

ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	157584
ToLID	fAcaTral
Species	Acanthurus tractus
Class	Actinopteri
Order	Acanthuriformes

Genome Traits	Expected	Observed
Haploid size (bp)	645,626,606	650,177,444
Haploid Number	24 (source: ancestor)	16
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:

- . Observed Haploid Number is different from Expected
- . Kmer completeness value is less than 90 for collapsed

Curator notes

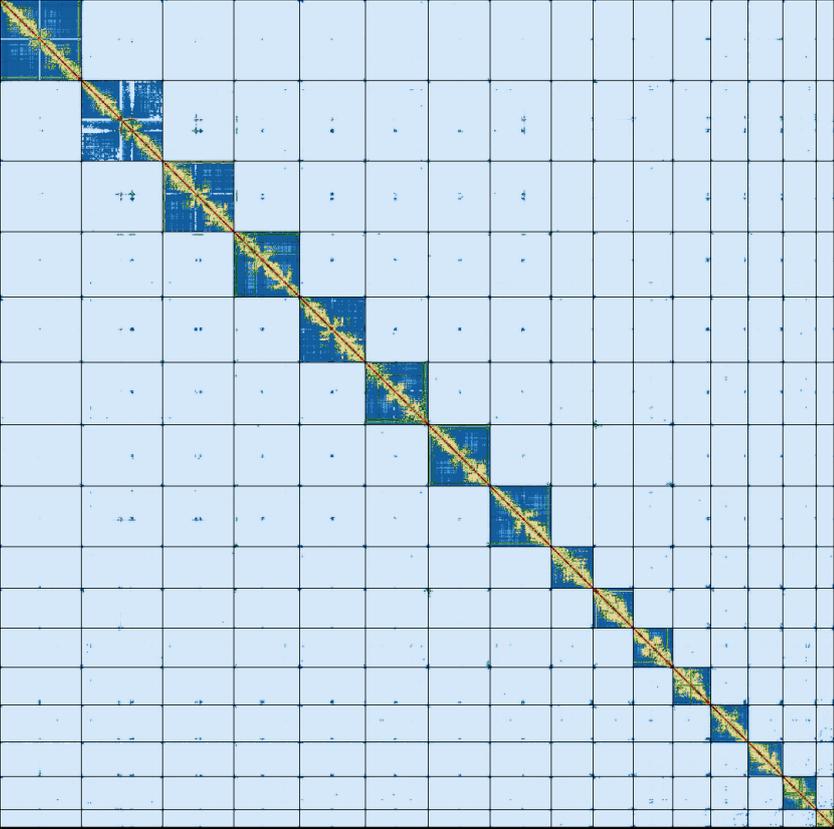
- . Interventions/Gb: 4
- . Contamination notes: ""
- . Other observations: "The assembly of Acanthurus tractus (fAcaTral) is based on 52X PacBio data and 194X Arima Hi-C data generated as part of the ATLASea programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial 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, 35 regions totaling 1.801 Mb (with the largest being 0.223 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. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. We note a difference in the number of chromosomes for this species (16 chromosomes) compared to the number of chromosomes in the species Acanthurus coeruleus (fAcaCoel), which has 24. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	650,177,044	650,177,444
GC %	40.57	40.57
Gaps/Gbp	32.3	35.37
Total gap bp	2,100	2,500
Scaffolds	29	28
Scaffold N50	47,825,794	48,961,085
Scaffold L50	7	6
Scaffold L90	15	14
Contigs	50	51
Contig N50	31,041,855	31,041,855
Contig L50	8	8
Contig L90	18	18
QV	49.4999	49.4999
Kmer compl.	84.4647	84.4647
BUSCO sing.	99.4%	99.4%
BUSCO dupl.	0.3%	0.3%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.3%	0.3%

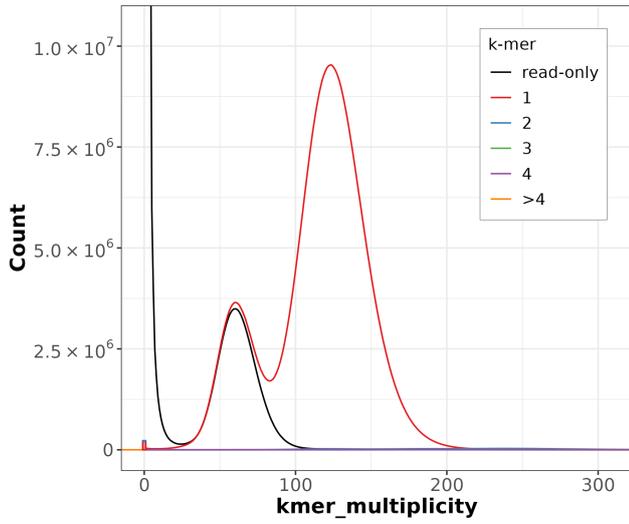
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

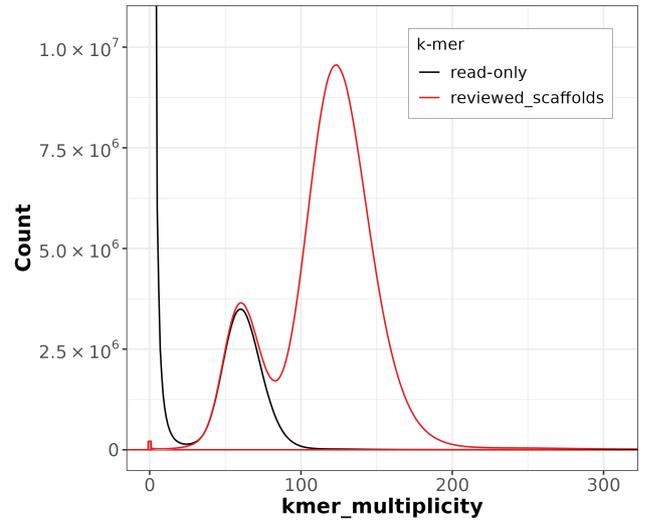


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

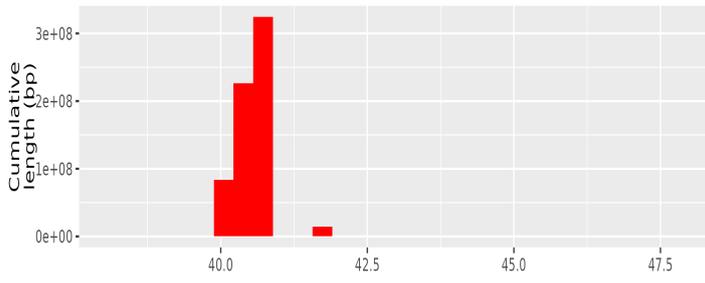


Distribution of k-mer counts per copy numbers found in asm

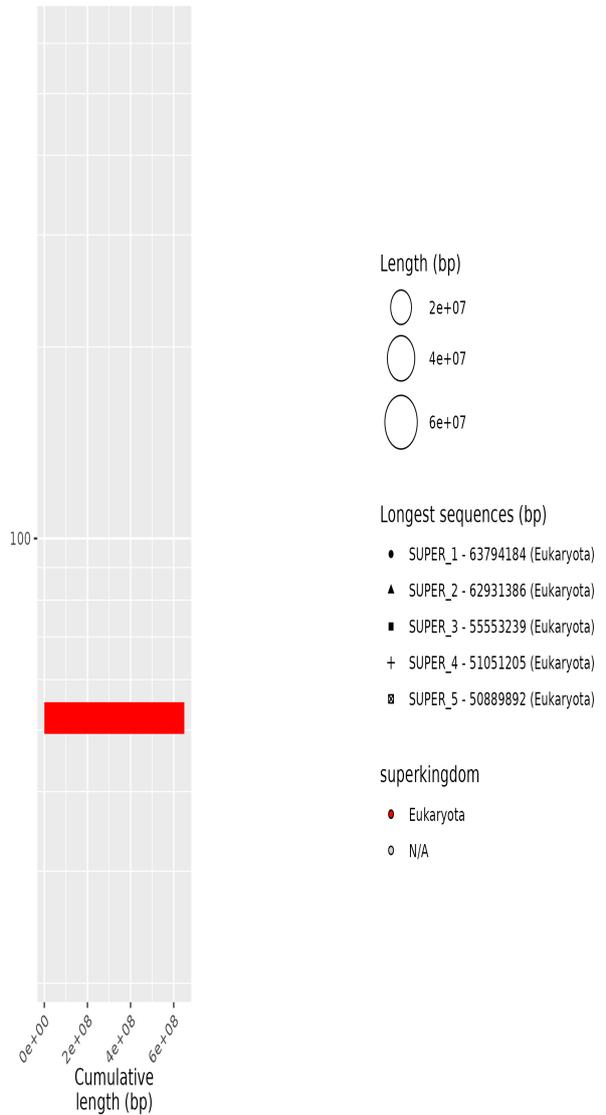
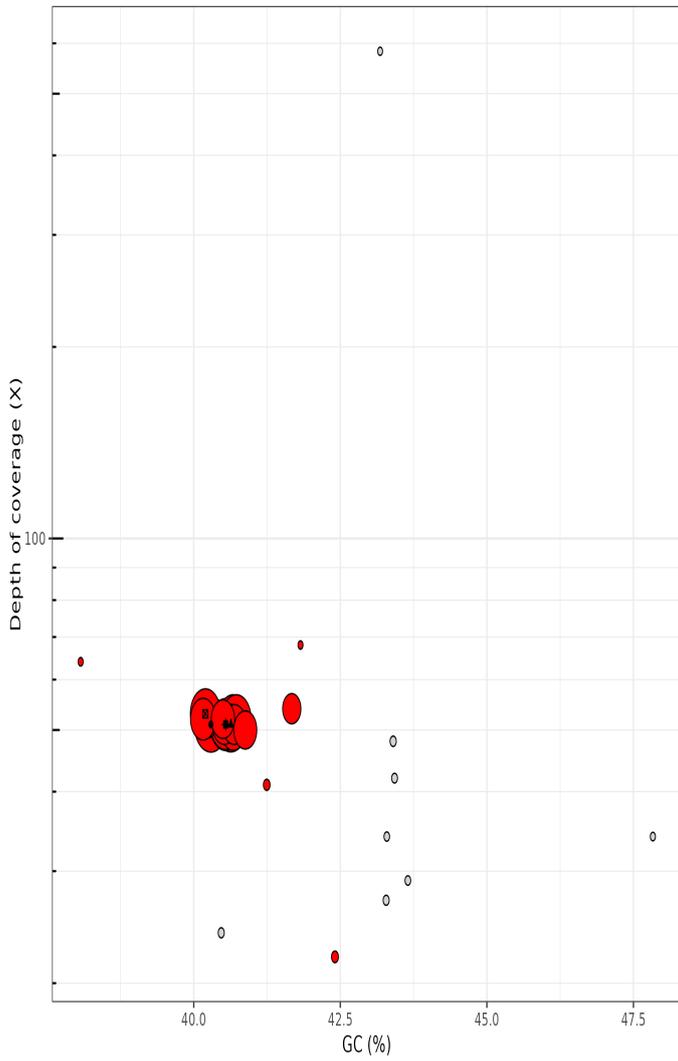


Distribution of k-mer counts coloured by their presence in reads/assemblies

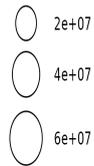
Post-curation contamination screening



TAPAs summary Graph



Length (bp)



Longest sequences (bp)

- SUPER_1 - 63794184 (Eukaryota)
- ▲ SUPER_2 - 62931386 (Eukaryota)
- SUPER_3 - 55553239 (Eukaryota)
- + SUPER_4 - 51051205 (Eukaryota)
- ▣ SUPER_5 - 50889892 (Eukaryota)

superkingdom

- Eukaryota
- N/A

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	194

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: 2026-02-06 17:12:20 CET