ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	120394
ToLID	**xgDisAtro1.1**
Species	Peltodoris atromaculata
Class	Gastropoda
Order	Nudibranchia

Genome Traits	Expected	Observed
Haploid size (bp)	1,150,992,015	1,335,860,655
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.Q51

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

. BUSCO single copy value is less than 90% for collapsed
. More than 1000 gaps/Gbp for collapsed

Curator notes

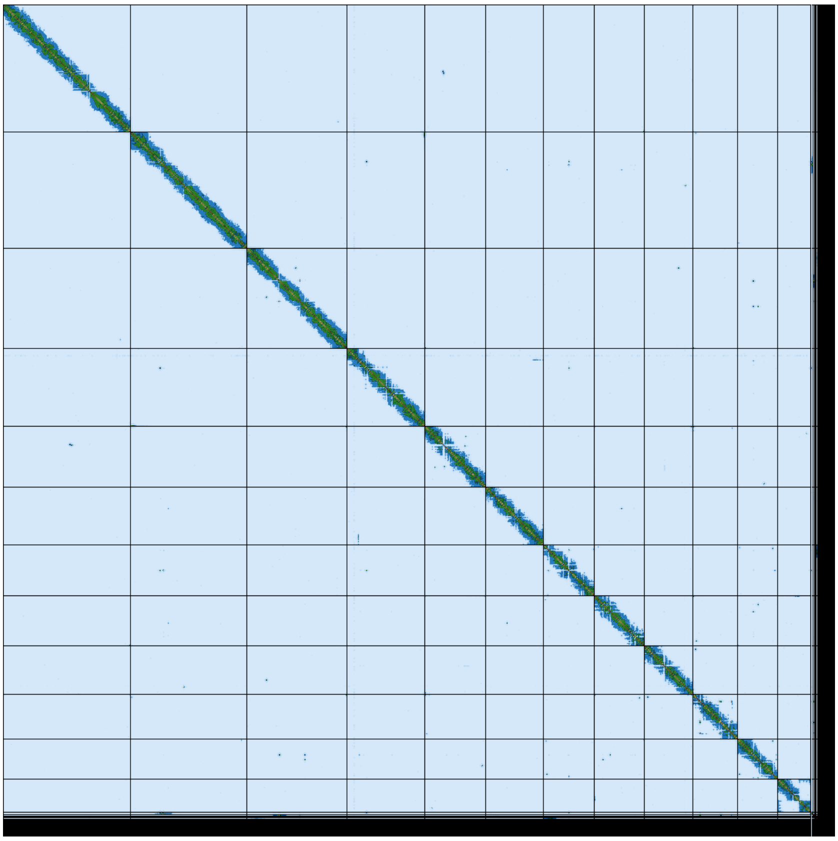
. Interventions/Gb: 108
. Contamination notes: ""
. Other observations: "The assembly of Peltodoris atromaculata (Discodoris atromaculata xgDisAtro1) is based on 28X PacBio data and Arima HighCoverage Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included 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, 981 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 32.9Mb (with the largest being 445 Kb). Additionally, 640 regions totaling 70 Mb (with the largest being 5.3 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 29 haplotypic regions were removed, totaling 5.1 Mb (with the largest being 0.657 Mb) "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,342,034,293	1,335,860,655
GC %	38.06	38.05
Gaps/Gbp	1,367.33	1,403.59
Total gap bp	183,500	193,900
Scaffolds	755	303
Scaffold N50	125,895,859	125,817,325
Scaffold L50	4	4
Scaffold L90	11	11
Contigs	2,590	2,178
Contig N50	1,446,143	1,427,876
Contig L50	274	273
Contig L90	1,026	1,027
QV	42.211	51.2678
Kmer compl.	92.359	93.5071
BUSCO sing.	86.5%	86.4%
BUSCO dupl.	0.9%	0.8%
BUSCO frag.	2.5%	2.5%
BUSCO miss.	10.1%	10.3%

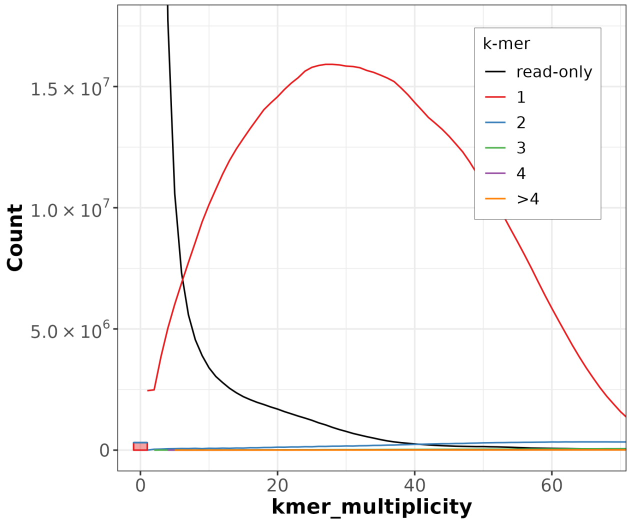
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: mollusca_odb10 (genomes:7, BUSCOs:5295)

HiC contact map of curated assembly

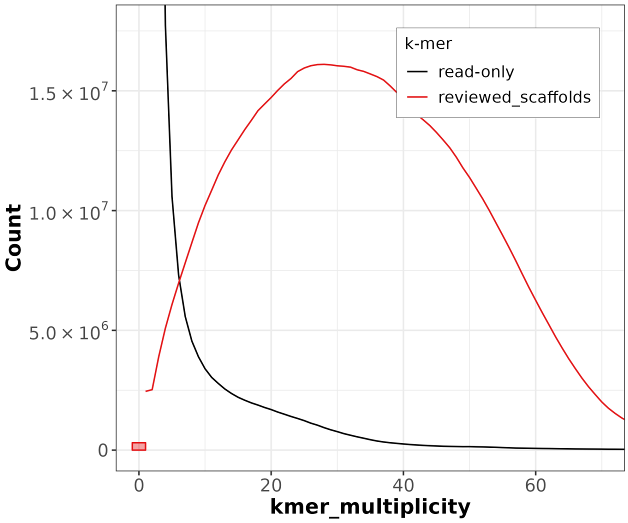


collapsed [LINK]

K-mer spectra of curated assembly

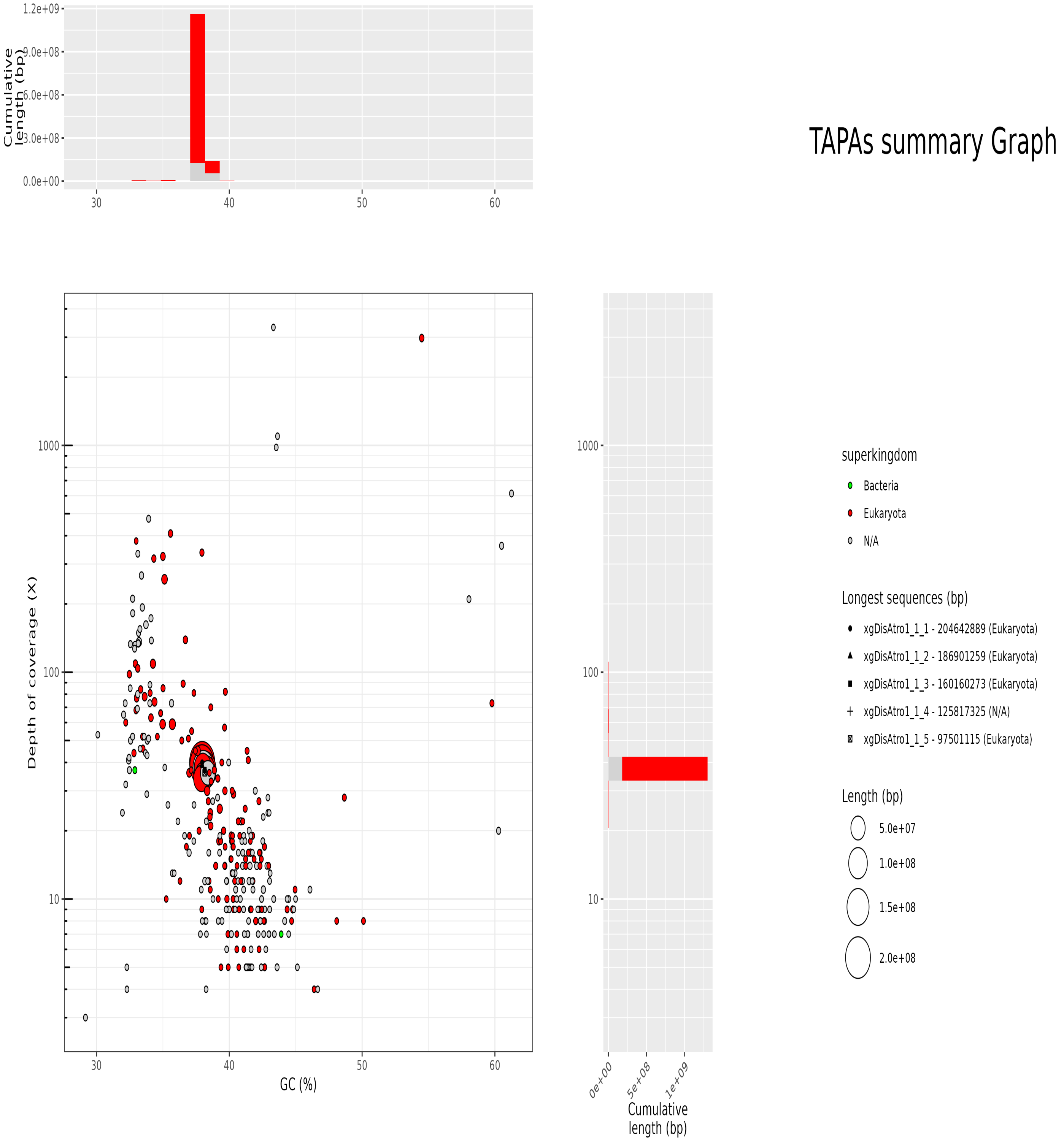


Distribution of k-mer counts per copy numbers found in asm



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	PACBIO Hifi	Arima
Coverage	40	185

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: Caroline Menguy
Affiliation: Genoscope

Date and time: 2025-01-27 11:52:01 CET