ERGA Assembly Report

v24.10.15

Tags: Other[INVALID TAG]

TxID	27695
ToLID	**fProLin1_hap2**
Species	Prochilodus lineatus
Class	Actinopteri
Order	Characiformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,179,458,285	1,421,922,085
Haploid Number	27 (source: direct)	27
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q57

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

. Observed Haploid size (bp) has >20% difference with Expected

. Kmer completeness value is less than 90 for collapsed

Curator notes

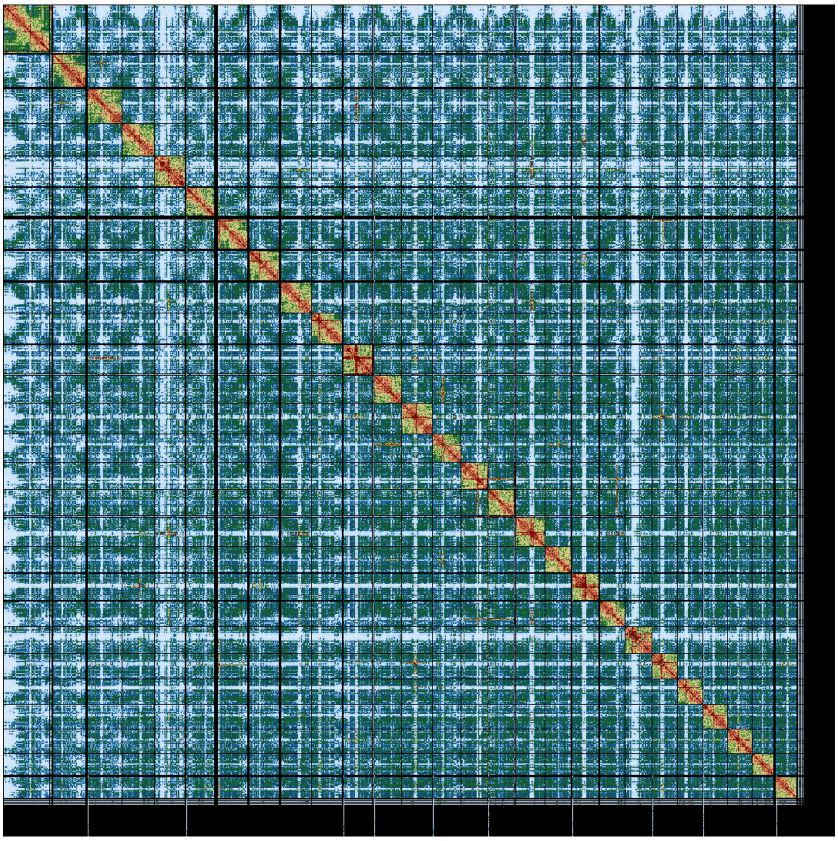
. Interventions/Gb: 23
. Contamination notes: ""
. Other observations: "The assembly of Prochilodus lineatus (fProLin1) was generated as part of the Vertebrate Genomes Project (https://vertebrategenomesproject.org/). Both haplotypes were analyzed and manually improved using Pretext. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,436,976,027	1,421,922,085
GC %	42.02	42.03
Gaps/Gbp	658.33	683.58
Total gap bp	189,200	194,400
Scaffolds	589	565
Scaffold N50	46,199,413	48,754,244
Scaffold L50	14	13
Scaffold L90	30	26
Contigs	1,535	1,537
Contig N50	6,324,622	6,445,832
Contig L50	59	58
Contig L90	466	459
QV	57.4096	57.1984
Kmer compl.	75.1856	74.8062
BUSCO sing.	94.5%	96.7%
BUSCO dupl.	1.3%	1.8%
BUSCO frag.	1.3%	0.3%
BUSCO miss.	3.0%	1.2%

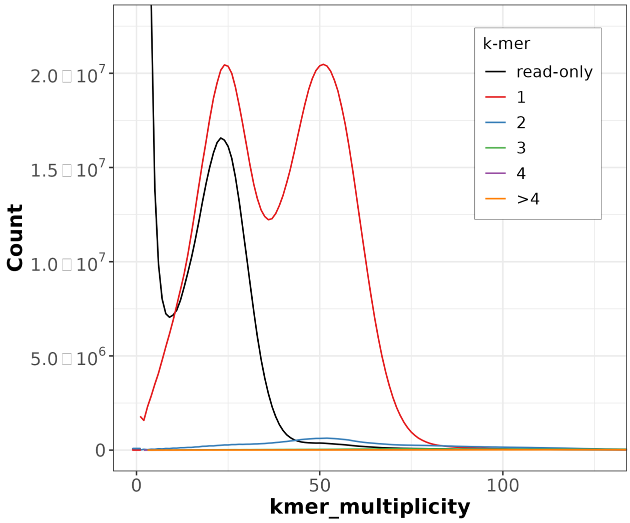
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

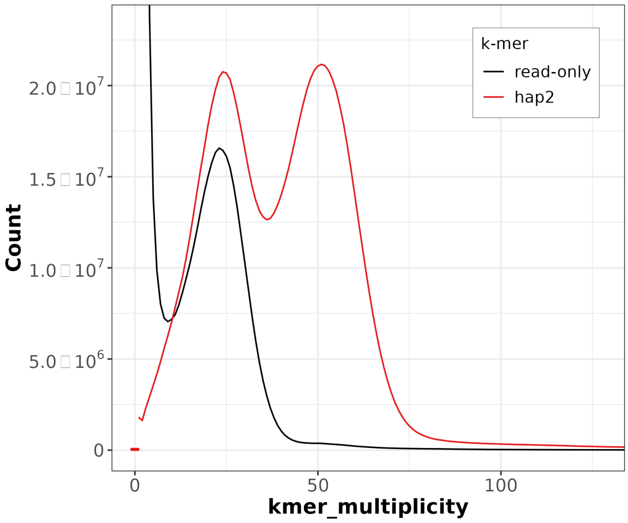


collapsed [LINK]

K-mer spectra of curated assembly

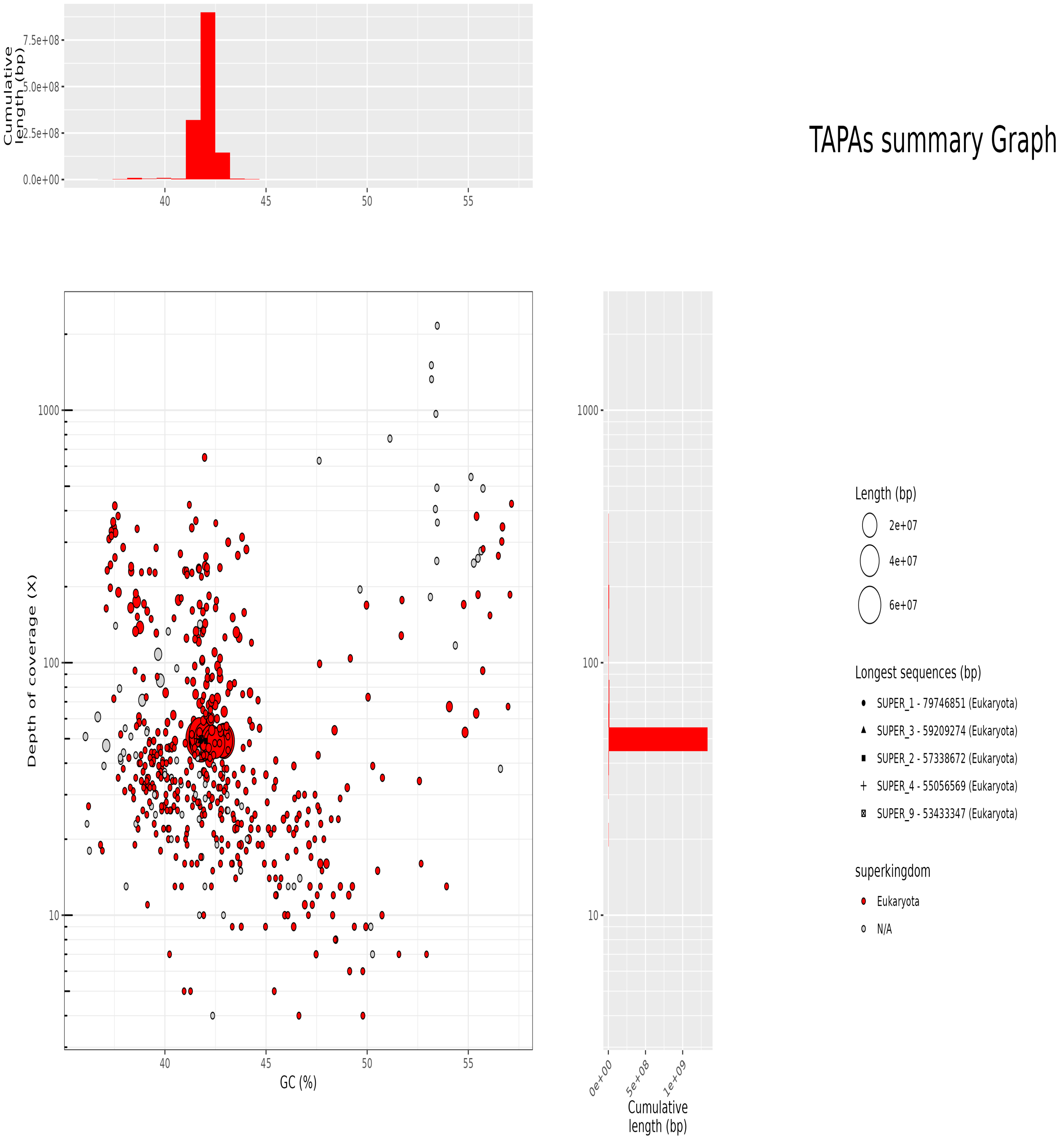


Distribution of k-mer counts per copy numbers found in asm



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: 2025-12-04 05:04:59 CET