ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1764286
ToLID	**ucChlSieb1**
Species	Chloropicon sieburthii
Class	Chloropicophyceae
Order	Chloropicales

Genome Traits	Expected	Observed
Haploid size (bp)	19,658,172	18,577,337
Haploid Number	7 (source: ancestor)	21
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.6.Q48

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
. Assembly length loss > 3% for collapsed

Curator notes

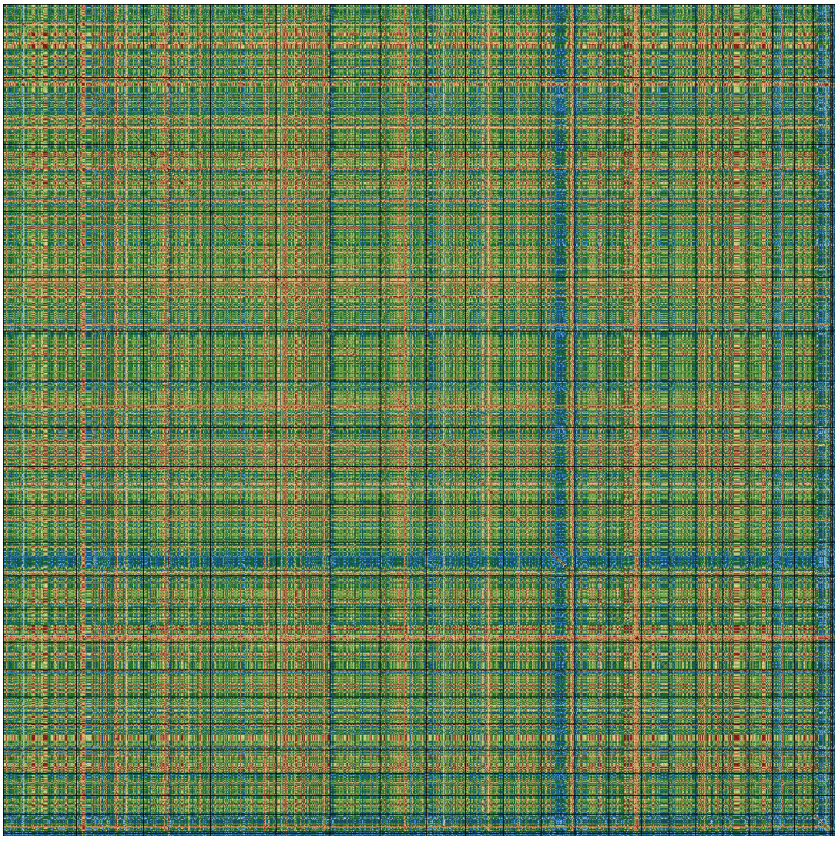
. Interventions/Gb: 364
. Contamination notes: ""
. Other observations: "The assembly of \'Chloropicon sieburthii str. RCC1043\' (ucChlSieb1) is based on 148X PacBio data and 196X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following 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, 56 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 5.752 Mb (with the largest being 3.896 Mb). Additionally, 9 regions totaling 0.854 Mb (with the largest being 0.451 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. During manual curation, 6 haplotypic regions were removed, totaling 2.208Mb (with the largest being 0.778Mb). The telomeric pattern TCTTTAGG 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	20,720,778	18,577,337
GC %	58.7	58.68
Gaps/Gbp	48.26	161.49
Total gap bp	100	500
Scaffolds	25	23
Scaffold N50	1,024,123	1,024,123
Scaffold L50	8	7
Scaffold L90	19	18
Contigs	26	26
Contig N50	1,024,123	1,024,123
Contig L50	8	7
Contig L90	19	18
QV	48.2326	48.1836
Kmer compl.	73.7961	72.9558
BUSCO sing.	86.9%	91.7%
BUSCO dupl.	6.2%	1.2%
BUSCO frag.	1.2%	1.1%
BUSCO miss.	5.7%	5.9%

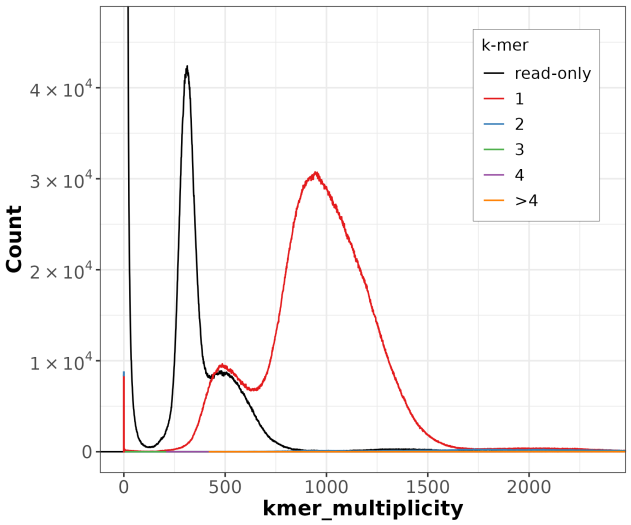
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: chlorophyta_odb12 (genomes:39, BUSCOs:1523)

HiC contact map of curated assembly

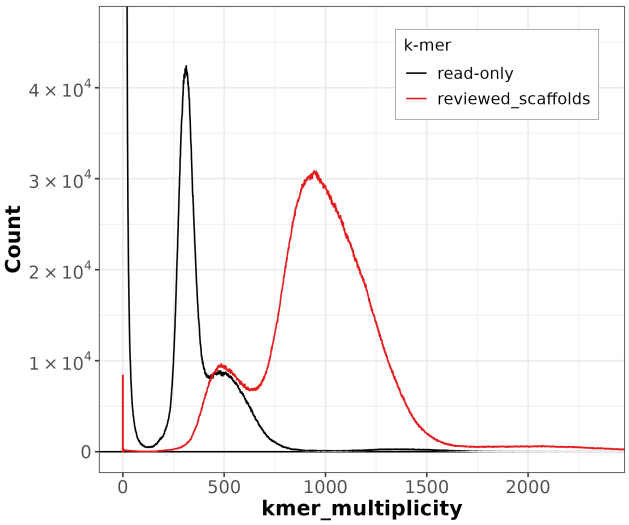


collapsed [LINK]

K-mer spectra of curated assembly

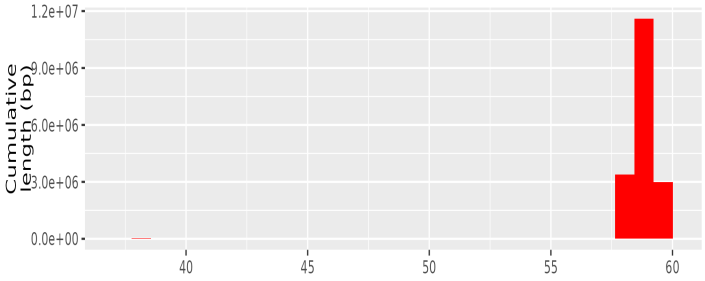


Distribution of k-mer counts per copy numbers found in asm

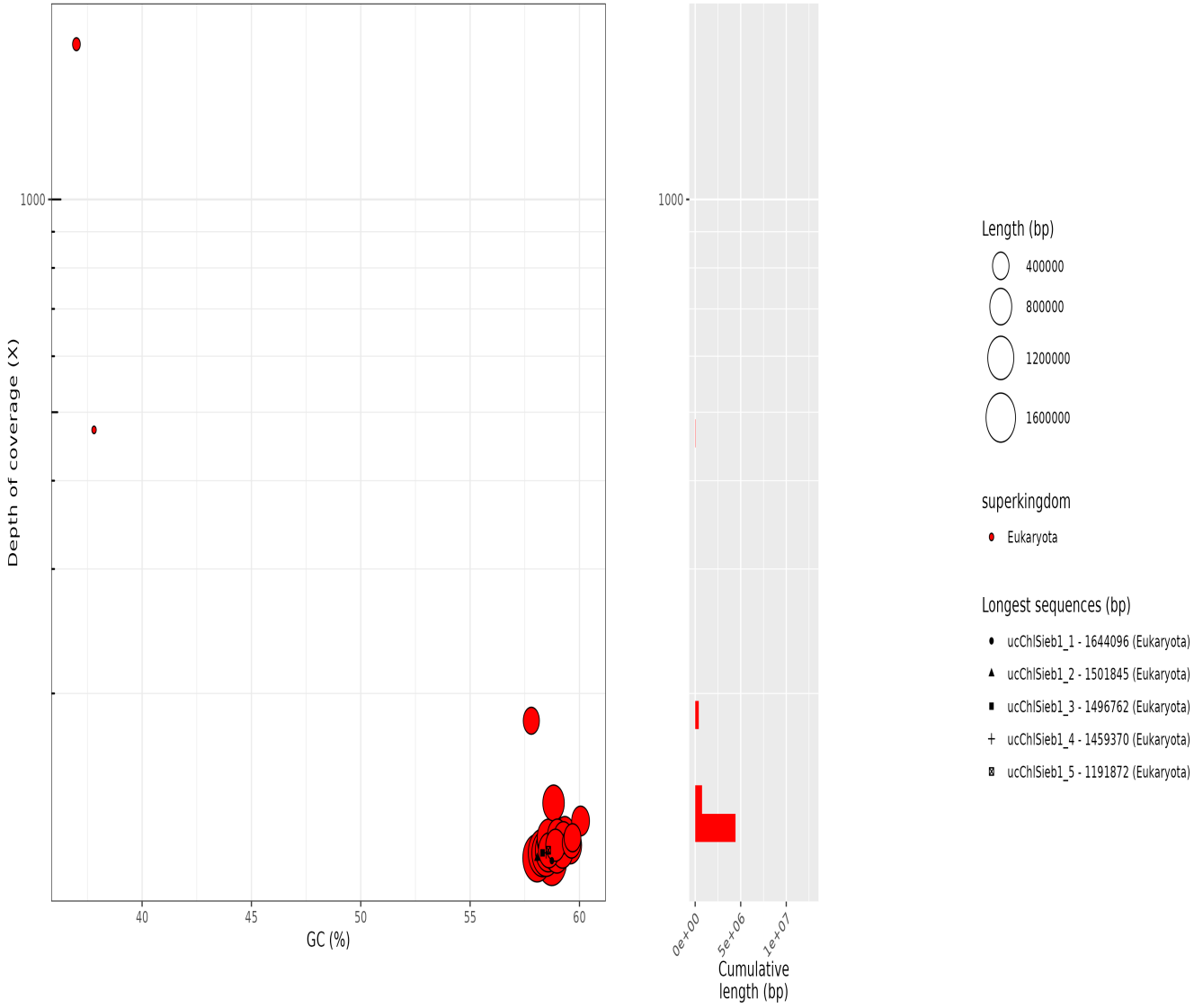


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	165	219

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: Simone Duprat
Affiliation: Genoscope

Date and time: 2026-05-03 12:22:36 CEST