ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	138986
ToLID	**xbPitRudi1**
Species	Pitar rudis
Class	Bivalvia
Order	Venerida

Genome Traits	Expected	Observed
Haploid size (bp)	1,125,445,532	1,138,664,234
Haploid Number	19 (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.Q52

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

. Kmer completeness value is less than 90 for collapsed
. Assembly length loss > 3% for collapsed

Curator notes

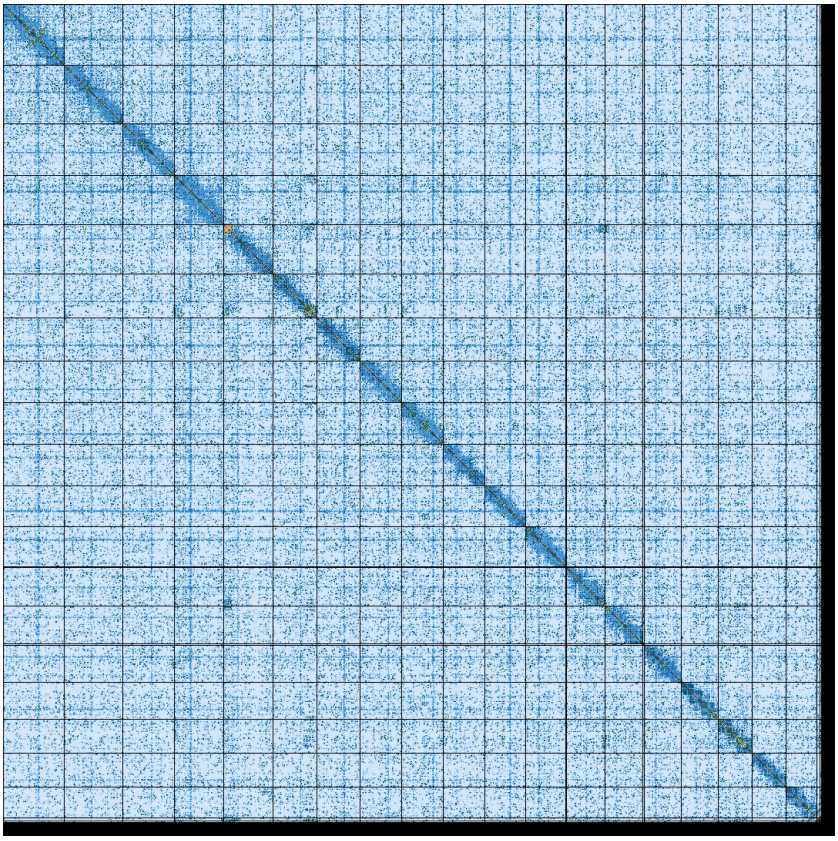
. Interventions/Gb: 115
. Contamination notes: ""
. Other observations: "The assembly of Pitar rudis (xbPitRudi1) is based on 46X PacBio data and 102X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 4 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.14 Mb (with the largest being 0.051 Mb). Additionally, 261 regions totaling 136.628 Mb (with the largest being 7.208 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, 65 haplotypic regions and 4 contaminant sequences were removed, totaling 483.593Mb and 0.245Mb, respectively (with the largest being 38.383Mb and 0.116Mb). The telomeric pattern TTAGGG was identified with TelFinder and used to generate the Pretext Telomeres tracks. 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	1,622,902,277	1,138,664,234
GC %	36.9	36.91
Gaps/Gbp	131.25	141.39
Total gap bp	21,300	20,500
Scaffolds	179	148
Scaffold N50	76,465,788	57,006,686
Scaffold L50	7	9
Scaffold L90	17	18
Contigs	392	309
Contig N50	18,156,000	17,919,265
Contig L50	31	21
Contig L90	102	69
QV	52.6192	52.6175
Kmer compl.	78.9253	59.974
BUSCO sing.	53.0%	94.7%
BUSCO dupl.	46.2%	0.9%
BUSCO frag.	0.3%	2.1%
BUSCO miss.	0.4%	2.3%

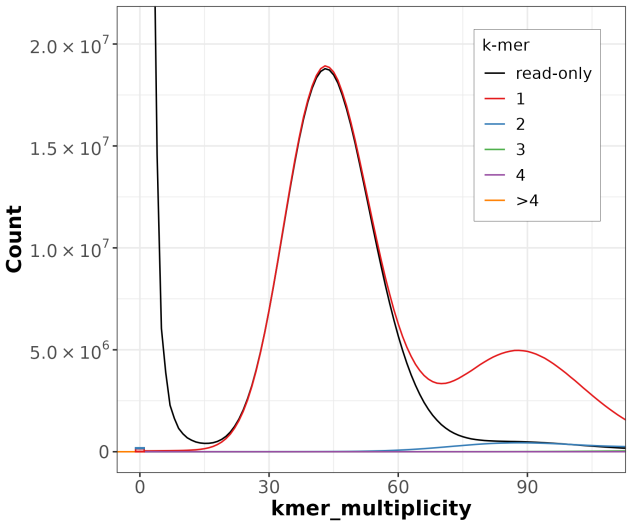
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

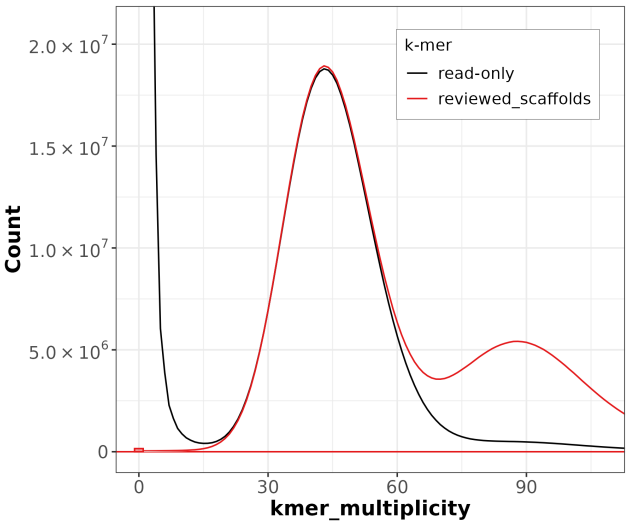


collapsed [LINK]

K-mer spectra of curated assembly

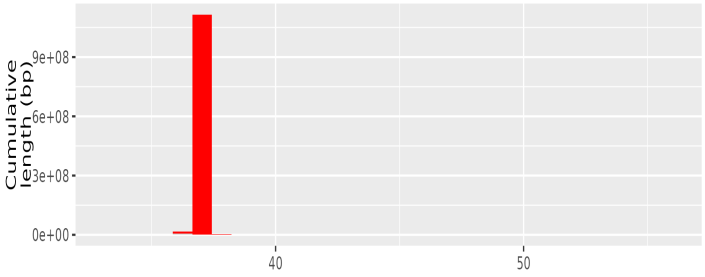


Distribution of k-mer counts per copy numbers found in asm

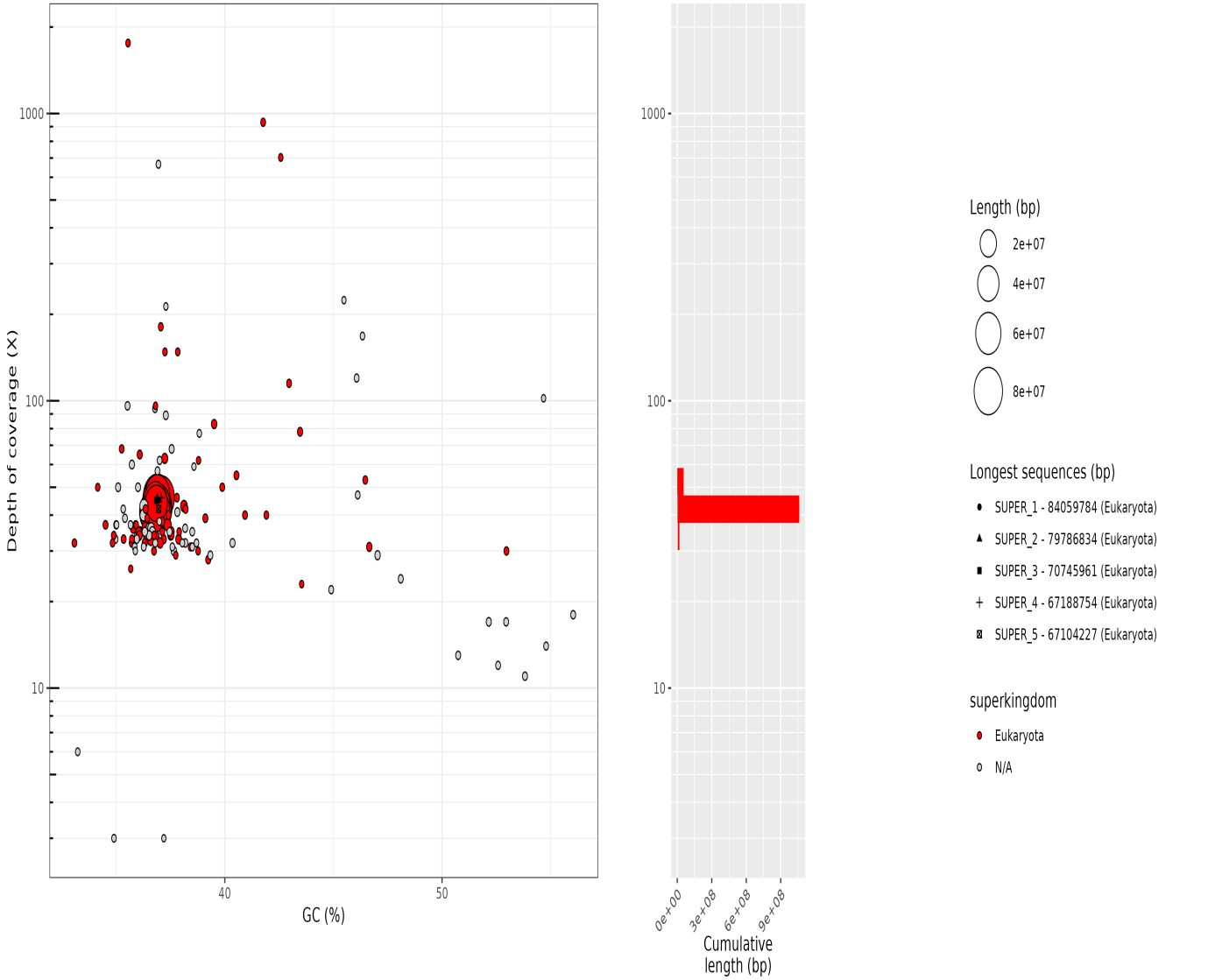


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



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	65	145

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: 2026-08-01 15:48:36 CEST