

ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	466607
ToLID	urLitStic1
Species	Lithophyllum stictiforme
Class	Florideophyceae
Order	Corallinales

Genome Traits	Expected	Observed
Haploid size (bp)	217,533,958	185,314,401
Haploid Number	7 (source: ancestor)	32
Ploidy	4 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 5.6.Q47

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
- . More than 1000 gaps/Gbp for collapsed

Curator notes

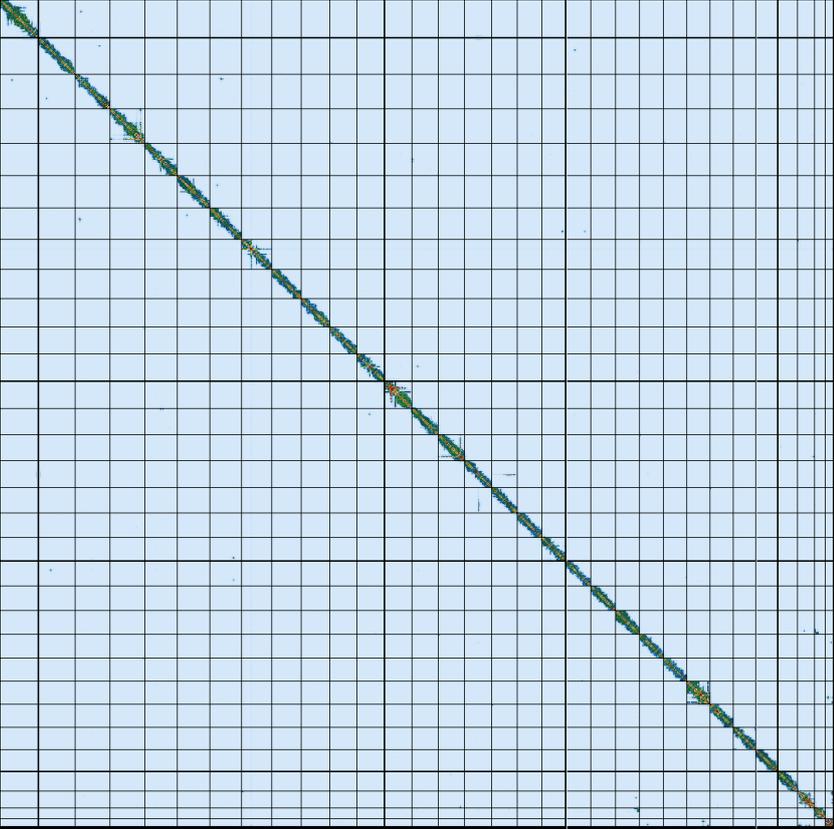
- . Interventions/Gb: 1019
- . Contamination notes: ""
- . Other observations: "The assembly of *Lithophyllum stictiforme* (urLitStic1) is based on 50X PacBio data and 220X Arima Hi-C data generated as part of the ATLASea programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial PacBio assembly generation with Hifiiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 226 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 4.841 Mb (with the largest being 0.168 Mb). Additionally, 367 regions totaling 8.898 Mb (with the largest being 0.164 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, 4 contaminant sequences were removed, totaling 0.195Mb (with the largest being 0.112Mb). 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	185,503,699	185,314,401
GC %	43.28	43.29
Gaps/Gbp	1,919.1	2,261.02
Total gap bp	35,600	53,200
Scaffolds	105	46
Scaffold N50	5,919,614	5,919,614
Scaffold L50	12	14
Scaffold L90	25	27
Contigs	461	465
Contig N50	655,459	640,000
Contig L50	89	91
Contig L90	276	284
QV	46.3475	47.16
Kmer compl.	78.6664	78.6282
BUSCO sing.	78.0%	78.1%
BUSCO dupl.	3.7%	3.6%
BUSCO frag.	3.2%	3.1%
BUSCO miss.	15.1%	15.1%

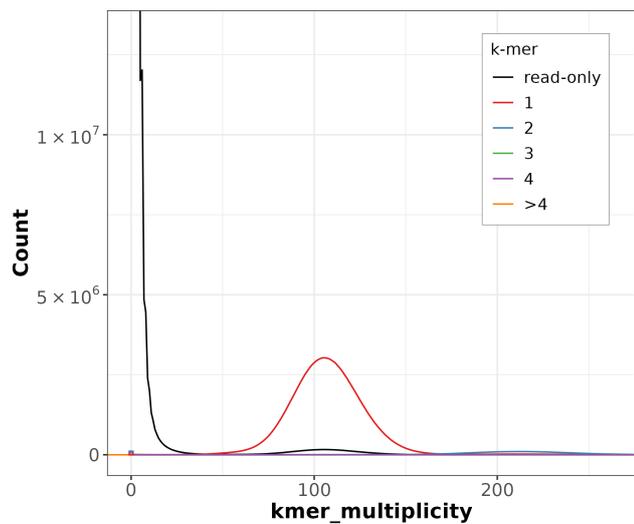
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: rhodophyta_odb12 (genomes:8, BUSCOs:1591)

HiC contact map of curated assembly

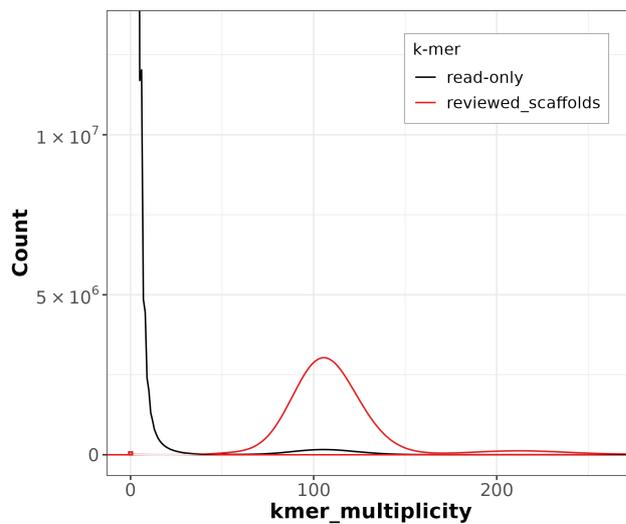


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

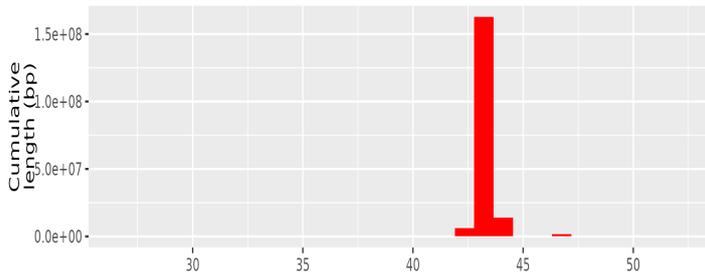


Distribution of k-mer counts per copy numbers found in asm

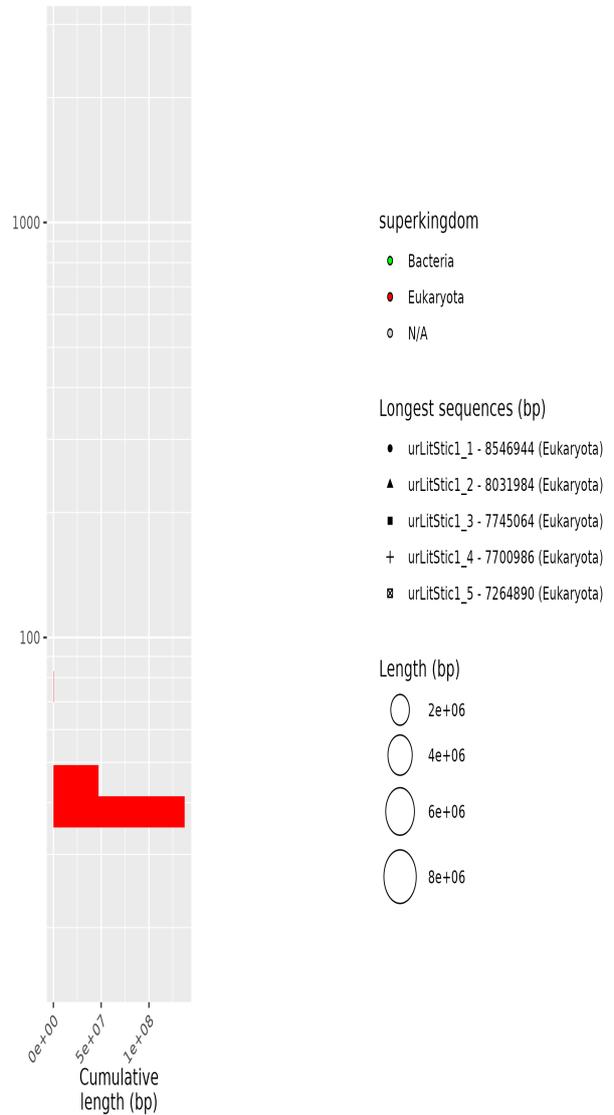
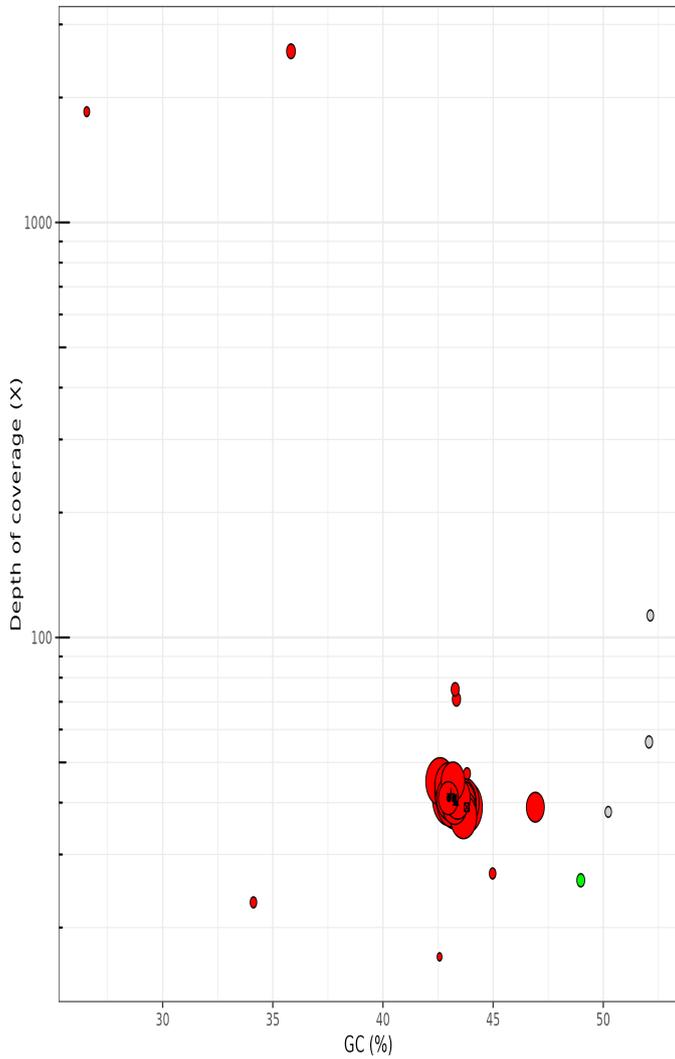


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



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	43	187

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: Mohamed Meftah Saoues

Affiliation: Genoscope

Date and time: 2026-02-06 03:21:56 CET