

ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	13608
ToLID	ucManSpeal
Species	Mantoniella squamata
Class	Mamiellophyceae
Order	Mamiellales

Genome Traits	Expected	Observed
Haploid size (bp)	331,181,783	61,339,814
Haploid Number	7 (source: ancestor)	18
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.6.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
- . Kmer completeness value is less than 90 for collapsed

Curator notes

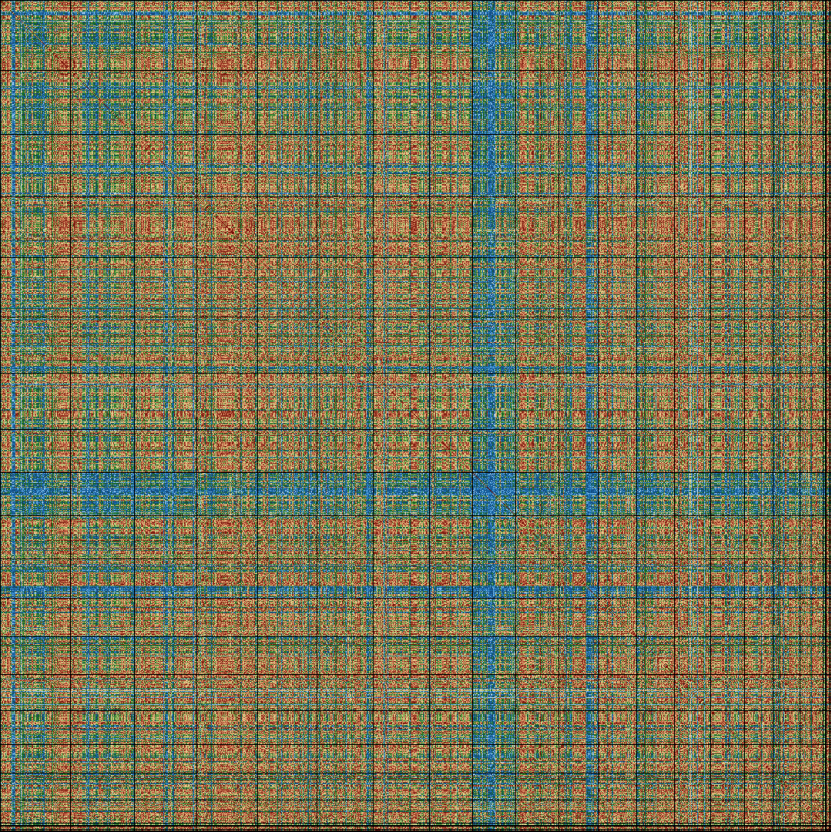
. Interventions/Gb: 0
. Contamination notes: ""
. Other observations: "The assembly of *Mantoniella squamata* (ucManSpeal) is based on 40X PACBIO data and 20X Arima Hi-C data generated as part of the ATLASea programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial 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, 199 contigs were identified as contaminants (bacterial), totaling 74.187 Mb (with the largest being 8.862 Mb). Additionally, 7 regions totaling 0.556 Mb (with the largest being 0.221 Mb) were identified as haplotypic duplications and removed. Mitochondrial and chloroplast genomes were assembled using OATK. A circular sequence of 135 Kb was obtained for the mitochondria. However, it is not a single copy. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 1 contaminant sequence of 0.318 Mb was removed. 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	61,785,136	61,339,814
GC %	58.43	58.42
Gaps/Gbp	80.93	81.51
Total gap bp	500	500
Scaffolds	26	24
Scaffold N50	4,134,351	4,134,351
Scaffold L50	7	7
Scaffold L90	16	16
Contigs	31	29
Contig N50	3,173,443	3,206,208
Contig L50	8	7
Contig L90	18	18
QV	43.9072	43.8834
Kmer compl.	72.1727	71.7606
BUSCO sing.	96.6%	96.7%
BUSCO dupl.	1.6%	1.6%
BUSCO frag.	0.5%	0.5%
BUSCO miss.	1.3%	1.3%

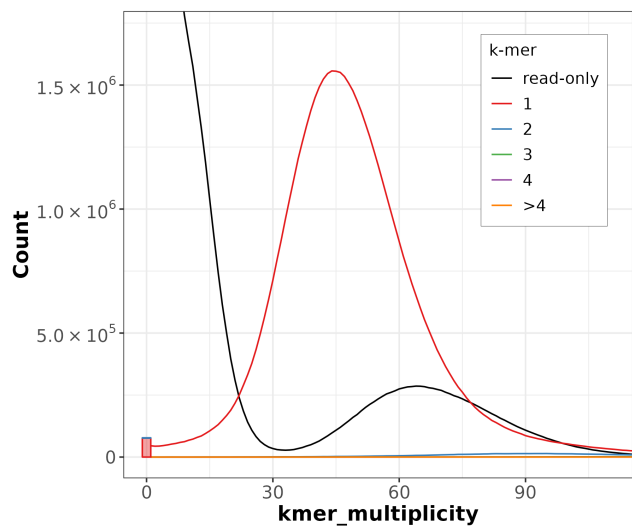
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: chlorophyta_odb12 (genomes:39, BUSCOs:1523)

HiC contact map of curated assembly

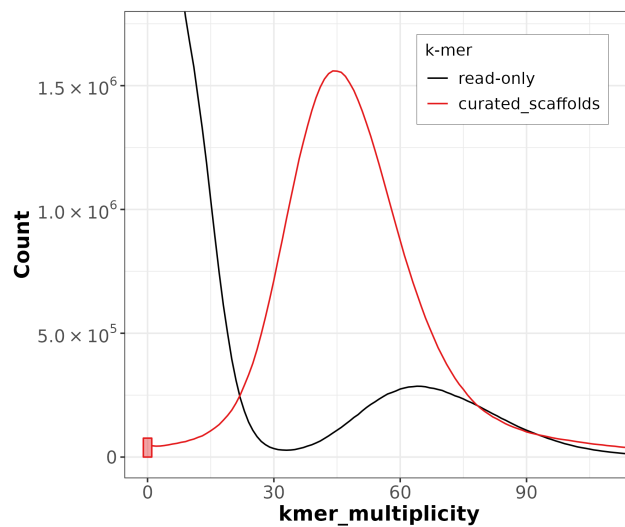


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

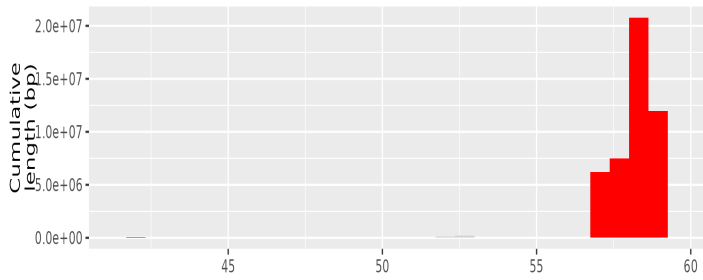


Distribution of k-mer counts per copy numbers found in asm

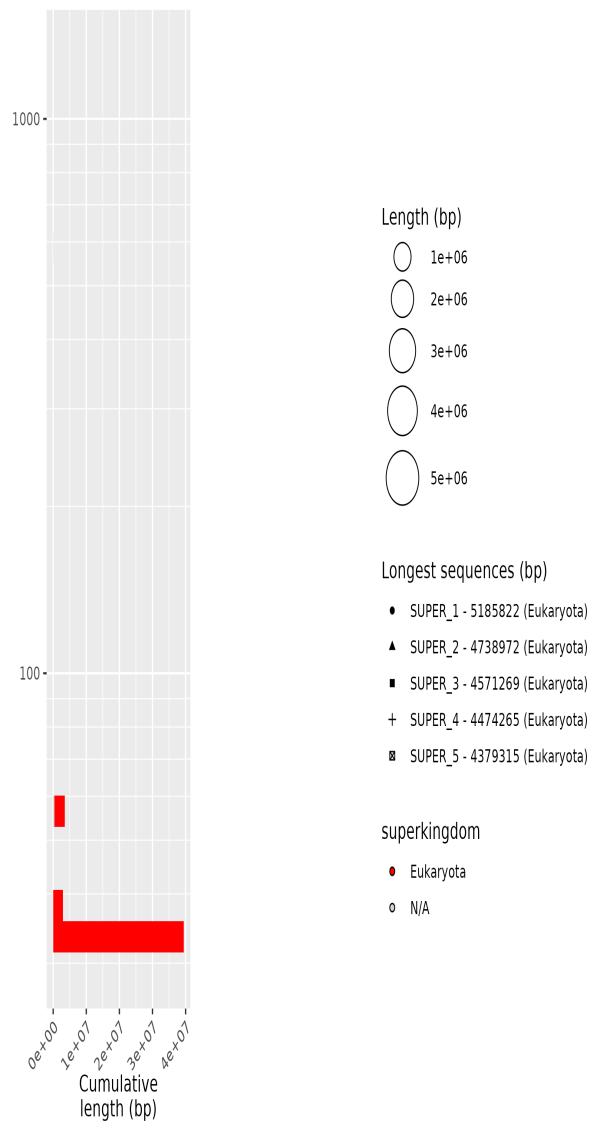
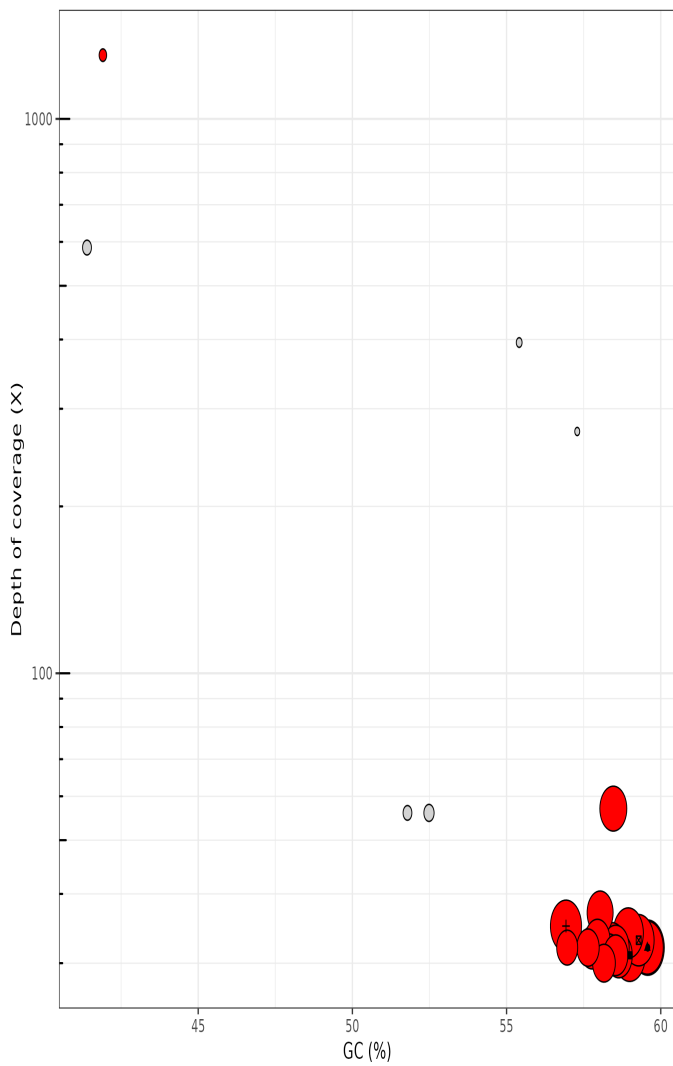


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	40	20

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: Adama Ndar
Affiliation: Genoscope

Date and time: 2025-11-12 18:01:24 CET