ERGA Assembly Report
v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	989235
ToLID	**xbModBarb1**
Species	Modiolus barbatus
Class	Bivalvia
Order	Mytilida

Genome Traits	Expected	Observed
Haploid size (bp)	2,098,531,296	2,192,276,050
Haploid Number	14 (source: ancestor)	16
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q50

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
. Assembly length loss > 3% for collapsed

Curator notes

. Interventions/Gb: 53
. Contamination notes: ""
. Other observations: "The assembly of Modiolus barbatus (xbModBarb1) is based on 39X PacBio data and 156X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio+Hi-C assembly generation with Hifiasm, removal of contaminant sequences using Context, and Hi-C-based scaffolding with YaHS of each haplotype. In total, 15 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.158 Mb (with the largest being 0.485 Mb). Mitochondrial genome was assembled using OATK. Both haplotypes were 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	2,375,172,333	2,192,276,050
GC %	34.52	34.62
Gaps/Gbp	407.13	296.04
Total gap bp	104,100	73,000
Scaffolds	302	875
Scaffold N50	124,750,922	136,183,202
Scaffold L50	8	7
Scaffold L90	16	15
Contigs	1,234	1,524
Contig N50	6,379,577	5,192,939
Contig L50	79	130
Contig L90	362	421
QV	50.9754	50.4626
Kmer compl.	69.0204	64.2297
BUSCO sing.	91.7%	98.5%
BUSCO dupl.	8.0%	1.1%
BUSCO frag.	0.1%	0.2%
BUSCO miss.	0.2%	0.2%

BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)
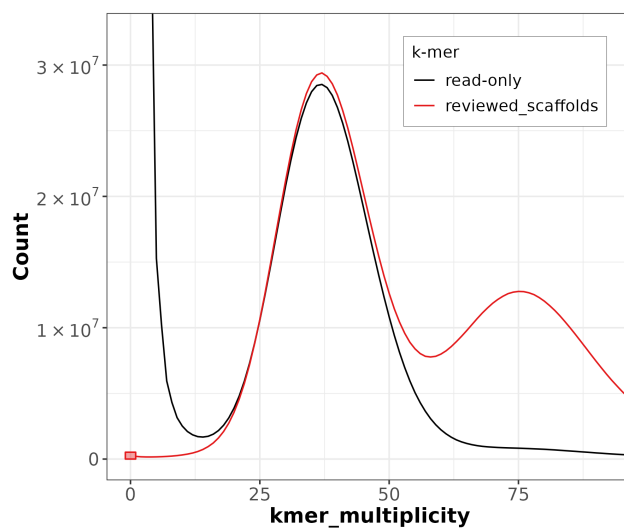
HiC contact map of curated assembly
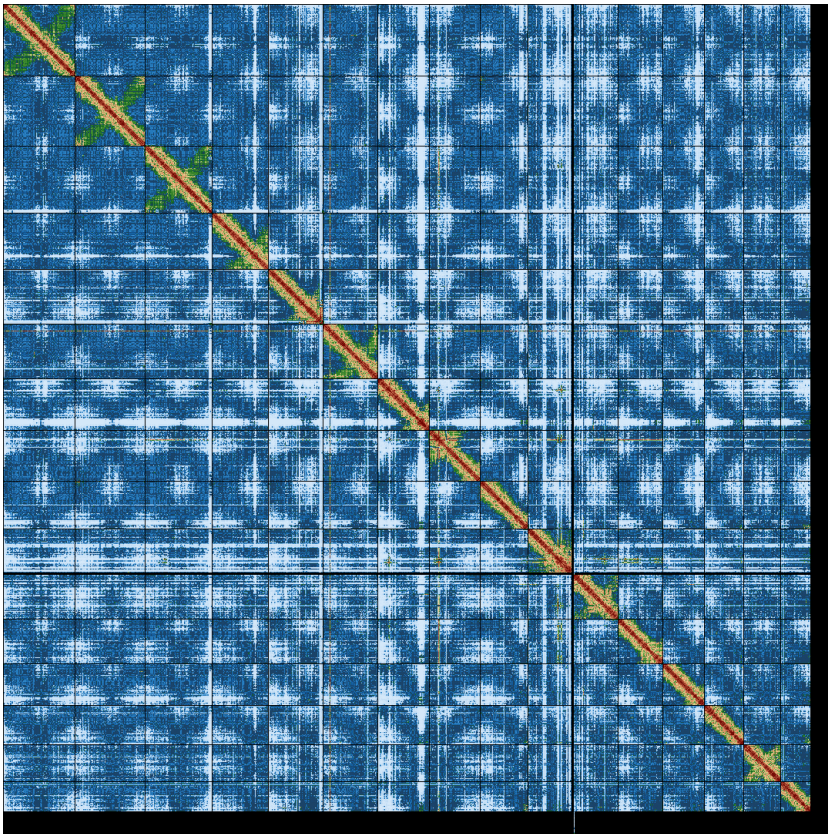
collapsed [LINK]
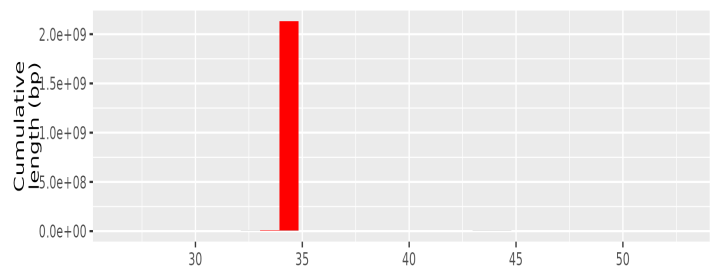
K-mer spectra of curated assembly
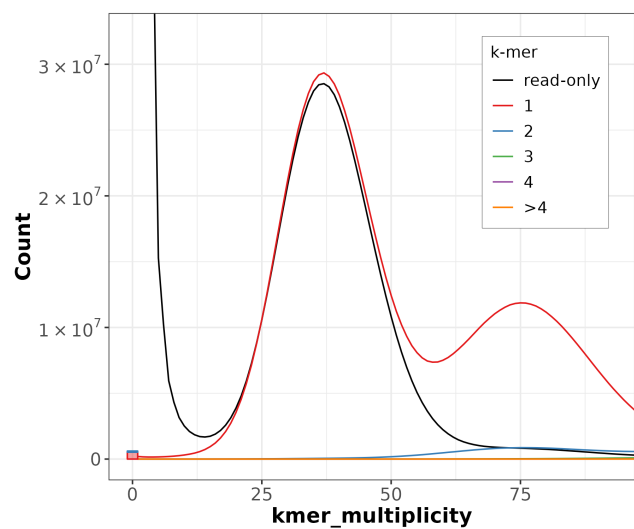
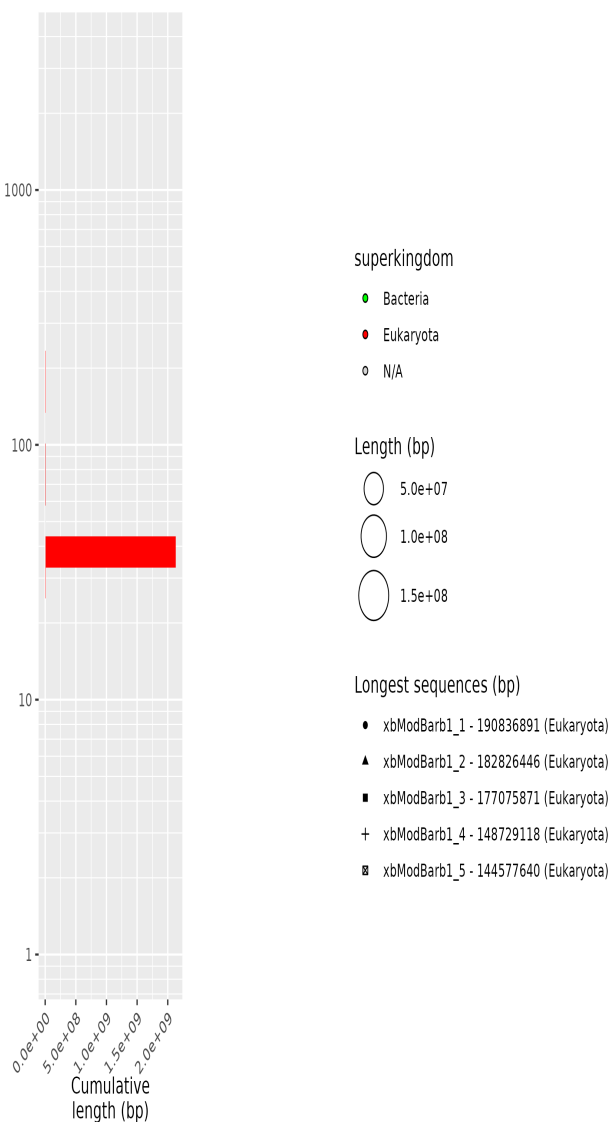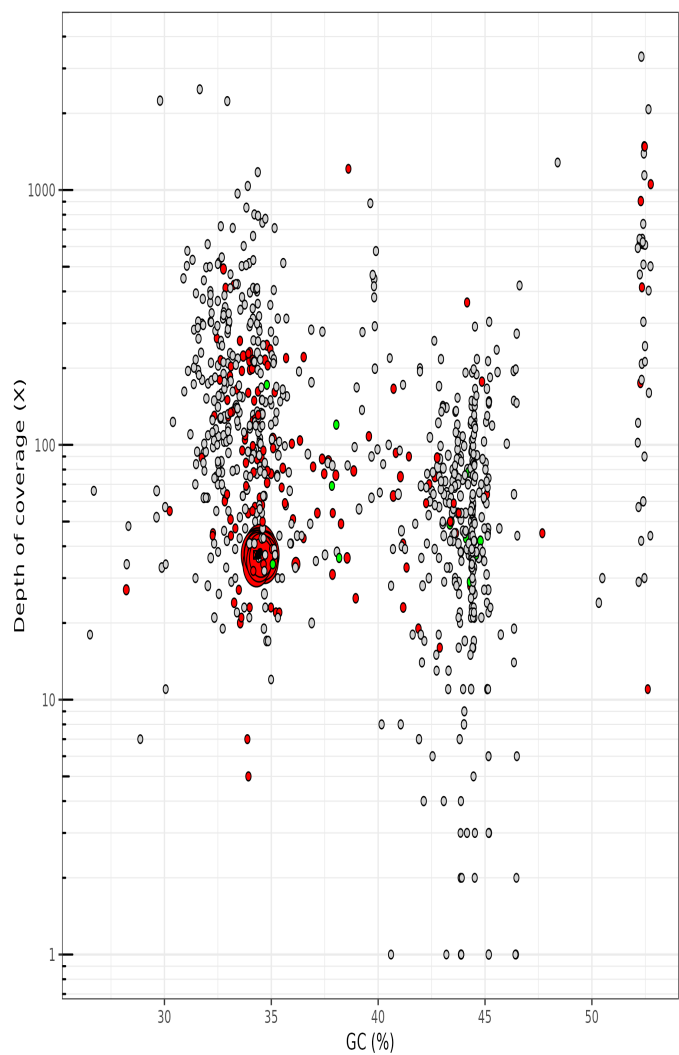Distribution of k-mer counts per copy
numbers found in asm

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

Post-curation contamination screening

TAPAs summary Graph

(9 0X contigs have been hidden)

superkingdom
- Bacteria
- Eukaryota
- N/A

Length (bp)
- 5.0e+07
- 1.0e+08
- 1.5e+08

Longest sequences (bp)
- xbModBarb1_1 - 190836891 (Eukaryota)
- xbModBarb1_2 - 182826446 (Eukaryota)
- xbModBarb1_3 - 177075871 (Eukaryota)
- xbModBarb1_4 - 148729118 (Eukaryota)
- xbModBarb1_5 - 144577640 (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	38	156

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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2025-12-07 22:22:33 CET