

# ERGA Assembly Report

v24.04.03\_beta

Tags: ERGA-BGE

TxID	1518524
ToLID	<b>qqMacCret1</b>
Species	Macrothele cretica
Class	Arachnida
Order	Araneae

Genome Traits	Expected	Observed
Haploid size (bp)	3,736,835,300	3,973,420,169
Haploid Number	23 (source: ancestor)	47
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q58

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

- . Observed Haploid Number is different from Expected

### Curator notes

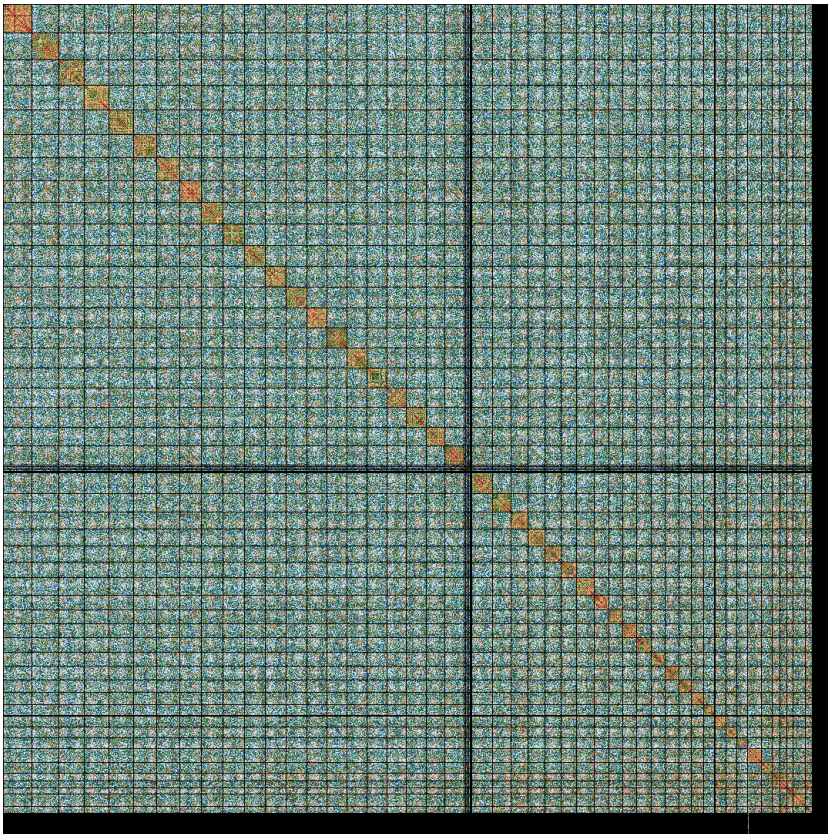
- . Interventions/Gb: 17
- . Contamination notes: "2 bacterial contigs removed after assembly "
- . Other observations: "11 scaffolds with lower coverage, maybe sexual chromosomes since mygalomorphs seem to exhibit more than two X chromosomes. Presence of 6 small contigs that seem to only interact with those low coverage scaffolds and were tagged as unloc to scaffold SUPER\_21 "

# Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	3,977,790,240	3,973,420,169
GC %	40.57	40.57
Gaps/Gbp	562.88	569.79
Total gap bp	223,900	230,500
Scaffolds	733	710
Scaffold N50	90,425,106	94,972,123
Scaffold L50	19	19
Scaffold L90	44	41
Contigs	2,972	2,974
Contig N50	3,118,366	3,118,366
Contig L50	368	368
Contig L90	1,354	1,354
QV	58.205	58.2019
Kmer compl.	96.3213	96.3133
BUSCO sing.	96.1%	96.1%
BUSCO dupl.	2.7%	2.7%
BUSCO frag.	1.2%	1.2%
BUSCO miss.	0.0%	0.0%

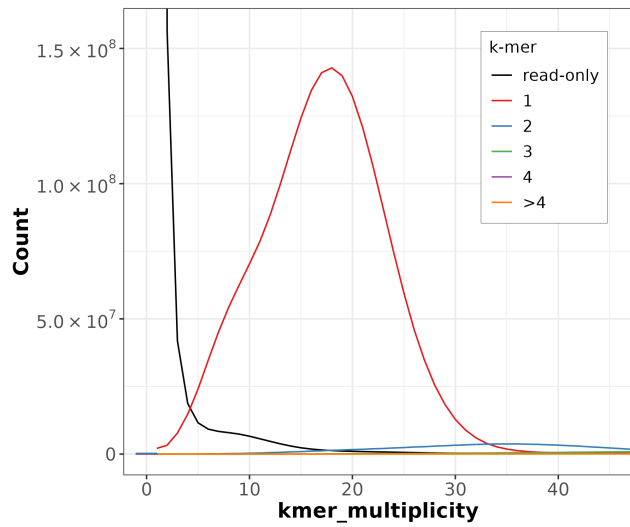
BUSCO 5.4.3 Lineage: eukaryota\_odb10 (genomes:70, BUSCOs:255)

# HiC contact map of curated assembly

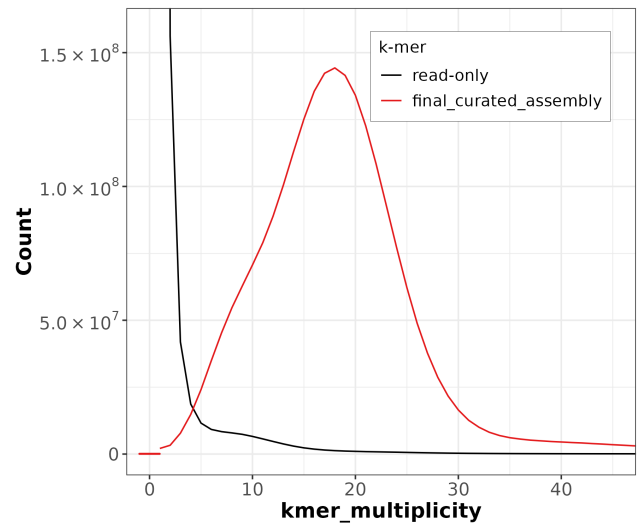


collapsed [\[LINK\]](#)

# K-mer spectra of curated assembly

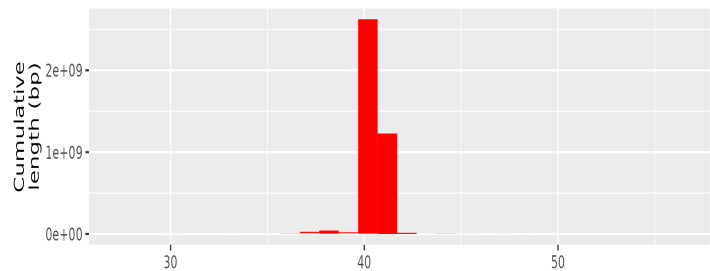


Distribution of k-mer counts per copy numbers found in asm



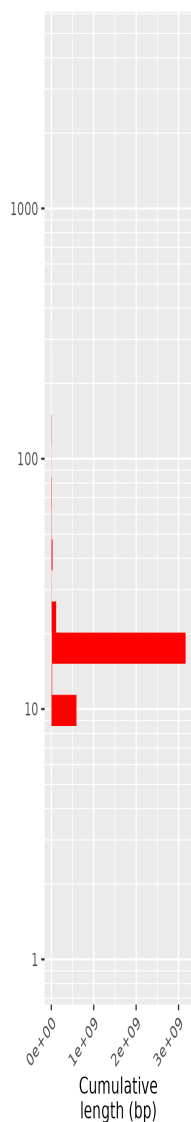
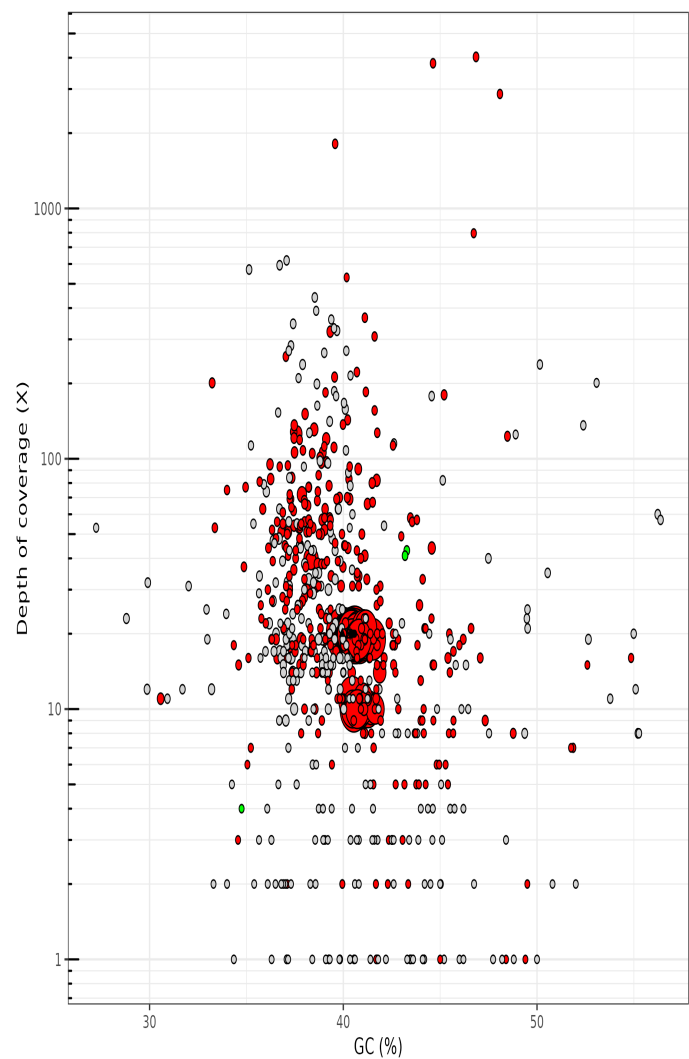
Distribution of k-mer counts coloured by their presence in reads/assemblies

# Post-curation contamination screening



## TAPAs summary Graph

(7 0X contigs have been hidden)



### Longest sequences (bp)

- SUPER\_1 - 138909147 (Eukaryota)
- ▲ SUPER\_2 - 128092427 (Eukaryota)
- SUPER\_3 - 120350022 (Eukaryota)
- + SUPER\_4 - 119240569 (Eukaryota)
- ▣ SUPER\_5 - 117438305 (Eukaryota)

### Length (bp)

- 5e+07
- 1e+08

### superkingdom

- Bacteria
- Eukaryota
- N/A

**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 (4-enz)
Coverage	62	15

## Assembly pipeline

- **Hifiasm**
  - |\_ *ver*: 0.19.5-r593
  - |\_ *key param*: NA
- **purge\_dups**
  - |\_ *ver*: 1.2.5
  - |\_ *key param*: NA

## Curation 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

Submitter: Emilie Teodori

Affiliation: Genoscope

Date and time: 2024-09-04 16:15:46 CEST