

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	408200
ToLID	urRhoRamo1
Species	Rhodogorgon ramosissima
Class	Florideophyceae
Order	Rhodogorgonales

Genome Traits	Expected	Observed
Haploid size (bp)	102,490,527	37,658,053
Haploid Number	4 (source: ancestor)	31
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 size (bp) has >20% difference with Expected
- . Observed Haploid Number is different from Expected
- . Observed Ploidy is different from Expected
- . Kmer completeness value is less than 90 for collapsed
- . Assembly length loss > 3% for collapsed

Curator notes

- . Interventions/Gb: 123
- . Contamination notes: ""
- . Other observations: "The assembly of Rhodogorgon ramosissima (urRhoRamo1) is based on 209X PacBio data and 161X 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, 371 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 31.947 Mb (with the largest being 4.655 Mb). Additionally, 633 regions totaling 12.485 Mb (with the largest being 0.074 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, 367 haplotypic regions and 9 contaminant sequences were removed, totaling 43.625Mb and 0.407Mb, respectively (with the largest being 0.734Mb and 0.069Mb). Chromosome-scale scaffolds

confirmed by Hi-C data were named in order of size. "

Quality metrics table

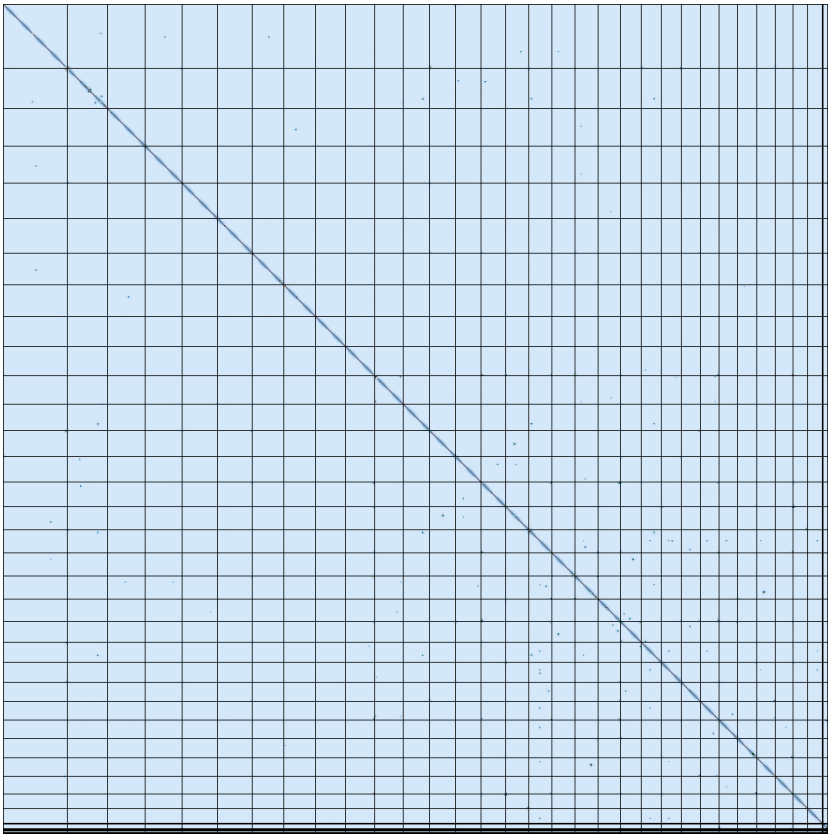
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	110,600,393	37,658,053
GC %	45.59	50.18
Gaps/Gbp	9.04	79.66
Total gap bp	100	600
Scaffolds	1,103	42
Scaffold N50	218,833	1,208,824
Scaffold L50	96	12
Scaffold L90	764	27
Contigs	1,104	45
Contig N50	218,833	1,180,493
Contig L50	96	13
Contig L90	765	28
QV	39.9206	46.9263
Kmer compl.	70.3203	69.8834
BUSCO sing.	16.5%	92.8%
BUSCO dupl.	76.7%	0.4%
BUSCO frag.	3.3%	3.1%
BUSCO miss.	3.6%	3.6%

Warning! BUSCO versions or lineage datasets are not the same across results:

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

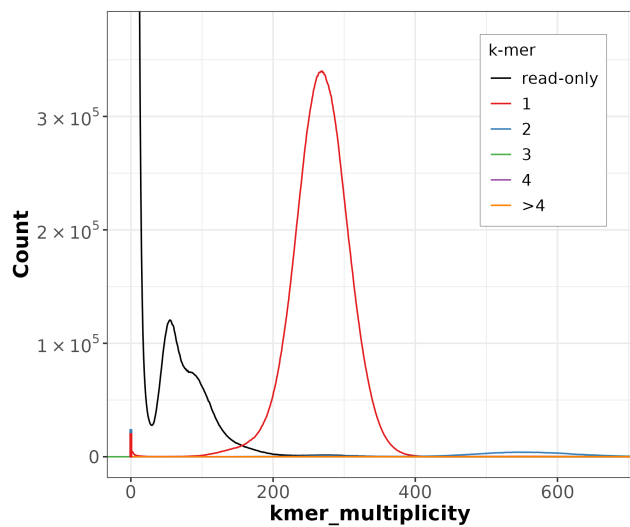
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: rhodophyta_odb12 (genomes:8, BUSCOs:1591)

HiC contact map of curated assembly

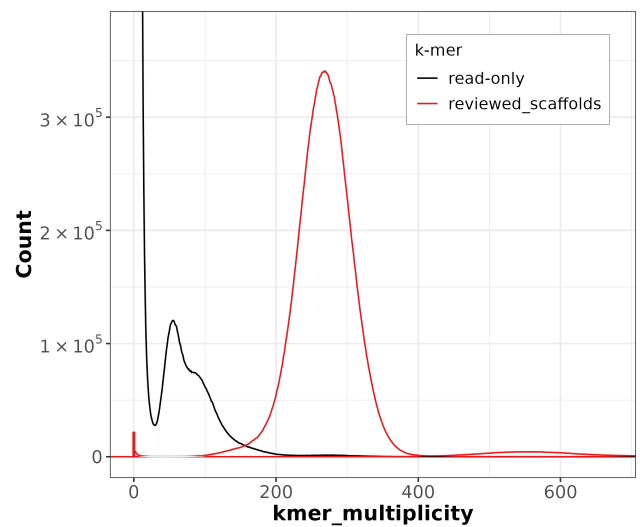


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

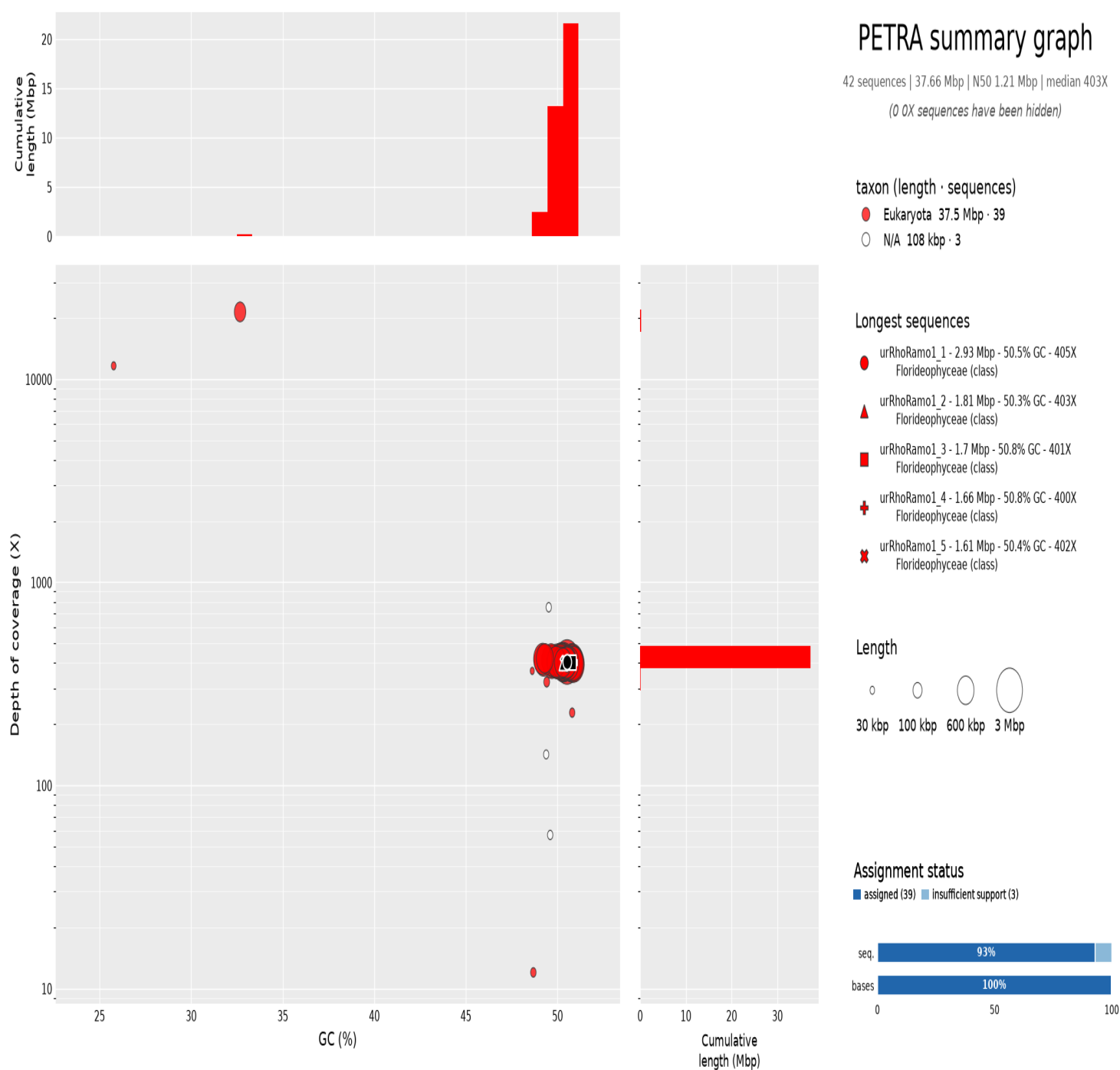


Distribution of k-mer counts per copy numbers found in asm



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	615	472

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*: 0.2.4
 - |_ *key param*: NA
- **PretextView**
 - |_ *ver*: 1.0.7
 - |_ *key param*: NA

Submitter: Jean-Marc Aury

Affiliation: Genoscope

Date and time: 2026-10-06 10:27:49 CEST