

ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	926810
ToLID	xbAtrSem1
Species	<i>Atrina seminuda</i>
Class	Bivalvia
Order	Pterioida

Genome Traits	Expected	Observed
Haploid size (bp)	792,011,696	820,524,418
Haploid Number	14 (source: ancestor)	17
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:

- . Observed Haploid Number is different from Expected
- . Kmer completeness value is less than 90 for collapsed
- . Assembly length loss > 3% for collapsed

Curator notes

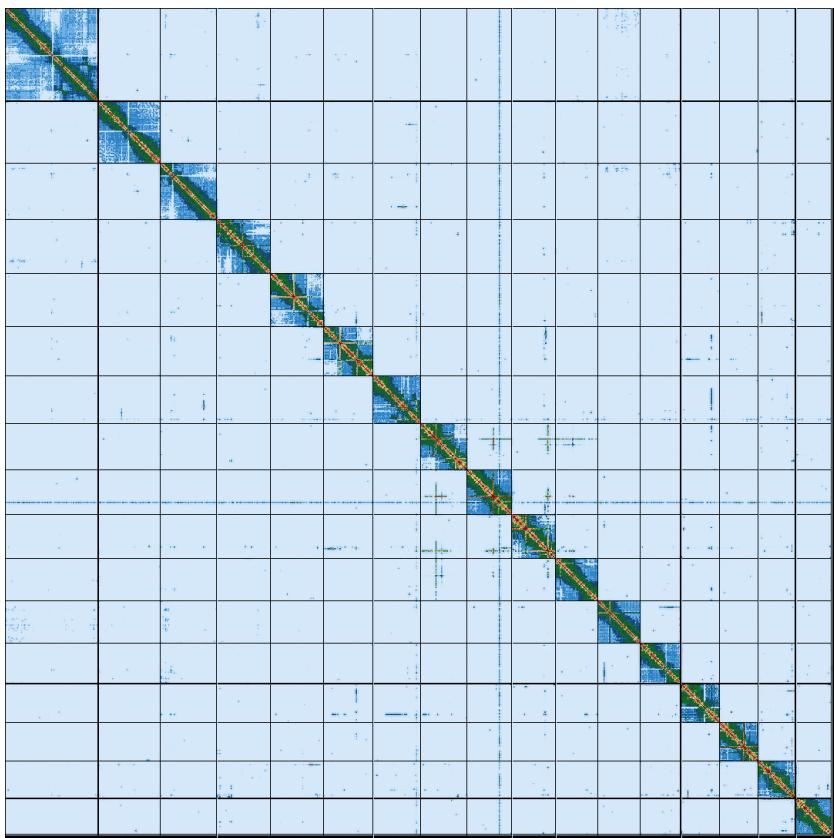
. Interventions/Gb: 101
. Contamination notes: ""
. Other observations: "The assembly of *Atrina seminuda* (xbAtrSem1) is based on 90X PacBio data and 211X Arima Hi-C data generated as part of the ATLASea programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial 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, 11 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 3.616 Mb (with the largest being 3.115 Mb). Additionally, 139 regions totaling 57.041 Mb (with the largest being 4.864 Mb) were identified as haplotypic duplications and removed. During manual curation, 17 haplotypic regions were removed, totaling 41Mb (with the largest being 1.4Mb). 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	861,733,375	820,524,418
GC %	35.93	35.93
Gaps/Gbp	94	137.72
Total gap bp	8,100	15,400
Scaffolds	61	46
Scaffold N50	44,629,212	46,145,226
Scaffold L50	8	8
Scaffold L90	17	15
Contigs	142	159
Contig N50	13,270,883	13,135,000
Contig L50	20	19
Contig L90	66	63
QV	48.4393	48.4318
Kmer compl.	67.2342	65.6785
BUSCO sing.	93.6%	98.4%
BUSCO dupl.	5.4%	0.5%
BUSCO frag.	0.3%	0.3%
BUSCO miss.	0.7%	0.7%

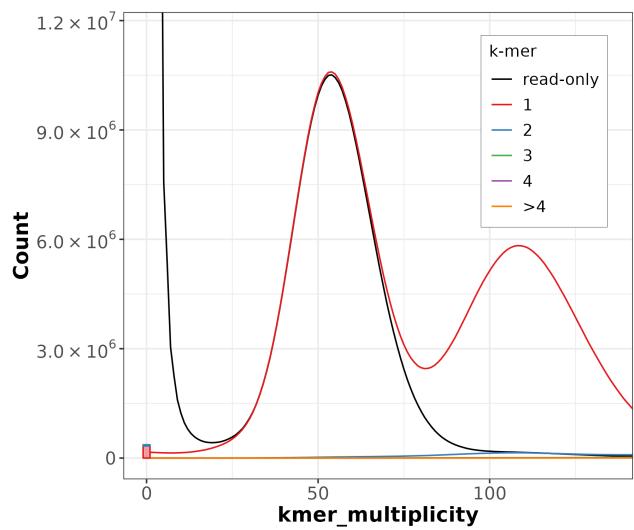
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

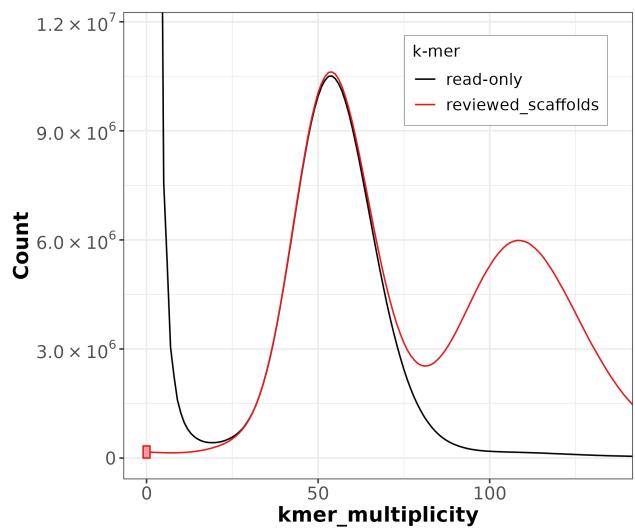


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

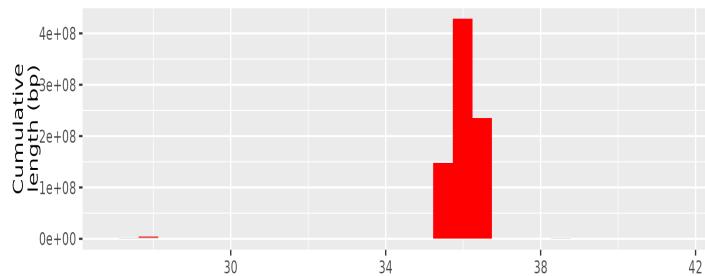


Distribution of k-mer counts per copy numbers found in asm

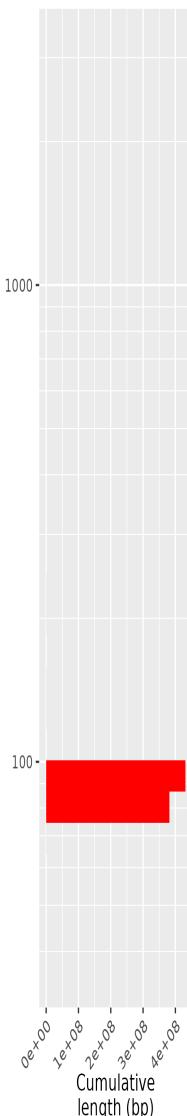
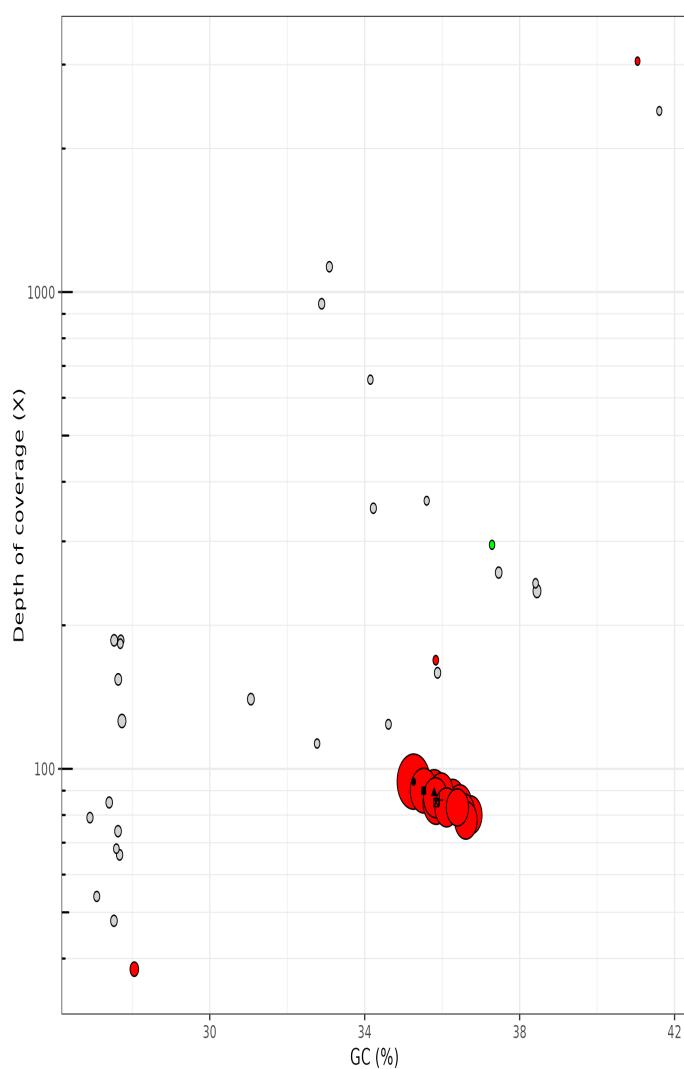


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



superkingdom

- Bacteria
- Eukaryota
- N/A

Length (bp)

- 2.5e+07
- 5.0e+07
- 7.5e+07

Longest sequences (bp)

- xbAtrSemi1_1 - 91791366 (Eukaryota)
- ▲ xbAtrSemi1_2 - 61069450 (Eukaryota)
- xbAtrSemi1_3 - 55713748 (Eukaryota)
- + xbAtrSemi1_4 - 53134211 (Eukaryota)
- ☒ xbAtrSemi1_5 - 52594949 (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	90	211

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: Benjamin Istace
Affiliation: Genoscope

Date and time: 2025-12-07 21:49:42 CET