ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1277942
ToLID	**urPhyCalc1**
Species	Phymatolithon calcareum
Class	Florideophyceae
Order	Hapalidiales

Genome Traits	Expected	Observed
Haploid size (bp)	109,152,907	97,261,408
Haploid Number	7 (source: ancestor)	34
Ploidy	4 (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
. Observed Ploidy is different from Expected

. Kmer completeness value is less than 90 for collapsed
. BUSCO single copy value is less than 90% for collapsed

Curator notes

. Interventions/Gb: 103
. Contamination notes: ""
. Other observations: "The assembly of Phymatolithon calcareum (urPhyCalc1) is based on 53X PacBio data and 141X 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, 2079 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 73.293 Mb (with the largest being 0.197 Mb). Additionally, 65 regions totaling 4.799 Mb (with the largest being 0.551 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. During manual curation, 1 haplotypic regions and 8 contaminant sequences were removed, totaling 0.630Mb and 0.333Mb, respectively (with the largest being 0.63Mb and 0.11Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. Note that urPhyCalc1_34 shows a different pattern

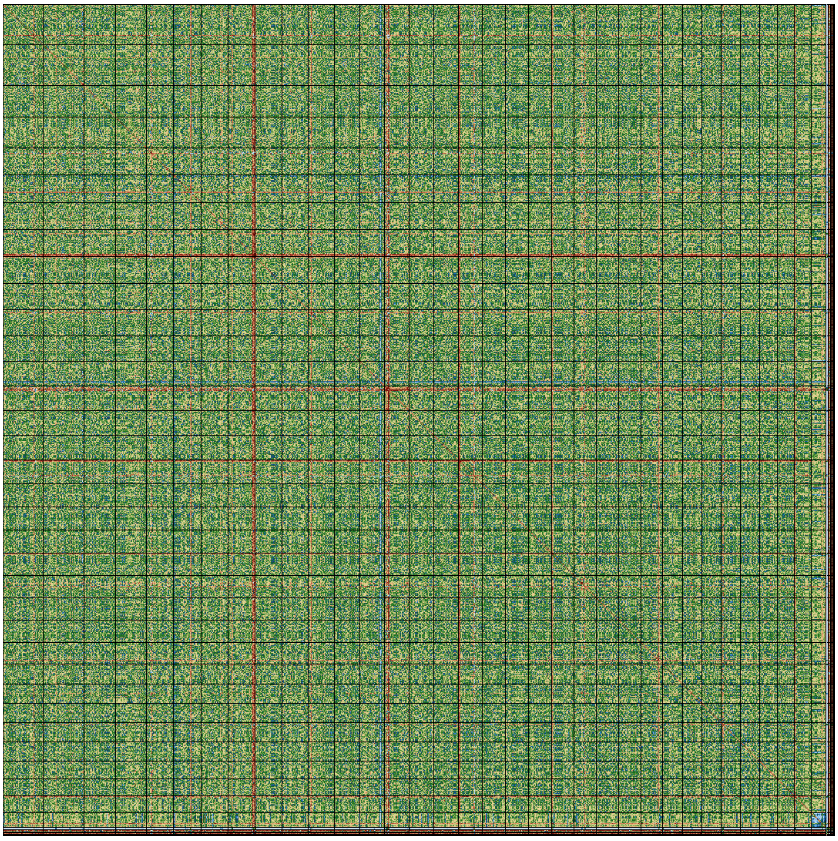
on the Hi-C compared to the other chromosomes. In addition, coverage is higher than for the other chromosomes, and a duplication was eliminated at one extremity, suggesting a circular sequence that may correspond to a plastid genome. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	98,120,205	97,261,408
GC %	47.8	47.81
Gaps/Gbp	81.53	82.25
Total gap bp	800	1,200
Scaffolds	49	43
Scaffold N50	2,956,594	2,873,000
Scaffold L50	12	15
Scaffold L90	27	30
Contigs	57	51
Contig N50	2,822,023	2,822,023
Contig L50	15	15
Contig L90	31	31
QV	45.9157	46.9421
Kmer compl.	83.9793	84.1243
BUSCO sing.	85.5%	88.1%
BUSCO dupl.	2.6%	2.8%
BUSCO frag.	7.0%	3.5%
BUSCO miss.	4.9%	5.7%

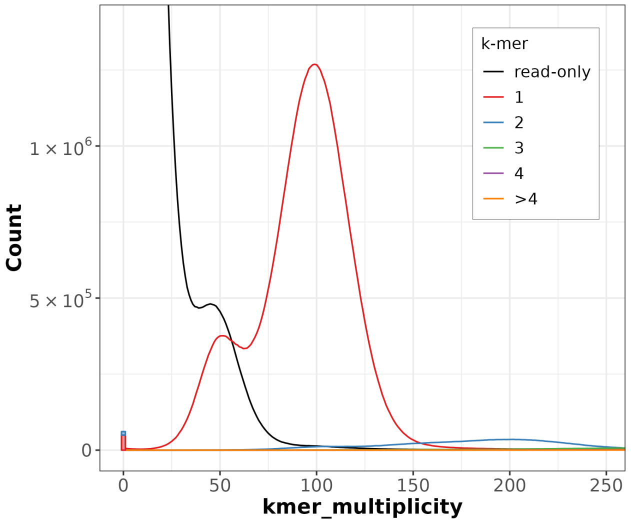
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: rhodophyta_odb12 (genomes:8, BUSCOs:1591)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: rhodophyta_odb12 (genomes:8, BUSCOs:1591)

HiC contact map of curated assembly

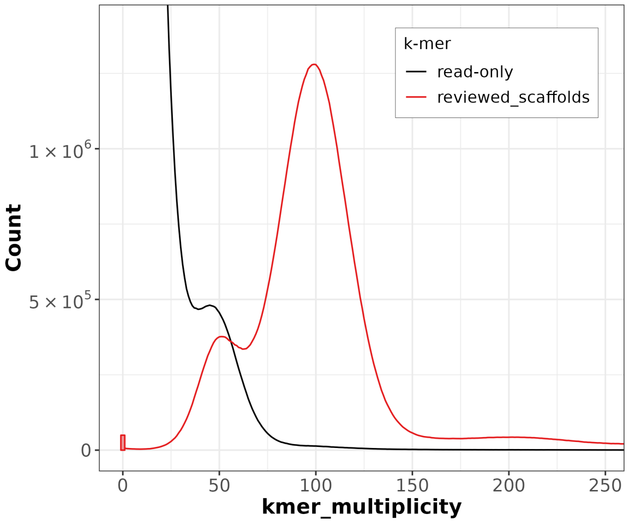


collapsed [LINK]

K-mer spectra of curated assembly

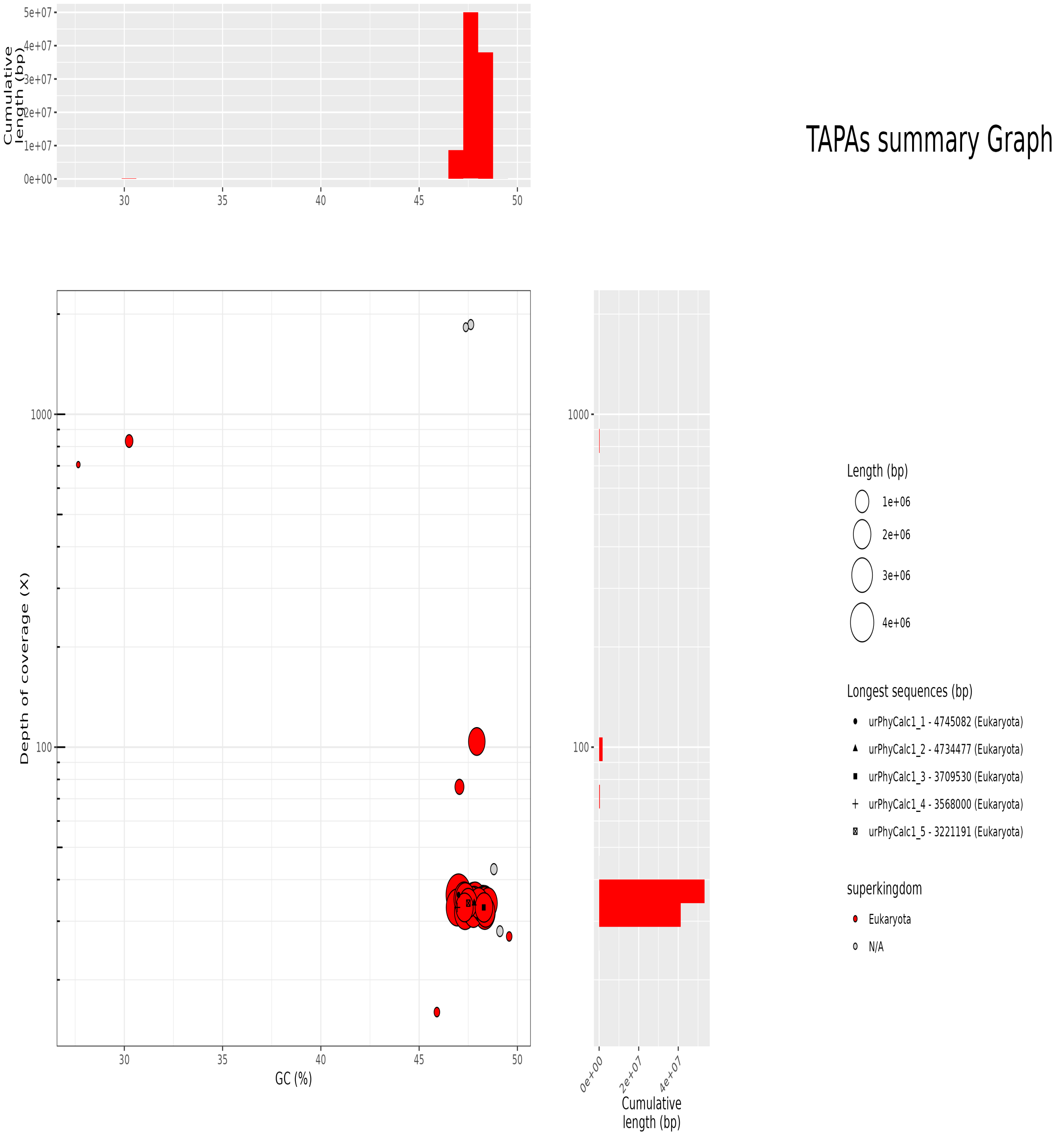


Distribution of k-mer counts per copy numbers found in asm



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	53	140

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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2025-10-08 13:01:29 CEST