ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	941138
ToLID	**eoAstMuri1**
Species	Astrophyton muricatum
Class	Ophiuroidea
Order	Euryalida

Genome Traits	Expected	Observed
Haploid size (bp)	3,083,036,455	3,126,075,112
Haploid Number	18 (source: ancestor)	27
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.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
. Not 90% of assembly in chromosomes for collapsed

Curator notes

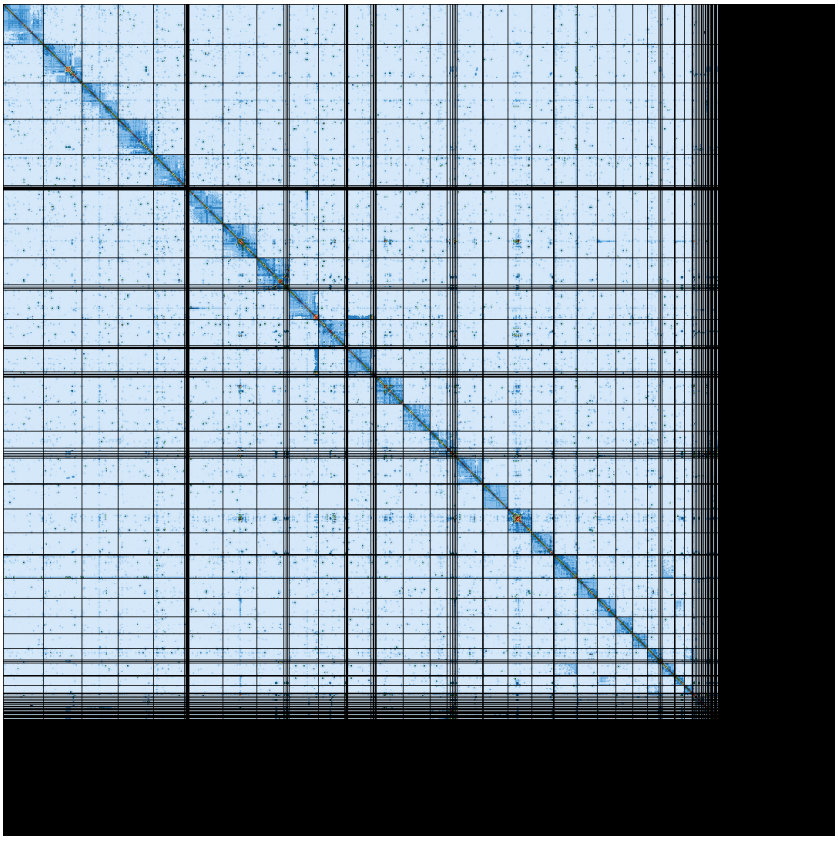
. Interventions/Gb: 17
. Contamination notes: ""
. Other observations: "The assembly of Astrophyton muricatum (eoAstMuri1) is based on 31X PacBio data and 148X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 642 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 60.709 Mb (with the largest being 2.268 Mb). Additionally, 1405 regions totaling 277.892 Mb (with the largest being 2.977 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, 3 haplotypic regions were removed, totaling 7.946Mb (with the largest being 3.352Mb). 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	3,134,028,465	3,126,075,112
GC %	36.75	36.76
Gaps/Gbp	127.95	138.51
Total gap bp	40,100	47,500
Scaffolds	1,069	1,035
Scaffold N50	75,908,373	94,380,292
Scaffold L50	15	14
Scaffold L90	158	127
Contigs	1,470	1,468
Contig N50	7,067,387	7,168,327
Contig L50	122	121
Contig L90	498	497
QV	49.5652	49.5638
Kmer compl.	80.5049	80.3949
BUSCO sing.	97.2%	97.2%
BUSCO dupl.	1.4%	1.4%
BUSCO frag.	0.8%	0.8%
BUSCO miss.	0.6%	0.6%

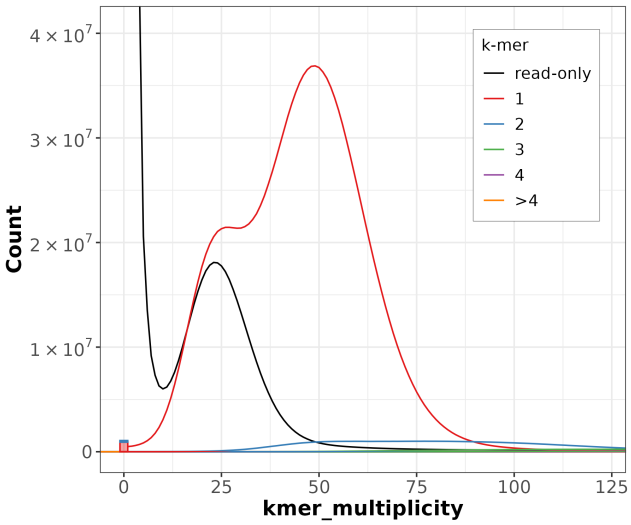
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12.2 (genomes:182, BUSCOs:932)

HiC contact map of curated assembly

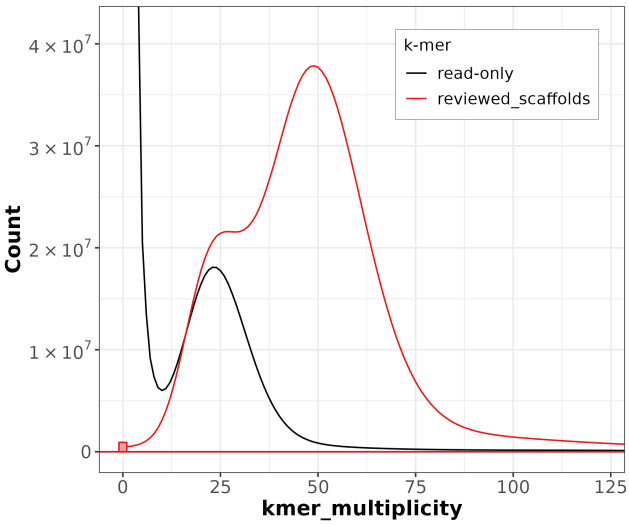


collapsed [LINK]

K-mer spectra of curated assembly

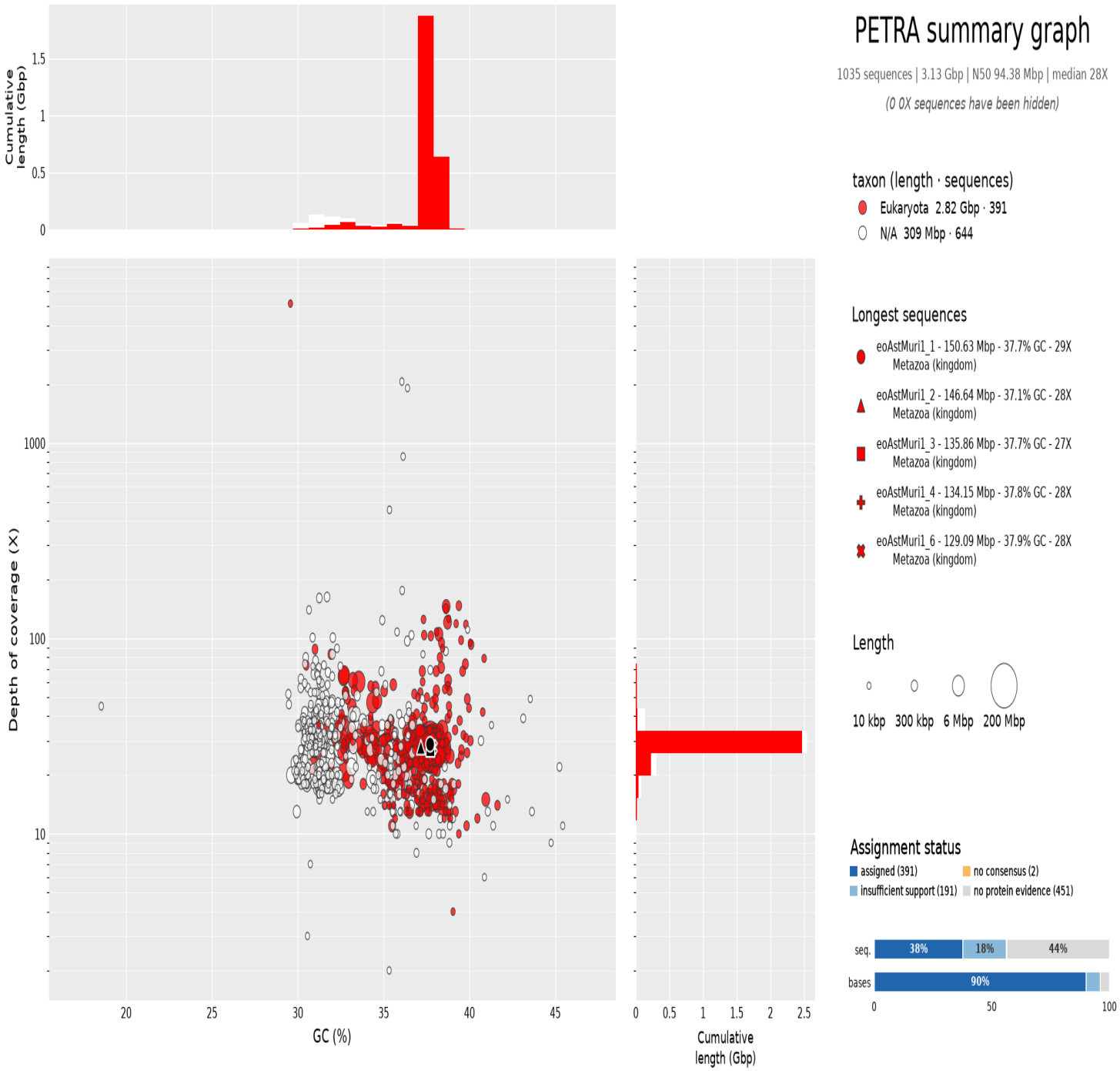


Distribution of k-mer counts per copy numbers found in asm



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	Long reads	Arima
Coverage	31	148

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:* 0.2.4
  |_ *key param:* NA
- **PretextView**
  |_ *ver:* 1.0.7
  |_ *key param:* NA

Submitter: Arnaud Couloux
Affiliation: Genoscope

Date and time: 2026-10-28 17:27:19 CET