ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	55704
ToLID	**xpAcaFasc5**
Species	Acanthochitona fascicularis
Class	Polyplacophora
Order	Chitonida

Genome Traits	Expected	Observed
Haploid size (bp)	1,663,967,742	1,702,392,739
Haploid Number	9 (source: ancestor)	9
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q47

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

. Kmer completeness value is less than 90 for collapsed
. BUSCO single copy value is less than 90% for collapsed

Curator notes

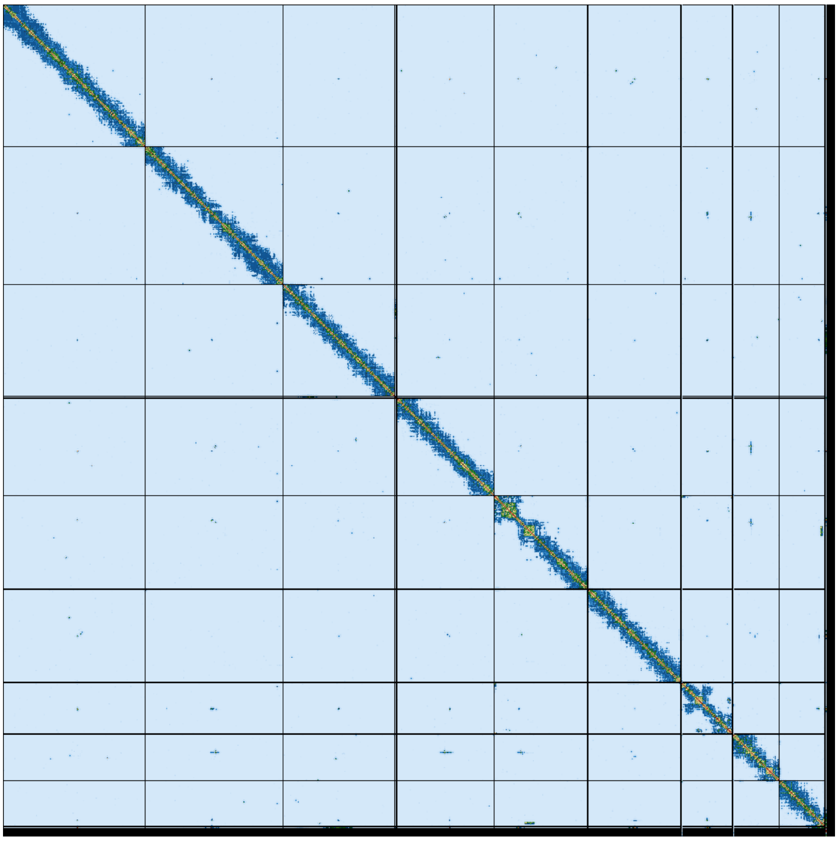
. Interventions/Gb: 69
. Contamination notes: ""
. Other observations: "The assembly of Acanthochitona fascicularis (xpAcaFasc5) is based on 37X PacBio data and 171X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: 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, 462 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 19.1 Mb (with the largest being 2.6 Mb). Additionally, 904 regions regions totaling 336 Mb (with the largest being 10 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 50 haplotypic regions and 10 contaminant sequences were removed, totaling 21 Mb and 0.15 Mb, respectively (with the largest being 1.5 Mb and 21 kb). 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	1,733,151,989	1,702,392,739
GC %	41.08	41.08
Gaps/Gbp	200.79	234.38
Total gap bp	34,800	46,300
Scaffolds	166	121
Scaffold N50	233,284,280	198,769,152
Scaffold L50	4	4
Scaffold L90	8	8
Contigs	514	520
Contig N50	7,984,887	7,997,289
Contig L50	64	58
Contig L90	219	209
QV	47.1095	47.2885
Kmer compl.	67.6049	66.8404
BUSCO sing.	75.1%	76.1%
BUSCO dupl.	3.1%	2.0%
BUSCO frag.	4.7%	4.7%
BUSCO miss.	17.1%	17.2%

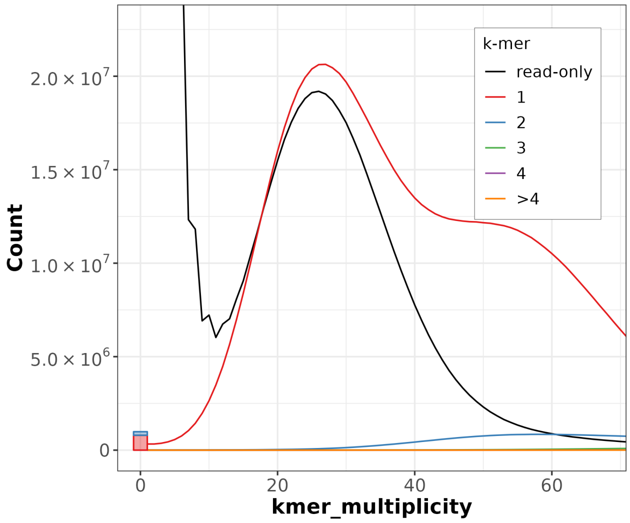
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: mollusca_odb10 (genomes:7, BUSCOs:5295)

HiC contact map of curated assembly

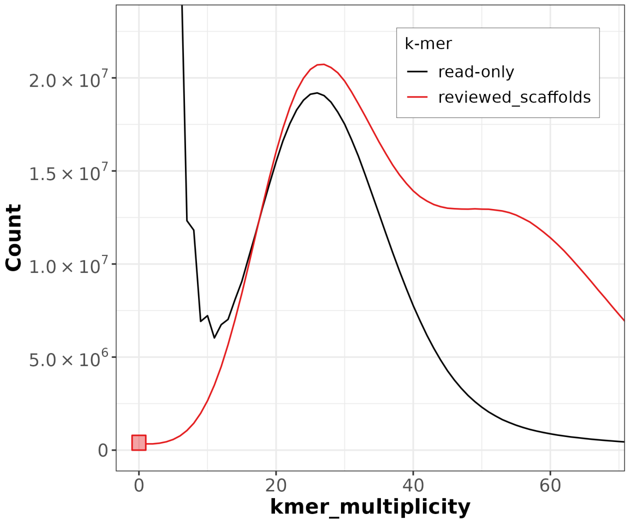


collapsed [LINK]

K-mer spectra of curated assembly

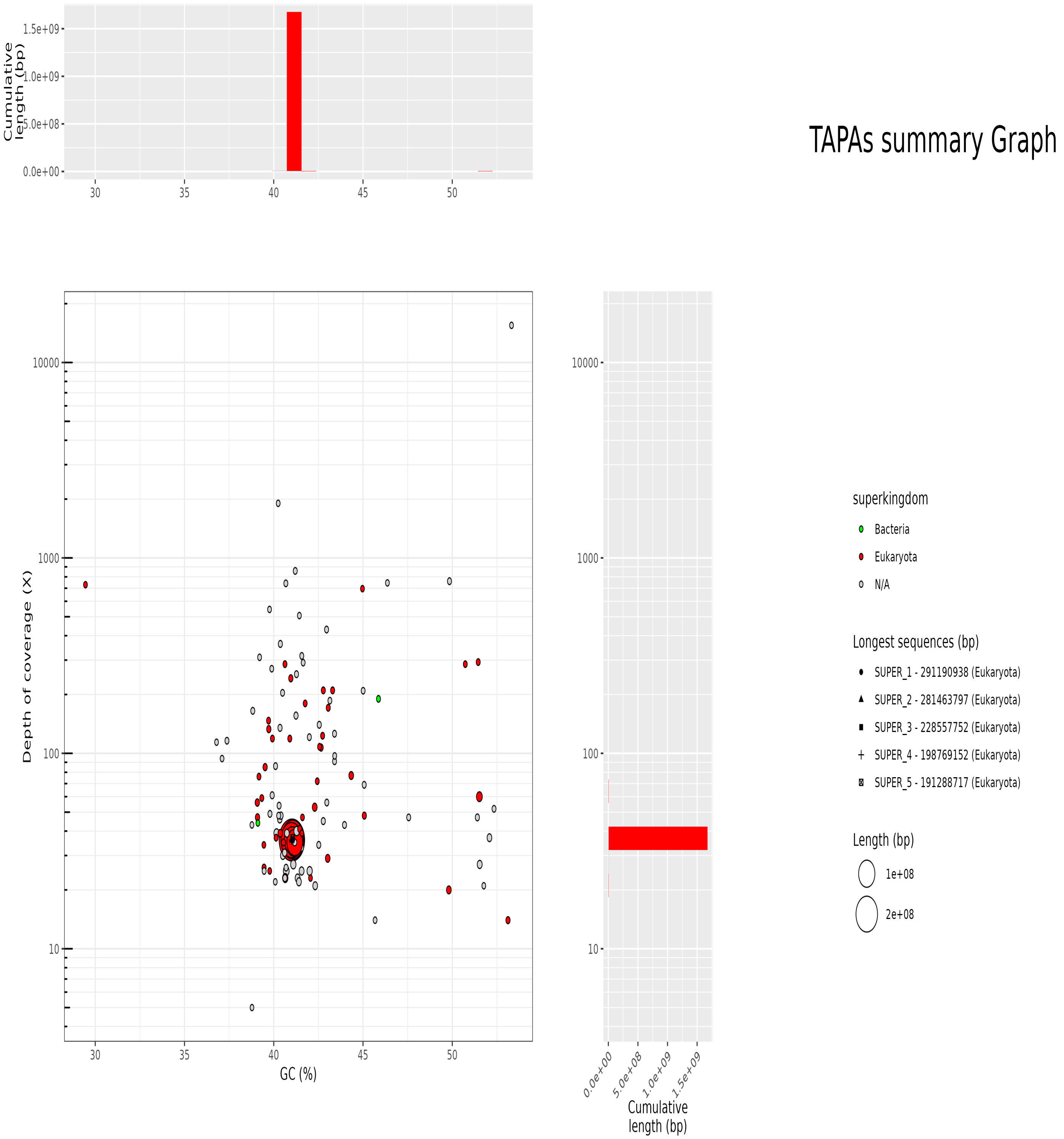


Distribution of k-mer counts per copy numbers found in asm



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	37	155

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: 2025-05-05 10:57:50 CEST