ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	303740
ToLID	**fCanros1**
Species	Canthigaster rostrata
Class	Actinopteri
Order	Tetraodontiformes

Genome Traits	Expected	Observed
Haploid size (bp)	342,100,910	361,488,908
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: 6.7.Q43

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

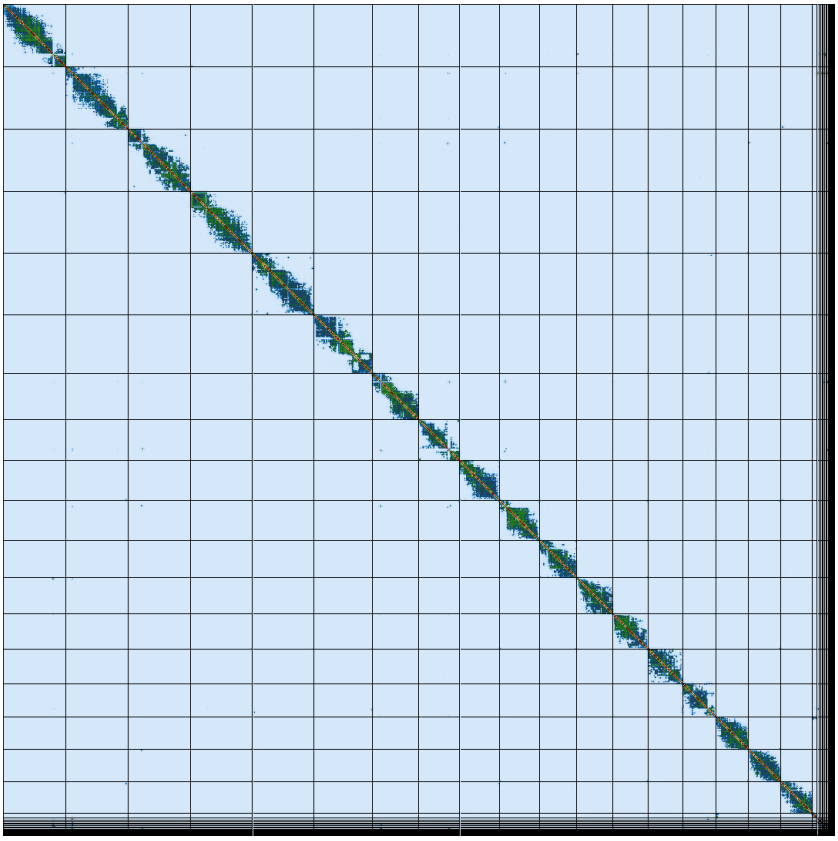
. Interventions/Gb: 89
. Contamination notes: ""
. Other observations: "The assembly of Canthigaster rostrata (fCanRos1) is based on 33X PacBio data and 194X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 113 regions totaling 6.029 Mb (with the largest being 0.415 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	361,595,787	361,488,908
GC %	46.9	46.9
Gaps/Gbp	724.57	832.67
Total gap bp	27,200	39,400
Scaffolds	120	60
Scaffold N50	17,482,471	17,593,143
Scaffold L50	8	8
Scaffold L90	17	17
Contigs	376	361
Contig N50	5,095,325	5,095,325
Contig L50	18	18
Contig L90	93	94
QV	43.4684	43.4738
Kmer compl.	83.3494	83.3396
BUSCO sing.	93.6%	93.6%
BUSCO dupl.	0.5%	0.5%
BUSCO frag.	1.5%	1.5%
BUSCO miss.	4.4%	4.4%

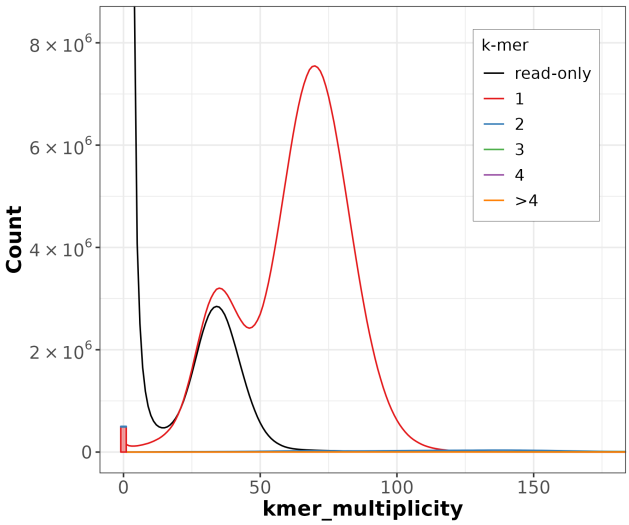
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

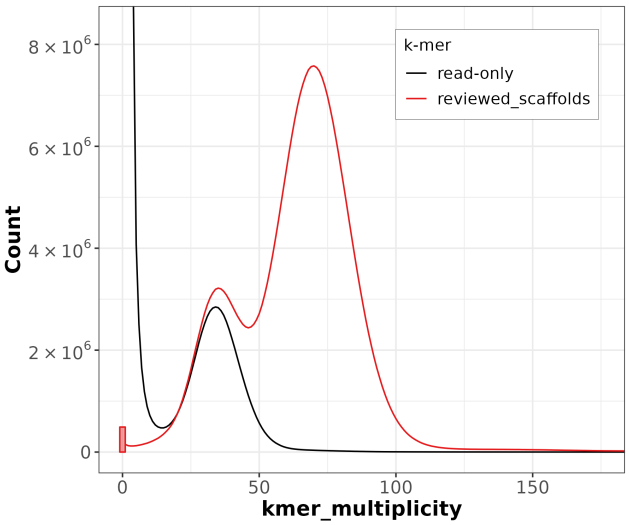


collapsed [LINK]

K-mer spectra of curated assembly

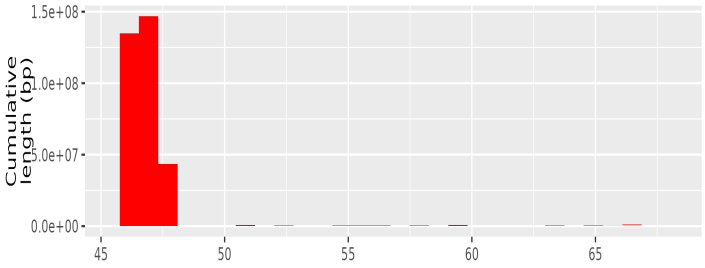


Distribution of k-mer counts per copy numbers found in asm

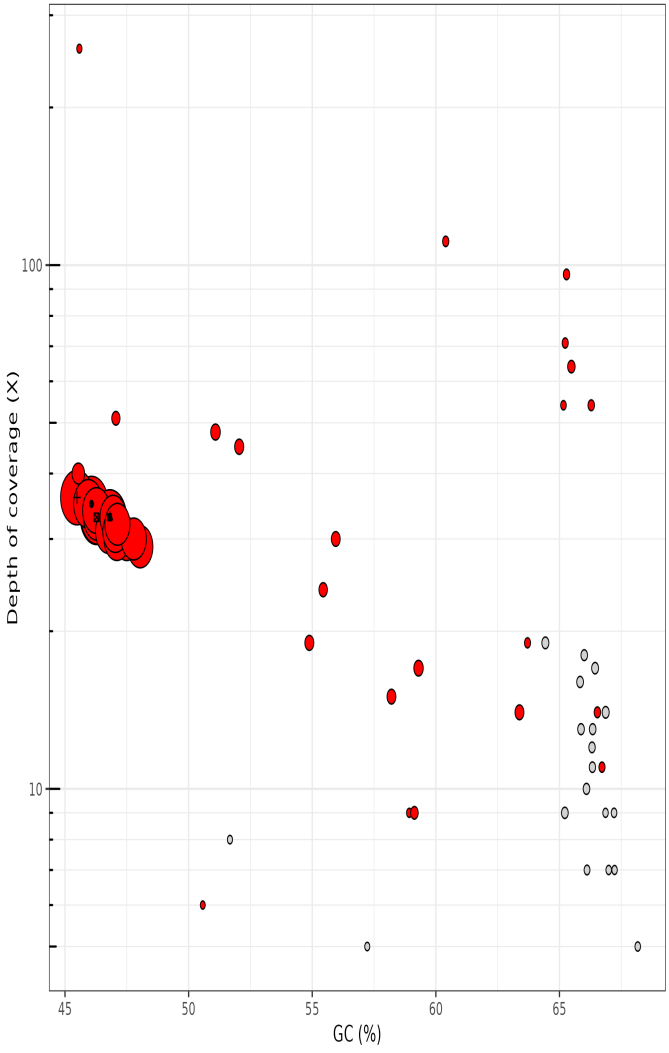


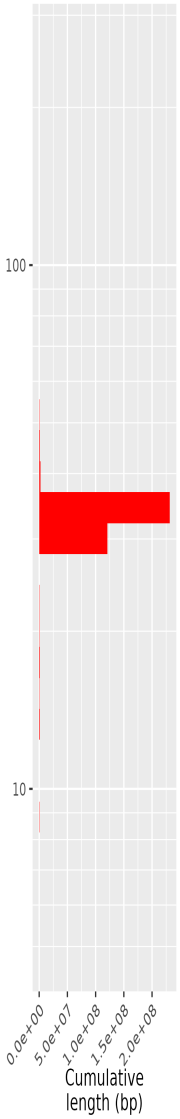
Distribution of k-mer counts coloured by their presence in reads/assemblies

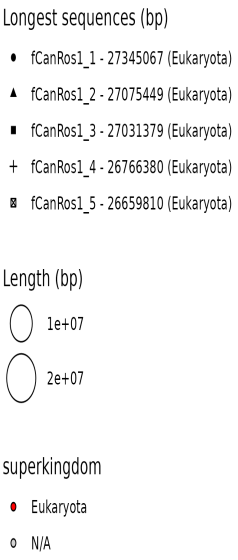
Post-curation contamination screening



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	33	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: Benjamin Istace
Affiliation: Genoscope

Date and time: 2025-09-25 07:07:10 CEST