ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	102291
ToLID	**xbAnoEphi3**
Species	Anomia ephippium
Class	Bivalvia
Order	Pectinida

Genome Traits	Expected	Observed
Haploid size (bp)	597,719,894	606,706,885
Haploid Number	19 (source: ancestor)	7
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

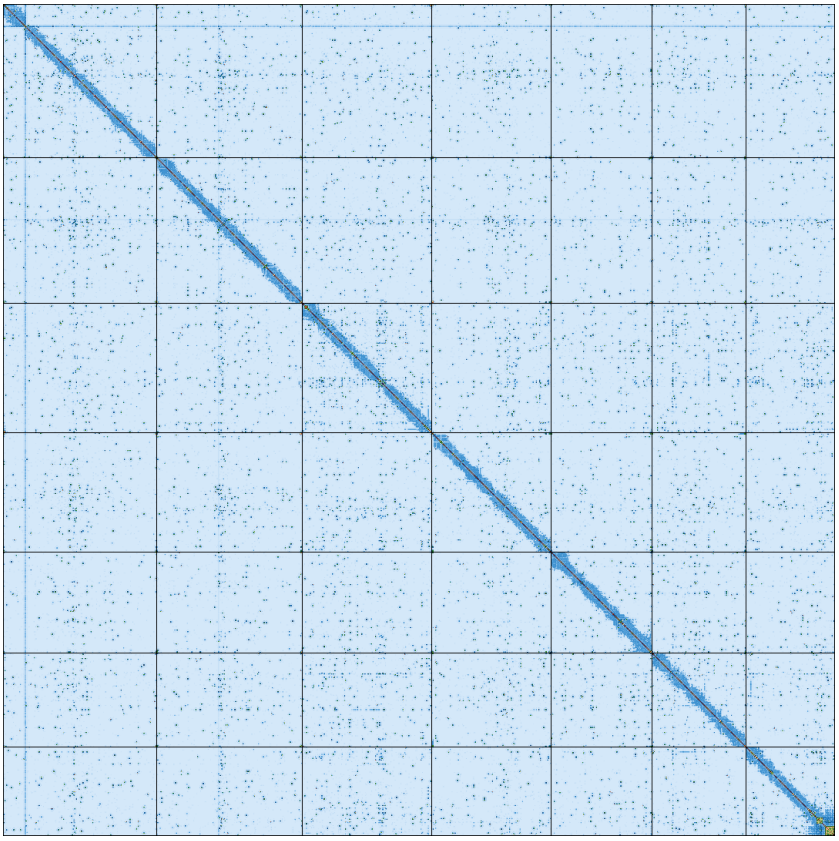
. Interventions/Gb: 26
. Contamination notes: ""
. Other observations: "The assembly of \'Anomia ephippium\' (xbAnoEphi3) is based on 87X PacBio data and 391X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 38 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.739 Mb (with the largest being 0.211 Mb). Additionally, 68 regions totaling 16.358 Mb (with the largest being 4.318 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using Oatk and shows an unusual size compared to currently available reference. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 7 haplotypic regions were removed, totaling 12.967Mb (with the largest being 6.193Mb). 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	619,620,686	606,706,885
GC %	34.2	34.2
Gaps/Gbp	25.82	44.5
Total gap bp	1,600	4,000
Scaffolds	30	13
Scaffold N50	94,060,247	94,258,050
Scaffold L50	3	3
Scaffold L90	7	7
Contigs	46	40
Contig N50	45,744,000	45,744,000
Contig L50	6	6
Contig L90	14	13
QV	50.0759	50.0921
Kmer compl.	75.3942	75.0188
BUSCO sing.	95.0%	96.1%
BUSCO dupl.	1.5%	0.5%
BUSCO frag.	0.9%	0.9%
BUSCO miss.	2.6%	2.6%

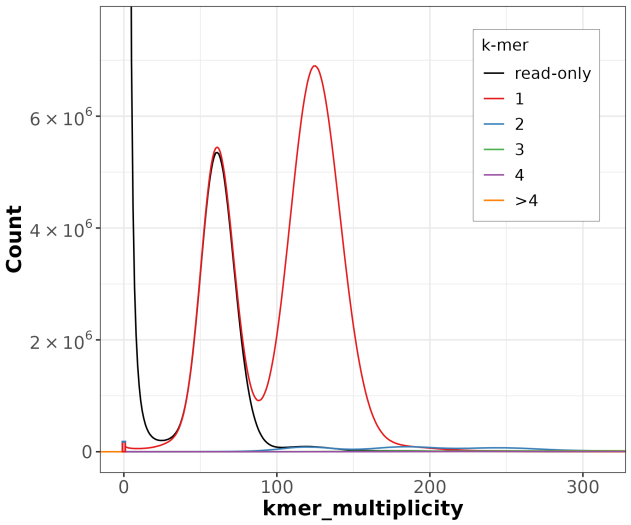
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

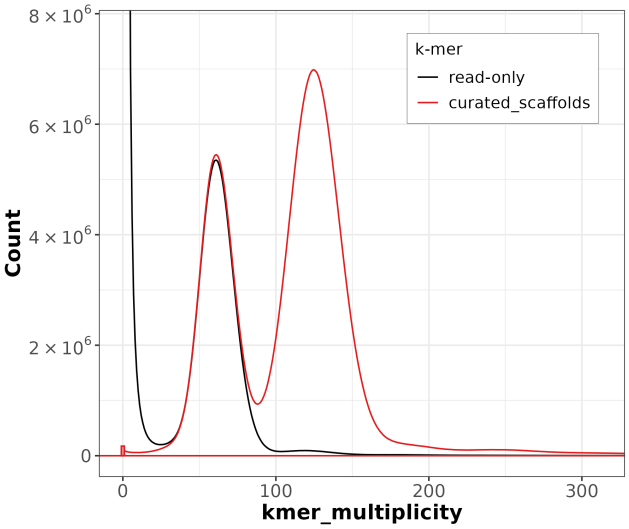


collapsed [LINK]

K-mer spectra of curated assembly

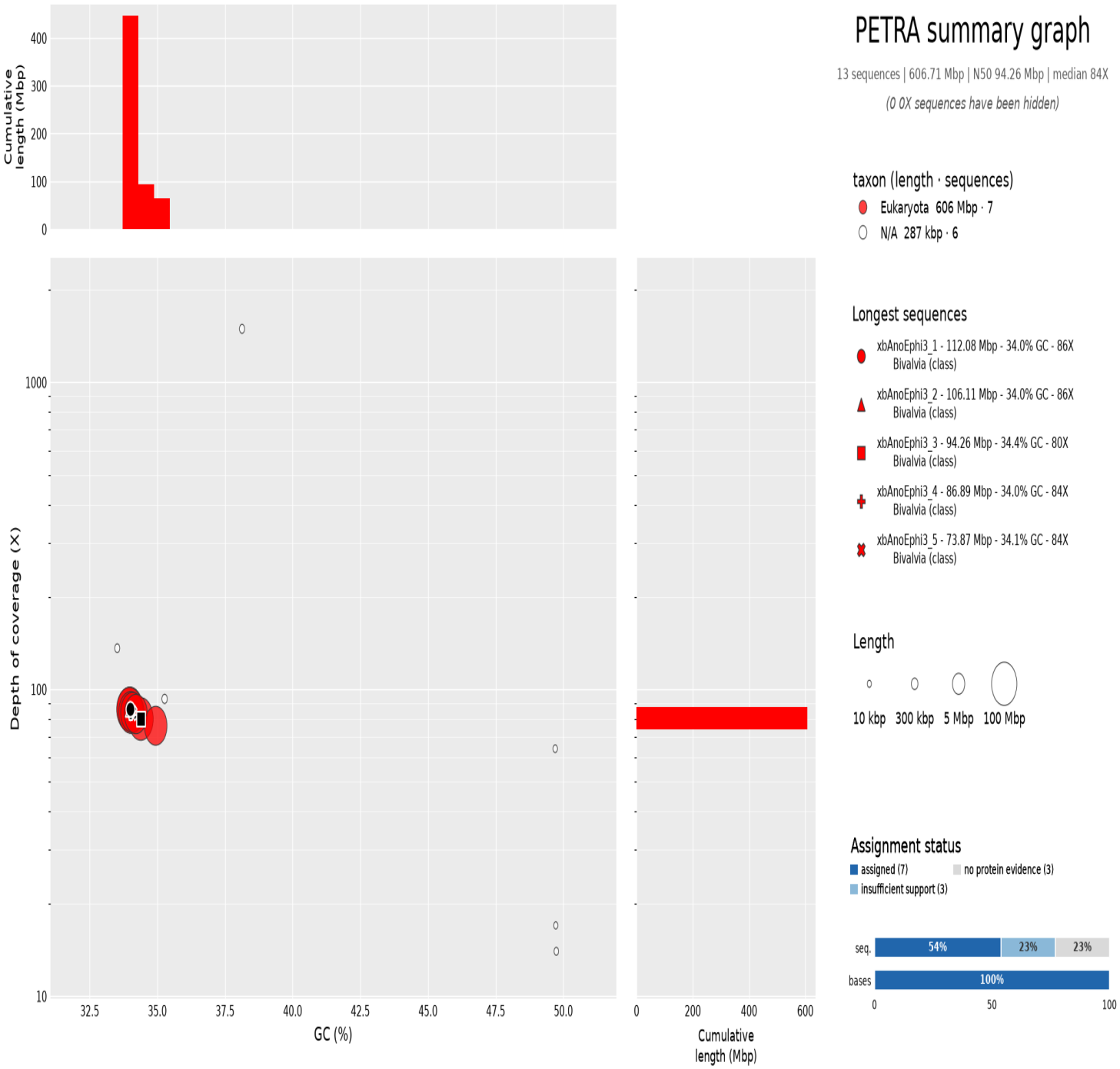


Distribution of k-mer counts per copy numbers found in asm



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	89	399

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: Simone Duprat
Affiliation: Genoscope

Date and time: 2026-10-23 11:39:35 CEST