ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	798309
ToLID	**odCymCant1**
Species	Cymbastela cantharella
Class	Demospongiae
Order	Axinellida

Genome Traits	Expected	Observed
Haploid size (bp)	276,060,298	263,515,416
Haploid Number	5 (source: ancestor)	14
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q45

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

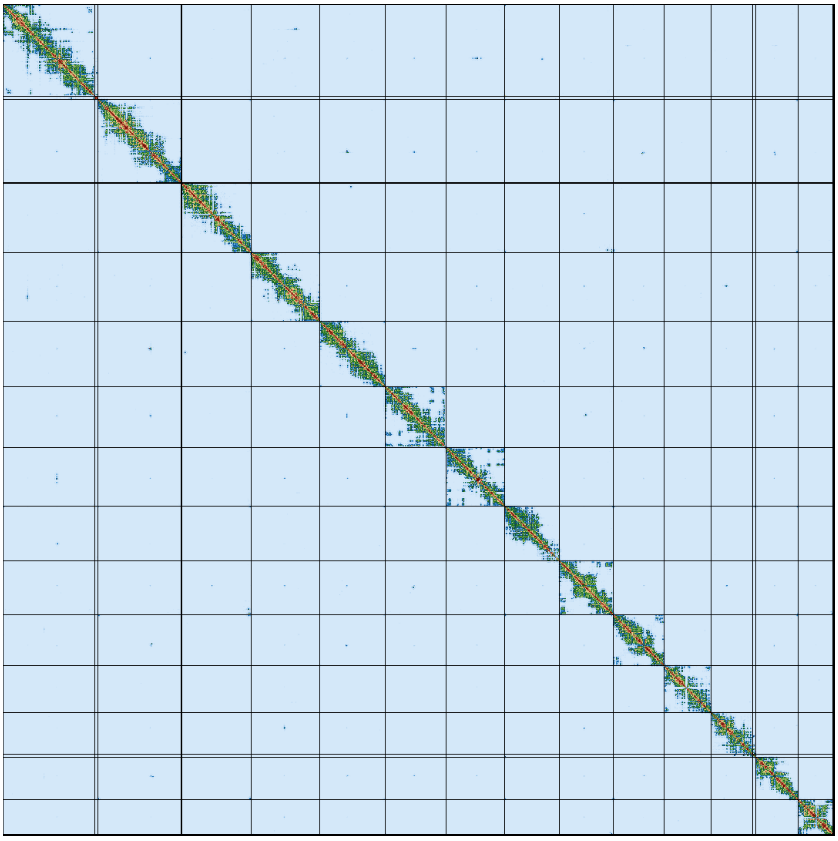
. Interventions/Gb: 49
. Contamination notes: ""
. Other observations: "The assembly of Cymbastela cantharella (odCymCant1) is based on 55X PacBio data and 168X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 293 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 48.567 Mb (with the largest being 6.372 Mb). Additionally, 170 regions totaling 14.599 Mb (with the largest being 0.609 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. During manual curation, 2 haplotypic regions and 52 contaminant sequences were removed, totaling 1.164Mb and 2.457 Mb respectively (with the largest being 0.741Mb and 0.449Mb). 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	267,111,038	263,515,416
GC %	44.16	44.24
Gaps/Gbp	505.41	466.77
Total gap bp	14,800	14,100
Scaffolds	84	29
Scaffold N50	19,224,645	19,224,645
Scaffold L50	6	6
Scaffold L90	13	13
Contigs	211	152
Contig N50	4,078,554	4,291,112
Contig L50	21	20
Contig L90	77	73
QV	43.3031	45.3589
Kmer compl.	70.6597	70.1697
BUSCO sing.	85.9%	86.2%
BUSCO dupl.	2.7%	2.4%
BUSCO frag.	4.2%	4.2%
BUSCO miss.	7.3%	7.3%

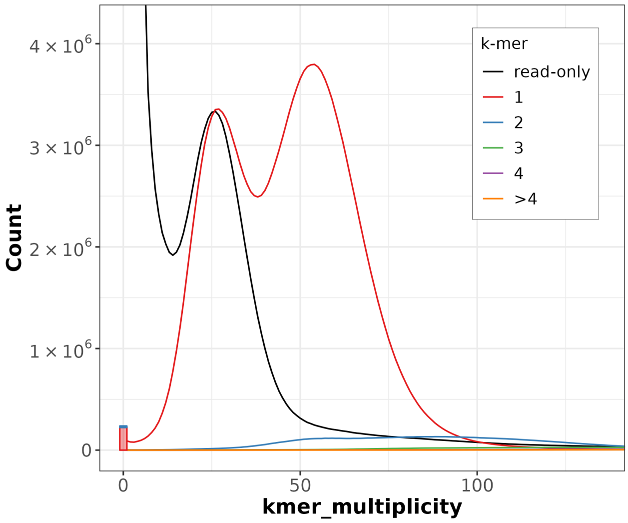
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

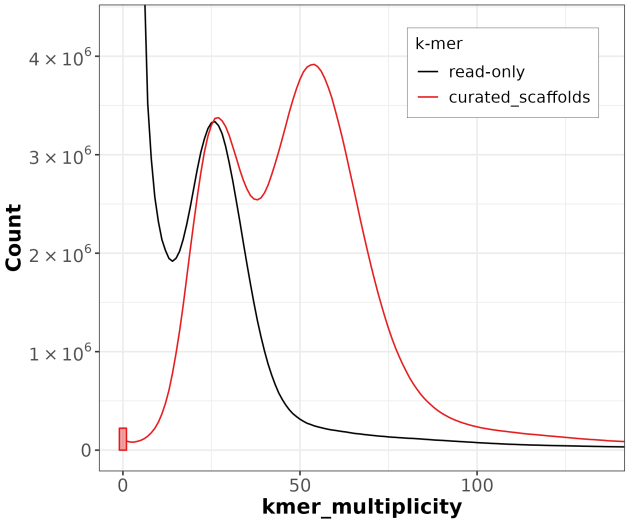


collapsed [LINK]

K-mer spectra of curated assembly

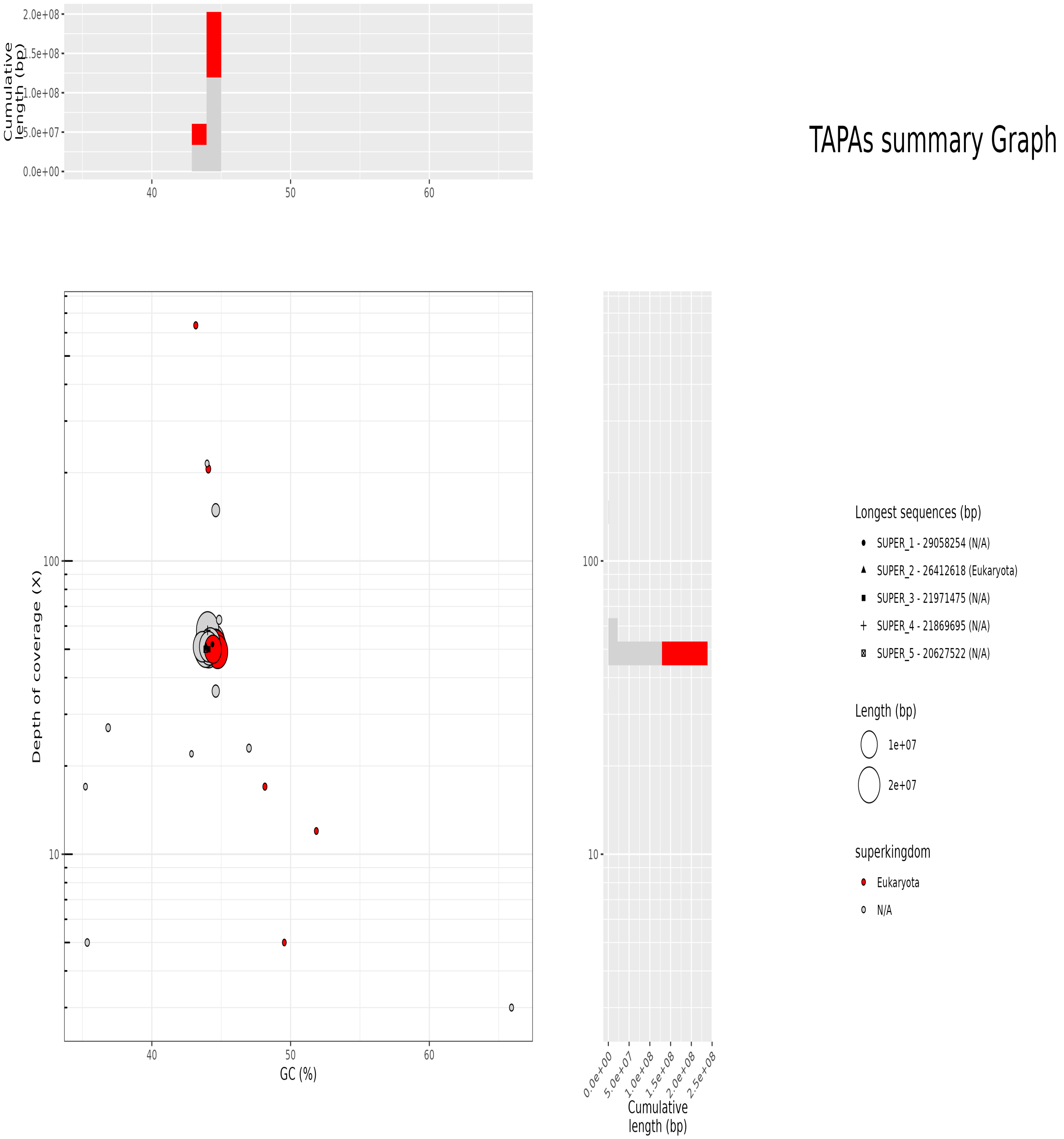


Distribution of k-mer counts per copy numbers found in asm



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	55	168

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: 2025-11-12 15:02:40 CET