ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	60560
ToLID	**eaAstPoly1**
Species	Astropecten polyacanthus
Class	Asteroidea
Order	Paxillosida

Genome Traits	Expected	Observed
Haploid size (bp)	472,123,037	488,081,401
Haploid Number	20 (source: ancestor)	22
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q50

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

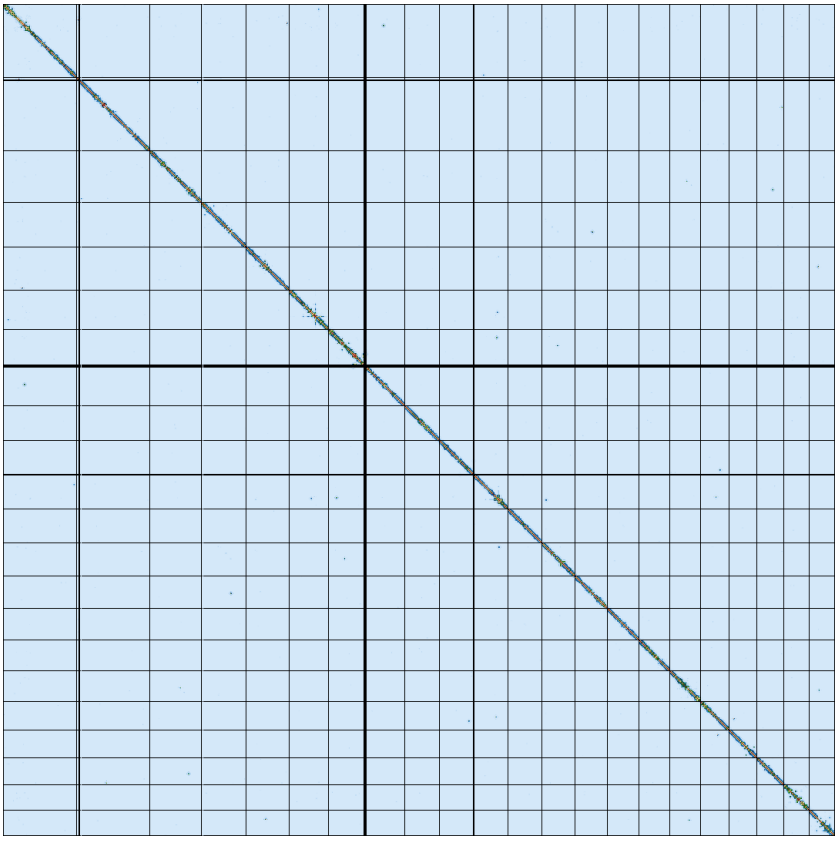
. Interventions/Gb: 100
. Contamination notes: ""
. Other observations: "The assembly of Astropecten polyacanthus (eaAstPoly1) is based on 63X PacBio data and 203X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 1 contigs was identified as contaminants (bacterial, archaeal, or viral), totaling 0.268 Mb (with the largest being 0.268 Mb). Additionally, 148 regions totaling 13.166 Mb (with the largest being 1.303 Mb) were identified as haplotypic duplications and removed. During manual curation, 4 haplotypic regions were removed, totaling 2.26Mb (with the largest being 0.981Mb). 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	490,328,783	488,081,401
GC %	41.92	41.92
Gaps/Gbp	59.14	108.59
Total gap bp	2,900	9,000
Scaffolds	76	49
Scaffold N50	19,913,332	20,675,991
Scaffold L50	9	9
Scaffold L90	20	20
Contigs	105	102
Contig N50	13,798,000	13,798,000
Contig L50	14	14
Contig L90	36	35
QV	50.3798	50.3767
Kmer compl.	68.7293	68.5863
BUSCO sing.	98.1%	98.1%
BUSCO dupl.	0.3%	0.3%
BUSCO frag.	1.2%	1.2%
BUSCO miss.	0.4%	0.4%

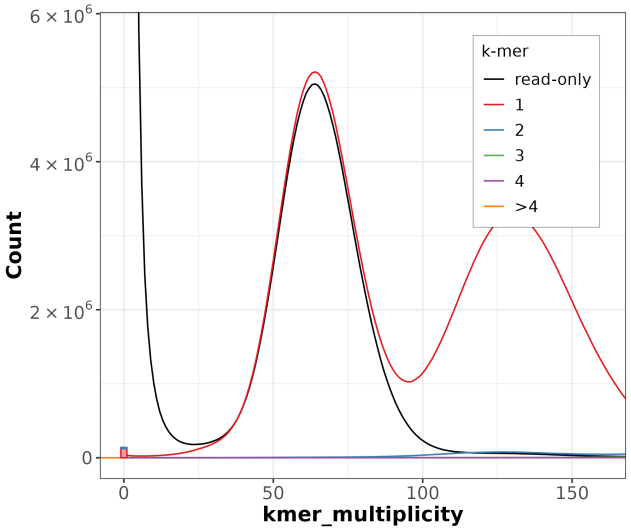
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

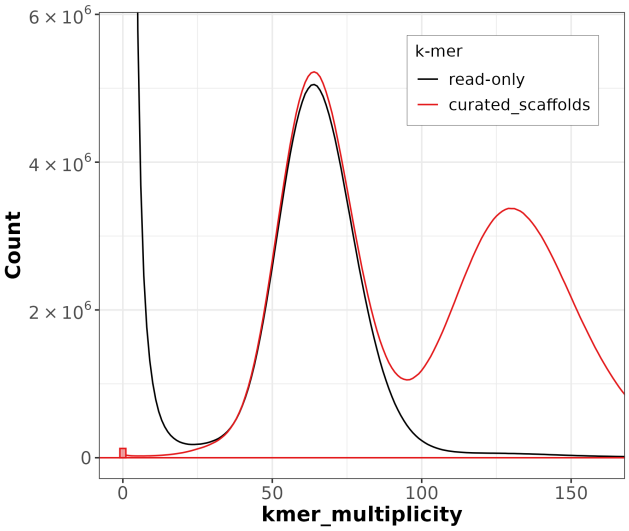


collapsed [LINK]

K-mer spectra of curated assembly

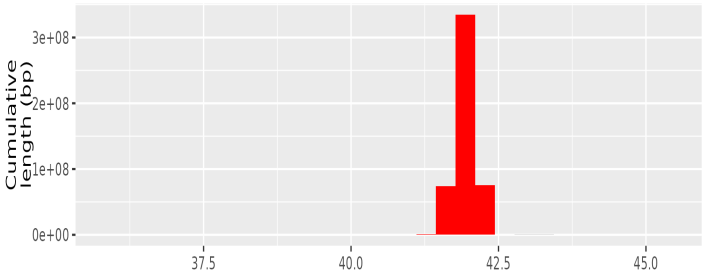


Distribution of k-mer counts per copy numbers found in asm

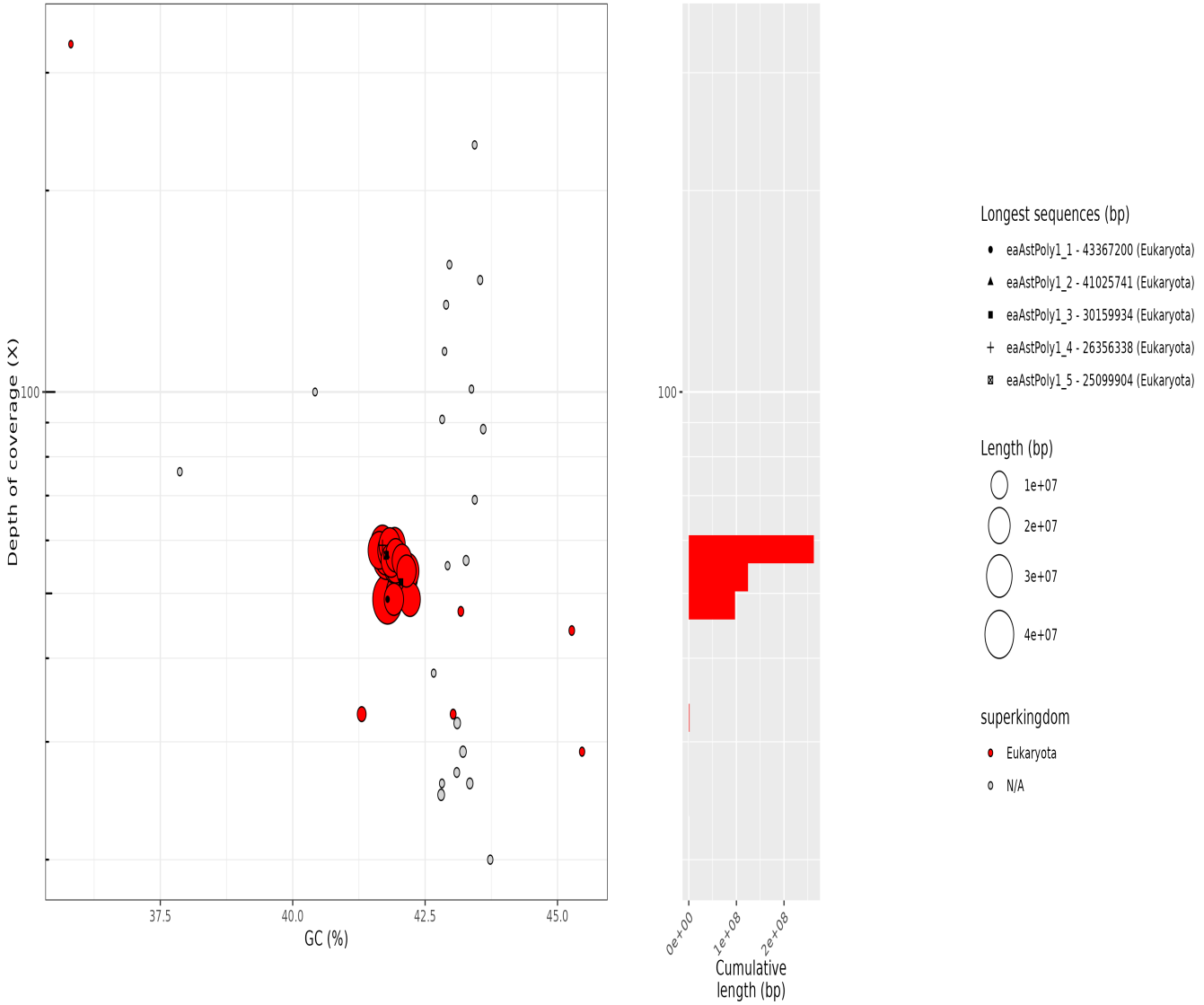


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	61	197

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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2026-03-14 10:04:45 CET