

ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	203290
ToLID	fElaChal
Species	Elacatinus chancei
Class	Actinopteri
Order	Gobiiformes

Genome Traits	Expected	Observed
Haploid size (bp)	820,476,022	944,324,776
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.Q47

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

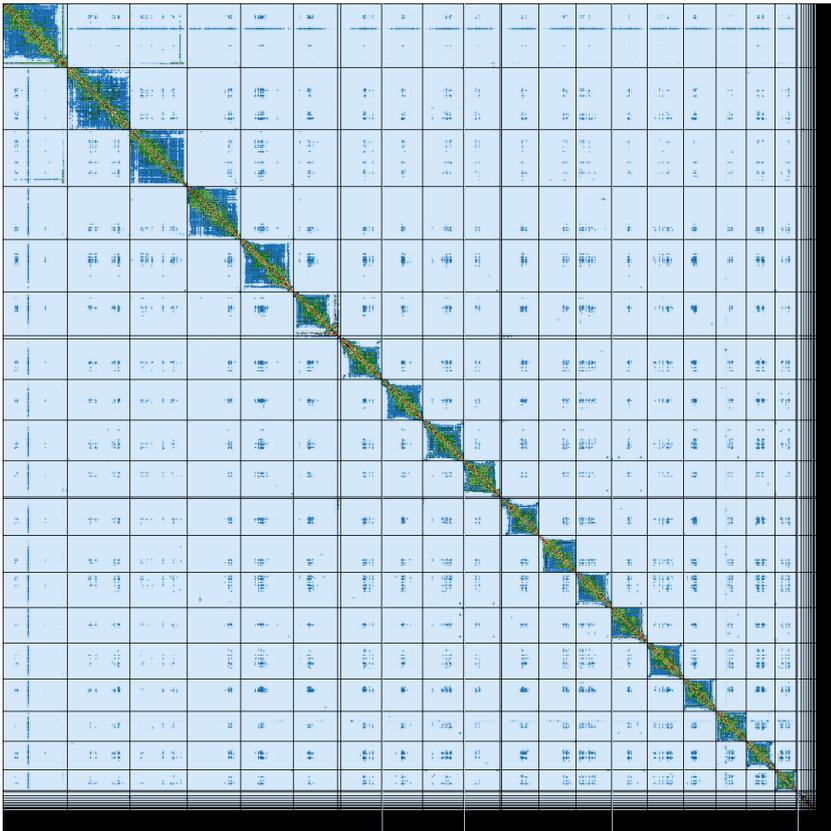
. Interventions/Gb: 62
. Contamination notes: ""
. Other observations: "The assembly of *Elacatinus chancei* (fElaChal) is based on 52X PACBIO data and 136X Arima Hi-C data generated as part of the ATLASea programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial 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), totaling 0.047 Mb (with the largest being 0.030 Mb). Additionally, 164 regions totaling 30.325 Mb (with the largest being 1.406 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 were removed, totaling 2.919 Mb (with the largest being 1.132 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	947,234,681	944,324,776
GC %	43.42	43.42
Gaps/Gbp	186.86	223.44
Total gap bp	17,700	25,600
Scaffolds	188	144
Scaffold N50	40,446,751	45,871,026
Scaffold L50	9	9
Scaffold L90	22	18
Contigs	365	355
Contig N50	16,946,356	16,946,356
Contig L50	15	15
Contig L90	104	104
QV	47.1954	47.2064
Kmer compl.	84.4727	84.4348
BUSCO sing.	96.7%	97.1%
BUSCO dupl.	0.9%	0.5%
BUSCO frag.	0.6%	0.6%
BUSCO miss.	1.8%	1.8%

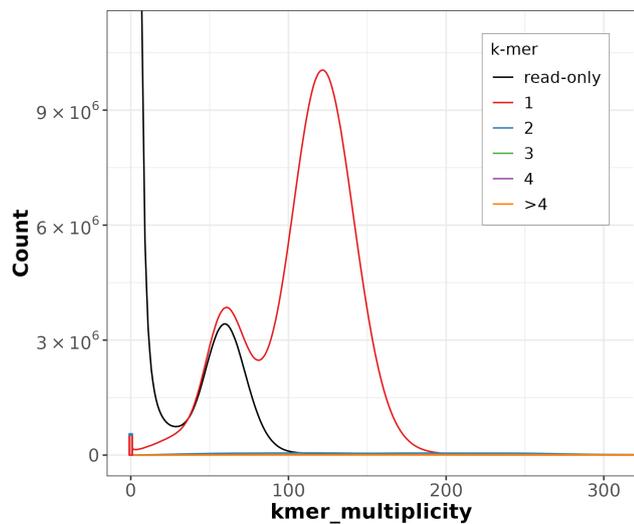
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

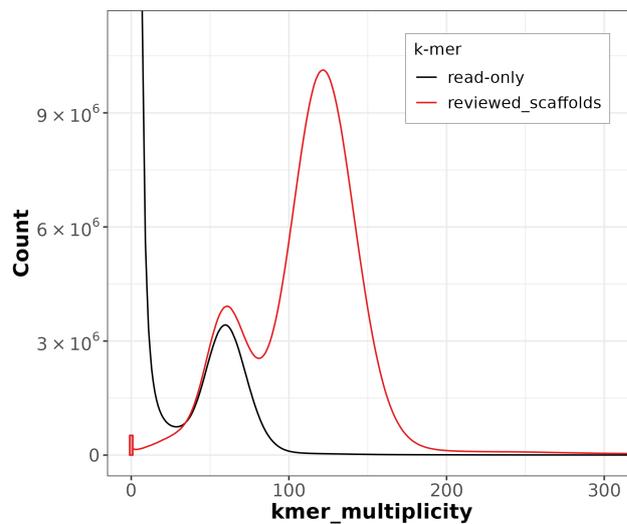


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

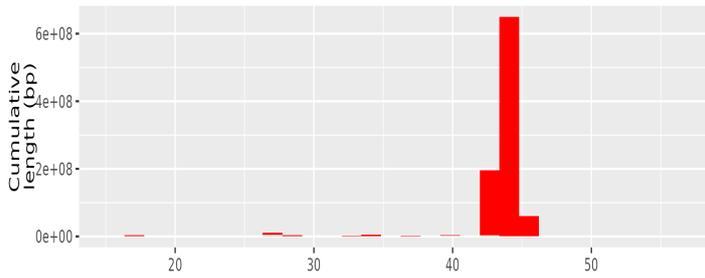


Distribution of k-mer counts per copy numbers found in asm

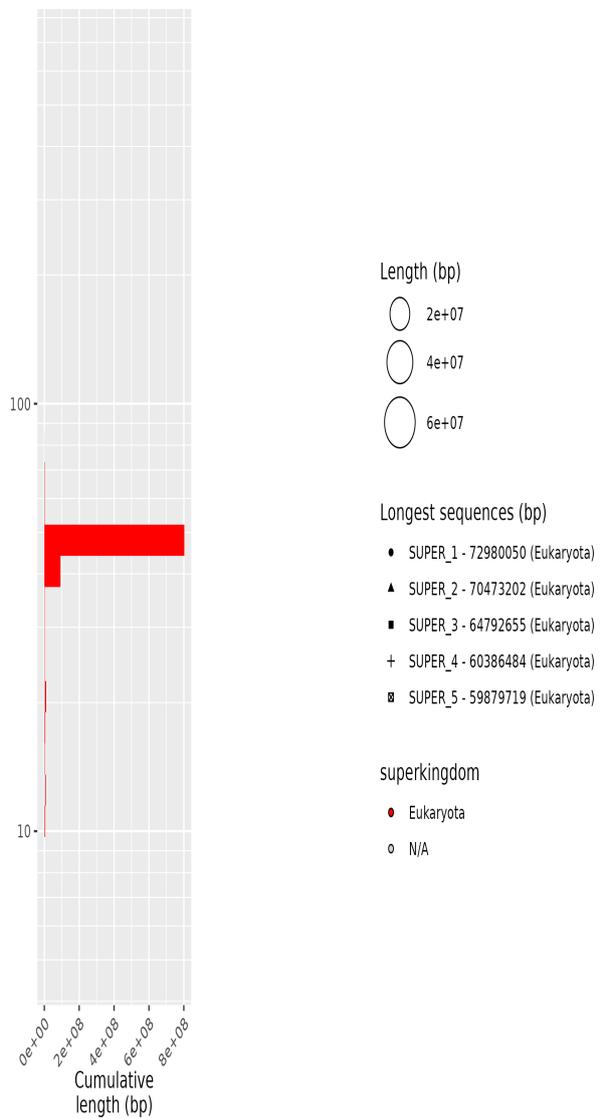
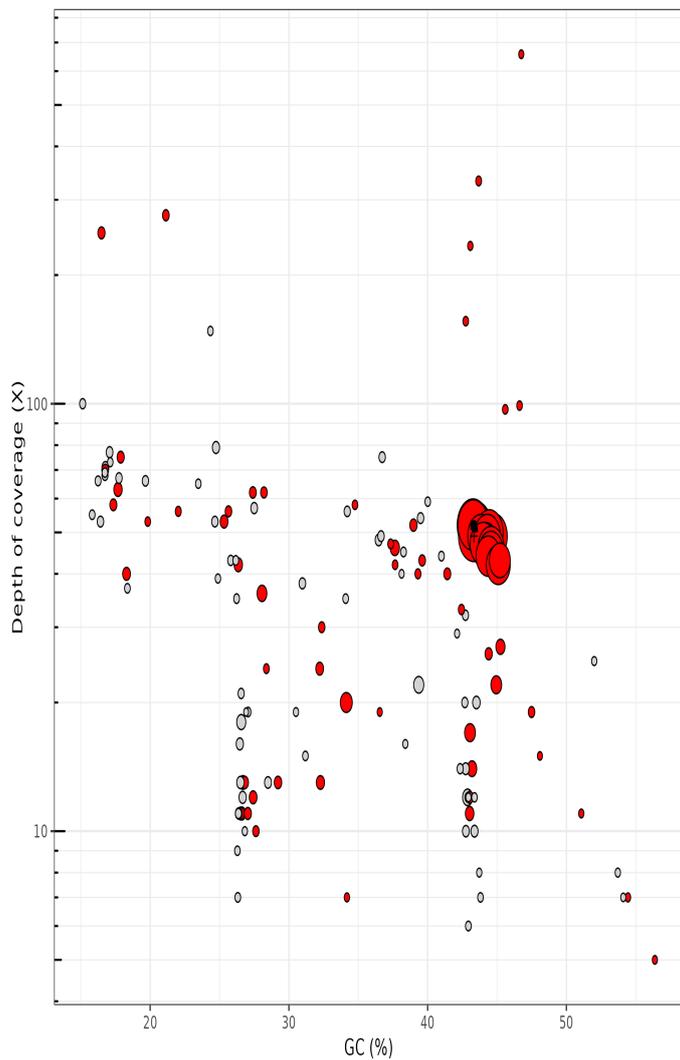


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	52	136

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: Sophie Layac

Affiliation: Genoscope

Date and time: 2026-01-21 06:05:40 CET