

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	154653
ToLID	xgPlaArgo1
Species	Platydoris argo
Class	Gastropoda
Order	Nudibranchia

Genome Traits	Expected	Observed
Haploid size (bp)	992,156,513	1,158,465,175
Haploid Number	12 (source: ancestor)	12
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q42

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

- . Kmer completeness value is less than 90 for collapsed

Curator notes

- . Interventions/Gb: 62
- . Contamination notes: ""
- . Other observations: "The assembly of *Platydoris argo* (xgPlaArgo1) is based on 33X PacBio data and 213X 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, 9 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.297 Mb (with the largest being 0.053 Mb). Additionally, 526 regions totaling 43.246 Mb (with the largest being 0.678 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 13 haplotypic regions were removed, totaling 2.991Mb (with the largest being 0.603Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

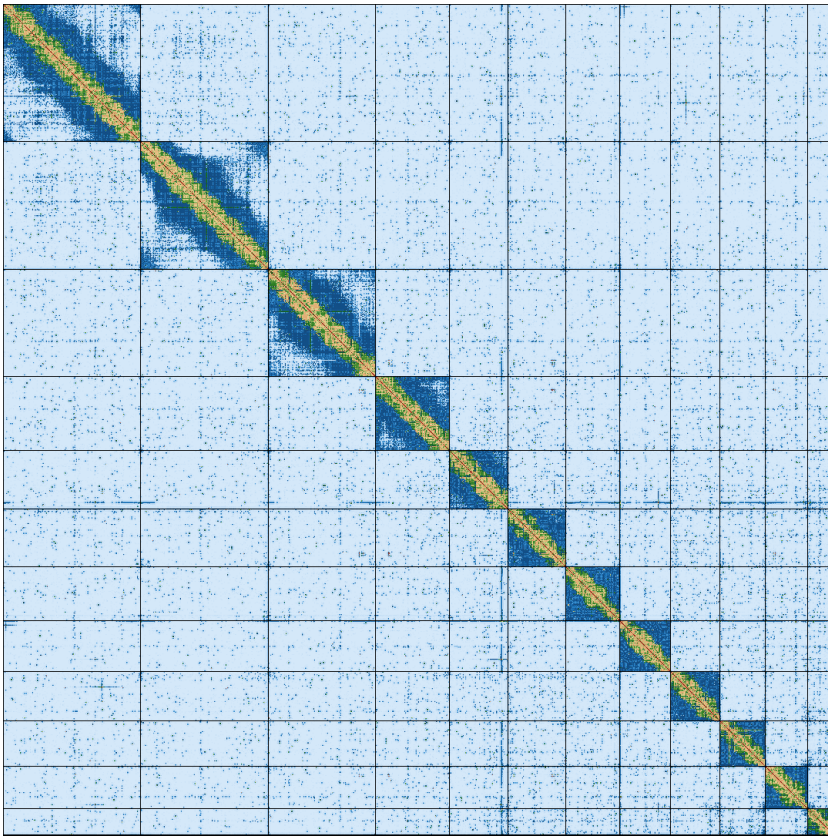
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,161,466,071	1,158,465,175
GC %	38.09	38.09
Gaps/Gbp	317.7	326.29
Total gap bp	36,900	41,000
Scaffolds	85	76
Scaffold N50	102,823,341	102,259,910
Scaffold L50	4	4
Scaffold L90	10	10
Contigs	454	454
Contig N50	4,706,071	4,706,071
Contig L50	80	80
Contig L90	239	238
QV	42.4144	42.4143
Kmer compl.	81.5803	81.5292
BUSCO sing.	97.9%	96.3%
BUSCO dupl.	1.0%	0.8%
BUSCO frag.	0.4%	2.1%
BUSCO miss.	0.7%	0.7%

Warning! BUSCO versions or lineage datasets are not the same across results:

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

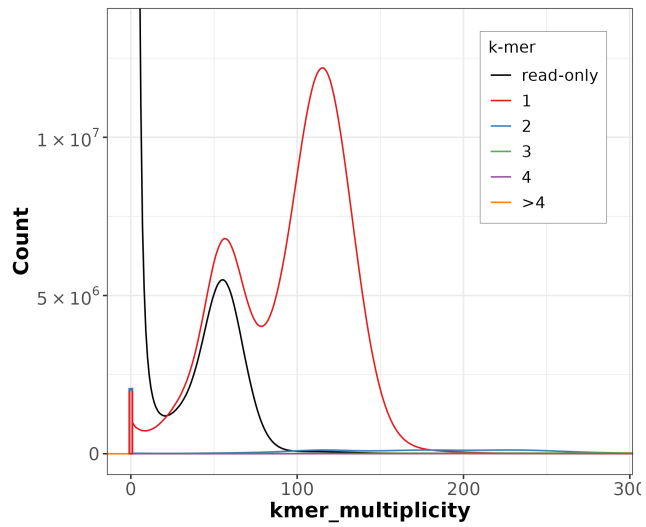
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

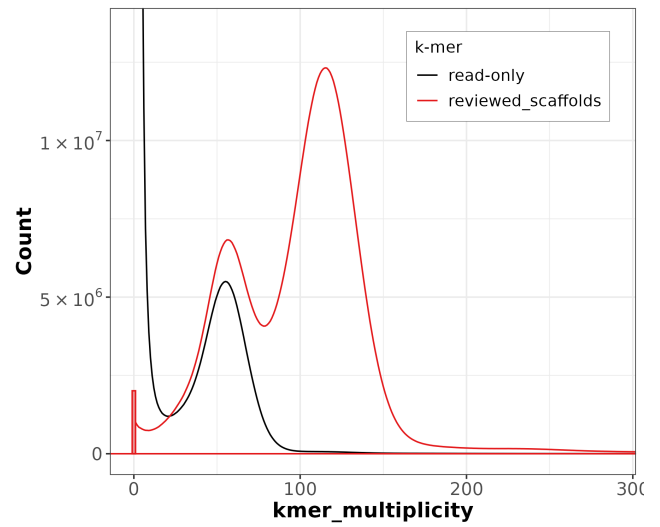


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

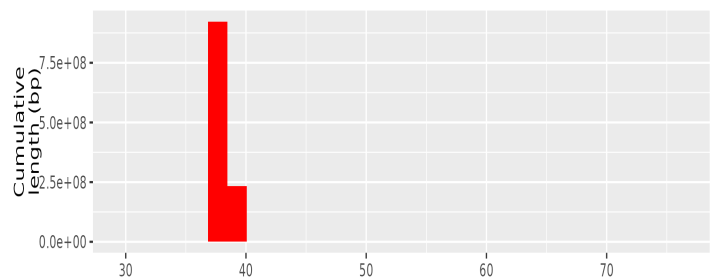


Distribution of k-mer counts per copy numbers found in asm



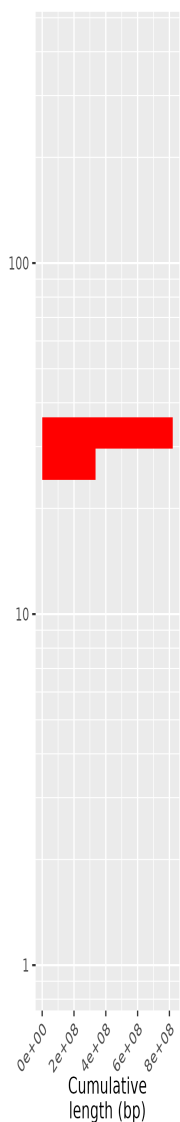
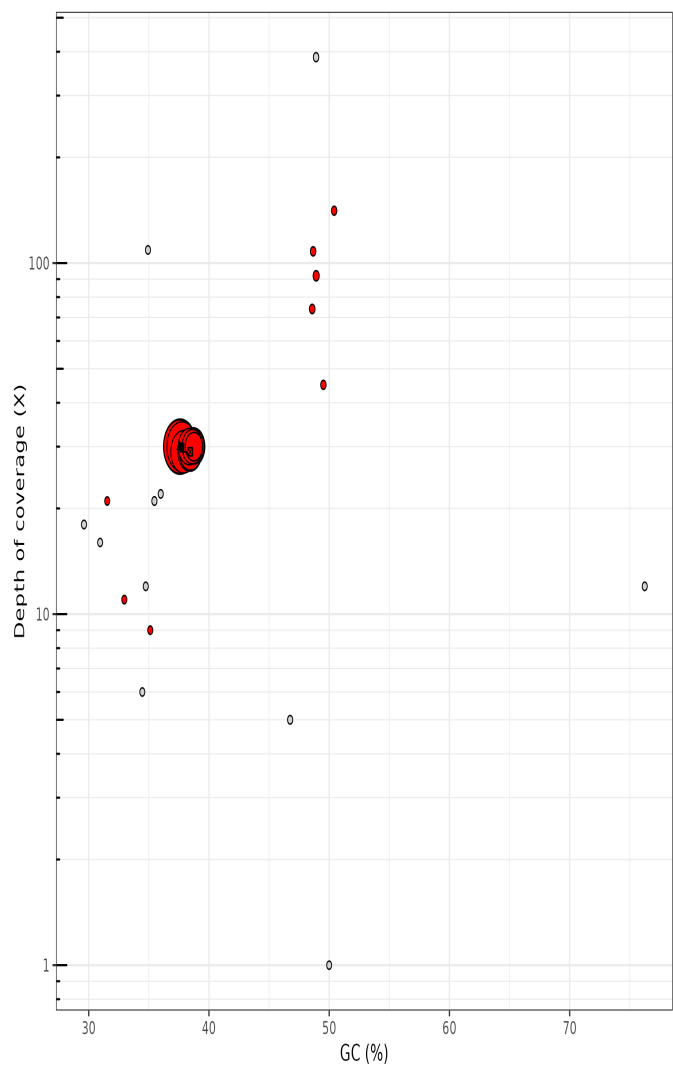
Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph

(45 0X contigs have been hidden)



Longest sequences (bp)

- xgPlaArgo1_1 - 191614605 (Eukaryota)
- ▲ xgPlaArgo1_2 - 177841216 (Eukaryota)
- xgPlaArgo1_3 - 149555881 (Eukaryota)
- + xgPlaArgo1_4 - 102259910 (Eukaryota)
- ▣ xgPlaArgo1_5 - 81370516 (Eukaryota)

Length (bp)

- 5.0e+07
- 1.0e+08
- 1.5e+08

superkingdom

- Eukaryota
- N/A

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	33	213

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: Mohamed Meftah Saoues

Affiliation: Genoscope

Date and time: 2026-07-14 08:21:27 CEST