

# ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	217773
ToLID	<b>xgOvuOvum1</b>
Species	Ovula ovum
Class	Gastropoda
Order	Littorinimorpha

Genome Traits	Expected	Observed
Haploid size (bp)	1,550,539,792	1,635,348,073
Haploid Number	17 (source: ancestor)	36
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q37

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

- . Observed Haploid Number is different from Expected
- . QV value is less than 40 for collapsed
- . Kmer completeness value is less than 90 for collapsed
- . BUSCO single copy value is less than 90% for collapsed
- . BUSCO duplicated value is more than 5% for collapsed
- . More than 1000 gaps/Gbp for collapsed

### Curator notes

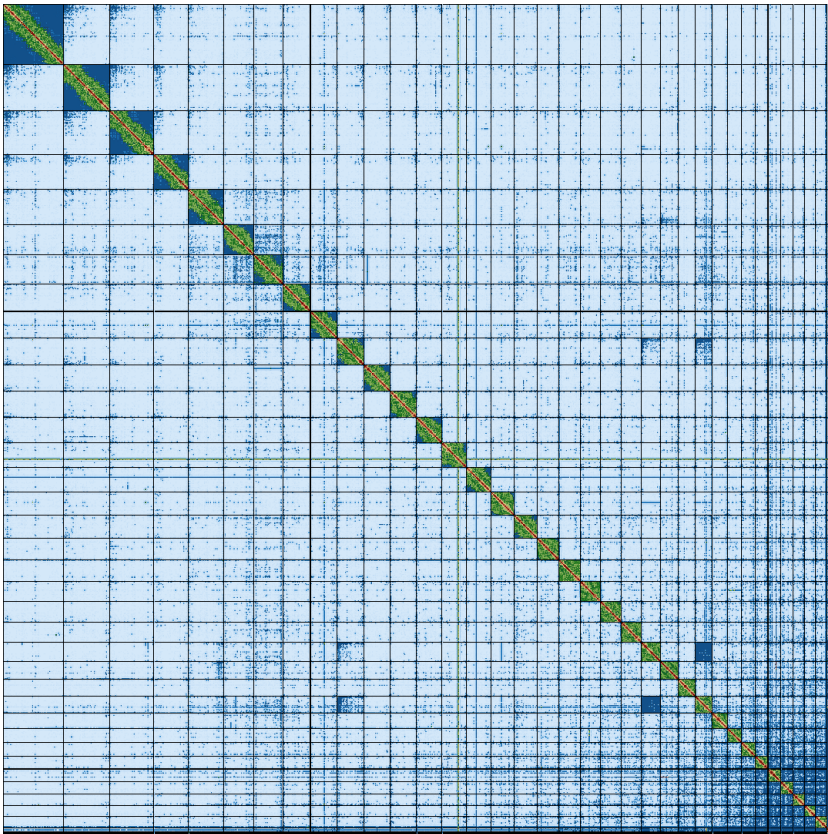
- . Interventions/Gb: 21
- . Contamination notes: ""
- . Other observations: "The assembly of Ovula ovum (xgOvuOvum1) is based on 32X ONT data and 146X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial ONT assembly generation with Flye, removal of contaminant sequences using Context, removal of haplotypic duplications using purge\_dups, and Hi-C-based scaffolding with YaHS. In total, 65 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 7.082 Mb (with the largest being 4.422 Mb). Additionally, 1367 regions totaling 46.158 Mb (with the largest being 0.27 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. 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	1,635,396,278	1,635,348,073
GC %	41.91	41.91
Gaps/Gbp	1,114.1	1,120.25
Total gap bp	182,200	185,400
Scaffolds	470	425
Scaffold N50	49,931,715	49,931,715
Scaffold L50	13	13
Scaffold L90	29	29
Contigs	2,292	2,257
Contig N50	2,008,797	2,008,797
Contig L50	245	245
Contig L90	852	852
QV	37.2516	37.252
Kmer compl.	89.49	89.4882
BUSCO sing.	78.6%	78.6%
BUSCO dupl.	18.3%	18.3%
BUSCO frag.	0.9%	1.0%
BUSCO miss.	2.2%	2.2%

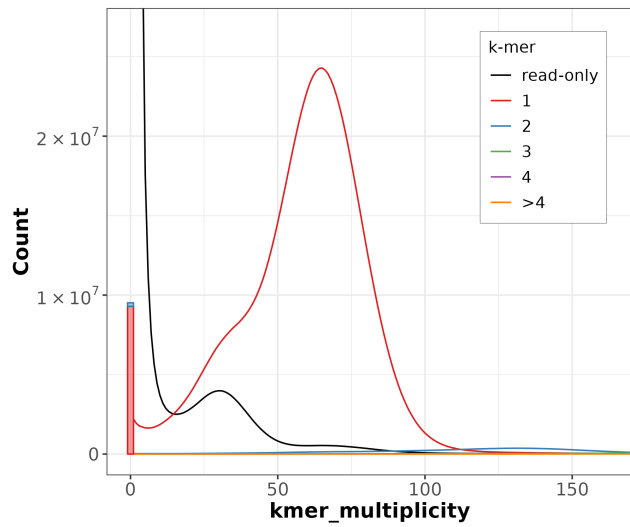
BUSCO: 6.0.0 (euk\_genome\_min, miniprot) / Lineage: mollusca\_odb12 (genomes:36, BUSCOs:4421)

# HiC contact map of curated assembly

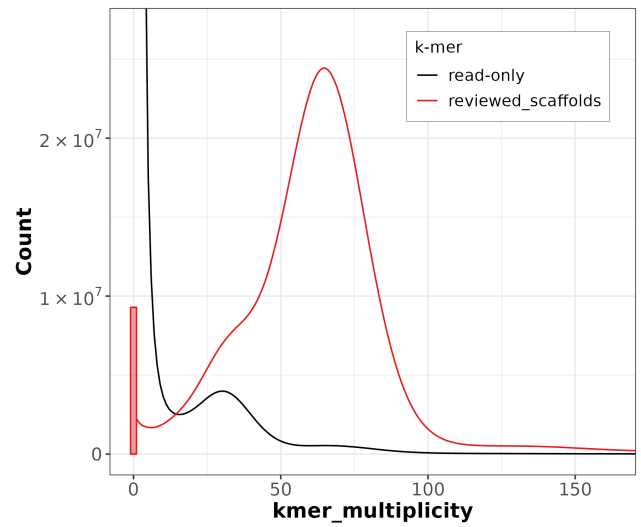


collapsed [\[LINK\]](#)

# K-mer spectra of curated assembly

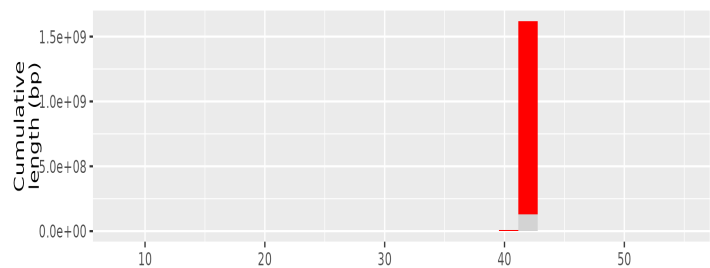


Distribution of k-mer counts per copy numbers found in asm

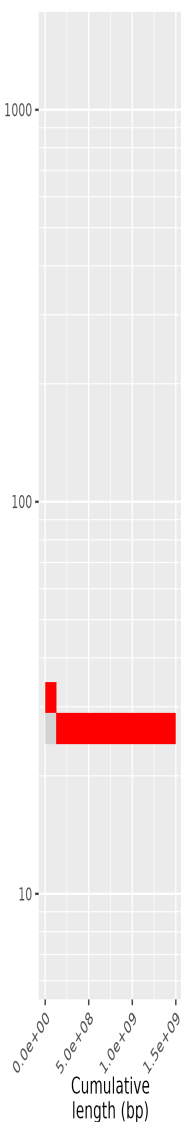
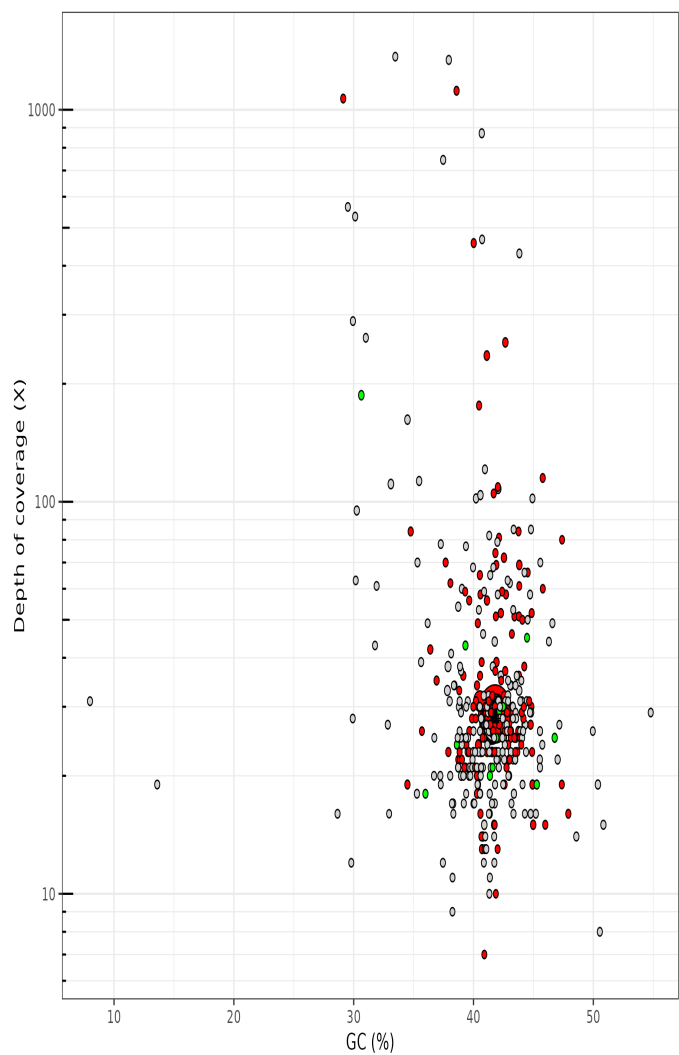


Distribution of k-mer counts coloured by their presence in reads/assemblies

# Post-curation contamination screening



TAPAs summary Graph



superkingdom

- Bacteria
- Eukaryota
- N/A

Length (bp)

- 3.0e+07
- 6.0e+07
- 9.0e+07

Longest sequences (bp)

- xgOvuOvum1\_1 - 119289278 (Eukaryota)
- ▲ xgOvuOvum1\_2 - 90465760 (Eukaryota)
- xgOvuOvum1\_3 - 86160184 (Eukaryota)
- + xgOvuOvum1\_5 - 68580655 (N/A)
- ⊠ xgOvuOvum1\_4 - 68493758 (Eukaryota)

**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	32	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: Emilie Teodori

Affiliation: Genoscope

Date and time: 2025-11-10 15:46:37 CET