ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	179535
ToLID	**fEpiMar1**
Species	Epinephelus marginatus
Class	Actinopteri
Order	Perciformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,094,448,048	1,092,847,383
Haploid Number	24 (source: direct)	24
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q49

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

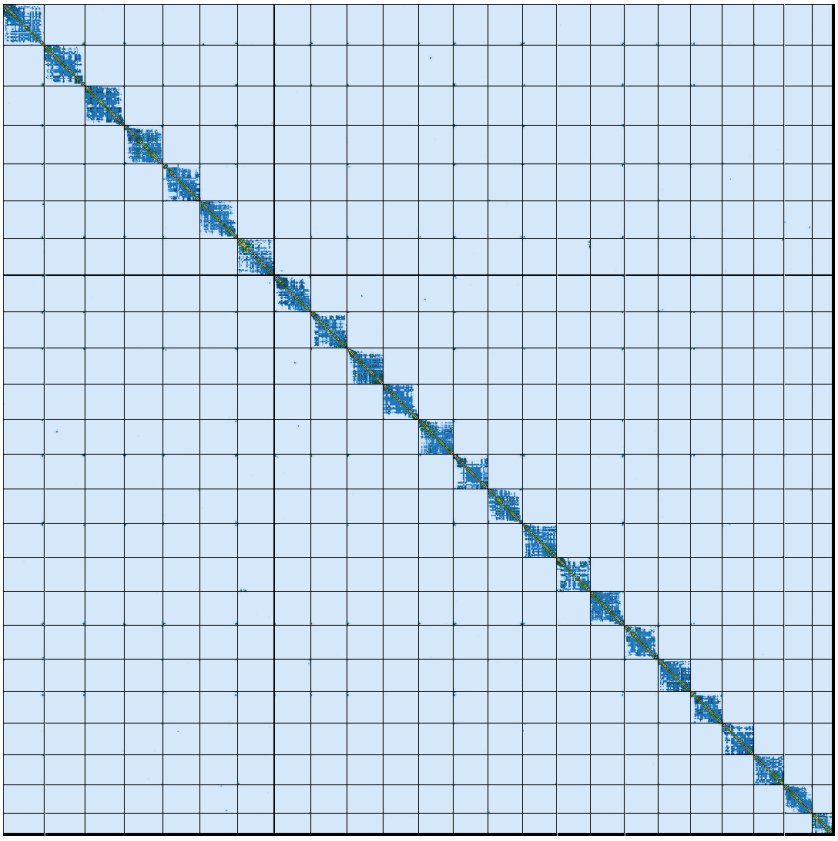
. Interventions/Gb: 71
. Contamination notes: ""
. Other observations: "The assembly of Epinephelus marginatus (fEpiMar1) is based on 53X PacBio data and 130X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included the following steps: initial 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 identified as contaminants (bacterial, archaeal, or viral). Additionally, 133 regions totaling 27.2 Mb (with the largest being 4 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, 4 haplotypic regions and 1 contaminant sequence were removed, totaling 6.5Mb and 0.058Mb, respectively (with the largest being 5.7Mb and 0.058Mb). 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	1,099,454,344	1,092,847,383
GC %	41.44	41.45
Gaps/Gbp	57.3	85.1
Total gap bp	6,300	14,000
Scaffolds	91	64
Scaffold N50	45,560,722	45,815,100
Scaffold L50	12	12
Scaffold L90	22	21
Contigs	154	157
Contig N50	29,909,000	29,601,000
Contig L50	14	14
Contig L90	43	44
QV	48.8144	49.4095
Kmer compl.	89.5503	89.4743
BUSCO sing.	98.2%	99.2%
BUSCO dupl.	1.4%	0.4%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.4%	0.4%

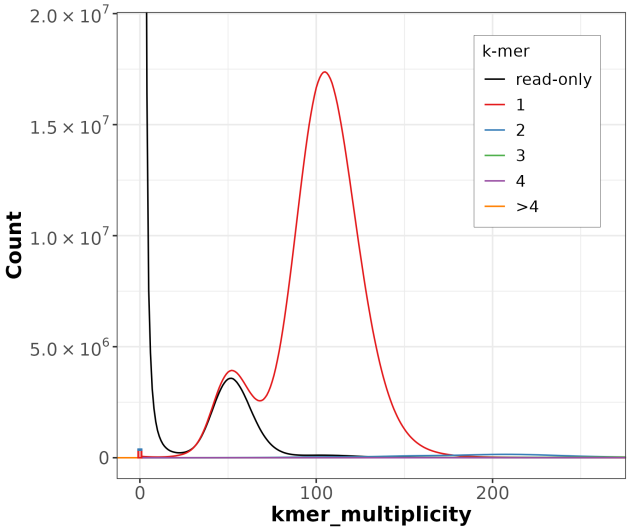
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

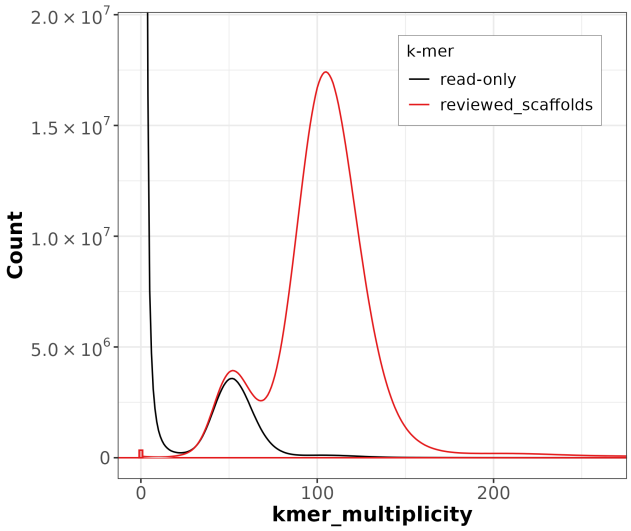


collapsed [LINK]

K-mer spectra of curated assembly

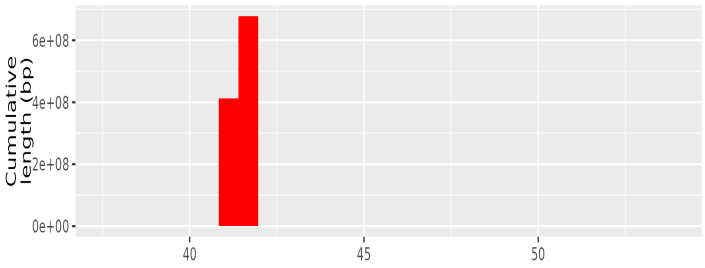


Distribution of k-mer counts per copy numbers found in asm

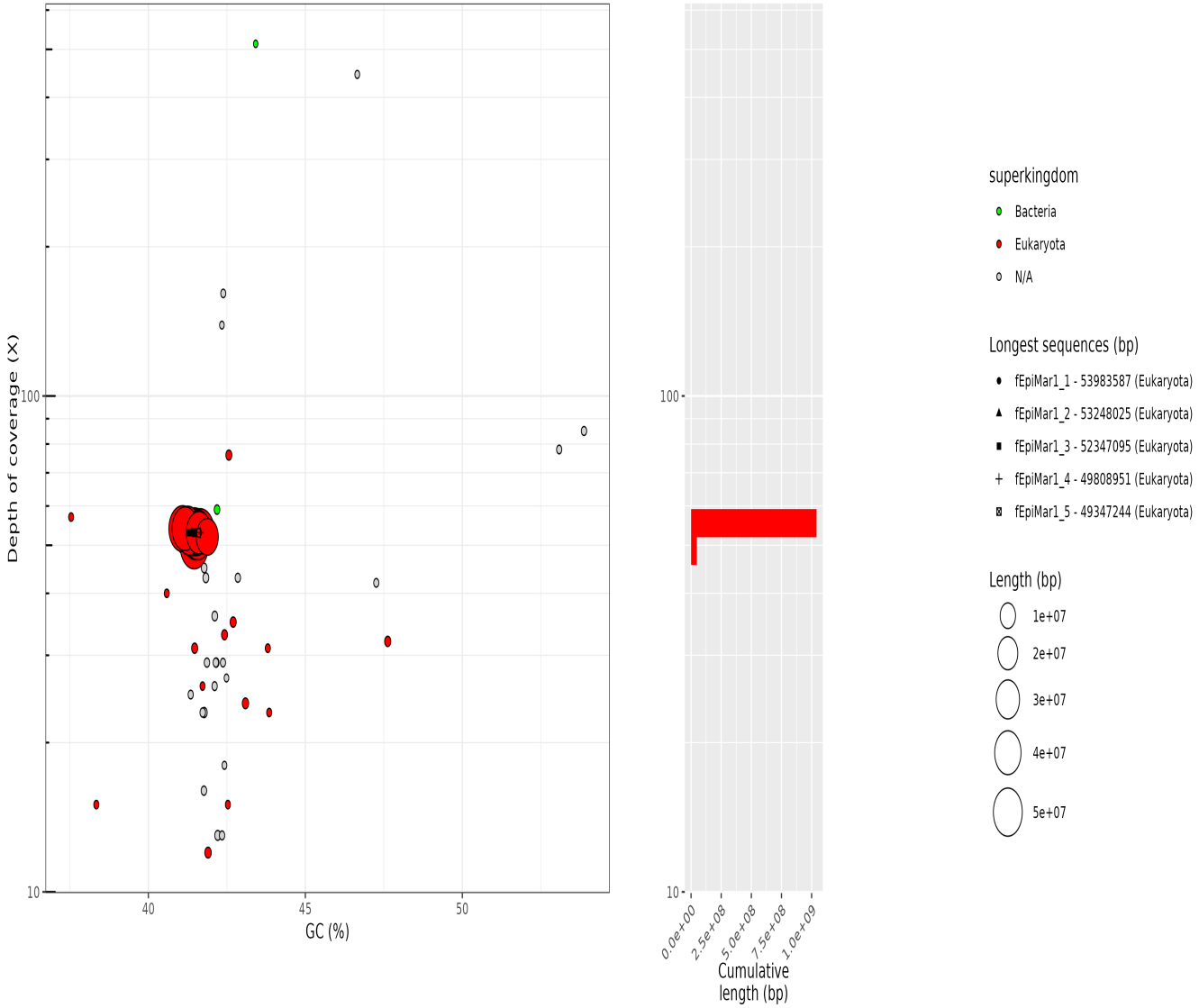


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	53	131

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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2026-05-04 05:36:37 CEST