DATA NOTE

The genome sequence of the brown trout, *Salmo trutta* Linnaeus 1758 [version 1; peer review: 3 approved]

Tom Hansen[1], Per Gunnar Fjelldal[1], Sigbjørn Lien[2], Michelle Smith[3], Craig Corton[3], Karen Oliver[3], Jason Skelton[3], Emma Betteridge[3], Jale Doulcan[3,4], Olivier Fedrigo[5], Jacquelyn Mountcastle[5], Erich Jarvis[5,6], Shane A. McCarthy[3,7], William Chow[3], Kerstin Howe[3], James Torrance[3], Jonathan Wood[3], Ying Sims[3], Leanne Haggerty[8], Richard Challis[3], Jonathan Threlfall[3], Daniel Mead[3,9], Richard Durbin[3,7], Mark Blaxter[3]

[1]Institute of Marine Research (IMR), Matredal, Norway
[2]Norwegian University of Life Sciences, Ås, 1432, Norway
[3]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK
[4]Achilles Therapeutics plc, London, W6 8PW, UK
[5]The Rockefeller University, New York, New York, 10065, USA
[6]Howard Hughes Medical Institute, Chevy Chase, Maryland, 20815, USA
[7]Department of Genetics, University of Cambridge, Cambridge, CB2 3EH, UK
[8]EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK
[9]Owlstone Medical, Cambridge Science Park, Cambridge, CB4 0GJ, UK

v1 **First published:** 13 May 2021, **6**:108
https://doi.org/10.12688/wellcomeopenres.16838.1
Latest published: 13 May 2021, **6**:108
https://doi.org/10.12688/wellcomeopenres.16838.1

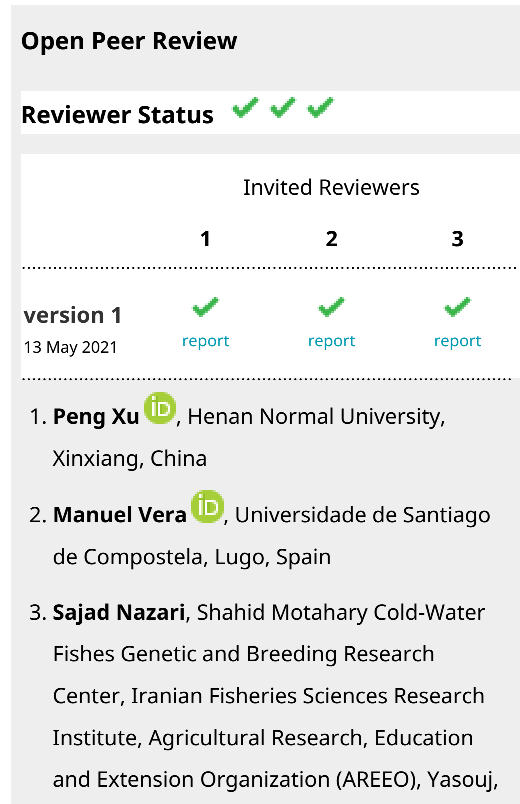
Abstract

We present a genome assembly from an individual female *Salmo trutta* (the brown trout; Chordata; Actinopteri; Salmoniformes; Salmonidae). The genome sequence is 2.37 gigabases in span. The majority of the assembly is scaffolded into 40 chromosomal pseudomolecules. Gene annotation of this assembly on Ensembl has identified 43,935 protein coding genes.

Keywords

Salmo trutta, brown trout, genome sequence, chromosomal

This article is included in the Tree of Life gateway.



Open Peer Review

Reviewer Status ✔ ✔ ✔

	Invited Reviewers		
	1	**2**	**3**
version 1 13 May 2021	✔ report	✔ report	✔ report

1. **Peng Xu**, Henan Normal University, Xinxiang, China
2. **Manuel Vera**, Universidade de Santiago de Compostela, Lugo, Spain
3. **Sajad Nazari**, Shahid Motahary Cold-Water Fishes Genetic and Breeding Research Center, Iranian Fisheries Sciences Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Yasouj,

Iran

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Mark Blaxter (mark.blaxter@sanger.ac.uk)

Author roles: Hansen T: Data Curation, Formal Analysis, Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Fjelldal PG**: Data Curation, Formal Analysis, Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Lien S**: Data Curation, Formal Analysis, Investigation, Resources, Writing – Review & Editing; **Smith M**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Corton C**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Oliver K**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Skelton J**: Formal Analysis, Investigation, Resources, Writing – Review & Editing; **Betteridge E**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Doulcan J**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Fedrigo O**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Mountcastle J**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Jarvis E**: Conceptualization, Project Administration, Resources, Supervision, Validation, Writing – Review & Editing; **McCarthy SA**: Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Review & Editing; **Chow W**: Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Review & Editing; **Howe K**: Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Review & Editing; **Torrance J**: Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Review & Editing; **Wood J**: Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Review & Editing; **Sims Y**: Formal Analysis, Software, Validation, Visualization, Writing – Review & Editing; **Haggerty L**: Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Review & Editing; **Challis R**: Formal Analysis, Methodology, Software, Validation, Visualization, Writing – Review & Editing; **Threlfall J**: Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Mead D**: Conceptualization, Investigation, Project Administration, Writing – Review & Editing; **Durbin R**: Conceptualization, Data Curation, Funding Acquisition, Supervision, Writing – Review & Editing; **Blaxter M**: Conceptualization, Data Curation, Funding Acquisition, Supervision, Writing – Review & Editing

Competing interests: J. Threlfall was a previous employee at F1000Research up until January 2021.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (206194) and the Darwin Tree of Life Discretionary Award (218328). SAM and RD are supported by Wellcome (207492).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Hansen T *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Hansen T, Fjelldal PG, Lien S *et al.* **The genome sequence of the brown trout, *Salmo trutta* Linnaeus 1758 [version 1; peer review: 3 approved]** Wellcome Open Research 2021, **6**:108 https://doi.org/10.12688/wellcomeopenres.16838.1

First published: 13 May 2021, **6**:108 https://doi.org/10.12688/wellcomeopenres.16838.1

Species taxonomy

Metazoa; Chordata; Craniata; Actinopterygii; Actinopteri; Neopterygii; Teleostei; Euteleosteomorpha; Salmoniformes; Salmonidae; Salmoninae; Salmo; *Salmo trutta* Linnaeus 1758 (NCBItxid:8032).

Introduction

The brown trout, *Salmo trutta*, is native to Europe, western Asia and North Africa; however, the species has been successfully introduced to a multitude of other geographical locations (Klemetsen *et al.*, 2003). Genetically similar *S. trutta* can be freshwater residents, freshwater migrants or anadromous (migrating to the sea to feed, only returning to freshwater to breed), leading taxonomists initially to believe that these were multiple independent species. This phenotypic difference has a genetic component but is also partly caused by environmental factors, such as food availability, which lead to changes in gene expression and drives migration and adaptation to different environments (Ferguson *et al.*, 2019). *S. trutta* also exhibit considerable genetic variation within migratory or resident populations; these differences can be seen by populations in different habitats (Ferguson, 1989) or in the same habitat (Andersson *et al.*, 2017). This genetic diversity can allow populations to occupy different environments, such as those with high levels of acidity (Prodöhl *et al.*, 2019).

This reference genome sequence will be of utility for researchers that wish to sample and analyse the genetics of *S. trutta* populations, helping to understand genetic drivers behind migration and the reasons why different populations of brown trout are so well adapted to different conditions. As increases in atmospheric CO_2 continue to increase temperatures and acidify oceans, this information will help conservation of *S. trutta* and other species by revealing which genetic components allow populations to adapt to warmer and more acidic environments.

Genome sequence report

The genome was sequenced from a single female *Salmo trutta* bred at the Institute of Marine Research, Bergen, Norway. A total of 52-fold coverage in Pacific Biosciences single-molecule long reads (N50 19 kb) and 70-fold coverage in 10X Genomics read clouds (from molecules with an estimated N50 of 65 kb) were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data, and 67-fold coverage of Bionano optical maps. Manual assembly curation corrected 175 missing/misjoins, reducing the scaffold number by 4.8% and the assembly length by 0.5%. The final assembly has a total length of 2.37 Gb in 1,441 sequence scaffolds with a scaffold N50 of 52.21 Mb (Table 1). The majority, 91.5%, of the assembly sequence was assigned to 40 chromosomal-level scaffolds, representing 40 autosomes (numbered by sequence length). No sex chromosomes could be identified (Figure 1; Table 2). The assembly has a BUSCO (Simão *et al.*, 2015) completeness of 97.2% using the actinopterygii_odb10 reference set. Genome assembly

Table 1. Genome data for *Salmo trutta*, fSalTru1.1.

Project accession data	
Assembly identifier	fSalTru1.1
Species	*Salmo trutta*
Specimen	fSalTru1
NCBI taxonomy ID	txid8032
BioProject	PRJEB32115
BioSample ID	SAMEA994732
Isolate information	Female, muscle
Raw data accessions	
PacificBiosciences SEQUEL I	ERX3245920, ERX3253848-ERX3253850, ERX3279922-ERX3279929, ERX3288373, ERX3311049-ERX3311054, ERX3311066, ERX3318044-ERX3318049, ERX3338928, ERX3338929
10X Genomics Illumina	ERX3341615-ERX3341622
Hi-C Illumina	ERX4142808-ERX4142812
BioNano	ERZ1395486
Genome assembly	
Assembly accession	GCA_901001165.1
Span (Mb)	2,372
Number of contigs	5,378
Contig N50 length (Mb)	1.7
Number of scaffolds	1441
Scaffold N50 length (Mb)	52.2
Longest scaffold (Mb)	81.5
BUSCO* genome score	C:94.7%[S:49.4%,D:45.3%],F:1.8%, M:3.5%,n:4584
Genome annotation	
Number of protein-coding genes	43,935
Average coding sequence length (bp)	2,058
Average number of exons per gene	13
Average exon size (bp)	210
Average intron size (bp)	2,770

*BUSCO scores based on the actinopterygii_odb10 BUSCO set using v5.0.0. C= complete [S= single copy, D=duplicated], F=fragmented, M=missing, n=number of orthologues in comparison. A full set of BUSCO scores is available at https://blobtoolkit.genomehubs.org/view/Salmo%20trutta/dataset/CAAJIE01/busco.

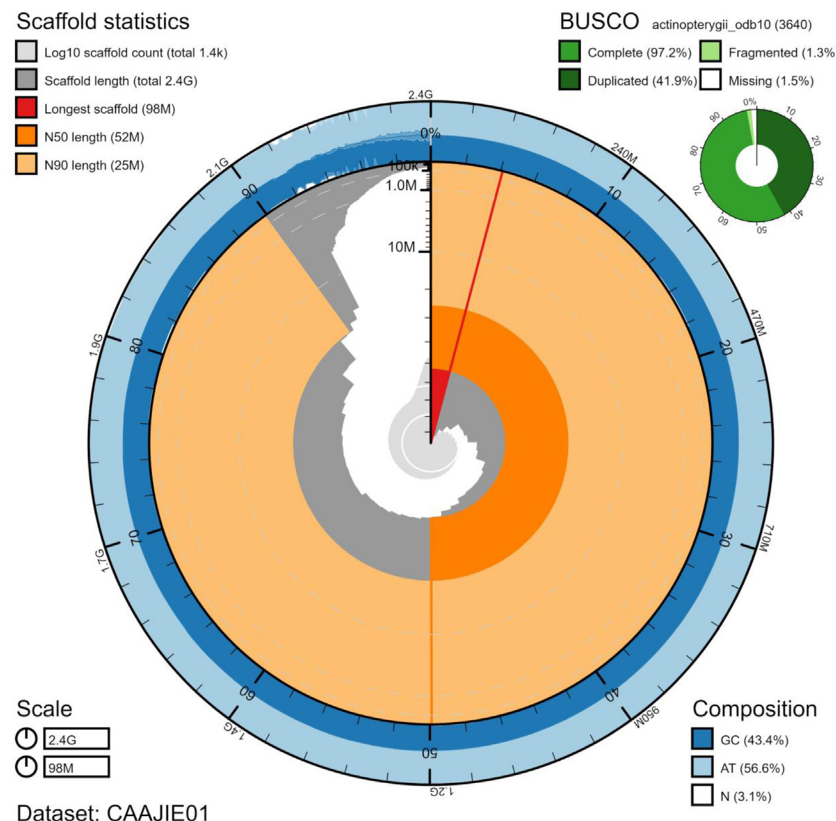


Figure 1. Genome assembly of *Salmo trutta*, fSalTru1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Salmo%20trutta/dataset/CAAJIE01/snail.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Salmo trutta*, fSalTru1.1.

INSDC accession	Chromosome	Size (Mb)	GC%
LR584410.1	1	81.54	43.8
LR584445.1	2	75.35	43.6
LR584416.1	3	74.75	43.6
LR584420.1	4	73.17	43.2
LR584433.1	5	67.76	43.1
LR584406.1	6	60.1	43.5
LR584430.1	7	59.84	43.1
LR584407.1	8	51.19	43.8
LR584409.1	9	49.36	43.5
LR584419.1	10	46.6	43.2
LR584438.1	11	22.96	43.8
LR584441.1	12	97.53	43.8
LR584428.1	13	91.49	43.9
LR584411.1	14	86.25	43.3
LR584415.1	15	66.9	42.9
LR584431.1	16	61.35	43.1
LR584426.1	17	59.76	43.1
LR584435.1	18	59.14	43.1
LR584427.1	19	56.58	43.2
LR584429.1	20	55.16	43.2
LR584437.1	21	52.73	43.4
LR584440.1	22	52.21	43.6
LR584421.1	23	51.49	43.5
LR584412.1	24	50.33	43.2
LR584436.1	25	48.97	43.6
LR584439.1	26	48.7	44
LR584424.1	27	46.41	43.4
LR584422.1	28	46.38	43.5
LR584418.1	29	46.06	43.7
LR584432.1	30	45.79	43.7
LR584423.1	31	45.59	43.1
LR584408.1	32	44.95	43.9
LR584414.1	33	44.89	43.5
LR584434.1	34	42.9	43.9
LR584444.1	35	41.92	43.5
LR584442.1	36	41.68	43.9
LR584417.1	37	35.21	43.8
LR584425.1	38	34.89	43.3
LR584413.1	39	25.83	43.6
LR584443.1	40	25.48	44.1

metrics, GC coverage, cumulative sequence and the Hi-C contact map are visualised in Figure 1–Figure 4, respectively.

Gene annotation

The Ensembl gene annotation system (Aken *et al.*, 2016) was used to generate annotation for the fSalTru1.1 assembly (GCA_901001165.1) (Table 1). Annotation was created primarily through alignment of transcriptomic data to the genome, with gap filling via protein-to-genome alignments of a select set of vertebrate proteins from UniProt (UniProt Consortium, 2019). The resulting Ensembl annotation includes 122,381 transcripts assigned to 43,935 coding and 4,441 non-coding genes (Salmo trutta - Ensembl Rapid Release).

Methods

Owing to the high genetic diversity of brown trout and the variable chromosome numbers (*S. trutta* have 38-42 chromosomes, with multiple copies of these chromosomes), doubled haploid specimens were bred for sequencing and generation of the assembly. The doubled haploid female used in this study was bred on 26 November 2015 at the Institute of Marine Research using a protocol optimized for Atlantic salmon, *Salmo salar* (see (Hansen *et al.*, 2020)). In summary, eggs from one *Salmo trutta* female from a domestic stock that originated from Lake Tunhovd in eastern Norway were fertilized with UV irradiated milt (brown trout sperm diluted 1:40 with sperm fluid and irradiated (254 nm) for 8 mins at

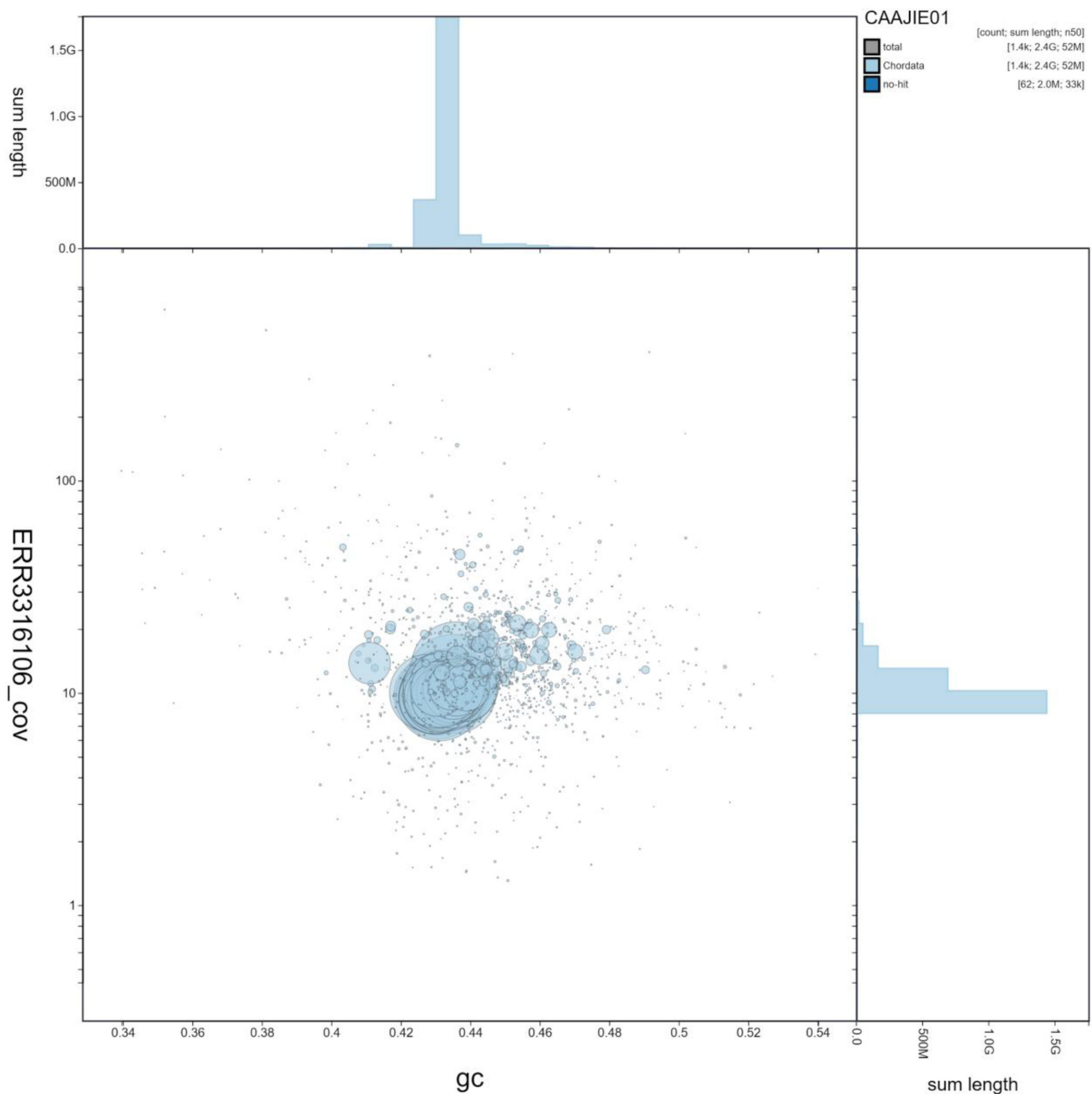


Figure 2. Genome assembly of *Salmo trutta*, fSalTru1.1: GC coverage. BlobToolKit GC-coverage plot. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Salmo%20trutta/dataset/CAAJIE01/blob?plotShape=circle.

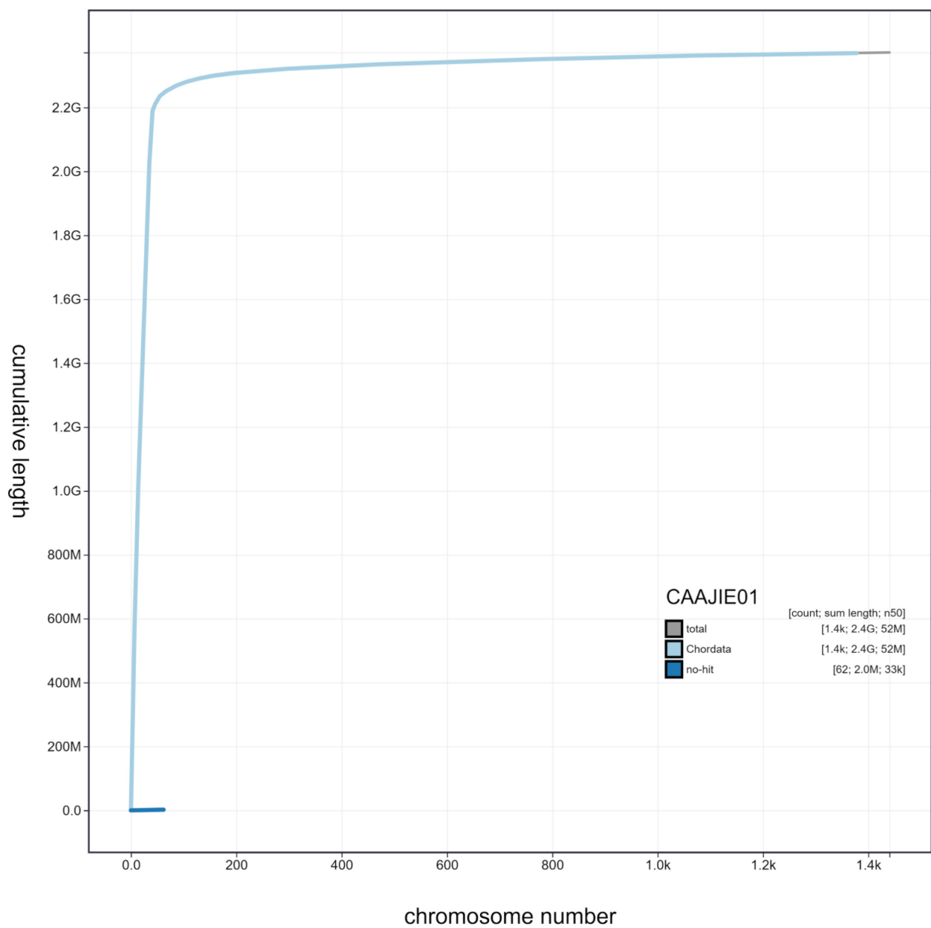


Figure 3. Genome assembly of *Salmo trutta*, fSalTru1.1: cumulative sequence. BlobToolKit cumulative sequence plot. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Salmo%20trutta/dataset/CAAJIE01/cumulative.

0.48 mWcm², activated and left to hydrate in 8°C freshwater in a polyethylene (PE) container. After 4700 min.°C irradiation, the PE bottle was transferred to a pressure chamber and the eggs were subjected to a hydrostatic pressure of 655 bar for 5 mins. The eggs were incubated at approximately 6°C and surviving larvae were fed at 12°C and continuous light until June 2016 when temperature and photoperiod was changed to ambient conditions. On 16 January 2018, one female individual was euthanized (500 mgL− 1 Finquel® (MS 222) and sampled.

The specimen was transferred to the Wellcome Sanger Institute and DNA was extracted using an agarose plug extraction from spleen tissue following the Bionano Prep Animal Tissue DNA Isolation Soft Tissue Protocol.

Sequencing was performed by the Scientific Operations core at the Wellcome Sanger Institute on Pacific Biosciences SEQUEL I and Illumina HiSeq X instruments. Hi-C data were generated using the Arima Hi-C kit v1 by Arima Genomics, San Diego, USA, and sequenced on Illumina HiSeqX. BioNano data were generated in the Rockefeller University Vertebrate Genome laboratory using the Saphyr instrument. Ultra-high molecular weight DNA was extracted using the Bionano Prep Animal Tissue BioNano data were generated in the Rockefeller University Vertebrate Genome laboratory using the Saphyr instrument. Ultra-high molecular weight DNA was extracted using the Bionano Prep Animal Tissue DNA Isolation Fibrous-Tissue Protocol and assessed by pulsed field gel and Qubit 3 fluorimetry. DNA was labeled for Bionano Genomics optical mapping following the Bionano Prep Direct Label and Stain (DLS) Protocol and run on one Saphyr instrument chip flowcell. The total yield of tagged molecules ≥ 150 kb with at least 9 sites was 272.3 Gb (N50 0.28 Mb). A CMAP (Bionano assembly consensus genome map) was *de-novo* assembled using Bionano Solve (see Table 3 for software versions and sources) a total map length of 2.62 Gb and a map N50 of 29.37 Mb.

Assembly was carried out following the Vertebrate Genome Project pipeline v1.0 (Rhie *et al.*, 2020) with Falcon-unzip (Chin *et al.*, 2016) and a first round of scaffolding carried out with 10X Genomics read clouds using scaff10x. Hybrid

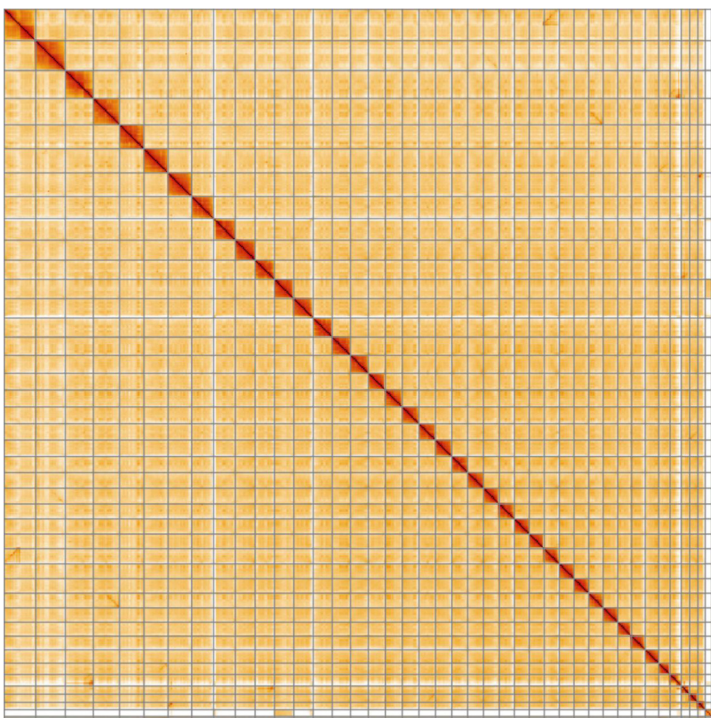


Figure 4. Genome assembly of *Salmo trutta*, fSalTru1.1: Hi-C contact map. Hi-C contact map of the fSalTru1.1 assembly, visualised in HiGlass.

Table 3. Software tools used.

Software tool	Version	Source
Falcon-unzip	falcon-kit 1.2.1	(Chin *et al.*, 2016)
SALSA2	2.1	(Ghurye *et al.*, 2019)
scaff10x	3.0	https://github.com/wtsi-hpag/Scaff10X
arrow	GenomicConsensus 2.2.2	https://github.com/PacificBiosciences/GenomicConsensus
longranger align	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
freebayes	1.1.0-3-g961e5f3	(Garrison & Marth, 2012)
bcftools consensus	1.9	http://samtools.github.io/bcftools/bcftools.html
Bionano Solve	3.2.2_08222018	https://bionanogenomics.com/downloads/bionano-solve/
HiGlass	1.11.6	(Kerpedjiev *et al.*, 2018)
PretextViewer	0.0.4	https://github.com/wtsi-hpag/PretextView
gEVAL	N/A	(Chow *et al.*, 2016)
BlobToolKit	1.2	(Challis *et al.*, 2020)

scaffolding was performed using the BioNano DLE-1 data and BioNano Solve. Scaffolding with Hi-C data (Rao *et al.*, 2014) was carried out with SALSA2 (Ghurye *et al.*, 2019). The Hi-C scaffolded assembly was polished with arrow using the PacBio data, then polished with the 10X Genomics Illumina data by aligning to the assembly with longranger align, calling variants with freebayes (Garrison & Marth, 2012) and applying homozygous non-reference edits using

bcftools consensus. Two rounds of the Illumina polishing were applied. The assembly was checked for contamination and corrected. Manual curation was performed as described previously (Howe *et al.*, 2021) using the gEVAL system (Chow *et al.*, 2016), Bionano Access, HiGlass and Pretext. Figure 1–Figure 3 and BUSCO values were generated using BlobToolKit (Challis *et al.*, 2020).

Data availability

Underlying data

BioProject: Salmo trutta RefSeq Genome, Accession number PRJNA550988: https://www.ncbi.nlm.nih.gov/bioproject/550988

The genome sequence is released openly for reuse. The *S. trutta* genome sequencing initiative is part of the Darwin Tree of Life (DToL) project and the Vertebrate Genome Project (VGP) ordinal references programme. All raw data and the assembly have been deposited in INSDC databases. Raw data and assembly accession identifiers are reported in Table 1.

Acknowledgements

We thank Mike Stratton and Julia Wilson for their support for the 25 genomes for 25 years project.

References

Aken BL, Ayling S, Barrell D, *et al.*: **The Ensembl Gene Annotation System.** *Database (Oxford).* 2016; **2016**: baw093.
PubMed Abstract | Publisher Full Text | Free Full Text

Andersson A, Jansson E, Wennerström L, *et al.*: **Complex Genetic Diversity Patterns of Cryptic, Sympatric Brown Trout (*Salmo Trutta*) Populations in Tiny Mountain Lakes.** *Conserv Genet.* 2017; **18**(5): 1213–27.
Publisher Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit - Interactive Quality Assessment of Genome Assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–74.
PubMed Abstract | Publisher Full Text | Free Full Text

Chin CS, Peluso P, Sedlazeck FJ, *et al.*: **Phased Diploid Genome Assembly with Single-Molecule Real-Time Sequencing.** *Nat Methods.* 2016; **13**(12): 1050–54.
PubMed Abstract | Publisher Full Text | Free Full Text

Chow W, Brugger K, Caccamo M, *et al.*: **gEVAL - a Web-Based Browser for Evaluating Genome Assemblies.** *Bioinformatics.* 2016; **32**(16): 2508–10.
PubMed Abstract | Publisher Full Text | Free Full Text

Ferguson A: **Genetic Differences among Brown Trout, *Salmo Trutta*, Stocks and Their Importance for the Conservation and Management of the Species.** *Freshw Biol.* 1989; **21**(1): 35–46.
Publisher Full Text

Ferguson A, Reed TE, Cross TF, *et al.*: **Anadromy, Potamodromy and Residency in Brown Trout *Salmo Trutta*: The Role of Genes and the Environment.** *J Fish Biol.* 2019; **95**(3): 692–718.
PubMed Abstract | Publisher Full Text | Free Full Text

Garrison E, Marth G: **Haplotype-Based Variant Detection from Short-Read Sequencing.** arXiv: 1207.3907. 2012.
Reference Source

Ghurye J, Rhie A, Walenz BP, *et al.*: **Integrating Hi-C Links with Assembly Graphs for Chromosome-Scale Assembly.** *PLoS Comput Biol.* 2019; **15**(8): e1007273.
PubMed Abstract | Publisher Full Text | Free Full Text

Hansen TJ, Penman D, Glover KA, *et al.*: **Production and Verification of the First Atlantic Salmon (*Salmo Salar* L.) Clonal Lines.** *BMC Genet.* 2020; **21**(1): 71.
PubMed Abstract | Publisher Full Text | Free Full Text

Howe K, Chow W, Collins J, *et al.*: **Significantly Improving the Quality of Genome Assemblies through Curation.** *Gigascience.* 2021; **10**(1): giaa153.
PubMed Abstract | Publisher Full Text | Free Full Text

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: Web-Based Visual Exploration and Analysis of Genome Interaction Maps.** *Genome Biol.* 2018; **19**(1): 125.
PubMed Abstract | Publisher Full Text | Free Full Text

Klemetsen A, Amundsen PA, Dempson JB, *et al.*: **Atlantic Salmon *Salmo Salar* L., Brown Trout *Salmo Trutta* L. and Arctic Charr *Salvelinus Alpinus* (L.): A Review of Aspects of Their Life Histories.** *Ecol Freshw Fish.* 2003; **12**(1): 1–59.
Publisher Full Text

Prodöhl PA, Ferguson A, Bradley CR, *et al.*: **Impacts of Acidification on Brown Trout *Salmo Trutta* Populations and the Contribution of Stocking to Population Recovery and Genetic Diversity.** *J Fish Biol.* 2019; **95**(3): 719–42.
PubMed Abstract | Publisher Full Text | Free Full Text

Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping.** *Cell.* 2014; **159**(7): 1665–80.
PubMed Abstract | Publisher Full Text | Free Full Text

Rhie A, McCarthy SA, Fedrigo O, *et al.*: **Towards Complete and Error-Free Genome Assemblies of All Vertebrate Species.** *bioRxiv.* 2020; 2020.05.22.110833.
Publisher Full Text

Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs.** *Bioinformatics.* 2015; **31**(19): 3210–12.
PubMed Abstract | Publisher Full Text

UniProt Consortium: **UniProt: A Worldwide Hub of Protein Knowledge.** *Nucleic Acids Res.* 2019; **47**(D1): D506–15.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 14 October 2021

https://doi.org/10.21956/wellcomeopenres.18573.r46064

© **2021 Nazari S.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The author(s) is/are employees of the US Government and therefore domestic copyright protection in USA does not apply to this work. The work may be protected under the copyright laws of other jurisdictions when used in those jurisdictions.

Sajad Nazari
Shahid Motahary Cold-Water Fishes Genetic and Breeding Research Center, Iranian Fisheries Sciences Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Yasouj, Iran

The brown trout, *Salmo trutta*, is native to Europe, western Asia and North Africa and is an important fish across these regions. The authors improved the reference genome of *Salmo trutta* using PacBio, Hi-C sequencing technologies which means a much more complete chromosome-level assembly can be feasibly obtained.
Based on the new assembly, genome analysis was then performed on a female individual.
The manuscript did a great job demonstrating successful high-quality chromosome level analysis in a non-model species. In particular, the Introduction provides an excellent backdrop to the findings of the paper. Some figures are clear and concise, and the analyses are sufficiently well described in the methods to enable the reader to fully understand what was done.
In general, the manuscript was clearly written and the analytical methods were sound. I have only a few minor concerns about the paper.

1. Sex determination should be an extremely simple trait. Is it due to genetics or assembly error?

2. There are quite a few inconsistencies between the genetic map and the assembly. Careful checking is needed to make sure the inconsistencies are not due to assembly errors

3. The BUSCO is outdated version. Please consider using BUSCO (Manni *et al.*, 2021).[1]

References

1. Manni M, Berkeley MR, Seppey M, Simão FA, et al.: BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes.*Mol Biol Evol*. 2021; **38** (10): 4647-4654 PubMed Abstract |

Publisher Full Text

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Analysis of the next-generation DNA sequence data, genome assembly, genome annotation

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 01 October 2021

https://doi.org/10.21956/wellcomeopenres.18573.r46069

© 2021 Vera M. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuel Vera
Department of Zoology, Genetics and Physical Anthropology, ACUIGEN group, Faculty of Veterinary, Universidade de Santiago de Compostela, Lugo, Spain

The present data note represents the reference genome sequence of the brown trout, *Salmo trutta*. The population/conservation genetics of this species has been widely studied for the last four decades. The next generation sequencing (NGS) techniques together with Genotyping-by-Sequencing (GBS) methodologies (e.g. RADseq) have allowed the studies of population genomics in the species, but its reference genome was necessary. Thus, this genomic resource, with a very high quality, represents a valuable tool for geneticists interested in *S. trutta* and other salmonids. The interest and workflow (including the protocols) of the manuscript are well presented. Methodology is well described and detailed, allowing their reproducibility by other researchers. Finally, all the data is freely available with useful links (some of them interactive) which facilitates its use and increases the interest of the manuscript.

I have just a few comments to the authors:

- Page 3, "Introduction" section: Replace "...(Ferguson *et al.,* 2019). *S. trutta* also exhibit..." by "...(Ferguson *et al.,* 2019). *Salmo trutta* also exhibit..."

- Page 3, "Genome sequence report" section: Authors stated that "The majority, 91.5%, of the assembly sequence was assigned to 40 chromosomal-level scaffolds, representing 40 autosomes (numbered by sequence length)". However, chromosome names shown on Table 2 are not numbered by size (e.g. The length of chromosome 12 (97.53 Mb) is longer than those shown for chromosomes 1-11. Do the authors have information about the relationship among their chromosomal pseudomolecules and the physical chromosomes?

- Page 6, second column: Replace "...was extracted using the Bionano Prep Animal Tissue BioNano data were generated..." by "...was extracted using the Bionano Prep Animal Tissue. BioNano data were generated..." (a dot between the two sentences is missed).

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Population/Conservation genetics of aquatic organisms

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 01 June 2021

https://doi.org/10.21956/wellcomeopenres.18573.r43865

© 2021 Xu P. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peng Xu
College of Fishery, Henan Normal University, Xinxiang, China

This Data Note reported a high-quality chromosome level *Salmo trutta* genome used PacBio, Hi-C and Biono optical maps. The quality of *Salmo trutta* genome is very sound and it will be helpful in the analysis of the genetics of *S. trutta* populations. In this manuscript, the rationale for creating

the dataset(s) was described clearly. The protocols and work are technically appropriate. The details of the method and materials are sufficient to allow replication by others. The data links were clearly presented in a useable and accessible format.

I have only one suggestion as follows:

- In the section of “Genome sequence report” Line 8, please explain the method that you identify the sex chromosome?

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Fish genomics and genetics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.