ERGA Assembly Report
v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	2528241
ToLID	**ucTetSpeb1**
Species	Tetraselmis sp. RCC6047
Class	Chlorodendrophyceae
Order	Chlorodendrales

Genome Traits	Expected	Observed
Haploid size (bp)	448,352,665	337,868,294
Haploid Number	7 (source: ancestor)	12
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 5.7.Q42

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
. BUSCO single copy value is less than 90% for collapsed
. Assembly length loss > 3% for collapsed
. More than 1000 gaps/Gbp for collapsed
. Not 90% of assembly in chromosomes for collapsed

Curator notes

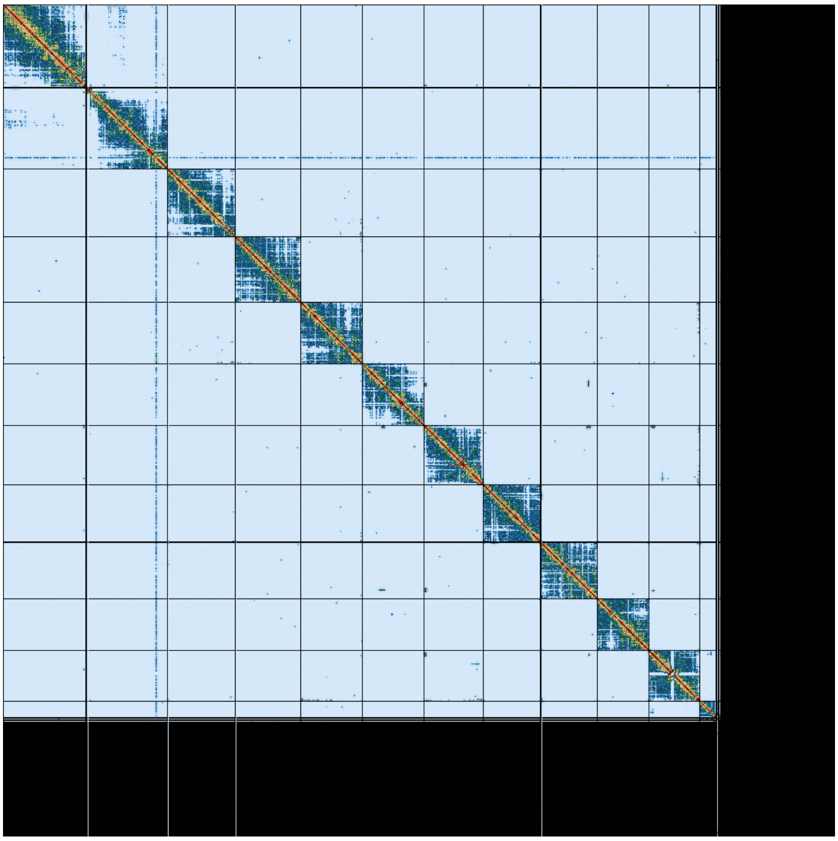
. Interventions/Gb: 637
. Contamination notes: ""
. Other observations: "The assembly of Tetraselmis sp. RCC6047 (ucTetSpeb1) is based on 138X PacBio data and 254X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio+HiC assembly generation with Hifiasm, removal of contaminant sequences using Context, and Hi-C-based scaffolding with YaHS of each haplotype. In total, 153 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 3.961 Mb (with the largest being 0.442 Mb). Mitochondrial and chloroplastic genomes were assembled using OATK. During manual curation, 22 haplotypic regions were removed, totaling 3.9Mb (with the largest being 0.82Mb).. 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	413,998,823	337,868,294
GC %	54.85	56.74
Gaps/Gbp	1,958.94	3,060.36
Total gap bp	81,100	119,200
Scaffolds	1,854	2,054
Scaffold N50	25,642,300	24,750,372
Scaffold L50	7	6
Scaffold L90	343	439
Contigs	2,665	3,088
Contig N50	616,354	404,845
Contig L50	193	238
Contig L90	1,137	1,472
QV	41.5578	42.0826
Kmer compl.	82.2861	73.1689
BUSCO sing.	81.5%	87.5%
BUSCO dupl.	12.7%	3.5%
BUSCO frag.	1.2%	2.0%
BUSCO miss.	4.5%	7.0%

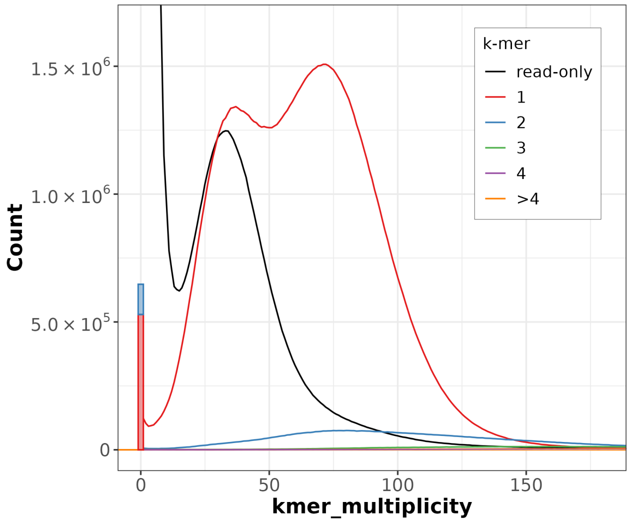
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: chlorophyta_odb12 (genomes:39, BUSCOs:1523)

HiC contact map of curated assembly

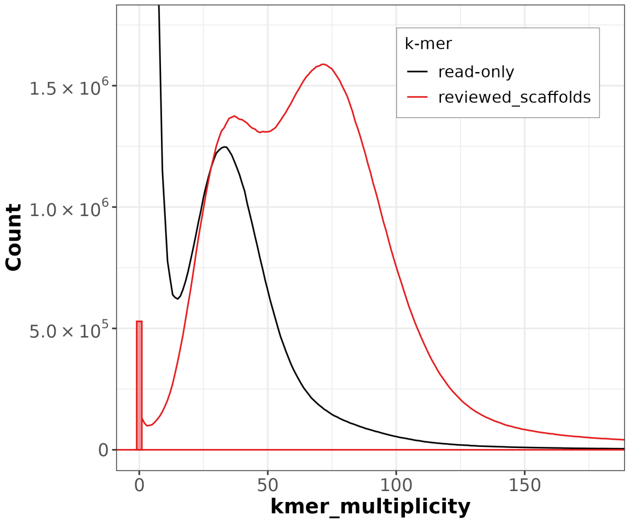


collapsed [LINK]

K-mer spectra of curated assembly

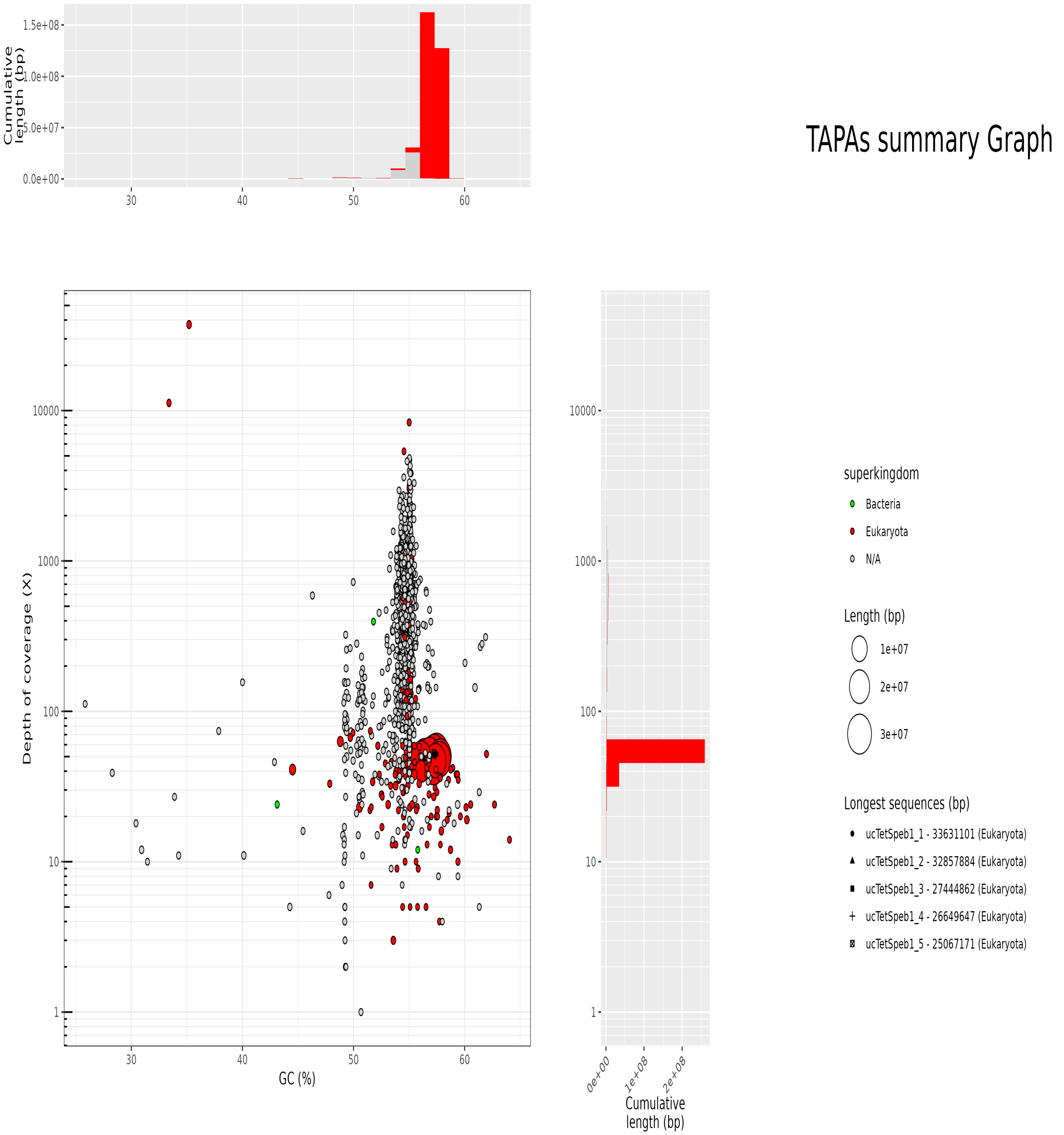


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	112	206

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: Emilie Teodori
Affiliation: Genoscope

Date and time: 2026-02-12 23:45:52 CET