ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	203291
ToLID	**fElaEvl1**
Species	Elacatinus evelynae
Class	Actinopteri
Order	Gobiiformes

Genome Traits	Expected	Observed
Haploid size (bp)	823,185,473	953,763,457
Haploid Number	21 (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.Q45

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
. Not 90% of assembly in chromosomes for collapsed

Curator notes

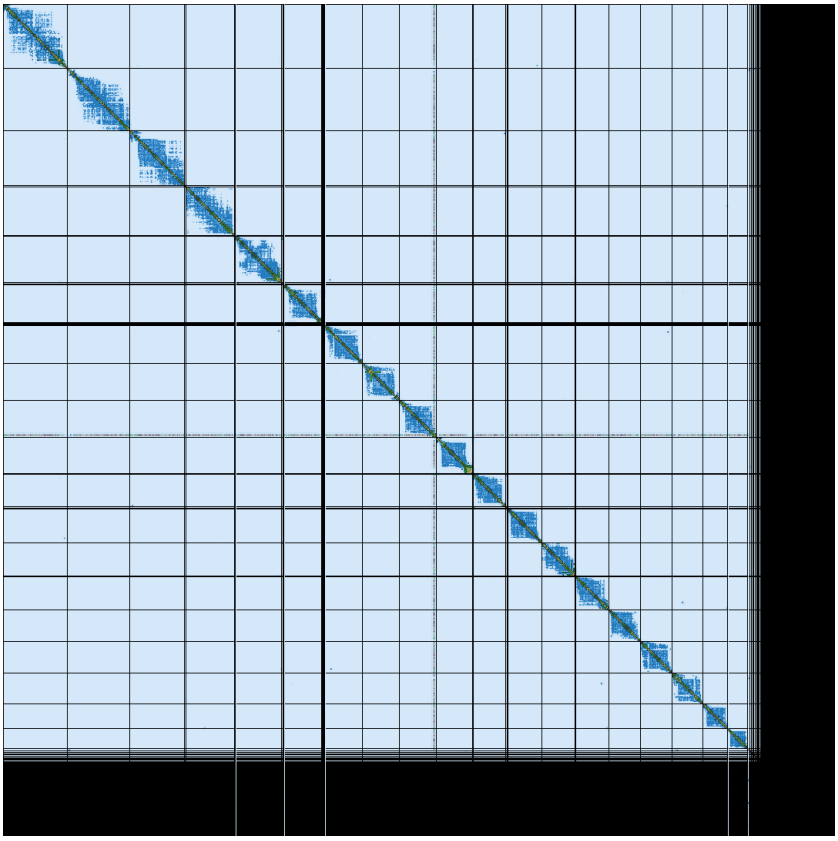
. Interventions/Gb: 32
. Contamination notes: ""
. Other observations: "The assembly of Elacatinus evelynae (fElaEvl1) is based on 33X PacBio data and 107X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 639 regions totaling 55.59 Mb (with the largest being 0.333 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. During manual curation, 13 haplotypic regions were removed, totaling 10Mb (with the largest being 3.17Mb). 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	966,535,168	953,763,457
GC %	44.26	44.25
Gaps/Gbp	518.35	501.17
Total gap bp	56,400	59,000
Scaffolds	655	494
Scaffold N50	37,225,694	42,235,630
Scaffold L50	10	9
Scaffold L90	61	31
Contigs	1,103	972
Contig N50	11,708,270	11,923,904
Contig L50	21	20
Contig L90	267	255
QV	45.4757	45.5061
Kmer compl.	77.2488	76.9671
BUSCO sing.	95.7%	96.6%
BUSCO dupl.	1.8%	0.9%
BUSCO frag.	0.5%	0.5%
BUSCO miss.	2.0%	2.0%

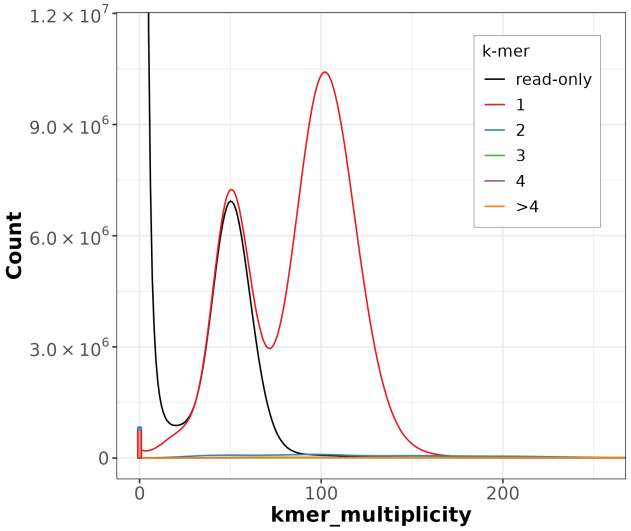
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

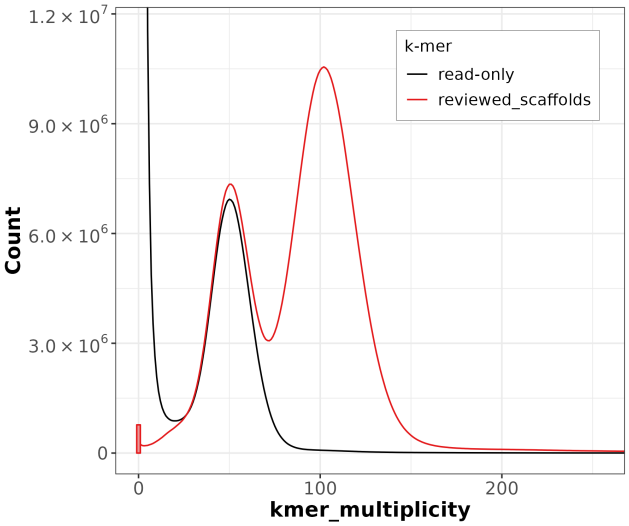


collapsed [LINK]

K-mer spectra of curated assembly

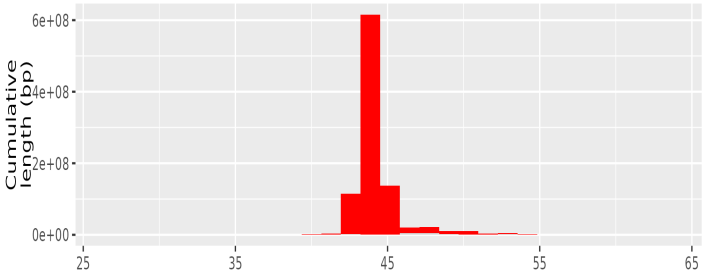


Distribution of k-mer counts per copy numbers found in asm

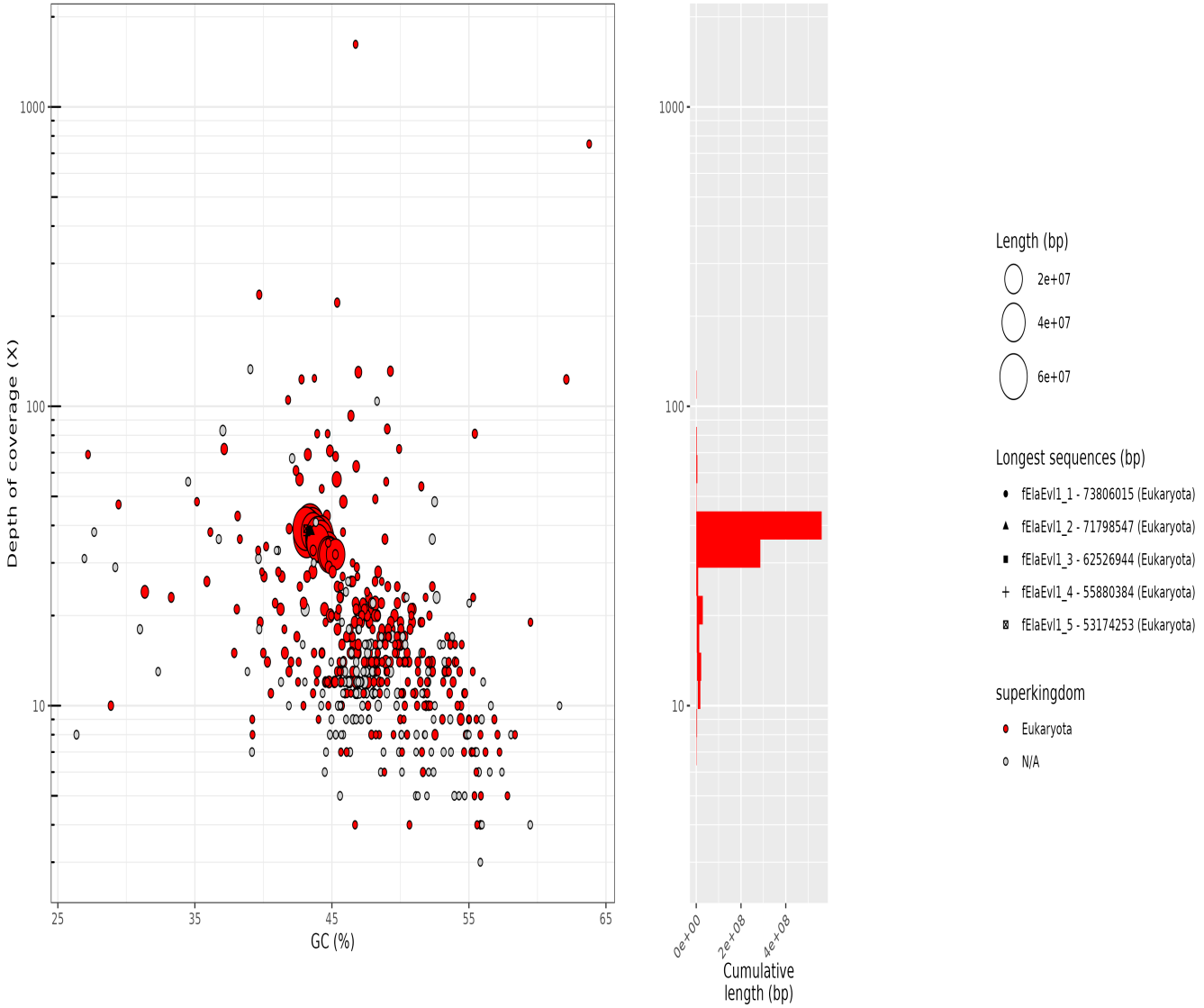


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	38	125

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: 2025-12-13 10:42:15 CET