

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	1582122
ToLID	xbCenAcul1
Species	Centrocardita aculeata
Class	Bivalvia
Order	Carditida

Genome Traits	Expected	Observed
Haploid size (bp)	3,502,046,961	3,766,937,780
Haploid Number	19 (source: ancestor)	15
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q44

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

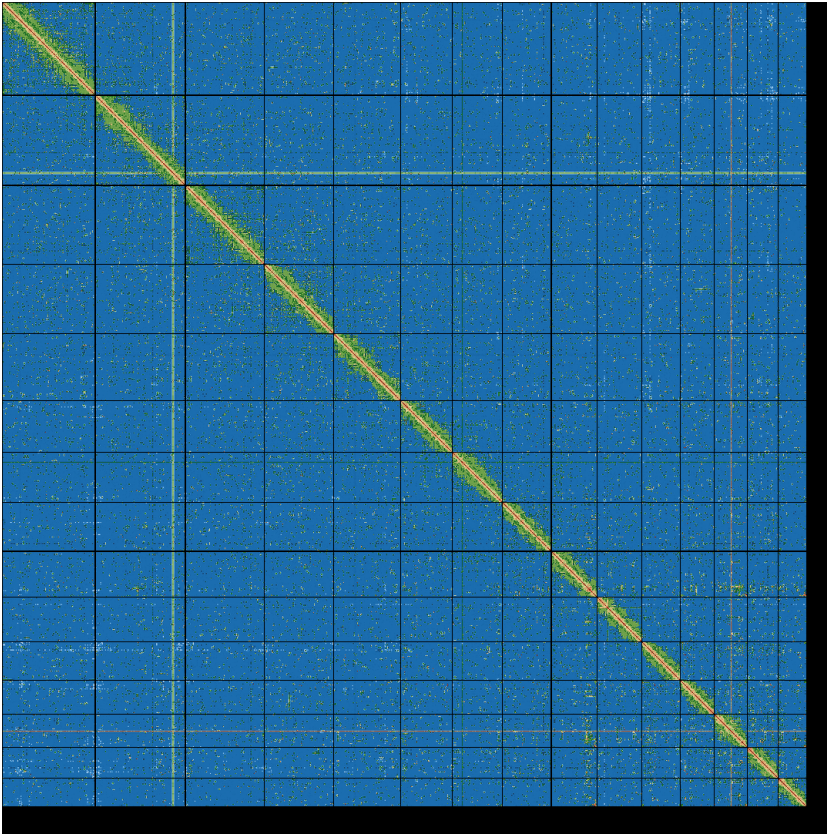
. Interventions/Gb: 55
. Contamination notes: ""
. Other observations: "The assembly of *Centrocardita aculeata* (xbCenAcul1) is based on 33.64X PacBio data and 199.46X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: 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, 76 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 17.768 Mb (with the largest being 3.664 Mb). Additionally, 2471 regions totaling 188.163 Mb (with the largest being 1.159 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 64 haplotypic regions were removed, totaling 79.272Mb (with the largest being 4.568Mb). 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	3,852,658,183	3,766,937,780
GC %	37.81	37.81
Gaps/Gbp	938.57	929.93
Total gap bp	361,600	361,800
Scaffolds	848	722
Scaffold N50	238,076,307	233,999,345
Scaffold L50	6	6
Scaffold L90	14	14
Contigs	4,464	4,225
Contig N50	1,780,197	1,803,430
Contig L50	651	631
Contig L90	2,273	2,171
QV	44.617	44.6689
Kmer compl.	90.6588	89.2027
BUSCO sing.	96.0%	96.6%
BUSCO dupl.	2.3%	1.7%
BUSCO frag.	0.5%	0.6%
BUSCO miss.	1.2%	1.2%

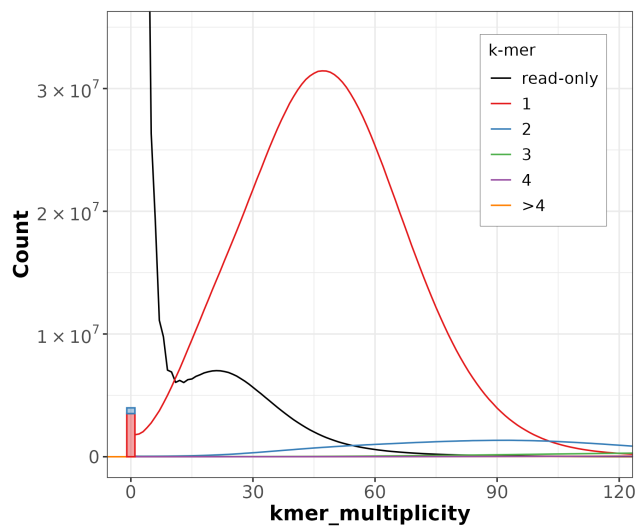
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12.2 (genomes:31, BUSCOs:4236)

HiC contact map of curated assembly

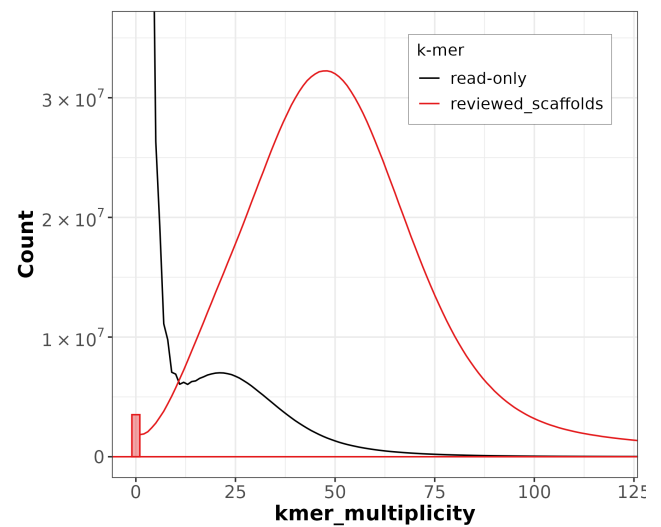


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

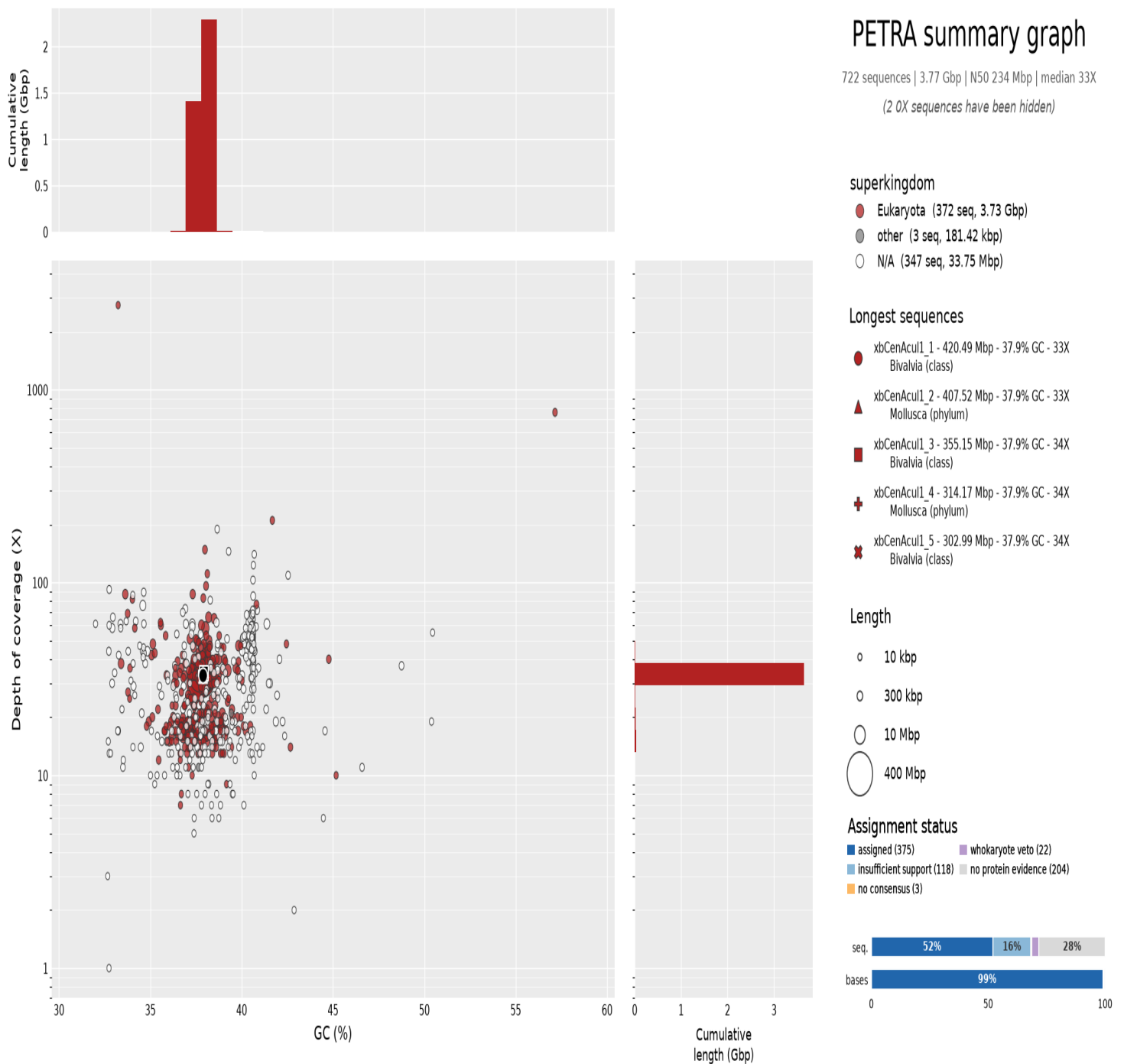


Distribution of k-mer counts per copy numbers found in asm



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	34	204

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: Emilie Teodori

Affiliation: Genoscope

Date and time: 2026-09-10 00:54:55 CEST