ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	144766
ToLID	**xgAplDact1**
Species	Aplysia dactylomela
Class	Gastropoda
Order	Aplysiida

Genome Traits	Expected	Observed
Haploid size (bp)	1,191,658,192	1,316,545,377
Haploid Number	12 (source: ancestor)	17
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 5.7.Q32

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

. Observed Haploid Number is different from Expected

. QV value is less than 40 for collapsed
. Kmer completeness value is less than 90 for collapsed
. More than 1000 gaps/Gbp for collapsed

Curator notes

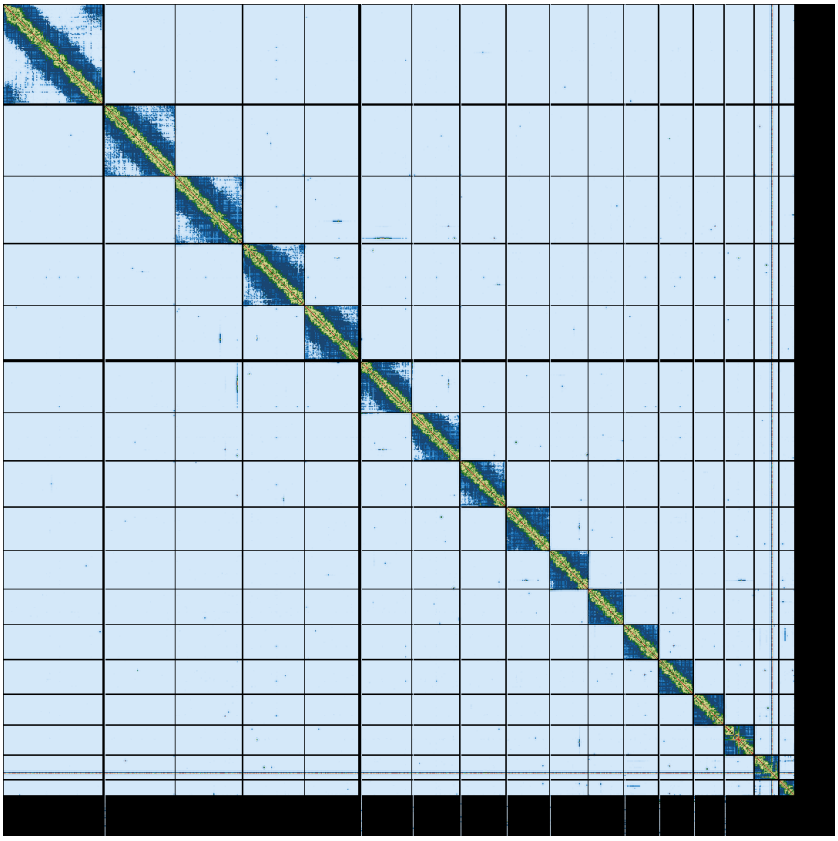
. Interventions/Gb: 228
. Contamination notes: ""
. Other observations: "The assembly of Aplysia dactylomela (xgAplDact1) is based on 25X ONT data and 142X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial ONT assembly generation with Flye, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 110 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 8.376 Mb (with the largest being 3.41 Mb). Additionally, 3983 regions totaling 89.683 Mb (with the largest being 0.17 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using ptGAUL. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 5 haplotypic regions and 3 contaminant sequences were removed, totaling 0.994Mb and 0.071Mb, respectively (with the largest being 0.459Mb and 0.032Mb). 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	1,317,650,019	1,316,545,377
GC %	41.39	41.39
Gaps/Gbp	3,588.21	3,651.98
Total gap bp	472,800	499,700
Scaffolds	2,055	1,958
Scaffold N50	73,464,299	75,072,562
Scaffold L50	7	7
Scaffold L90	17	16
Contigs	6,783	6,766
Contig N50	522,726	523,390
Contig L50	657	656
Contig L90	3,255	3,249
QV	32.795	32.7943
Kmer compl.	79.9759	79.9423
BUSCO sing.	96.7%	96.9%
BUSCO dupl.	1.3%	1.2%
BUSCO frag.	1.5%	1.4%
BUSCO miss.	0.5%	0.5%

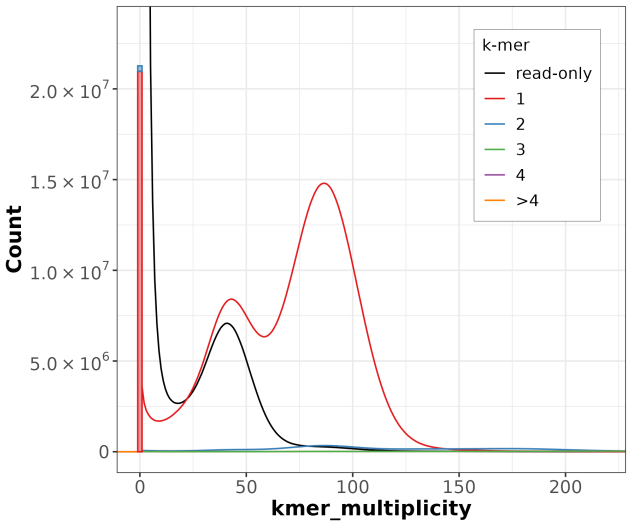
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

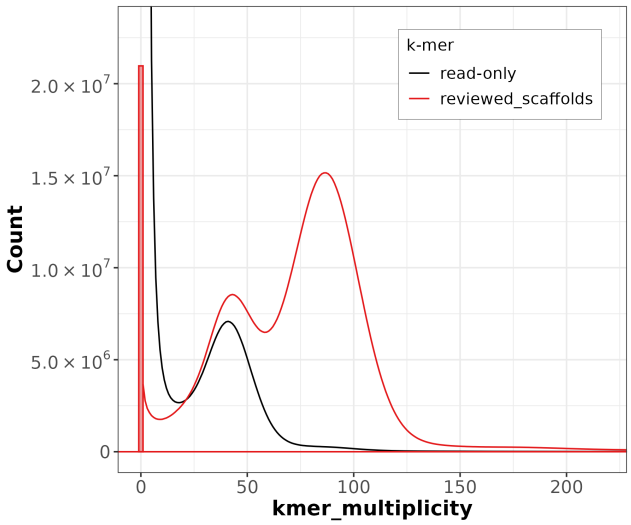


collapsed [LINK]

K-mer spectra of curated assembly

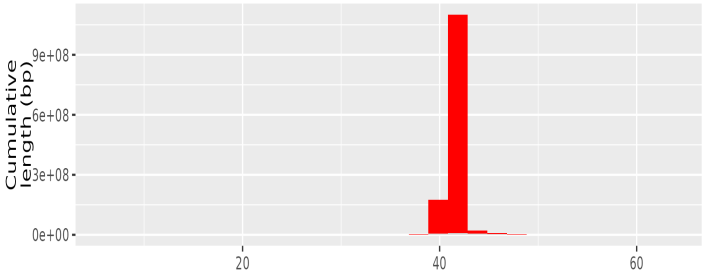


Distribution of k-mer counts per copy numbers found in asm

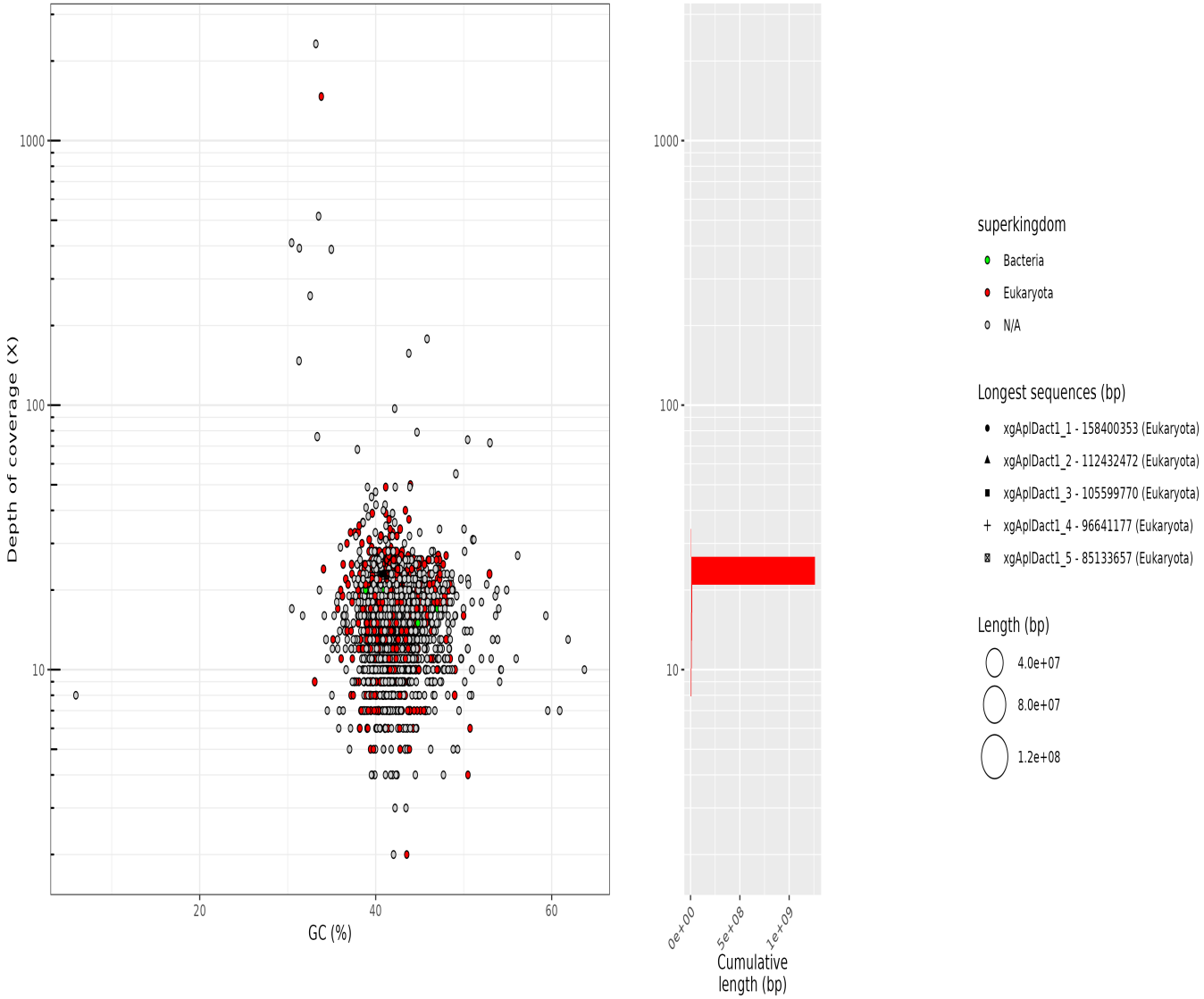


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	25	143

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: Emilie Teodori
Affiliation: Genoscope

Date and time: 2026-05-28 03:18:30 CEST