ERGA Assembly Report

v24.10.15

Tags: Other[INVALID TAG]

TxID	8036
ToLID	**fSalAlp3_combined**
Species	Salvelinus alpinus
Class	Actinopteri
Order	Salmoniformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,914,103,308	4,273,955,860
Haploid Number	40 (source: direct)	37
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 5.7.Q47

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

. Observed Haploid size (bp) has >20% difference with Expected
. Observed Haploid Number is different from Expected

. BUSCO single copy value is less than 90% for collapsed
. BUSCO duplicated value is more than 5% for collapsed
. More than 1000 gaps/Gbp for collapsed
. Not 90% of assembly in chromosomes for collapsed

Curator notes

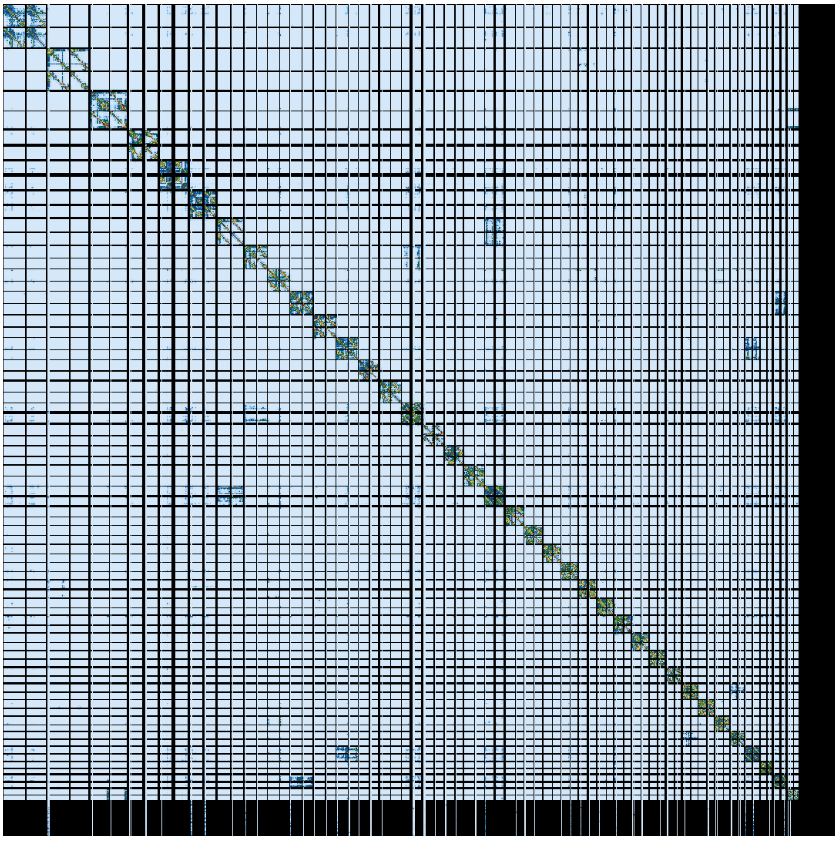
. Interventions/Gb: 424
. Contamination notes: ""
. Other observations: "The assembly of Salvelinus alpinus (fSalAlp3) was generated as part of the Vertebrate Genomes Project (https://vertebrategenomesproject.org/). Both haplotypes were analyzed and manuallyimproved using Pretext. Chromosomes 7 and 19, 8 and 15, 30 and 33 were initially scaffolded together due to their high similarity and were manually separated. As a result, their exact structure is uncertain. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	4,273,902,426	4,273,955,860
GC %	43.38	43.38
Gaps/Gbp	1,881.42	1,949.01
Total gap bp	1,608,200	1,666,000
Scaffolds	3,905	3,633
Scaffold N50	50,125,689	50,032,745
Scaffold L50	30	32
Scaffold L90	255	81
Contigs	11,946	11,963
Contig N50	973,147	967,602
Contig L50	1,019	1,024
Contig L90	5,537	5,570
QV	47.7576	47.7562
Kmer compl.	95.5836	95.5836
BUSCO sing.	6.5%	6.5%
BUSCO dupl.	90.0%	90.2%
BUSCO frag.	1.2%	1.2%
BUSCO miss.	2.3%	2.2%

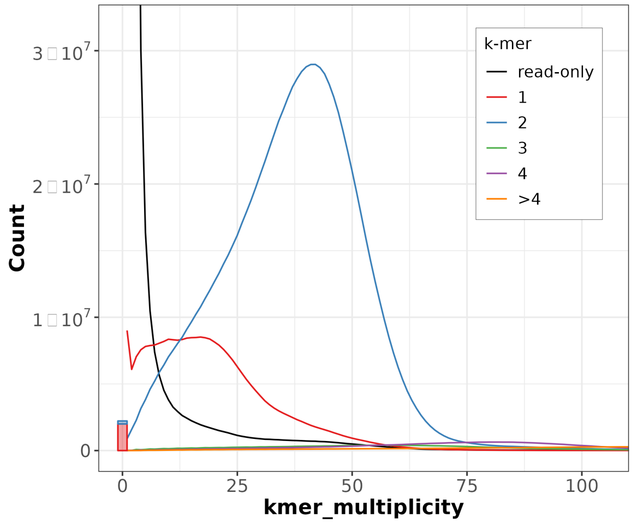
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

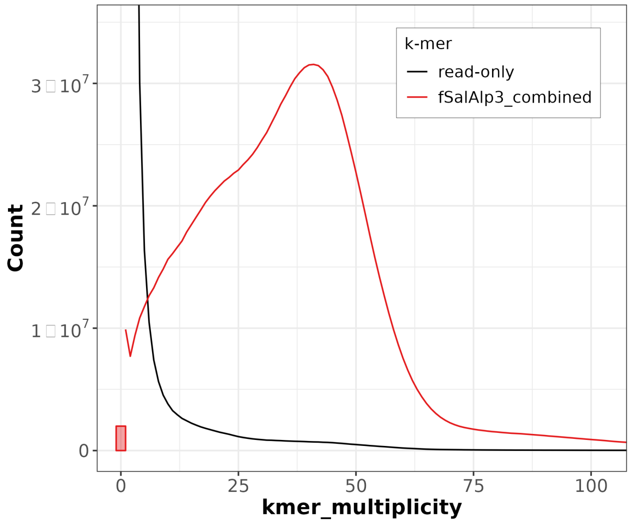


collapsed [LINK]

K-mer spectra of curated assembly

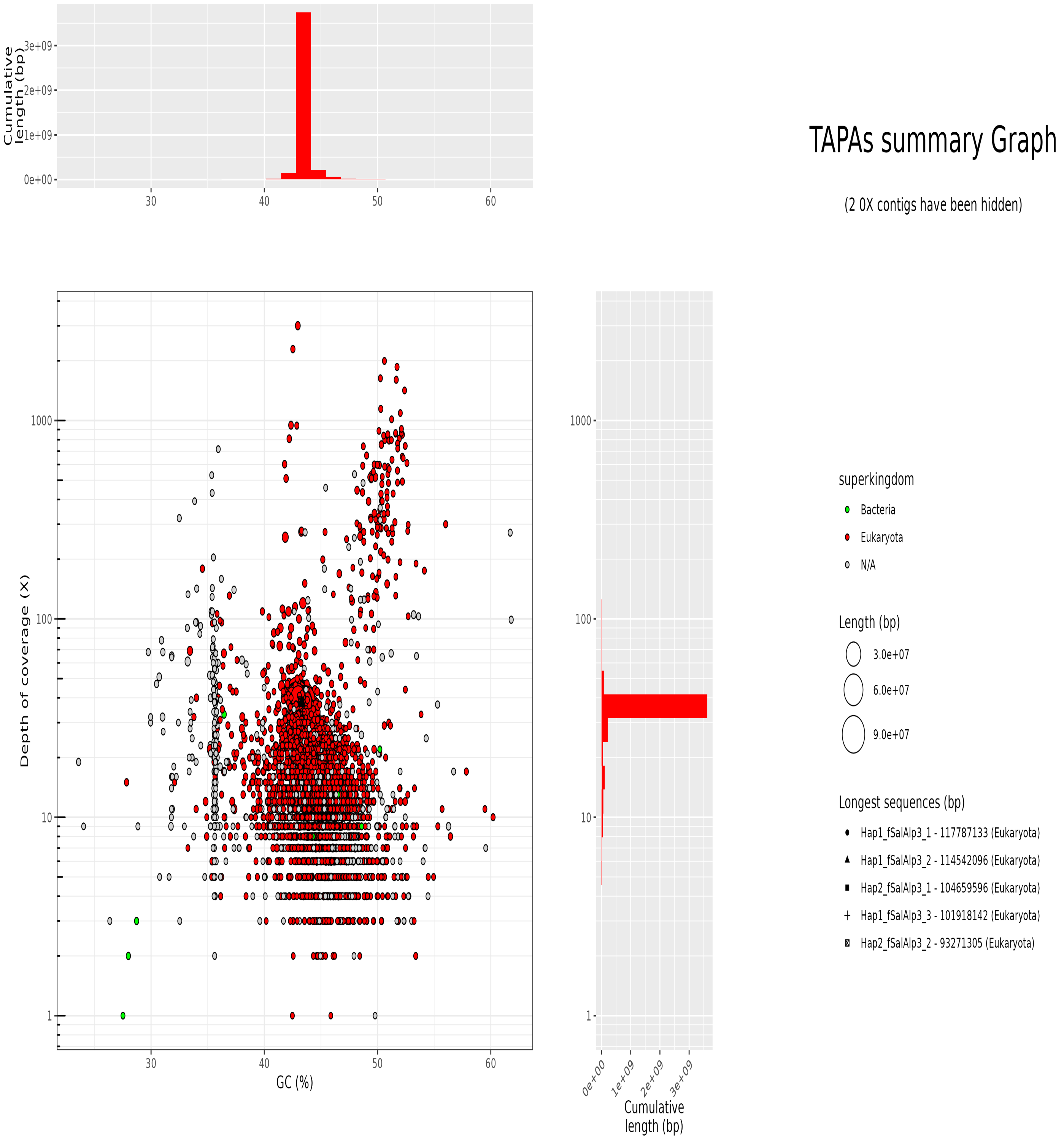


Distribution of k-mer counts per copy numbers found in asm



Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



collapsed. Bubble plot circles are scaled by sequence length, positioned by coverage and GC proportion, and coloured by taxonomy. Histograms show total assembly length distribution on each axis.

Data profile

Data	PACBIO Hifi	Arima
Coverage	NA	NA

Assembly pipeline

- **Hifiasm**
    |_ *ver:* 0.19.5-r593
    |_ *key param:* NA
- **purge_dups**
    |_ *ver:* 1.2.5
    |_ *key param:* NA
- **YaHS**
    |_ *ver:* 1.2
    |_ *key param:* NA

Curation pipeline

- **PretextMap**
    |_ *ver:* 0.1.9
    |_ *key param:* NA
- **PretextView**
    |_ *ver:* 0.2.5
    |_ *key param:* NA

Submitter: Benjamin Istace
Affiliation: Genoscope

Date and time: 2025-09-08 08:37:17 CEST