# ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

| | |
|---|---|
| TxID | 286151 |
| ToLID | **kaAplEleg1** |
| Species | Aplidium elegans |
| Class | Ascidiacea |
| Order | Aplousobranchia |

| Genome Traits | Expected | Observed |
|---|---|---|
| Haploid size (bp) | 592,467,604 | 559,751,178 |
| Haploid Number | 9 (source: ancestor) | 19 |
| Ploidy | 2 (source: ancestor) | 2 |
| Sample Sex | Unknown | Unknown |

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q48

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
. Assembly length loss > 3% for collapsed

### Curator notes

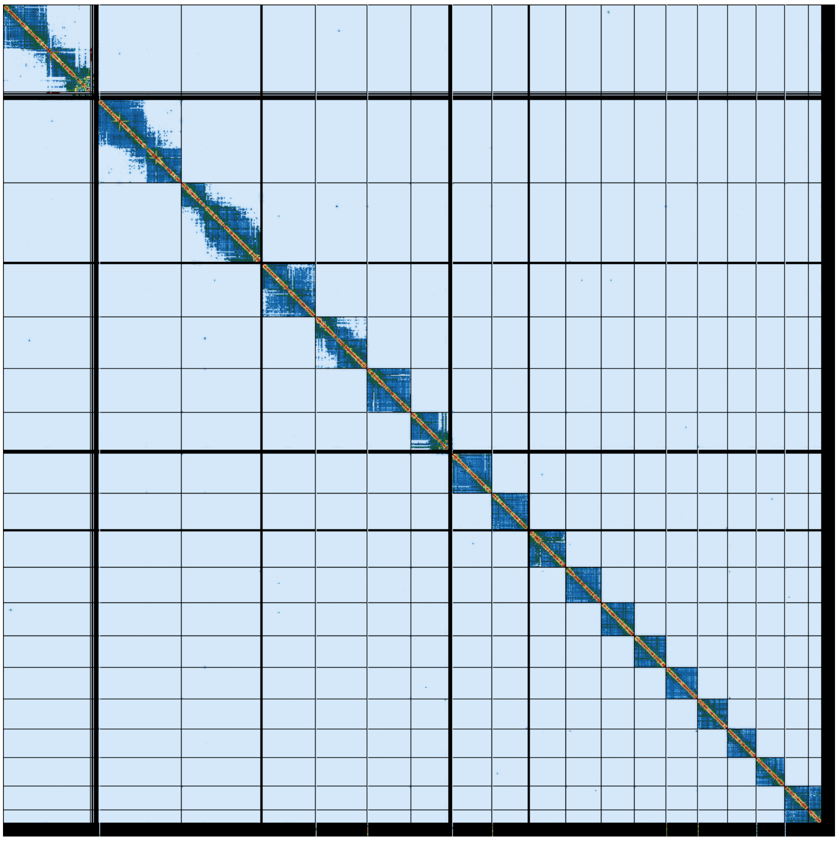
. Interventions/Gb: 824
. Contamination notes: ""
. Other observations: "The assembly of Aplidium elegans (kaAplEleg1) is based on 44X PacBio data and 197X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based sccaffolding with YaHS. In total, 11 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 3.995 Mb (with the largest being 3.651 Mb). Additionally, 918 regions totaling 90.806 Mb (with the largest being 1.095 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. During manual curation, 342 haplotypic regions were removed, totaling 126,719,920 pb (with the largest being 2,745,081 pb). 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 | 686,465,898 | 559,751,178 |
| GC % | 37.72 | 37.71 |
| Gaps/Gbp | 425.37 | 491.29 |
| Total gap bp | 29,200 | 37,200 |
| Scaffolds | 428 | 179 |
| Scaffold N50 | 27,203,573 | 26,586,232 |
| Scaffold L50 | 9 | 7 |
| Scaffold L90 | 59 | 17 |
| Contigs | 720 | 454 |
| Contig N50 | 2,686,587 | 3,267,206 |
| Contig L50 | 75 | 56 |
| Contig L90 | 309 | 186 |
| QV | 48.3513 | 48.6576 |
| Kmer compl. | 91.8213 | 88.3206 |
| BUSCO sing. | 73.4% | 91.5% |
| BUSCO dupl. | 23.1% | 4.2% |
| BUSCO frag. | 1.9% | 1.9% |
| BUSCO miss. | 1.6% | 2.4% |

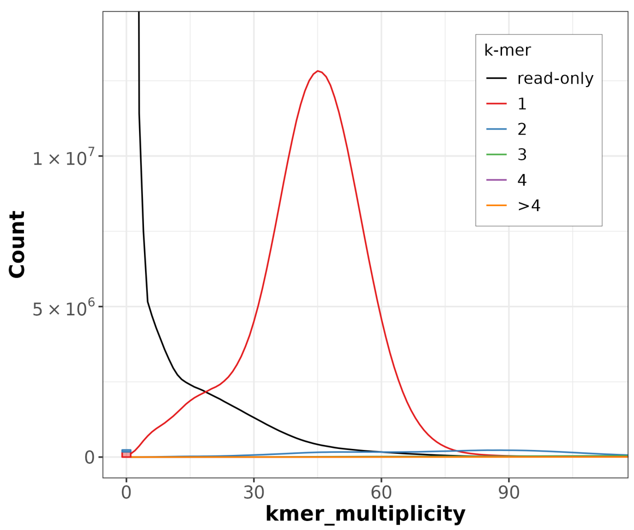
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

# HiC contact map of curated assembly

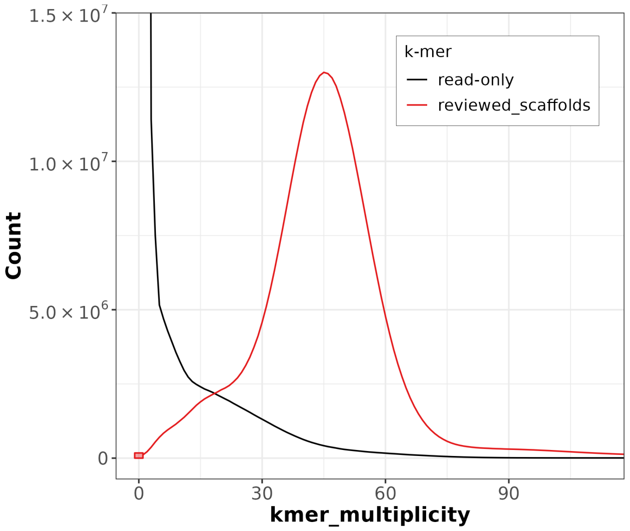


**collapsed** [LINK]

# K-mer spectra of curated assembly

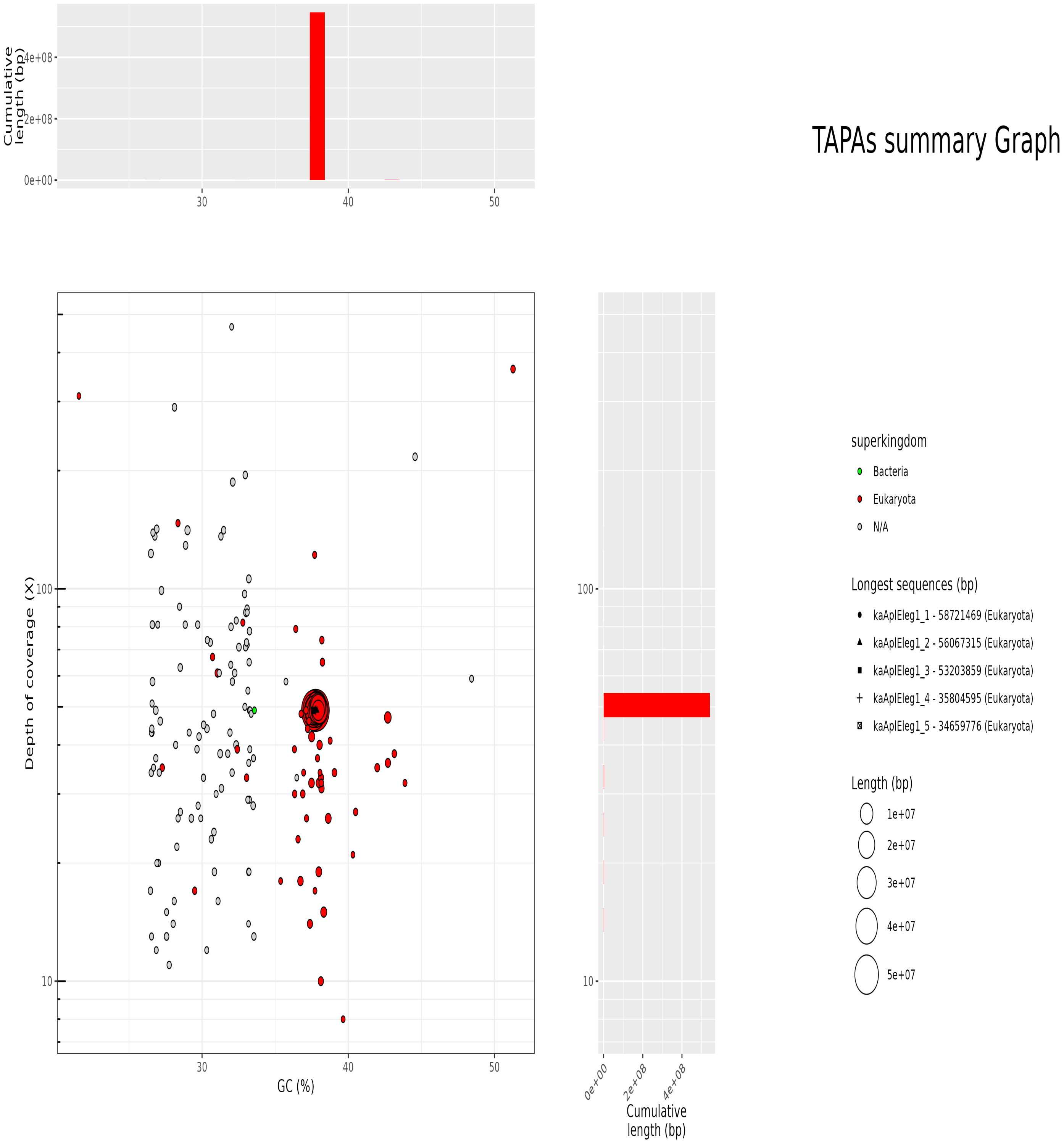


Distribution of k-mer counts per copy numbers found in asm



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 | 44 | 197 |

# 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: Lola Demirdjian
Affiliation: Genoscope

Date and time: 2025-12-20 01:36:41 CET