

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	157585
ToLID	fAcaCoe1
Species	Acanthurus coeruleus
Class	Actinopteri
Order	Acanthuriformes

Genome Traits	Expected	Observed
Haploid size (bp)	656,716,702	660,715,304
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.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

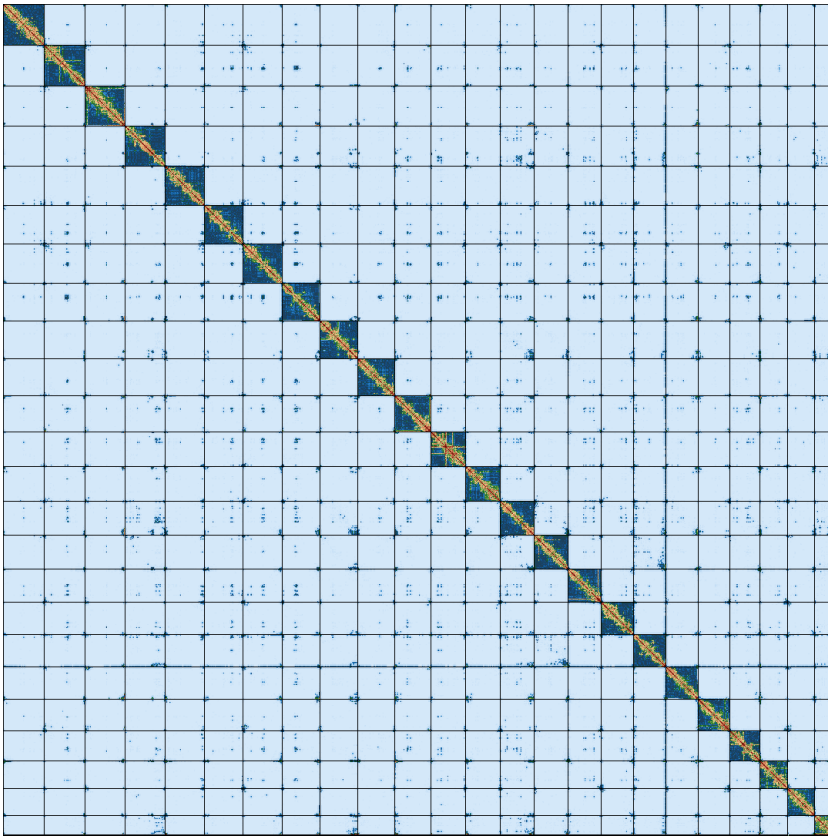
- . Interventions/Gb: 8
- . Contamination notes: ""
- . Other observations: "The assembly of Acanthurus coeruleus (fAcaCoe1.1) is based on 104X PACBIO data and 190X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: 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, 4 contigs were identified as contaminants (bacterial), totaling 0.116 Mb (with the largest being 0.048 Mb). Additionally, 63 regions totaling 4.682 Mb (with the largest being 0.532 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. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	660,714,604	660,715,304
GC %	40.04	40.04
Gaps/Gbp	4.54	9.08
Total gap bp	300	1,000
Scaffolds	40	37
Scaffold N50	28,675,853	28,675,853
Scaffold L50	11	11
Scaffold L90	21	21
Contigs	43	43
Contig N50	28,675,853	28,675,853
Contig L50	11	11
Contig L90	22	22
QV	49.6066	49.6066
Kmer compl.	88.1984	88.1984
BUSCO sing.	99.4%	99.4%
BUSCO dupl.	0.3%	0.3%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.3%	0.3%

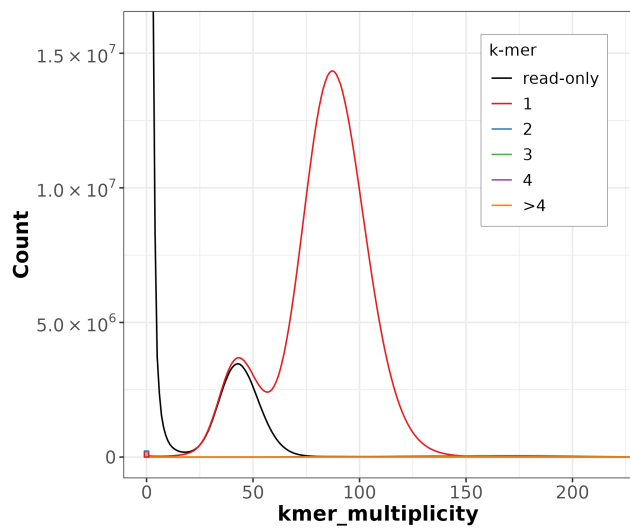
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

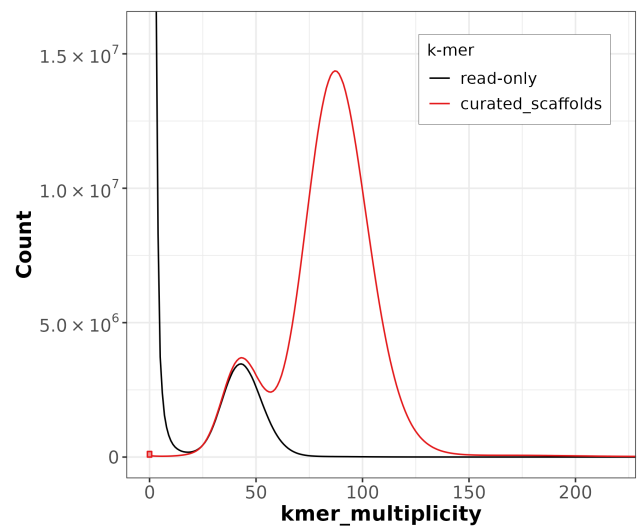


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

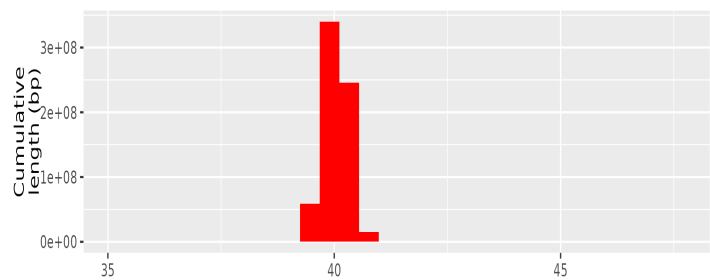


Distribution of k-mer counts per copy numbers found in asm

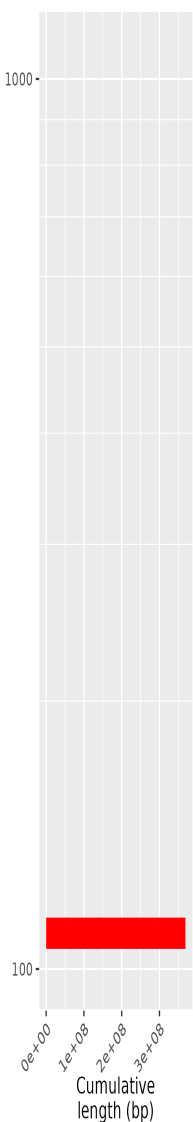
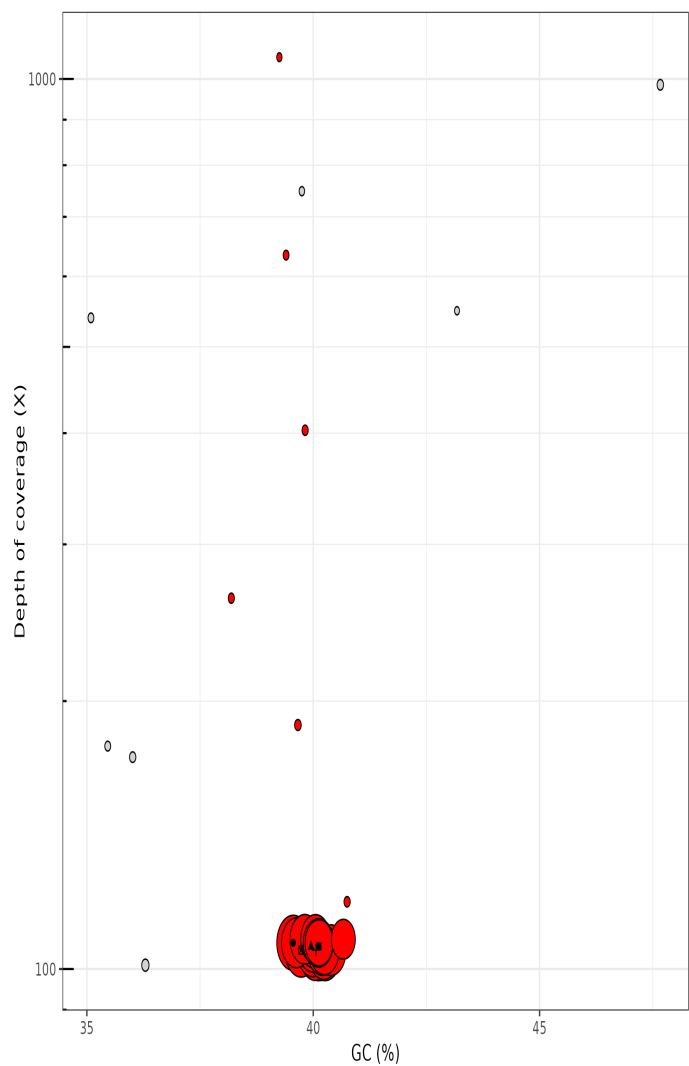


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



Length (bp)

1e+07

2e+07

3e+07

Longest sequences (bp)

• SUPER_1 - 32976313 (Eukaryota)

▲ SUPER_2 - 32323337 (Eukaryota)

■ SUPER_3 - 31758736 (Eukaryota)

+ SUPER_4 - 31495237 (Eukaryota)

▣ SUPER_5 - 31192075 (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	104	190

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: Adama Ndar

Affiliation: Genoscope

Date and time: 2025-11-12 10:40:54 CET