ERGA Assembly Report

v24.10.15

Tags: Other[INVALID TAG]

TxID	77931
ToLID	**uyTelAmph1**
Species	Teleaulax amphioxeia
Class	Cryptophyceae
Order	Pyrenomonadales

Genome Traits	Expected	Observed
Haploid size (bp)	205,225,348	264,668,554
Haploid Number	5 (source: ancestor)	101
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 5.6.Q57

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

. BUSCO single copy value is less than 90% for collapsed
. More than 1000 gaps/Gbp for collapsed

Curator notes

. Interventions/Gb: 1093
. Contamination notes: ""
. Other observations: "The assembly of Teleaulax amphioxeia (uyTelAmph1) is based on 57X PacBio data and 1106X Arima Hi-C data. 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, 73 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 10.131 Mb (with the largest being 5.82 Mb). Additionally, 286 regions totaling 8.023 Mb (with the largest being 0.062 Mb) were identified as haplotypic duplications and removed. The mitochondrial and chloroplastic genomes were assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 6 haplotypic regions and 184 contaminant sequences were removed, totaling 0.271 Mb and 0.126 Mb, respectively (with the largest being 0.091 Mb and 5.26 Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. The nuclear genome is composed of 101 chromosomes, while chromosomes 102, 103, and 104

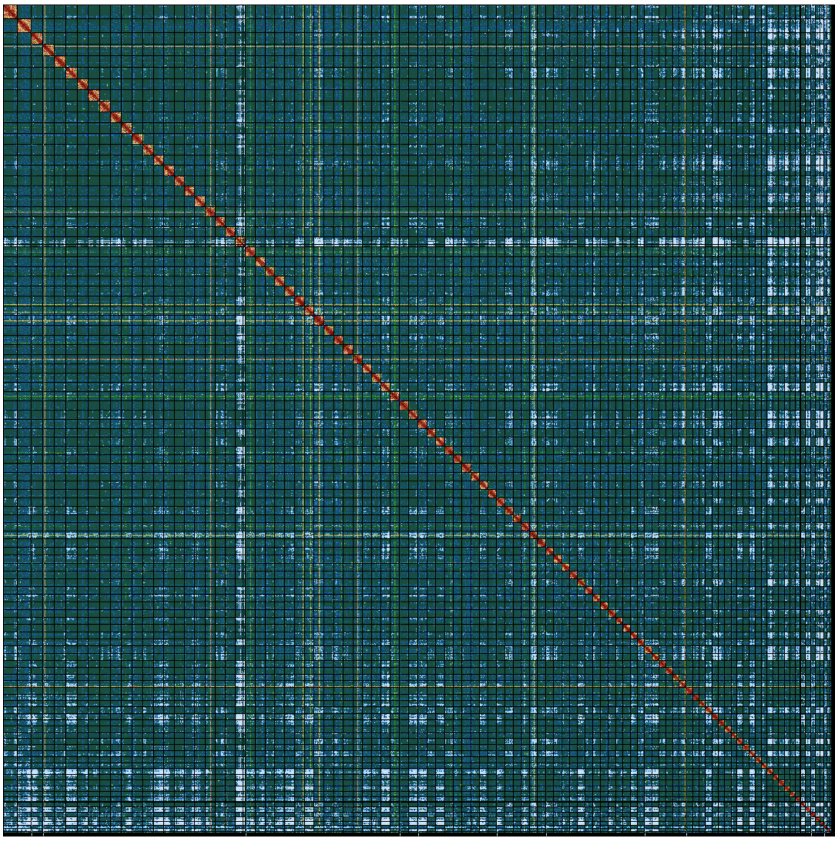
correspond to the nucleomorph genome. Note that 183 contigs that correspond to another eukaryotic species (Thraustochytrida order) were filtered out. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	270,385,418	264,668,554
GC %	54.34	54.28
Gaps/Gbp	4,600.84	5,089.38
Total gap bp	124,700	154,500
Scaffolds	469	135
Scaffold N50	2,793,894	2,821,441
Scaffold L50	42	41
Scaffold L90	89	85
Contigs	1,708	1,482
Contig N50	296,653	304,539
Contig L50	265	256
Contig L90	949	886
QV	57.0572	57.232
Kmer compl.	95.7313	95.6425
BUSCO sing.	52.7%	53.5%
BUSCO dupl.	3.9%	0.0%
BUSCO frag.	31.8%	32.6%
BUSCO miss.	11.6%	14.0%

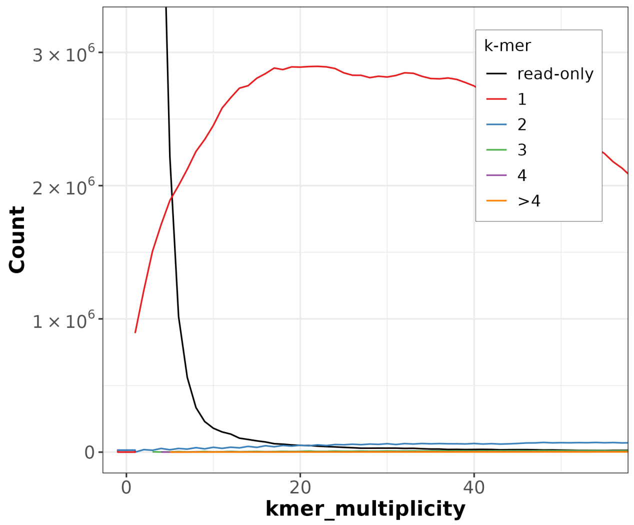
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: eukaryota_odb12 (genomes:456, BUSCOs:129)

HiC contact map of curated assembly

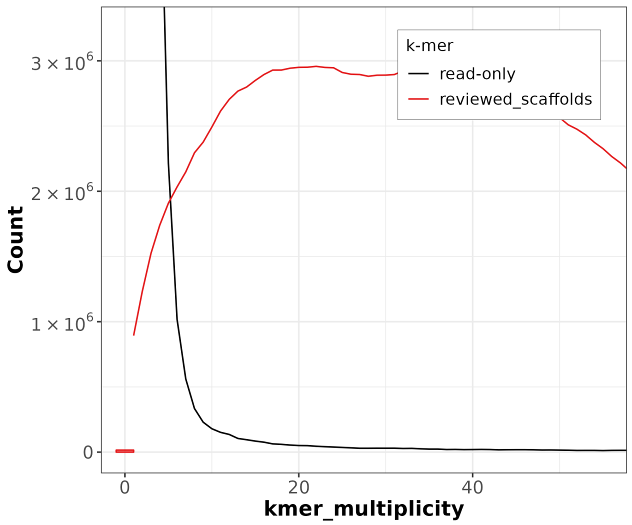


collapsed [LINK]

K-mer spectra of curated assembly

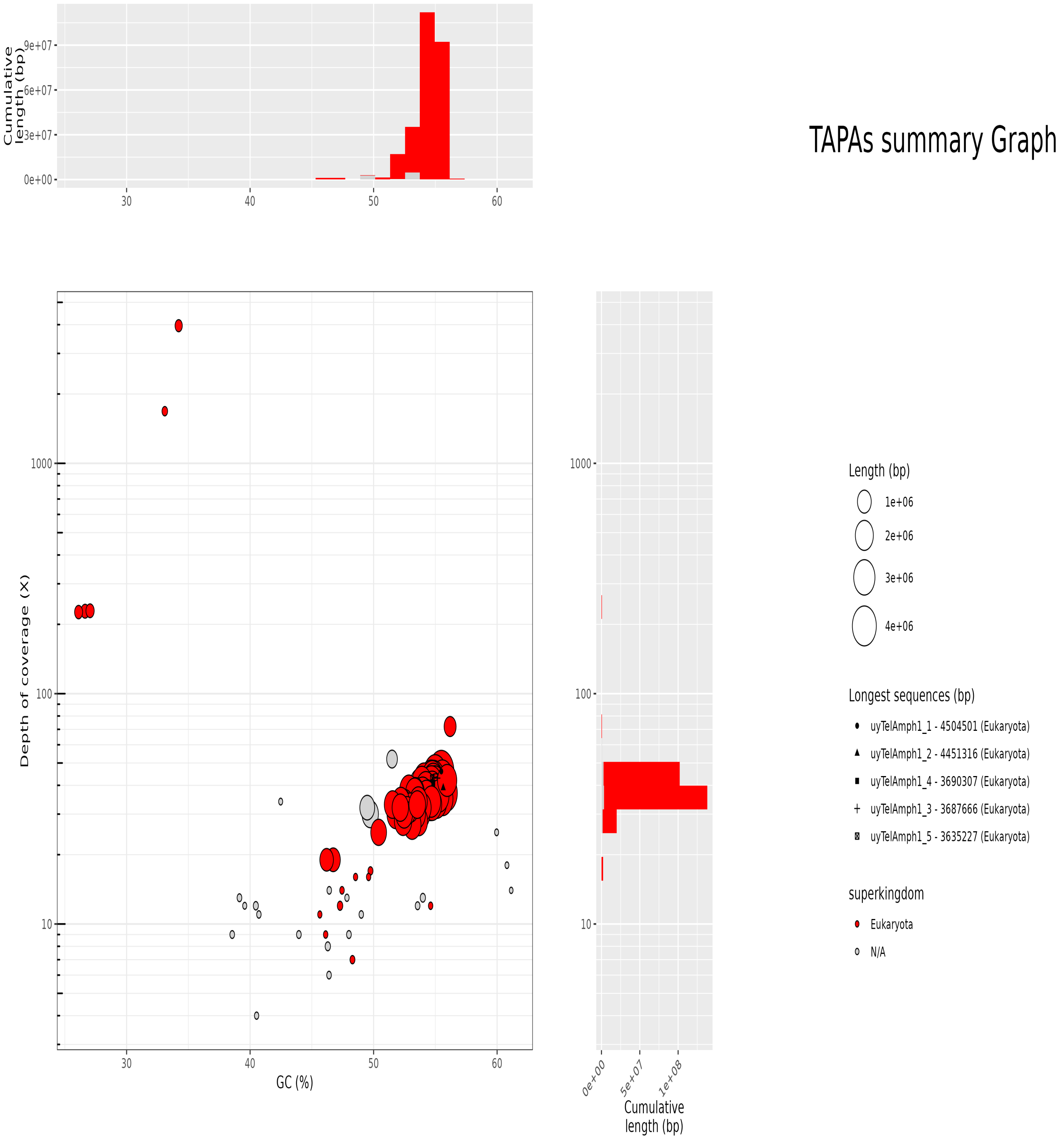


Distribution of k-mer counts per copy numbers found in asm



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	PACBIO Hifi	Arima
Coverage	56	1105

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: Emilie Teodori
Affiliation: Genoscope

Date and time: 2025-09-09 22:18:56 CEST