Arlee and Swanson; domesticated) share SVs enriched in genes linked with growth, reproduction and adaptation to domestication, such as GTP binding and ECM-receptor interaction. In comparison, the other two strains (WR and KC; wild origin) share SVs associated with reproductive timing, such as the GnRH signalling pathway. Both Arlee and WR also have unique SVs potentially related to their geographic origin and unique life history. Additionally, we identified SVs in key regions, such as a QTL for fillet yield on Omy17 and the maturation-associated *six6/erβ-gphb5* locus on Omy25q, suggesting the importance of considering SVs when investigating the genomics of complex traits. Together, these assemblies and comparative analyses establish a foundation for a rainbow trout pangenome reference, illuminating how they can be utilized to reveal the structural genomic basis of domestication, adaptation, and other complex traits in *O. mykiss*.

1 | Introduction

Rainbow trout (*Oncorhynchus mykiss*) originating from the west coast of North America have been introduced globally for aquaculture and recreational fishing (Savini et al. 2010). It is one of the most widely cultivated fish species, where hatchery artificial selection dates back to the 1870s in California. While selective breeding for aquaculture and sport fishing has shaped the genetic makeup of domesticated stocks, anthropogenic and natural barriers (e.g., waterfalls, landslides, dams) have contributed to the geographic isolation of many of the wild populations. All these forces have contributed to the remarkable genomic and ecological diversity of the species (Leeds et al. 2010; Rexroad Iii et al. 2006; Parkinson et al. 1999; Deiner et al. 2007).

As a highly adaptable salmonid, rainbow trout has significant ecological, economic and evolutionary importance (Ali et al. 2025; Pearse et al. 2019). The species has been widely used as a model for studying adaptation, domestication and population genetics. Rainbow trout genetic lines have demonstrated differences in temperature tolerance (Ineno et al. 2018), degree of domestication (Campbell et al. 2015), behaviour (Lucas et al. 2004), development time (Robison and Thorgaard 2004) and disease resistance (Zwollo et al. 2015). For example, genomic differences were found in IgL repertoire between lines of trout which could affect immune responses during disease (Rego et al. 2020). Additionally, trout clonal lines exhibited significant differences in the relative expression of HSP70 under thermal stress, suggesting genetic differences underlying the variation (Heredia-Middleton et al. 2008). Understanding the genomic basis of the variation is crucial for conservation efforts, sustainable aquaculture and improving hatchery breeding programmes.

Structural genome variation is broadly defined as insertion/deletions, inversions, duplications and translocations that are larger than 50 bp (Alkan et al. 2011). While sequence-level (e.g., single-nucleotide polymorphism or SNP) variation has long been studied as a source of genetic differences between and within brood lines or populations, a growing body of research is demonstrating the large importance of structural variation in shaping important phenotypic variation in humans and in numerous domesticated and wild vertebrate species, including fishes (Pearse et al. 2019; Collins and Talkowski 2025; Chen, Khan, et al. 2024; Lecomte et al. 2024). For example, a strong association between two adjacent inversions and migratory ecotypes was documented in Atlantic cod (*Gadus morhua*) (Kirubakaran et al. 2016). Similarly, strong associations between inversions and adaptation to different habitats like freshwater and seawater were also found in other fish species like the threespine stickleback (Jones et al. 2012; Terekhanova et al. 2019). Additionally, structural deletion variants were found to influence phenotypic variation in fish; for example, in sticklebacks, recurrent deletions of a *Pitx1* enhancer underlie pelvic reduction, highlighting how regulatory SVs can drive adaptive morphological changes (Chan et al. 2010). Perhaps the most well documented association between structural genome variation and evolutionary adaptation in rainbow trout is the one between the large double inversion on chromosome 5 (Omy05) and sex dependent migratory behaviour (Pearse et al. 2019). However, despite their large impact on genomic and genetic diversity in model organisms and the growing list of cases in wild fish populations, SVs are still understudied in most species, including rainbow trout. This is primarily due to the larger technical challenge and large investment in genome sequencing that is required for the detection and genotyping of SVs. To that end, multiple de novo genome assemblies and the generation of *pan-genome* references can provide invaluable resources for the detection of SVs in rainbow trout and for research on their association and impact on traits of interest. While originally applied to microbial species which exhibit a 'core' set of genes present across all individuals and a large library of variably present 'accessory' genes, the term *pan-genome* has been increasingly used in reference to plant and animals genomic resources as the unexpected degree of structural variation among individuals has become apparent. Herein, we use *pan-genome* to refer to a utilitarian compendium of the known small-scale (e.g., SNP) and large-scale (e.g., SV) variation among several to many individual genome assemblies of a species (Gong et al. 2023).

Recent advances in sequencing technologies have enabled the de novo assembly of high-quality reference genomes of non-model organisms. The rainbow trout genome assemblies from the Swanson and Arlee lines are notable examples that have been assembled using advanced sequencing methods (Ali et al. 2025, 2021; Pearse et al. 2019; Salem et al. 2010; Gao et al. 2021). This provided valuable insights into the species genomic architecture. The Swanson and Arlee genome assemblies have been used for studying SNPs, SVs and other genomic features that contribute to phenotypic diversity and population differentiation (Ali et al. 2021; Gao et al. 2021, 2018; Salem et al. 2018; Liu et al. 2021; Vallejo et al. 2017; Palti et al. 2024; Larson et al. 2017). For example, alignments of the chromosome sequences from the Swanson and Arlee genome assemblies have revealed karyotype differences due to fissions of chromosome arms in the metacentric chromosomes Omy04, Omy14 and Omy25 and major inversions on chromosomes Omy05, Omy20 and Omy26 (Ali et al. 2025; Pearse et al. 2019; Gao et al. 2021). However, those two reference genomes do not fully represent the species' genetic diversity, which evolved as a result of multiple events leading to geographic isolation, domestication and local adaptation (Hale et al. 2024; de los Ríos-Pérez et al. 2020). To fully capture the genomic landscape of rainbow trout, more efforts are still needed to assemble multiple reference genomes that will provide a more complete representation of the collective genetic diversity of the species.

In this study, we describe two new chromosome-level de novo genome assemblies for the Whale Rock (WR) and Keithley Creek (KC) genetic lines and present an initial examination of the breadth of genomic diversity across rainbow trout strains through comparative analysis of the four representative genomes. The strains from which those genetic lines originated have distinct ecological and evolutionary histories. Therefore, their genomes are expected to have unique genetic variants associated with adaptation to specific environments. The WR doubled haploid homozygous line originated from a wild landlocked steelhead population from the WR reservoir in coastal central California. The original anadromous steelhead

population became landlocked by the dam construction that created the WR reservoir in 1961 and steelhead were recorded migrating from the tributaries into the reservoir as 1- and 2-year-olds in the 1970s and 1980s (Titus et al. 2010). In addition, maturing adult fish from the WR line exhibit silver body coloration that is consistent with smoltification in steelhead (Thorgaard and Wheeler, Unpublished). Therefore, it is likely that the population from which the WR line was derived has retained migratory behaviour and tolerance to marine or estuarine conditions, similar to the observed ability of landlocked steelhead descendants to produce viable smolts and adults after 70 years in freshwater (Thrower and Joyce 2005). On the other hand, the wild Redband trout from KC is native to inland river systems, and thus it represents the inland lineage of rainbow trout (*O. m. gairdneri*) in contrast to the three previous assemblies that were done using fish from the coastal lineage (*O. m. irideus*) (Andrews et al. 2023). By comparing these genomes with the previously assembled Swanson and Arlee strain genomes, we identify SVs, changes in gene content and other genomic features that may underlie differences in their history of domestication, phylogeographic divergence or local adaptation. Moreover, several recent studies have described genomic regions associated with life history variants important for population resilience in exploited or imperilled anadromous populations of *O. mykiss* (steelhead), including migration-timing, age-at-maturity and repeat spawning phenology (Willis et al. 2020, 2023). However, these analyses relied exclusively on examinations of small-scale sequence variation (SNP and short insertion/deletions) based on whole genome data aligned to a single genome assembly. To further demonstrate the potential utility of a pan-genome reference for rainbow trout, we used comparative analysis of the four de-novo genome assemblies in those previously described genomic regions with the aim of identifying larger structural variations that are co-localized with the smaller sequence variants and that were otherwise undetected to date.

The overall aim of the comparative analyses we described in this manuscript was to demonstrate how de novo genome assemblies in rainbow trout can be used for the identification of conserved and strain-specific genomic regions, with the goal of providing the basis of a *pan-genome* for rainbow trout that can be used for future studies on the association between genome variants and phenotypes as well as for exploring the genetic basis of adaptation, domestication and population structure.

2 | Materials and Methods

2.1 | Rainbow Trout Sampled for the Genome Assemblies

The WR rainbow trout we used for this de novo genome assembly was sampled from the established doubled haploid YY male line at Washington State University (Gao et al. 2018; Palti et al. 2014). Similar to the Arlee line, it has $2N = 64$ chromosomes and it originated from the Central California Coast. The one known major difference is that it originated from a wild fish from a landlocked steelhead population while the Arlee line was established from a domestic freshwater hatchery stock.

The KC rainbow trout line was prepared using the established protocol to induce androgenesis with the aim of starting a doubled haploid line at Washington State University through irradiation of eggs obtained from an aquaculture stock and fertilization of the irradiated eggs with milt from a single wild KC rainbow trout male, which was followed by heat shock of 31.5°C for 5 min that was applied 260 min post fertilization to block the first cell cleavage of the fertilized eggs (Komen and Thorgaard 2007). However, since our DNA sequences mapping analysis in the process of the genome assembly indicated a higher level of heterozygosity than in the established doubled haploid rainbow trout lines (Palti et al. 2014), we are obligated to caution that in the process that was used to establish the KC line at Washington State University, there might have been an inadvertent maternal contribution from the donor of the irradiated eggs, which was an aquaculture strain of rainbow trout possibly from a coastal origin.

2.2 | WR Line Genome Assembly

The assembly workflow for the WR line was composed of these steps: Genome sequencing → Contig assembly → BioNano scaffolding → Hi-C scaffolding → Chromosome-scale ordering/orientation using linkage map information. Software default parameters were used unless otherwise noted.

2.2.1 | Genome Sequencing

DNA from a single fish from the WR DH line was extracted from a fresh blood sample collected in EDTA and kept on ice prior to extraction (within 24 h) using the Qiagen DNeasy kit (Qiagen, Germantown, MD). This approach ensured a high yield of high molecular weight DNA suitable for long-read sequencing, following previously described methods (Gao et al. 2021) (NCBI Bio-sample Accession SAMN32933077). Genomic DNA was quantified using Qubit DNA Assay (Thermo Fisher Scientific, Waltham, MA) and quality of the high molecular weight DNA was verified on 1% agarose gel and with the Agilent 2100 Bioanalyzer System (Agilent, Santa Clara, CA). Long-read DNA sequencing was performed using the continuous long reads (CLR) mode on the PacBio RS-II system (CD Genomics, Shirley, NY), generating 14.2 million reads with 254.7 Gb of DNA sequence data (106× coverage; NCBI SRA accession SRX19198672). An additional 5.5 million high fidelity (HiFi) reads with total length of 74.8 Gb (31× coverage; accessions SRX27459513, SRX27459514, SRX27459515, SRX27459516) were generated using circular consensus sequencing (CCS) mode on the PacBio Sequel II system at the USDA-ARS Genomics and Bioinformatics Research Unit (GBRU) in Stoneville, MS. In addition, Illumina whole-genome sequence data was used for polishing and indel corrections of the assembly sequence scaffolds. A total of 558 million whole genome Illumina paired-end reads (2 × 150 bp; 70× coverage) were generated by a third party service provider (Admera Health, South Plainfield, NJ) using the Kapa Hyper Prep PCR free kit, and submitted to the NCBI SRA database (SRX21517362). The short-read sequences were processed with the programme Trimmomatic (v0.38) to remove the Illumina sequencing adaptors.

2.2.2 | Assembly of Contigs

The WR line has a doubled haploid genome that is nearly 100% homozygous. Hence we followed the approach that we previously used for the genome assemblies of the Arlee (Gao et al. 2021) and Swanson (Ali et al. 2025) doubled haploid lines. This approach used the CLR PacBio mode that generates longer reads with a higher base calling error rate and the Canu assembler to achieve an overall higher contiguity. PacBio long reads were assembled into contigs using the Canu assembler (v2.2) (Koren et al. 2017). The raw PacBio reads were aligned for error correction and were trimmed of poor-quality segments and overlapping sequences, after which Canu retained 2.5 million high-quality reads with 84.6 Gb of DNA sequence data (35.3× coverage).

2.2.3 | Assembly of Scaffolds

2.2.3.1 | Bionano. The Bionano optical mapping with the Saphyr platform (Bionano Genomics, San Diego, CA, USA) was used to improve the contiguity of the assembly by generating the Bionano scaffolds assembly. DNA was extracted from a blood sample of the WR Male DH line fish and then labelled with Direct Label Enzyme (DLE-1). The Bionano Solve software (Solve3.3_10252018) (https://bionanogenomics.com/support-page/bionano-access-software/) was used to process the raw mapping data yielding a de novo Bionano assembly guided by the Canu contigs assembly. The Bionano hybrid scaffolds pipeline was then used to scaffold the Canu assembled contigs with the Bionano genome map. We then used the PacBio HiFi sequence data to polish and correct sequence errors in the scaffold sequences. In the first step, we removed all the scaffolds that were smaller than 5 Kb from the Bionano hybrid scaffolds assembly to retain 1183 scaffolds. In the second step, the HiFi reads were mapped to the remaining scaffolds using pbmm2 (PacBio minimap2), and in the third step, we used Racon (Vaser et al. 2017) to correct sequencing errors in the scaffolds assembly. The Illumina short-read data were mapped to the assembly scaffolds to correct indel errors following the methods we have previously described for the Arlee line genome assembly (Gao et al. 2021).

2.2.3.2 | Hi-C. Larger scaffolds were generated from the Bionano hybrid scaffolds assembly using Hi-C proximity ligation sequencing data. The Hi-C library was prepared from a frozen blood sample and sequenced by a commercial vendor (Phase Genomics, Seattle, WA) to produce 780 million Illumina paired-end reads (2×150 bp) (SRA accessions SRX23202418 and SRX23202419). The scaffolding programme Salsa (Ghurye et al. 2019) was used in conjunction with the Hi-C sequence data to merge contigs and scaffolds from the Bionano hybrid assembly and to break chimeric scaffolds following the same methods as we have previously described for the assemblies of the Arlee and the Swanson genomes (Ali et al. 2025; Gao et al. 2021).

2.2.4 | Assembly of Chromosome Sequences

In order to produce chromosomal sequences from the Hi-C assembly scaffolds, linkage data from a SNP genetic map were used as previously described for the Omyk_1.0 (Pearse et al. 2019) and OmykA_1.1 (Gao et al. 2021) chromosomal sequences assemblies. Briefly, the flanking sequences of each SNP marker were mapped to the scaffolds using the programme NovoAlign (http://www.novocraft.com/products/novoalign/), which was then used to order, orient and concatenate the scaffolds into 32 chromosomes using the genetic linkage data. Using the linkage information, eight large Hi-C scaffolds were broken into 19 smaller scaffolds. The linkage map that was used to anchor the scaffolds to chromosomes and guide the ordering of scaffolds within the chromosomes was generated from a robust dataset based of two large aquaculture population from the US and Norway (Pearse et al. 2019). It is important to acknowledge that the use of linkage data that was generated from different strains has some limitations as it may introduce errors in breaking, ordering and orienting the de novo scaffolds. However, it is important to also note that the Hi-C scaffolds that were broken due to disagreement with the linkage map information were primarily due to being aligned to two or three chromosomes with obvious breaking points within the scaffold, indicating higher likelihood of errors in the bridging of contigs into scaffolds by the optical map or by the Hi-C scaffolding. In general, the closer the scaffolds are to complete chromosome length as they were with this WR assembly, the smaller the risk of introducing errors due to strain differences with the linkage map.

2.3 | KC Line Genome Assembly

The assembly workflow for the KC line was composed of these steps: Genome sequencing → Contig assembly → Hi-C scaffolding → Chromosome-scale ordering/orientation using linkage map information. Software default parameters were used unless otherwise noted.

2.3.1 | Genome Sequencing

DNA was extracted from a fresh blood sample of a single rainbow trout male fish from the KC line as described above for the WR line (NCBI Bio-sample Accession SAMN37796906). HiFi reads were produced using the CCS mode on the PacBio Sequel II system (GBRU Stoneville, MS). A total of 84.2 Gb sequence data (35× genome coverage) was generated from 588,679 HiFi reads (NCBI SRA Accessions SRX22156623, SRX22156624, SRX22156625, SRX22156626). In addition, Illumina whole-genome sequencing data were generated for quality control of the genome assembly and to assess assembly completeness. A total of 649 million paired-end Illumina reads (2×150 bp; 82× coverage) were generated by a third party service provider (Admera Health, South Plainfield, NJ) using the Kapa Hyper Prep PCR free kit and were submitted to the NCBI SRA database (SRX22157130).

2.3.2 | Assembly of Contigs

For the KC contigs assembly we needed to generate two haplotypes since this genome was from a heterozygous diploid fish. Therefore, for the KC contigs assembly we used the Hifiasm assembler with the HiFi reads from the CCS mode, which are shorter than the CLR mode reads but contain a much lower error rate to enable haplotypes separation in heterozygous regions of

the genome. The genome assembler Hifiasm (Cheng et al. 2021) was used to generate two genome-wide haplotypes (primary and alternative) for the outbred diploid KC male rainbow trout using input data from the PacBio HiFi long read sequences and the HiC Illumina short reads.

2.3.3 | Assembly of Scaffolds

The Hi-C library was prepared from a frozen blood sample (Phase Genomics, Seattle, WA) to produce 762 million Illumina paired-end reads (2 × 150 bp) (SRA accessions SRX23159327 and SRX23159328). Hi-C scaffolding was prepared from the primary haplotype contigs assembly that was a bit larger than the alternative haplotype and was found to include the sdY gene and thus the Y chromosome. The contigs were checked with the NCBI fcs_adaptor programme and 17 contigs were found to have the Biosciences Blunt Adapters. These adapters were trimmed at the end of nine contigs, and in eight contigs, they were removed by breaking the contig before and after the adapter. Here again, the programme Salsa (Ghurye et al. 2019) was used to scaffold contigs, as we have previously described for the assemblies of the Arlee and the Swanson genomes (Ali et al. 2025; Gao et al. 2021).

2.3.4 | Assembly of Chromosome Sequences

As described above for the WR assembly, the genetic map linkage information was used to anchor, order, orient and concatenate the scaffolds from the KC Hi-C assembly onto chromosome sequences. In the process, 90 Hi-C scaffolds that were in large disagreement with the linkage map information were broken into 402 smaller scaffolds.

2.4 | K-Mer Completeness and Multiplicity Assessment

Assembly quality was assessed using k-mer-based approach implemented in the Merqury (v1.3) software package (Rhie et al. 2020). For this analysis, illumina paired-end read sets generated from the WR line (SRX21517362) and the KC line (SRX22157130) were used with corresponding genome assemblies (GCA_029834435.1_USDA_OmykWR_1.0 and GCA_034753235.1_USDA_OmykKC_1.0, respectively).

The k-mers ($k = 21$) were counted from the Illumina reads using Meryl (Miller et al. 2008). Next, Merqury was run using the read-based k-mer database and the genome assembly file in FASTA format. Assembly quality metrics including consensus quality value (QV), k-mer completeness and copy number spectrum were obtained. In particular, the spectra-cn plots were used to assess the completeness and heterozygosity in the assembly.

2.5 | Annotation of WR, KC and Swanson Assemblies

The WR, KC and new Swanson genome assemblies were annotated using Liftoff (v1.6.3) (Shumate and Salzberg 2021). The Arlee genome assembly (USDA_OmykA_1.1) (Gao et al. 2021), and its corresponding gene annotation (GFF) file was used as reference to lift genes from. Annotation parameters used were -s 0.9 (for minimum sequence identity) and -a 0.9 (for minimum alignment coverage). The -copies option was used to search for extra copies of genes that are not annotated in the reference after the initial lift over. The -chroms option was used to lift-over the annotations chromosome by chromosome and then genes that did not map were aligned to the whole genome. A list of unplaced sequences was provided with the -unplaced option to map them to the target assembly after genes on the main chromosomes have been mapped. Unmapped features were then saved in a separate file with (-u). The -polish option was used to maximize alignments and improve annotation quality.

2.6 | Whole-Genome Alignment and SV Detection Using MUMmer and SyRI

To identify SVs (≥ 50 bp in length) and complex structural rearrangements among rainbow trout strains, we used MUMmer (v3.23) (Kurtz et al. 2004) and SyRI (v1.6.3) (Goel et al. 2019) for whole-genome alignment and variant calling. NUCmer (NUCleotide MUMmer; v3.1) was applied for pairwise alignments of genomes using the --maxmatch option under the following: minimum cluster length (-c) of 500 bp, break length (-b) of 500 bp and minimum match length (-l) of 100 bp. Settings were selected based on prior studies (Dhakal et al. 2024; Jain et al. 2018), and all values are reported in DNA base pairs (bp). Arlee was used as the primary reference genome in this study, except for three cases where the WR genome served as a reference to identify wild-specific syntenic blocks, WR un-aligned/unique regions and Arlee-unique variants. SVs and complex structural rearrangements were filtered based on size thresholds using a custom Python script. Insertions, deletions and tandem repeats were retained if ≥ 50 bp (Alkan et al. 2011), whereas other complex structural rearrangements were filtered using a more stringent threshold of 100 bp to reduce false positive calls (Stephens et al. 2018; Quinlan and Hall 2012; Sedlazeck et al. 2018). For downstream analysis using DNAdiff (v1.3), delta files were filtered using delta-filter with the following identity (-i 90) and minimum alignment length (-l 100) of 90% and 100 bp, respectively. Filtered delta files were used as input to dnadiff for alignment statistics and genomic differences. For calling genomic variants, delta files were subsequently filtered with the -m option to retain 1-to-1 alignments, and coordinates were retrieved using show-coords with the -THrd option. SyRI was then run with filtered delta and coordinate files using the reference and query FASTA sequences for the detection and annotation of syntenic regions, structural rearrangements and local sequence variants. All SVs included in downstream analyses were at least ≥ 50 bp in length. SyRI analysis commands were executed with at least two computational threads (-nc 2) and the logging configured for debugging (-log DEBUG). All SyRI analyses were conducted within the Anaconda environment for reproducibility and for managing dependencies.

2.7 | Long-Read Mapping–Based SV Validation

SVs were independently called from long-read alignments using SVIM (Heller and Vingron 2019) and Sniffles2 (Smolka

et al. 2024), followed by merging with SURVIVOR, requiring support from both callers (SUPP = 2). We then intersected mapping-based SV calls with assembly-based SVs identified by SyRI, applying stringent reciprocal overlap criteria (≥ 50%) to assess support. This validation approach allowed us to: (1) provide orthogonal evidence for assembly-derived SVs and (2) quantify support rates specifically for small-sized SVs (50–500 bp).

2.8 | SV Annotation and Filtering

For each line alignment with the Arlee or WR as the reference, chromosome-level VCF files that were generated by SyRI were compressed, indexed and then concatenated with bcftools concat (v1.3.1). To functionally annotate SVs, we used an in-house Python pipeline to simultaneously annotate multiple VCF files using SnpEff (v5.2). The script went through a specified directory for indexed VCF files. For each input VCF, the pipeline produced an annotated VCF as well as complementary HTML and CSV summary reports. Annotations were performed using a custom SnpEff database derived from the *O. mykiss* Arlee reference genome assembly (GCF_013265735.2) (Gao et al. 2021) and corresponding GTF annotation file. SnpEff annotated the variants as MODIFIER (variants in intergenic, intronic or regulatory regions with no predicted functional impact) or functional with LOW (typically synonymous or minor effects), MODERATE (non-disruptive coding changes such as missense or in-frame variants) or HIGH (predicted disruptive effects including frameshifts, stop-gained, splice-site alterations and transcript ablation) effects. To identify SVs with potential functional consequences, we used SnpSift (v5.2) to filter high-impact insertions, deletions, inversions and duplications across the three query genomes. The six VCF files with filtered SVs for each line based on alignment to the Arlee and WR genome as the reference were deposited in Dryad (https://doi.org/10.5061/dryad.1rn8pk17m) and will be publicly released after peer review and acceptance of the manuscript.

2.9 | Gene Enrichment Analysis

To investigate the biological functions associated with genes affected by SVs, gene enrichment analysis was performed using ShinyGO (v0.77) (https://bioinformatics.sdstate.edu/go77/) (Ge et al. 2020). Lists of unique gene identifiers affected by a specific type or category of SVs were uploaded to ShinyGO, setting *O. mykiss* as the target species. Default parameters were used, including a false discovery rate (FDR) cutoff of 0.05 for assessing statistical significance. The software generated interactive plots and tables of enriched pathways and terms, utilized to understand the potential biological impacts/consequences of SVs of interest.

2.10 | Genomic Regions Associated With *O. mykiss* Life History Variants

Several analyses were conducted to examine the degree to which structural variation may be identified in genomic regions that were previously found to be associated with life history variations in *O. mykiss*. First, we made alignments of trait-associated genomic regions across the four assemblies, which reside in chromosomes Omy28 (migration-timing) and the q-arm of Omy25 (age-at-maturity and repeat spawning phenology), using Mummer (v4) (nucmer --nosimplify -c 100 -g 1000) to examine the presence of structural differences in these regions. Second, for the Omy25 region, we selected 16 stocks (populations) for which low coverage whole genomic data were available and which represent the diversity of steelhead stocks within the Columbia River Basin (Willis et al. 2024). We mapped this sequence data to each of the four assemblies using *bwa* v0.7.18 (Li 2013), filtered discordant and improperly paired alignments using *samtools* (Danecek et al. 2021; Li et al. 2009), and examined the mean coverage across the adjacent regions in Omy25q associated with age-at-maturity and repeat spawning phenology, respectively. Within these regions, we calculated in R (R Corp.), from mpileup files of each stock produced by *samtools*, the mean coverage in 500 bp sequential windows, and then normalized this coverage to the mean coverage for each stock across the entire region.

3 | Results and Discussion

3.1 | De Novo Genome Assemblies

The de novo genome assemblies analyzed in this study were derived from four genetic lines of *O. mykiss* (rainbow trout). The four lines exhibited variation in chromosome numbers, geographic origin, domestication and life history (Table 1), making them valuable resources for comparative genomics and the study of complex traits in *O. mykiss*. For example, variation in chromosome number (ranging from 58 to 64) highlights the diversity of karyotypes that was captured by the selection of these four lines as the foundation toward developing a pan-genome reference panel for rainbow trout.

Among these four assemblies, two were newly generated and reported here for the first time (WR and KC lines). The final

TABLE 1 | Genetic lines used for the de novo genome assemblies.

Line name	2*N*[a]	Sex	Geographic origin	Life history type	Origin
Arlee	64	YY Male	Northern California	Resident	Domesticated
Swanson	58	YY Male	Kenai Peninsula, Alaska	Resident	Semi-domesticated[b]
WR	64	YY Male	Central California Coast	Landlocked Steelhead	Wild
KC	60	XY Male	Columbia Basin, Snake River	Resident	Wild

[a]A diploid chromosome number (2*N*).
[b]Fish have undergone partial domestication after one generation in the hatchery environment.

chromosome-level genome assembly for the WR line (USDA_OmykWR_1.0; GCA_029834435.1) contains 102 scaffolds that were placed on 32 chromosomes and 1036 unplaced smaller scaffolds, with a total sequence length of 2.3 Gb and N50 of 50.8 Mb. The karyotype of $N = 32$ we identified for the WR male genome is consistent with the previously reported karyotype for the WR female doubled haploid line (Miller et al. 2012) and the coastal California origin of this genetic line. The final chromosome-level genome assembly for the KC line (USDA_OmykKC_1.0; GCA_034753235.1) contains 1079 scaffolds that were placed on 30 chromosomes and 2042 unplaced smaller scaffolds, with a total sequence length of 2.3 Gb and N50 of 3.4 Mb. There is no specific information about the expected karyotype of the KC population. Some work that was done on other populations of the inland rainbow trout lineage (*O. m. gairdneri*) suggested a haploid karyotype of $N = 29$ (Thorgaard 1983). However, based on the number of major scaffolds and their unique mapping pattern to genetic linkage groups, and their sequence alignments with chromosome arms from the Arlee and Swanson genomes (Ali et al. 2025; Gao et al. 2021), we concluded that the haplotype number of chromosomes for the individual fish that was sampled for this genome assembly is $N = 30$. To that end, it is worth noting here that there might have been an inadvertent maternal contribution from an aquaculture strain of rainbow trout possibly from a coastal origin in the cross that was used to establish the KC line at Washington State University. Hence, we advise caution in drawing definitive conclusions regarding the karyotype and the haploid chromosome number of native rainbow trout from the inland lineage.

The assembly statistics of the four rainbow trout genomes revealed notable similarities and differences in assembly quality and continuity (Table 2). All four assemblies have the same genome size (2.3 Gb), ungapped length (2.3 Gb), GC content (43.5%) and chromosome level sequences that were guided using the robust genetic linkage information. This indicates consistency in the overall structure and completeness of the four assemblies. However, the KC genome assembly is the most fragmented, likely due to the lower genome coverage (35.0×), the use of a heterozygous individual fish that is not a doubled haploid, and the application of shorter PacBio HiFi reads instead of high-coverage PacBio CLR sequencing to generate the initial contigs level assembly. The KC assembly has significantly more gaps (1049), scaffolds (3121), and contigs (6120), as well as the lowest scaffold N50 (3.4 Mb) and contig N50 (1.1 Mb) values. The Swanson and WR are the most contiguous assemblies. Both assemblies have high scaffold N50 values (52.4 and 50.8 Mb, respectively) and low L50 values (16), indicating fewer and larger scaffolds. The WR assembly benefited from the high homozygosity in this doubled haploid line and was enhanced by our ability to use the PacBio CLR long reads for the contigs assembly, supplemented by multi-platform scaffolding. The Arlee assembly also exhibited high continuity, with the fewest scaffolds (938) and contigs (1228), as well as the highest contig N50 (15.6 Mb).

The near completeness of the genome assemblies was evident from the BUSCO (Benchmarking Universal Single-Copy Orthologs) results. The Complete BUSCO scores were 99.9% and 99.8% for the Arlee and WR assemblies, respectively, and 98.9% for both the Swanson and KC assemblies. A high-quality assembly typically shows > 95% 'Complete' BUSCO score. To further assess the completeness of the two new assemblies from the WR and KC lines, we used k-mer analysis with Merqury. In the k-mer multiplicity plot for the WR assembly (Figure 1A), a single dominant k-mer peak centred at ~55–60× coverage was observed, with minimal read-only k-mers, indicating strong concordance between the Illumina reads data and the assembly. The small proportion of missing k-mers is consistent with the

TABLE 2 | Genome assemblies' statistics.

Feature	OmykA_1.1 (GCA_013265735.3) (Arlee)	Omyk_2.0 (GCA_025558465.1) (Swanson)	OmykWR_1.0 (GCA_029834435.1) (WR)	OmykKC_1.0 (GCA_034753235.1) (KC)
Number of chromosomes	32	29	32	30
Total length	2.3 Gb	2.3 Gb	2.3 Gb	2.3 Gb
Number of scaffolds	938	1278	1138	3121
Gaps in scaffolds	486	460	382	2999
Gaps between scaffolds	196	70	109	1049
Scaffold N50	39.2 Mb	52.4 Mb	50.8 Mb	3.4 Mb
Scaffold L50	22	16	16	182
Number of contigs	1228	1738	1520	6120
Contig N50	15.6 Mb	10.7 Mb	12.2 Mb	1.1 Mb
Contig L50	41	57	55	541
Genome coverage[a]	111.6×	106.0×	94.3×	35.0×
BUSCO score[b]	C: 99.9% (S: 54.9%; D: 44.9%)	C: 98.9% (S: 55.7%; D: 43.2%)	C: 99.8% (S: 54.7%; D: 45.1%)	C: 98.9% (S: 55.8%; D: 43.1%)

[a]Sequencing depth with PacBio long reads or HiFi reads.
[b]Benchmarking Universal Single-Copy Orthologs (BUSCO) using version 6.0.0 and lineage actinopterygii_odb10 (2024-01-08). C: Percent of complete protein-coding genes detected. S: Percent in single copy. D: Percent of duplicated genes.

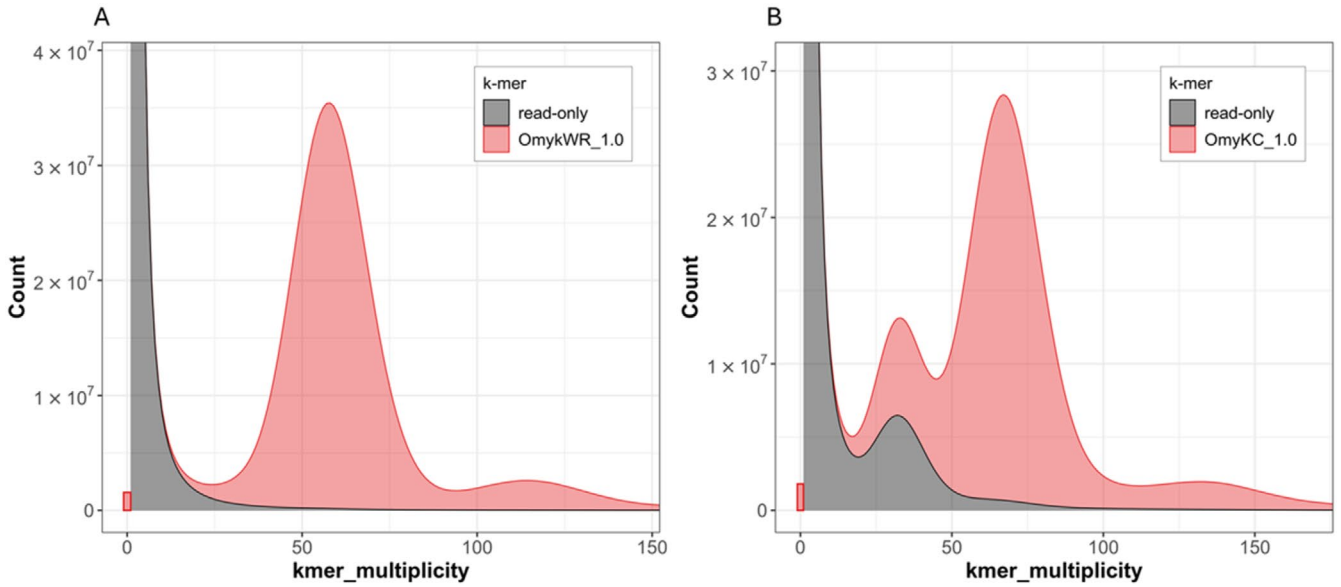

FIGURE 1 | K-mer multiplicity plots used to assess the quality of the new de novo genome assemblies of (A) the Whale Rock (WR) rainbow trout line and (B) the Keithley Creek (KC) line.

high k-mer completeness (98.36%) and high QVof 43.88. The KC assembly exhibited a dominant k-mer peak at ~65× coverage and a secondary shoulder around ~30–35× (Figure 1B), indicating the presence of heterozygous k-mers. This pattern is consistent with our prior mapping analysis that suggested elevated heterozygosity relative to the other three established doubled haploid trout lines. As we described in the methods, an inadvertent partial maternal contribution in the process aimed to induce androgenesis is the likely source of the residual heterozygosity or admixture in the KC line. The lower k-mer completeness score of 86.31% for the KC line is also consistent with its lower assembly statistics relative to the WR, Swanson and Arlee genome assemblies. However, the high consensus k-mer QV score of 44.13 indicates that the base-level accuracy of the KC assembly remains very good despite the biological complexity. In addition, we used the TeloExplorer module of the quarTeT programme (Lin, Ye, et al. 2023) to detect telomeres on chromosome ends in both assemblies. In the WR assembly the quarTeT analysis detected that five chromosomes had both telomeres, 20 had one and seven had none. In the KC assembly it detected that seven chromosomes had both telomeres, 18 had one and five had none.

3.2 | Intronic and Intergenic SVs Dominate the Landscape Across the Rainbow Trout Lines

To identify SVs between the lines we aligned the genome assemblies of KC, Swanson and WR to the Arlee reference genome (Table S1). The alignment statistics revealed notable differences in the variants landscape between the lines. All three lines showed high sequence conservation with the reference genome. The average identity values range from 97% to 98%. However, SV analysis revealed a wide range of genomic rearrangements. Overall, the Swanson genome assembly exhibited the highest number of indels, breakpoints and relocations, as well as frequent translocations and insertions relative to the Arlee genome. In contrast, KC and WR are more conserved relative to the Arlee genome, exhibiting fewer relocations and genomic variations such as inversions. These differences suggest that the Swanson line may represent a more genetically distinct population, but it is important to note here that the difference in number of variants between Swanson and the other three genomes is inflated by the fission of Omy04, Omy14 and Omy25, as well as the large double inversion on Omy05, for which Swanson is the only genome with the Rearranged (R) haplotype, while the other three genomes exhibit the Ancestral (A) haplotype.

SyRI annotated a range of genomic features, including structural rearrangements (inversions, translocations and duplications) and local sequence variation (insertions and deletions) (Figure 2a), as well as syntenic (co-linear) regions and unaligned (NOTAL) regions. Notably, deletions and insertions were the most abundant variant types across all comparisons. The counts ranged from 49,533 to 55,334 for deletions and 48,310 to 53,807 for insertions. Such high counts reflect wide sequence variation and may imply mobile element activity variation between the three query strains and the reference genome, as has been shown in human genomics (Han et al. 2008; Paterson et al. 2015). However, further sequence characterization is needed to confirm the presence of transposable element-related sequences in the regions associated with high counts of insertions and deletions. Additionally, a high frequency of duplications and unaligned regions was detected. This might be an indication of potential lineage-specific expansions or reference collapse (Hartasanchez et al. 2018; Wang et al. 2012), but since the genome assemblies are not yet complete we also have to caution here that those unaligned regions are not classified as SVs by SyRI and may simply reflect assembly completeness limitations rather than true biological lineage-specific expansions or reference collapse. A complete list of SV counts in the three rainbow trout lines is provided in Table S2. As a positive control to validate the SV detection pipeline, we aligned the X chromosome (Omy29) from the recently assembled hybrid-derived haploid genome of a rainbow trout from the domesticated Hayspur hatchery strain in Idaho (GCA_049308825.1) (Flores et al. 2025) to the Y chromosome of the Arlee reference genome.

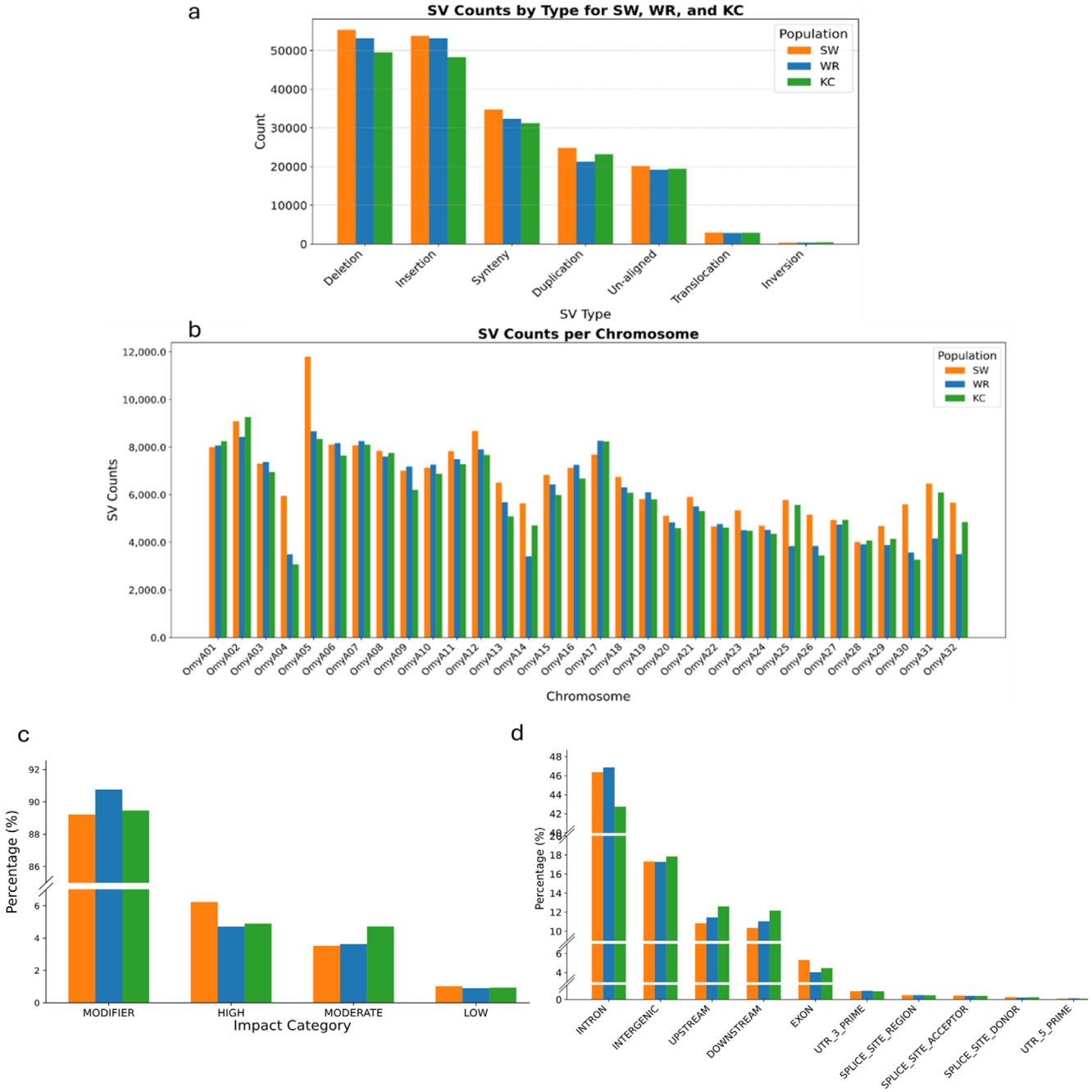

FIGURE 2 | Characterization of SVs identified in the Swanson, WR and KC lines in reference to the Arlee genome assembly. (a) Distribution of filtered SV types annotated by SyRI in the reference-query alignments reveals a predominance of insertions and deletions. (b) Total SVs counts per chromosome. (c) Annotated SV frequencies in each effect class, which were dominated by the MODIFIER class. (d) Genomic features affected by SVs, indicating that the majority of variants are in intronic and intergenic regions with a smaller fraction impacting coding and regulatory regions.

The MUMmer/SyRI-based approach successfully identified a 169,949 bp deletion encompassing the sdY gene, which is known to be absent in XX rainbow trout females (Yano et al. 2013).

Functional annotation of the SVs revealed consistent patterns in variant type, impact and genomic distribution among the three non-reference rainbow trout lines—Swanson, WR and KC (Figure 2b–d). The three lines showed a consistent pattern in variant distribution across chromosomes. Population-specific differences in variant counts were observed on Omy05 and chromosomes that underwent fission (Omy04, Omy14 and Omy25). Swanson and WR tend to show slightly elevated variant counts on some chromosomes compared to KC (Figure 2b). Most of the variants were from the MODIFIER impact category (> 89%) and localized in non-coding regions without predicted functional consequence. HIGH-impact variants occurred in a minor fraction (~4.7%–6.2%) and LOW- and MODERATE-impact variants were even more infrequent (~0.9%–3.5%) (Figure 2c). Genomic feature analysis identified that SVs largely affected intronic (~42.8%–46.9%) and then intergenic regions (~17%), and fewer being in exons, untranslated regions (UTRs) or splice sites (Figure 2d). Collectively, these findings

demonstrate a highly conserved SV landscape across the three lines, with most variants likely to have low functional effects but with a subset potentially having regulatory or coding effects of interest. These results are similar to those of Liu et al. (2021) who conducted SVs discovery and characterization in rainbow trout using short reads whole genome sequencing and found that 95% of the SVs in their study were classified in the MODIFIER category and genomic locations identified predominantly as intronic (45.5%) or intergenic (25.3%) based on alignments to the previous reference genome assembly (Omyk_1.0) (Pearse et al. 2019). Furthermore, in a previous study we also showed that intron retention is the most frequent alternative splicing event in rainbow trout (Ali et al. 2021), which suggests that genomic variation is more likely to influence gene regulation or splicing rather than disrupting or modifying coding regions directly in this species.

Further analysis of high-impact SVs revealed differences in SV characteristics and burden among the three rainbow trout lines when aligned to the Arlee reference genome (Table 3). A total of 1300 high-impact SVs were identified in KC, compared to 1302 in WR and 1221 in Swanson. Across all three lines, duplications and deletions were the predominant variant types, accounting for the majority of high-impact SVs (approximately 34%–46%), followed by insertions (14.6%–16.4%), while inversions were relatively rare (2.2%–5.2%). As expected, Swanson displayed the largest SV (35,109,550 bp inversion) located on Omy05. Chromosomal hotspots for variants accumulation differed among populations: Omy17 in KC (7.8%) and WR (6.8%) and Omy13 in Swanson (5.81%). Functional impact analysis indicated that frameshift variants, duplications, gene fusions and stop-gained mutations with duplications were among the most frequent high-impact effect types.

3.3 | Long-Reads Mapping–Based SVs Validation

For the WR assembly, a total of 130,711 assembly-based SVs (≥ 50 bp) were identified (Figure 2a). Among these, 129,068 (98.7%) were supported by simple overlap with long-reads mapping-based SV calls and 71,725 (54.87%) were supported by mapping-based SV calls with a more stringent reciprocal overlap criterion (≥ 50%). Similarly, for the small-size SVs (50–500 bp), 98,173 variants were identified, of which 97,024 (98.8%) were validated by long-read mapping using the simple overlap criterion, and 55,584 (56.62%) were validated by long-read mapping using the more stringent reciprocal overlap criterion.

For the KC assembly, a total of 121,508 assembly-based SVs (≥ 50 bp) were identified. Of these, 93,671 (77.1%) or 64,230 (52.86%) were supported by long-reads mapping–based SV calls using a simple overlap or a stringent reciprocal overlap criterion (≥ 50%) between the assembly-based SVs and the mapping-based SV, respectively. Similarly, for the smaller SVs (50–500 bp), 89,832 variants were identified, of which 69,169 (77.0%) were supported by simple long-reads mapping evidence and 48,731 (54.25%) were supported by long-reads mapping evidence using the more stringent criterion.

These results are consistent with the overall assembly quality metrics: WR shows higher validation rates, which aligns with its higher completeness and lower residual heterozygosity compared

TABLE 3 | Summary of high-impact SVs identified in KC, WR and Swanson compared to the Arlee reference genome.

Category	Swanson	WR	KC
Total high-impact SVs	1221	1302	1300
SV type distributions			
Duplications	514 (42.1%)	549 (42.2%)	599 (46.1%)
Inversions	32 (2.6%)	28 (2.2%)	68 (5.2%)
Deletions	477 (39.1%)	512 (39.3%)	443 (34.1%)
Insertions	198 (16.2%)	213 (16.4%)	190 (14.6%)
Chromosome hotspots			
Highest SV burden	Omy13 (5.81%)	Omy17 (6.8%)	Omy17 (7.8%)
Top effect types			
Frameshift variants	181	205	201
Duplications	159	148	181
Frameshift variants and stop-gained	125	124	122
Gene fusions	118	147	124
Stop-gained and duplications	113	136	133
Largest high-impact variants[a]			
Inversions	Omy05: 52,392,628 (35,109,550 bp)	Omy17: 36,299,758 (7,148,210 bp)	Omy21: 3,822,308 (7,227,944 bp)

[a]Variants reached a size of up to 35 Mb and span numerous genes.

to KC. Validation rates were also consistent between size classes for each assembly, indicating that the assembly-based detection was not disproportionately biased toward larger variants. The concordance between assembly-based and mapping-based approaches provides strong support for the robustness of SV identification in this study, but also indicates a relationship between the quality of the assembly and the reliability of the SV calls.

Regarding the lower validation rates that were calculated using a stringent reciprocal overlap criterion (≥50%) between assembly-based SVs and mapping-based SV calls. This threshold is quite conservative, particularly given breakpoint uncertainty and representation differences between assembly-based and mapping-based approaches. Small shifts in breakpoints or differences in how complex variants are represented can cause true positives to fail the ≥50% reciprocal overlap requirement. Two recent studies that compared methods for SV detections in other species have shown similar results of concordance between the assembly alignments-based and the long-reads mapping-based approach with either the strict reciprocal overlap criteria or when using the more permissive distance-based matching (Liu et al. 2024; Lin, Jia, et al. 2023).

3.4 | Comparative Structural Rearrangement and Sequence Variant Profiles Between the Rainbow Trout Lines

To further investigate the biological relevance of structural rearrangements and sequence variants among strains, we grouped variants into their presence or absence within domesticated and wild lineages, as well as between the WR line that represents a landlocked steelhead population and the other three lines that represent freshwater resident strains of origin. *Syntenic* and un-aligned regions, which are not classified as SVs, were included to provide comparative genomic context. This grouping enabled a preliminary identification of genomic variants for strains that differ by (1) domestication history and (2) life history background. Four general groups (Table 4) were formed:

3.4.1 | Genomic Regions Present in Arlee and Swanson but Not in the Lines Derived From Wild Strains (WR and KC)

This group had a combined total of 404 syntenic blocks (non-variant, co-linear regions) shared by Arlee and Swanson but not by WR and KC (Table S3). These regions had a maximum length of 175,589 bp (Average size 21,205 bp). Notably, chromosome Omy13 had the highest proportion of these regions, at 27.5%, which reflects a region of genomic divergence among strains.

A total of 289 syntenic regions (~72%) shared only by Arlee and Swanson overlapped 339 genes, among which the coding sequences (CDS) of 245 genes were completely encompassed by 195 of these syntenic regions. Functional annotation of the 245 genes revealed several enriched pathways (Table S4). Interestingly, enrichment for nucleotide binding functions, including GTP binding, ribonucleotide guanyl binding and guanine nucleotide binding, was detected. Guanylate binding proteins (GBPs) are involved in cell proliferation and have been associated with growth traits in cattle (Zhang et al. 2018), which may suggest a potential role in selective breeding practices in captivity. Additionally, variations in G protein-coupled receptors have been linked to domestication traits, including reproduction, development and stress responses (Kleinau et al. 2024). However, further examination with larger sample sizes from domesticated vs. wild strains will be necessary before inferences can be made regarding domestication effects.

3.4.2 | Arlee-Unique Regions Not Found in Other Strains, Including Un-Aligned Regions (NOTAL) and Lineage-Specific Variants

We found 4269 genes annotated in 7298 Arlee-specific un-aligned regions (NOTAL; not classified as SVs) to Swanson, WR or KC (Table S5). Of them, the CDS of 2095 genes were encompassed by the un-aligned regions (> 95% coverage). The

TABLE 4 | Classification of genomic regions and variants by strain of origin or shared life history.

Group	Description	Ref.[a]	Present	Absent	SyRI annotation
1	Shared: Arlee & Swanson, not wild	Arlee	Arlee, Swanson	WR, KC	Syntenic regions[b]
2	Unique Arlee, not in others	Arlee	Arlee	Swanson, WR, KC	Un-aligned regions[b]
	Arlee-only vs. others (using WR ref)	WR	Arlee	Swanson, WR, KC	SVs (Inversions, translocations, insertions, deletions, etc.)
3	Shared wild (WR, KC), not domestic (Swanson, Arlee)	WR	WR, KC	Arlee, Swanson	Syntenic regions
	Wild-only SVs (by Arlee ref)	Arlee	WR, KC	Arlee, Swanson	SVs (Inversions, translocations, insertions, deletions, etc.)
4	WR un-aligned (not in KC, Arlee, Swanson)	WR	WR	KC, Arlee, Swanson	Un-aligned regions
	Unique WR SVs (by Arlee ref)	Arlee	WR	KC, Swanson, Arlee	SVs (Inversions, translocations, insertions, deletions, etc.)

[a]Reference genome assembly used in the analysis.
[b]Syntenic regions represent collinear alignments between genomes, whereas un-aligned regions (NOTAL) represent genomic segments lacking detectable alignment. Neither category represents SVs as defined by SyRI; however, both were retained to provide comparative genomic context and to support interpretation of genome-wide alignment structure.

highest proportion of the Arlee-unique blocks was on chromosome Omy13 (7.0%). Notably, two genes located within un-aligned regions—LOC118940322 (*cd22*) on Omy17 and LOC118943595 (*taar6*) on Omy22—matched candidate gene presence/absence variations (PAVs) previously identified in American-selected rainbow trout (Bao et al. 2025). The *taar6* (Trace Amine-Associated Receptor 6) is related to the G protein-coupled receptor family that was found enriched in group 1. However, gene set enrichment analysis of all Arlee-specific genes within un-aligned regions (Table S6) identified the significant overrepresentation of the ECM-receptor interaction pathway (FDR = 0.0417; Fold Enrichment = 6.26), which contains genes such as *itgb3b*, *sv2a* and *lamc3*. The ECM-receptor interaction pathway has been linked to reproductive traits in chickens, which affect the efficiency of spermatogonial stem cell development and quality of meat (Hu et al. 2022; San et al. 2021). Efficiency in reproduction and meat quality are preferred traits in domesticated animals, and hence ECM pathways may play a role in these traits.

Additionally, we detected 169,030 distinct Arlee line SVs (inversions, translocations, duplications, insertions, deletions and other complex SVs) using the WR as the reference for genome alignment. The average length of these SVs is approximately 1691 bp, with the largest variant reaching 4,089,394 bp. The CDS of 41 genes associated with fatty acid synthesis, energy metabolism and enzymatic activity were fully encompassed within the inserted loci. Similarly, the CDS of 10 other genes were completely overlapping deleted loci in Arlee. Among them, the Semaphorin receptor, *plxna1*, which regulates axon guidance, perhaps due to less need for complex environmental navigation. No enriched pathways were identified for genes in which < 95% of their CDS were affected by indels.

Several key biological pathways were recurrently enriched across diverse SV categories. This included insertions (Figure 3a), deletions (Figure 3b), duplications, inversions, translocations and other complex SVs. Of note, Focal adhesion, MAPK signalling, Neuroactive ligand–receptor interaction, Endocytosis and Adrenergic signalling pathways were each associated with at least seven different SV types (Table S7). The Focal Adhesion pathway (FDR = 2.36E-41) was significantly enriched, involving 131 genes in cell adhesion and signal transduction. This is consistent with findings in avian models, where elevated activity in the liver of line S fowl under artificial selection for growth was mainly involved in focal adhesion, ECM-receptor interaction, cell adhesion molecules and signal transduction, collectively contributing to enhanced muscle development and lipid metabolism in Guangxi Partridge chickens (Shao et al. 2022). In addition, we noticed enrichment of pathways with potential impacts on energy metabolism in the heart and cardiac function like the Adrenergic signalling in cardiomyocytes (FDR = 3.24E-25). The presence of overlapping genes such as *akt1*, *pik3r1*, *raf1*, *src* and *grb2* across multiple pathways underscores their central regulatory roles. These pathways combined collectively indicate participation in vascular development, cell communication, growth and stress responses (Chen et al. 2010; He et al. 2006; Bers 2008; Beghi et al. 2022; Dale Abel 2021; Xie et al. 2015; Novakov et al. 2015; Wang et al. 2016; Papa et al. 2022). Overall, this comparative analysis of the de novo genome assemblies demonstrates how structural rearrangements and sequence variants, which may have profound impact on phenotypic diversity, could be associated with the processes of local adaptation or artificial selection.

3.4.3 | Genomic Regions and Variants Present in WR and KC, Absent in Arlee and Swanson

We compared syntenic regions (non-variant, co-linear regions) that are shared by the wild strains (KC and WR) but not in the domesticated lines (Arlee and Swanson) (Table S8). There were 62 such regions ranging from 880 to 55,501 bp (median: 6660.5 bp; mean: 8779.8 bp). Chromosome Omy17 had the highest proportion of such wild-specific syntenic regions that accounted for 32.2% of the total. A total of 28 syntenic regions shared only by wild strains overlapped 110 genes, among which the CDS of 17 genes were completely encompassed by 18 of these syntenic regions. Functional annotation revealed several pathways, including Adherens junction, Tight junction, Focal adhesion and Regulation of actin cytoskeleton (Figure 4a).

Other than un-aligned and syntenic regions (non-variant categories), we detected 56,479 SVs (deletions, insertions, inversions, translocations, etc.) shared between wild strains WR and KC but not in the domesticated strains Arlee and Swanson. The average size of these SVs is approximately 2629 bp and the longest form covers 3,652,591 bp, showing a broad variation in size distribution. These wild-restricted variants may reflect the limited diversity represented by only four fish or natural variation that has been lost to domestication or evolution to the wild environment.

Similar to group 2, pathways related to environmental responsiveness, neurosensory processing and stress adaptation (Table S9) were recurrently enriched across diverse SV categories such as insertions (Figure 4b), deletions (Figure 4c), duplications and other complex SVs. These pathways include adrenergic signalling in cardiomyocytes, Calcium signalling pathway, Gap junction, Metabolic pathways, MAPK signalling pathway, Neuroactive ligand-receptor interaction. Notably, MAPK signalling pathway (FDR = 1.14E-19) and Calcium signalling pathway (FDR = 1.53E-16) are highly enriched, both of which are key to transducing extracellular environmental signals into appropriate cellular responses, such as those required for thermal tolerance, osmoregulation and immune defence (Cowan and Storey 2003; Xu et al. 2022; Ólafsson et al. 2021; Haddad 2005; Jaggi 2018). The significant enrichment of the Neuroactive ligand-receptor interaction pathway (FDR = 3.80E-15) and Adrenergic signalling in cardiomyocytes (FDR = 8.63E-12) suggests enhanced neurosensory and cardiovascular regulation (Chottová and Slavíková 2011; Fu and Xiang 2015; Sanchez-Alonso et al. 2023; Franzoso et al. 2022; Bardsley et al. 2018) for predator evasion, foraging and migration in natural populations.

Insertions encompassed 3234 genes, but only the CDS of 66 genes were affected. Similarly, deletions spanned 3606 genes, of which the CDS of 98 genes were impacted. This suggests that noncoding regions are impacted by indels. Interestingly, genes associated with the GnRH signalling pathway, vital for reproductive timing, were enriched in both insertion and deletion categories, suggesting potential evolutionary divergence

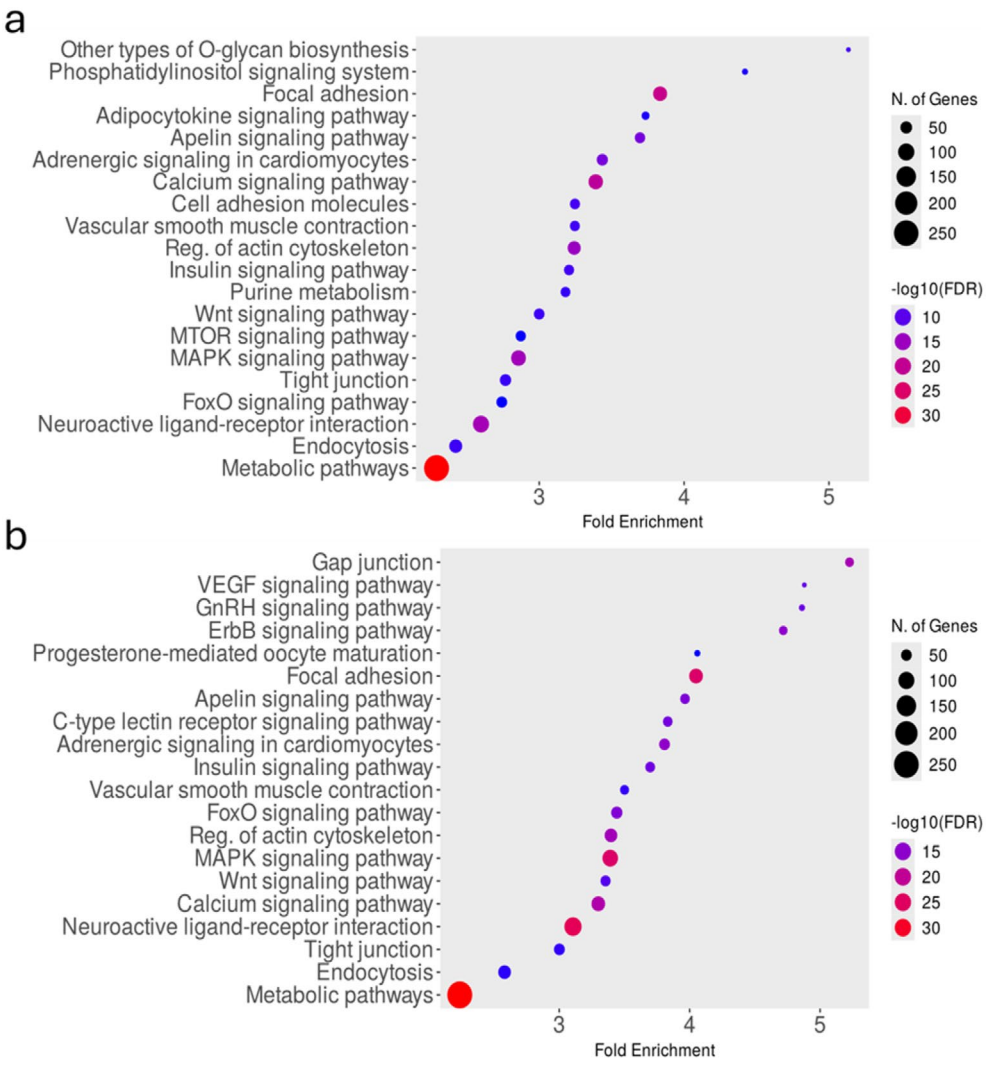

FIGURE 3 | Gene enrichment analyses of genes affected by insertions (a) and deletions (b) unique to the Arlee line.

in developmental and physiological regulation among the lines represented. Collectively, these findings demonstrate how the comparative analysis of de novo genome assemblies may provide insights on differences in the genomic organization between populations of rainbow trout.

3.4.4 | WR–Specific Genomic Regions and Variants

To explore further genomic features potentially linked to life history of the WR strain, we found genomic regions in the WR genome that were un-aligned to the resident strains (KC, Arlee and Swanson) (Table S10). There were 4850 unique un-aligned regions with sizes up to 539,026 bp (median: 2136 bp; mean: 14,763.9 bp). Chromosome 17 has the highest percentage of these regions (6.6%). WR-specific sequences could be genomic components that are associated with life history traits involved in local adaptation, or lineage-specific structural divergence, or merely a reflection of limited representation of genomic variation due to the small sample size used in this study. Further characterization of the 2894 genes located within these genomic intervals revealed significant enrichment in pathways such as phototransduction, cell adhesion molecules and endocytosis (Figure 3d). These findings demonstrate how comparative analysis of de novo genome assemblies may help uncover mechanisms underlying ancestral migratory behaviour or other attributes resulting from the unique ecological pressures to which the steelhead populations are exposed.

Additionally, we identified 126,411 WR line unique SVs (inversions, translocations, duplications, insertions, deletions and other complex SVs). These SVs have a maximum size of 1,224,483 bp, indicating that there are some extremely large structural alterations that potentially would have significant impacts on gene function or regulation.

Several signalling pathways showed significant enrichment among the SVs from the WR line (Figure 4e,f; Table S11). The MAPK signalling pathway, VEGF signalling and Focal adhesion pathways ranked among the top enriched pathways, suggesting important functions in stress response, vascular remodelling and cell adhesion. These processes are essential for physiological transformation during freshwater–saltwater transitions and long-distance migration (Monette and Velotta 2022; Sundell et al. 2003; Yada and Abe 2024; Foddai et al. 2023; Bystriansky and Schulte 2011; Sundell and Sundh 2012). The MAPK signalling pathway (FDR = 3.79E-28), which comprises over 120 genes such as *mapk14*, *raf1*, *akt1* and *nfatc1*, is responsible

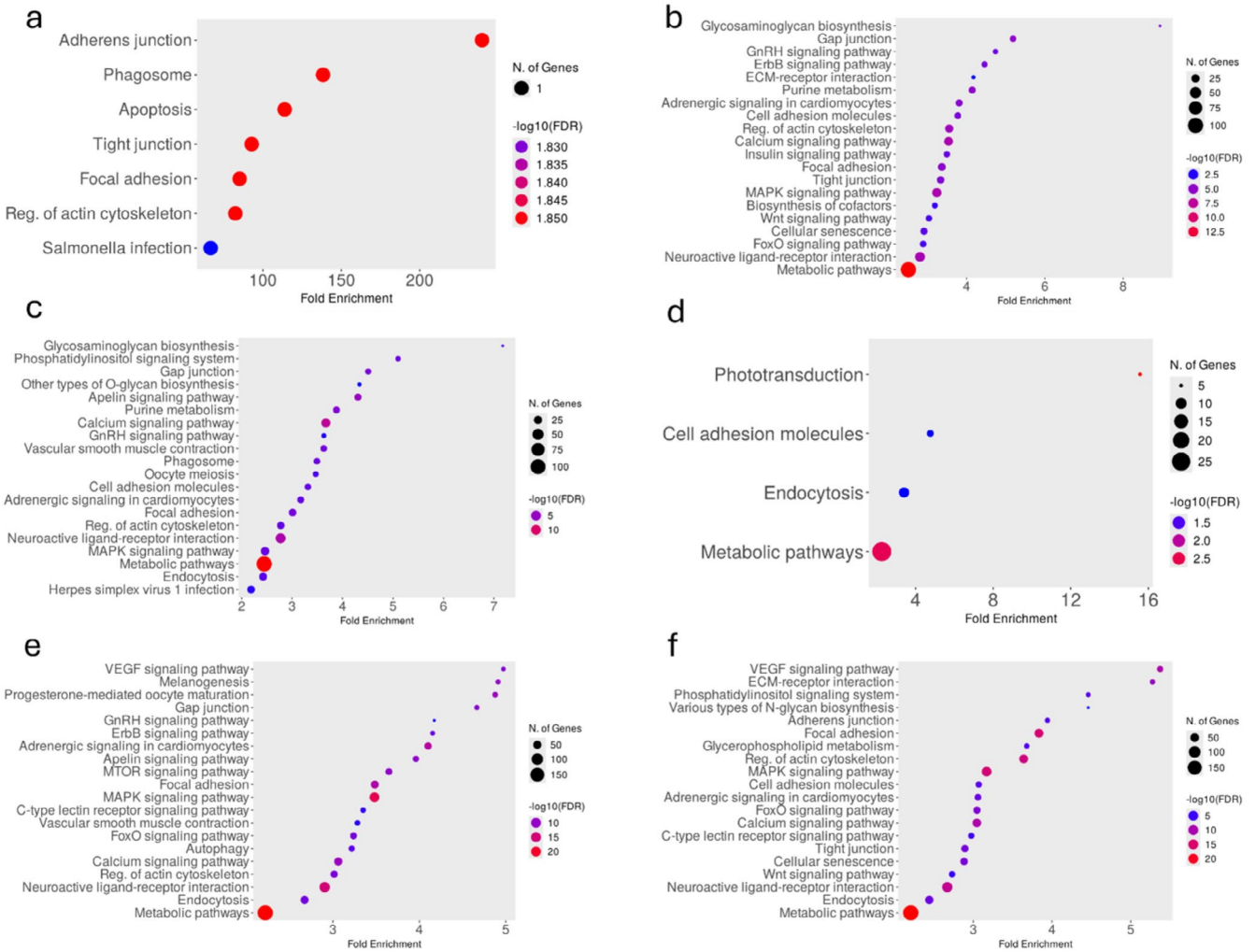

FIGURE 4 | Gene enrichment analyses of genes encompassed or affected by syntenic regions (a), insertions (b), and deletions (c) unique to WR and KC, the lines with 'wild' strains of origin, and absent or not affected in Swanson and Arlee, the lines derived from the more domesticated strains. (d–f) Gene enrichment analyses of genes encompassed or affected by un-aligned regions (d), insertions (e) and deletions (f) unique only to the WR strain.

for cell response to environmental stresses like osmotic stress and temperature changes during migration (Ding et al. 2022; Zhou et al. 2021; Tse et al. 2011; Yang et al. 2022; Johnston and Gillis 2020). Similarly, the VEGF signalling pathway (FDR = 4.07E-20), including *vegfaa*, *src* and *pik3r1*, is crucial for angiogenesis and vascular permeability (Ribatti 2019; Shibuya and Claesson-Welsh 2006) which are important for sustaining elevated metabolic demands and oxygen delivery during periods of high activity (Chen et al. 2025; Chen, Zhou, et al. 2024).

Focal adhesion (FDR = 3.79E-31) was also highly enriched. The *pxn*, *itga6* and *rac1* genes from this pathway can be suspected of involvement in muscle plasticity and tissue remodelling, which are critical for swimming movement (Miller et al. 2006; Klossner et al. 2013; Morash et al. 2014; Macqueen et al. 2010). Regulation of actin cytoskeleton and Apelin signalling pathway suggests additional importance of cardiovascular functionality, and contractility of muscles, and osmoregulatory control (Linlin et al. 2024; Chapman et al. 2014; Corvol 2009). Genes such as *mylk3*, *pdgfra* and *pak6b* may reflect the physiological acclimatization of anadromous steelhead to fluctuating salinity environments. Activation of the ErbB, Insulin and FoxO pathways also indicates neuroendocrine and metabolic regulation of migratory behaviour (Culbert et al. 2024, 2022; Volkoff 2016). These pathways regulate energy homeostasis, cell growth and survival and are in agreement with known migratory traits such as enhanced feeding before seaward migration and increased stress resistance. Finally, Adrenergic signalling and Gap junction pathways (*adrb1*, *adcy5* and *gjd2*) point to cardiac function, communication between neurons and arousal behaviour, each of which is essential for initiating and sustaining migratory behaviour (Ojima and Iwata 2009, 2010; Haraguchi et al. 2015). Collectively, the enriched pathways may represent preliminary evidence for local adaptation or divergent lineages, but we cannot exclude that they may merely be the result of limited representation of genomic variation in this sample.

Together, structural rearrangements and sequence variants observed among these four lines may represent drivers of genomic and phenotypic diversity across highly divergent lineages of rainbow trout. Several enriched pathways such as MAPK signalling, neuroactive ligand-receptor interaction and

calcium signalling are common among all comparison groups. This might occur when different variants affect different genes within the same pathway, particularly in large pathways which involve many genes such as neuroactive ligand-receptor interaction (488 genes) and MAPK signalling pathway (366 genes). Such overlap suggests the versatility/adaptability of conserved pathways to potentially mediate artificial or natural selection, where variant architecture and regulatory divergence may account for lineage-specific adaptation. The findings are in agreement with studies showing that salmonid genomic variation has a tendency to repurpose ancient pathways for novel ecological or domestication-related traits (Lecomte et al. 2024; Bertolotti et al. 2020).

3.5 | Exploring the Overlap Between Identified SVs and QTLs for Muscle Yield

We previously identified a quantitative trait locus (QTL) on Omy17, in which 50-SNP genomic windows accounted for ≥1% of additive genetic variance for fillet yield (Salem et al. 2018). Notably, alignment of the KC and WR genome assemblies to the Arlee reference identified a chromosomal inversion on Omy17, not found in the Swanson assembly (Figure 5a–d). The inversion interval fully overlaps the QTL associated with variation in fillet yield in a domesticated aquaculture population. This suggests a potential functional link between the inversion and phenotypic variation observed. Similarly, previous studies reported inversions on Omy05 and Omy20 associated with genetic divergence between different ecotypes in rainbow trout (Pearse et al. 2019; Arostegui et al. 2019; Campbell et al. 2021). In that context, large inversions can be the source of fixed allelic-based variation with impact on the phenotype, and the gene content of the inversion region can provide insight to the metabolic pathways that may be affected by that genetic variation. Furthermore, a recently published study in Drosophila revealed how an inversion can introduce polymorphism in the expression profile of the genes bound by it with potential impact on the fitness of the fly (Durmaz Mitchell et al. 2025). Gene annotation in the inverted region (Arlee Omy17: 36,299,758–43,447,967) uncovered several

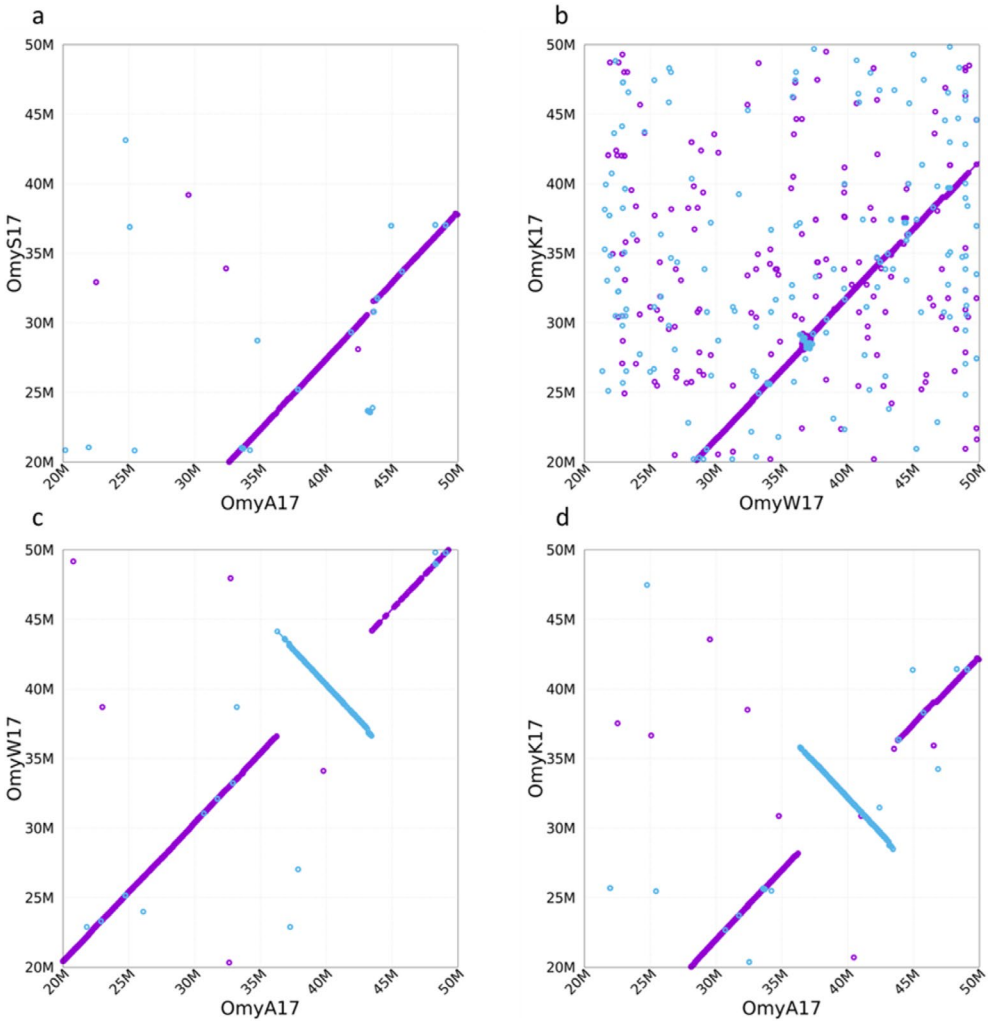

FIGURE 5 | (a) Chromosome Omy17 of the Omyk_2.0 Swanson assembly (OmyS17) aligned to the Arlee assembly (OmyA17) and showing no inversion in the Omy17 QTL region. (b) No inversion between the two lines from wild or natural strains KC (OmyK17) and WR (OmyW17). However, an approximately 7 Mb inversion overlapping with the muscle yield QTL region is found from chromosome sequence alignments between (c) OmyW17 and (d) OmyK17 with the Arlee OmyA17 chromosome. Chromosome identifiers include strain-specific prefixes added to 'Omy' to distinguish assemblies: A = Arlee, S = Swanson, K = Keithley Creek and W = Whale Rock.

biologically relevant candidates (Table S12), including *st3gal2* (glycosylation), *cox6c1* (mitochondrial respiration) *and pck1* (a gluconeogenesis critical enzyme). In addition, the region has genes for cellular stress response pathways (*romo1*), proteasomal degradation (*itch*) and transcriptional regulation (*znfx1* and *cbfa2t2*). The presence of *cyp24a1*, a vitamin D metabolism gene, also indicates the potential metabolic relevance of this inversion.

Notably, the inversion is found in the wild KC and WR strains but not the fully domesticated Arlee and semi-domesticated Swanson lines. These findings suggest the Omy17 inversion might be a candidate SV underpinning physiological or other traits that warrant further investigation with a larger sample size of domesticated and wild rainbow trout populations to establish the significance of the potential association of this SV with domestication or with artificial selection in rainbow trout.

3.6 | Exploring Structural Variation in Loci Associated With Life History Traits

Previous genome-wide association analyses for salmonids used small variants (SNPs and small INDELs) inferred from a single reference genome assembly to identify associations between genomic regions and important life history traits. Here, we looked for structural variation in two regions previously found to be associated with migration timing and age-at-maturity/spawning phenology on chromosomes Omy28 and Omy25q, respectively. Though we did not explicitly test for association of SVs with these traits, we discovered potentially important variation that would go unrecognized and untested in a typical GWAS analysis.

Notable differences were observed from sequence alignments of the loci on chromosomes Omy28 and Omy25q (Figure S1). In

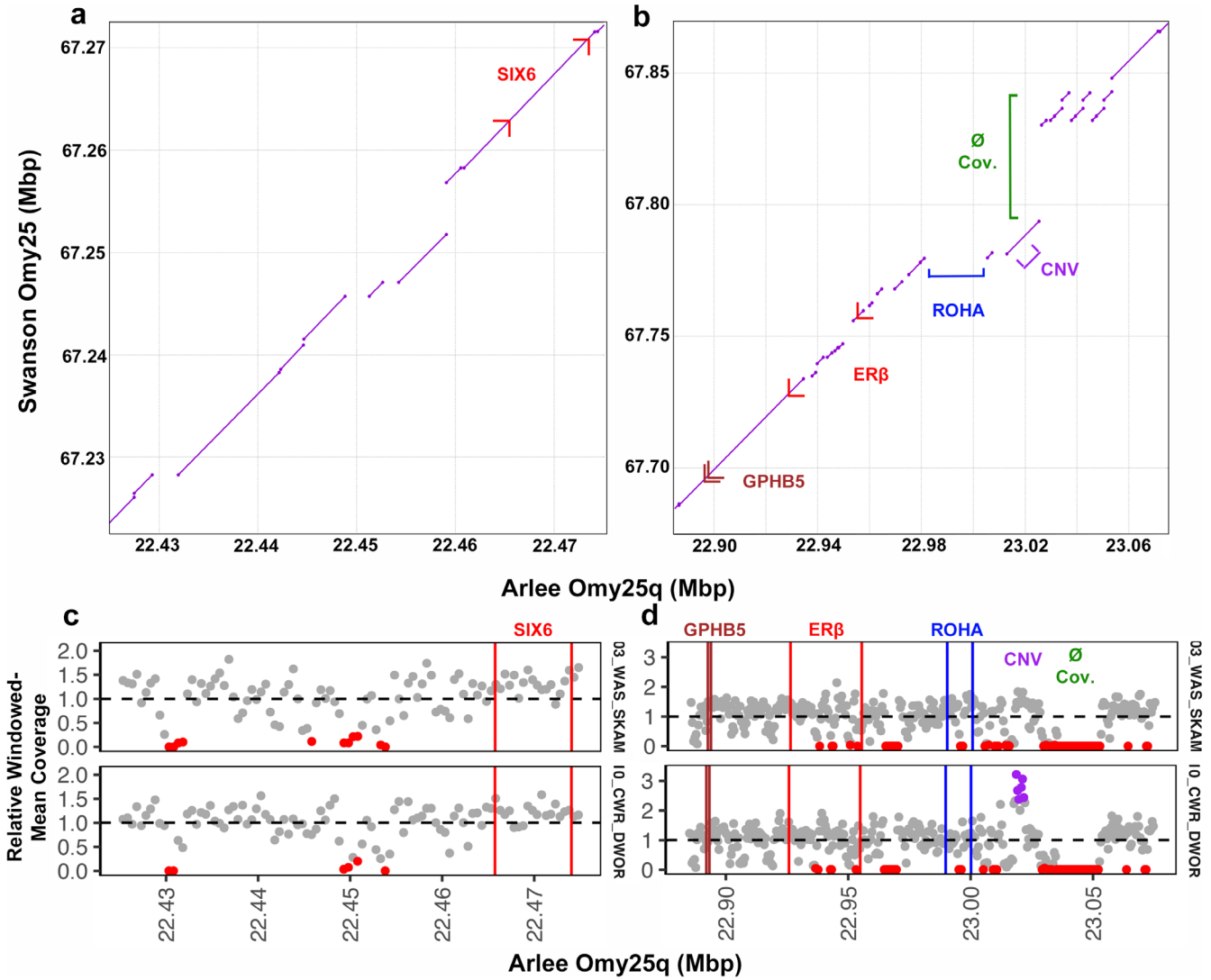

FIGURE 6 | Structural variation in the chromosome Omy25q regions association with age-at-maturity (left) and repeat spawning morphology (right). (a, b) Mummer alignments of subsets of the Omy25q chromosome arm from the Arlee and Swanson genome assemblies, with structural features labelled (genes are shown with right angles indicated the direction of transcription; CNV, copy number variant; ØCOV, zero read coverageROHA, region of highest association). (c, d) Plots of relative windowed-mean sequence coverage for two Columbia Basin stocks on the Arlee genome assembly, with structural features labelled (Skamania Hatchery, Washougal River; Dworshak Hatchery, Clearwater River). Note that for (a, c) the entire region reflects the ROHA. The figure depicts structural genome variations detected from the alignment of the Arlee genome assembly to the Swanson assembly, but similar structural variations were detected in the alignment of the Arlee assembly to the other two de novo genome assemblies (Figure S3).

the Omy28 region, which includes the *greb1l* and *rock1* genes, there were some INDEL discontinuities between the assemblies, which lay in the intergenic region between the two genes or within the introns of *greb1l*. While these regions do not obviously cause major changes to the function of either of these genes, association studies (based on SNPs in previous genome assemblies) have shown that the intergenic and upstream (centre-adjacent) portion of the *greb1l* gene shows the strongest association with migration timing in Columbia Basin steelhead (Willis et al. 2020), suggesting the intergenic structural variation could be similarly associated with this trait.

The region of Omy25q presents an even more complex case than Omy28. Willis et al. (2023) found that two aspects of spawning in steelhead, age-at-maturity (age at maiden spawning) and repeat spawning phenology (proclivity to skip or spawn consecutively following maiden spawning), were separately associated with adjacent regions on Omy25q that occupy a contiguous island of linkage disequilibrium in two wild populations. Similarly, Willis et al. (2024) noted that these regions exhibited exceptional allele frequency variation among 74 population samples from the Columbia Basin (Figure S2). Alignments of this region among the four de novo genome assemblies showed more complex structural variation than the Omy28 region (Figure 6, Figures S3–S6). In the region near the *six6* gene, associated with age-at-maturity, structural variation (INDELs) was evident in the upstream intergenic region, which was also the region of highest association in the SNP-based GWAS analyses (Figure 6, Figures S3–S6). Notably, this was also reflected in plots of relative windowed-mean coverage as differing areas of low or zero coverage across samples. While there was some coverage variation evident among population samples, it was small relative to the differences among genome assemblies (Figure 6, Figure S7).

Alignments of the four assemblies near the Omy25q region associated with repeat spawning phenology showed very complex variation upstream of two genes, *erβ* and *gphb5*, including both INDELs and duplications (Figure 6, Figure S8), as well as a prominent INDEL in the first intron of the *erβ* gene. Notably, the region most strongly associated with this trait is only present in the Arlee assembly and absent from the other three assemblies. Similarly, plots of relative windowed-mean coverage across populations revealed not only that the positions of zero coverage differed among assemblies, but also that there exists an adjacent copy-number variant (CNV) that varies in frequency among populations. In contrast, while plots of SNP frequency across this region show differences among stocks (Figure S9), they do not capture the differences in structural variation, which may show a different pattern from SNP variants in mediating the adjacent genes and associated traits.

4 | Conclusions

This study provides a comparative analysis of de novo genome assemblies of four genetically diverse rainbow trout lines, an important foundational step toward the construction of a robust pangenome reference for *O. mykiss*. Our results show pervasive SVs including deletions, insertions and complex rearrangements with lineage-specific patterns. Specifically, the genomic regions that harbour these SVs were found to overlap with genes that are functionally enriched for growth, reproduction and environmental adaptation pathways, implicating their potential roles in genetic diversity across divergent lines of this species. SVs were identified in key regions such as the Omy17 muscle yield QTL and the maturation-associated *six6*/*erβ*-*gphb5* region, which highlights the role that they may have in the genetics of complex traits in rainbow trout. Our findings demonstrate the importance of high-quality, strain-specific genome assemblies that will provide comprehensive representation of intra-specific genetic diversity. This work paves the way for follow-up functional validation and evolutionary studies, and broadens the applicability of a pangenome framework for basic and applied research in salmonid genomics by highlighting the need to incorporate additional de novo genome assemblies from strains shaped by distinct life histories or other traits of interest.

Author Contributions

Y.P. conceived and planned the study. G.H.T. conceived the idea of generating doubled haploid lines that represent the species diversity for genetic and genomic analyses and supervised the collection, generation and propagation of the lines. P.A.W. was involved in gametes collection and generated and propagated the lines. G.H.T., P.A.W. and S.R.N. provided the genetic lines and samples for this study. A.A. planned and conducted the SV analyses and drafted the manuscript including tables, figures and supplemental information. Y.P. planned, coordinated and supervised the DNA sequencing and data collection, handling and storage of the data, genome assemblies and analyses of the data and contributed to the draft manuscript writing. G.C.W., R.C.Y. and B.E.S. conducted optical mapping and DNA sequencing and contributed resources. S.W. and S.R.N. conducted data analyses and contributed to the manuscript writing and revising and to the presentation of the results. M.S. provided resources and supervision, contributed to data interpretation and revising the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Dr. Guangtu Gao with our utmost gratitude. Sadly, Dr. Gao passed away last year. Before his passing, he co-managed the sequence data collection and archiving and generated the de-novo genome assemblies of the WR and KC lines. We would like to take this opportunity to acknowledge his incredible talents and most impactful contributions to this work and to the past decade of salmonids genomics research. We thank Roseanna Long from USDA-ARS in Leetown, WV for helping with processing the DNA samples and Sheron Simpson from USDA-ARS in Stoneville, M.S. for technical assistance with the PacBio sequencing of the KC DNA sample. This study was supported by funding from Agricultural Research Service (ARS) in-house project number 8082-10600-002-000D. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. government. USDA is an equal opportunity provider and employer.

Funding

This work was supported by Agricultural Research Service, 8082-10600-002-000D.

Disclosure

Benefit-sharing statement: Benefits from this research accrue from the sharing of our data and results on public databases as described above.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All the WR line raw genome sequence data and the final genome assembly are associated with the NCBI SRA BioProject Accession: PRJNA928430. All the KC line genome sequence data and the genome assembly are associated with the NCBI SRA BioProject Accession: PRJNA1027447. The genome assembly accessions are: USDA_OmykA_1.1: GCA_013265735.3; Omyk_2.0: GCA_025558465.1; USDA_OmykWR_1.0: GCA_029834435.1; USDA_OmykKC_1.0: GCA_034753235.1. The six VCF files with filtered SVs for each line based on alignment to the Arlee and WR genome as the reference were deposited in Dryad (https://doi.org/10.5061/dryad.1rn8pk17m) and will be publicly released after peer review and acceptance of the manuscript. All the data analysis results are provided in the Supporting Information files.

References

Ali, A., G. Gao, R. Al-Tobasei, et al. 2025. "Chromosome Level Genome Assembly and Annotation of the Swanson Rainbow Trout Homozygous Line." *Scientific Data* 12, no. 1: 345.

Ali, A., G. H. Thorgaard, and M. Salem. 2021. "PacBio Iso-Seq Improves the Rainbow Trout Genome Annotation and Identifies Alternative Splicing Associated With Economically Important Phenotypes." *Frontiers in Genetics* 12, no. 1194: 683408.

Alkan, C., B. P. Coe, and E. E. Eichler. 2011. "Genome Structural Variation Discovery and Genotyping." *Nature Reviews Genetics* 12, no. 5: 363–376.

Andrews, K. R., T. Seaborn, J. P. Egan, et al. 2023. "Whole Genome Resequencing Identifies Local Adaptation Associated With Environmental Variation for Redband Trout." *Molecular Ecology* 32, no. 4: 800–818.

Arostegui, M. C., T. P. Quinn, L. W. Seeb, J. E. Seeb, and G. J. McKinney. 2019. "Retention of a Chromosomal Inversion From an Anadromous Ancestor Provides the Genetic Basis for Alternative Freshwater Ecotypes in Rainbow Trout." *Molecular Ecology* 28, no. 6: 1412–1427.

Bao, H., N. Xue, B. Wang, et al. 2025. "Exploration of Gene Presence/Absence Variations in *Oncorhynchus mykiss* and Their Differentiation Between Wild and Selection Populations." *Open Biology* 15, no. 5: 240382.

Bardsley, E. N., H. Davis, K. J. Buckler, and D. J. Paterson. 2018. "Neurotransmitter Switching Coupled to β-Adrenergic Signaling in Sympathetic Neurons in Prehypertensive States." *Hypertension* 71, no. 6: 1226–1238.

Beghi, S., M. Furmanik, A. Jaminon, et al. 2022. "Calcium Signalling in Heart and Vessels: Role of Calmodulin and Downstream Calmodulin-Dependent Protein Kinases." *International Journal of Molecular Sciences* 23, no. 24: 16139.

Bers, D. M. 2008. "Calcium Cycling and Signaling in Cardiac Myocytes." *Annual Review of Physiology* 70: 23–49.

Bertolotti, A. C., R. M. Layer, M. K. Gundappa, et al. 2020. "The Structural Variation Landscape in 492 Atlantic Salmon Genomes." *Nature Communications* 11, no. 1: 5176.

Bystriansky, J. S., and P. M. Schulte. 2011. "Changes in Gill H+-ATPase and Na+/K +-ATPase Expression and Activity During Freshwater Acclimation of Atlantic Salmon (*Salmo salar*)." *Journal of Experimental Biology* 214, no. 14: 2435–2442.

Campbell, J. M., P. A. Carter, P. A. Wheeler, and G. H. Thorgaard. 2015. "Aggressive Behavior, Brain Size and Domestication in Clonal Rainbow Trout Lines." *Behavior Genetics* 45, no. 2: 245–254.

Campbell, M. A., E. C. Anderson, J. C. Garza, and D. E. Pearse. 2021. "Polygenic Basis and the Role of Genome Duplication in Adaptation to Similar Selective Environments." *Journal of Heredity* 112, no. 7: 614–625.

Chan, Y. F., M. E. Marks, F. C. Jones, et al. 2010. "Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer." *Science* 327: 302–305.

Chapman, N. A., D. J. Dupré, and J. K. Rainey. 2014. "The Apelin Receptor: Physiology, Pathology, Cell Signalling, and Ligand Modulation of a Peptide-Activated Class A GPCR." *Biochemistry and Cell Biology* 92, no. 6: 431–440.

Chen, T., G. Zhou, Q. Zhu, et al. 2010. "Overexpression of Vascular Endothelial Growth Factor 165 (VEGF165) Protects Cardiomyocytes Against Doxorubicin-Induced Apoptosis." *Journal of Chemotherapy* 22, no. 6: 402–406.

Chen, X. Q., Y. G. Zhou, K. T. Jia, et al. 2024. "Comparative Study on Growth, Biochemical Index, and Gene Expression of Rainbow Trout and Steelhead Under Chronic Hypoxia." *Acta Hydrobiologica Sinica* 48, no. 11: 1789–1801.

Chen, X. Q., Y. G. Zhou, K. T. Jia, et al. 2025. "The Histological and Molecular Response of Acute Hypoxia and Reoxygenation in Rainbow Trout and Steelhead." *Comparative Biochemistry and Physiology - Part A: Molecular and Integrative Physiology* 305: 111854.

Chen, Y., M. Z. Khan, X. Wang, et al. 2024. "Structural Variations in Livestock Genomes and Their Associations With Phenotypic Traits: A Review." *Frontiers in Veterinary Science* 11: 1416220.

Cheng, H., G. T. Concepcion, X. Feng, H. Zhang, and H. Li. 2021. "Haplotype-Resolved De Novo Assembly Using Phased Assembly Graphs With Hifiasm." *Nature Methods* 18, no. 2: 170–175.

Chottová, D. M., and J. Slavíková. 2011. "Adrenergic Regulation of the Mammalian Heart." *Ceskoslovenská Fysiologie / Ústredni ústav biologický* 60, no. 1: 14–19.

Collins, R. L., and M. E. Talkowski. 2025. "Diversity and Consequences of Structural Variation in the Human Genome." *Nature Reviews Genetics* 26, no. 7: 443–462.

Corvol, P. 2009. "Effect of Apelin in Cardiovascular System and Water Metabolism." *Thérapie* 64, no. 4: 233–239.

Cowan, K. J., and K. B. Storey. 2003. "Mitogen-Activated Protein Kinases: New Signaling Pathways Functioning in Cellular Responses to Environmental Stress." *Journal of Experimental Biology* 206, no. 7: 1107–1115.

Culbert, B. M., S. D. McCormick, and N. J. Bernier. 2024. "Regulation of a Unique Neuroendocrine Complex by Environmental Salinity in Salmonid Fishes." bioRxiv:2024.2012.2013.627795.

Culbert, B. M., A. M. Regish, D. J. Hall, S. D. McCormick, and N. J. Bernier. 2022. "Neuroendocrine Regulation of Plasma Cortisol Levels During Smoltification and Seawater Acclimation of Atlantic Salmon." *Frontiers in Endocrinology* 13: 859817.

Dale Abel, E. 2021. "Insulin Signaling in the Heart." *American Journal of Physiology—Endocrinology and Metabolism* 321, no. 1: 130–145.

Danecek, P., J. K. Bonfield, J. Liddle, et al. 2021. "Twelve Years of SAMtools and BCFtools." *GigaScience* 10, no. 2: giab008.

de los Ríos-Pérez, L., R. M. Brunner, F. Hadlich, et al. 2020. "Comparative Analysis of the Transcriptome and Distribution of Putative Snps in Two Rainbow Trout (*Oncorhynchus mykiss*) Breeding Strains by Using Next-Generation Sequencing." *Genes* 11, no. 8: 1–16.

Deiner, K., J. C. Garza, R. Coey, and D. J. Girman. 2007. "Population Structure and Genetic Diversity of Trout (*Oncorhynchus mykiss*) Above and Below Natural and Man-Made Barriers in the Russian River, California." *Conservation Genetics* 8, no. 2: 437–454.

Dhakal, U., H.-S. Kim, and C. Toomajian. 2024. "The Landscape and Predicted Roles of Structural Variants in *Fusarium graminearum* Genomes." *G3: Genes, Genomes, Genetics* 14, no. 6: jkae065.

Ding, Y., E. F. Johnston, and T. E. Gillis. 2022. "Mitogen-Activated Protein Kinases Contribute to Temperature-Induced Cardiac Remodelling in Rainbow Trout (*Oncorhynchus mykiss*)." *Journal of Comparative Physiology B Biochemical, Systemic, and Environmental Physiology* 192, no. 1: 61–76.

Durmaz Mitchell, E., E. Kerdaffrec, P. Schmidt, T. Flatt, and S. Kittelmann. 2025. "A Balanced Inversion Polymorphism Exhibits a Dominance Reversal at the Gene Expression Level That Depends on Developmental Context." *Molecular Biology and Evolution* 42, no. 9: msaf209.

Flores, A.-M., K. A. Christensen, T. Godin, et al. 2025. "The Genome Assembly of the Westslope Cutthroat Trout, *Oncorhynchus lewisi*, Reveals Interspecific Chromosomal Rearrangements With the Rainbow Trout, *Oncorhynchus mykiss*." *G3: Genes, Genomes, Genetics* 15, no. 5: jkaf064.

Foddai, M., C. G. Carter, K. Anderson, et al. 2023. "Impact of Temperature and Dietary Replacement of Fishmeal on Cardiovascular Remodelling and Growth Performance of Adult Atlantic Salmon (*Salmo salar* L.)." *Aquaculture* 573: 739590.

Franzoso, M., L. Dokshokova, L. Vitiello, T. Zaglia, and M. Mongillo. 2022. "Tuning the Consonance of Microscopic Neuro-Cardiac Interactions Allows the Heart Beats to Play Countless Genres." *Frontiers in Physiology* 13: 841740.

Fu, Q., and Y. K. Xiang. 2015. "Trafficking of β-Adrenergic Receptors: Implications in Intracellular Receptor Signaling." *Progress in Molecular Biology and Translational Science* 132: 151–188.

Gao, G., S. Magadan, G. C. Waldbieser, et al. 2021. "A Long Reads-Based De-Novo Assembly of the Genome of the Arlee Homozygous Line Reveals Chromosomal Rearrangements in Rainbow Trout." *G3 (Bethesda)* 11: jkab052.

Gao, G., T. Nome, D. E. Pearse, et al. 2018. "A New Single Nucleotide Polymorphism Database for Rainbow Trout Generated Through Whole Genome Resequencing." *Frontiers in Genetics* 9: 147.

Ge, S. X., D. Jung, and R. Yao. 2020. "ShinyGO: A Graphical Gene-Set Enrichment Tool for Animals and Plants." *Bioinformatics* 36, no. 8: 2628–2629.

Ghurye, J., A. Rhie, B. P. Walenz, et al. 2019. "Integrating Hi-C Links With Assembly Graphs for Chromosome-Scale Assembly." *PLoS Computational Biology* 15, no. 8: e1007273.

Goel, M., H. Sun, W.-B. Jiao, and K. Schneeberger. 2019. "SyRI: Finding Genomic Rearrangements and Local Sequence Differences From Whole-Genome Assemblies." *Genome Biology* 20, no. 1: 277.

Gong, Y., Y. Li, X. Liu, Y. Ma, and L. Jiang. 2023. "A Review of the Pangenome: How It Affects Our Understanding of Genomic Variation, Selection and Breeding in Domestic Animals?" *Journal of Animal Science and Biotechnology* 14, no. 1: 73.

Haddad, J. J. 2005. "N-Methyl-D-Aspartate (NMDA) and the Regulation of Mitogen-Activated Protein Kinase (MAPK) Signaling Pathways: A Revolving Neurochemical Axis for Therapeutic Intervention?" *Progress in Neurobiology* 77, no. 4: 252–282.

Hale, M. C., D. E. Pearse, and M. A. Campbell. 2024. "Characterization and Distribution of a 14-Mb Chromosomal Inversion in Native Populations of Rainbow Trout (*Oncorhynchus mykiss*)." *G3: Genes, Genomes, Genetics* 14, no. 7: jkae100.

Han, K., J. Lee, T. J. Meyer, P. Remedios, L. Goodwin, and M. A. Batzer. 2008. "L1 Recombination-Associated Deletions Generate Human Genomic Variation." *Proceedings of the National Academy of Sciences of the United States of America* 105, no. 49: 19366–19371.

Haraguchi, S., Y. Yamamoto, Y. Suzuki, et al. 2015. "7α-Hydroxypregnenolone, a Key Neuronal Modulator of Locomotion, Stimulates Upstream Migration by Means of the Dopaminergic System in Salmon." *Scientific Reports* 5: 12546.

Hartasanchez, D. A., M. Braso-Vives, J. M. Heredia-Genestar, M. Pybus, and A. Navarro. 2018. "Effect of Collapsed Duplications on Diversity Estimates: What to Expect." *Genome Biology and Evolution* 10, no. 11: 2899–2905.

He, Z., D. M. Opland, K. J. Way, et al. 2006. "Regulation of Vascular Endothelial Growth Factor Expression and Vascularization in the Myocardium by Insulin Receptor and PI3K/Akt Pathways in Insulin Resistance and Ischemia." *Arteriosclerosis, Thrombosis, and Vascular Biology* 26, no. 4: 787–793.

Heller, D., and M. Vingron. 2019. "SVIM: Structural Variant Identification Using Mapped Long Reads." *Bioinformatics* 35, no. 17: 2907–2915.

Heredia-Middleton, P., J. Brunelli, R. E. Drew, and G. H. Thorgaard. 2008. "Heat Shock Protein (HSP70) RNA Expression Differs Among Rainbow Trout (*Oncorhynchus mykiss*) Clonal Lines." *Comparative Biochemistry and Physiology. Part B, Biochemistry & Molecular Biology* 149, no. 4: 552–556.

Hu, C., Q. Zuo, K. Jin, et al. 2022. "Retinoic Acid Promotes Formation of Chicken (*Gallus gallus*) Spermatogonial Stem Cells by Regulating the ECM-Receptor Interaction Signaling Pathway." *Gene* 820: 146227.

Ineno, T., K. Tamaki, K. Yamada, et al. 2018. "Thermal Tolerance of a Thermally Selected Strain of Rainbow Trout Oncorhynchus Mykiss and the Pedigrees of Its F1 and F2 Generations Indicated by Their Critical Thermal Maxima." *Fisheries Science* 84, no. 4: 671–679.

Jaggi, M. 2018. "Recent Advancement on Map Kinase Cascade in Biotic Stress." In *Molecular Aspects of Plant-Pathogen Interaction*, edited by Singh and I. K. Singh, 139–158. Springer Nature Link.

Jain, M., S. Koren, K. H. Miga, et al. 2018. "Nanopore Sequencing and Assembly of a Human Genome With Ultra-Long Reads." *Nature Biotechnology* 36, no. 4: 338–345.

Johnston, E. F., and T. E. Gillis. 2020. "Short-Term Cyclical Stretch Phosphorylates p38 and ERK1/2 MAPKs in Cultured Fibroblasts From the Hearts of Rainbow Trout, *Oncorhynchus mykiss*." *Biology Open* 9, no. 1: bio049296.

Jones, F. C., M. G. Grabherr, Y. F. Chan, et al. 2012. "The Genomic Basis of Adaptive Evolution in Threespine Sticklebacks." *Nature* 484, no. 7392: 55–61.

Kirubakaran, T. G., H. Grove, M. P. Kent, et al. 2016. "Two Adjacent Inversions Maintain Genomic Differentiation Between Migratory and Stationary Ecotypes of Atlantic Cod." *Molecular Ecology* 25: 2130–2143.

Kleinau, G., B. Chini, L. Andersson, and P. Scheerer. 2024. "The Role of G Protein-Coupled Receptors and Their Ligands in Animal Domestication." *Animal Genetics* 55, no. 6: 893–906.

Klossner, S., R. Li, S. Ruoss, A. C. Durieux, and M. Flück. 2013. "Quantitative Changes in Focal Adhesion Kinase and Its Inhibitor, FRNK, Drive Load-Dependent Expression of Costamere Components." *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology* 305, no. 6: R647–R657.

Komen, H., and G. H. Thorgaard. 2007. "Androgenesis, Gynogenesis and the Production of Clones in Fishes: A Review." *Aquaculture* 269, no. 1: 150–173.

Koren, S., B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, and A. M. Phillippy. 2017. "Canu: Scalable and Accurate Long-Read Assembly via Adaptive k-Mer Weighting and Repeat Separation." *Genome Research* 27, no. 5: 722–736.

Kurtz, S., A. Phillippy, A. L. Delcher, et al. 2004. "Versatile and Open Software for Comparing Large Genomes." *Genome Biology* 5, no. 2: R12.

Larson, W. A., Y. Palti, G. Gao, K. I. Warheit, and J. E. Seeb. 2017. "Rapid Discovery of SNPs That Differentiate Hatchery Steelhead Trout From ESA-Listed Natural-Origin Steelhead Trout Using a 57K SNP Array." *Canadian Journal of Fisheries and Aquatic Sciences* 75, no. 7: 1160–1168.

Lecomte, L., M. Árnyasi, A. L. Ferchaud, et al. 2024. "Investigating Structural Variant, Indel and Single Nucleotide Polymorphism Differentiation Between Locally Adapted Atlantic Salmon Populations." *Evolutionary Applications* 17, no. 3: e13653.

Leeds, T. D., J. T. Silverstein, G. M. Weber, et al. 2010. "Response to Selection for Bacterial Cold Water Disease Resistance in Rainbow Trout." *Journal of Animal Science* 88, no. 6: 1936–1946.

Li, H. 2013. "Aligning Sequence Reads, Clone Sequences and Assembly Contigs With BWA-MEM." arXiv: Genomics.

Li, H., B. Handsaker, A. Wysoker, et al. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25, no. 16: 2078–2079.

Lin, J., P. Jia, S. Wang, W. Kosters, and K. Ye. 2023. "Comparison and Benchmark of Structural Variants Detected From Long Read and Long-Read Assembly." *Briefings in Bioinformatics* 24, no. 4: bbad188.

Lin, Y., C. Ye, X. Li, et al. 2023. "quarTeT: A Telomere-to-Telomere Toolkit for Gap-Free Genome Assembly and Centromeric Repeat Identification." *Horticulture Research* 10, no. 8: uhad127.

Linlin, A., Y. Zhang, B. Xu, et al. 2024. "Comprehensive Analyses of Annexins in Naked Carp (*Gymnocypris przewalskii*) Unveil Their Roles in Saline-Alkaline Stress." *Aquaculture* 579: 740175.

Liu, S., G. Gao, R. M. Layer, et al. 2021. "Identification of High-Confidence Structural Variants in Domesticated Rainbow Trout Using Whole-Genome Sequencing." *Frontiers in Genetics* 12: 639355.

Liu, Y. H., C. Luo, S. G. Golding, J. B. Ioffe, and X. M. Zhou. 2024. "Tradeoffs in Alignment and Assembly-Based Methods for Structural Variant Detection With Long-Read Sequencing Data." *Nature Communications* 15, no. 1: 2447.

Lucas, M. D., R. E. Drew, P. A. Wheeler, P. A. Verrell, and G. H. Thorgaard. 2004. "Behavioral Differences Among Rainbow Trout Clonal Lines." *Behavior Genetics* 34, no. 3: 355–365.

Macqueen, D. J., N. I. Bower, and I. A. Johnston. 2010. "Positioning the Expanded Akirin Gene Family of Atlantic Salmon Within the Transcriptional Networks of Myogenesis." *Biochemical and Biophysical Research Communications* 400, no. 4: 599–605.

Miller, J. R., A. L. Delcher, S. Koren, et al. 2008. "Aggressive Assembly of Pyrosequencing Reads With Mates." *Bioinformatics* 24, no. 24: 2818–2824.

Miller, M. R., J. P. Brunelli, P. A. Wheeler, et al. 2012. "A Conserved Haplotype Controls Parallel Adaptation in Geographically Distant Salmonid Populations." *Molecular Ecology* 21, no. 2: 237–249.

Miller, R. K., H. Qadota, M. L. Landsverk, K. B. Mercer, H. F. Epstein, and G. M. Benian. 2006. "UNC-98 Links an Integrin-Associated Complex to Thick Filaments in *Caenorhabditis elegans* Muscle." *Journal of Cell Biology* 175, no. 6: 853–859.

Monette, M. Y., and J. P. Velotta. 2022. "Network Analysis Reveals That Acute Stress Exacerbates Gene Regulatory Responses of the Gill to Seawater in Atlantic Salmon." *Journal of Experimental Biology* 225, no. 8: jeb.242424.

Morash, A. J., M. Vanderveken, and M. C. GB. 2014. "Muscle Metabolic Remodeling in Response to Endurance Exercise in Salmonids." *Frontiers in Physiology* 5: 452.

Novakov, V., G. S. Sandhu, D. Dragomir-Daescu, and M. Klabusay. 2015. "Apelinergic System in Endothelial Cells and Its Role in Angiogenesis in Myocardial Ischemia." *Vascular Pharmacology* 76: 1–10.

Ojima, D., and M. Iwata. 2009. "Central Administration of Growth Hormone-Releasing Hormone Triggers Downstream Movement and Schooling Behavior of Chum Salmon (*Oncorhynchus keta*) Fry in an Artificial Stream." *Comparative Biochemistry and Physiology—A Molecular and Integrative Physiology* 152, no. 3: 293–298.

Ojima, D., and M. Iwata. 2010. "Central Administration of Growth Hormone-Releasing Hormone and Corticotropin-Releasing Hormone Stimulate Downstream Movement and Thyroxine Secretion in Fall-Smolting Coho Salmon (*Oncorhynchus kisutch*)." *General and Comparative Endocrinology* 168, no. 1: 82–87.

Ólafsson, E. B., A. L. ten Hoeve, X. Li-Wang, L. Westermark, M. Varas-Godoy, and A. Barragan. 2021. "Convergent Met and Voltage-Gated Ca2+ Channel Signaling Drives Hypermigration of Toxoplasma-Infected Dendritic Cells." *Journal of Cell Science* 134, no. 5: jcs241752.

Palti, Y., G. Gao, M. R. Miller, et al. 2014. "A Resource of Single-Nucleotide Polymorphisms for Rainbow Trout Generated by Restriction-Site Associated DNA Sequencing of Doubled Haploids." *Molecular Ecology Resources* 14, no. 3: 588–596.

Palti, Y., R. L. Vallejo, M. K. Purcell, et al. 2024. "Genome-Wide Association Analysis of the Resistance to Infectious Hematopoietic Necrosis Virus in Two Rainbow Trout Aquaculture Lines Confirms Oligogenic Architecture With Several Moderate Effect Quantitative Trait Loci." *Frontiers in Genetics* 15: 1394656.

Papa, A., J. Kushner, and S. O. Marx. 2022. "Adrenergic Regulation of Calcium Channels in the Heart." *Annual Review of Physiology* 84: 285–306.

Parkinson, E., E. Taylor, and E. Keeley. 1999. "Conserving Genetic Diversity in Rainbow Trout." In: Proceedings of a Conference on the Biology and Management of Species and Habitats at Risk: 1999; Kamloops, BC: BC Ministry of Environment, Lands and Parks and University College of the Cariboo.

Paterson, A. L., J. M. J. Weaver, M. D. Eldridge, S. Tavaré, R. C. Fitzgerald, and P. A. W. Edwards. 2015. "Mobile Element Insertions Are Frequent in Oesophageal Adenocarcinomas and Can Mislead Paired-End Sequencing Analysis." *BMC Genomics* 16, no. 1: 473.

Pearse, D. E., N. J. Barson, T. Nome, et al. 2019. "Sex-Dependent Dominance Maintains Migration Supergene in Rainbow Trout." *Nature Ecology & Evolution* 3, no. 12: 1731–1742.

Quinlan, A. R., and I. M. Hall. 2012. "Characterizing Complex Structural Variation in Germline and Somatic Genomes." *Trends in Genetics* 28, no. 1: 43–53.

Rego, K., J. D. Hansen, and E. S. Bromage. 2020. "Genomic Architecture and Repertoire of the Rainbow Trout Immunoglobulin Light Chain Genes." *Developmental and Comparative Immunology* 113: 103776.

Rexroad Iii, C. E., Y. Palti, R. L. Vallejo, and J. T. Silverstein. 2006. "Integration of Molecular Genetic Information Into the NCCCWA Selective Breeding Program for Rainbow Trout." *Israeli Journal of Aquaculture—Bamidgeh* 2006: 323–327.

Rhie, A., B. P. Walenz, S. Koren, and A. M. Phillippy. 2020. "Merqury: Reference-Free Quality, Completeness, and Phasing Assessment for Genome Assemblies." *Genome Biology* 21, no. 1: 245.

Ribatti, D. 2019. "The Discovery of the Fundamental Role of VEGF in the Development of the Vascular System." *Mechanisms of Development* 160: 103579.

Robison, B. D., and G. H. Thorgaard. 2004. "The Phenotypic Relationship of a Clonal Line to Its Population of Origin: Rapid Embryonic Development in an Alaskan Population of Rainbow Trout." *Transactions of the American Fisheries Society* 133, no. 2: 455–461.

Salem, M., R. Al-Tobasei, A. Ali, et al. 2018. "Genome-Wide Association Analysis With a 50K Transcribed Gene SNP-Chip Identifies QTL Affecting Muscle Yield in Rainbow Trout." *Frontiers in Genetics* 9: 387.

Salem, M., C. E. Rexroad 3rd, J. Wang, G. H. Thorgaard, and J. Yao. 2010. "Characterization of the Rainbow Trout Transcriptome Using Sanger and 454-Pyrosequencing Approaches." *BMC Genomics* 11: 564.

San, J., Y. Du, G. Wu, R. Xu, J. Yang, and J. Hu. 2021. "Transcriptome Analysis Identifies Signaling Pathways Related to Meat Quality in Broiler Chickens—The Extracellular Matrix (ECM) Receptor Interaction Signaling Pathway." *Poultry Science* 100, no. 6: 101135.

Sanchez-Alonso, J. L., L. Fedele, J. S. Copier, et al. 2023. "Functional LTCC-β2AR Complex Needs Caveolin-3 and Is Disrupted in Heart Failure." *Circulation Research* 133, no. 2: 120–137.

Savini, D., A. Occhipinti-Ambrogi, A. Marchini, et al. 2010. "The Top 27 Animal Alien Species Introduced Into Europe for Aquaculture and Related Activities." *Journal of Applied Ichthyology* 26, no. Suppl. 2: 1–7.

Sedlazeck, F. J., P. Rescheneder, M. Smolka, et al. 2018. "Accurate Detection of Complex Structural Variations Using Single-Molecule Sequencing." *Nature Methods* 15, no. 6: 461–468.

Shao, M., K. Shi, Q. Zhao, et al. 2022. "Transcriptome Analysis Reveals the Differentially Expressed Genes Associated With Growth in Guangxi Partridge Chickens." *Genes* 13: 798.

Shibuya, M., and L. Claesson-Welsh. 2006. "Signal Transduction by VEGF Receptors in Regulation of Angiogenesis and Lymphangiogenesis." *Experimental Cell Research* 312, no. 5: 549–560.

Shumate, A., and S. L. Salzberg. 2021. "Liftoff: Accurate Mapping of Gene Annotations." *Bioinformatics* 37, no. 12: 1639–1643.

Smolka, M., L. F. Paulin, C. M. Grochowski, et al. 2024. "Detection of Mosaic and Population-Level Structural Variants With Sniffles2." *Nature Biotechnology* 42, no. 10: 1571–1580.

Stephens, Z., C. Wang, R. K. Iyer, and J.-P. Kocher. 2018. "Detection and Visualization of Complex Structural Variants From Long Reads." *BMC Bioinformatics* 19, no. 20: 508.

Sundell, K., F. Jutfelt, T. Ágústsson, et al. 2003. "Intestinal Transport Mechanisms and Plasma Cortisol Levels During Normal and Out-of-Season Parr-Smolt Transformation of Atlantic Salmon, *Salmo salar*." *Aquaculture* 222, no. 1–4: 265–285.

Sundell, K. S., and H. Sundh. 2012. "Intestinal Fluid Absorption in Anadromous Salmonids: Importance of Tight Junctions and Aquaporins." *Frontiers in Physiology* 3: 388.

Terekhanova, N. V., A. E. Barmintseva, A. S. Kondrashov, G. A. Bazykin, and N. S. Mugue. 2019. "Architecture of Parallel Adaptation in Ten Lacustrine Threespine Stickleback Populations From the White Sea Area." *Genome Biology and Evolution* 11, no. 9: 2605–2618.

Thorgaard, G. H. 1983. "Chromosomal Differences Among Rainbow Trout Populations." *Copeia* 1983, no. 3: 650–662.

Thrower, F. P., and J. E. Joyce. 2005. "Effects of 70 Years of Freshwater Residency on Survival, Growth, Early Maturation, and Smolting in a Stock of Anadromous Rainbow Trout From Southeast Alaska." *American Fisheries Society Symposium* 44: 485–496.

Titus, R. G., D. C. Erman, and W. M. Snider. 2010. "History and Status of Steelhead in California Coastal Drainages South of San Francisco Bay." In Draft for Publication as a Department of Fish and Game, Fish Bulletin.

Tse, W. K. F., K. P. Lai, and Y. Takei. 2011. "Medaka Osmotic Stress Transcription Factor 1b (Ostf1b/TSC22D3-2) Triggers Hyperosmotic Responses of Different Ion Transporters in Medaka Gill and Human Embryonic Kidney Cells via the JNK Signalling Pathway." *International Journal of Biochemistry and Cell Biology* 43, no. 12: 1764–1775.

Vallejo, R. L., S. Liu, G. Gao, et al. 2017. "Similar Genetic Architecture With Shared and Unique Quantitative Trait Loci for Bacterial Cold Water Disease Resistance in Two Rainbow Trout Breeding Populations." *Frontiers in Genetics* 8: 156.

Vaser, R., I. Sović, N. Nagarajan, and M. Šikić. 2017. "Fast and Accurate De Novo Genome Assembly From Long Uncorrected Reads." *Genome Research* 27, no. 5: 737–746.

Volkoff, H. 2016. "The Neuroendocrine Regulation of Food Intake in Fish: A Review of Current Knowledge." 10: 540.

Wang, D., Y. Su, X. Wang, H. Lei, and J. Yu. 2012. "Transposon-Derived and Satellite-Derived Repetitive Sequences Play Distinct Functional Roles in Mammalian Intron Size Expansion." *Evolutionary Bioinformatics* 2012, no. 8: 301–319.

Wang, L., Z. M. Zhu, N. K. Zhang, et al. 2016. "Apelin: An Endogenous Peptide Essential for Cardiomyogenic Differentiation of Mesenchymal Stem Cells via Activating Extracellular Signal-Regulated Kinase 1/2 and 5." *Cell Biology International* 40, no. 5: 501–514.

Willis, S., D. K. Coykendall, M. R. Campbell, and S. Narum. 2024. "Contrasting Patterns of Sequence Variation in Steelhead Populations Reflect Distinct Evolutionary Processes." *Evolutionary Applications* 17, no. 1: e13623.

Willis, S., J. Stephenson, A. Pierce, et al. 2023. "A Genomic Region Associated With Iteroparous Spawning Phenology is Linked With Age-at-Maturity in Female Steelhead Trout." *Evolutionary Applications* 17: e13622.

Willis, S. C., J. E. Hess, J. K. Fryer, et al. 2020. "Steelhead (*Oncorhynchus mykiss*) Lineages and Sexes Show Variable Patterns of Association of Adult Migration Timing and Age-At-Maturity Traits With Two Genomic Regions." *Evolutionary Applications* 13, no. 10: 2836–2856.

Xie, F., W. Liu, F. Feng, et al. 2015. "Apelin-13 Promotes Cardiomyocyte Hypertrophy via PI3K-Akt-ERK1/2-p70S6K and PI3K-Induced Autophagy." *Acta Biochimica et Biophysica Sinica* 47, no. 12: 969–980.

Xu, J. F., H. F. Chen, N. Wang, and J. Liu. 2022. "Research Advances in Hog1 MAPK Signaling Pathway in Fungi." *Biotechnology Bulletin* 38, no. 11: 32–40.

Yada, T., and M. Abe. 2024. "Comparison of Osmoregulatory and Endocrine Factors in Steelhead and Rainbow Trout *Oncorhynchus mykiss* Following Free Selection of Environmental Salinity." *Journal of Fish Biology*: 1–11.

Yang, X. D., Z. S. Hou, M. Q. Liu, et al. 2022. "Identification and Characterization of Mkk Genes and Their Expression Profiles in Rainbow Trout (*Oncorhynchus mykiss*) Symptomatically or Asymptomatically Infected With *Vibrio anguillarum*." *Fish and Shellfish Immunology* 121: 1–11.

Yano, A., B. Nicol, E. Jouanno, et al. 2013. "The Sexually Dimorphic on the Y-Chromosome Gene (sdY) is a Conserved Male-Specific Y-Chromosome Sequence in Many Salmonids." *Evolutionary Applications* 6: 486–496.

Zhang, G. M., L. Zheng, H. He, et al. 2018. "Associations of GBP2 Gene Copy Number Variations With Growth Traits and Transcriptional Expression in Chinese Cattle." *Gene* 647: 101–106.

Zhou, C. Q., W. Ka, W. K. Yuan, and J. L. Wang. 2021. "The Effect of Acute Heat Stress on the Innate Immune Function of Rainbow Trout Based on the Transcriptome." *Journal of Thermal Biology* 96: 102834.

Zwollo, P., J. C. Ray, M. Sestito, et al. 2015. "B Cell Signatures of BCWD-Resistant and Susceptible Lines of Rainbow Trout: A Shift Towards More EBF-Expressing Progenitors and Fewer Mature B Cells in Resistant Animals." *Developmental and Comparative Immunology* 48, no. 1: 1–12.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Table S1:** Whole-genome alignment statistics of Swanson, Keithley Creek and Whale Rock assemblies against the Arlee reference genome. **Table S2:** Counts of SyRI-annotated genomic regions and variants in rainbow trout strains before and after size-based filtration. **Table S3:** Syntenic regions present in domesticated strains (Arlee and Swanson) but absent in wild (Keithley Creek and Whale Rock). **Table S4:** Functional enrichment analysis of genes with CDS completely encompassed by syntenic blocks shared by Arlee and Swanson but absent in wild strains (Whale Rock and Keithley Creek). **Table S5:**

Un-aligned genomic regions specific to Arlee not shared with Keithley Creek, Whale Rock or Swanson. **Table S6:** Gene enrichment analysis of genes annotated in Arlee-unique structural regions not aligned to Swanson, Keithley Creek or Whale Rock. **Table S7:** Functional enrichment analysis of genes located within regions of structural rearrangements and local sequence variants unique to Arlee strain and absent in Swanson, Keithley Creek and Whale Rock. **Table S8:** Syntenic regions present in wild (Keithley Creek and Whale Rock) but absent in domesticated strains (Arlee and Swanson). **Table S9:** Functional enrichment analysis of genes located within regions of structural rearrangements and local sequence variants shared by wild strains (Whale Rock and Keithley Creek) but absent in Arlee and Swanson. **Table S10:** Unaligned genomic regions specific to Whale Rock not shared with Keithley Creek, Arlee or Swanson. **Table S11:** Functional enrichment analysis of genes located within regions of structural rearrangements and local sequence variants specific to Whale Rock and absent in Swanson, Arlee and Keithley Creek. **Table S12:** Genes located within the Omy17 inversion interval overlapping the fillet yield QTL. **Figure S1:** Alignment of the *greb1l-rock1* region of Chr. 28 across assemblies. **Figure S2:** Allele frequency differences on Chr. 25q among 74 populations of steelhead [Modified from Willis et al. (2023)]. **Figure S3:** Alignment of the *six6-erβ* region of Chr. 25q across assemblies. **Figure S4:** Annotation and structural variation of the *six6* and *erβ* region of Chr. 25q in alignment of the Swanson (2.0) and Arlee assemblies. **Figure S5:** Annotation and structural variation of the *six6* and *erβ* region of Chr. 25q in alignment of the Arlee and Whale Rock assemblies. **Figure S6:** Annotation and structural variation of the *six6* and *erβ* region of Chr. 25q in alignment of the Arlee and KC assemblies. **Figure S7:** Relative sequence coverage by population across the four assemblies. Regions of low coverage are shown in red. **Figure S8:** Relative sequence coverage by population across the four assemblies. Regions of low coverage are shown in red. **Figure S9:** Allele frequencies across stocks at the *erβ-gpdh5* region for each Omy assembly.