

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	1677693
ToLID	xgDorPseu5
Species	Doris pseudoargus
Class	Gastropoda
Order	Nudibranchia

Genome Traits	Expected	Observed
Haploid size (bp)	1,528,329,540	1,580,293,460
Haploid Number	12 (source: ancestor)	13
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q43

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

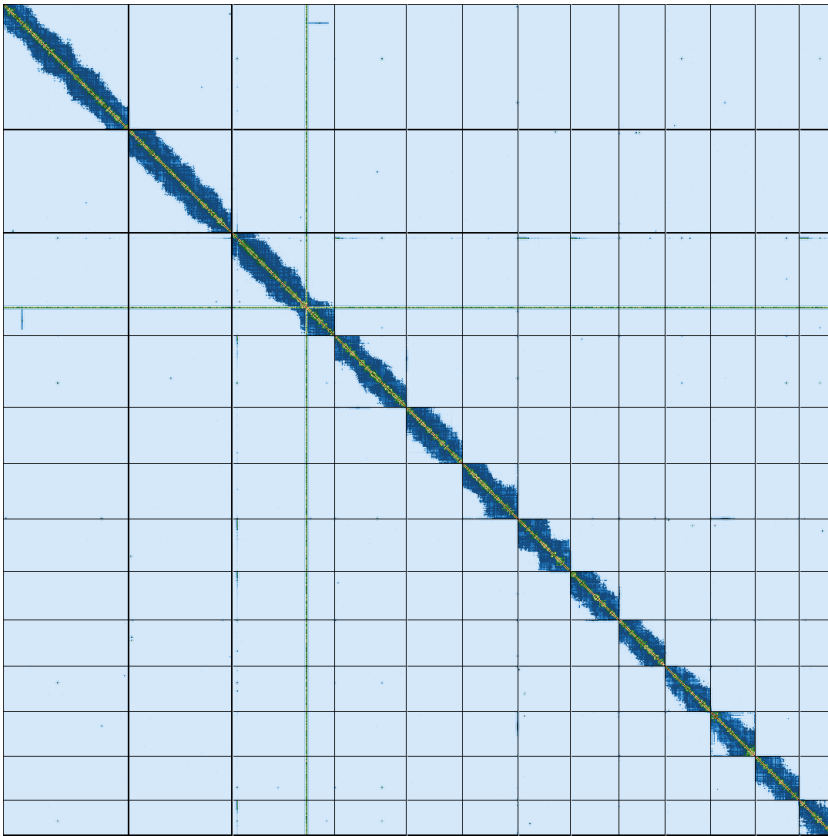
. Interventions/Gb: 51
. Contamination notes: ""
. Other observations: "The assembly of *Doris pseudoargus* (xgDorPseu5) is based on 36X ONT data and 283X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial 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, 3 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.358 Mb (with the largest being 0.155 Mb). Additionally, 327 regions totaling 26.735 Mb (with the largest being 0.296 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using ptGAUL. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 1 haplotypic region was removed, totaling 0.146Mb. 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	1,580,784,540	1,580,293,460
GC %	39.05	39.05
Gaps/Gbp	850.84	868.19
Total gap bp	134,500	142,500
Scaffolds	206	77
Scaffold N50	105,440,723	106,163,881
Scaffold L50	5	5
Scaffold L90	11	11
Contigs	1,551	1,449
Contig N50	2,482,702	2,482,702
Contig L50	190	190
Contig L90	702	701
QV	43.0674	43.0691
Kmer compl.	89.7195	89.7045
BUSCO sing.	90.0%	90.0%
BUSCO dupl.	0.6%	0.6%
BUSCO frag.	6.7%	6.7%
BUSCO miss.	2.7%	2.7%

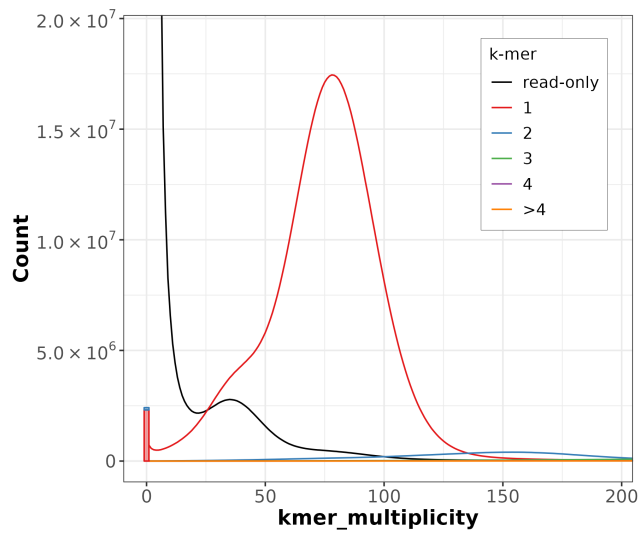
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

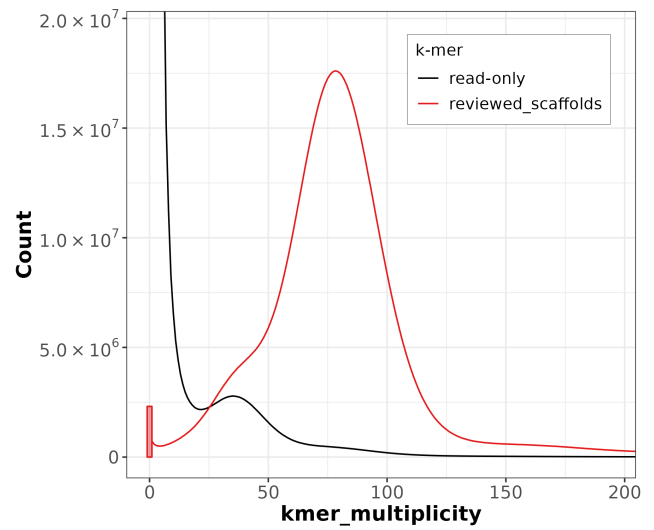


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

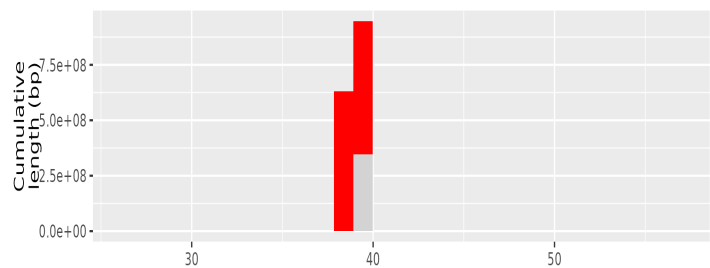


Distribution of k-mer counts per copy numbers found in asm

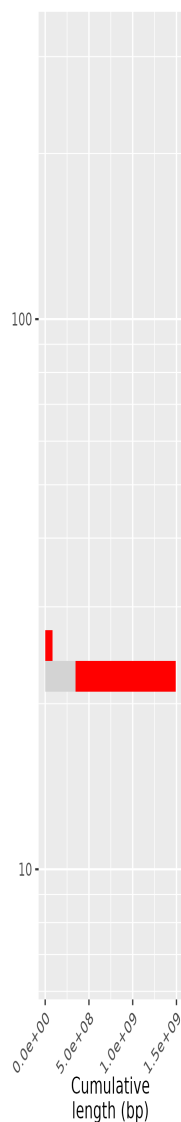
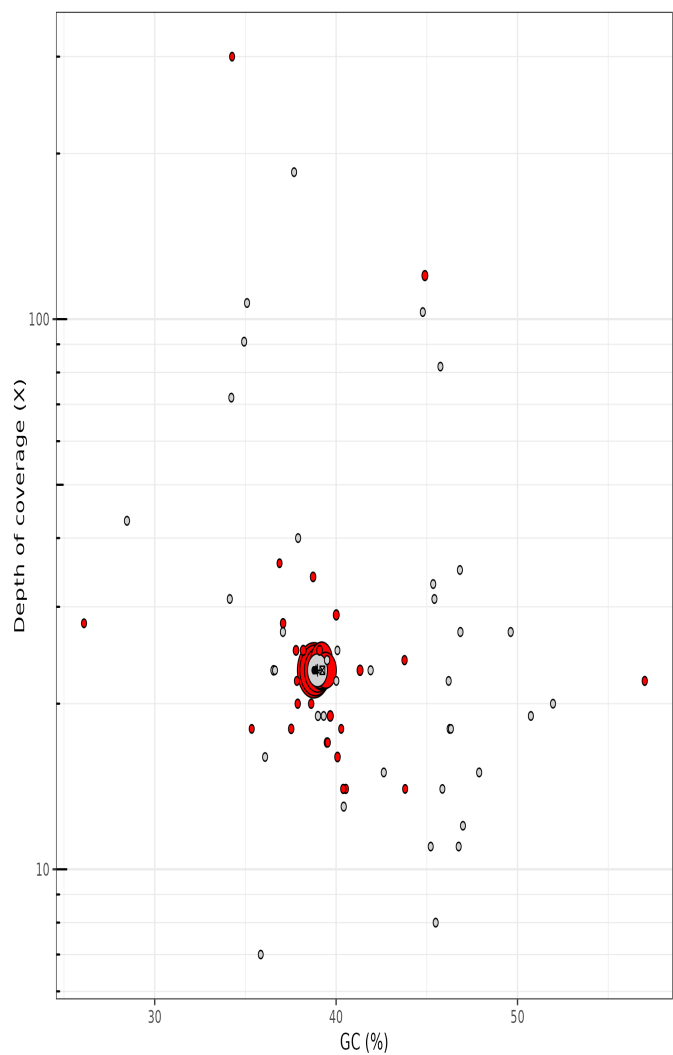


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



- Longest sequences (bp)
- xgDorPseu5_1 - 238457566 (Eukaryota)
 - ▲ xgDorPseu5_2 - 195503930 (Eukaryota)
 - xgDorPseu5_3 - 195352333 (Eukaryota)
 - + xgDorPseu5_4 - 136022056 (Eukaryota)
 - ▣ xgDorPseu5_5 - 106163881 (Eukaryota)

- superkingdom
- Eukaryota
 - N/A

- Length (bp)
- 5.0e+07
 - 1.0e+08
 - 1.5e+08
 - 2.0e+08

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	26	283

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: 2025-08-09 11:02:17 CEST