

Chinook salmon genome assembly pipeline

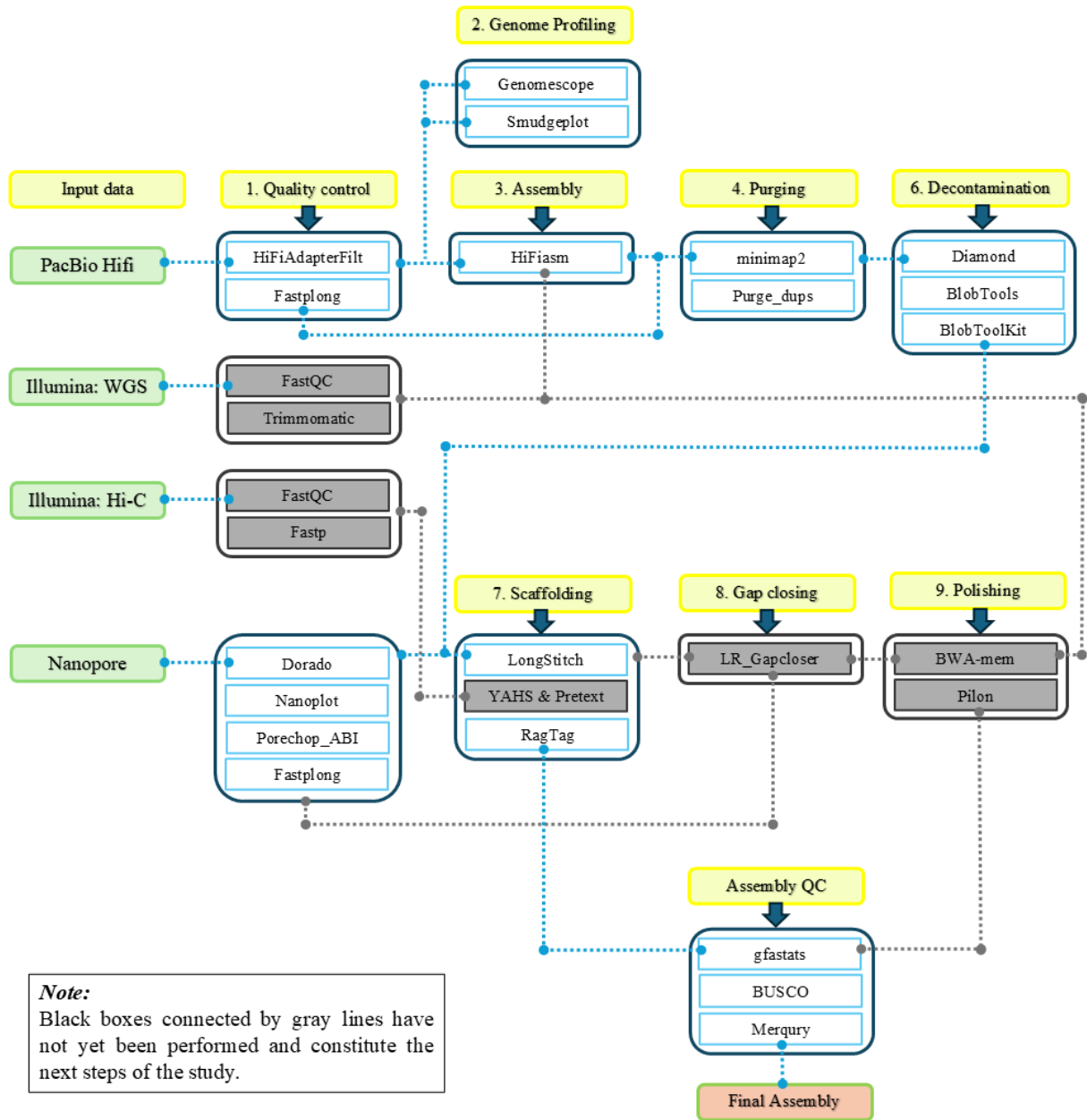


Table 1. List of softwares used to assemble the genome of Chinook salmon.

Step	Software	Version	Parameters	References
Quality control	HiFiAdapterFilt	v3.0.0	default	Sim et al., 2022
	Fastplong	v0.5.0	PacBio HiFi reads: -q 20, -l 5000 Nanopore reads: -q 10 -l, 10000	Chen, 2023
	Dorado	0.9.6+0949eb8d	basecaller --no-trim, --device cuda:all, --emit-fastq, --recursive dna_r10.4.1_e8.2_400bps_sup@v5.0.0	Oxford Nanopore Technologies, 2025
	Nanoplot		default	De Coster et al., 2018
	Porechop_ABI	0.5.0	-abi, --guess_adapter_only	Bonenfant et al., 2022
Genome profiling	Genomescope	v2.0	-k 51, -p 2	Ranallo-Benavidez et al., 2020
	Smudgeplot	v0.4.0	default	Ranallo-Benavidez et al., 2020
Assembly	HiFiasm	0.21.0-r686	--primary, -k 51	Cheng et al., 2021
Purging	Minimap2	2.30-r1287	read depth analysis: -I 4g, -x map-hifi self-alignment: -x asm5	Li, 2018
	Purge_dups	v1.2.6	cutoffs: -m 30, -u 200	Guan et al., 2020
Decontamination	Diamond	v2.0.13	blastx, --db diamondnr.dmnd, --evaluate 1e-5, --max-hsps 1, --max-target-seqs 1, -outfmt 6 qseqid sseqid bitscore, --more-sensitive	Buchfink et al., 2015
	BlobTools	v1.1.1	Taxify: -m prot.accession2taxid	Laetsch and Blaxter, 2017
	BlobToolKit	v4.4.4	default	Challis et al., 2020
Scaffolding	LongStitch	v1.0.5	k=32, w=500 (with seven rounds of scaffolding)	Coombi et al., 2021
	RagTag	v.2.1.0	default	Alonge et al., 2022
Assembly QC	Gfastats	v1.3.11	default	Formenti et al., 2022
	BUSCO	v6.0.0	-l actinopterygii_odb10, -m geno, --augustus	Tegenfeldt et al., 2025
	Merquy	v1.3	default	Rhie et al., 2020

References

- Alonge M., Lebeigle L., Kirsche M., Jenike K., Ou S., et al. (2022) Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.*, 23, Article 258. <https://doi.org/10.1186/s13059-022-02823-7>
- Bonenfant Q., Noé L., and Touzet H. (2022) Porechop_ABI: Discovering unknown adapters in Oxford Nanopore Technology sequencing reads for downstream trimming. *Bioinform. Adv.*, 3(1), vbac085. <https://doi.org/10.1093/bioadv/vbac085>
- Buchfink B., Xie C., and Huson D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12(1), 59–60. <https://doi.org/10.1038/nmeth.3176>
- Challis R., Richards E., Rajan J., Cochrane G., and Blaxter M. (2020) BlobToolKit – interactive quality assessment of genome assemblies. *G3 (Bethesda)*, 10(4), 1361–1374. <https://doi.org/10.1534/g3.119.400908>
- Chen S. (2023) Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta*, 2, e107. <https://doi.org/10.1002/imt2.107>
- Cheng H., Concepcion G.T., Feng X., Zhang H., and Li H. (2021) Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods*, 18:170–175. <https://doi.org/10.1038/s41592-020-01056-5>
- Coombe L., Li J.X., Lo T., Wong J., Nikolic V., et al. (2021) LongStitch: high-quality genome assembly correction and scaffolding using long reads. *BMC Bioinformatics*, 22, 534. <https://doi.org/10.1186/s12859-021-04451-7>
- De Coster W., D’Hert S., Schultz D.T., Cruts M., and Van Broeckhoven, C. (2018) NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics*, 34(15), 2666–2669. <https://doi.org/10.1093/bioinformatics/bty149>
- Formenti G., Abueg L., Brajuka A., Brajuka N., Gallardo-Alba C., et al. (2022) Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs. *Bioinformatics*, 38(17), 4214–4216. <https://doi.org/10.1093/bioinformatics/btac460>
- Guan D., McCarthy S.A., Wood J., Howe K., Wang Y., et al. (2020) Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, 36(9), 2896–2898. <https://doi.org/10.1093/bioinformatics/btaa025>
- Laetsch D.R., and Blaxter M. L. (2017) BlobTools: Interrogation of genome assemblies. *F1000Research*, 6, 1287. <https://doi.org/10.12688/f1000research.12232.1>
- Li H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34, 3094–3100. [doi:10.1093/bioinformatics/bty191](https://doi.org/10.1093/bioinformatics/bty191)
- Oxford Nanopore Technologies (2025). Dorado (Version 1.3.1) [Basecalling software]. <https://nanoporetech.com/software/other/dorado>
- Ranallo-Benavidez T.R., Jaron K.S. and Schatz M.C. (2020) GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.*, 11, 1432. <https://doi.org/10.1038/s41467-020-14998-3>
- Rhie A., Walenz B.P., Koren S., and Phillippy A.M. (2020) Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.*, 21, Article 245. <https://doi.org/10.1186/s13059-020-02134-9>
- Sim S.B., Corpuz R.L., Simmonds T.J., and Geib S.M. (2022) HiFiAdapterFilt, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. *BMC Genomics*, 23(1), 157. <https://doi.org/10.1186/s12864-022-08375-1>
- Simão F.A., Waterhouse R.M., Ioannidis P., Kriventseva E.V., and Zdobnov E.M. (2015) BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Tegenfeldt F., Kuznetsov D., Manni M., Berkeley M., Zdobnov E. M., et al. (2025) OrthoDB and BUSCO update: Annotation of orthologs with wider sampling of genomes. *Nucleic Acids Res.*, 53(D1), D516–D522. <https://doi.org/10.1093/nar/gkae987>