ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	36099
ToLID	**xgHalOvin1**
Species	Haliotis ovina
Class	Gastropoda
Order	Lepetellida

Genome Traits	Expected	Observed
Haploid size (bp)	1,200,844,967	1,295,783,543
Haploid Number	9 (source: ancestor)	16
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.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

Curator notes

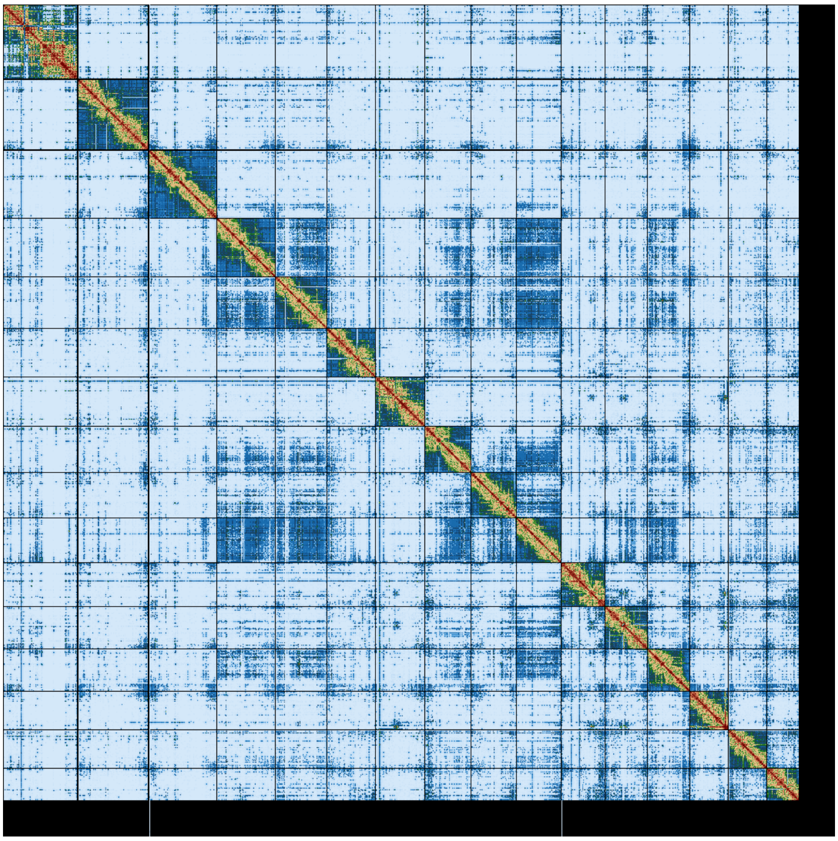
. Interventions/Gb: 53
. Contamination notes: ""
. Other observations: "The assembly of Haliotis ovina (xgHalOvin1) is based on 40X PacBio data and 165X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included the following steps: initial PacBio 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, 219 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 38 Mb (with the largest being 2 Mb). Additionally, 220 regions totaling 61 Mb (with the largest being 7.2 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using oatk. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 10 supplementary haplotypic regions were removed, totaling 8.3 Mb (with the largest being 1.5 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	1,304,206,168	1,295,783,543
GC %	40.24	40.24
Gaps/Gbp	72.84	88.75
Total gap bp	9,500	15,500
Scaffolds	362	337
Scaffold N50	70,296,476	76,376,157
Scaffold L50	8	7
Scaffold L90	16	15
Contigs	457	452
Contig N50	21,426,545	19,998,011
Contig L50	19	20
Contig L90	69	69
QV	50.8004	50.8041
Kmer compl.	80.6577	80.4945
BUSCO sing.	98.9%	99.4%
BUSCO dupl.	1.0%	0.4%
BUSCO frag.	0.0%	0.1%
BUSCO miss.	0.1%	0.1%

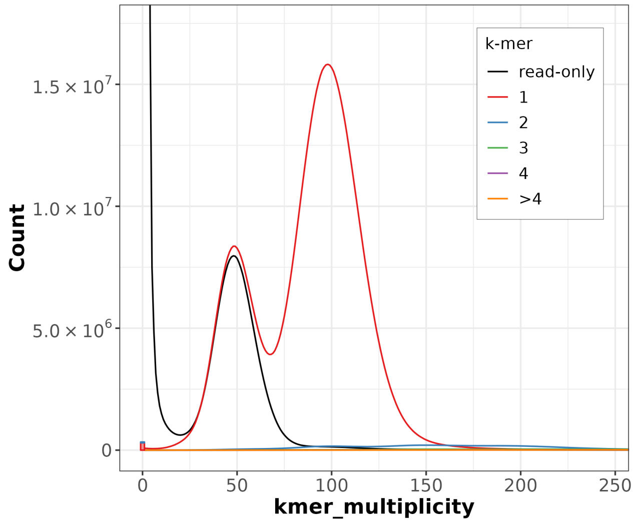
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

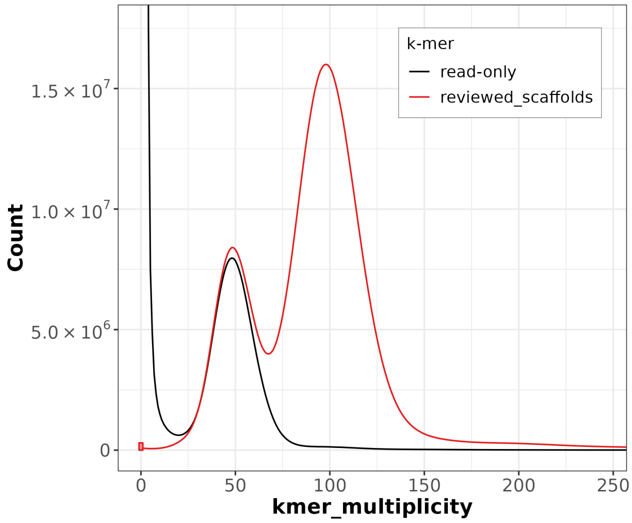


collapsed [LINK]

K-mer spectra of curated assembly

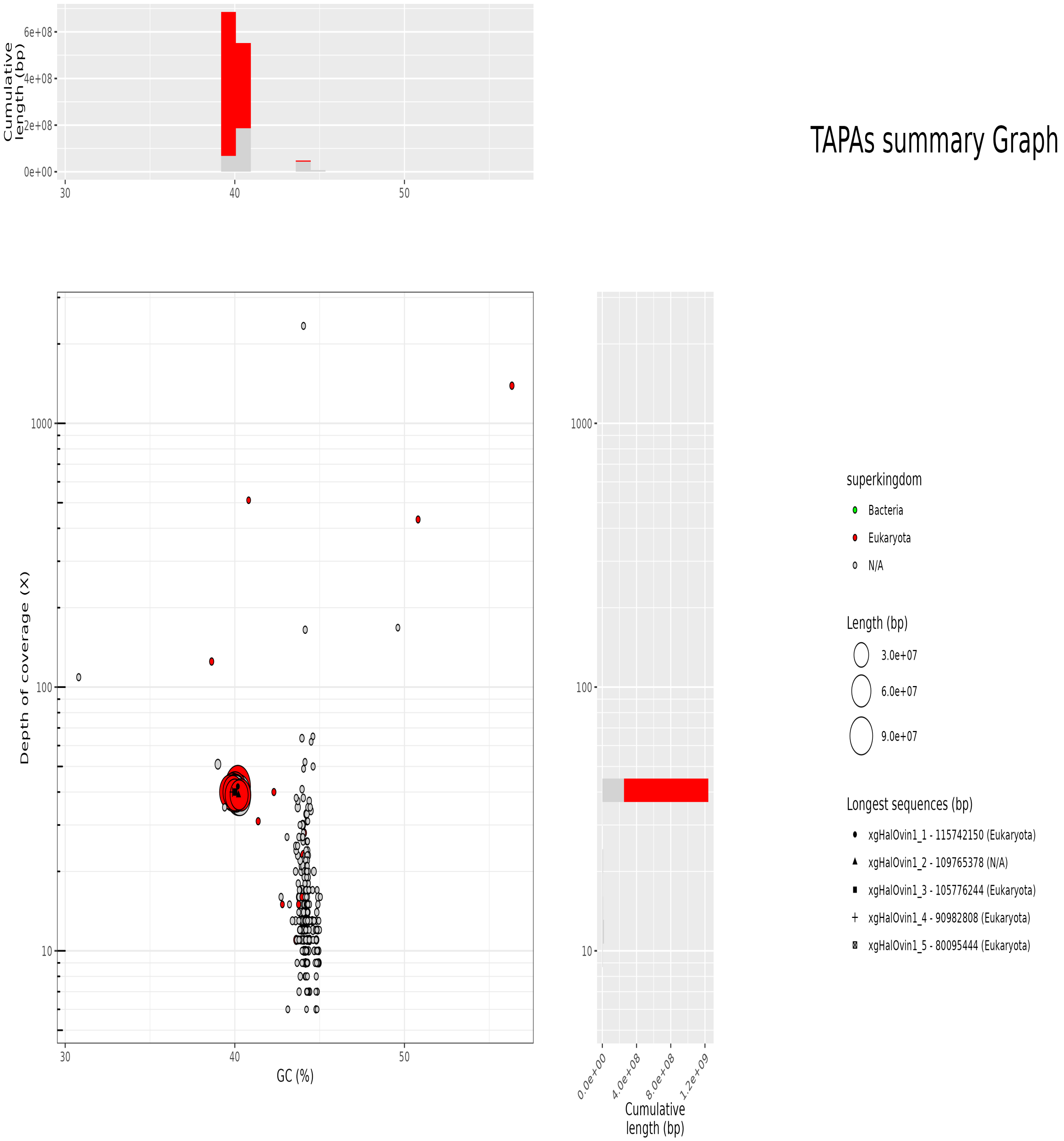


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	40	165

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: Caroline Menguy
Affiliation: Genoscope

Date and time: 2025-11-10 21:56:32 CET