

ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	2302900
ToLID	ucPyrSpeal
Species	Pyramimonas sp. RCC4209
Class	Pyramimonadophyceae
Order	Pyramimonadales

Genome Traits	Expected	Observed
Haploid size (bp)	336,581,351	132,013,778
Haploid Number	7 (source: ancestor)	14
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q34

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
- . QV value is less than 40 for collapsed
- . Kmer completeness value is less than 90 for collapsed
- . BUSCO single copy value is less than 90% for collapsed

Curator notes

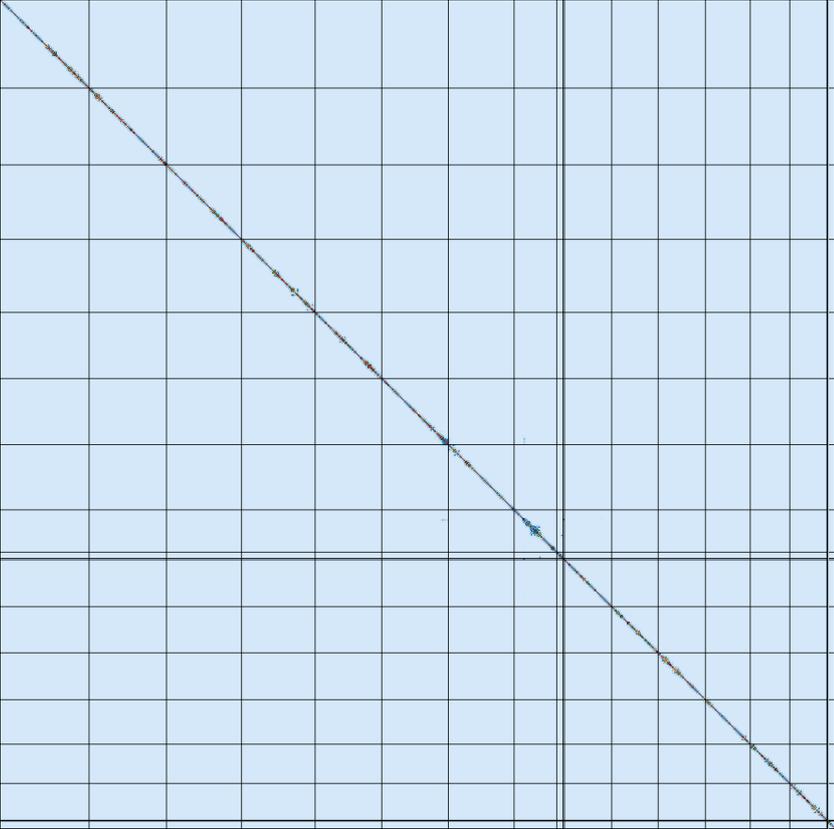
- . Interventions/Gb: 91
- . Contamination notes: ""
- . Other observations: "The assembly of Pyramimonas sp. RCC4209 (ucPyrSpeal) is based on 127X PacBio data and 97X Arima Hi-C data generated as part of the ATLASea programme (<https://www.atlasea.fr>). The assembly process included the following steps: 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, 184 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 50.291 Mb (with the largest being 4.481 Mb). Additionally, 81 regions totaling 2.246 Mb (with the largest being 0.128 Mb) were identified as haplotypic duplications and removed. Mitochondrial and chloroplastic genomes were assembled using OATK. During manual curation, 1 haplotypic region was removed, totaling 0.64Mb. 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	132,663,887	132,013,778
GC %	60.65	60.65
Gaps/Gbp	278.9	303
Total gap bp	3,700	4,700
Scaffolds	24	20
Scaffold N50	10,528,380	10,528,380
Scaffold L50	6	6
Scaffold L90	13	13
Contigs	61	60
Contig N50	4,011,752	4,011,752
Contig L50	11	11
Contig L90	32	32
QV	34.7734	34.7756
Kmer compl.	75.3684	75.3997
BUSCO sing.	88.3%	88.6%
BUSCO dupl.	2.8%	2.4%
BUSCO frag.	1.4%	1.5%
BUSCO miss.	7.5%	7.5%

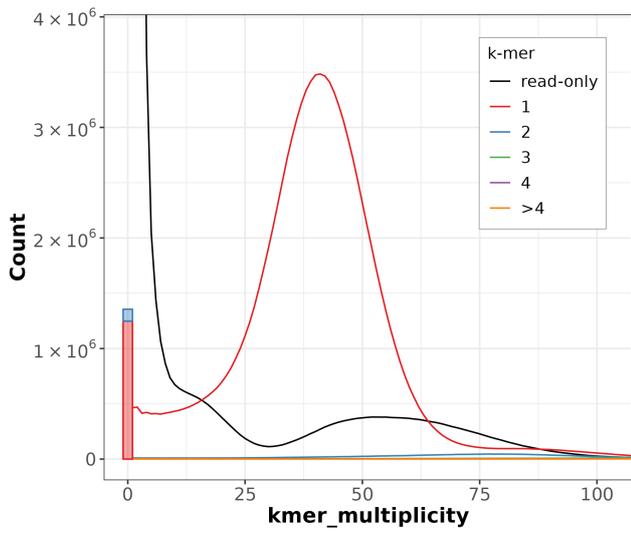
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: chlorophyta_odb12 (genomes:39, BUSCOs:1523)

HiC contact map of curated assembly

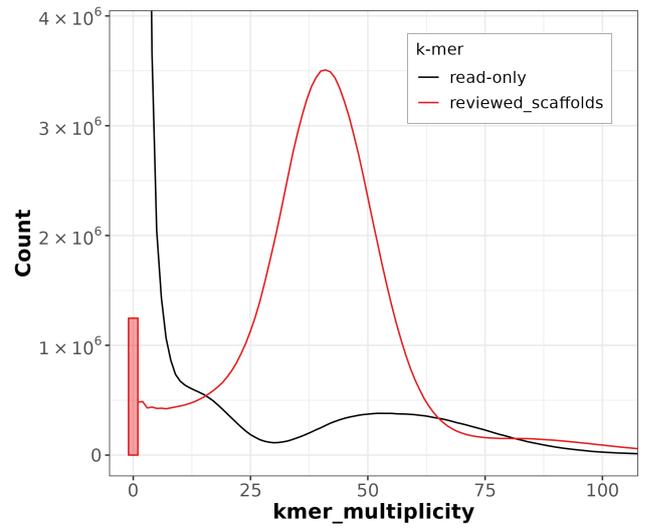


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

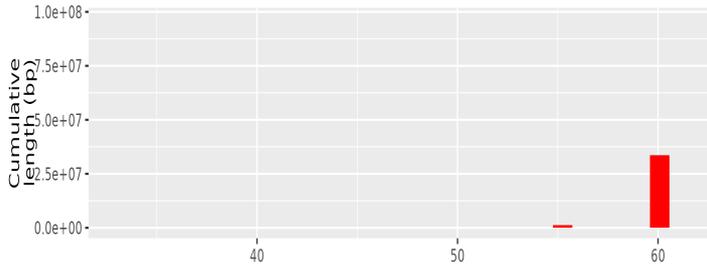


Distribution of k-mer counts per copy numbers found in asm

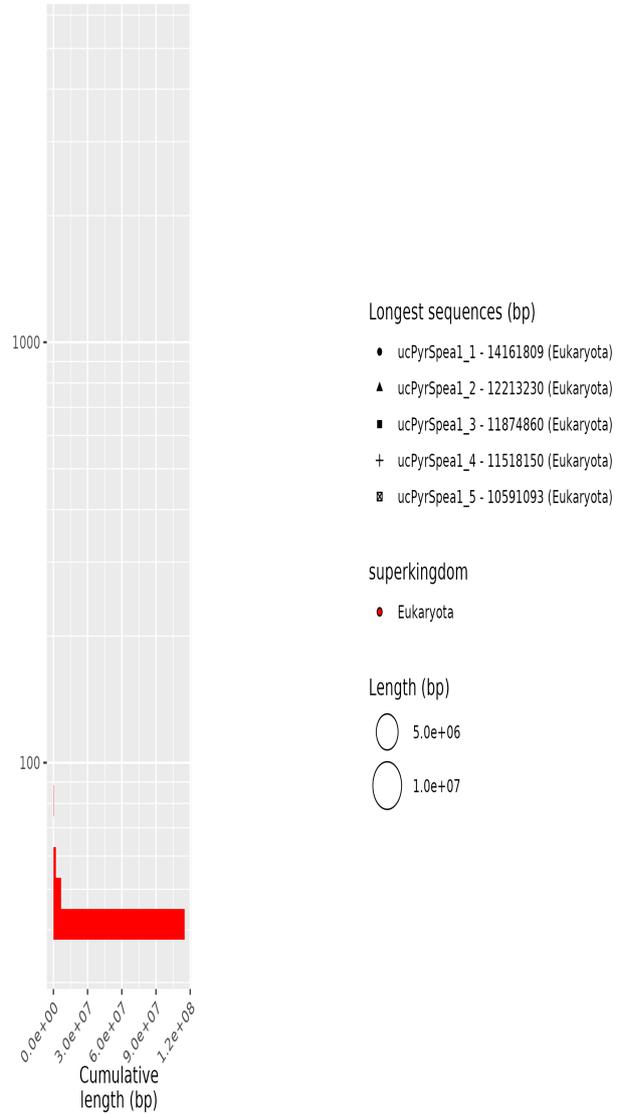
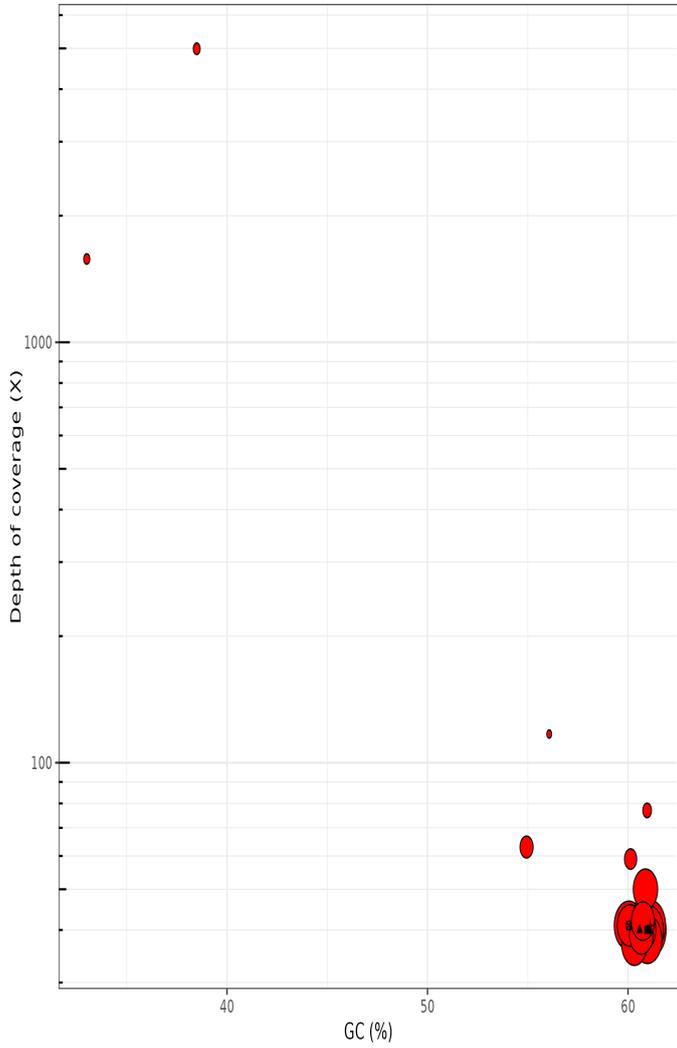


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	49	37

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: Arnaud Couloux

Affiliation: Genoscope

Date and time: 2026-02-05 08:28:36 CET