

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	3451528
ToLID	uoPelSpec1
Species	Pelagomonas sp. RCC985
Class	Pelagophyceae
Order	Pelagomonadales

Genome Traits	Expected	Observed
Haploid size (bp)	557,804,726	31,472,317
Haploid Number	4 (source: ancestor)	6
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.6.Q46

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
- . Kmer completeness value is less than 90 for collapsed
- . Assembly length loss > 3% for collapsed

Curator notes

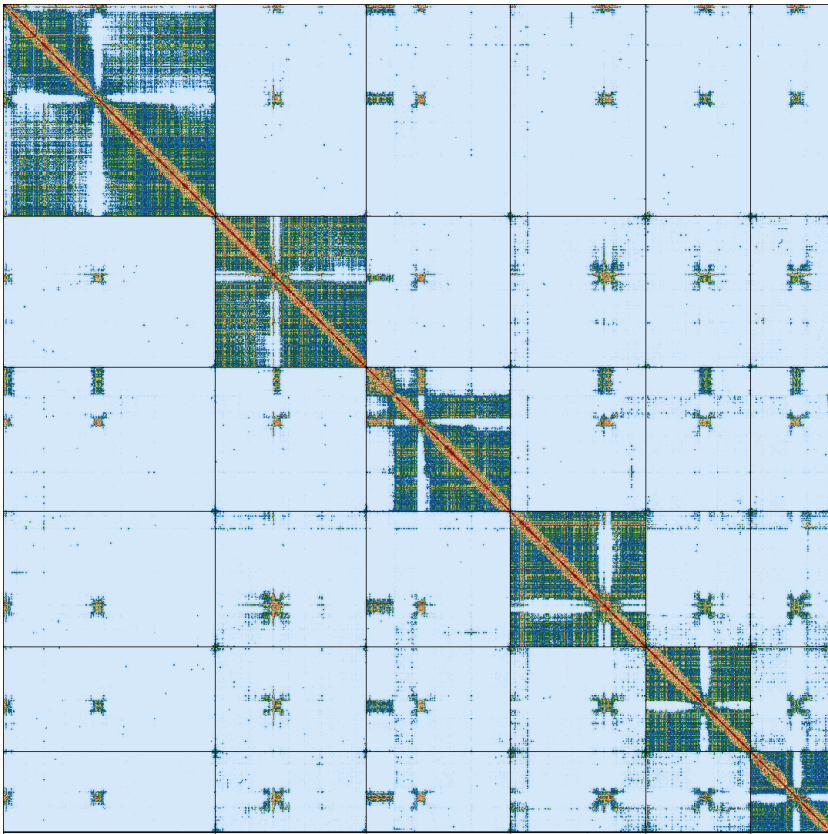
. Interventions/Gb: 0
. Contamination notes: ""
. Other observations: "The assembly of Pelagomonas sp.RCC985 (uoPelSpec1) is based on 587.94X PacBio data and 3860.73X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial PacBio+HiC assembly generation with Hifiiasm, removal of contaminant sequences using Context, and Hi-C-based scaffolding with YaHS of each haplotype. In total, 1092 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 70.266 Mb (with the largest being 8.581 Mb). The mitochondrial genome was assembled using Oatk. The chloroplast genome was assembled using Oatk. Finally, Both haplotypes were analyzed and manually improved using Pretext. During manual curation, 54 haplotypic regions were removed, totaling 1.098Mb (with the largest being 0.063Mb). The telomeric pattern AACCT was identified with TelFinder and used to generate the Pretext Telomeres tracks. 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	50,223,465	31,472,317
GC %	63.49	63.6
Gaps/Gbp	59.73	413.06
Total gap bp	300	2,600
Scaffolds	207	11
Scaffold N50	4,845,100	5,472,067
Scaffold L50	5	3
Scaffold L90	107	6
Contigs	210	24
Contig N50	1,938,000	5,234,000
Contig L50	6	3
Contig L90	110	10
QV	45.7786	46.8091
Kmer compl.	12.7471	12.6237
BUSCO sing.	55.8%	93.7%
BUSCO dupl.	39.9%	1.6%
BUSCO frag.	1.1%	1.1%
BUSCO miss.	3.2%	3.6%

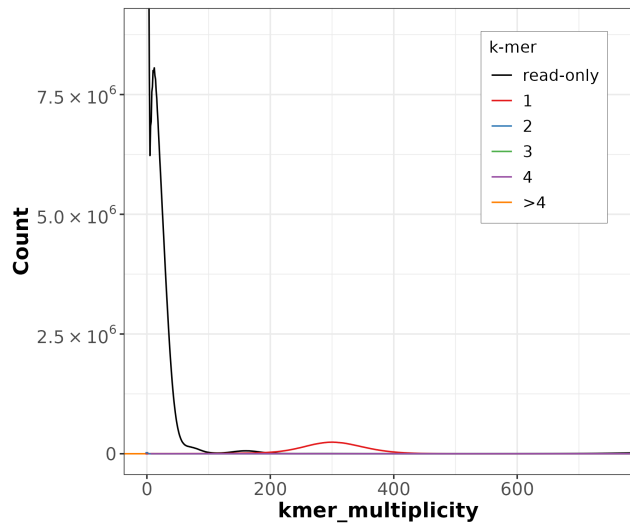
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: stramenopiles_odb12 (genomes:55, BUSCOs:697)

HiC contact map of curated assembly

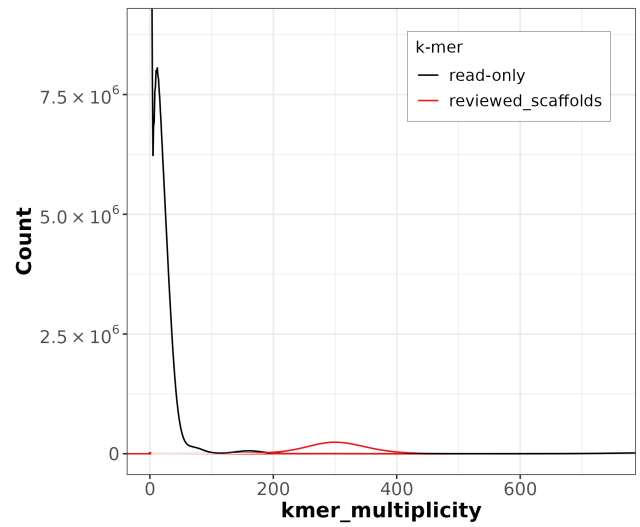


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

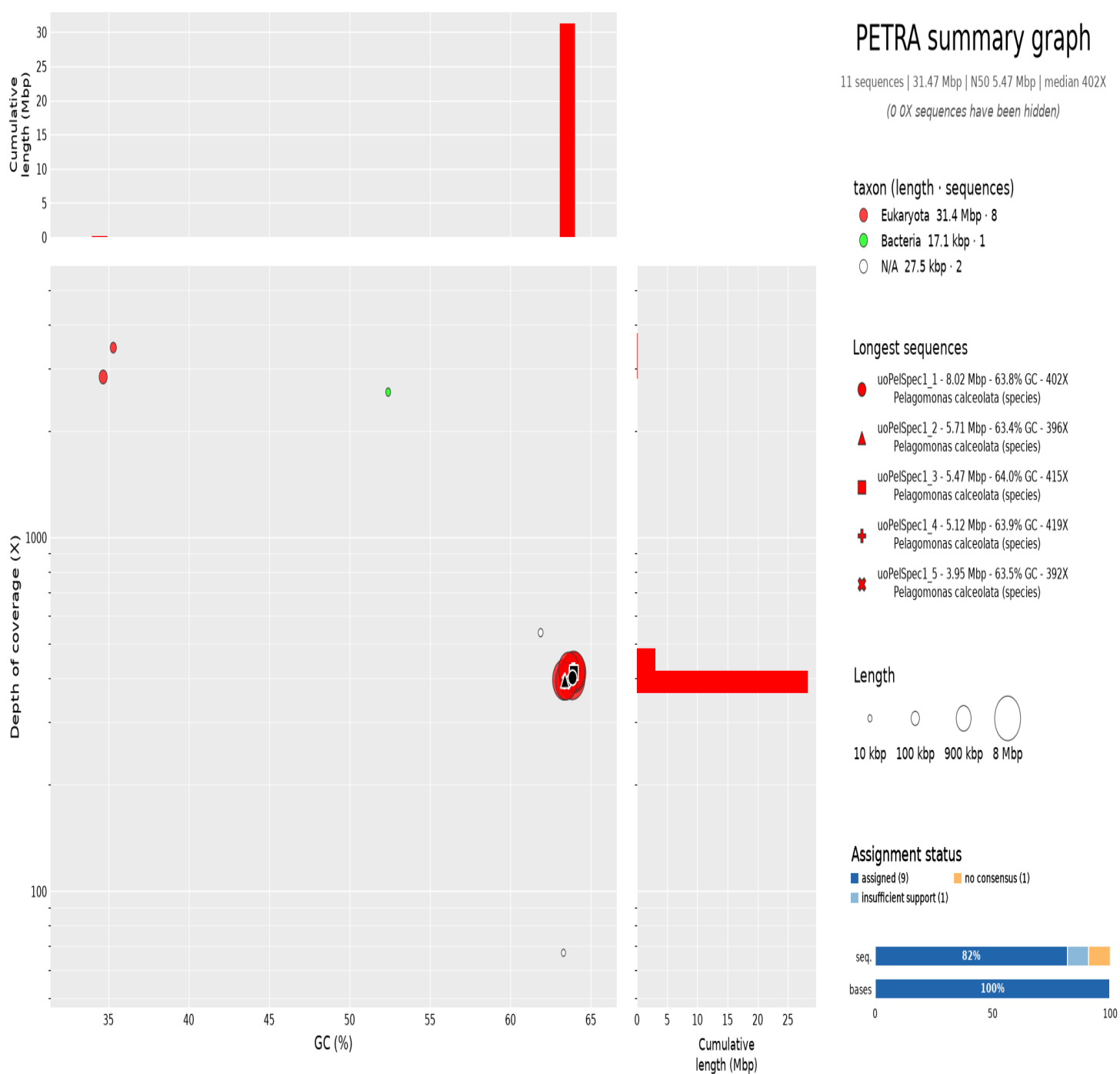


Distribution of k-mer counts per copy numbers found in asm



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	938	6160

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*: 0.2.4
 - |_ *key param*: NA
- **PretextView**
 - |_ *ver*: 1.0.7
 - |_ *key param*: NA

Submitter: Eleanore Lacoste

Affiliation: Genoscope

Date and time: 2026-10-08 02:12:37 CEST