

# ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1234821
ToLID	<b>uoPlaVanh1</b>
Species	Plagiogrammopsis vanheurckii
Class	Mediophyceae
Order	Cymatosirales

Genome Traits	Expected	Observed
Haploid size (bp)	83,934,846	77,437,806
Haploid Number	19 (source: ancestor)	20
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.6.Q70

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
- . Assembly length loss > 3% for collapsed

### Curator notes

- . Interventions/Gb: 12
- . Contamination notes: ""
- . Other observations: "The assembly of *Plagiogrammopsis vanheurckii* (uoPlaVanh1) is based on 48X PacBio data and 152X 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 Hifiiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge\_dups, and Hi-C-based scaffolding with YaHS. In total, 195 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 18.449 Mb (with the largest being 4.512 Mb). Additionally, 18 regions totaling 5.161 Mb (with the largest being 1.506 Mb) were identified as haplotypic duplications and removed. Mitochondrial and chloroplastic genomes were assembled using OATK. During manual curation, 4 haplotypic regions were removed, totaling 9.205Mb (with the largest being 3.705Mb). Chromosome 1 is assembled as a single contig; however, based on the Hi-C map, it may in fact correspond to two chromosomes, with a potential split around position ~2.5 Mb. The unlocalized segment, which contains telomeric repeats at its distal end, would then

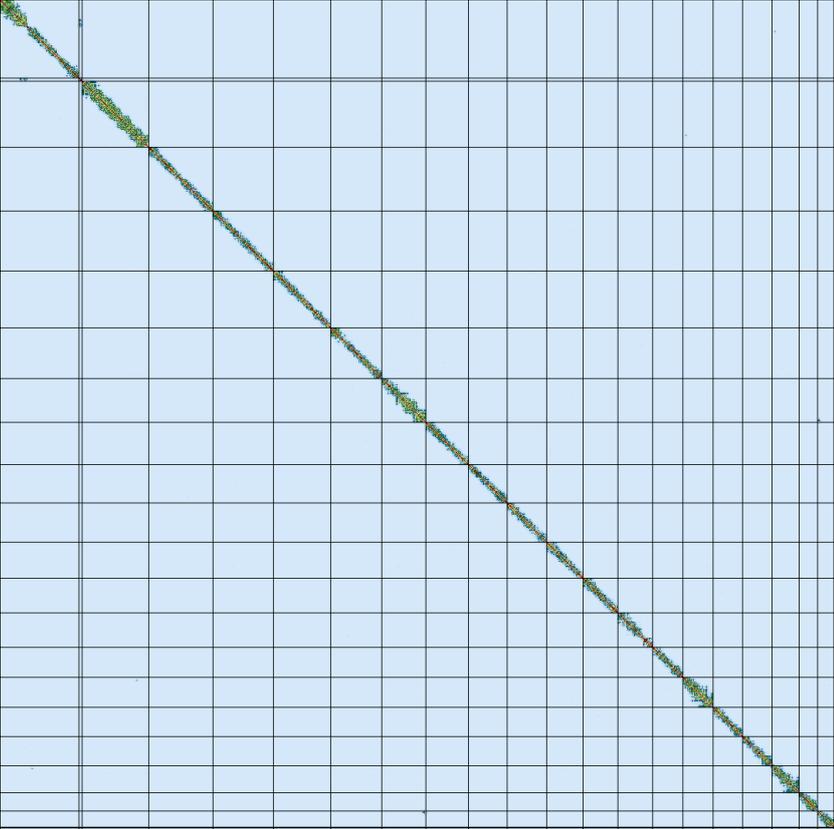
be placed at the extremity of the smaller chromosome (0-2.5 Mb). This contig was nevertheless retained as a single chromosome, as it contains no gaps and no telomeric repeats are detected in the region around 2.5Mb. Some chromosomes (chromosomes 2, 7, 15, and 18), as well as specific regions of chromosomes (0-2.5 Mb of chromosome 1), show approximately double sequencing coverage, which may indicate homozygous regions. 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	86,598,543	77,437,806
GC %	50.25	50.24
Gaps/Gbp	23.1	25.83
Total gap bp	200	200
Scaffolds	26	23
Scaffold N50	3,705,174	4,024,041
Scaffold L50	9	7
Scaffold L90	20	17
Contigs	28	25
Contig N50	3,443,668	3,617,401
Contig L50	10	8
Contig L90	22	19
QV	70.6151	70.1295
Kmer compl.	65.5126	62.0386
BUSCO sing.	76.6%	88.7%
BUSCO dupl.	13.8%	1.6%
BUSCO frag.	2.3%	2.4%
BUSCO miss.	7.4%	7.4%

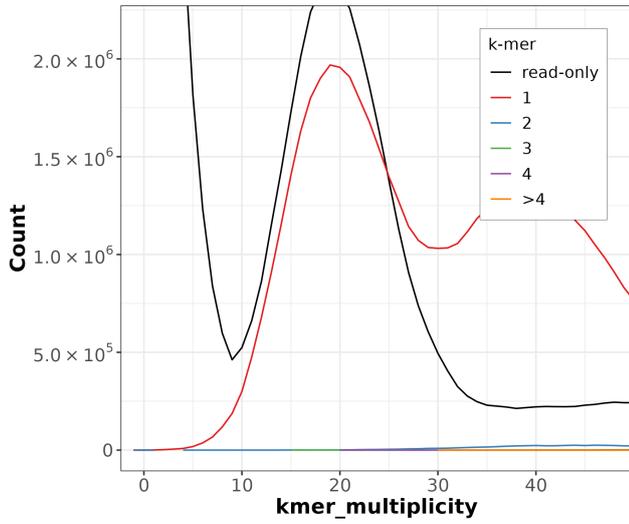
BUSCO: 6.0.0 (euk\_genome\_min, miniprot) / Lineage: bacillariophyta\_odb12 (genomes:8, BUSCOs:2944)

# HiC contact map of curated assembly

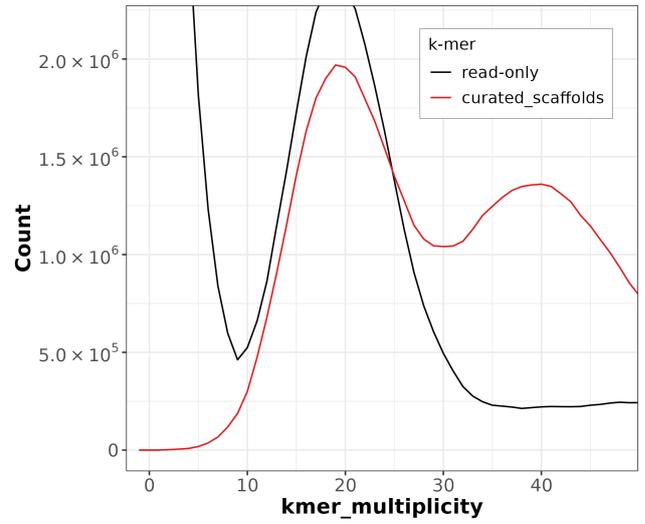


collapsed [\[LINK\]](#)

# K-mer spectra of curated assembly

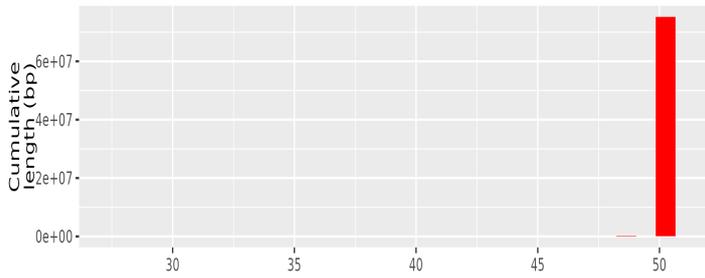


Distribution of k-mer counts per copy numbers found in asm

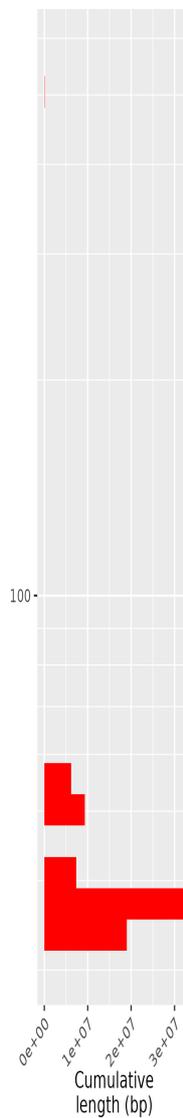
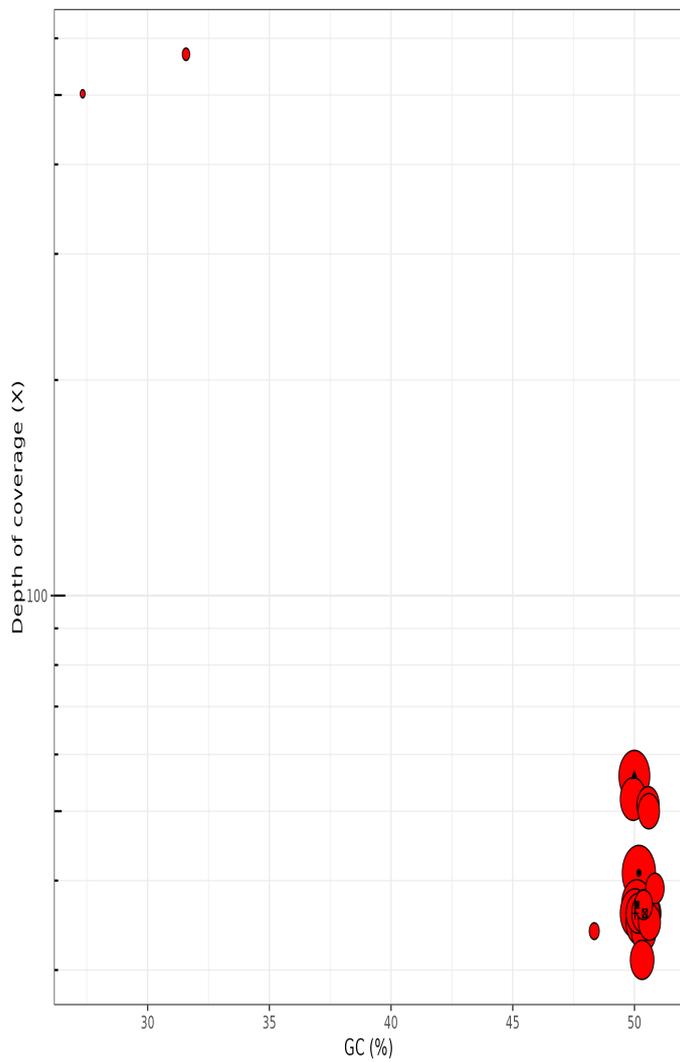


Distribution of k-mer counts coloured by their presence in reads/assemblies

# Post-curation contamination screening



## TAPAs summary Graph



Length (bp)

○ 2e+06

○ 4e+06

○ 6e+06

superkingdom

● Eukaryota

Longest sequences (bp)

● uoPlaVanh1\_1 - 7362269 (Eukaryota)

▲ uoPlaVanh1\_2 - 6164771 (Eukaryota)

■ uoPlaVanh1\_3 - 5958095 (Eukaryota)

+ uoPlaVanh1\_4 - 5588654 (Eukaryota)

⊠ uoPlaVanh1\_5 - 5258102 (Eukaryota)

**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	48	152

# 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: Jean-Marc Aury

Affiliation: Genoscope

Date and time: 2026-01-30 12:45:56 CET