ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	3451504
ToLID	**ucPseSpea1**
Species	Pseudoscourfieldia sp. RCC10579
Class	Pseudoscourfieldiophyceae
Order	Pseudoscourfieldiales

Genome Traits	Expected	Observed
Haploid size (bp)	26,722,634	28,608,579
Haploid Number	7 (source: ancestor)	44
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 5.5.Q45

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

. Observed Haploid Number 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

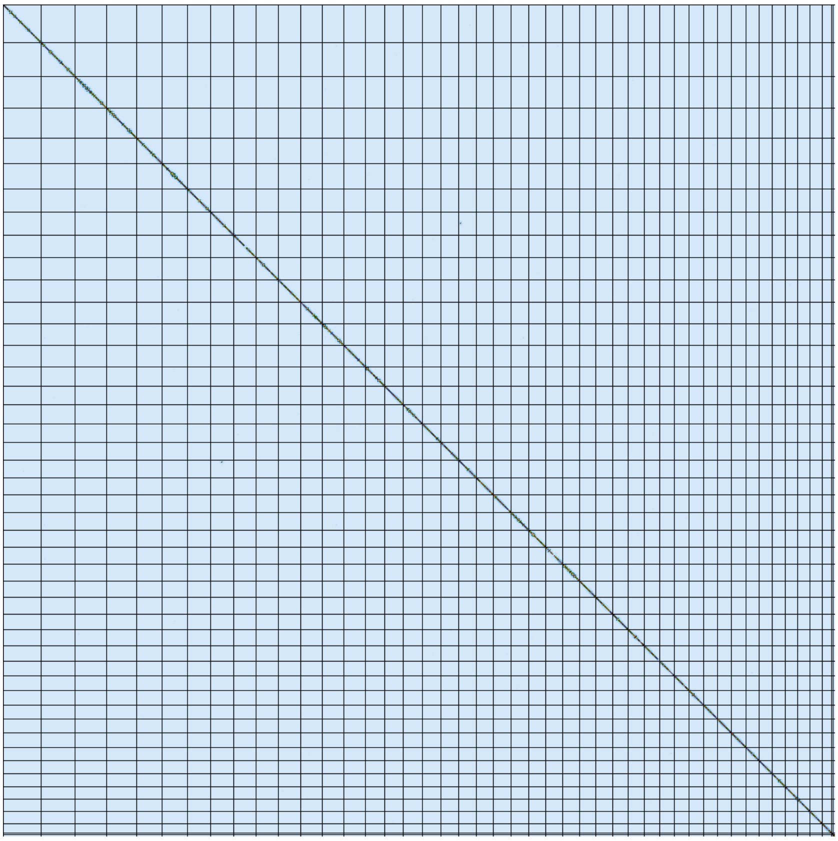
. Interventions/Gb: 0
. Contamination notes: ""
. Other observations: "The assembly of Pseudoscourfieldia sp. RCC10579 (ucPseSpea1) is based on 66X PacBio data and 515X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: 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. No contigs were found to be contaminants (bacterial, archaeal, or viral). Additionally, 3 regions totaling 0.064 Mb (with the largest being 0.027 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. The telomeric pattern AACCCT 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	28,629,173	28,608,579
GC %	67.65	67.66
Gaps/Gbp	34.93	0
Total gap bp	100	0
Scaffolds	45	46
Scaffold N50	648,132	648,132
Scaffold L50	17	17
Scaffold L90	38	38
Contigs	46	46
Contig N50	648,132	648,132
Contig L50	17	17
Contig L90	38	38
QV	45.5336	45.5305
Kmer compl.	89.0061	89.0051
BUSCO sing.	86.3%	87.2%
BUSCO dupl.	1.8%	1.5%
BUSCO frag.	3.6%	1.6%
BUSCO miss.	8.3%	9.7%

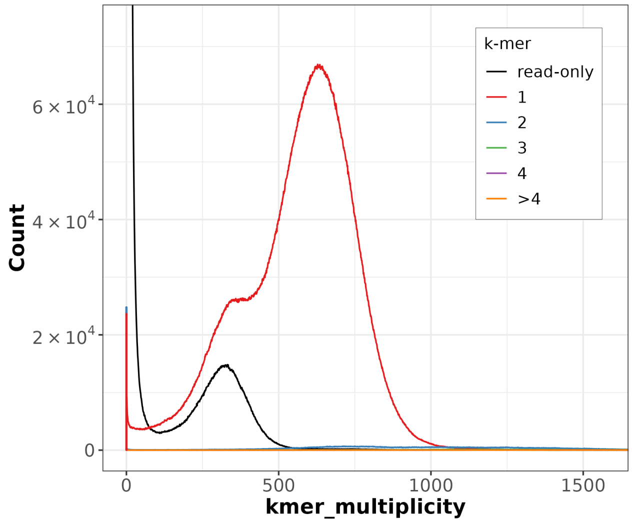
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: chlorophyta_odb12 (genomes:39, BUSCOs:1523)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: chlorophyta_odb12 (genomes:39, BUSCOs:1523)

HiC contact map of curated assembly

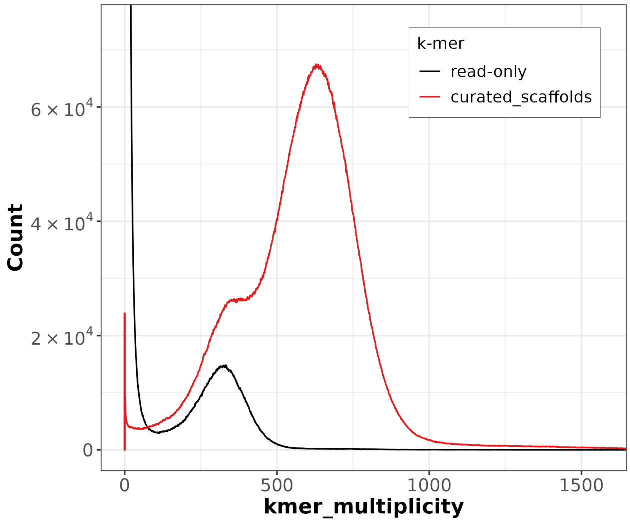


collapsed [LINK]

K-mer spectra of curated assembly

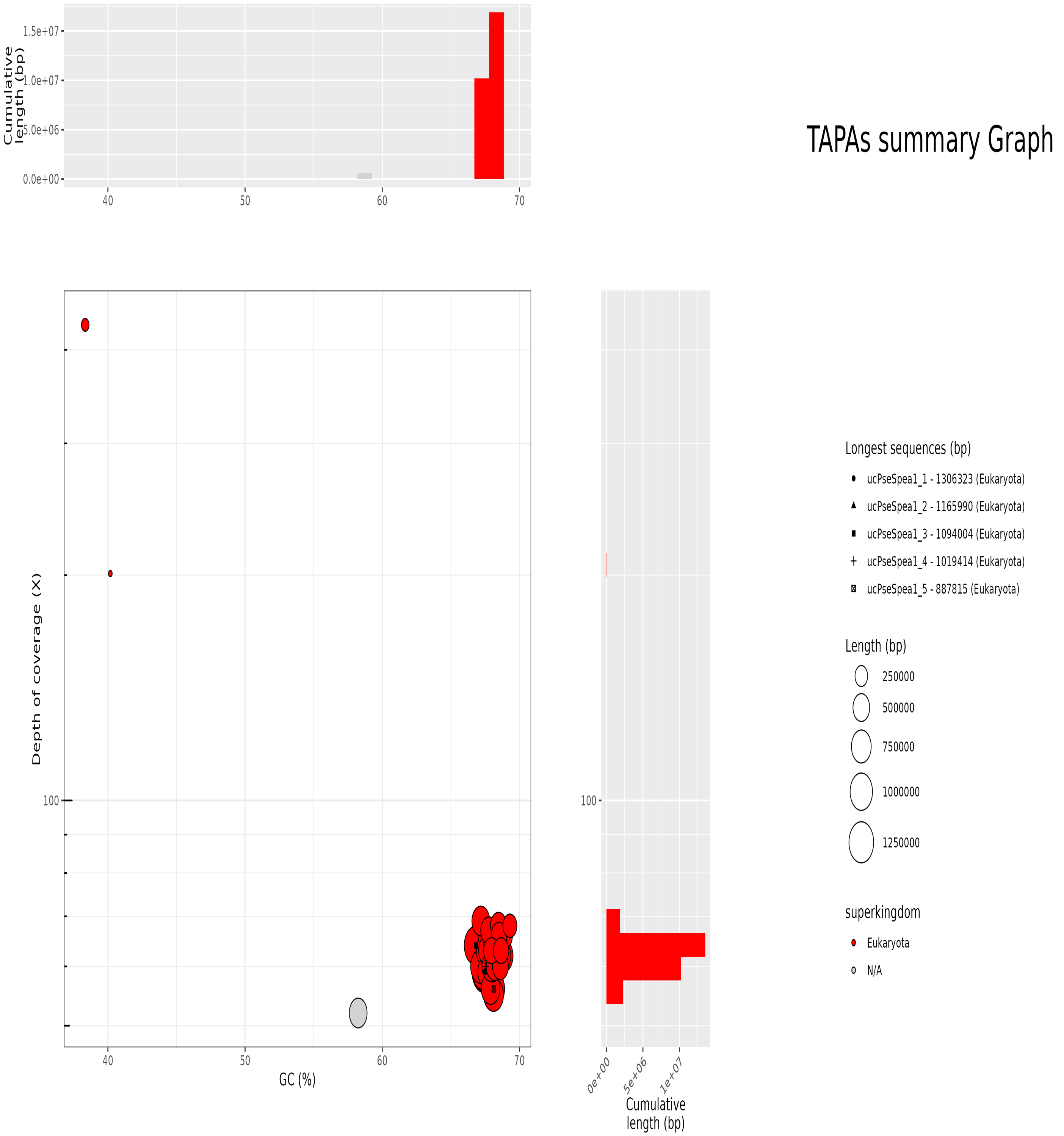


Distribution of k-mer counts per copy numbers found in asm



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	66	515

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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2025-12-04 16:05:44 CET