ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	121477
ToLID	**odAplCave1**
Species	Aplysina cavernicola
Class	Demospongiae
Order	Verongiida

Genome Traits	Expected	Observed
Haploid size (bp)	334,795,572	168,134,899
Haploid Number	5 (source: ancestor)	21
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.6.Q52

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

. Observed Haploid size (bp) has >20% difference with Expected
. Observed Haploid Number is different from Expected

. Kmer completeness value is less than 90 for collapsed
. BUSCO single copy value is less than 90% for collapsed
. Assembly length loss > 3% for collapsed

Curator notes

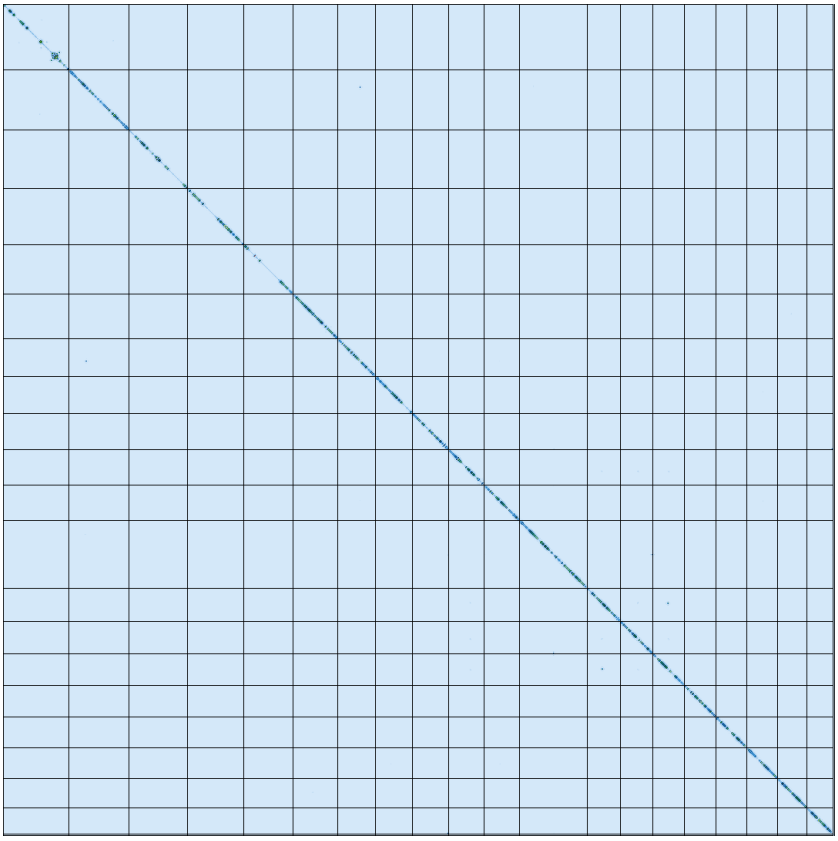
. Interventions/Gb: 81
. Contamination notes: ""
. Other observations: "The assembly of Aplysina cavernicola (odAplCave1) is based on 108X PacBio data and 200X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 3694 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 626 533 262 Mb (with the largest being 11 878 367 Mb). Additionally, 58 regions totaling 10 191 145 Mb (with the largest being 1 057 097 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, 25 contaminant sequences were removed, totaling 1 452 007 Mb (with the largest being 217 397 kb) . 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	351,189,821	168,134,899
GC %	54.94	46.61
Gaps/Gbp	1,816.68	398.49
Total gap bp	63,800	7,400
Scaffolds	2,147	24
Scaffold N50	4,084,083	7,589,756
Scaffold L50	20	8
Scaffold L90	571	18
Contigs	2,785	91
Contig N50	1,706,082	5,100,857
Contig L50	39	14
Contig L90	1,065	36
QV	28.9392	52.811
Kmer compl.	21.3865	23.2234
BUSCO sing.	65.2%	70.9%
BUSCO dupl.	4.3%	0.8%
BUSCO frag.	10.9%	11.8%
BUSCO miss.	19.6%	16.5%

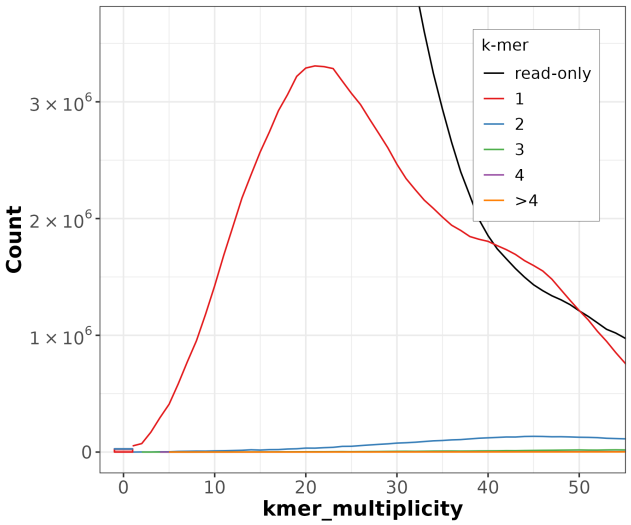
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

HiC contact map of curated assembly

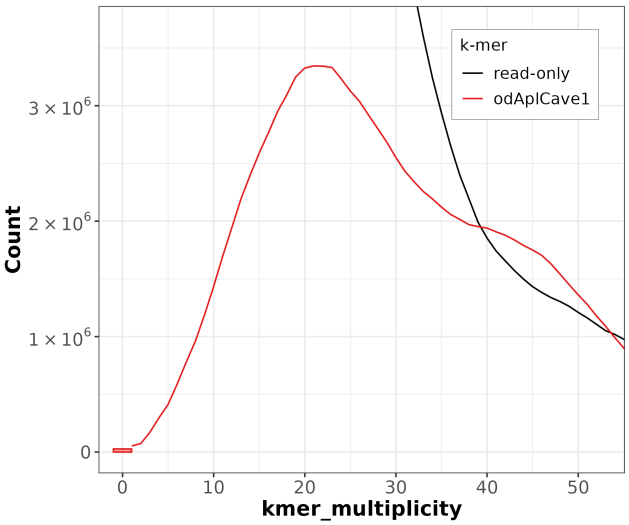


collapsed [LINK]

K-mer spectra of curated assembly

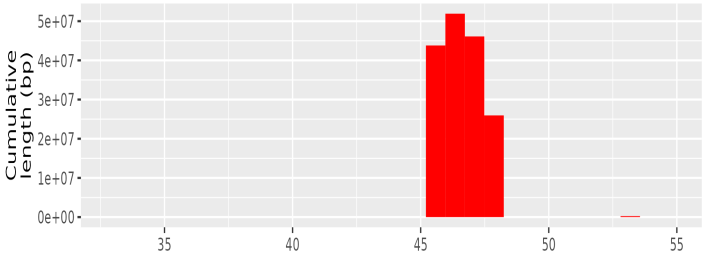


Distribution of k-mer counts per copy numbers found in asm

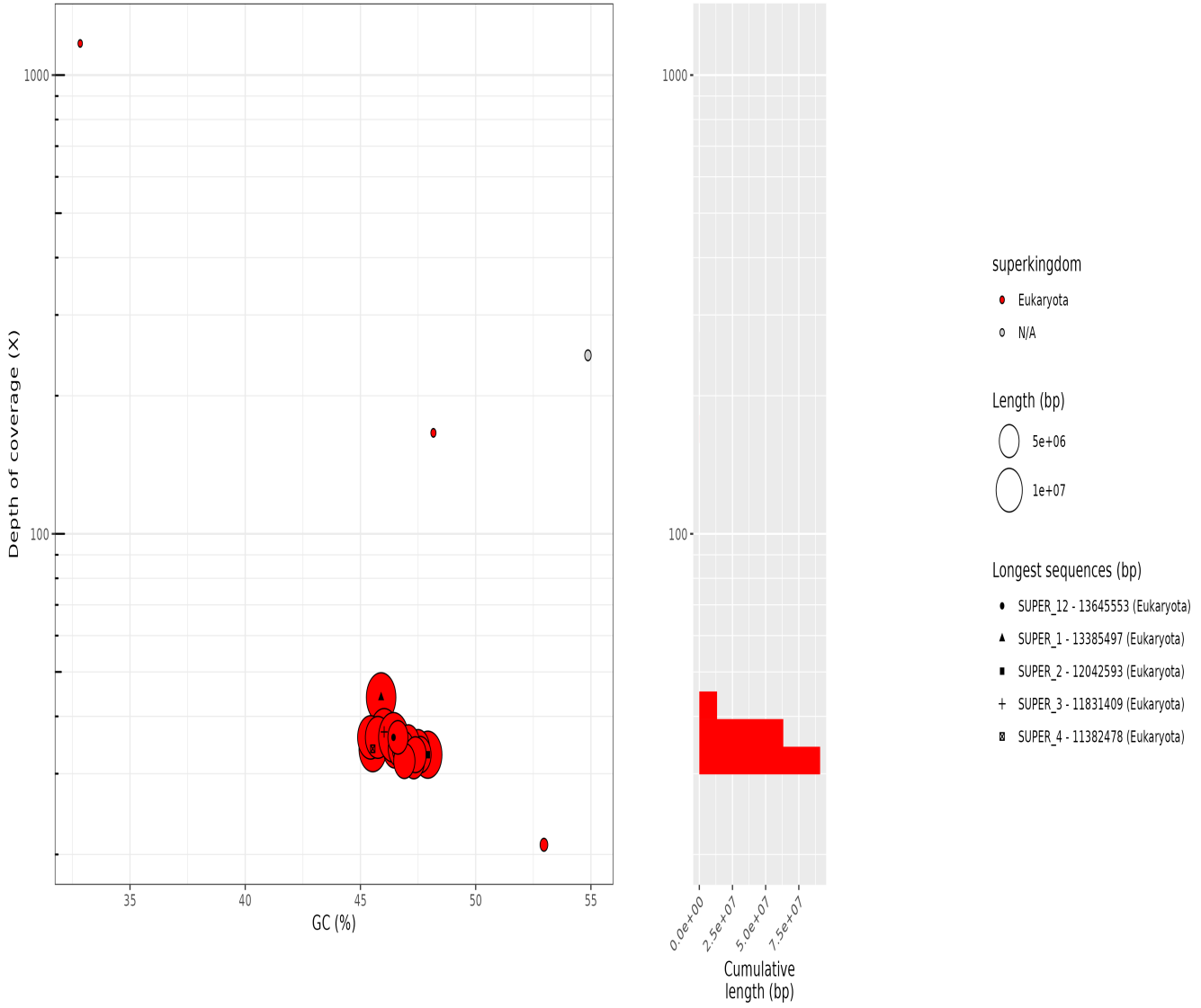


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	PACBIO Hifi	Arima
Coverage	108	200

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: Sophie Mangenot
Affiliation: Genoscope

Date and time: 2024-12-17 21:45:12 CET