ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	100776
ToLID	**xbAeqOper1**
Species	Aequipecten opercularis
Class	Bivalvia
Order	Pectinida

Genome Traits	Expected	Observed
Haploid size (bp)	932,996,933	995,840,645
Haploid Number	19 (source: ancestor)	13
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

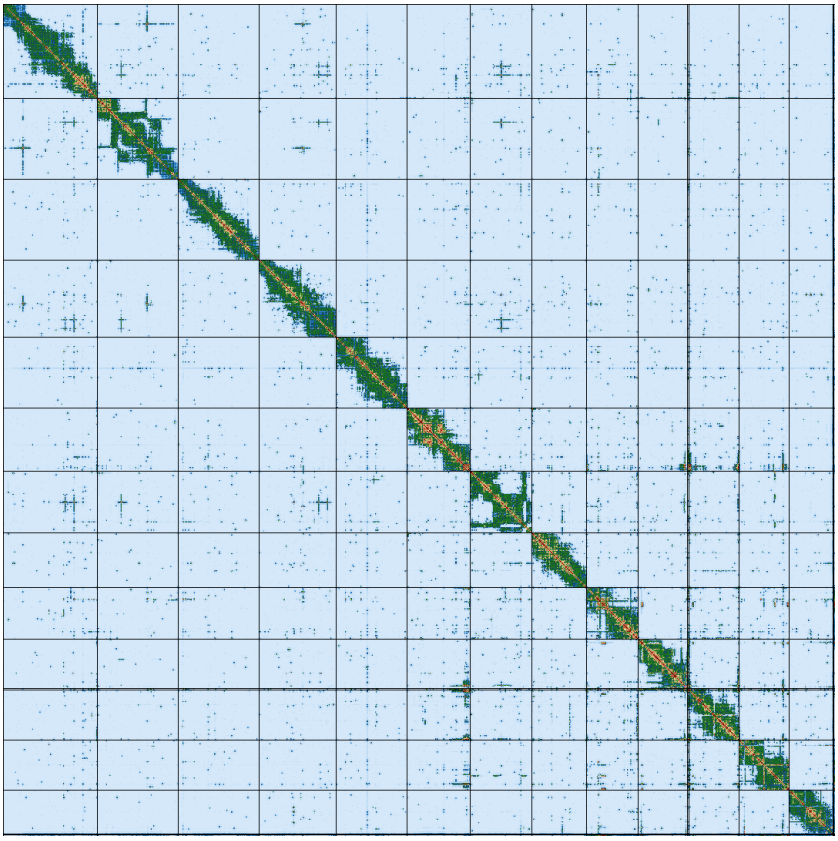
. Interventions/Gb: 0
. Contamination notes: ""
. Other observations: "The assembly of Aequipecten opercularis (xbAeqOper1) is based on 34.75X ONT data and 147.51X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial ONT assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. No contigs were found to be contaminants (bacterial, archaeal, or viral). Additionally, 150 regions totaling 65.248 Mb (with the largest being 5.104 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. The telomeric pattern AACCCT was identified with TelFinder and used to generate the Pretext Telomeres tracks. 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	1,059,557,788	995,840,645
GC %	35.74	35.69
Gaps/Gbp	32.09	42.18
Total gap bp	3,400	6,700
Scaffolds	79	17
Scaffold N50	75,625,093	75,525,905
Scaffold L50	6	6
Scaffold L90	13	12
Contigs	113	59
Contig N50	32,110,598	34,949,495
Contig L50	11	10
Contig L90	36	33
QV	45.0483	48.5489
Kmer compl.	75.9929	74.4505
BUSCO sing.	93.1%	99.1%
BUSCO dupl.	6.6%	0.5%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	0.2%	0.3%

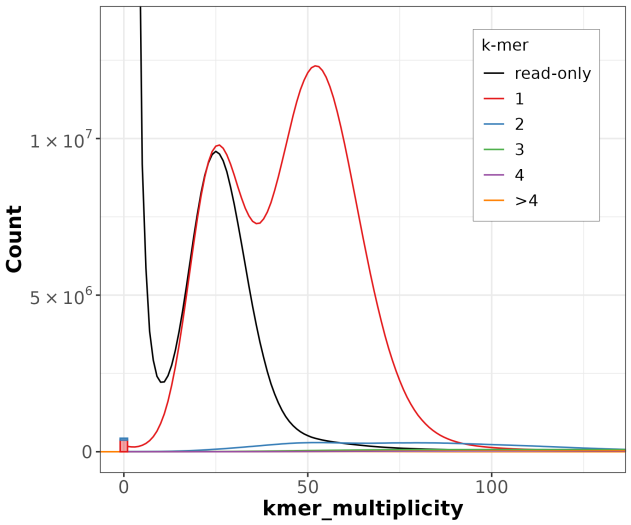
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12.2 (genomes:31, BUSCOs:4236)

HiC contact map of curated assembly

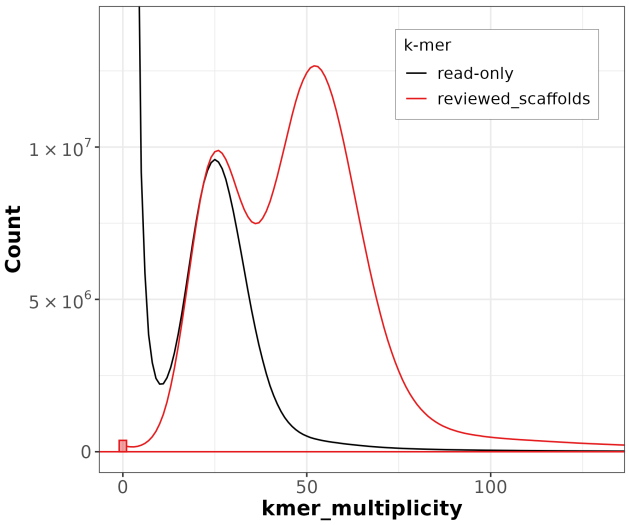


collapsed [LINK]

K-mer spectra of curated assembly

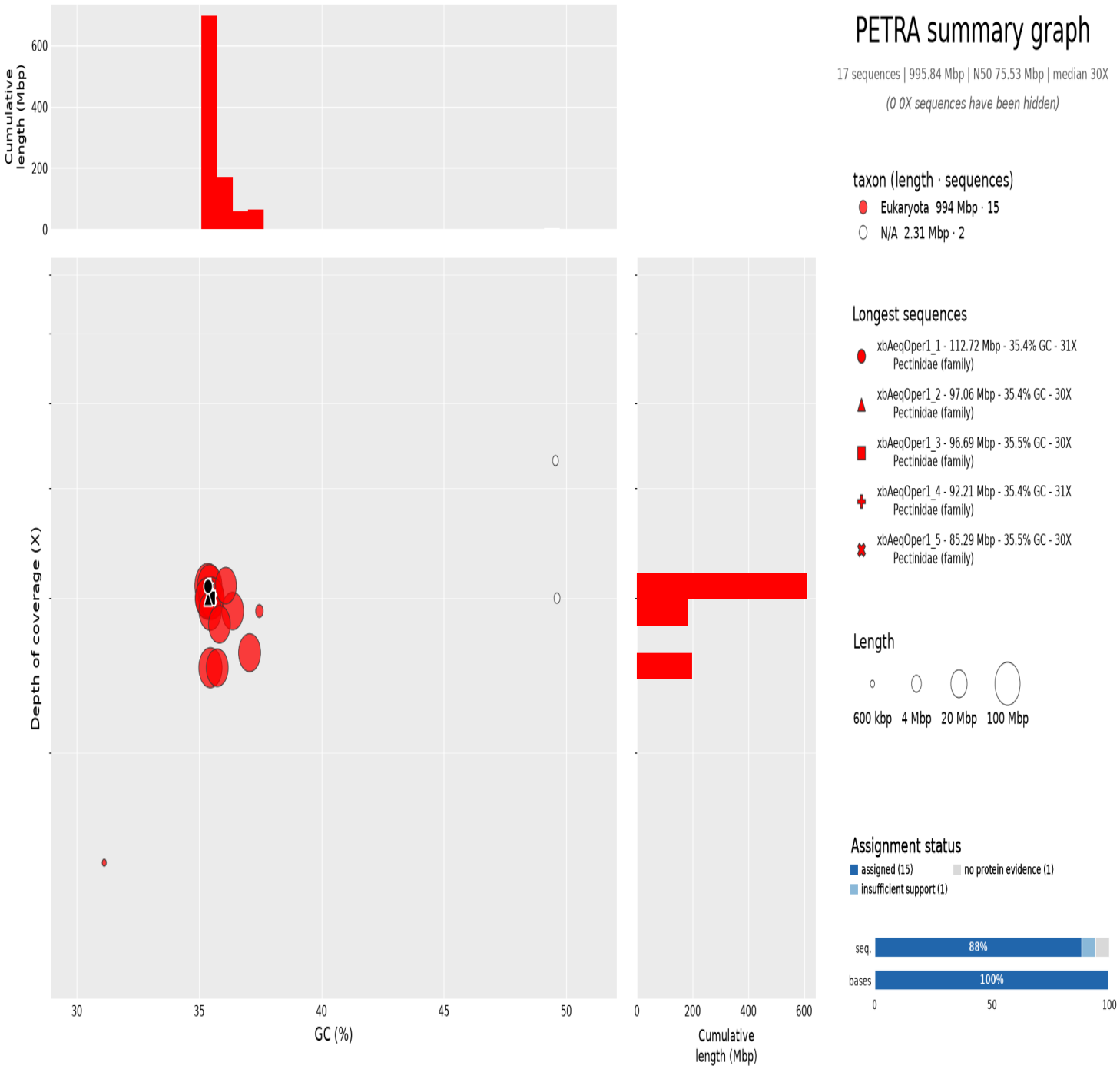


Distribution of k-mer counts per copy numbers found in asm



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	36	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:* 0.2.4
  |_ *key param:* NA
- **PretextView**
  |_ *ver:* 1.0.7
  |_ *key param:* NA

Submitter: Eleanore Lacoste
Affiliation: Genoscope

Date and time: 2026-11-07 20:17:06 CET