Supplemental Methods File 1.

Preparation of BAC libraries for genome assembly.

Blood was centrifuged and high molecular weight DNA was isolated for preparation of bacterial artificial chromosome (BAC) libraries. BAC libraries were prepared by Amplicon Express (Pullman, Washington, USA) with insert sizes of approximately 150 Kb and individual clones were arrayed into 384-well plates before BAC pools were created for sequencing. Sequencing of BAC pools was completed on 16 lanes (100bp paired-end) of an Illumina HiSeq 1500 instrument and produced a total of 2,518 million paired-end reads (2x100bp).

Preparation of synthetic long-read libraries for genome assembly.

High molecular weight DNA was isolated from fin clips to prepare libraries following manufacturers protocols for TruSeq Synthetic Long Reads (TSSLR; Illumina, California, USA). A total of 24 libraries were prepared with the TSSLR protocol and each was sequenced over 24 lanes (150bp paired-end) on an Illumina HiSeq 4000 instrument and produced a total of 4,521 million paired-end reads (2x150bp).

Preparation of mate-pair libraries for genome assembly.

High molecular weight DNA was isolated from fin clips to prepare mate-pair libraries ranging from 1 Kb to 40 Kb. Multiple small insert mate-pair libraries (600bp to 8 Kb) were prepared with Nextera Mate Pair Gel-Plus kits (Illumina) and sequenced on six lanes of an Illumina HiSeq 1500 instrument (100bp paired-end). Larger mate-pair libraries were prepared by Lucigen (Middleton, WI, USA) with either NxSeq 40 Kb Mate-Pair protocol (40 Kb fosmid library digested with BfaI and Csp6I) or NxSeq 20Kb Mate-Pair protocols (10Kb, 15Kb, and 20Kb libraries with NxSeq Clone Free Mate-Pair v2) and sequenced to saturation on five Illumina MiSeq runs (250bp paired-end). In total, mate-pair libraries produced a total of 2,250 million paired-end reads (2 x100bp small insert; 2 x 250bp large insert).

Preparation of RNA-seq libraries for transcriptome assembly.

Total RNA was isolated (Qiagen RNeasy) from tissues from eight organs (muscle, stomach, gill, liver, kidney, brain, heart, intestine) that had been stored at -80ºC, and mRNA was isolated with manufacturers recommendations from NEBNext Ultra and PolyA magenetic bead protocols. Libraries for RNA-seq were prepared with sense/antisense directional kits (NEBNext Ultra Directional and ScriptSeq V2) and sequenced on eight lanes (100bp paired-end) of an Illumina HiSeq 1500 instrument (two lanes per library).

De novo genome assembly approach and resources.

Illumina sequencing was performed in stages with various approaches including TruSeq Synthetic Long Reads (TSSLR), BAC clones, and mate-pair libraries. Collectively, our sequencing project generated a total of 9,294 million read pairs (1.58 TB) covering about 300X of the Chinook salmon genome (NCBI BioProject PRJNA402052). The short read data from synthetic long read libraries and BAC libraries were assembled with TRUSPADES and SPADES (Bankevich et al. 2016), respectively. A total of 10,752 assemblies were performed to produce more than 3.2 million synthetic long reads. The synthetic long reads had an N50 of 4.95 Kb and cover approximately 3.1-4.0X of the estimated Chinook salmon genome size (2.4-3.0 Gb Hardie et al. 2004; Gregory 2017). The observed repeat content in the local assemblies (subsample of 20 barcodes) ranged from 16-26%, which represents a reduction of at least 35% when compared to the global genome repeat content of a Atlantic salmon genome (60%; Lien et al. 2016). The 3.3 million synthetic long reads were used as input to produce an initial whole genome reconstruction using the CELERA assembler (Myers et al. 2000). The initial assembly had a contig N50 of 13.4 Kb and a total size of 2.2 Gb including degenerate contigs. The resulting contigs were scaffolded using mate-pairs libraries with FAST-SG (Di Genova et al. 2017) coupled with OPERA-LG (Gao et al. 2016). FAST-SG was run with 40-mers and using 1,827 million of error corrected mate pair reads as input. Mate pair libraries were error corrected using the merTrim program included in the CELERA assembler package. Then, OPERA-LG produced 114,446 scaffolds (> 1Kb) with N50 of 175.6 Kb and total size of 2.36 Gb. The resulting NGS scaffolds were gapfilled (Nadalin et al. 2012) and cleaned, removing adapter sequences and common DNA contaminations using the UniVec database.

The NGS scaffolds were error-corrected and re-scaffolded using optical maps from the same individual fish using BioNano technology to produce super-scaffolds. First, the optical maps were assembled *de novo* to produce longer molecules. The assembled maps had a total length of 1.9 Gb (N50 0.873 Mb) and 2.3Gb (N50 0.817 Mb) for the BSPQI and BSSSI enzymes, respectively. Both assembled maps were combined with the NGS scaffolds to produce hybrid scaffolds by using software tools provided by BioNano (Access). A total of 792 NGS scaffolds were corrected in the hybrid scaffolding process. The final hybrid assembly had a scaffold N50 of 153 Kb and a total length of 2.36 Gb.

Hybrid scaffolds were ordered into chromosomes using available genetic maps (Brieuc et al. 2014; McKinney et al. 2016) and whole genome alignments against a reference genome of a related species, rainbow trout (*O. mykiss*; Genbank accession MSJN01000000). The rainbow trout and Chinook salmon genomes were repeat masked using WINDOWMASKER (Morgulis et al. 2005). Then, the Chinook salmon scaffolds were aligned to existing rainbow trout chromosomes using MEGABLAST (Altschul et al. 1990). The MEGABLAST alignments were used to determine the best collinear regions for the Chinook salmon scaffolds in rainbow trout chromosomes. When the Chinook salmon scaffolds aligned to multiple rainbow trout chromosomes, the Chinook salmon scaffolds were split when supported by the Chinook salmon genetic maps (Brieuc et al. 2014; McKinney et al. 2016). Then, the Chinook salmon scaffolds were ordered, orientated and concatenated into chromosomes using a greedy algorithm (Moran et al. 1990) that iteratively assigned the Chinook scaffolds to rainbow trout chromosomes while prioritizing longer collinear blocks and avoiding overlaps among the anchored scaffolds. A total of 12,724 scaffolds were anchored to chromosomes. Finally, the Chinook salmon chromosome sequences were ordered and named following genetic maps (McKinney et al. 2016). A total of 1.70Gb of the Chinook assembly was anchored to chromosomes (72.2%). The final chromosome-level assembly has a N50 of 45.4Mb. The genome assembly was submitted to NCBI (accession # PIPH000000000).

Determination of large collinear homolog regions was performed by a self-alignment of the masked Chinook chromosome sequences. A total of 1.156 (34^2) chromosome pair alignments were performed using LASTZ (--identity=80..100 --nogapped --matchcount=100 --nochain --gfextend --format=general; Harris et al. 2007). Then, DAGchainer (-G 10.000 -O 0 -E -3 -S 10000 -D 200000; Haas et al. 2004) was used to identify collinear homolog regions among chromosomes pairs considering both strands. DAGchainer collinear blocks (1,077) were filtered by size (>1Mb) and ordered into non-overlapping regions along each chromosome by taking into account the best-scored blocks and both strands. A total of 184 collinear blocks (total 1.128Gb) were selected and plotted using the CIRCLIZE R-package (Gu et al. 2014). Finally, Identity between homeologous sequences were determined in 1Mb windows dividing the total number of match bases by the total alignment length of the selected LASTZ High-scoring Segment Pair (HSP) of each collinear block.

Transcriptome assembly for gene annotation.

A total of 1,736 million RNA-seq reads were sequenced from eight tissues collected from the same individual fish as used for the genome assembly. Chinook salmon gene models were built integrating evidence from transcriptome predictions (RNA-seq from eight tissues), mRNA alignments, protein alignments and *de novo* predictions. RNA-seq based gene models were built using the HISAT+STRINGTIE- +TRANSDECODER pipeline (Haas et al. 2013; Kim et al. 2015; Pertea et al. 2015) using the repeat masked genome assembly. A total of 41,171 models having a match to curated proteins were obtained. Trinity transcript (NCBI # GGDU00000000) were aligned with GMAP (Wu et al. 2005; min-identity ≥ 98% and coverage ≥ 50%) to the repeat masked genome assembly and reference based models were built with CUFFLINKS (Trapnell et al. 2010) to produce a total of 22,345 models with homology to curated Swissprot proteins.

Protein and mRNA models were built by aligning reference proteins and mRNAs from related salmonid species annotated by NCBI (*O. mykiss* and *Salmo salar*) using GENWISE (Birney et al. 2004) and GMAP. A total of 47,426 reference protein based models and 40,646 reference mRNA based models were obtained from the available salmonid genome annotations. Finally, a total of 31,916 *de novo* predictions covering 50% of a Swissprot protein were obtained using AUGUSTUS (Stanke and Morgenstern 2005).

All collected evidence for gene annotations was clustered in 47,139 non-redundant protein-coding loci using gene model boundaries with BEDTOOLS (Quinlan and Hall 2010) and then models having the largest score (in terms of transcript length and best match with curated swissprot proteins) within a locus were selected. A total of 47,139 protein-coding models were obtained and 35,696 high-quality models cover more than 50% of a curated Swissprot protein (Bairoch et al. 2000). To estimate completeness of the reference assembly, we analyzed the genome with BUSCO (benchmarking universal single-copy orthologs; Waterhouse et al. 2018) using the complete set of genes from the Actinopterygii database (parameters --mode genome --species zebrafish). To support annotation of the genome assembly, gene models were included in a gff3 file available on Dryad.

REFERENCES

Altschul, S. F. *et al.* 1990 Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Bankevich, A. & Pevzner, P. A. 2016 TruSPAdes: barcode assembly of TruSeq synthetic long reads. Nat. Methods 13, 248–250.

Bairoch, A. & Apweiler, R. 2000 The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* 28, 45–48.

Birney, E., Clamp, M. & Durbin, R. 2004 GeneWise and genomewise. *Genome Res.* 14, 988–995.

Brieuc, M.S., Waters, C.D., Seeb, J.E., & Naish, K.A. 2014 A dense linkage map for Chinook salmon (*Oncorhynchus tshawytscha*) reveals variable chromosomal divergence after an ancestral whole genome duplication event. *G3-Genes Genom. Genet.* 4, 447-460.

Di Genova, A. *et al.* 2017 FAST-SG: An alignment-free algorithm for hybrid assembly. *bioRxiv,* 209122.

Gao, S. *et al.* 2016 OPERA-LG: efficient and exact scaffolding of large, repeat-rich eukaryotic genomes with performance guarantees. *Genome Biol.* 17, 102.

Gregory, T. R. 2017 Animal genome size database. http://www.genomesize.com.

Gu, Z., *et al.* 2014 Circlize implements and enhances circular visualization in R. Bioinformatics 30, 2811-2812.

Haas, B.J., *et al.* 2004 DAGchainer: a tool for mining segmental genome duplications and synteny. Bioinformatics 20, 3643-3646.

Haas, B.J. *et al.* 2013 De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protoc.* 8, 1494–1512.

Harris, R.S. 2007. Improved pairwise alignment of genomic DNA. The Pennsylvania State University.

Hardie, D. C. & Hebert P. D. N. 2004 Genome-size evolution in fishes. *Can. J. Fish. Aquat. Sci.* 61, 1636-1646.

Kim, D., Langmead, B. & Salzberg, S. L. 2015 HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360.

McKinney, G.J. *et al.* 2016 An integrated linkage map reveals candidate genes underlying adaptive variation in Chinook salmon (*Oncorhynchus tshawytscha*). *Mol. Ecol. Res.* 16, 769-783.

Morgulis, A. *et al*. 2005 WindowMasker: window-based masker for sequenced genomes. *Bioinformatics* 22, 134–141.

Moran, S., Newman, I. & Wolfstahl, Y. 1990 Approximation algorithms for covering a graph by vertex‐disjoint paths of maximum total weight. *Networks* 20, 55-64.

Myers, E. W., *et al.* 2000 A whole-genome assembly of Drosophila. *Science* 287, 2196-2204.

Nadalin, F., F. Vezzi & Policriti, A. 2012 GapFiller: a de novo assembly approach to fill the gap within paired reads. *BMC Bioinformatics* 13, S8.

Pertea, M. *et al*. 2015 StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295.

Quinlan, A. R., & Hall, I. M. 2010 BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841-842.

Stanke, M. & Morgenstern, B. 2005 AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 33, W465–W467.

Trapnell, C. *et al.* 2010 Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol. 28*, 511-515.

Waterhouse, R.M. *et al.* 2018 BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35, 543-548.

Wu, T. D. & Watanabe, C. K. 2005 GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21, 1859–1875.