ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1266580
ToLID	**ooOscRubr1**
Species	Oscarella rubra
Class	Homoscleromorpha
Order	Homosclerophorida

Genome Traits	Expected	Observed
Haploid size (bp)	70,543,215	64,156,141
Haploid Number	5 (source: ancestor)	20
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.6.Q47

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

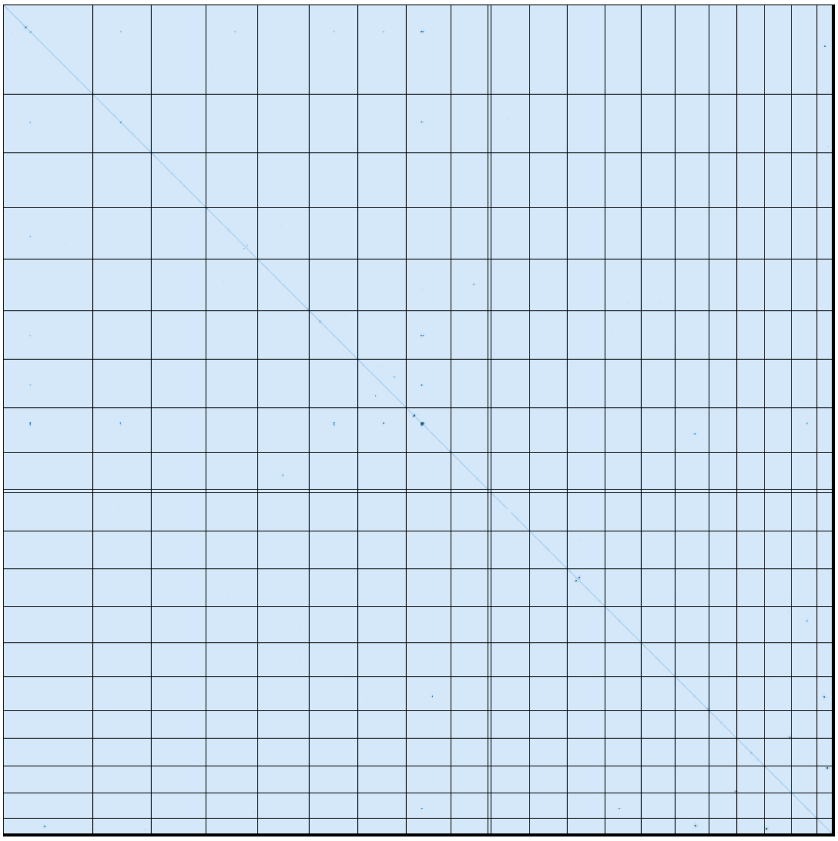
. Interventions/Gb: 0
. Contamination notes: ""
. Other observations: "The assembly of Oscarella rubra (ooOscRubr1) is based on 53X PacBio data and 163X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 297 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 14.37 Mb (with the largest being 5.145 Mb). Additionally, 22 regions totaling 1.45 Mb (with the largest being 0.208 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 30 contaminant sequences were removed, totaling 1.26Mb (with the largest being 0.207Mb). 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	65,392,040	64,156,141
GC %	46.21	46.4
Gaps/Gbp	0	0
Total gap bp	0	0
Scaffolds	55	26
Scaffold N50	3,448,099	3,448,099
Scaffold L50	8	8
Scaffold L90	18	17
Contigs	55	26
Contig N50	3,448,099	3,448,099
Contig L50	8	8
Contig L90	18	17
QV	45.5145	47.891
Kmer compl.	77.9231	77.857
BUSCO sing.	57.4%	84.1%
BUSCO dupl.	1.2%	2.1%
BUSCO frag.	25.1%	3.4%
BUSCO miss.	16.2%	10.4%

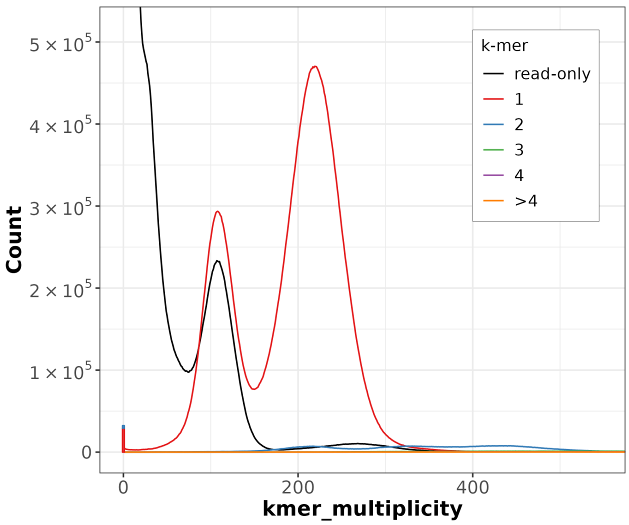
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

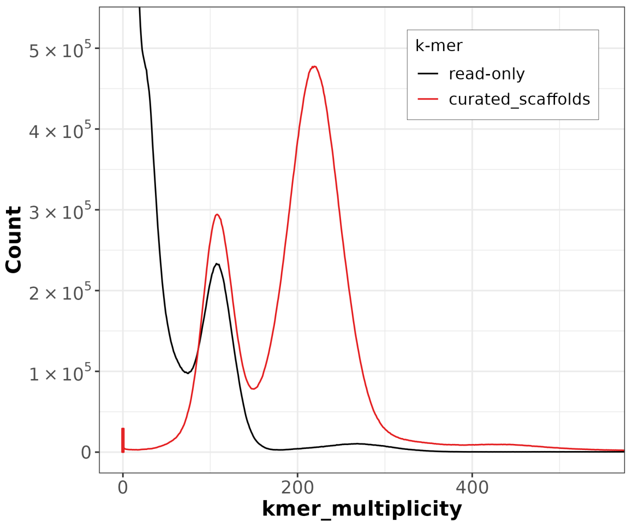


collapsed [LINK]

K-mer spectra of curated assembly

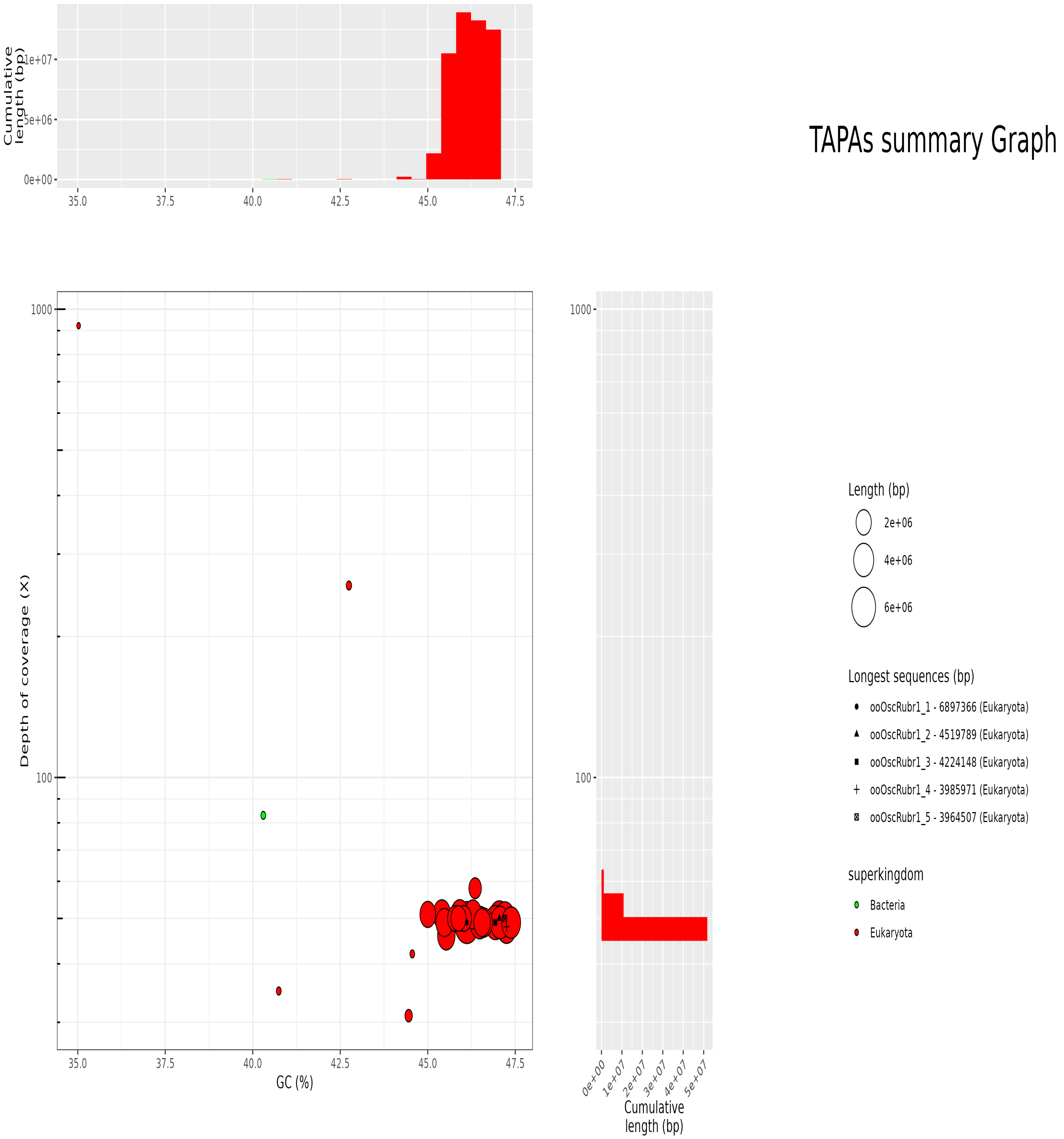


Distribution of k-mer counts per copy numbers found in asm



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	53	163

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: 2025-10-10 13:13:25 CEST