ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	167016
ToLID	**xpAcaGemm1**
Species	Acanthopleura gemmata
Class	Polyplacophora
Order	Chitonida

Genome Traits	Expected	Observed
Haploid size (bp)	703,882,946	719,731,633
Haploid Number	9 (source: ancestor)	12
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q48

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

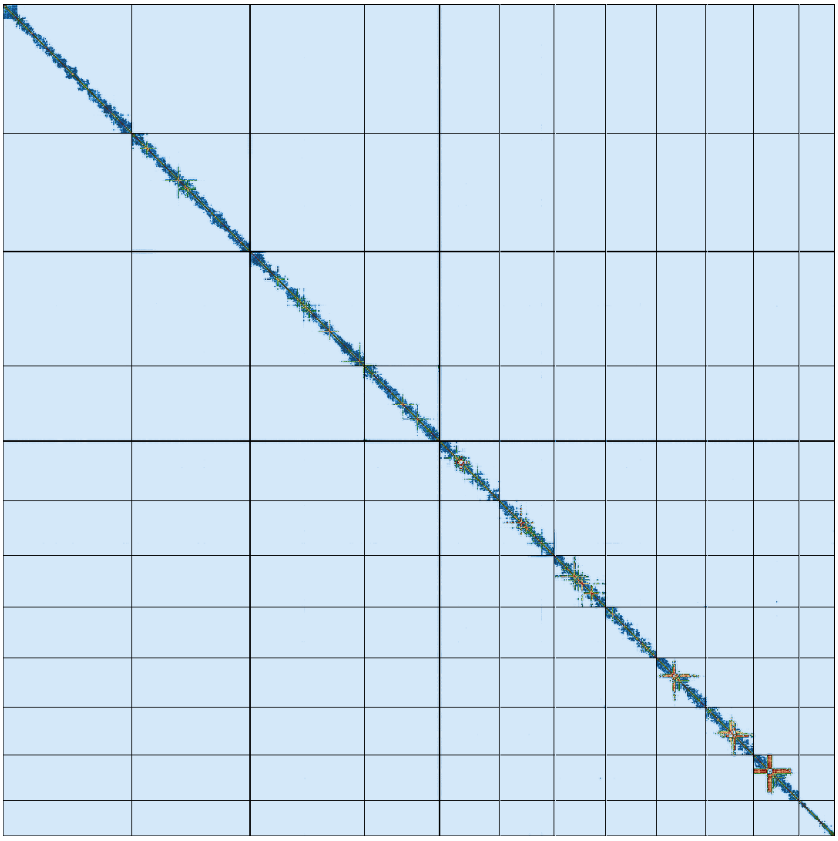
. Interventions/Gb: 41
. Contamination notes: ""
. Other observations: "The assembly of Acanthopleura gemmata (xpAcaGemm1) is based on 53X PacBio data and 146X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 7 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.227 Mb (with the largest being 0.064 Mb). Additionally, 133 regions totaling 28.649 Mb (with the largest being 5.174 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. During manual curation, 5 haplotypic regions and 2 contaminant sequences were removed, totaling 2.071Mb and 0.107 Mb respectively (with the largest being 0.682Mb and 0.076Mb). 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	721,893,155	719,731,633
GC %	40.64	40.64
Gaps/Gbp	27.7	41.68
Total gap bp	2,000	4,800
Scaffolds	49	36
Scaffold N50	64,603,748	64,603,748
Scaffold L50	4	4
Scaffold L90	11	10
Contigs	69	66
Contig N50	32,847,218	32,847,218
Contig L50	7	7
Contig L90	21	21
QV	48.9996	48.9971
Kmer compl.	85.4419	85.3763
BUSCO sing.	95.5%	95.8%
BUSCO dupl.	1.1%	0.7%
BUSCO frag.	1.1%	1.1%
BUSCO miss.	2.4%	2.4%

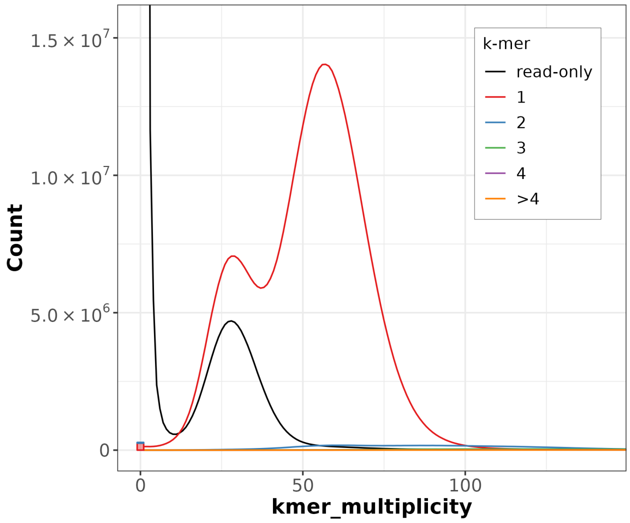
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

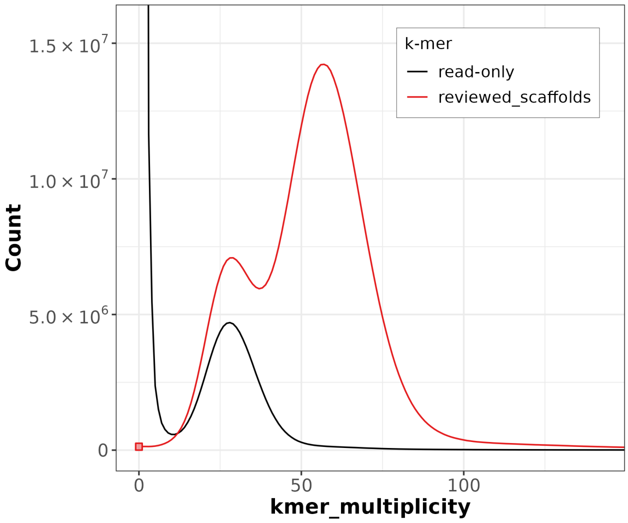


collapsed [LINK]

K-mer spectra of curated assembly

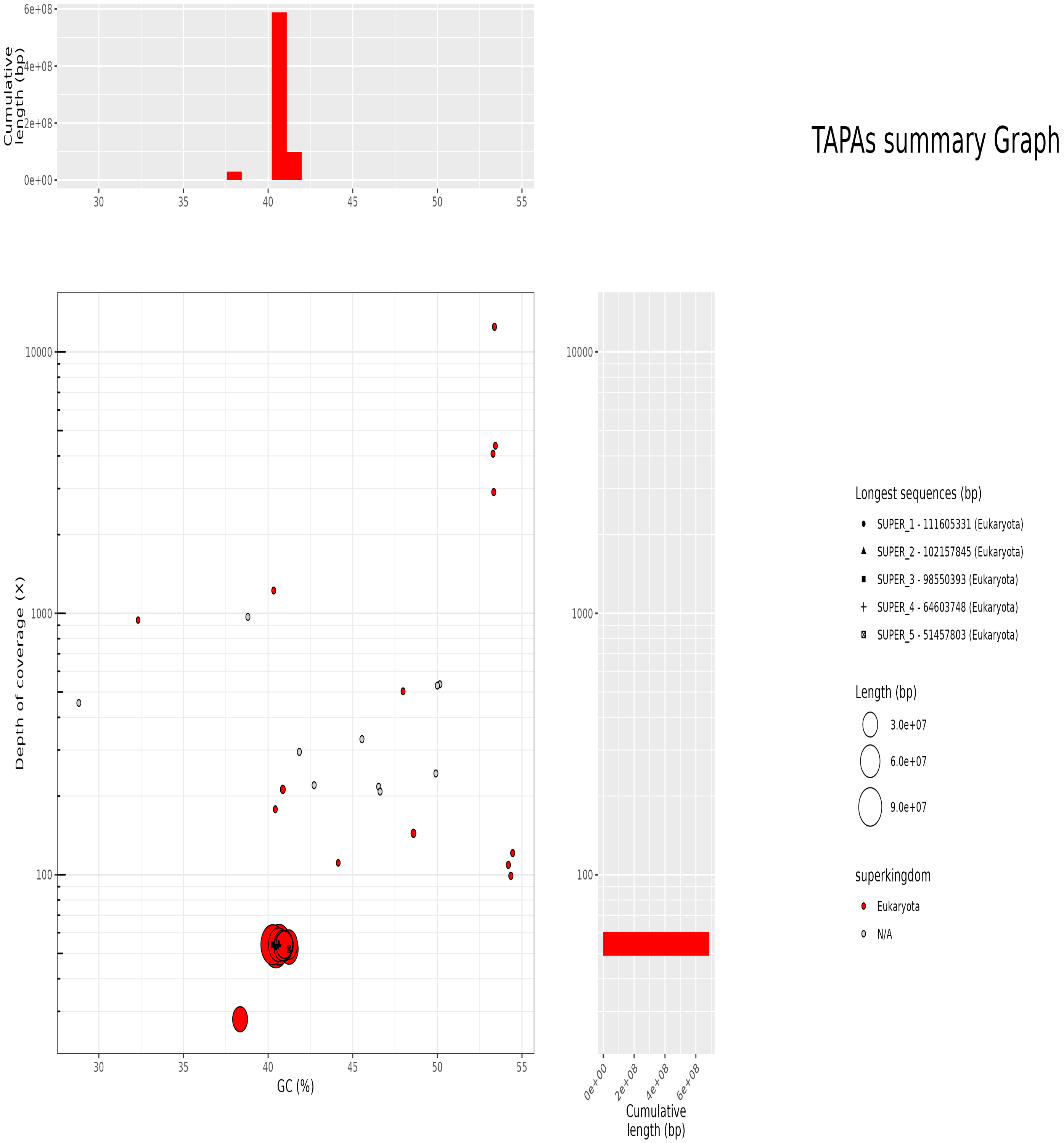


Distribution of k-mer counts per copy numbers found in asm



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	53	146

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: Sophie Layac
Affiliation: Genoscope

Date and time: 2025-11-14 07:53:24 CET