ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	296587
ToLID	**ucMicComm1**
Species	Micromonas commoda
Class	Mamiellophyceae
Order	Mamiellales

Genome Traits	Expected	Observed
Haploid size (bp)	29,348,810	21,428,823
Haploid Number	7 (source: ancestor)	17
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.6.Q50

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

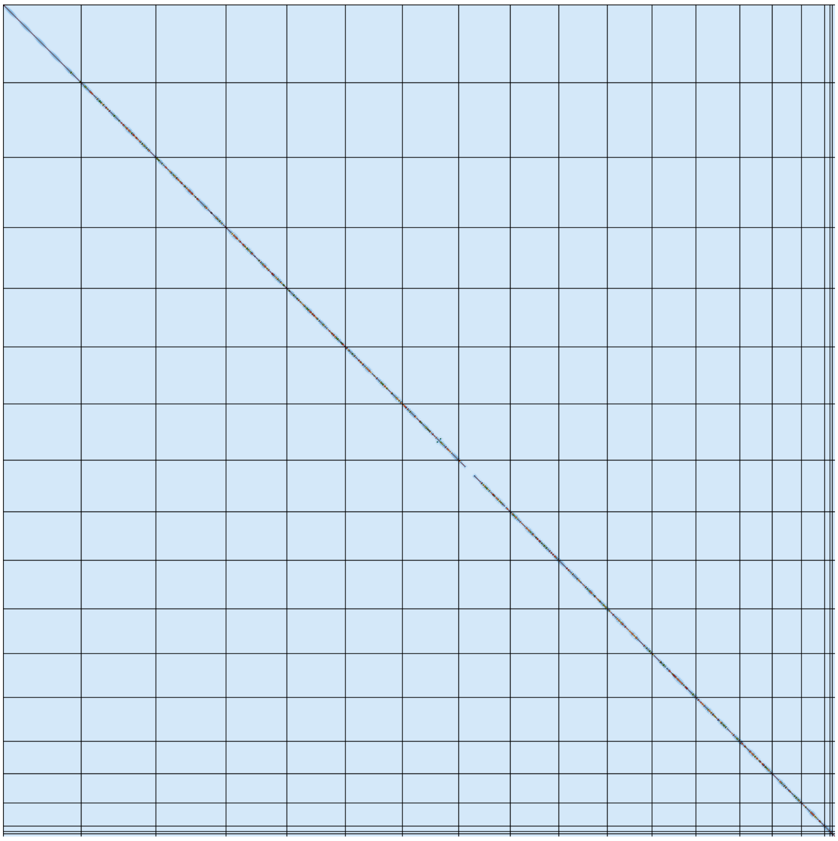
. Interventions/Gb: 46
. Contamination notes: ""
. Other observations: "The assembly of Micromonas commoda (ucMicComm1) is based on 87X PacBio data and 311X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 111 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 19.336 Mb (with the largest being 3.764 Mb). Additionally, 3 regions totaling 0.12 Mb (with the largest being 0.075 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. 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	21,736,753	21,428,823
GC %	61.72	62.06
Gaps/Gbp	0	46.67
Total gap bp	0	200
Scaffolds	20	19
Scaffold N50	1,325,195	1,446,641
Scaffold L50	7	7
Scaffold L90	15	14
Contigs	20	20
Contig N50	1,325,195	1,325,195
Contig L50	7	7
Contig L90	15	14
QV	50.5521	50.5195
Kmer compl.	59.3701	59.3698
BUSCO sing.	97.4%	97.4%
BUSCO dupl.	0.8%	0.8%
BUSCO frag.	0.6%	0.6%
BUSCO miss.	1.2%	1.2%

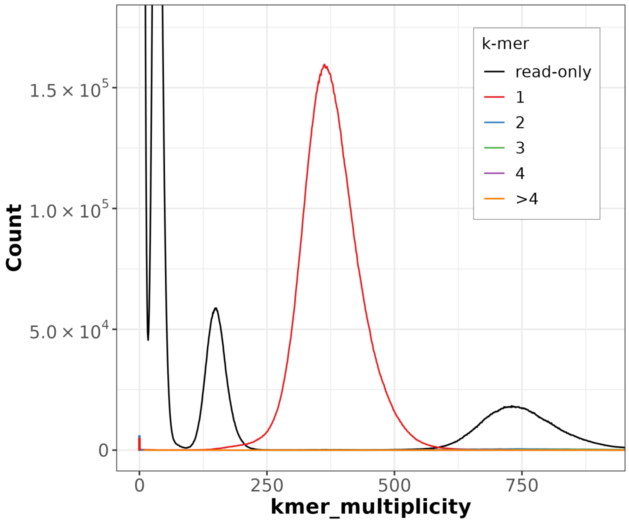
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: chlorophyta_odb12 (genomes:39, BUSCOs:1523)

HiC contact map of curated assembly

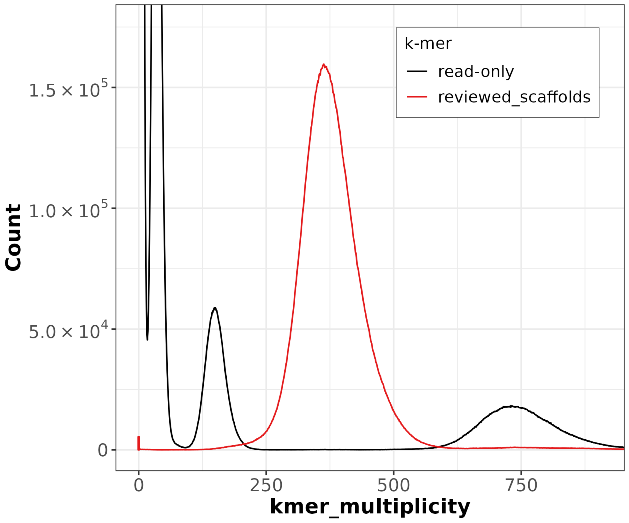


collapsed [LINK]

K-mer spectra of curated assembly

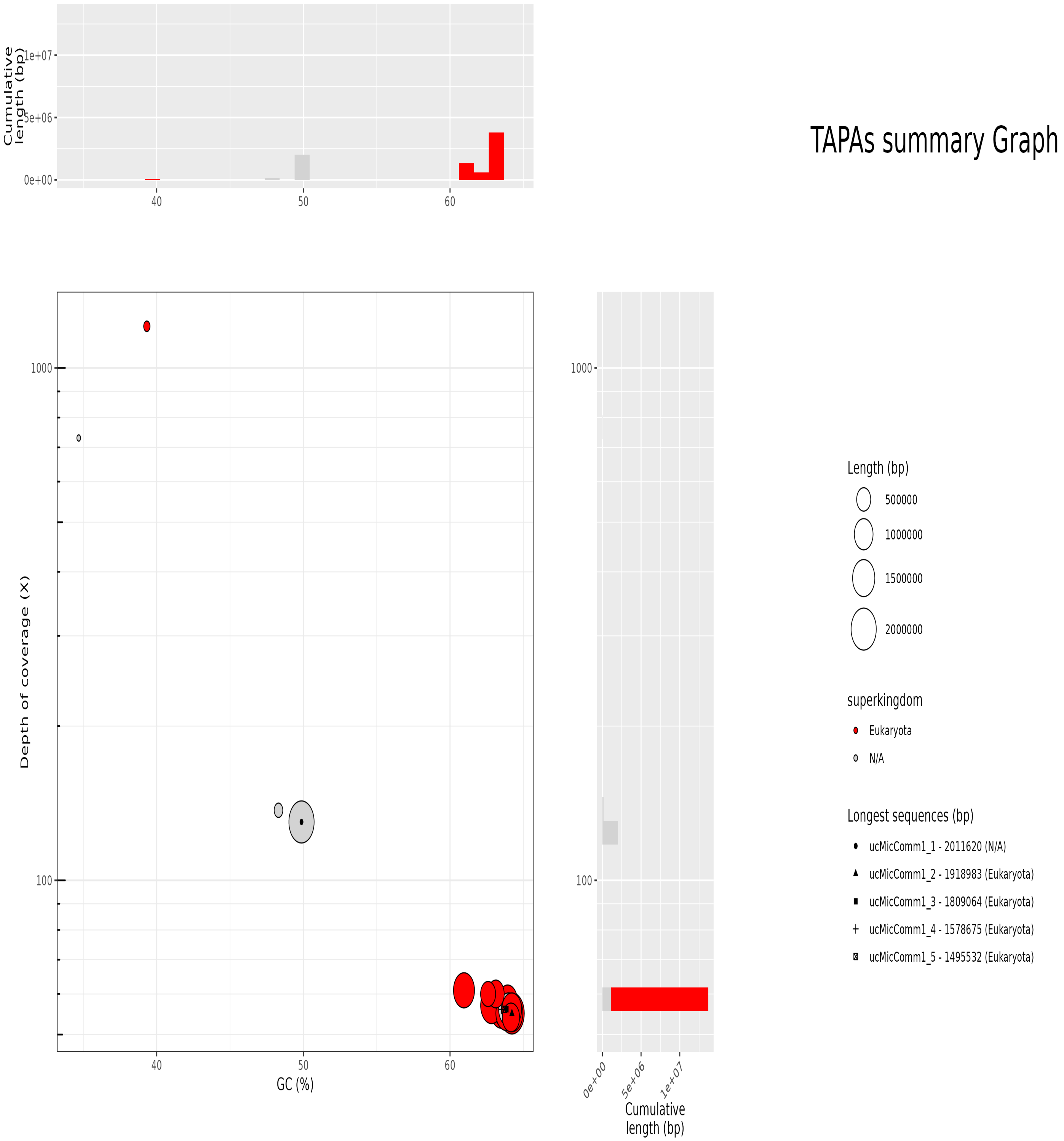


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	87	311

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: 2025-11-27 22:39:39 CET