ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	6511
ToLID	**xgDolAuri1**
Species	Dolabella auricularia
Class	Gastropoda
Order	Aplysiida

Genome Traits	Expected	Observed
Haploid size (bp)	1,040,347,989	1,169,812,557
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

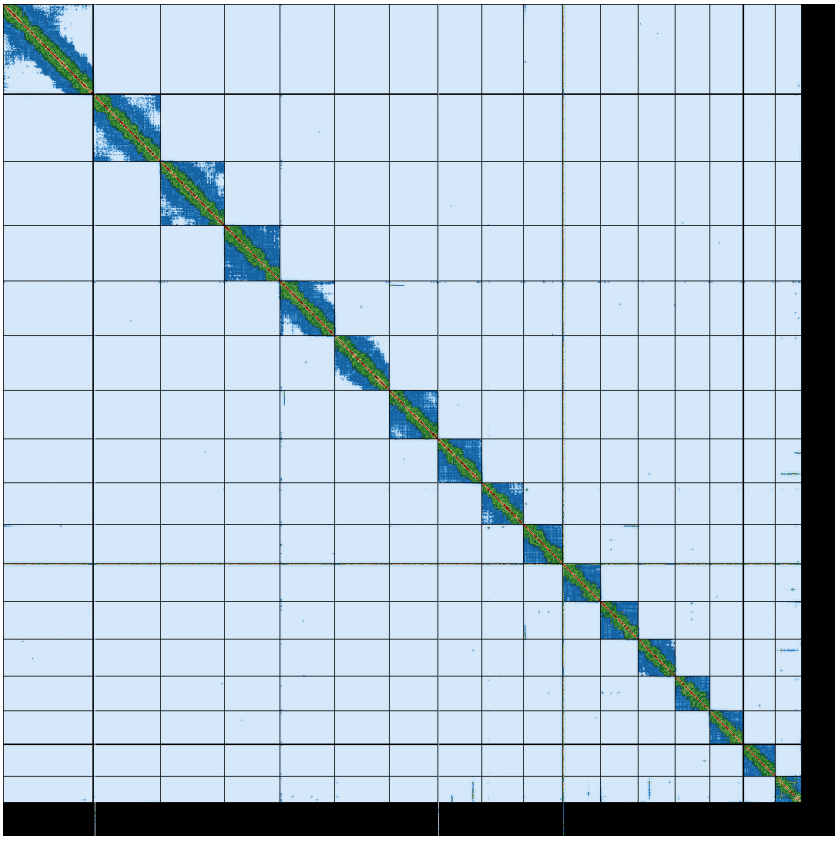
. Interventions/Gb: 431
. Contamination notes: ""
. Other observations: "The assembly of Dolabella auricularia (xgDolAuri1) is based on 31X ONT data and 146X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial ONT assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 96 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 4.295 Mb (with the largest being 0.273 Mb). Additionally, 4428 regions totaling 92.023 Mb (with the largest being 0.258 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, 9 haplotypic regions were removed, totaling 6.647Mb (with the largest being 1.454Mb). 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	1,176,648,313	1,169,812,557
GC %	43.3	43.29
Gaps/Gbp	3,370.59	3,567.24
Total gap bp	396,600	453,400
Scaffolds	1,649	1,404
Scaffold N50	66,322,656	67,472,486
Scaffold L50	7	7
Scaffold L90	17	16
Contigs	5,615	5,577
Contig N50	591,261	603,969
Contig L50	446	440
Contig L90	2,697	2,679
QV	32.5243	32.5078
Kmer compl.	80.519	80.144
BUSCO sing.	97.2%	97.5%
BUSCO dupl.	1.9%	1.7%
BUSCO frag.	0.6%	0.5%
BUSCO miss.	0.2%	0.3%

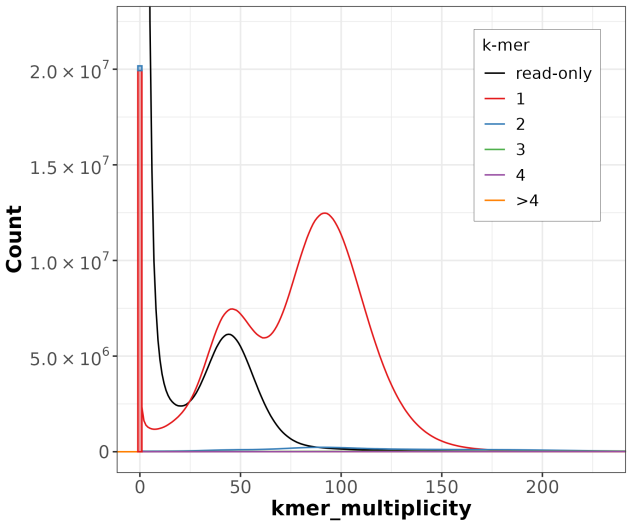
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

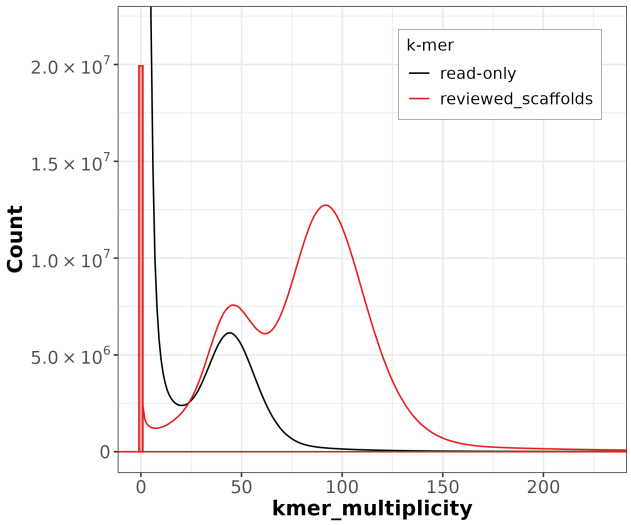


collapsed [LINK]

K-mer spectra of curated assembly

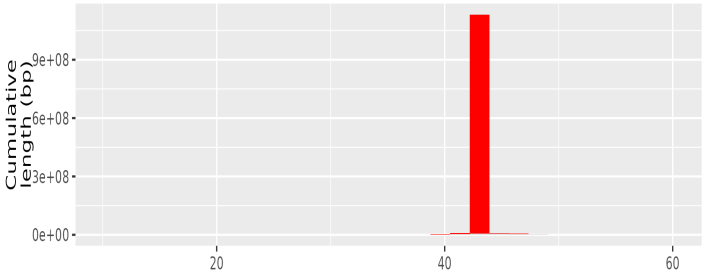


Distribution of k-mer counts per copy numbers found in asm

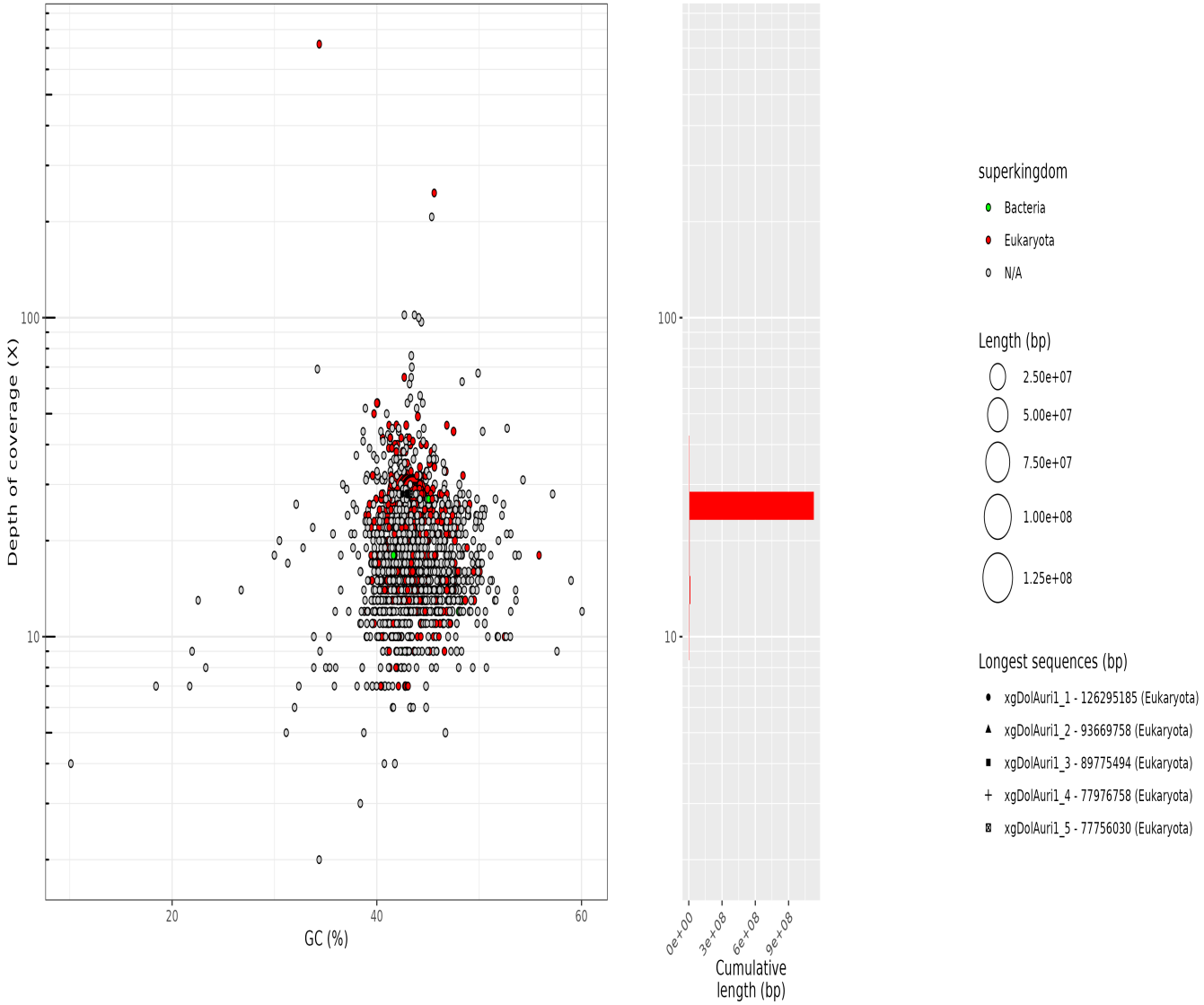


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	31	147

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: Arnaud Couloux
Affiliation: Genoscope

Date and time: 2026-04-03 02:06:47 CEST