ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	296588
ToLID	**ucMicBrav1**
Species	Micromonas bravo
Class	Mamiellophyceae
Order	Mamiellales

Genome Traits	Expected	Observed
Haploid size (bp)	11,061,724	21,493,393
Haploid Number	7 (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.Q48

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

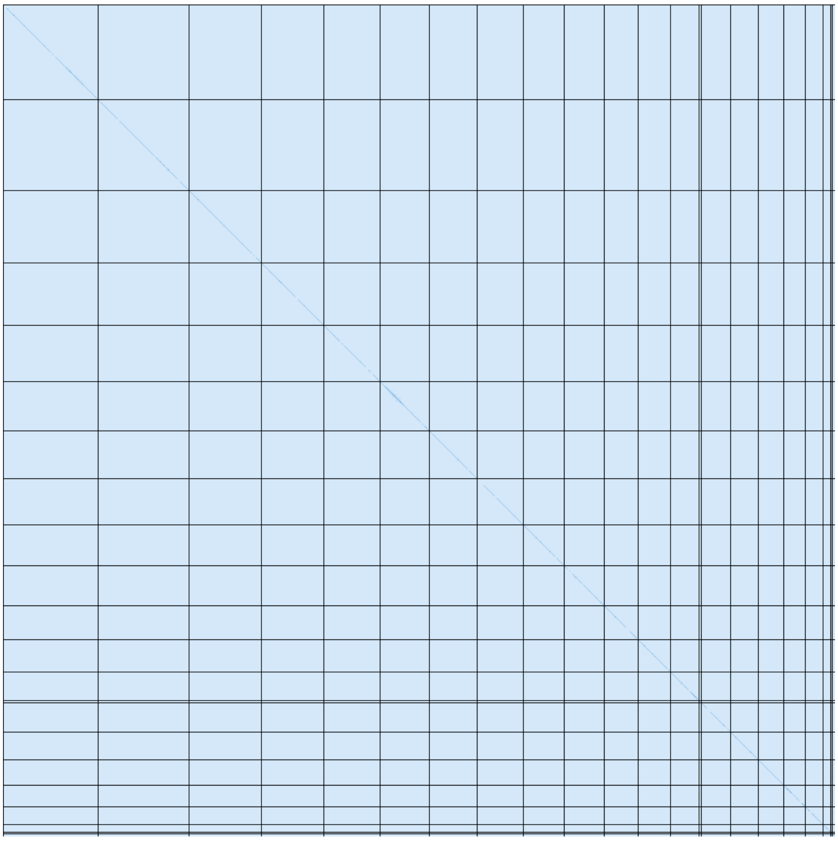
. Interventions/Gb: 278
. Contamination notes: ""
. Other observations: "The assembly of Micromonas bravo str. RCC418 (ucMicBrav1) is based on 560X PacBio data and Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included the following steps: initial 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, 93 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 8.8 Mb (with the largest being 3.1Mb). Additionally, 64 regions totaling 1.7 Mb (with the largest being 81 Kb) were identified as haplotypic duplications and removed. The mitochondrial and chloroplastic genomes were assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 5 haplotypic regions were removed,totaling 408 Kb (with the largest being 154 Kb. 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	21,953,997	21,493,393
GC %	68.16	68.28
Gaps/Gbp	227.75	418.73
Total gap bp	500	1,400
Scaffolds	30	22
Scaffold N50	1,230,736	1,271,127
Scaffold L50	6	6
Scaffold L90	17	15
Contigs	35	31
Contig N50	1,196,000	1,196,000
Contig L50	7	7
Contig L90	18	17
QV	46.6776	48.3141
Kmer compl.	86.0272	85.6016
BUSCO sing.	93.8%	94.8%
BUSCO dupl.	2.6%	1.6%
BUSCO frag.	2.4%	2.4%
BUSCO miss.	1.2%	1.2%

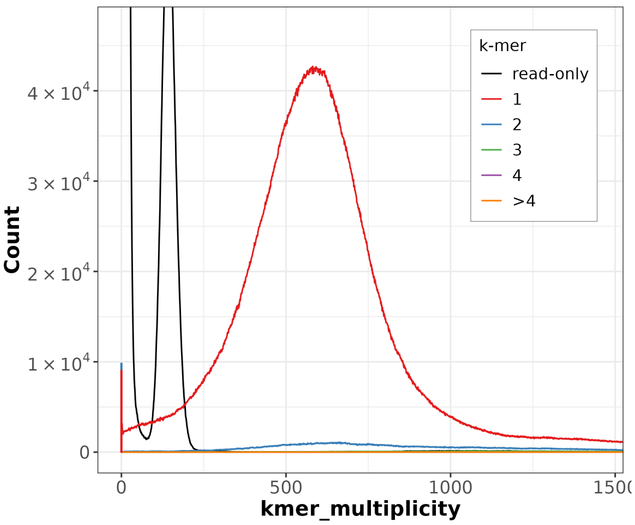
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: chlorophyta_odb12 (genomes:39, BUSCOs:1523)

HiC contact map of curated assembly

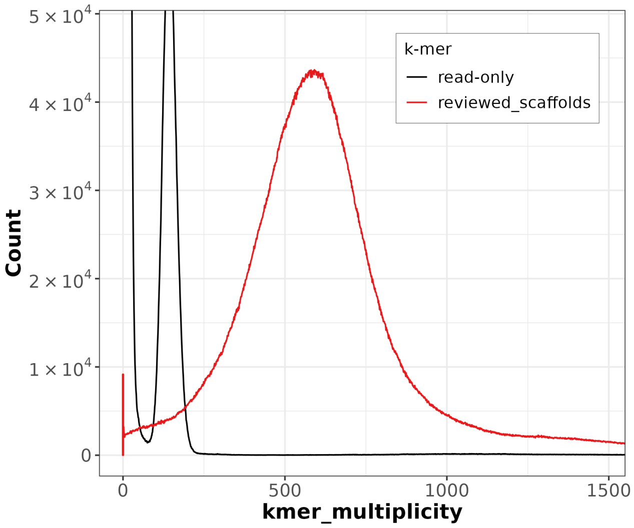


collapsed [LINK]

K-mer spectra of curated assembly

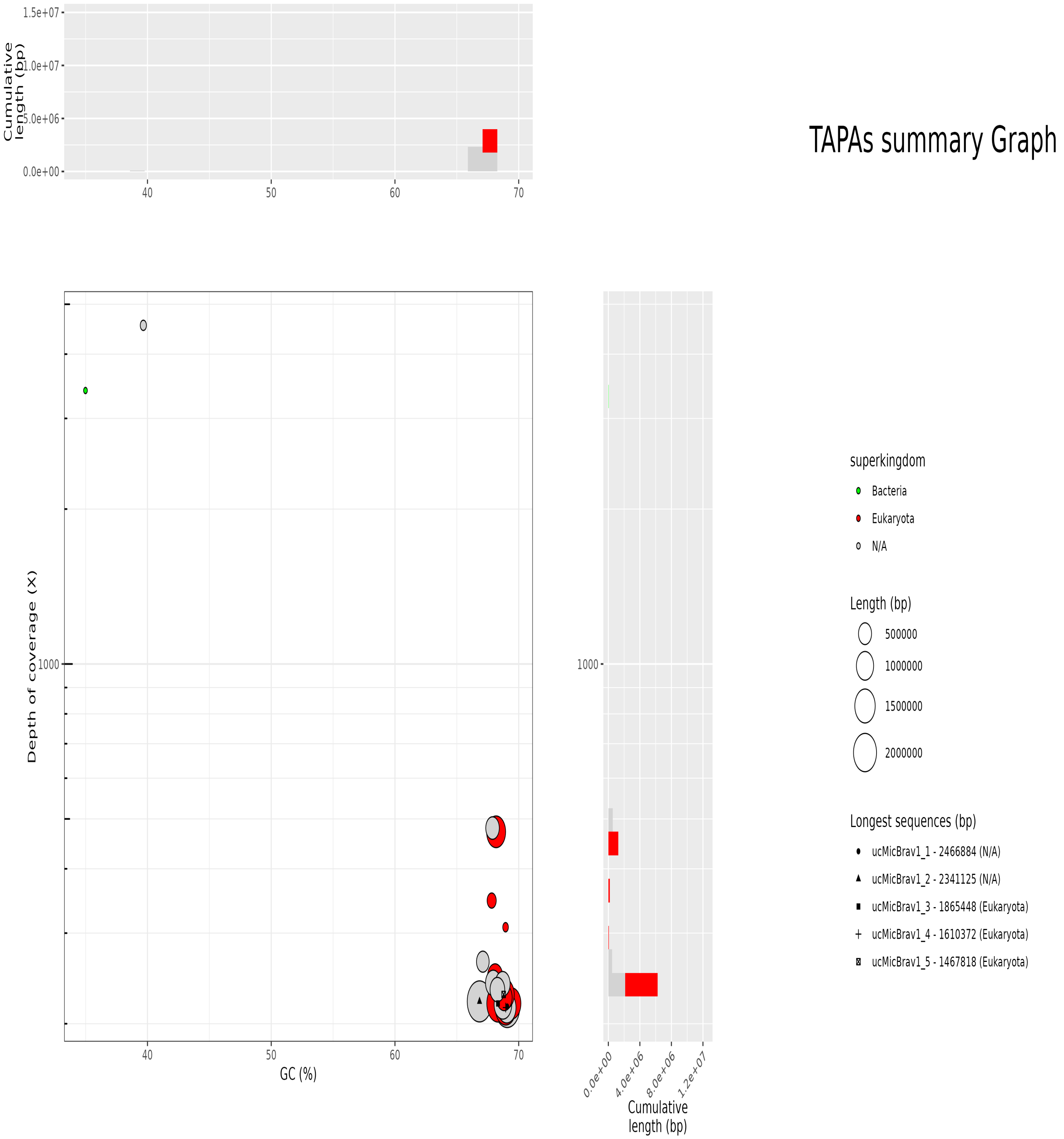


Distribution of k-mer counts per copy numbers found in asm



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	PACBIO Hifi	Arima
Coverage	560	383

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: Caroline Menguy
Affiliation: Genoscope

Date and time: 2025-08-01 23:03:37 CEST