ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	160705
ToLID	**fCepFul1**
Species	Cephalopholis fulva
Class	Actinopteri
Order	Perciformes

Genome Traits	Expected	Observed
Haploid size (bp)	981,328,045	1,000,380,871
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.Q50

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

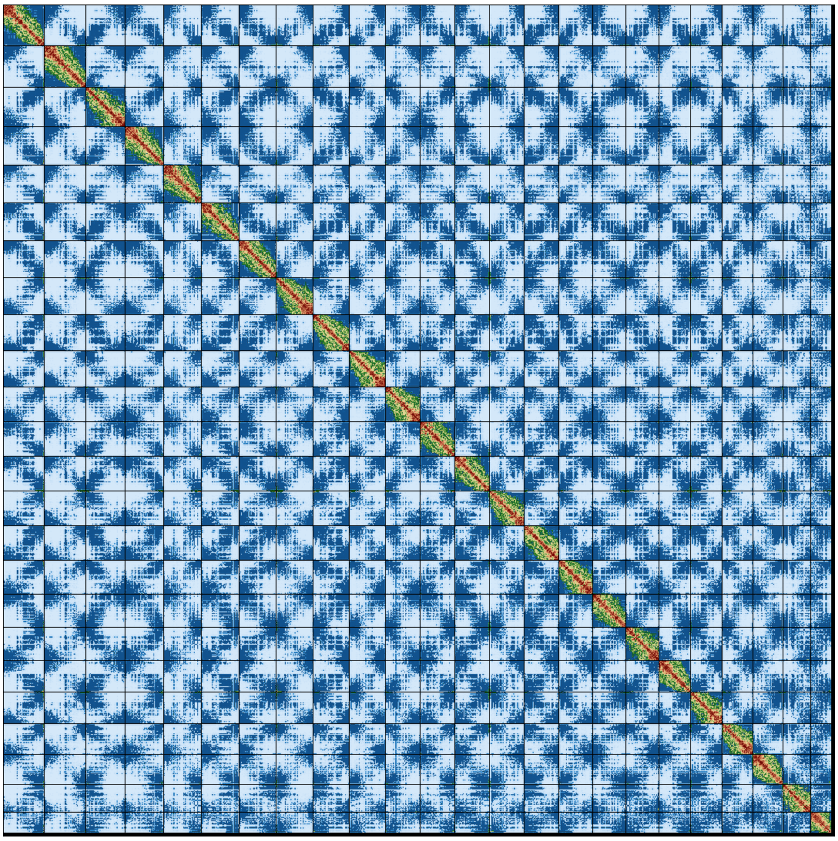
. Interventions/Gb: 6
. Contamination notes: ""
. Other observations: "The assembly of Cephalopholis fulva (fCepFul1) is based on 67X PacBio data and 214X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. No contigs were found to be contaminants (bacterial, archaeal, or viral). Additionally, 43 regions totaling 2.866 Mb (with the largest being 0.418 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. During manual curation, 1 haplotypic regions were removed, totaling 6.2Mb (with the largest being 6.2Mb). 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	1,006,663,017	1,000,380,871
GC %	41.01	41.02
Gaps/Gbp	16.89	13.99
Total gap bp	1,700	1,700
Scaffolds	46	44
Scaffold N50	42,142,236	42,142,236
Scaffold L50	11	11
Scaffold L90	22	21
Contigs	57	58
Contig N50	41,490,643	41,490,643
Contig L50	12	12
Contig L90	22	23
QV	50.7	50.6993
Kmer compl.	86.8488	86.7845
BUSCO sing.	96.5%	99.4%
BUSCO dupl.	1.2%	0.3%
BUSCO frag.	0.6%	0.0%
BUSCO miss.	1.7%	0.3%

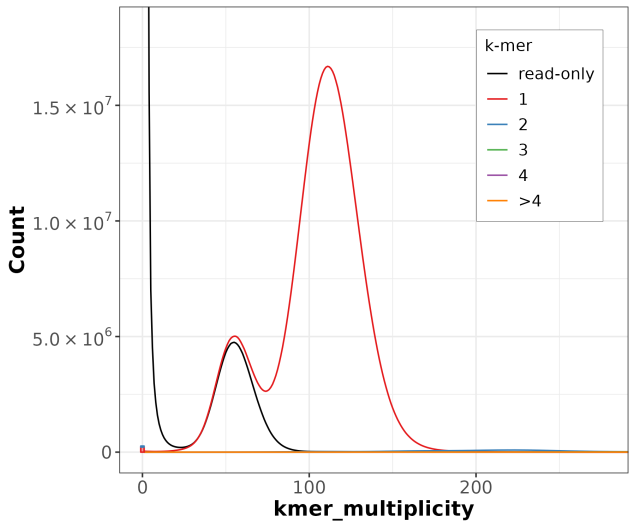
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

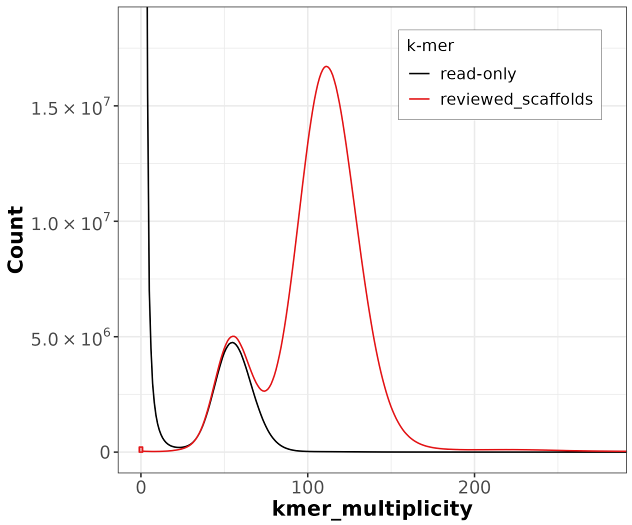


collapsed [LINK]

K-mer spectra of curated assembly

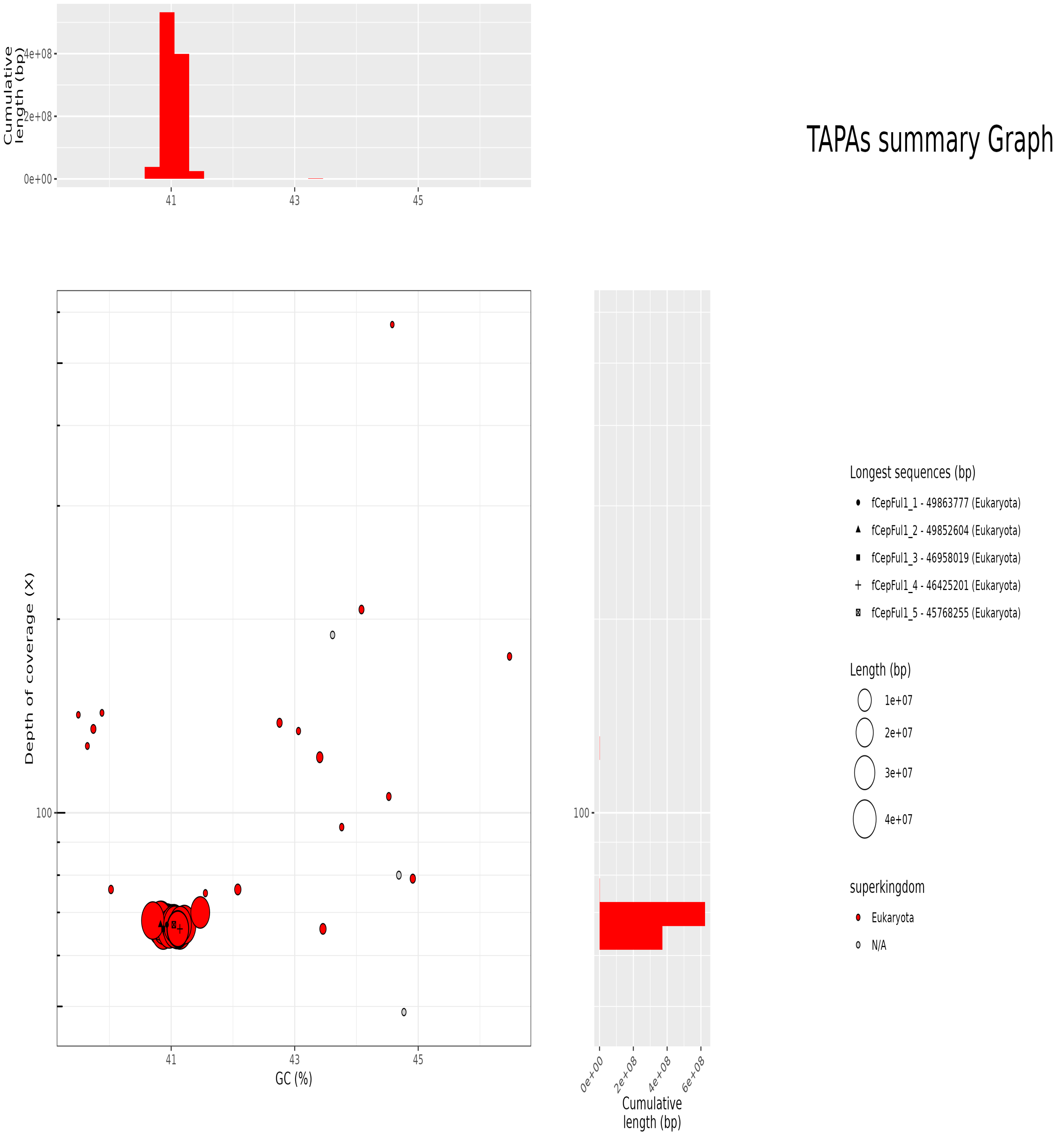


Distribution of k-mer counts per copy numbers found in asm



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	67	214

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: 2025-10-25 14:29:26 CEST