ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1310127
ToLID	**urGulNodu1**
Species	Gulsonia nodulosa
Class	Florideophyceae
Order	Ceramiales

Genome Traits	Expected	Observed
Haploid size (bp)	1,120,403,590	1,022,149,418
Haploid Number	5 (source: ancestor)	32
Ploidy	4 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q67

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

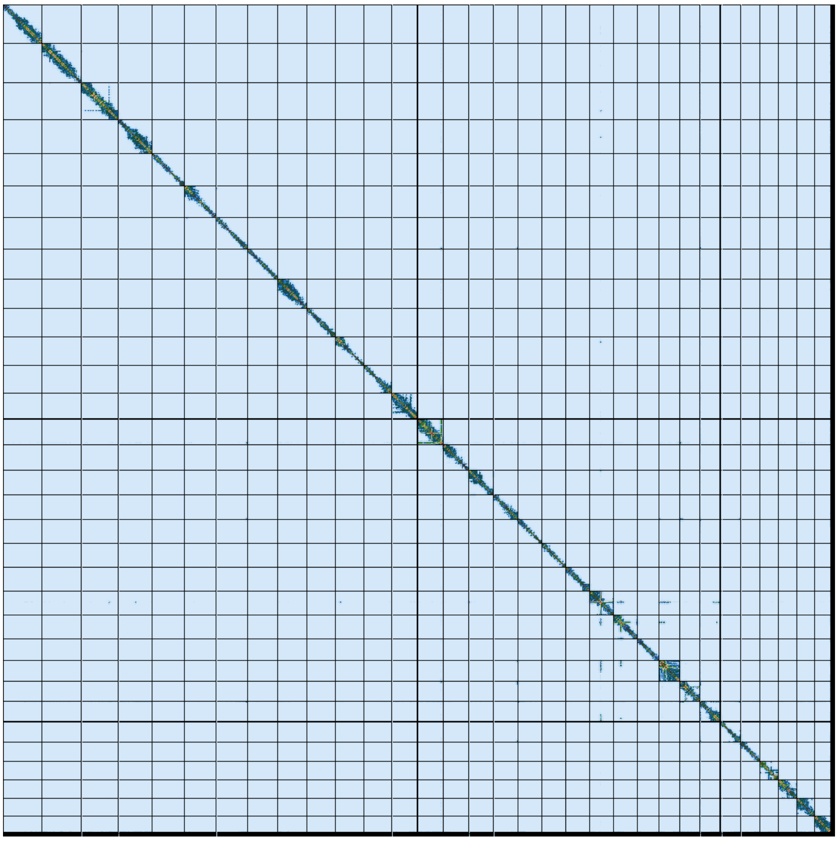
. Interventions/Gb: 62
. Contamination notes: ""
. Other observations: "The assembly of Gulsonia nodulosa (urGulNodu1) is based on 43X PacBio data and Arima HighCoverage Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included the following steps: initial 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, 607 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 60.4 Mb (with the largest being 4.4 Mb). Additionally, 619 regions totaling 348 Mb (with the largest being 34 Mb) were identified as haplotypic duplications. The contigs > 1Mb were reintroduced in the assembly. The mitochondrial and chloroplastic genomes were assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 24 haplotypic regions was removed,totaling 10.6 Mb (with the largest being 2.1 Mb) "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	716,820,763	1,022,149,418
GC %	56.56	56.52
Gaps/Gbp	47.43	77.29
Total gap bp	3,400	10,200
Scaffolds	133	135
Scaffold N50	26,795,376	31,295,068
Scaffold L50	11	14
Scaffold L90	23	28
Contigs	167	214
Contig N50	16,586,338	16,530,207
Contig L50	15	21
Contig L90	40	60
QV	43.148	67.0764
Kmer compl.	59.9284	83.1886
BUSCO sing.	54.1%	71.0%
BUSCO dupl.	2.7%	2.4%
BUSCO frag.	9.8%	9.8%
BUSCO miss.	33.4%	16.8%

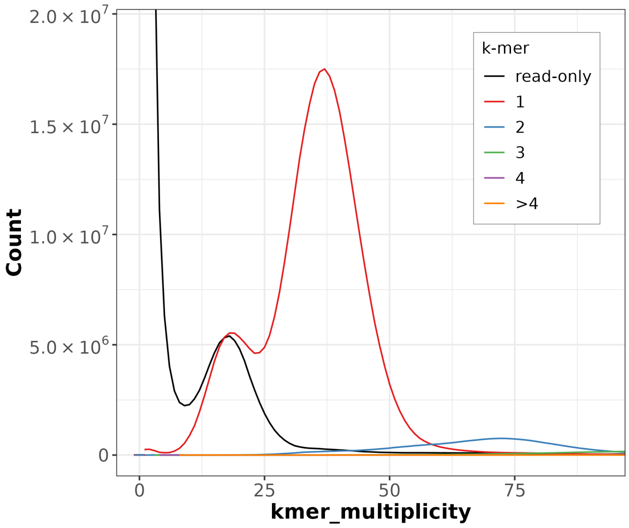
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: eukaryota_odb10 (genomes:70, BUSCOs:255)

HiC contact map of curated assembly

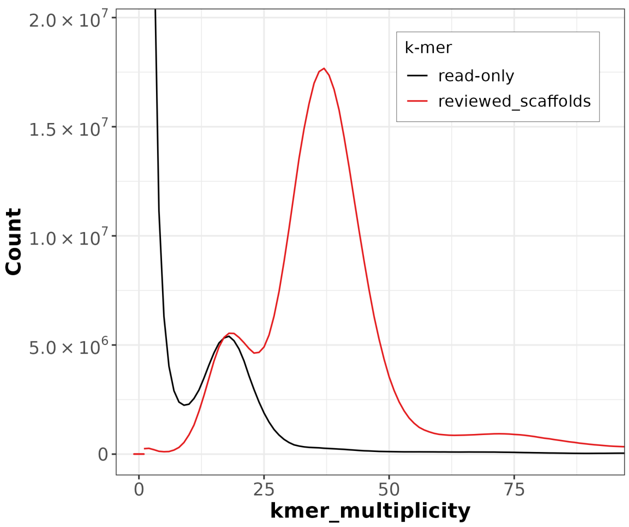


collapsed [LINK]

K-mer spectra of curated assembly

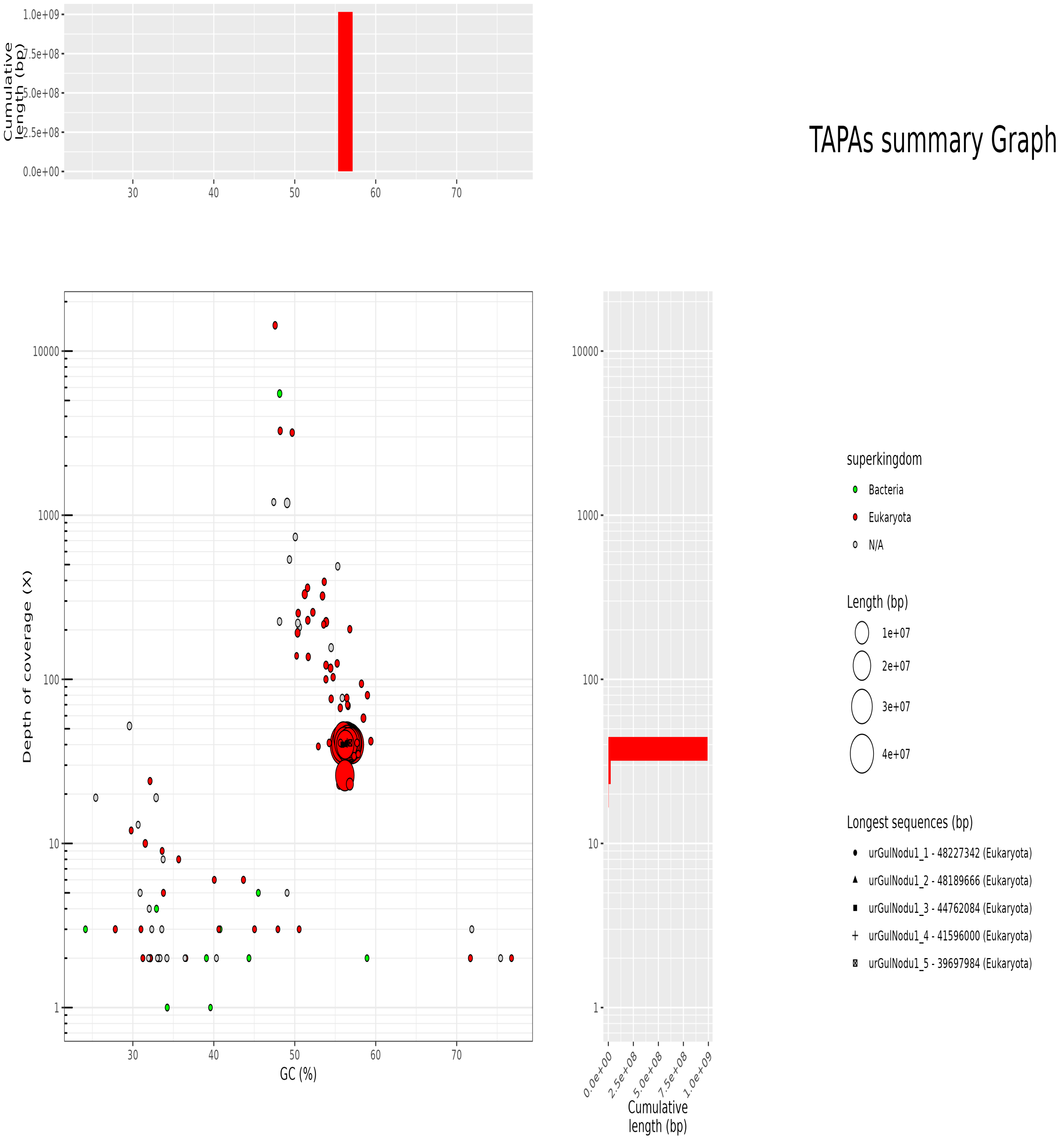


Distribution of k-mer counts per copy numbers found in asm



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	PACBIO Hifi	Arima
Coverage	39	86

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: Caroline Menguy
Affiliation: Genoscope

Date and time: 2025-04-13 15:09:36 CEST