ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1461544
ToLID	**ucChlRosc1**
Species	Chloropicon roscoffensis
Class	Chloropicophyceae
Order	Chloropicales

Genome Traits	Expected	Observed
Haploid size (bp)	17,373,551	17,268,302
Haploid Number	7 (source: ancestor)	19
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

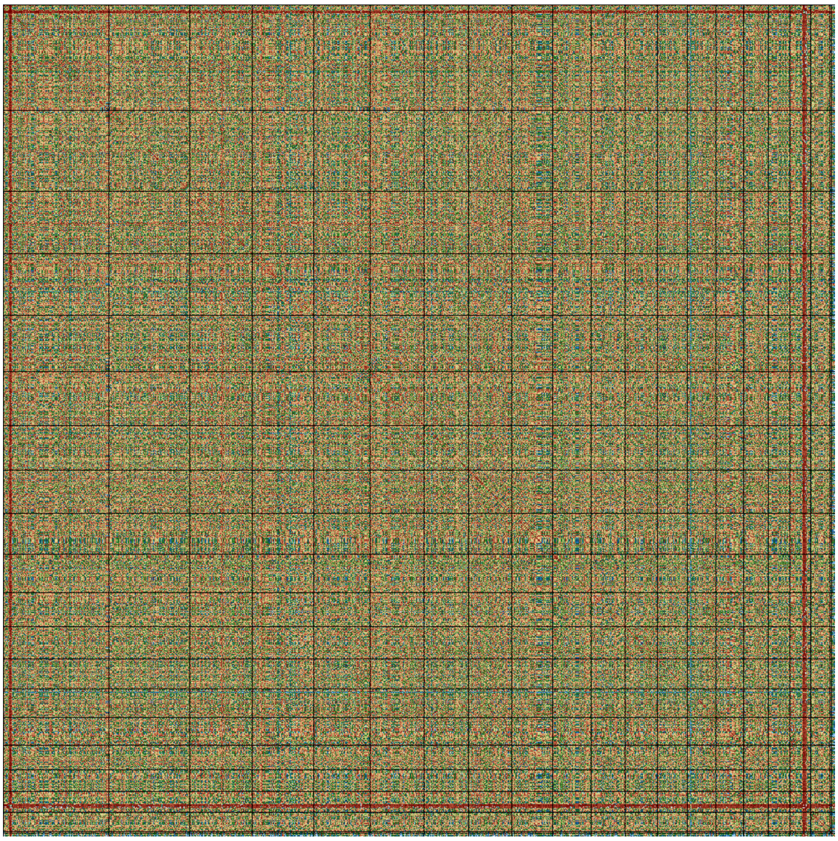
. Interventions/Gb: 347
. Contamination notes: ""
. Other observations: "The assembly of Chloropicon roscoffensis RCC7698 (ucChlRosc1) is based on 80X PacBio data and 169X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following 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, 2 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 12.141 Mb (with the largest being 8.39 Mb). Additionally, 6 regions totaling 0.882 Mb (with the largest being 0.395 Mb) were identified as haplotypic duplications and removed. Mitochondrial and chloroplastic genomes were assembled using OATK. During manual curation, 2 haplotypic regions were removed, totaling 0.008Mb (with the largest being 0.005Mb). The telomeric pattern AAACCTAAAA 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	17,302,316	17,268,302
GC %	60.19	60.23
Gaps/Gbp	0	115.82
Total gap bp	0	400
Scaffolds	21	21
Scaffold N50	1,113,234	1,113,234
Scaffold L50	6	6
Scaffold L90	16	16
Contigs	21	23
Contig N50	1,113,234	1,113,234
Contig L50	6	6
Contig L90	16	16
QV	46.1874	46.1787
Kmer compl.	51.9933	51.9859
BUSCO sing.	91.7%	91.7%
BUSCO dupl.	1.3%	1.3%
BUSCO frag.	1.2%	1.2%
BUSCO miss.	5.8%	5.8%

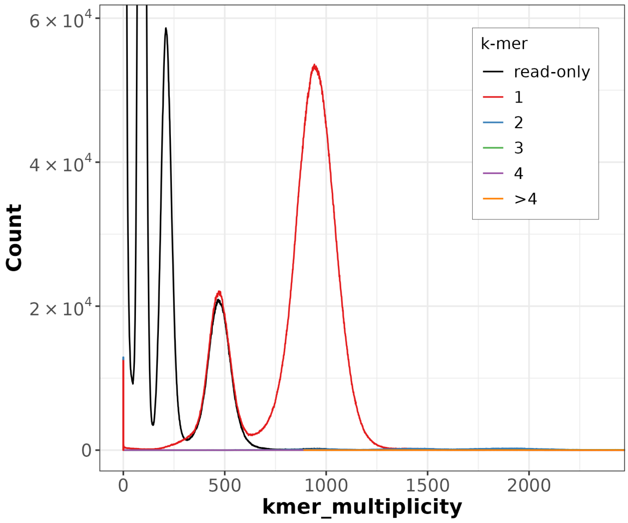
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: chlorophyta_odb12 (genomes:39, BUSCOs:1523)

HiC contact map of curated assembly

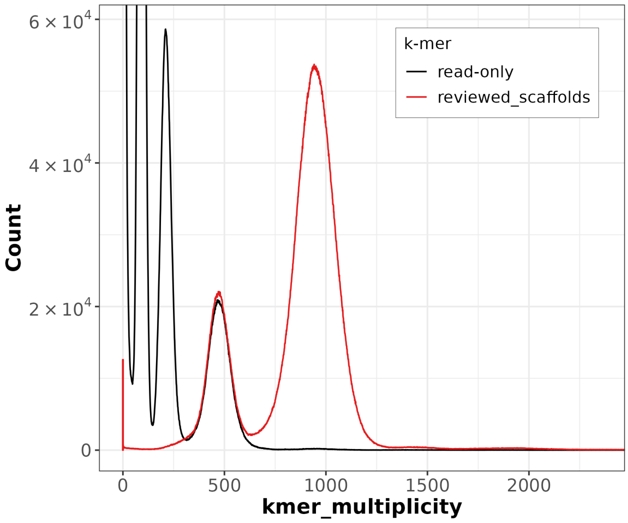


collapsed [LINK]

K-mer spectra of curated assembly

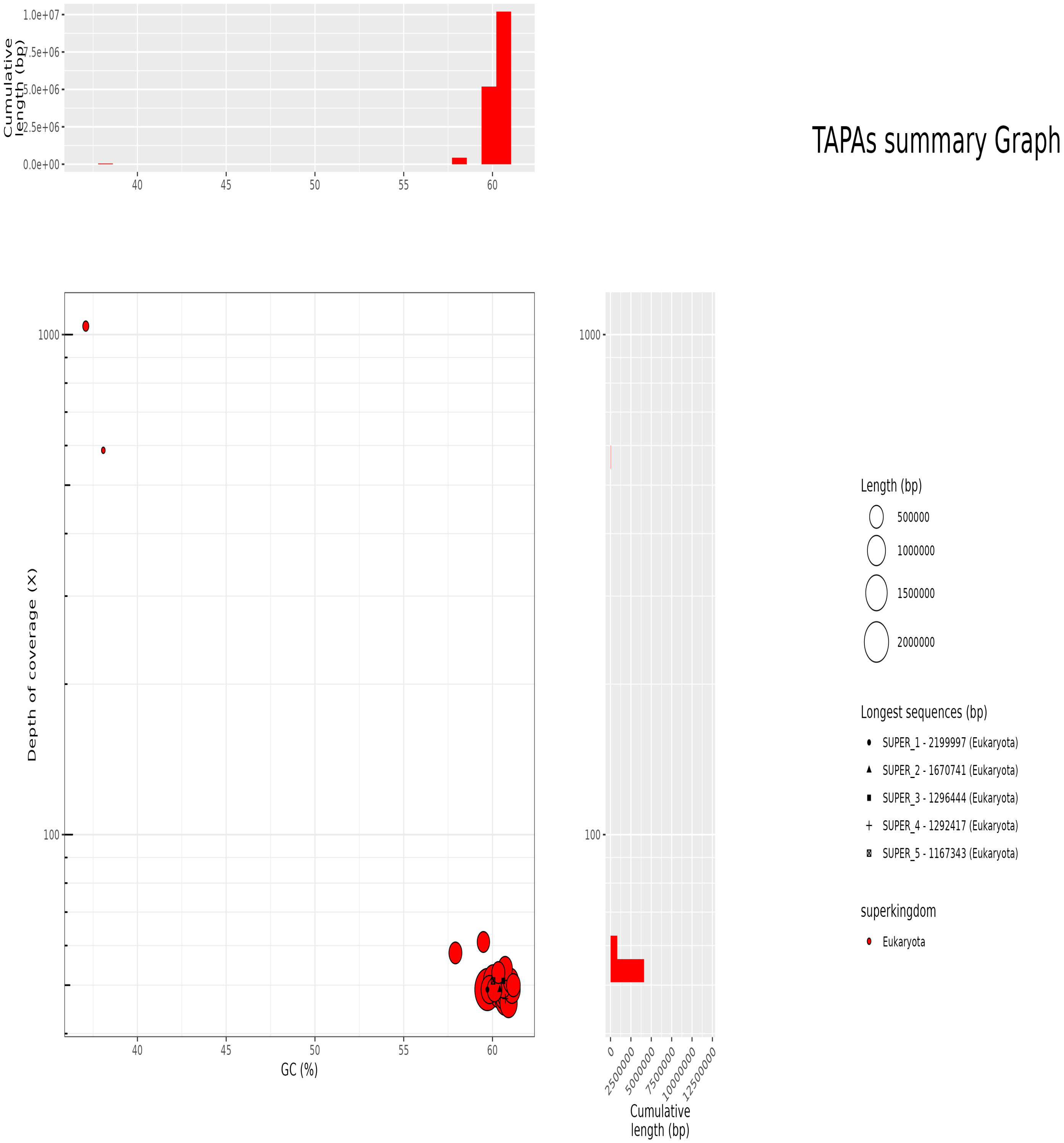


Distribution of k-mer counts per copy numbers found in asm



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	80	169

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: Benjamin Istace
Affiliation: Genoscope

Date and time: 2025-11-28 06:32:56 CET