ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1028561
ToLID	**xbGarDepr1**
Species	Gari depressa
Class	Bivalvia
Order	Cardiida

Genome Traits	Expected	Observed
Haploid size (bp)	1,852,107,467	1,966,443,956
Haploid Number	26 (source: ancestor)	19
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q47

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

Curator notes

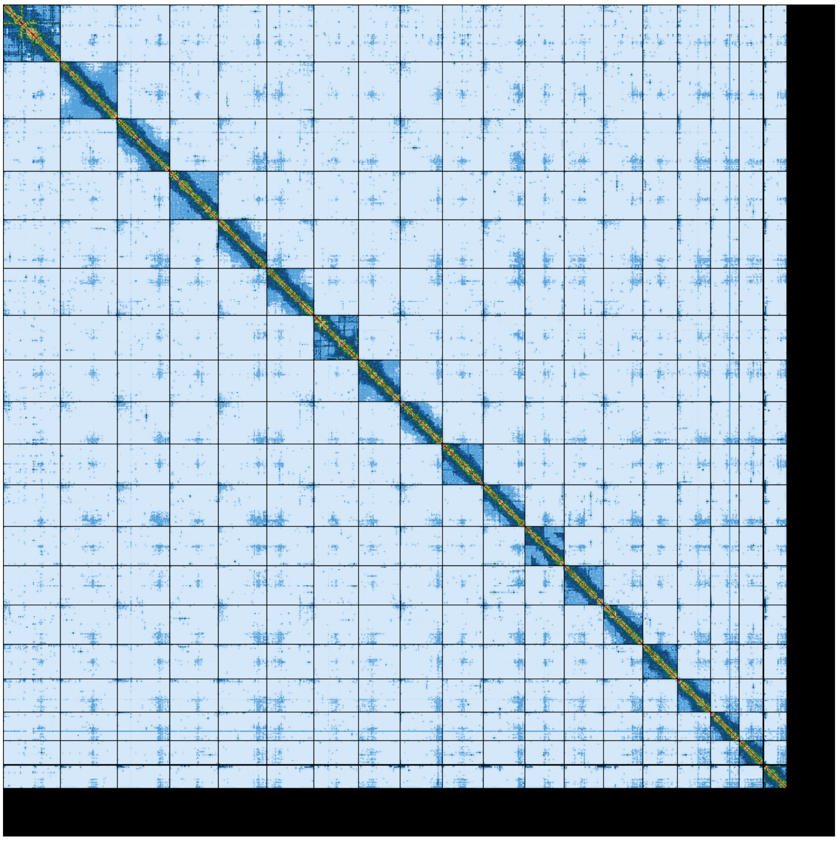
. Interventions/Gb: 66
. Contamination notes: ""
. Other observations: "The assembly of Gari depressa (xbGarDepr1) is based on 130X PacBio data and 178X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 368 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 7.094 Mb (with the largest being 0.53 Mb). Additionally, 6214 regions totaling 129.83 Mb (with the largest being 6.632 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using ptGaul. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 10 haplotypic regions were removed, totaling 12.7Mb, (with the largest being 4.4Mb). 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	2,015,514,819	1,966,443,956
GC %	36.67	36.58
Gaps/Gbp	27.29	30.51
Total gap bp	5,500	7,500
Scaffolds	14,811	1,754
Scaffold N50	98,752,486	98,752,486
Scaffold L50	9	9
Scaffold L90	19	18
Contigs	14,866	1,814
Contig N50	39,237,829	39,712,000
Contig L50	18	17
Contig L90	60	55
QV	45.6772	47.6133
Kmer compl.	70.1112	69.9607
BUSCO sing.	87.8%	88.2%
BUSCO dupl.	0.8%	0.5%
BUSCO frag.	8.9%	8.9%
BUSCO miss.	2.4%	2.4%

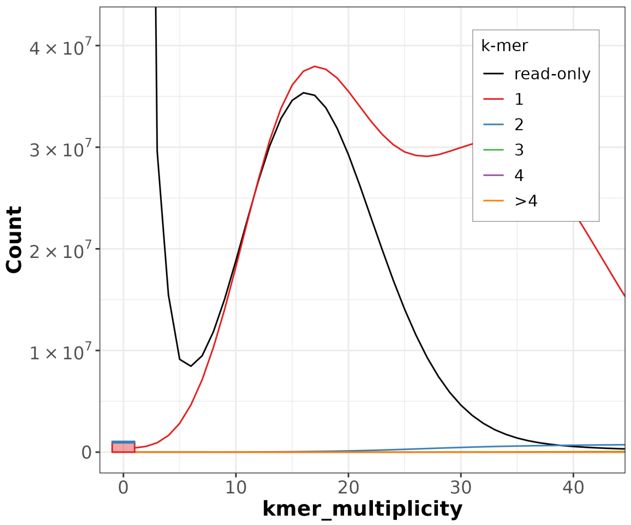
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

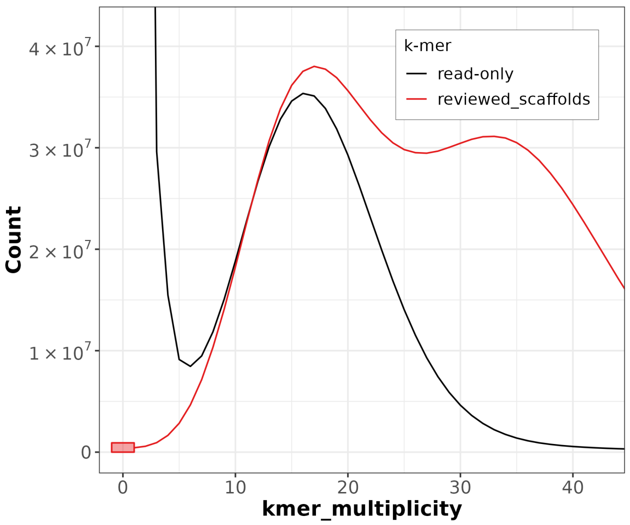


collapsed [LINK]

K-mer spectra of curated assembly

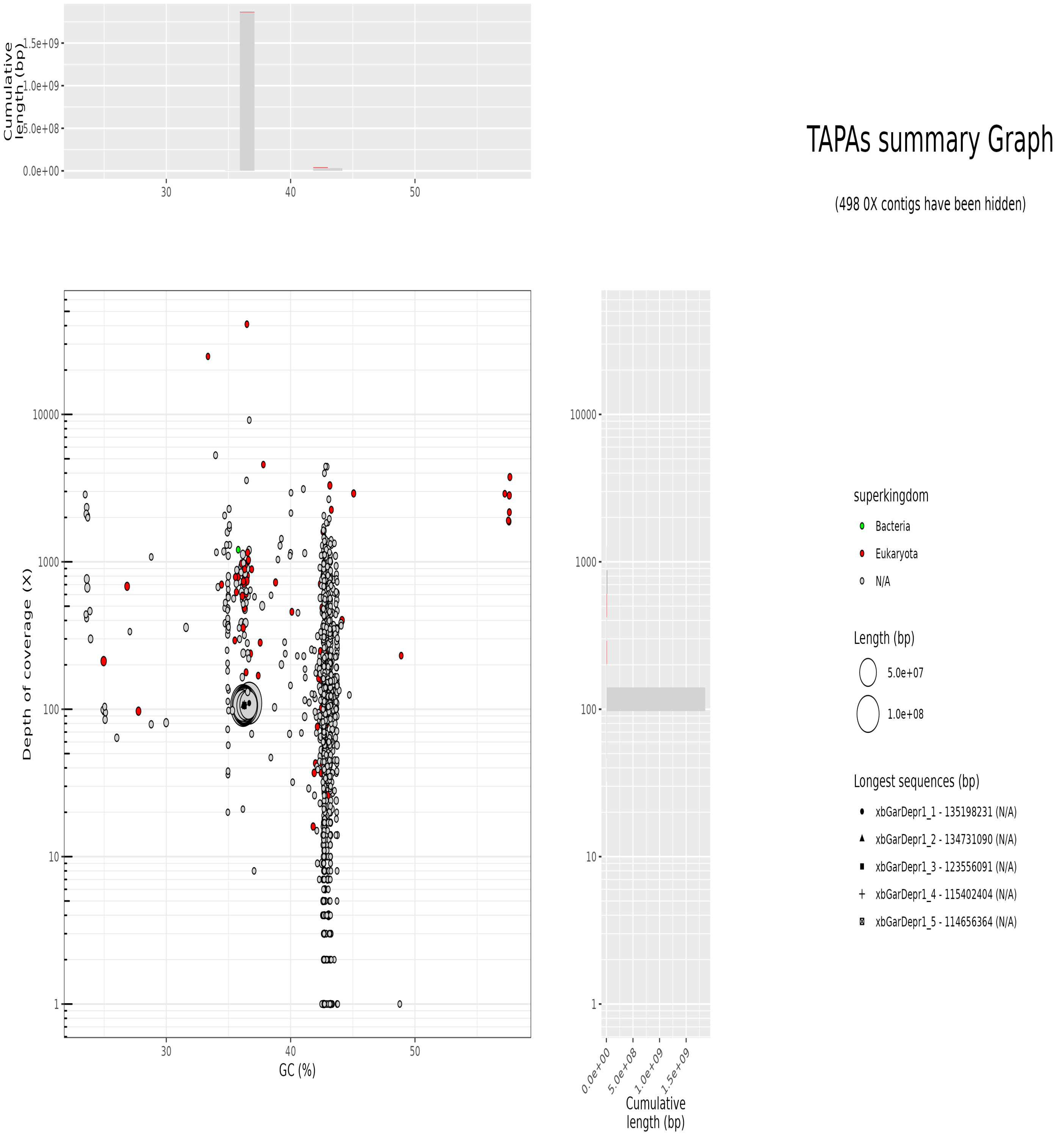


Distribution of k-mer counts per copy numbers found in asm



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	PACBIO Hifi	Arima
Coverage	130	178

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: Emilie Teodori
Affiliation: Genoscope

Date and time: 2025-07-25 22:00:18 CEST