ERGA Assembly Report

v24.10.15

Tags: Other[INVALID TAG]

TxID	8036
ToLID	**fSalAlp3_hap2**
Species	Salvelinus alpinus
Class	Actinopteri
Order	Salmoniformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,914,103,308	2,009,174,406
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.Q48

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

. Observed Haploid Number is different from Expected

. Kmer completeness value is less than 90 for collapsed
. 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

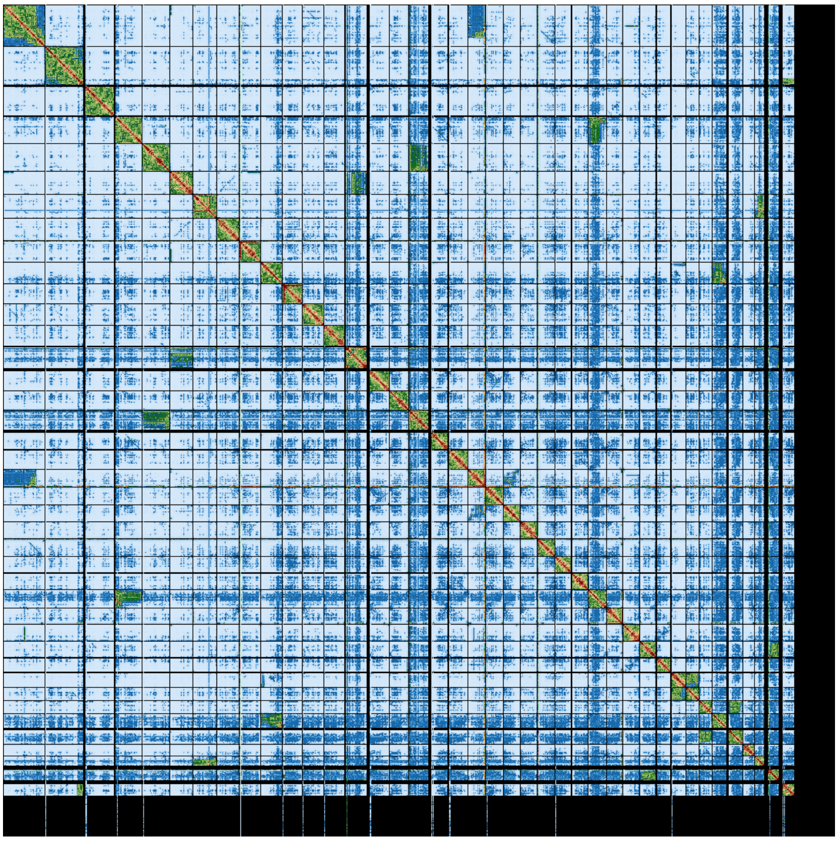
. Interventions/Gb: 116
. Contamination notes: ""
. Other observations: "The assembly of Salvelinus alpinus (fSalAlp3) was generated as part of the Vertebrate Genomes Project (https://vertebrategenomesproject.org/). Both haplotypes were analyzed and manually improved using Pretext. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	2,004,758,796	2,009,174,406
GC %	43.32	43.32
Gaps/Gbp	1,939.39	2,133.71
Total gap bp	777,600	857,400
Scaffolds	2,183	1,803
Scaffold N50	41,528,662	46,675,801
Scaffold L50	17	17
Scaffold L90	194	38
Contigs	6,071	6,090
Contig N50	914,978	914,612
Contig L50	501	504
Contig L90	2,830	2,859
QV	48.2022	48.1876
Kmer compl.	84.3092	84.0459
BUSCO sing.	55.6%	53.9%
BUSCO dupl.	36.2%	40.4%
BUSCO frag.	2.1%	1.3%
BUSCO miss.	6.1%	4.3%

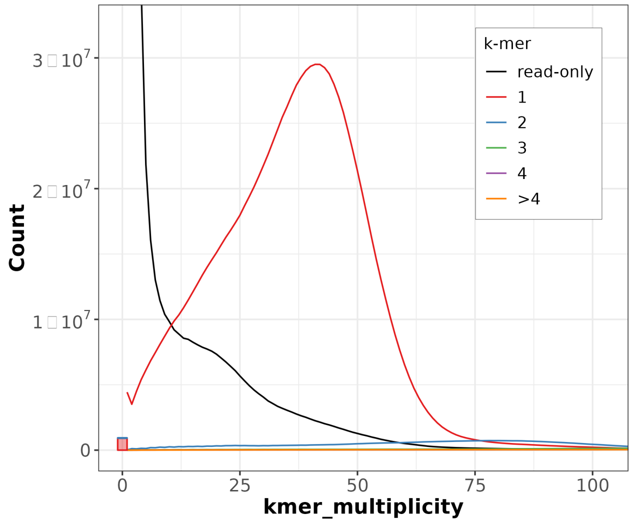
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

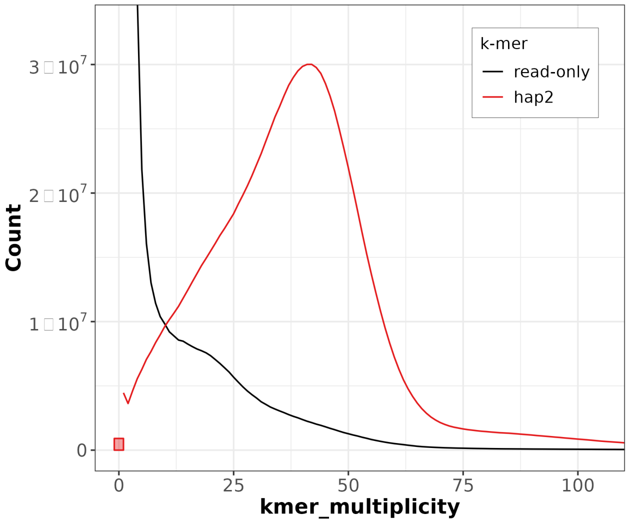


collapsed [LINK]

K-mer spectra of curated assembly

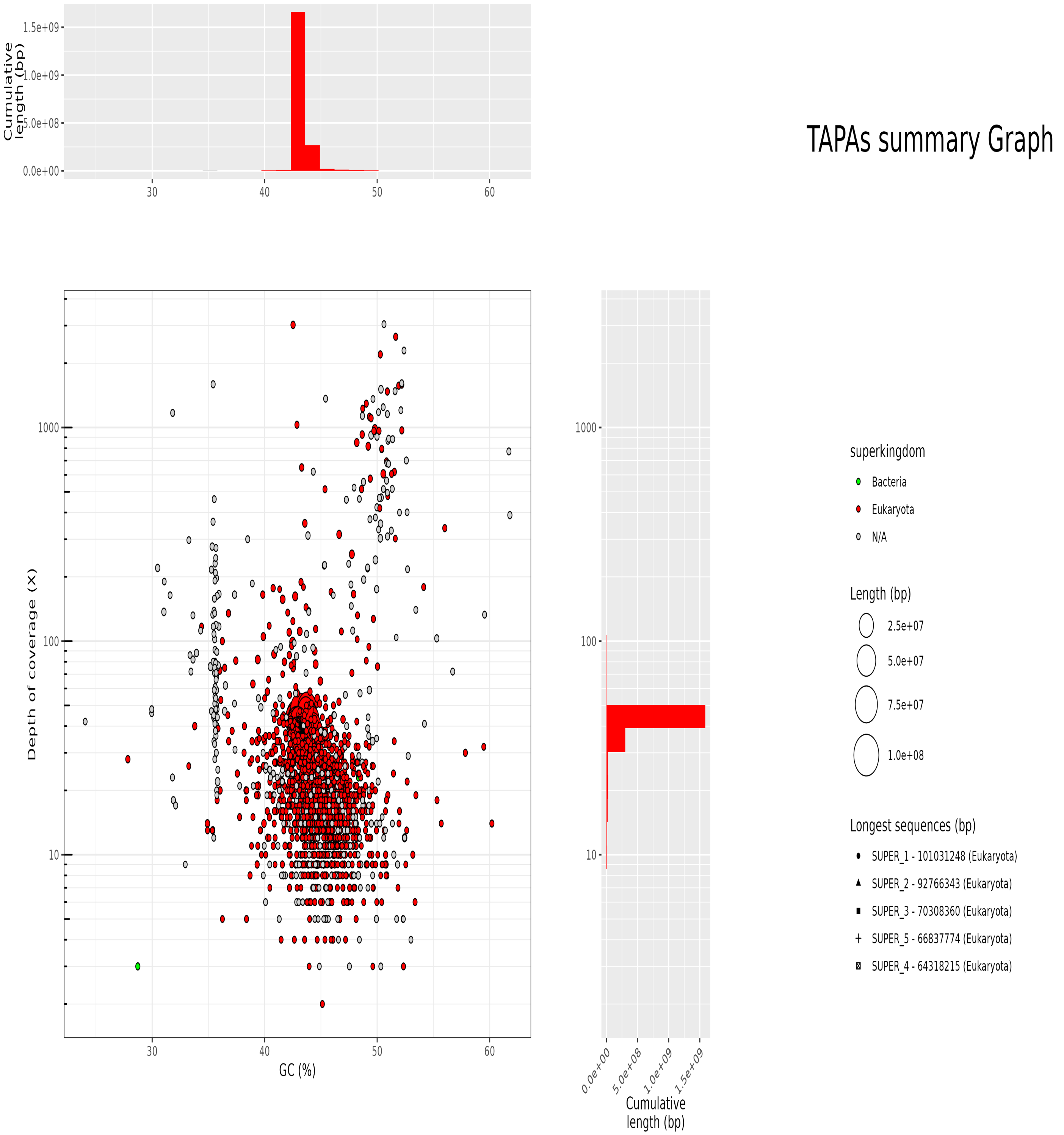


Distribution of k-mer counts per copy numbers found in asm



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	Long reads	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: 2025-11-29 00:56:24 CET