

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	2980509
ToLID	xgGeiPlan2
Species	Geitodoris planata
Class	Gastropoda
Order	Nudibranchia

Genome Traits	Expected	Observed
Haploid size (bp)	1,054,776,119	1,017,892,946
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: 5.7.Q39

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

- . QV value is less than 40 for collapsed
- . Kmer completeness value is less than 90 for collapsed
- . BUSCO single copy value is less than 90% for collapsed
- . More than 1000 gaps/Gbp for collapsed

Curator notes

- . Interventions/Gb: 149
- . Contamination notes: ""
- . Other observations: "The assembly of *Geitodoris planata* (xgGeiPlan2) is based on 29X ONT data and 144X 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 Nextdenovo, 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 1.116 Mb (with the largest being 0.279 Mb). Additionally, 662 regions totaling 37.058 Mb (with the largest being 0.167 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, 4 haplotypic regions and 13 contaminant sequences were removed, totaling 1.258Mb and 1.883Mb, respectively (with the largest being 0.416Mb and 0.334Mb). The telomeric pattern TTAGGG was identified with TelFinder and used to generate the Pretext Telomeres tracks. Chromosome-scale scaffolds confirmed

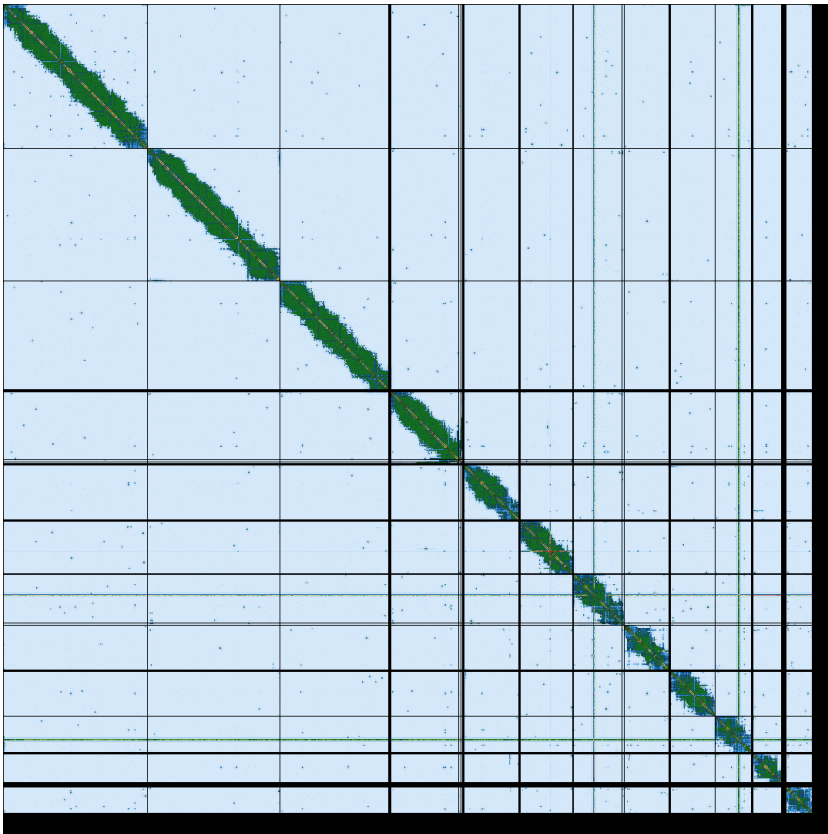
by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,021,016,221	1,017,892,946
GC %	39.55	39.55
Gaps/Gbp	3,283	3,325.5
Total gap bp	335,200	347,400
Scaffolds	320	269
Scaffold N50	123,612,349	82,390,730
Scaffold L50	4	4
Scaffold L90	11	11
Contigs	3,672	3,654
Contig N50	388,265	388,391
Contig L50	493	491
Contig L90	2,468	2,455
QV	39.0532	39.0546
Kmer compl.	69.5592	69.4284
BUSCO sing.	86.5%	86.5%
BUSCO dupl.	0.9%	0.9%
BUSCO frag.	1.9%	1.9%
BUSCO miss.	10.7%	10.8%

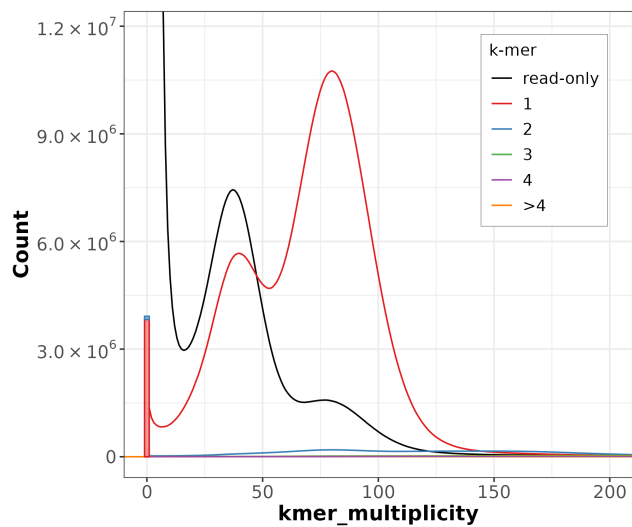
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

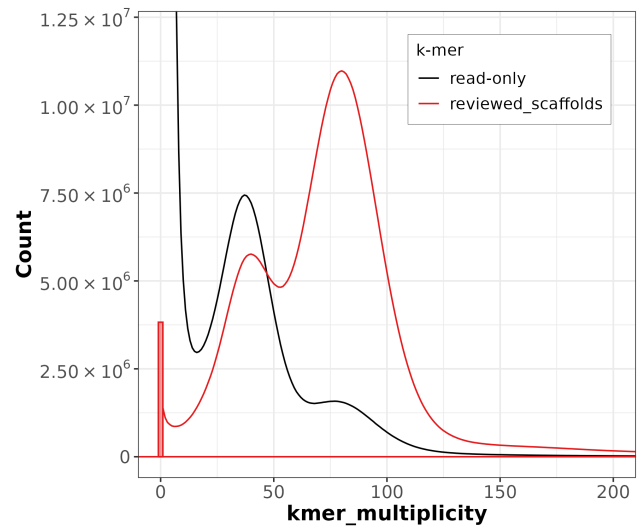


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

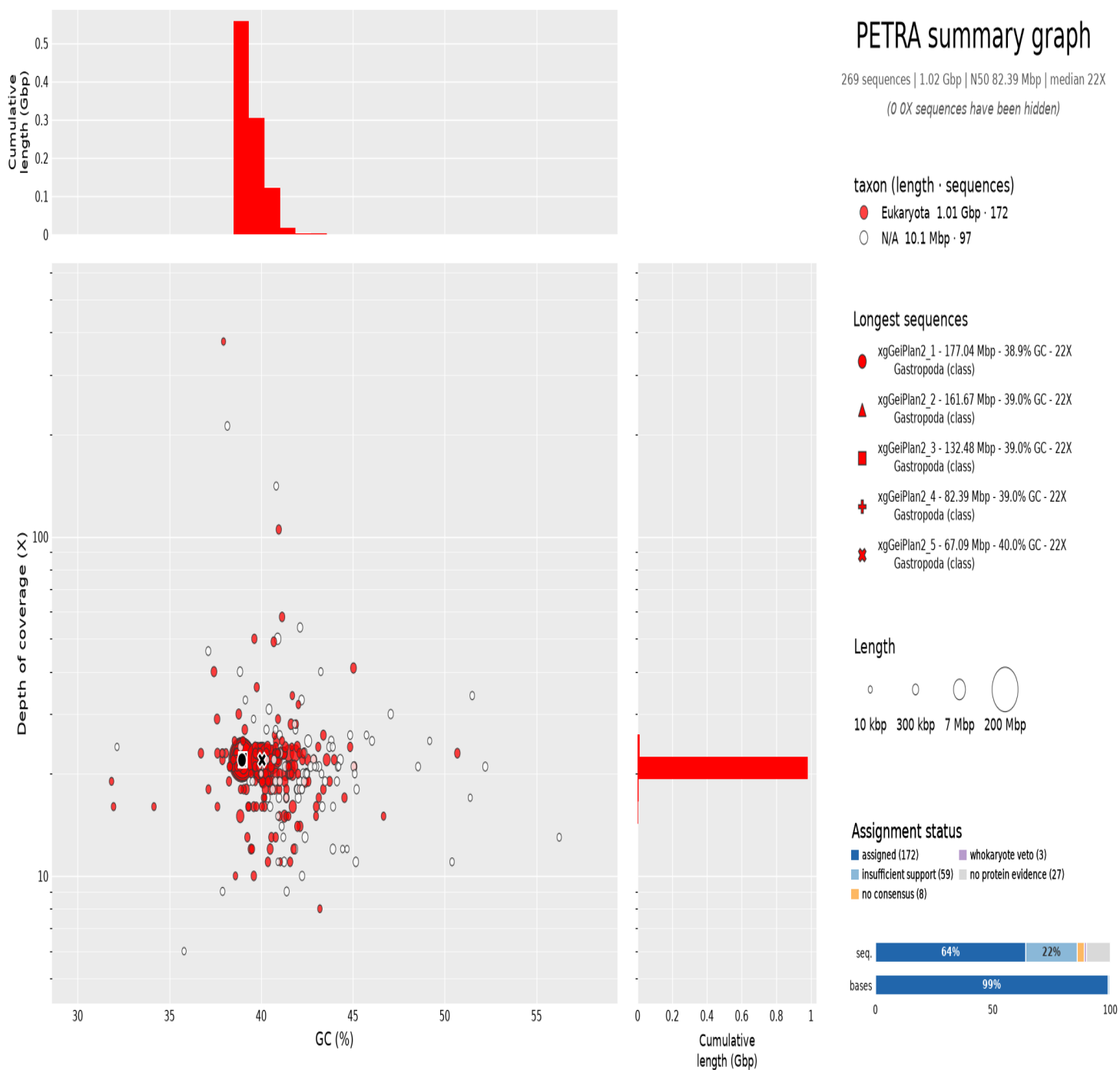


Distribution of k-mer counts per copy numbers found in asm



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	29	144

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: Sophie Layac

Affiliation: Genoscope

Date and time: 2026-10-08 16:37:35 CEST