ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	3818472
ToLID	**xgPhySpea1**
Species	Phyllidiella sp. A-5R5ARP2QW
Class	Gastropoda
Order	Nudibranchia

Genome Traits	Expected	Observed
Haploid size (bp)	2,134,297,663	2,142,176,149
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.Q41

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

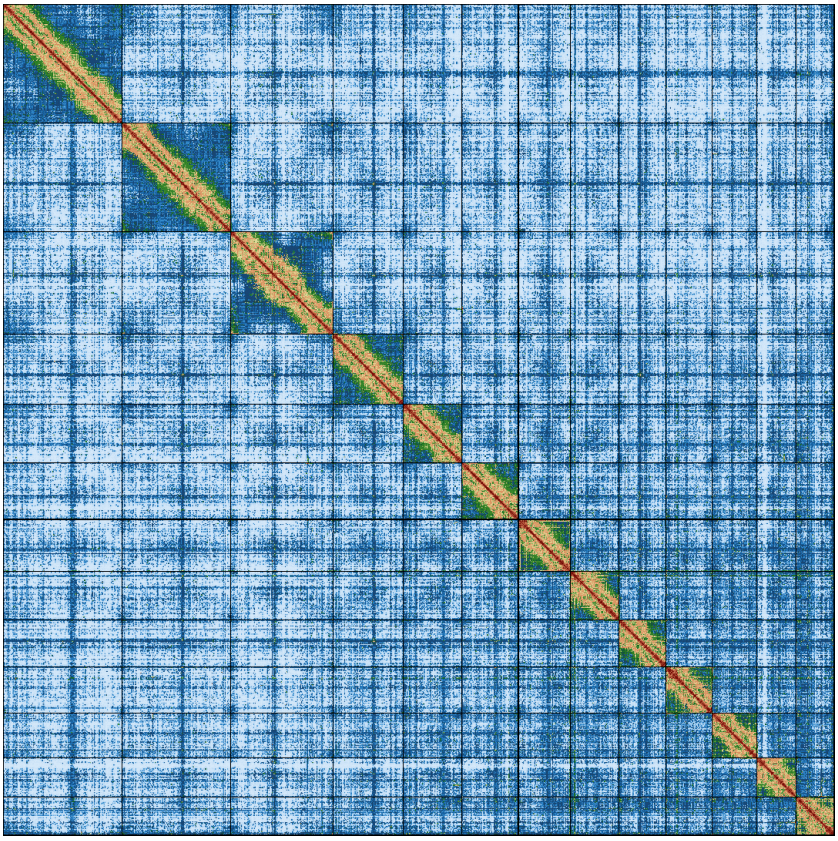
. Interventions/Gb: 64
. Contamination notes: ""
. Other observations: "The assembly of Phyllidiella sp. A-5R5ARP2QW (xgPhySpea1) is based on 46X ONT data and 159X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 20 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 7.373 Mb (with the largest being 3.767 Mb). Additionally, 993 regions totaling 77.738 Mb (with the largest being 0.983 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, 24 haplotypic regions were removed, totaling 6.838Mb (with the largest being 0.807Mb). 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	2,149,060,203	2,142,176,149
GC %	35.65	35.65
Gaps/Gbp	535.58	548.04
Total gap bp	115,100	124,700
Scaffolds	121	92
Scaffold N50	143,542,224	149,933,118
Scaffold L50	5	5
Scaffold L90	12	11
Contigs	1,272	1,266
Contig N50	3,058,331	3,058,331
Contig L50	212	211
Contig L90	697	695
QV	41.7388	41.7372
Kmer compl.	89.4209	89.3562
BUSCO sing.	97.2%	95.1%
BUSCO dupl.	1.2%	0.9%
BUSCO frag.	0.3%	2.5%
BUSCO miss.	1.2%	1.5%

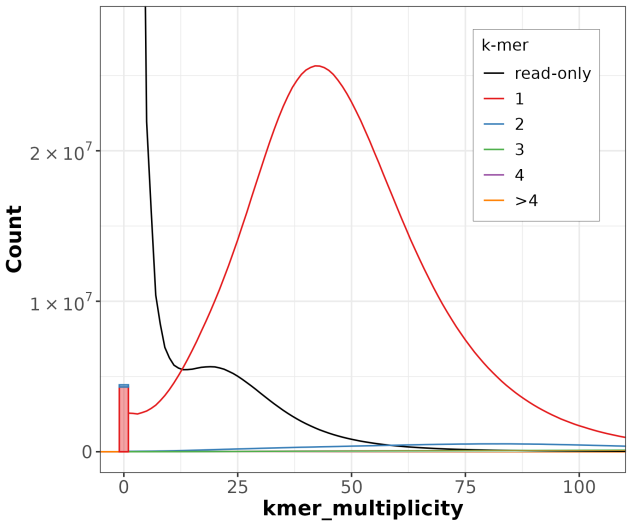
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

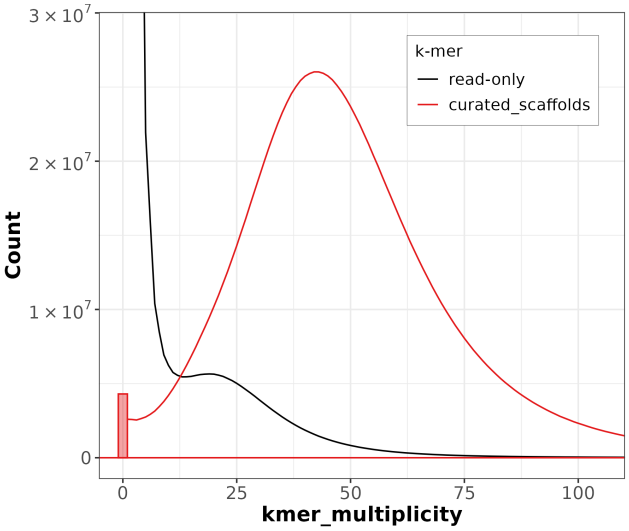


collapsed [LINK]

K-mer spectra of curated assembly

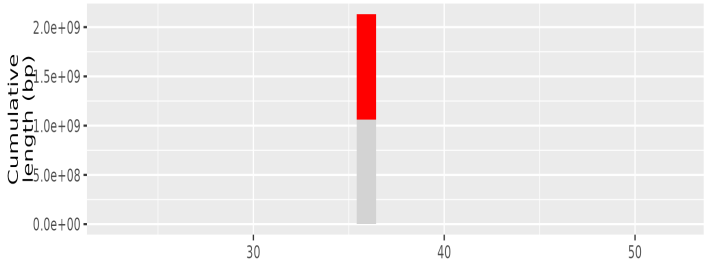


Distribution of k-mer counts per copy numbers found in asm



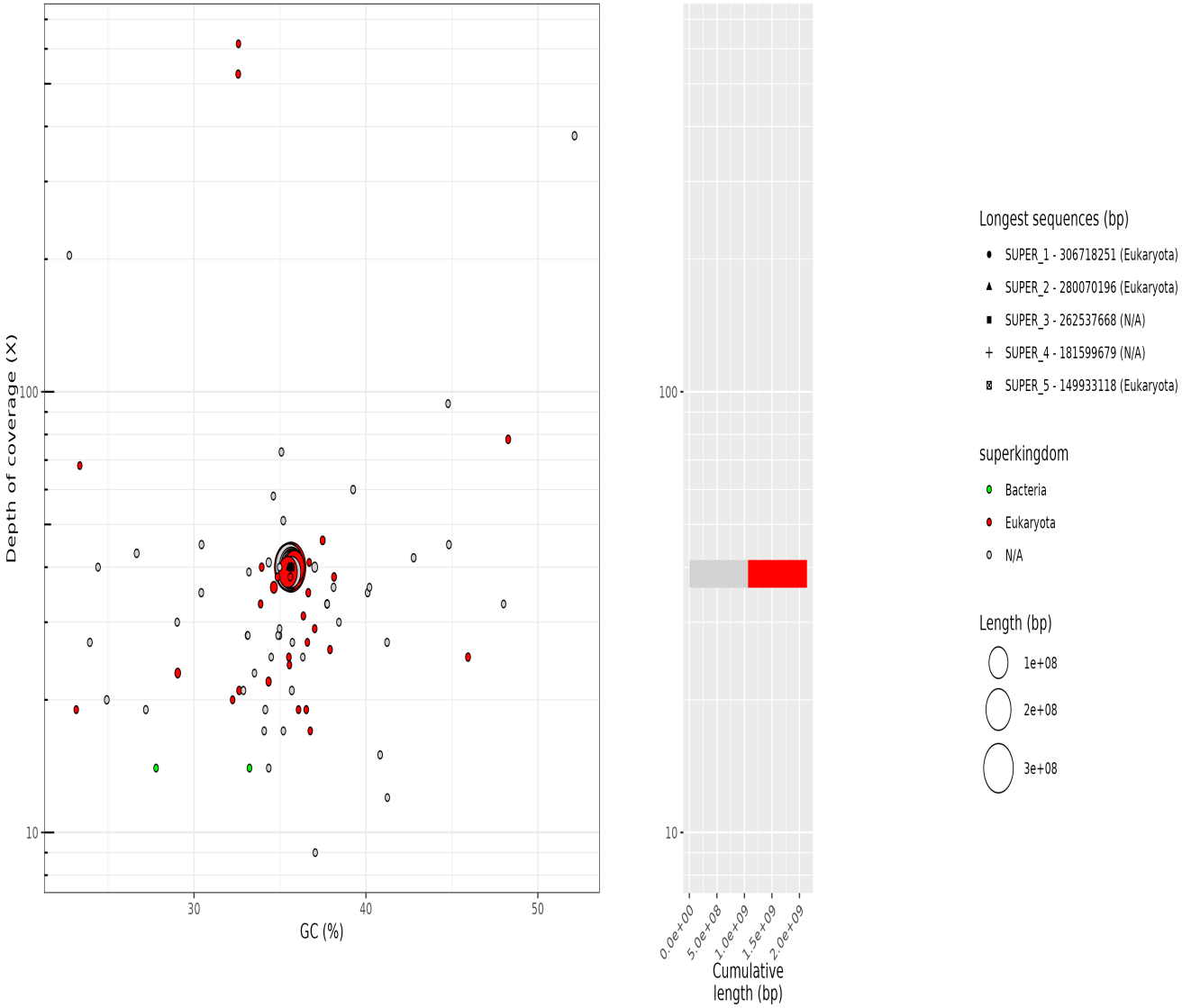
Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph

(1 0X contig has been hidden)



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	46	159

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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2026-09-29 15:00:05 CEST