ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	508027
ToLID	**urDudVert1**
Species	Dudresnaya verticillata
Class	Florideophyceae
Order	Gigartinales

Genome Traits	Expected	Observed
Haploid size (bp)	133,936,454	76,209,045
Haploid Number	4 (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.Q38

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

. QV value is less than 40 for collapsed
. Kmer completeness value is less than 90 for collapsed
. Assembly length loss > 3% for collapsed

Curator notes

. Interventions/Gb: 133
. Contamination notes: ""
. Other observations: "The assembly of \'Dudresnaya verticillata\' (urDudVert1) is based on 89X PacBio data and 147X 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, 116 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 18.557 Mb (with the largest being 3.221 Mb). Additionally, 292 regions totaling 9.659 Mb (with the largest being 0.147 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, 283 contaminant sequences were removed, totaling 10.319Mb (with the largest being 0.219Mb). The

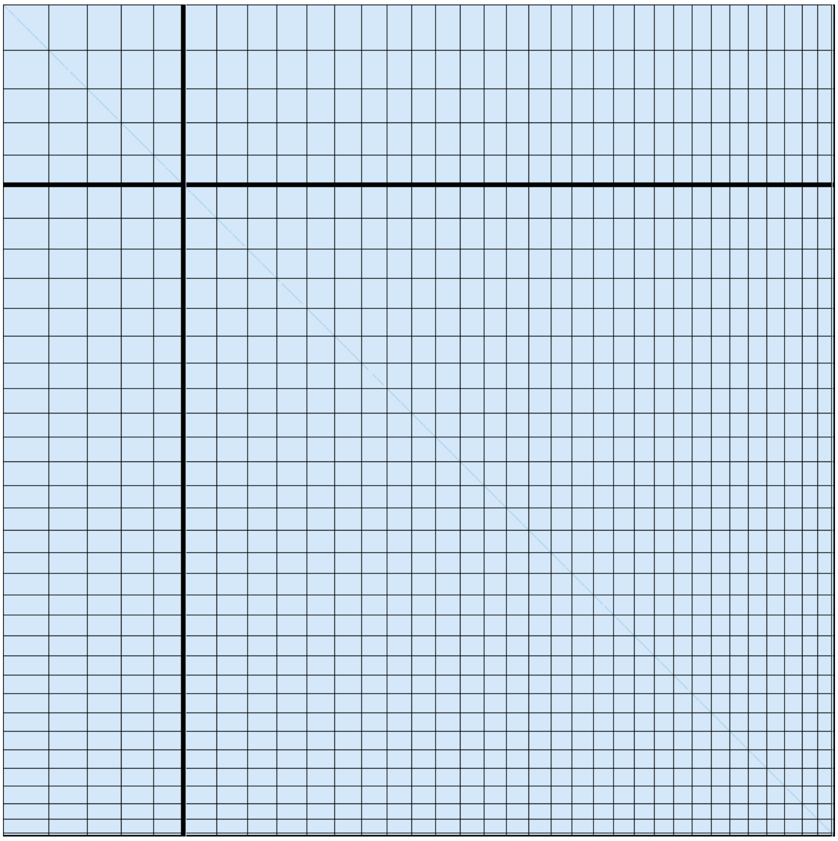
telomeric pattern TTTAGGGG was identified with TelFinder and used to generate the Pretext Telomeres tracks. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	135,821,376	76,209,045
GC %	43.07	53.82
Gaps/Gbp	66.26	183.71
Total gap bp	900	2,400
Scaffolds	1,141	45
Scaffold N50	1,628,858	2,225,360
Scaffold L50	28	14
Scaffold L90	753	30
Contigs	1,150	59
Contig N50	1,601,000	2,220,268
Contig L50	30	14
Contig L90	761	31
QV	33.696	38.7472
Kmer compl.	77.5518	75.3088
BUSCO sing.	90.4%	92.6%
BUSCO dupl.	1.0%	0.9%
BUSCO frag.	4.4%	2.1%
BUSCO miss.	4.1%	4.3%

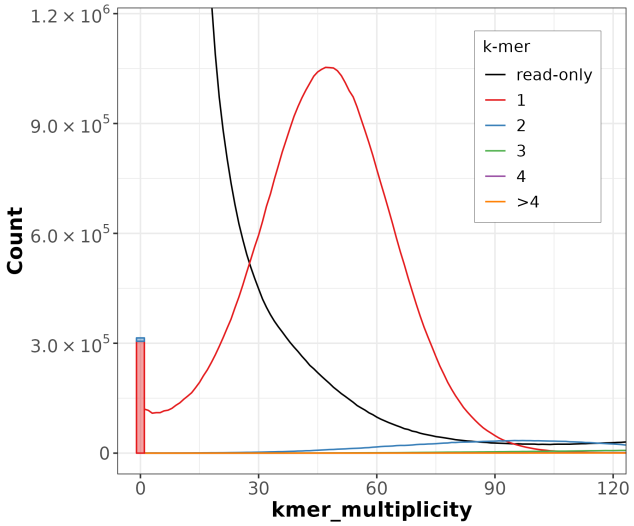
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: rhodophyta_odb12 (genomes:8, BUSCOs:1591)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: rhodophyta_odb12 (genomes:8, BUSCOs:1591)

HiC contact map of curated assembly

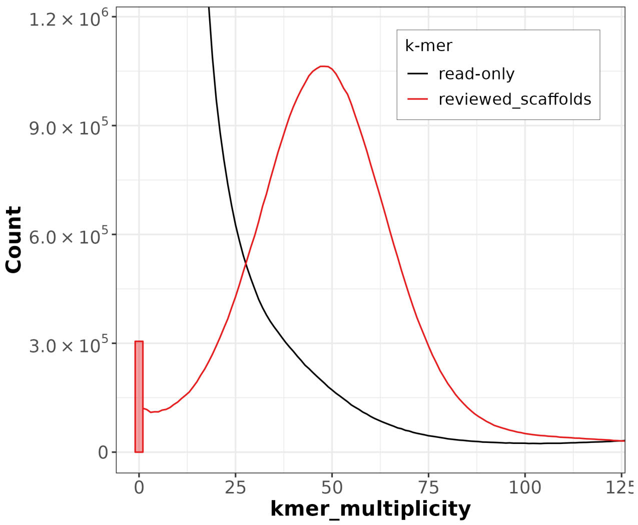


collapsed [LINK]

K-mer spectra of curated assembly

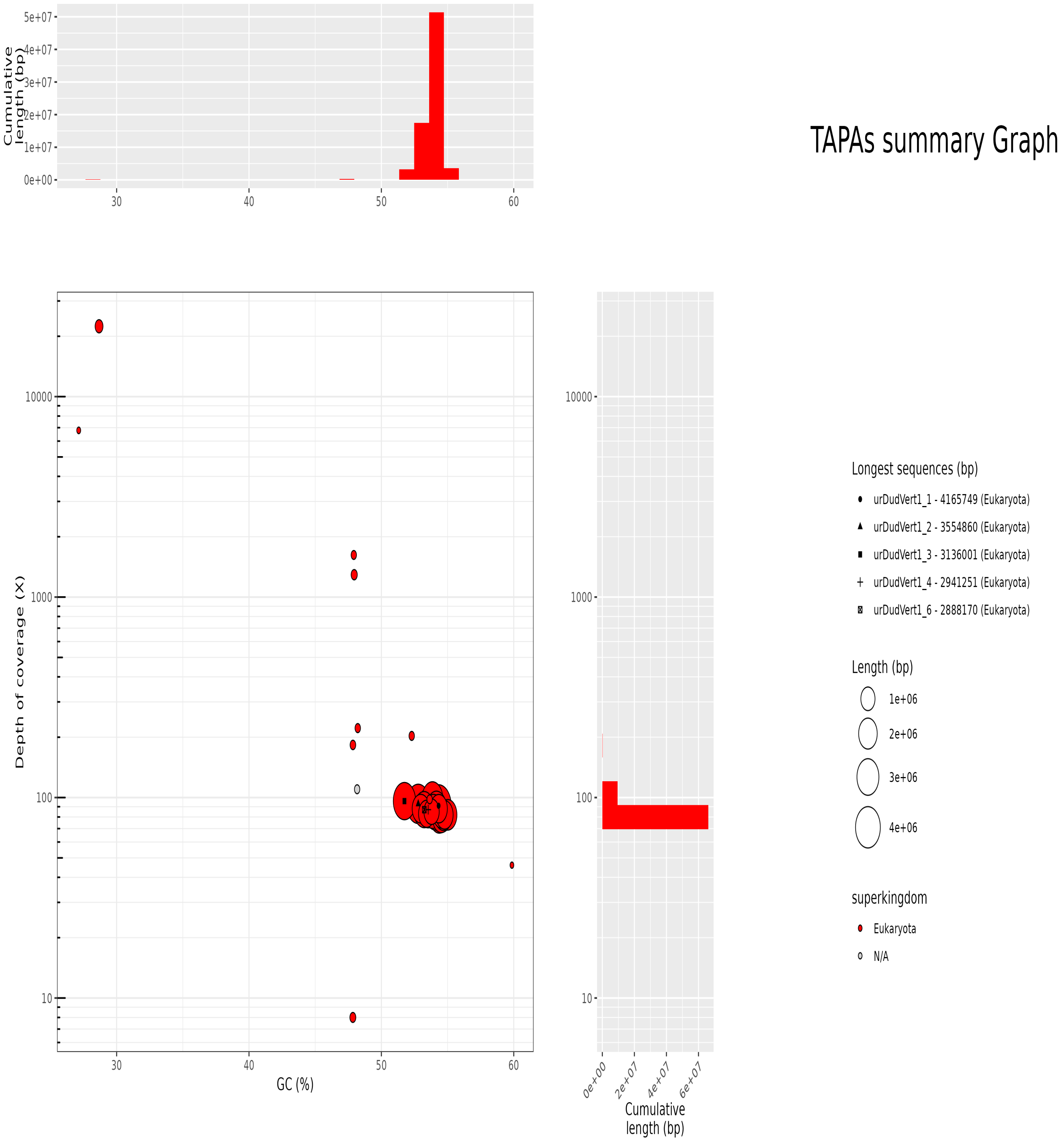


Distribution of k-mer counts per copy numbers found in asm



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	89	147

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: Simone Duprat
Affiliation: Genoscope

Date and time: 2025-11-27 00:32:53 CET