ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	137764
ToLID	**urOsmHybr1**
Species	Osmundea hybrida
Class	Florideophyceae
Order	Ceramiales

Genome Traits	Expected	Observed
Haploid size (bp)	589,987,396	795,532,140
Haploid Number	5 (source: ancestor)	28
Ploidy	4 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q43

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
. BUSCO single copy value is less than 90% for collapsed
. BUSCO duplicated value is more than 5% for collapsed
. Assembly length loss > 3% for collapsed

Curator notes

. Interventions/Gb: 438
. Contamination notes: ""
. Other observations: "The assembly of Osmundea hybrida (urOsmHybr1) is based on 58X PacBio data and 69X 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, 567 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 67.158 Mb (with the largest being 4.832 Mb). Additionally, 3722 regions totaling 741.053 Mb (with the largest being 5.167 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, 139 haplotypic regions, totaling

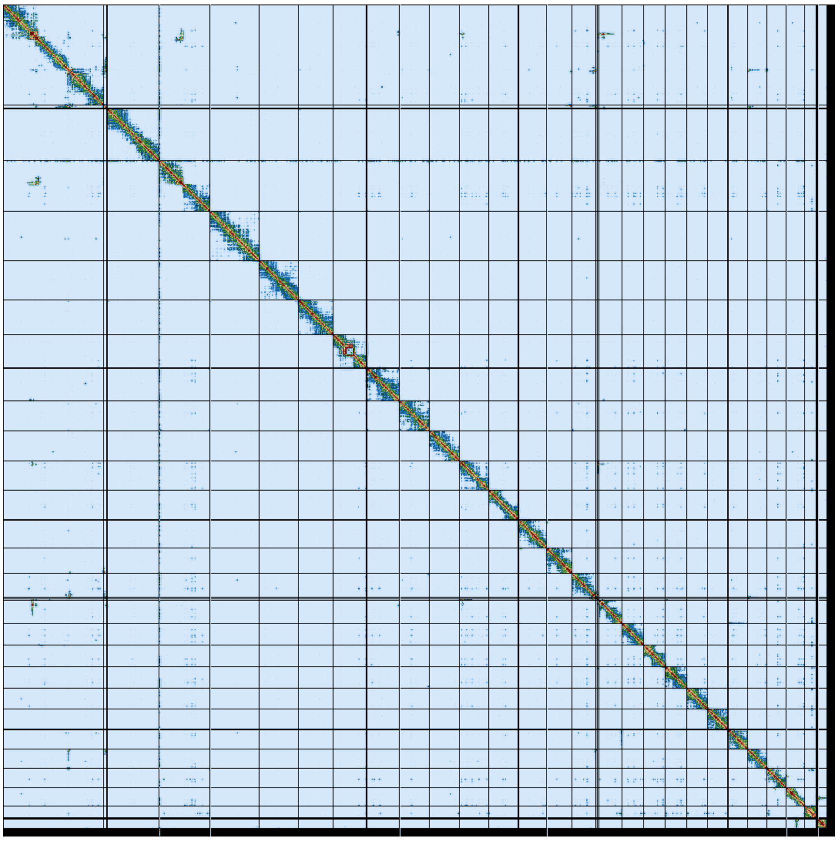
71.3Mb, (with the largest being 2.23Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	894,517,855	795,532,140
GC %	43.16	43.19
Gaps/Gbp	642.8	688.85
Total gap bp	57,500	66,100
Scaffolds	364	161
Scaffold N50	26,657,558	29,041,477
Scaffold L50	11	9
Scaffold L90	43	23
Contigs	939	709
Contig N50	1,766,460	2,133,222
Contig L50	165	117
Contig L90	491	374
QV	41.9129	43.2085
Kmer compl.	85.3553	84.0204
BUSCO sing.	61.2%	69.0%
BUSCO dupl.	13.3%	5.5%
BUSCO frag.	7.1%	7.1%
BUSCO miss.	18.4%	18.4%

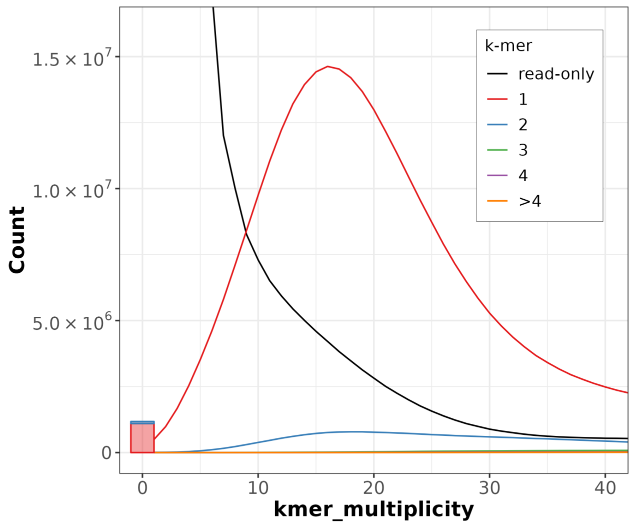
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: eukaryota_odb10 (genomes:70, BUSCOs:255)

HiC contact map of curated assembly

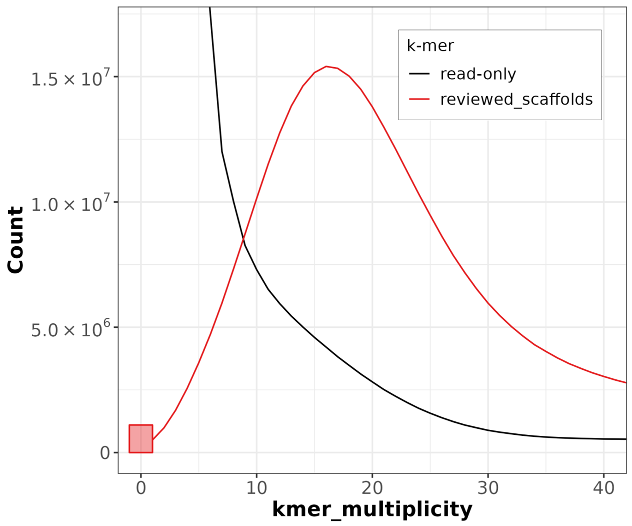


collapsed [LINK]

K-mer spectra of curated assembly

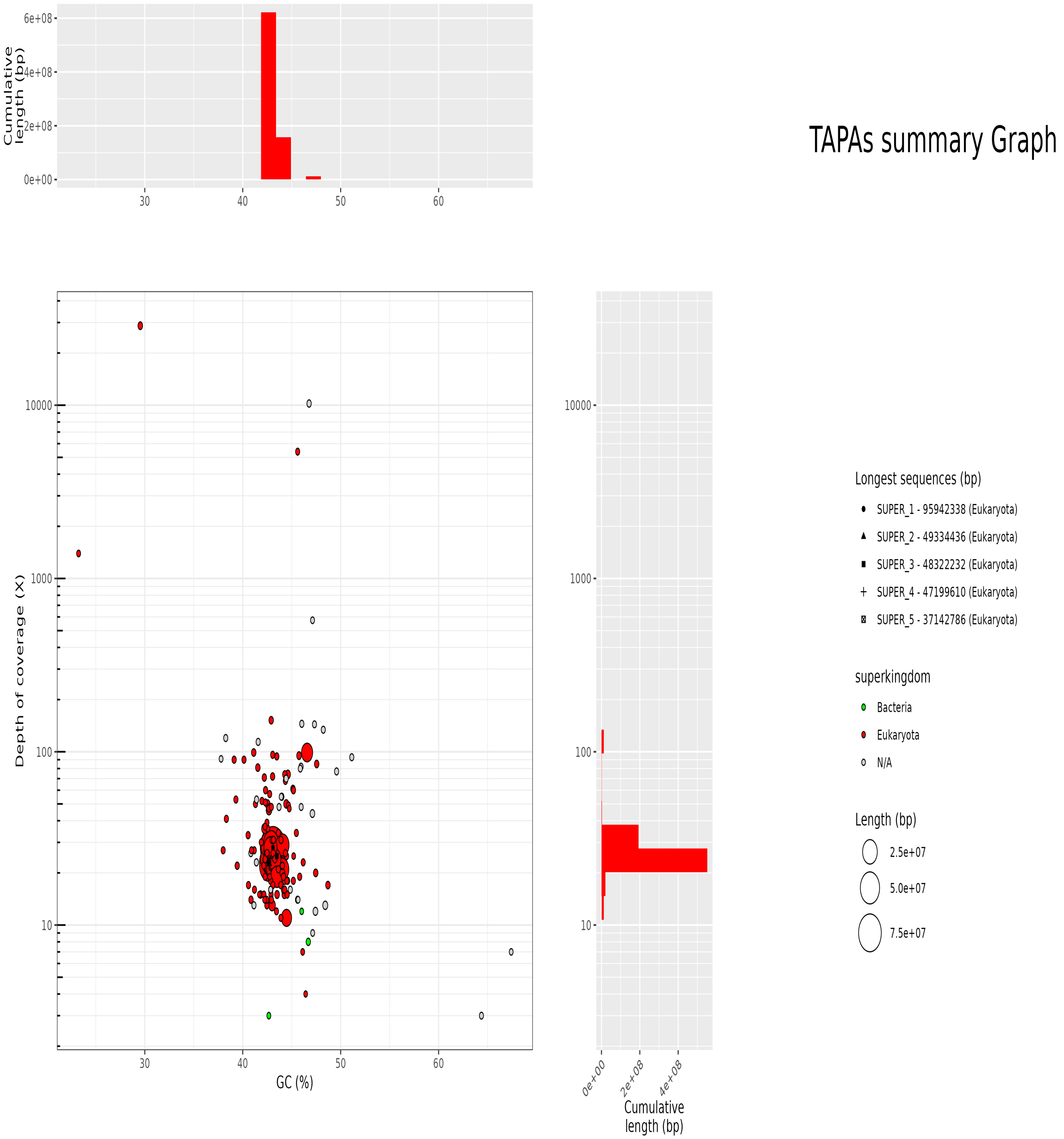


Distribution of k-mer counts per copy numbers found in asm



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	58	116

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: Benjamin Istace
Affiliation: Genoscope

Date and time: 2025-06-05 03:58:16 CEST