ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	206100
ToLID	**fCapApe1**
Species	Capros aper
Class	Actinopteri
Order	Caproiformes

Genome Traits	Expected	Observed
Haploid size (bp)	618,965,019	649,617,890
Haploid Number	24 (source: ancestor)	19
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q43

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

Curator notes

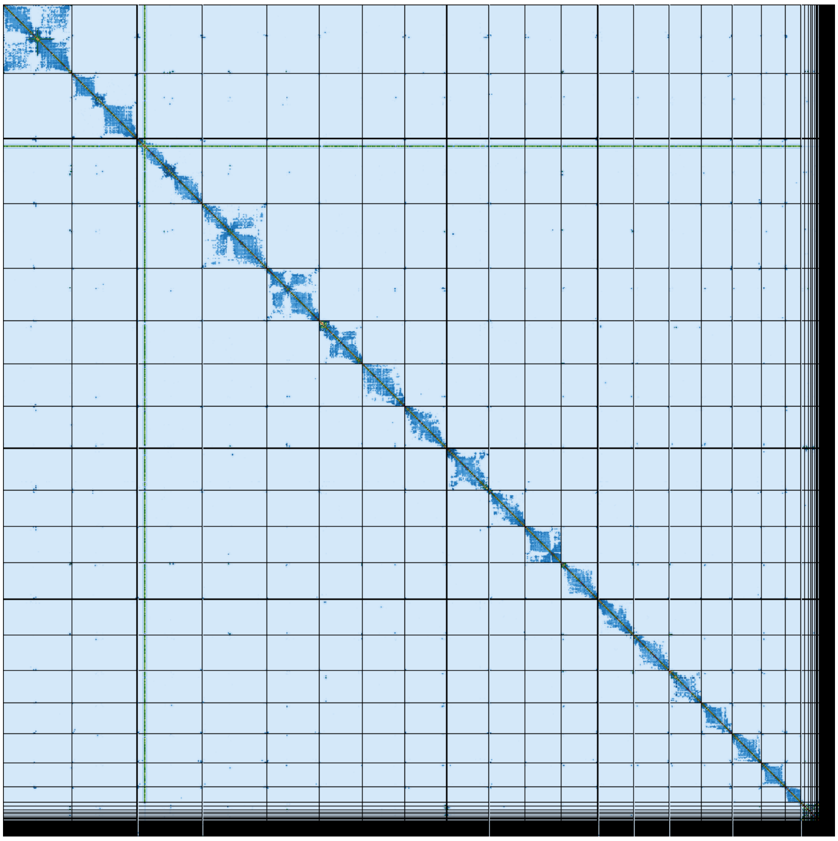
. Interventions/Gb: 39
. Contamination notes: ""
. Other observations: "The assembly of Capros aper (fCapApe1) is based on 38X PacBio data and 243X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 2 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.102 Mb (with the largest being 0.067 Mb). Additionally, 289 regions totaling 21.089 Mb (with the largest being 0.933 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 5 haplotypic regions and 38 contaminant sequences were removed, totaling 4.012 Mb and 1.298 Mb, respectively (with the largest being 2.855 Mb and 0.086 Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	654,971,367	649,617,890
GC %	43.26	43.31
Gaps/Gbp	511.47	532.62
Total gap bp	33,500	36,400
Scaffolds	227	147
Scaffold N50	32,022,220	32,570,917
Scaffold L50	8	8
Scaffold L90	18	17
Contigs	562	493
Contig N50	12,853,060	12,853,060
Contig L50	19	19
Contig L90	116	113
QV	42.7998	43.9731
Kmer compl.	85.6662	85.4999
BUSCO sing.	96.3%	96.6%
BUSCO dupl.	0.7%	0.4%
BUSCO frag.	0.7%	0.7%
BUSCO miss.	2.2%	2.2%

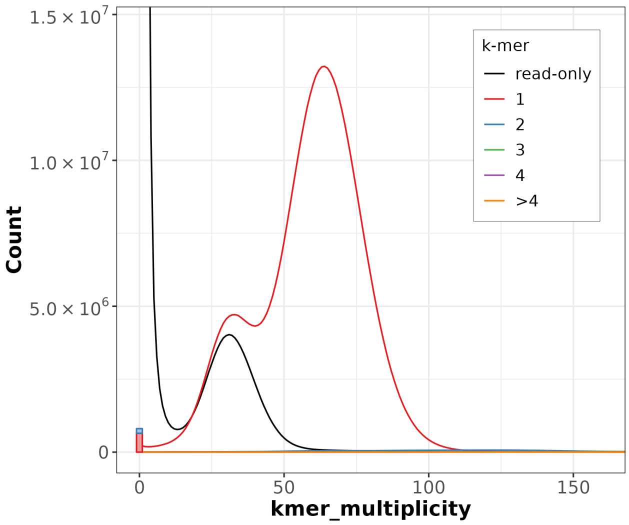
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

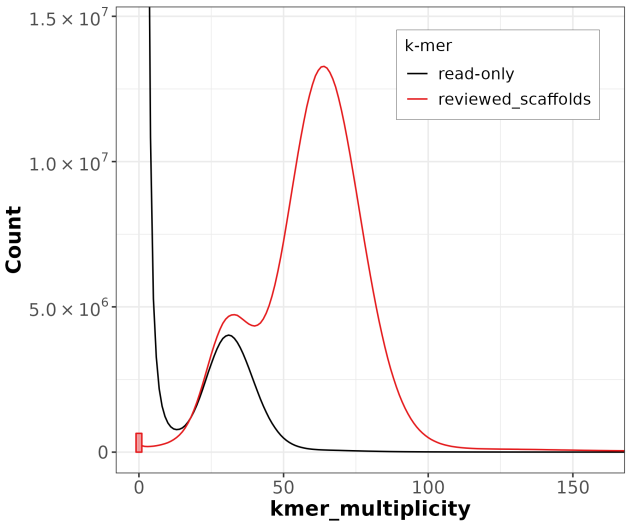


collapsed [LINK]

K-mer spectra of curated assembly

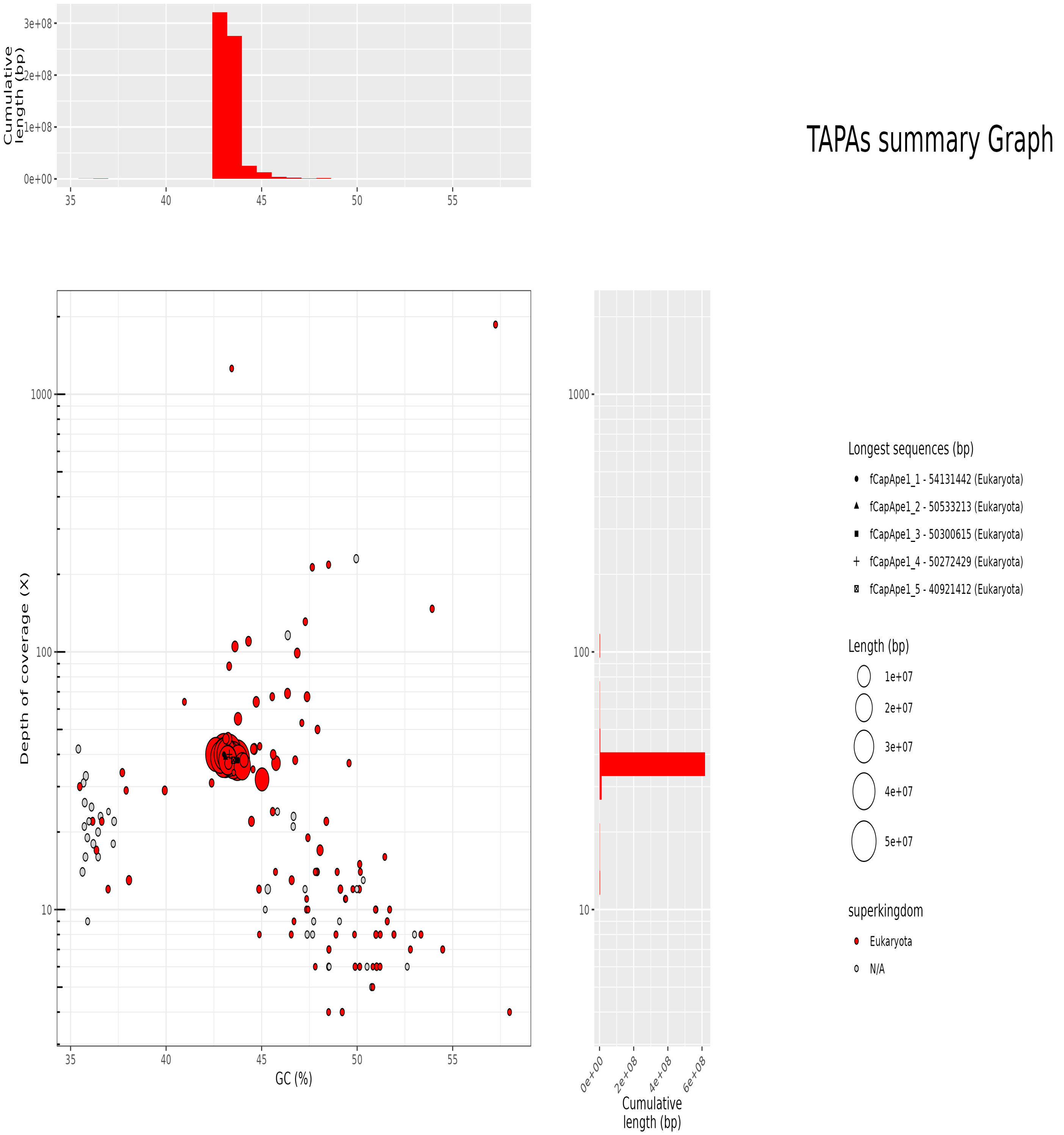


Distribution of k-mer counts per copy numbers found in asm



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	38	243

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: Simone Duprat
Affiliation: Genoscope

Date and time: 2025-06-07 16:37:31 CEST