

ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	2008650
ToLID	urAlsCoral
Species	Alsidium corallinum
Class	Florideophyceae
Order	Ceramiales

Genome Traits	Expected	Observed
Haploid size (bp)	394,807,981	286,235,878
Haploid Number	4 (source: ancestor)	32
Ploidy	4 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 5.7.Q44

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

- . Observed Haploid size (bp) has >20% difference with Expected
- . Observed Haploid Number is different from Expected
- . Observed Ploidy is different from Expected
- . Assembly length loss > 3% for collapsed
- . More than 1000 gaps/Gbp for collapsed

Curator notes

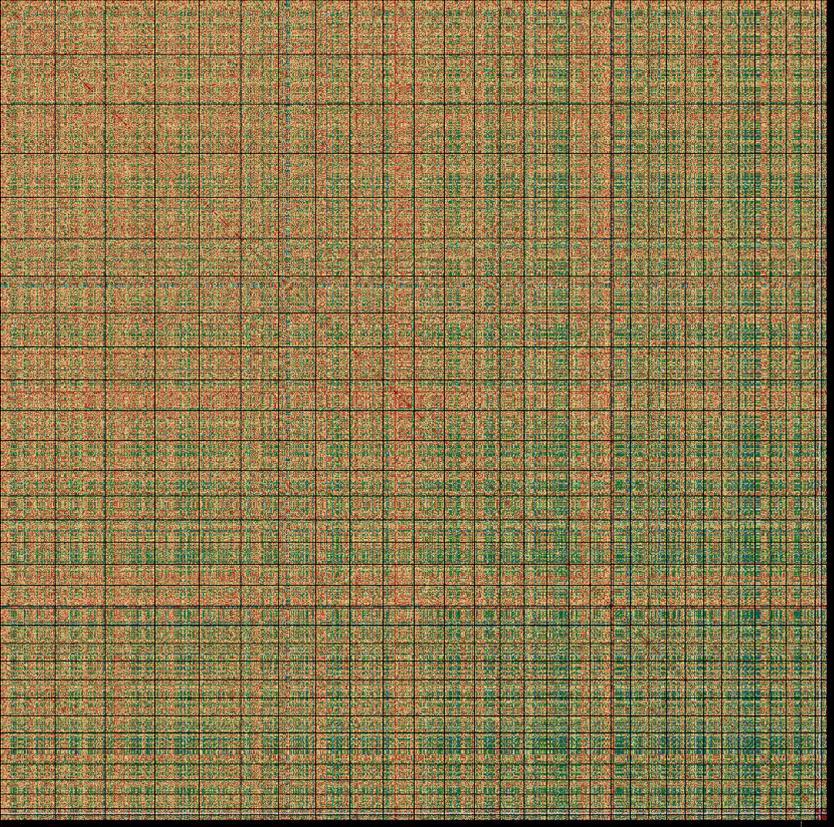
- . Interventions/Gb: 477
- . Contamination notes: ""
- . Other observations: "The assembly of Alsidium corallinum (urAlsCoral) is based on 71X PacBio data and 179X Arima Hi-C data generated as part of the ATLASea programme (<https://www.atlasea.fr>). The assembly process included the following steps: 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, 96 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.969 Mb (with the largest being 0.099 Mb). Additionally, 1172 regions totaling 26.098 Mb (with the largest being 0.17 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 84 haplotypic regions and 82 contaminant sequences were removed, totaling 9.412Mb and 2.436Mb, respectively (with the largest being 0.399Mb and 0.608Mb). 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	297,890,418	286,235,878
GC %	43.12	43.22
Gaps/Gbp	2,685.55	2,945.12
Total gap bp	80,000	97,000
Scaffolds	317	113
Scaffold N50	7,267,443	10,469,179
Scaffold L50	16	11
Scaffold L90	45	26
Contigs	1,117	956
Contig N50	518,055	524,455
Contig L50	176	169
Contig L90	611	561
QV	43.14	44.1057
Kmer compl.	92.8495	91.8915
BUSCO sing.	89.4%	90.1%
BUSCO dupl.	2.6%	1.5%
BUSCO frag.	2.6%	2.6%
BUSCO miss.	5.3%	5.8%

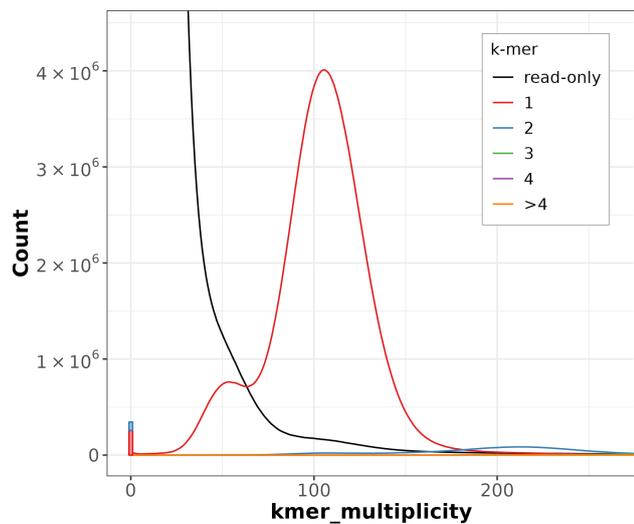
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: rhodophyta_odb12 (genomes:8, BUSCOs:1591)

HiC contact map of curated assembly

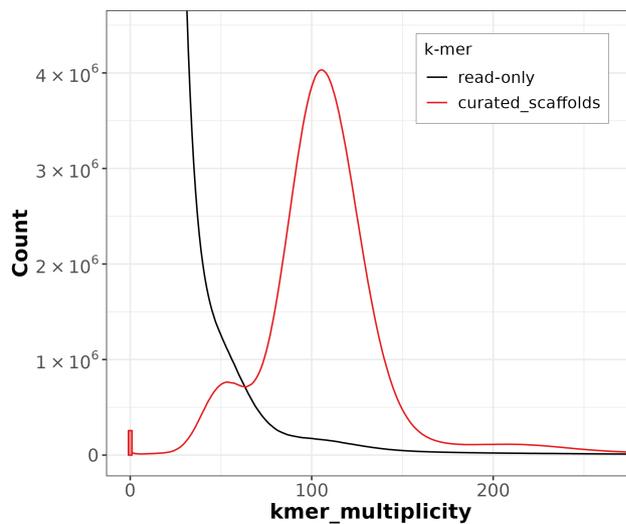


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

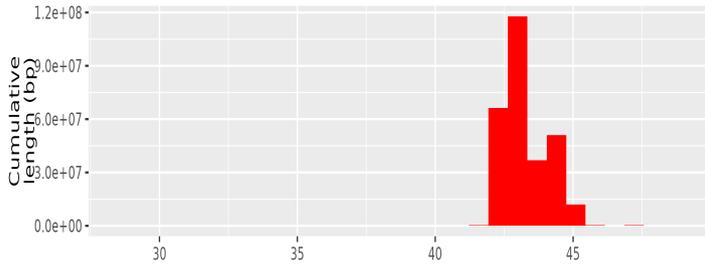


Distribution of k-mer counts per copy numbers found in asm

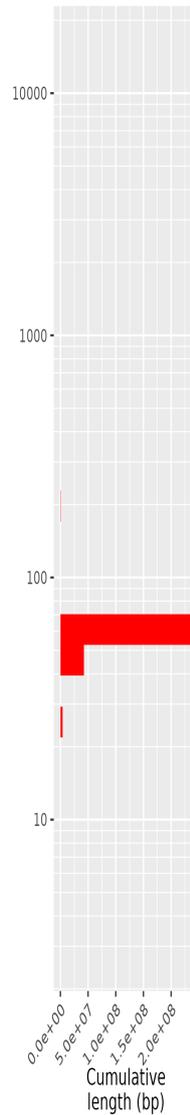
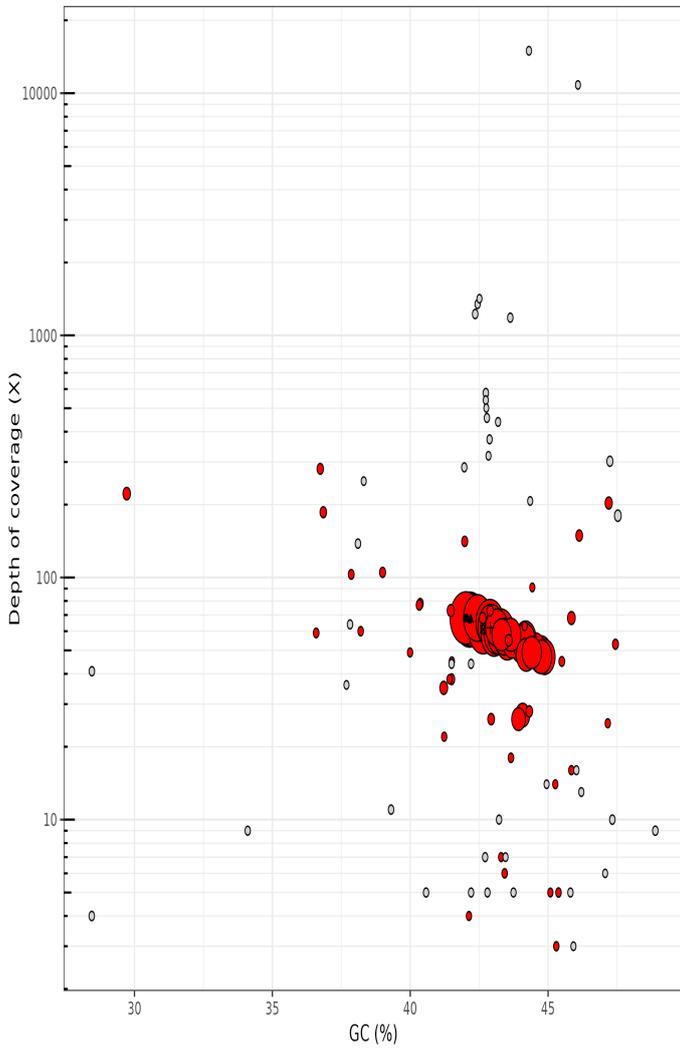


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



Longest sequences (bp)

- SUPER_1 - 19076079 (Eukaryota)
- ▲ SUPER_2 - 17010599 (Eukaryota)
- SUPER_3 - 16955296 (Eukaryota)
- + SUPER_4 - 15155698 (Eukaryota)
- ▣ SUPER_5 - 14330017 (Eukaryota)

superkingdom

- Eukaryota
- N/A

Length (bp)

- 5.0e+06
- 1.0e+07
- 1.5e+07

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	71	179

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: Sophie Layac

Affiliation: Genoscope

Date and time: 2026-01-28 17:01:18 CET