ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1764261
ToLID	**ucChlSped1**
Species	Chlamydomonas sp. RCC443
Class	Chlorophyceae
Order	Chlamydomonadales

Genome Traits	Expected	Observed
Haploid size (bp)	112,288,276	100,364,539
Haploid Number	7 (source: ancestor)	31
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.6.Q38

The following metrics were automatically flagged as below EBP recommended standards or different from 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

Curator notes

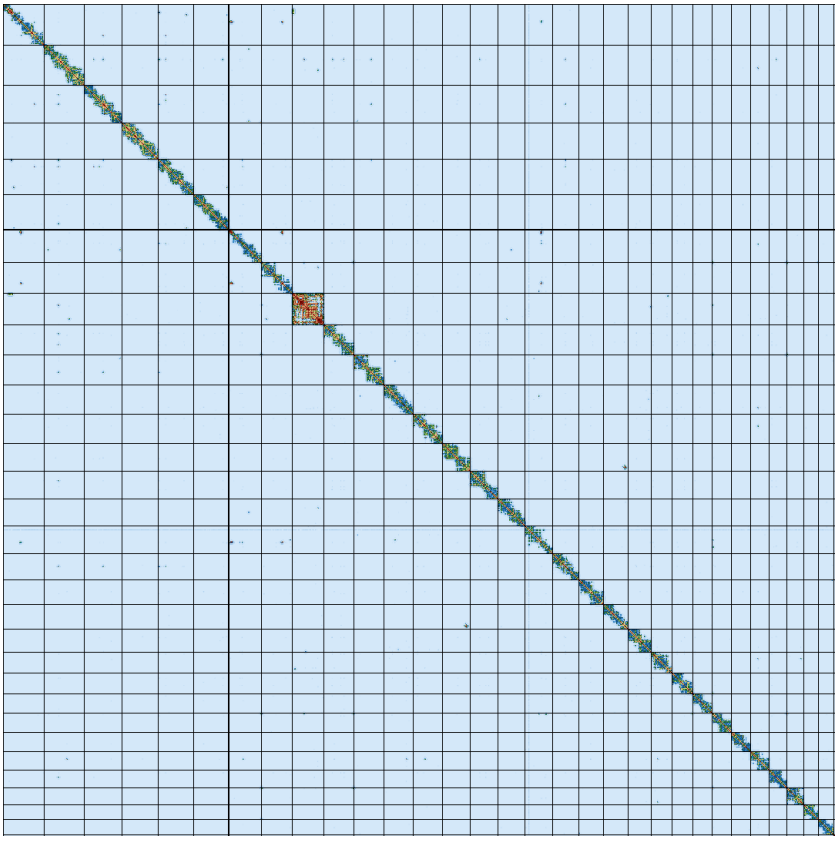
. Interventions/Gb: 70
. Contamination notes: ""
. Other observations: "The assembly of Chlamydomonas sp. RCC443 (ucChlSped1) is based on 32X PacBio data and 258X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 25 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 13.823 Mb (with the largest being 3.672 Mb). Additionally, 19 regions totaling 1.036 Mb (with the largest being 0.138 Mb) were identified as haplotypic duplications and removed. Mitochondrial and chloroplastic genomes were assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. 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	100,246,064	100,364,539
GC %	63.29	63.26
Gaps/Gbp	349.14	378.62
Total gap bp	3,500	4,300
Scaffolds	37	35
Scaffold N50	3,381,640	3,490,998
Scaffold L50	13	13
Scaffold L90	27	27
Contigs	72	73
Contig N50	2,200,160	2,200,160
Contig L50	16	16
Contig L90	44	44
QV	38.5434	38.5483
Kmer compl.	84.1489	84.2515
BUSCO sing.	90.3%	92.4%
BUSCO dupl.	4.9%	2.8%
BUSCO frag.	0.7%	0.7%
BUSCO miss.	4.1%	4.1%

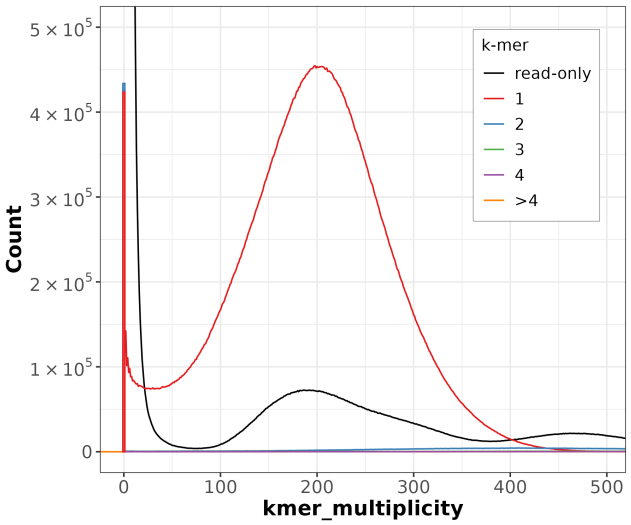
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: chlorophyceae_odb12 (genomes:18, BUSCOs:2432)
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: chlorophyceae_odb12 (genomes:18, BUSCOs:2432)

HiC contact map of curated assembly

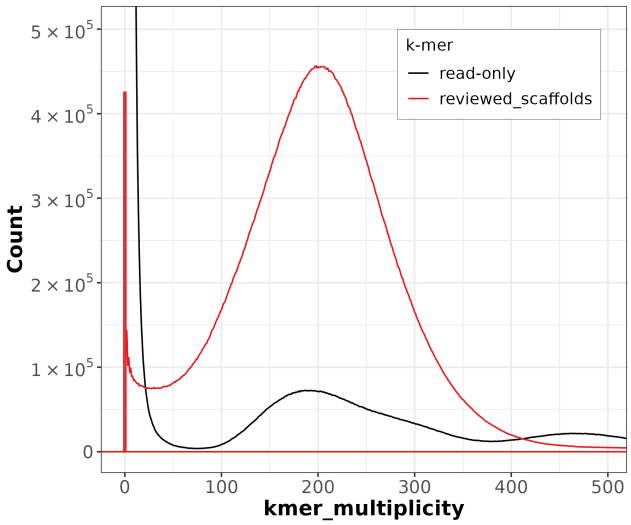


collapsed [LINK]

K-mer spectra of curated assembly

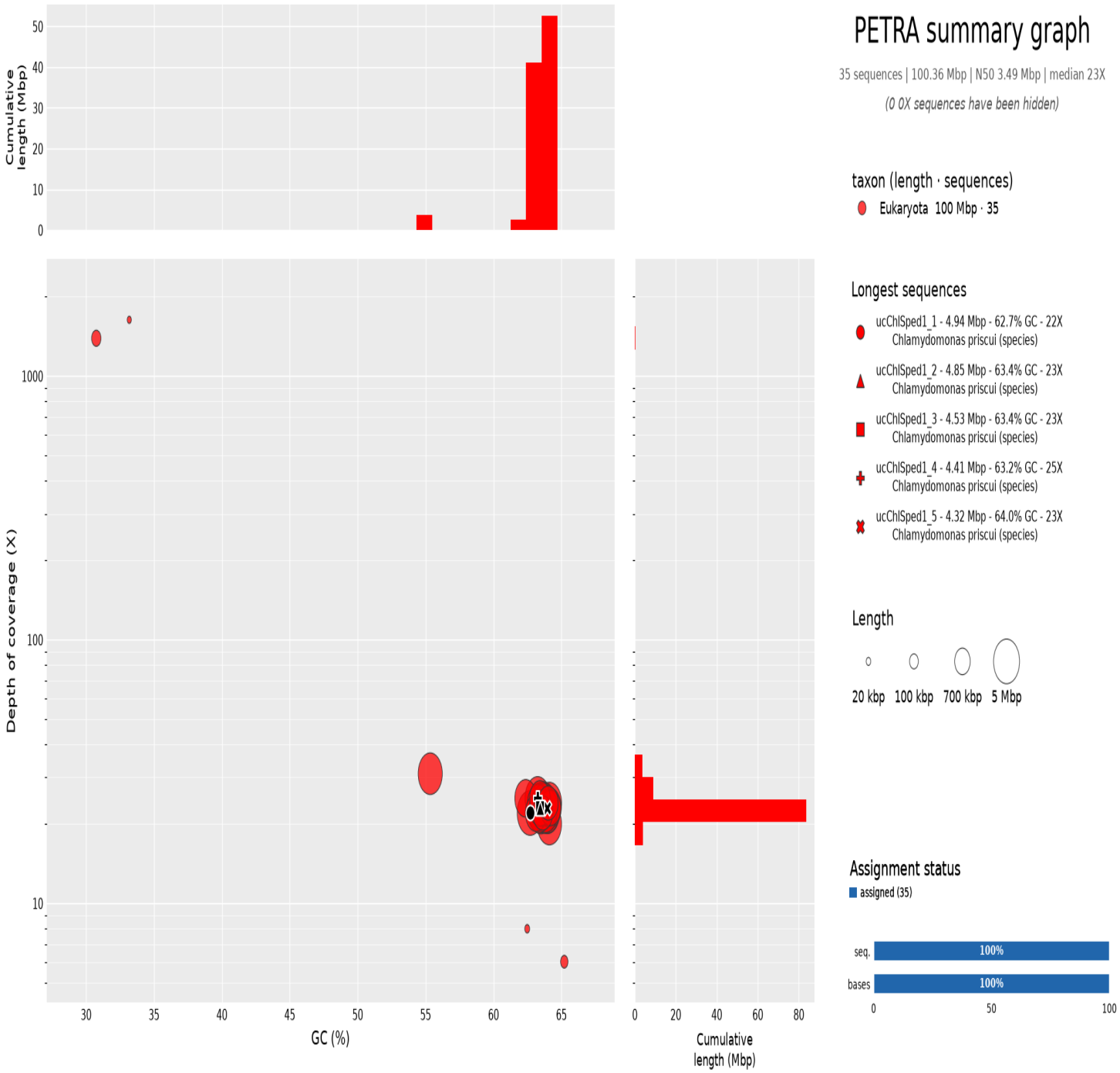


Distribution of k-mer counts per copy numbers found in asm



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	31	257

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: Mohamed Meftah Saoues
Affiliation: Genoscope

Date and time: 2026-10-06 10:23:15 CEST