ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	48191
ToLID	**fCheAur1**
Species	Chelon auratus
Class	Actinopteri
Order	Mugiliformes

Genome Traits	Expected	Observed
Haploid size (bp)	689,678,232	687,979,409
Haploid Number	24 (source: ancestor)	24
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q47

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

. Kmer completeness value is less than 90 for collapsed

Curator notes

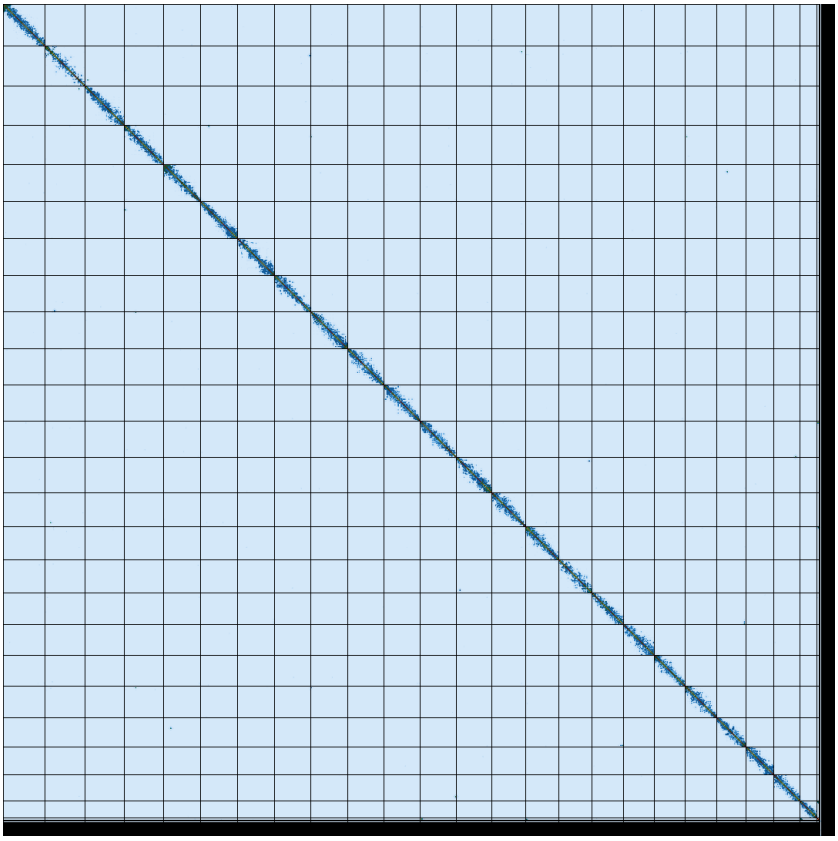
. Interventions/Gb: 6
. Contamination notes: ""
. Other observations: "The assembly of Chelon auratus (fCheAur1) is based on 48X PacBio data and 145X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 1 contigs was identified as contaminants (bacterial, archaeal, or viral), totaling 0.068 Mb (with the largest being 0.068 Mb). Additionally, 54 regions totaling 8.473 Mb (with the largest being 1.739 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. 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	687,962,005	687,979,409
GC %	42.31	42.31
Gaps/Gbp	31.98	34.88
Total gap bp	2,200	2,700
Scaffolds	96	95
Scaffold N50	29,772,500	29,772,500
Scaffold L50	11	11
Scaffold L90	22	22
Contigs	118	119
Contig N50	27,624,000	27,624,000
Contig L50	12	12
Contig L90	23	23
QV	47.8525	47.8526
Kmer compl.	86.5585	86.5585
BUSCO sing.	99.1%	99.1%
BUSCO dupl.	0.4%	0.4%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.5%	0.5%

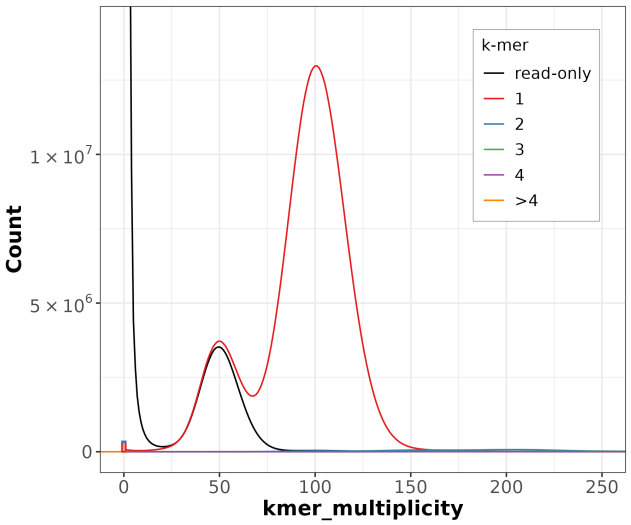
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

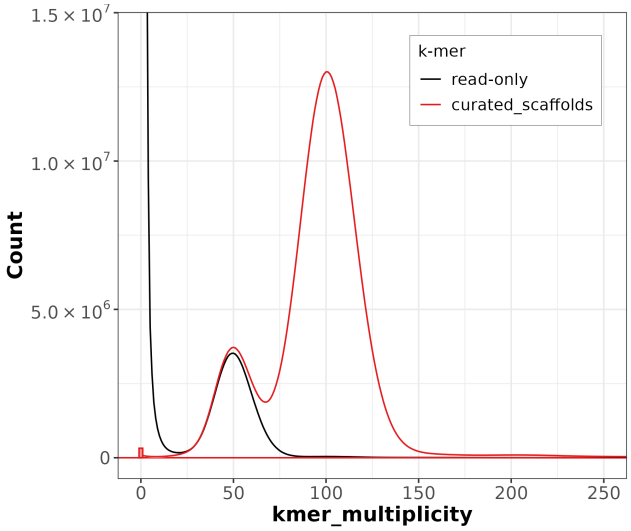


collapsed [LINK]

K-mer spectra of curated assembly

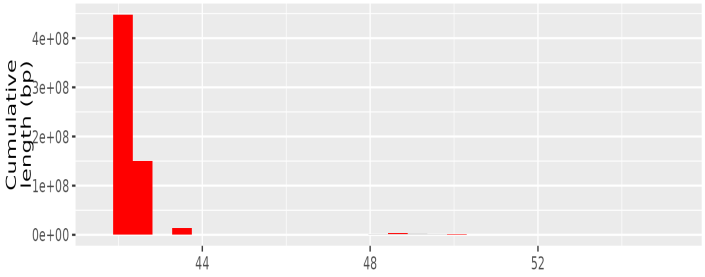


Distribution of k-mer counts per copy numbers found in asm

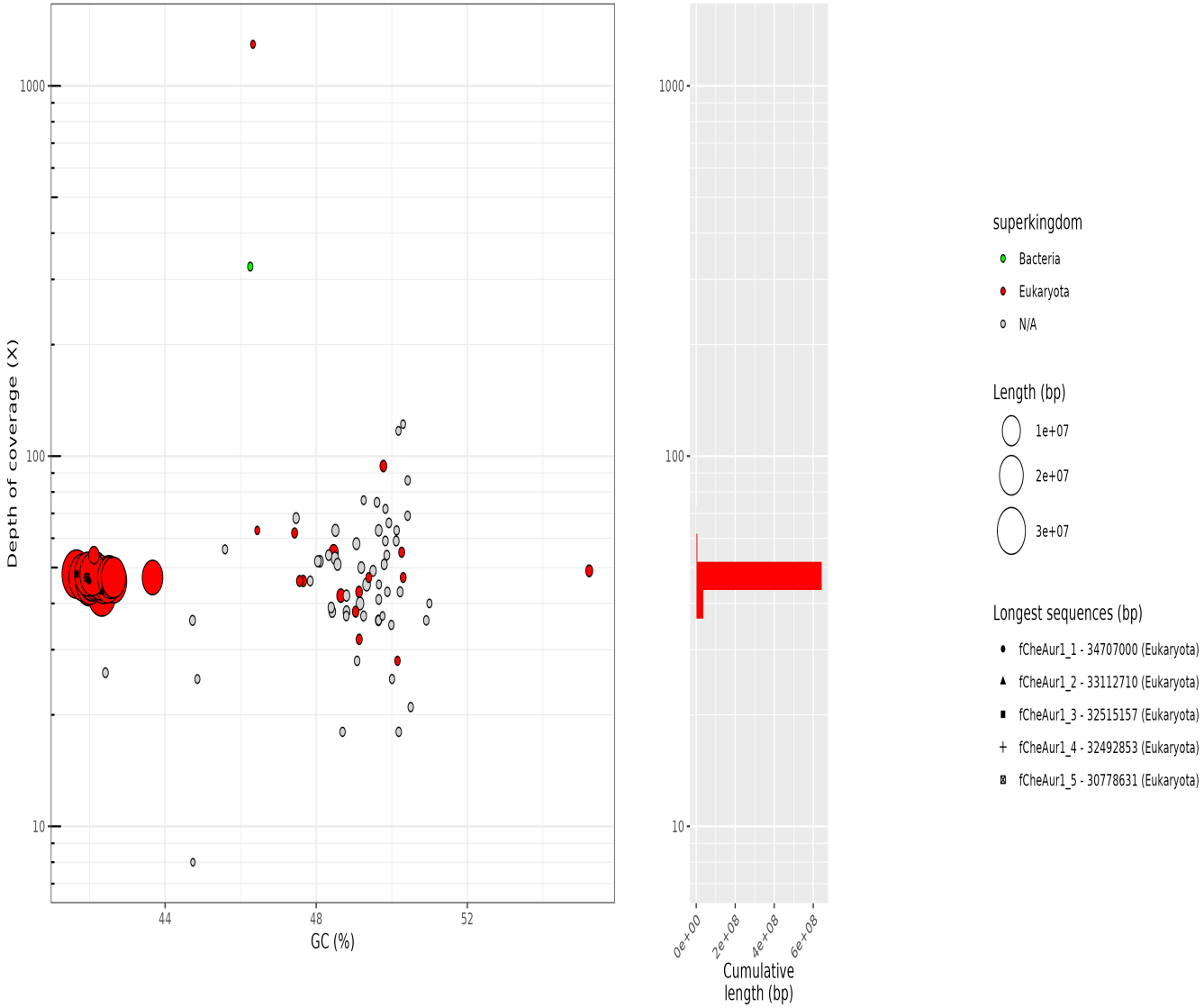


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	48	145

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: Emilie Teodori
Affiliation: Genoscope

Date and time: 2026-04-23 15:26:03 CEST