# ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

| | |
|---|---|
| TxID | 1774210 |
| ToLID | **kaLisPerf1** |
| Species | Lissoclinum perforatum |
| Class | Ascidiacea |
| Order | Aplousobranchia |

| Genome Traits | Expected | Observed |
|---|---|---|
| Haploid size (bp) | 386,688,480 | 382,996,208 |
| Haploid Number | 9 (source: ancestor) | 21 |
| Ploidy | 2 (source: ancestor) | 2 |
| Sample Sex | Unknown | Unknown |

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q49

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

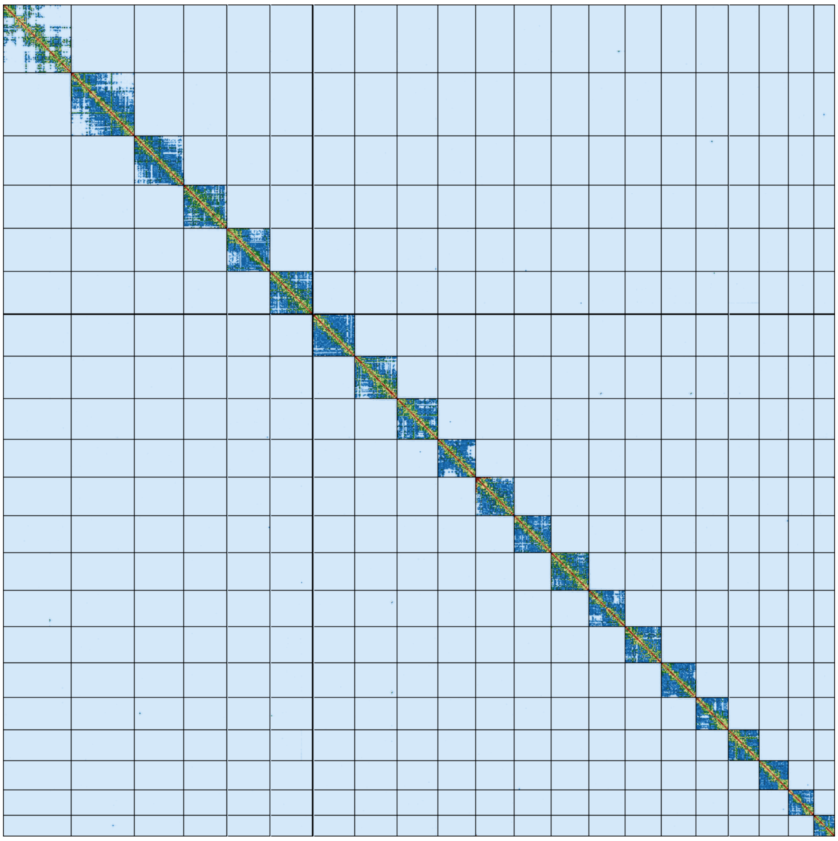
. Interventions/Gb: 90
. Contamination notes: ""
. Other observations: "The assembly of Lissoclinum perforatum (kaLisPerf1.1) is based on 78X PACBIO data and 174X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 20 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 5.749 Mb (with the largest being 2.112 Mb). Additionally, 76 regions totaling 35.559 Mb (with the largest being 3.994 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using ptGAUL. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 9 haplotypic regions and 2 contaminant sequences were removed, totaling 5.428 Mb and 0.053 Mb, respectively (with the largest being 2.104 Mb and 0.033 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 | 388,461,378 | 382,996,208 |
| GC % | 39.17 | 39.17 |
| Gaps/Gbp | 92.67 | 117.49 |
| Total gap bp | 3,900 | 6,700 |
| Scaffolds | 47 | 34 |
| Scaffold N50 | 18,761,558 | 18,761,558 |
| Scaffold L50 | 9 | 9 |
| Scaffold L90 | 19 | 18 |
| Contigs | 82 | 79 |
| Contig N50 | 10,875,095 | 10,502,235 |
| Contig L50 | 13 | 13 |
| Contig L90 | 37 | 36 |
| QV | 48.8554 | 49.1808 |
| Kmer compl. | 73.6749 | 73.24 |
| BUSCO sing. | 87.1% | 88.7% |
| BUSCO dupl. | 2.5% | 1.0% |
| BUSCO frag. | 2.8% | 2.7% |
| BUSCO miss. | 7.6% | 7.6% |

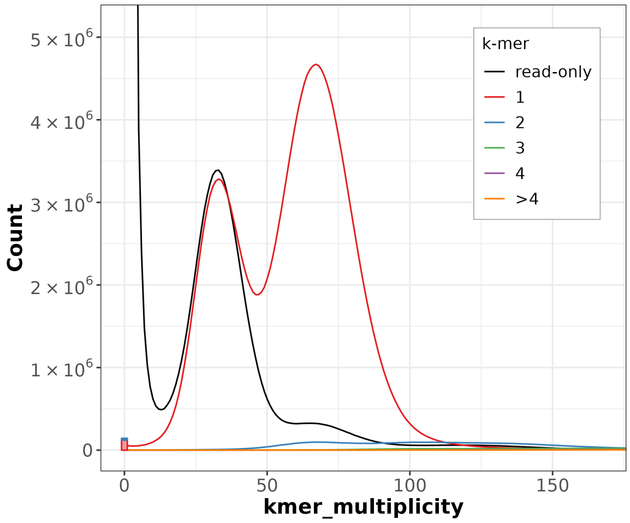
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

# HiC contact map of curated assembly

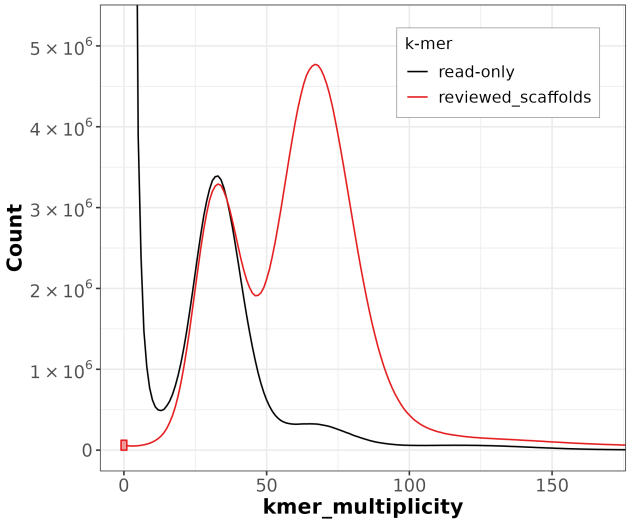


**collapsed** [LINK]

# K-mer spectra of curated assembly

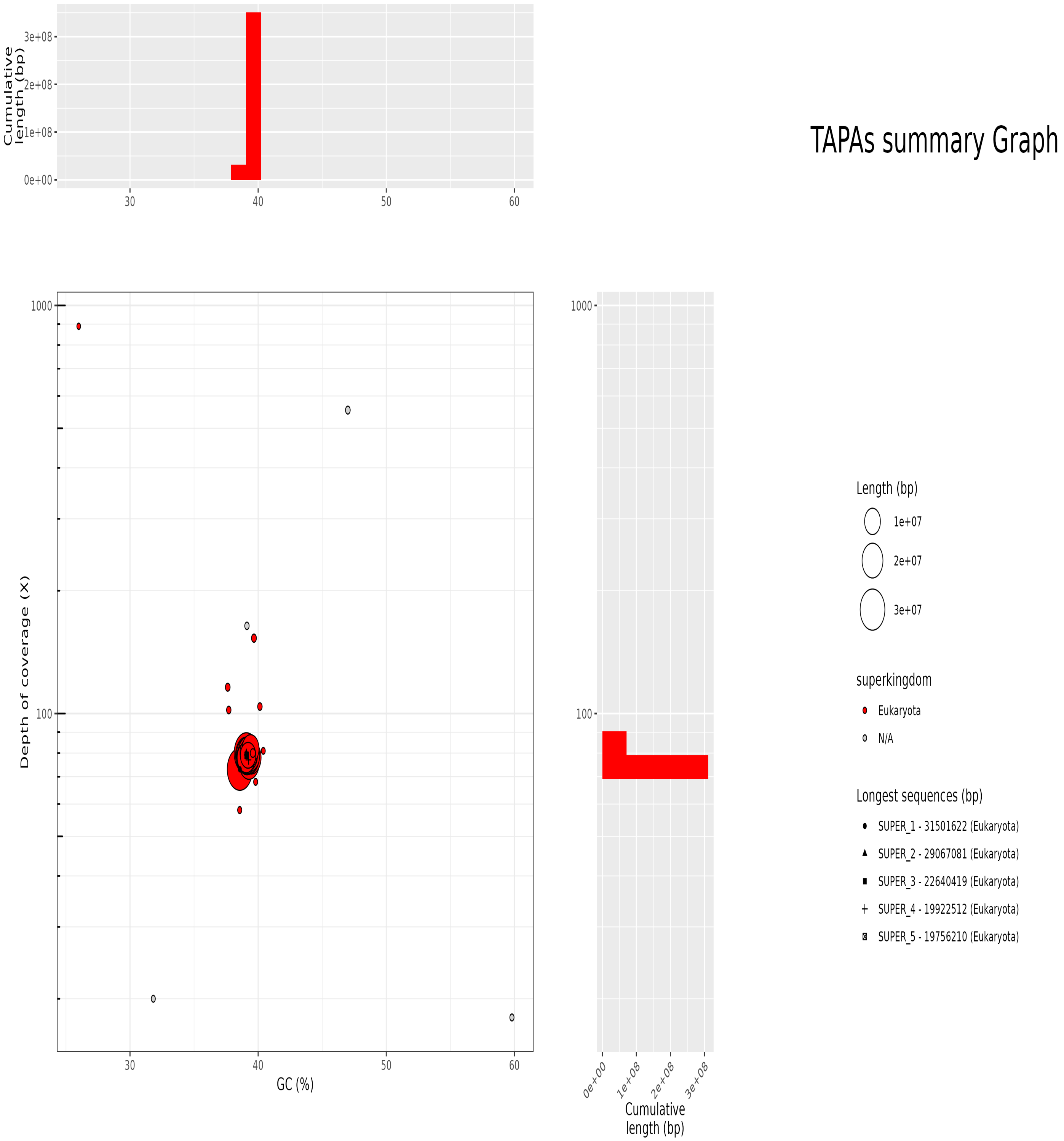


Distribution of k-mer counts per copy numbers found in asm



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 | 78 | 174 |

# 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: Adama Ndar
Affiliation: Genoscope

Date and time: 2025-11-23 04:06:40 CET