ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	798306
ToLID	**odAxiDiss2**
Species	Axinella dissimilis
Class	Demospongiae
Order	Axinellida

Genome Traits	Expected	Observed
Haploid size (bp)	187,652,185	153,628,644
Haploid Number	5 (source: ancestor)	27
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.6.Q42

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
. BUSCO single copy value is less than 90% for collapsed

Curator notes

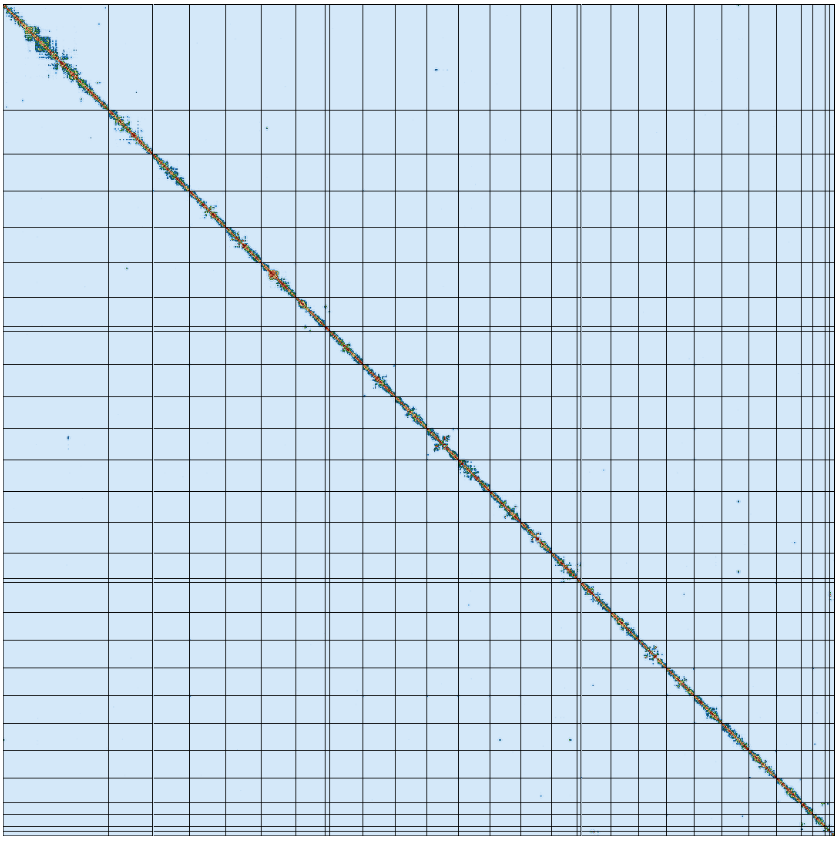
. Interventions/Gb: 437
. Contamination notes: ""
. Other observations: "The assembly of Axinella dissimilis (odAxiDiss2) is based on 135X PacBio data and 178X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 1521 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 171.614 Mb (with the largest being 5.868 Mb). Additionally, 155 regions totaling 10.097 Mb (with the largest being 0.516 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, 7 haplotypic regions and 84 contaminant sequences were removed, totaling 0.629 and 2.435Mb, respectively (with the largest being 0.161 Mb and 0.195 Mb).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	157,391,708	153,628,644
GC %	43.87	44.09
Gaps/Gbp	438.4	572.81
Total gap bp	6,900	11,000
Scaffolds	125	34
Scaffold N50	5,727,185	5,800,342
Scaffold L50	12	10
Scaffold L90	27	22
Contigs	194	122
Contig N50	2,473,000	2,400,692
Contig L50	20	19
Contig L90	65	65
QV	36.0086	42.5492
Kmer compl.	60.0721	59.8612
BUSCO sing.	57.4%	84.2%
BUSCO dupl.	0.6%	2.5%
BUSCO frag.	26.8%	4.9%
BUSCO miss.	15.2%	8.3%

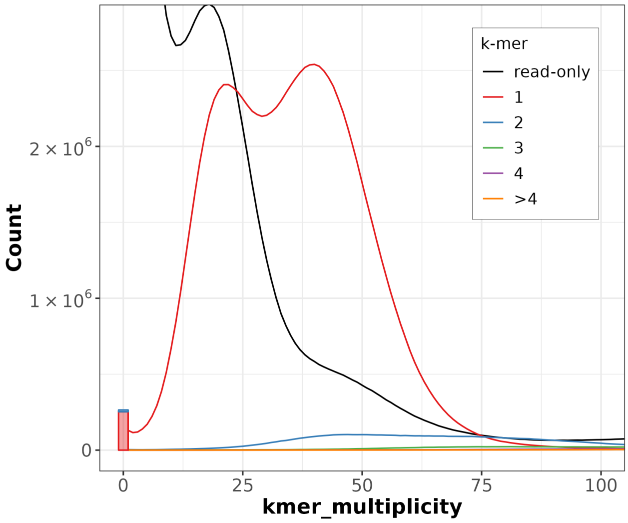
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

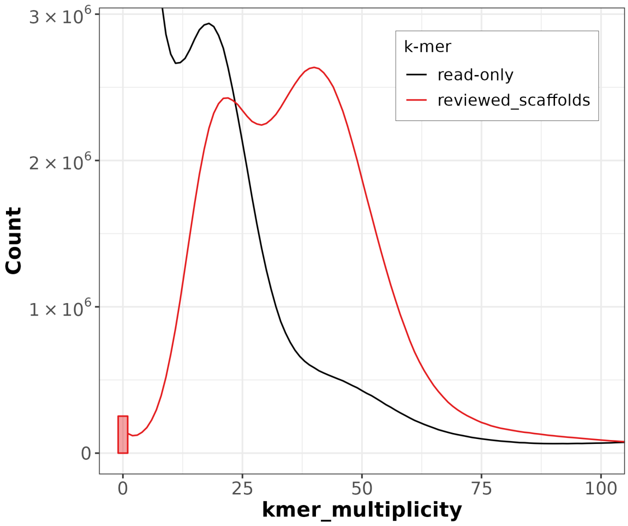


collapsed [LINK]

K-mer spectra of curated assembly

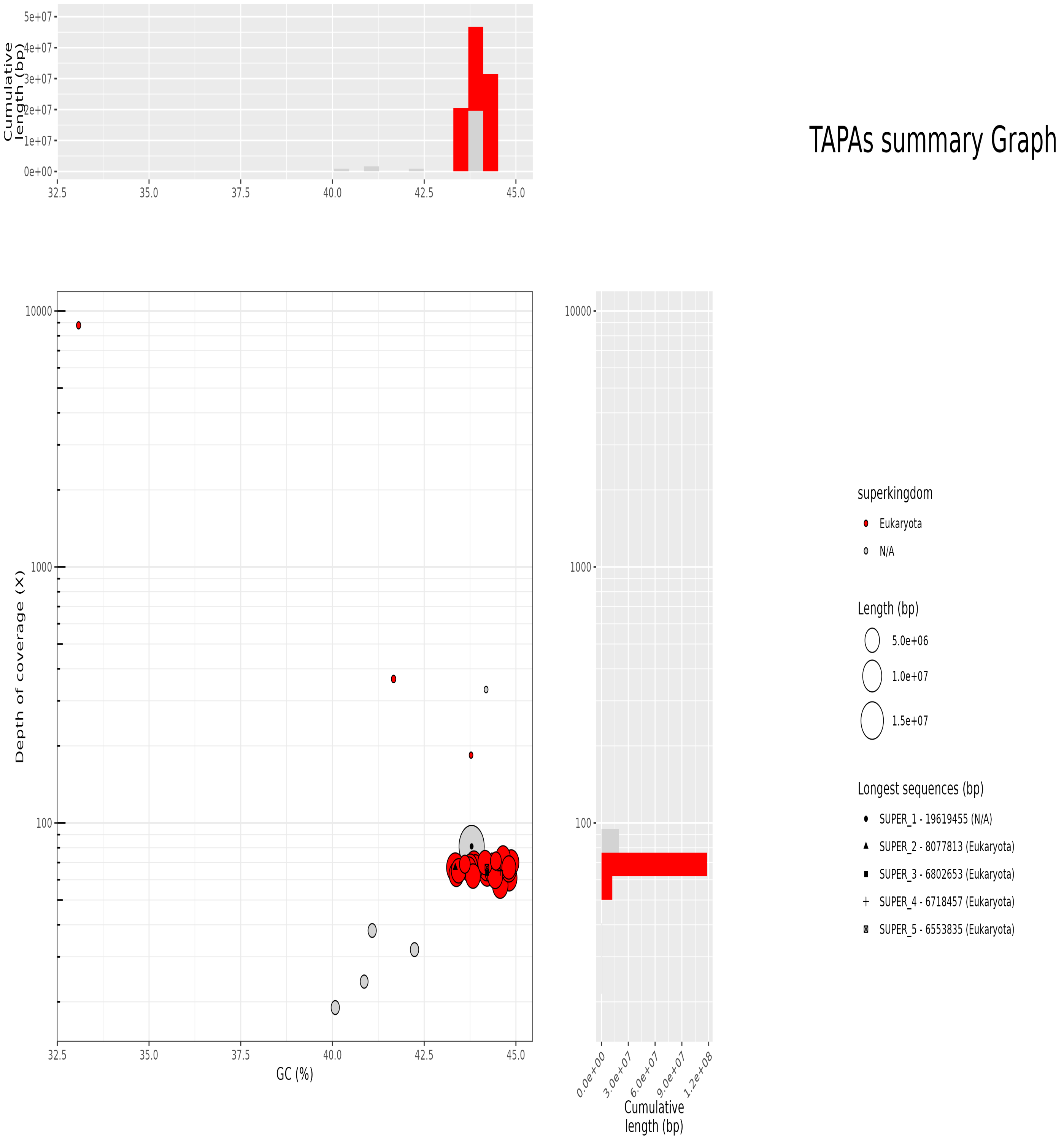


Distribution of k-mer counts per copy numbers found in asm



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	Long reads	Arima
Coverage	135	178

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 Layac
Affiliation: Genoscope

Date and time: 2026-01-20 06:22:12 CET