ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	3451518
ToLID	**ucChlSpeb1**
Species	Chloropicon sp. RCC7702
Class	Chloropicophyceae
Order	Chloropicales

Genome Traits	Expected	Observed
Haploid size (bp)	26,181,793	22,185,955
Haploid Number	7 (source: ancestor)	23
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 Number is different from Expected

. Kmer completeness value is less than 90 for collapsed

Curator notes

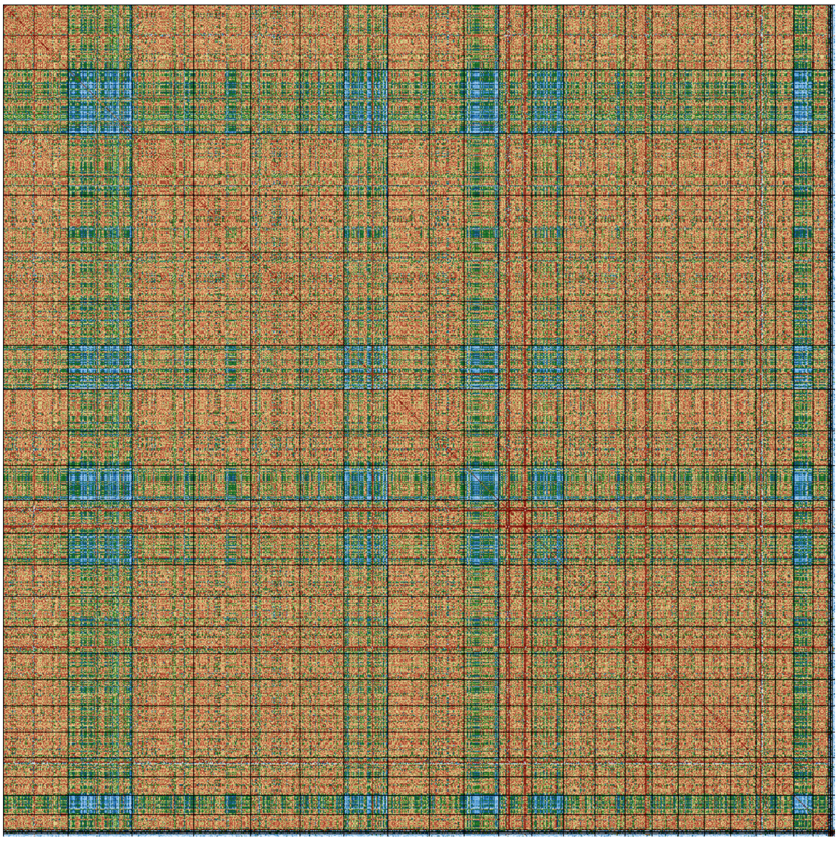
. Interventions/Gb: 0
. Contamination notes: ""
. Other observations: "The assembly of Chloropicon sp. RCC7702 (ucChlSpeb1) is based on 123X PacBio data and 127X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 438 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 53.606 Mb (with the largest being 4.728 Mb). Additionally, 1 region totaling 0.065 Mb (with the largest being 0.065 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was 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	22,083,409	22,185,955
GC %	47.94	47.89
Gaps/Gbp	0	0
Total gap bp	0	0
Scaffolds	25	27
Scaffold N50	1,096,114	1,096,114
Scaffold L50	8	8
Scaffold L90	19	19
Contigs	25	27
Contig N50	1,096,114	1,096,114
Contig L50	8	8
Contig L90	19	19
QV	46.622	46.6357
Kmer compl.	23.7146	23.7496
BUSCO sing.	92.7%	92.7%
BUSCO dupl.	1.4%	1.4%
BUSCO frag.	0.9%	0.9%
BUSCO miss.	5.1%	5.1%

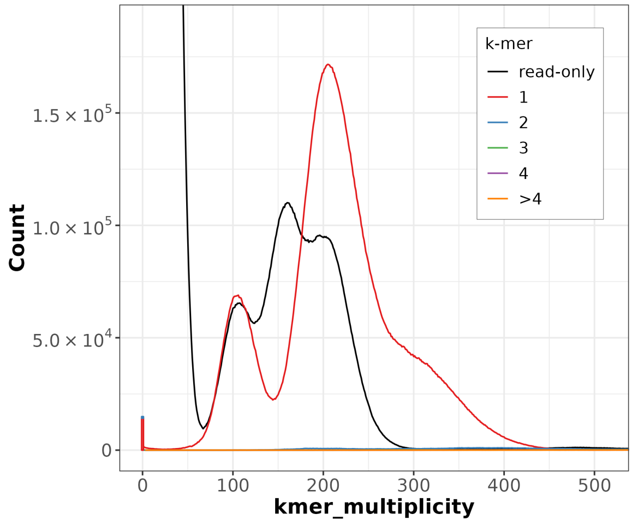
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: chlorophyta_odb12 (genomes:39, BUSCOs:1523)

HiC contact map of curated assembly

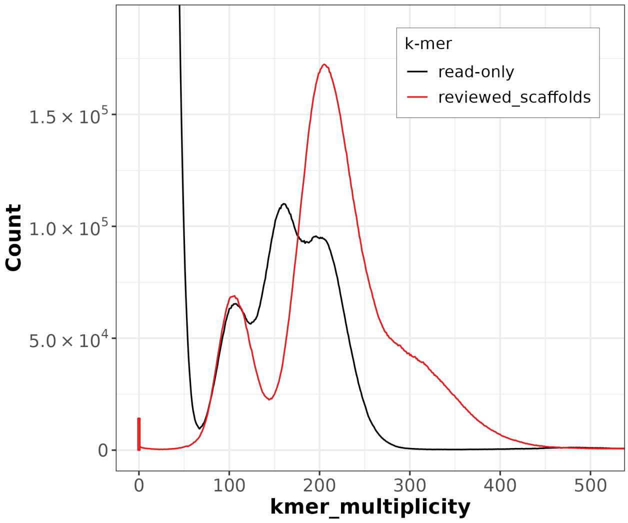


collapsed [LINK]

K-mer spectra of curated assembly

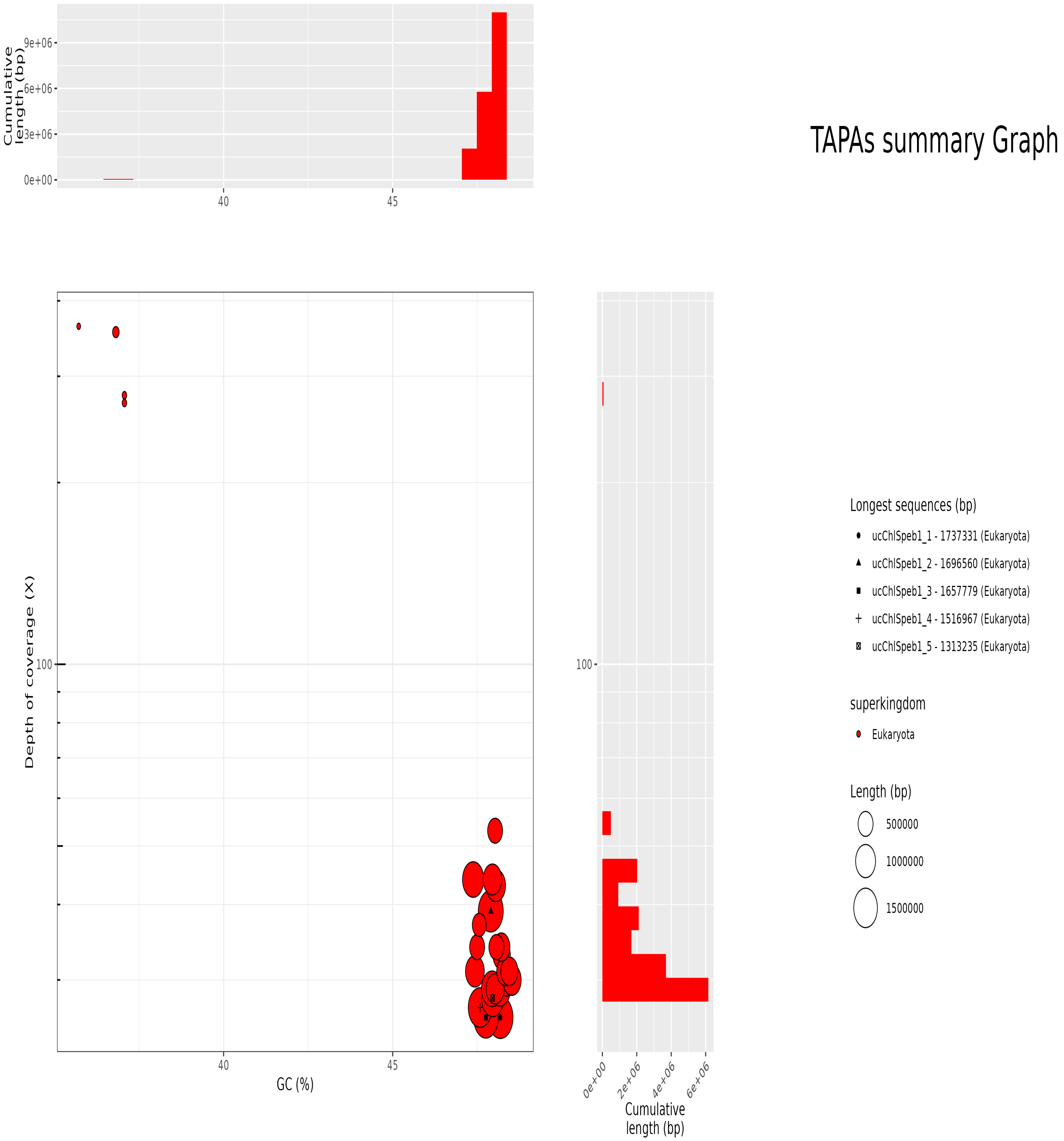


Distribution of k-mer counts per copy numbers found in asm



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	123	127

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: Lola Demirdjian
Affiliation: Genoscope

Date and time: 2025-12-03 05:12:12 CET