

# ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	184865
ToLID	<b>xbTalPusil</b>
Species	Talochlamys pusio
Class	Bivalvia
Order	Pectinida

Genome Traits	Expected	Observed
Haploid size (bp)	1,033,791,279	1,126,867,543
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.Q48

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: 113  
. Contamination notes: ""  
. Other observations: "The assembly of *Talochlamys pusio* (xbTalPusil) is based on 76X PacBio data and 153X 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, 6 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.463 Mb (with the largest being 0.219 Mb). Additionally, 205 regions totaling 97.776 Mb (with the largest being 5.107 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. During manual curation, 29 haplotypic regions and 20 contaminant sequences were removed, totaling 69.180 Mb and 0.937 Mb respectively (with the largest being 12.024Mb and 0.138 Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

# Quality metrics table

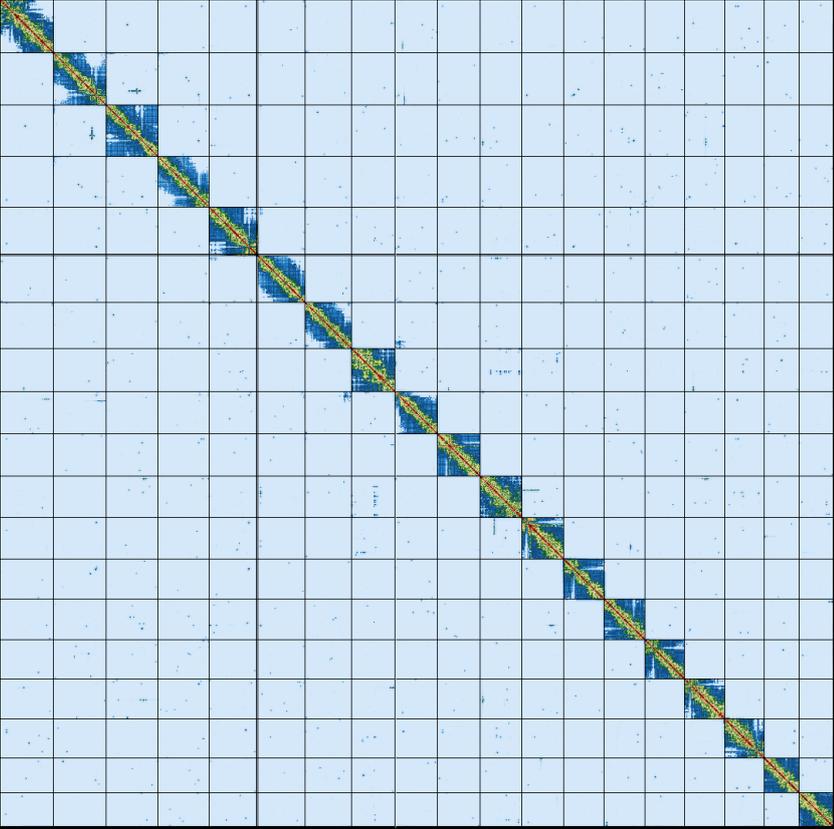
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,196,954,404	1,126,867,543
GC %	36.77	36.76
Gaps/Gbp	55.98	64.78
Total gap bp	6,700	11,200
Scaffolds	84	51
Scaffold N50	63,830,560	57,117,304
Scaffold L50	9	9
Scaffold L90	18	17
Contigs	151	124
Contig N50	28,879,202	31,685,418
Contig L50	15	13
Contig L90	53	46
QV	47.5108	48.7702
Kmer compl.	66.7397	64.2906
BUSCO sing.	94.1%	98.3%
BUSCO dupl.	4.2%	1.1%
BUSCO frag.	1.0%	0.1%
BUSCO miss.	0.7%	0.5%

Warning! BUSCO versions or lineage datasets are not the same across results:

BUSCO: 5.8.2 (euk\_genome\_met, metaeuk) / Lineage: mollusca\_odb12 (genomes:36, BUSCOs:4421)

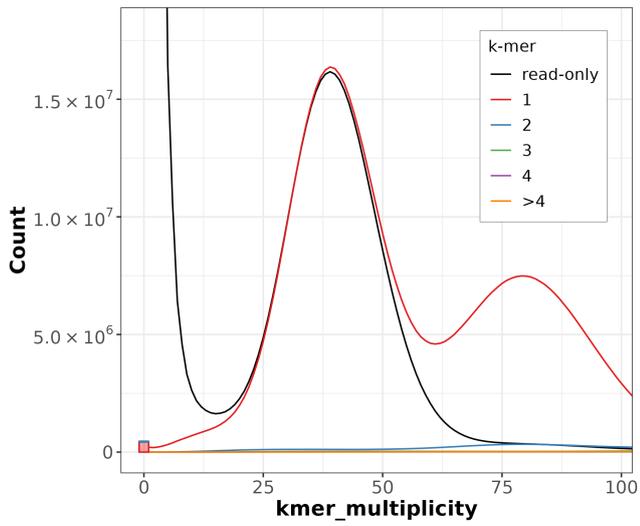
BUSCO: 6.0.0 (euk\_genome\_min, miniprot) / Lineage: mollusca\_odb12 (genomes:36, BUSCOs:4421)

# HiC contact map of curated assembly

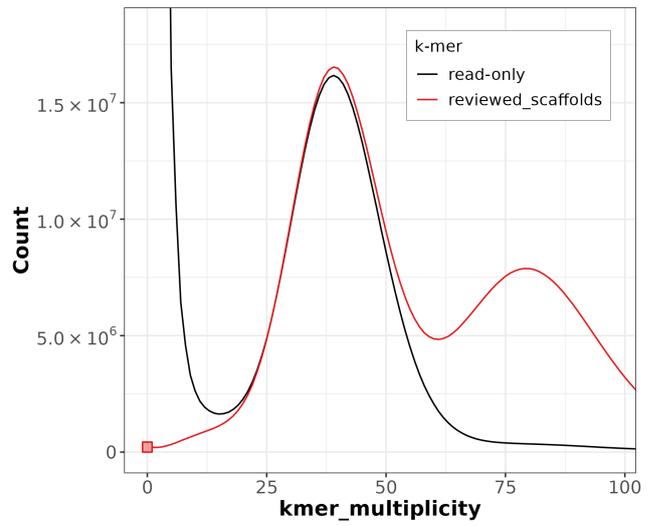


collapsed [\[LINK\]](#)

# K-mer spectra of curated assembly

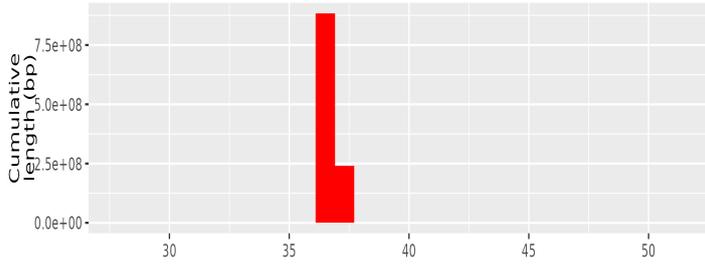


Distribution of k-mer counts per copy numbers found in asm

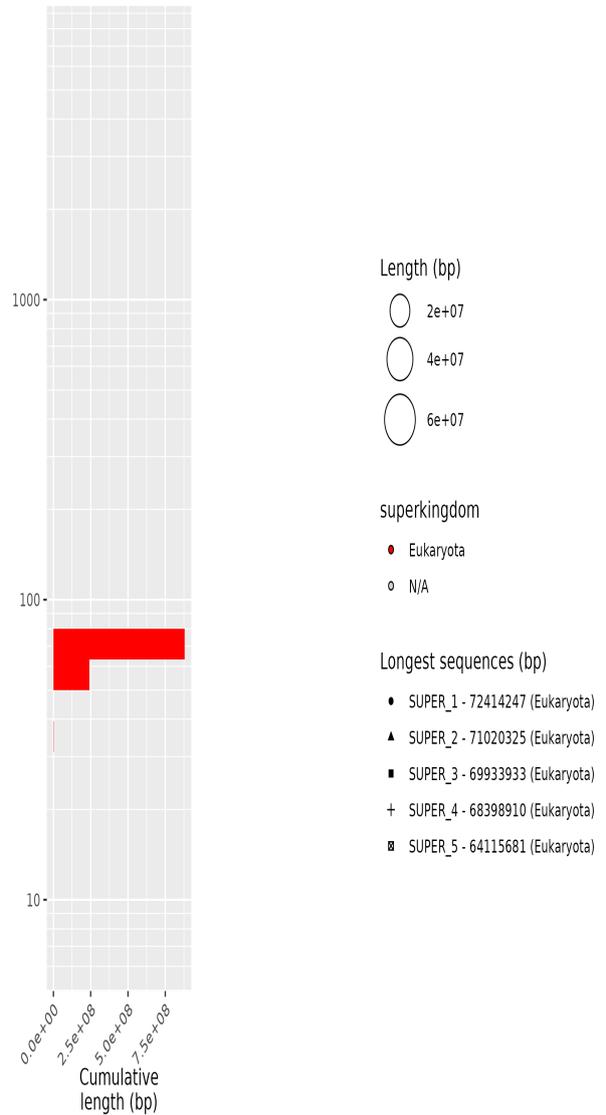
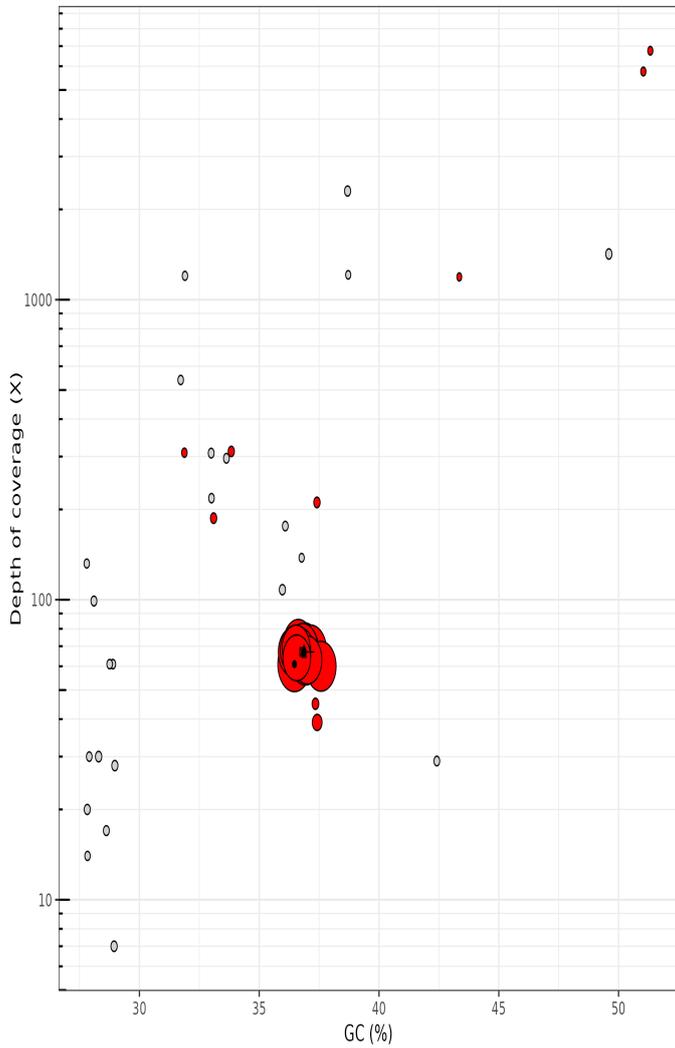


Distribution of k-mer counts coloured by their presence in reads/assemblies

# Post-curation contamination screening



## TAPAs summary Graph



**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	76	153

# 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: 2025-12-03 04:22:57 CET