ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	342441
ToLID	**fHetCru1**
Species	Heteropriacanthus cruentatus
Class	Actinopteri
Order	Priacanthiformes

Genome Traits	Expected	Observed
Haploid size (bp)	678,282,360	681,333,371
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:

Curator notes

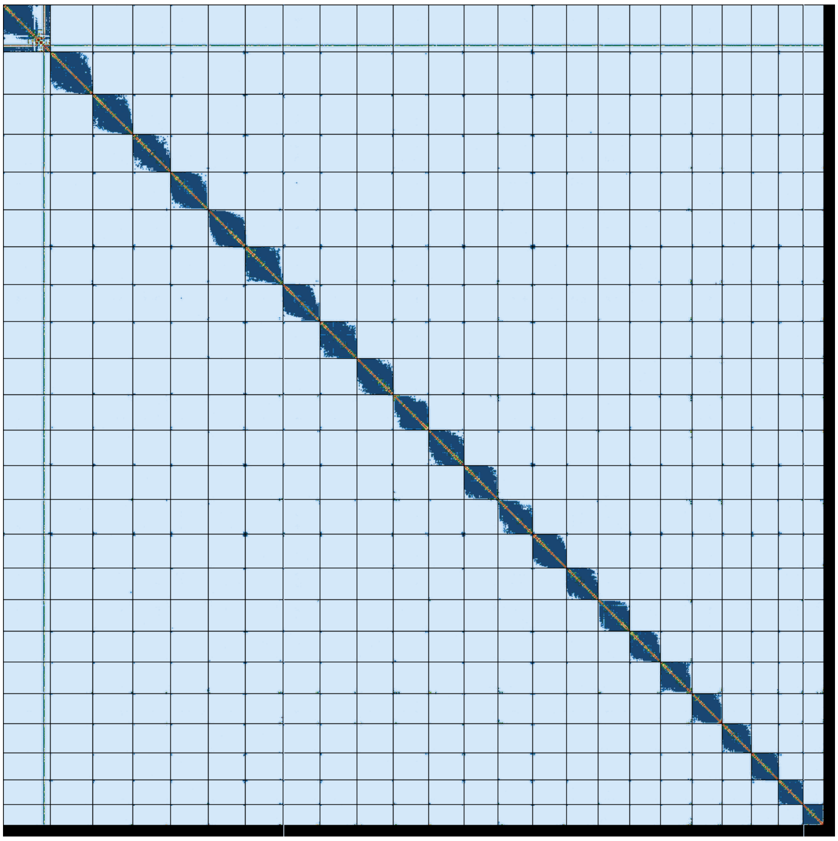
. Interventions/Gb: 10
. Contamination notes: ""
. Other observations: "The assembly of Heteropriacanthus cruentatus (fHetCru1.1) is based on 54X of PacBio data and 161X of Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following 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, 69 regions totaling 16 Mb (with the largest being 2.13 Mb) were identified as haplotypic duplications and removed. No contaminant sequences were detected. 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	681,336,671	681,333,371
GC %	42.27	42.27
Gaps/Gbp	82.19	86.59
Total gap bp	6,400	7,100
Scaffolds	71	66
Scaffold N50	29,071,764	29,071,764
Scaffold L50	11	11
Scaffold L90	22	22
Contigs	127	125
Contig N50	24,158,803	24,158,803
Contig L50	13	13
Contig L90	28	28
QV	49.652	49.6545
Kmer compl.	93.2629	93.2626
BUSCO sing.	97.3%	97.3%
BUSCO dupl.	0.1%	0.1%
BUSCO frag.	0.5%	0.5%
BUSCO miss.	2.0%	2.1%

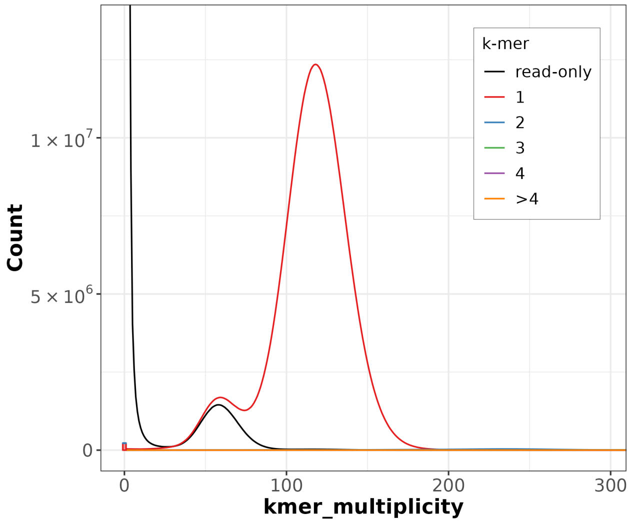
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

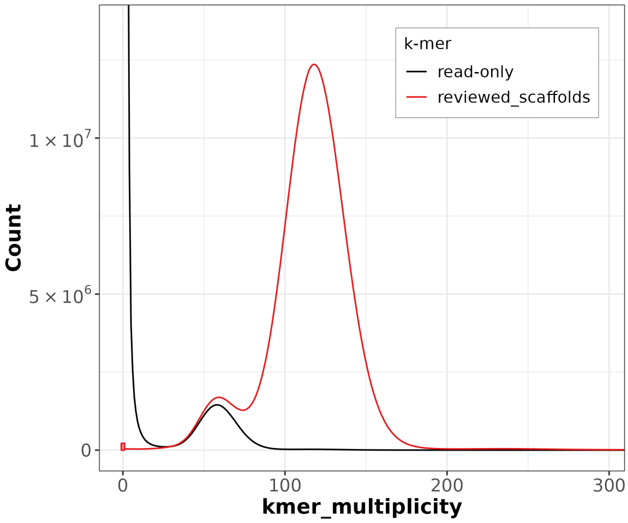


collapsed [LINK]

K-mer spectra of curated assembly

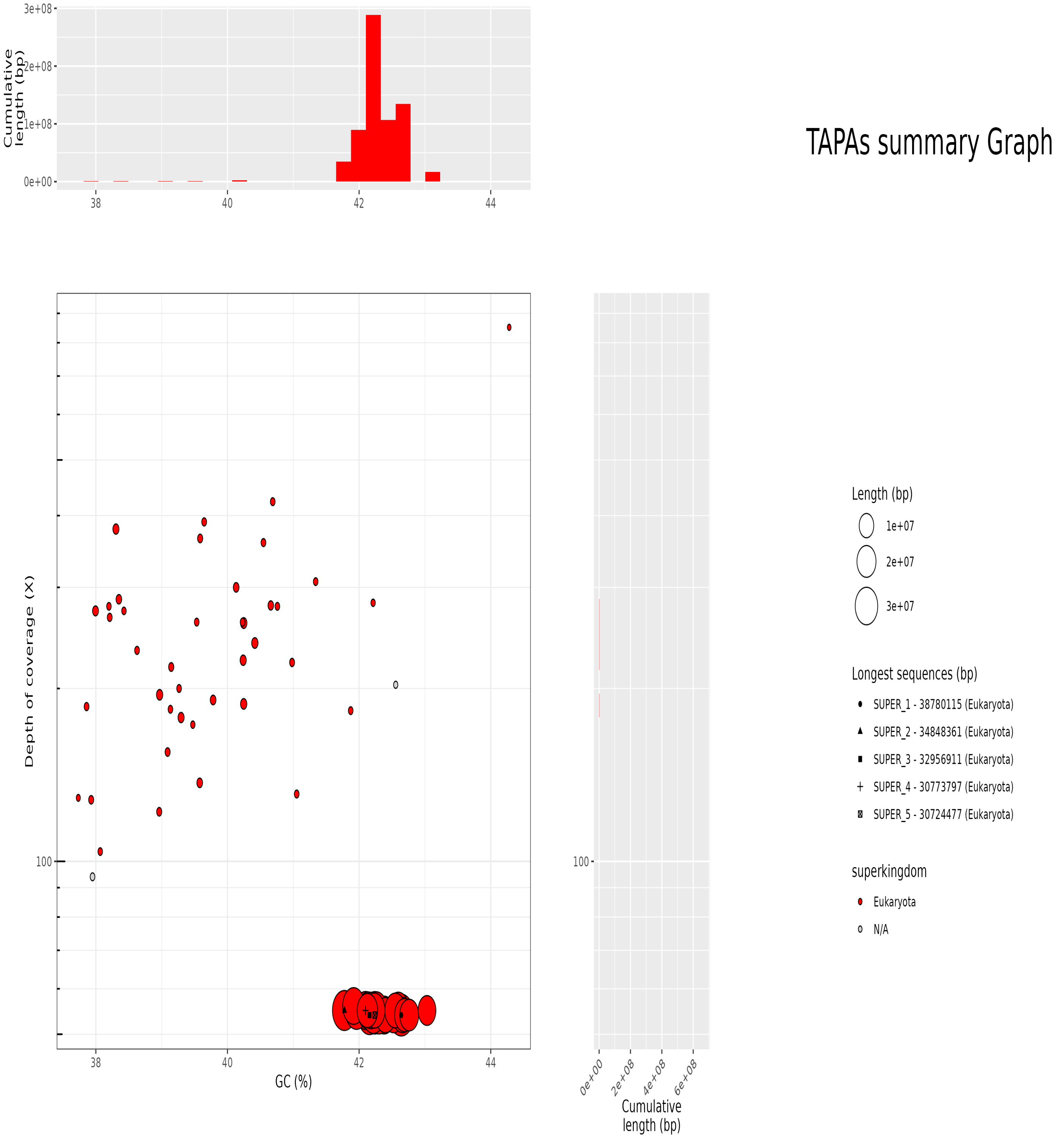


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	PACBIO Hifi	Arima
Coverage	54	161

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: 2025-09-09 20:55:16 CEST