

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	1518524
ToLID	qqMacCret1
Species	Macrothele cretica
Class	Arachnida
Order	Araneae

Genome Traits	Expected	Observed
Haploid size (bp)	3,736,835,300	3,980,132,586
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.Q62

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

- . Observed Haploid Number is different from Expected
- . BUSCO single copy value is less than 90% for collapsed
- . BUSCO duplicated value is more than 5% for collapsed

Curator notes

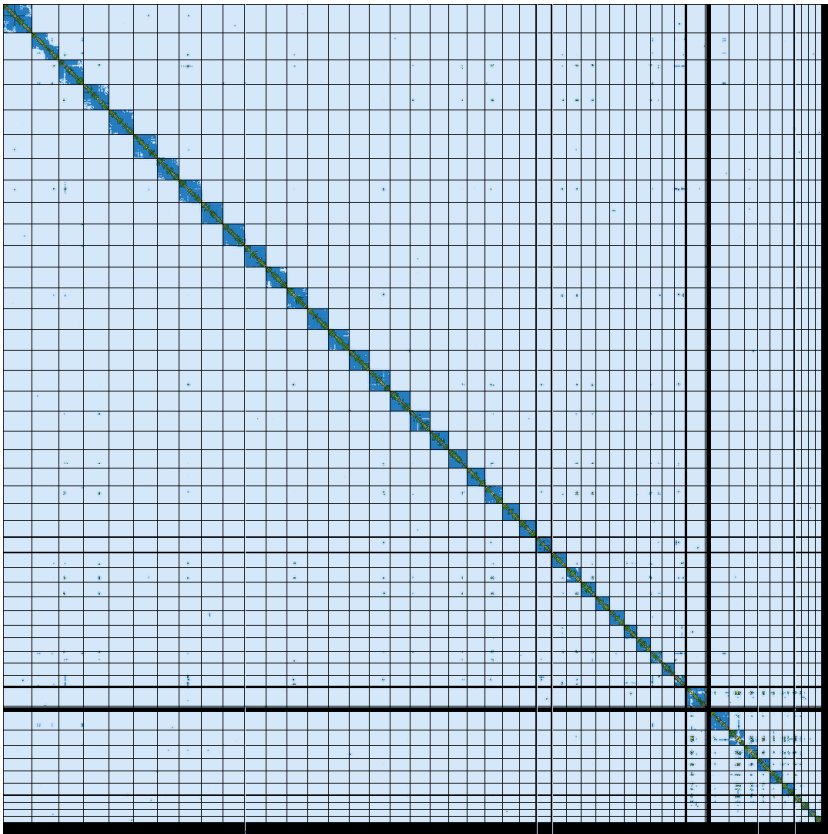
. Interventions/Gb: 44
. Contamination notes: ""
. Other observations: "The assembly of *Macrothele cretica* (qqMacCret1) is based on 19.7X PacBio data and 92X Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA, <https://www.erga-biodiversity.eu/>) via the Biodiversity Genomics Europe project (BGE, <https://biodiversitygenomics.eu/>). The assembly process included the following steps: initial PacBio assembly generation with Hifiiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 4 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.613 Mb (with the largest being 1.279 Mb). Additionally, 1161 regions totaling 239.165 Mb (with the largest being 8.15 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. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. Sexual chromosomes renamed X1 to X11 according to length. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	3,981,048,477	3,980,132,586
GC %	40.58	40.58
Gaps/Gbp	570.96	585.16
Total gap bp	227,300	244,500
Scaffolds	961	643
Scaffold N50	94,552,240	95,622,158
Scaffold L50	19	19
Scaffold L90	43	40
Contigs	3,234	2,972
Contig N50	3,255,254	3,255,254
Contig L50	358	358
Contig L90	1,306	1,307
QV	62.8509	62.8643
Kmer compl.	96.3119	96.3003
BUSCO sing.	89.5%	89.5%
BUSCO dupl.	6.7%	6.7%
BUSCO frag.	1.6%	1.6%
BUSCO miss.	2.2%	2.2%

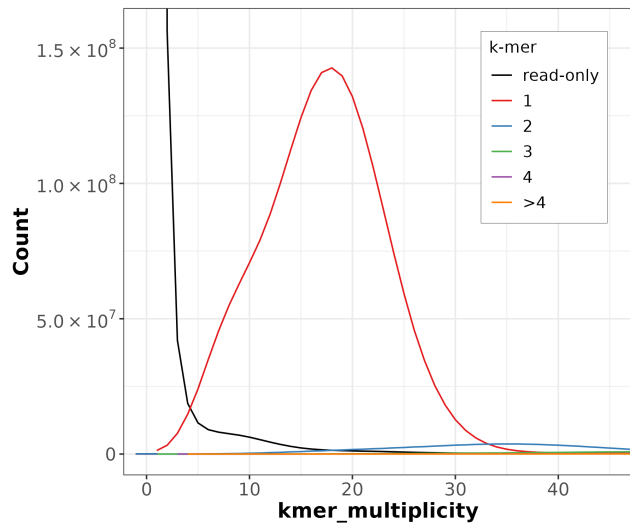
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: arachnida_odb10 (genomes:10, BUSCOs:2934)

HiC contact map of curated assembly

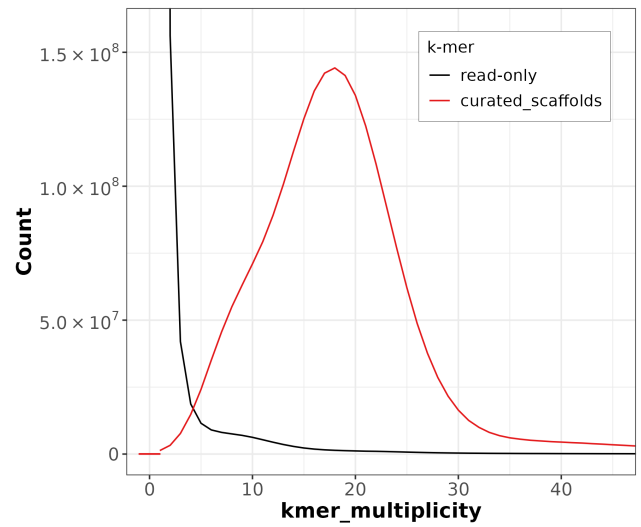


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

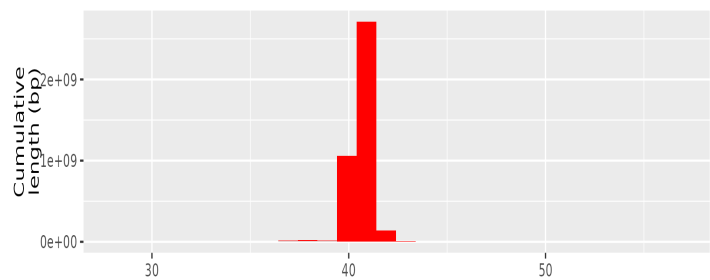


Distribution of k-mer counts per copy numbers found in asm

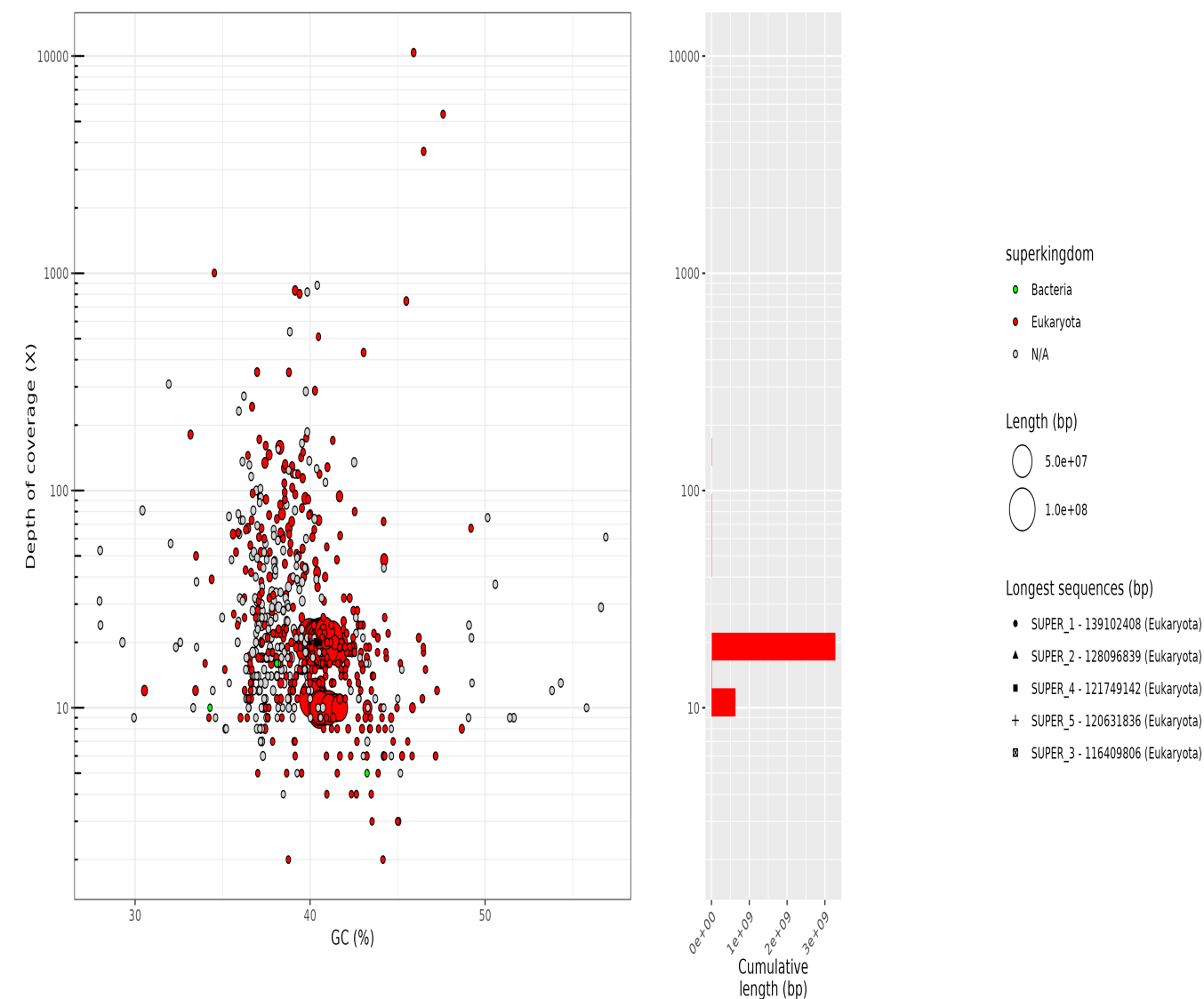


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	PACBIO Hifi	Arima
Coverage	19	97

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: Emilie Teodori

Affiliation: Genoscope

Date and time: 2025-02-16 00:42:24 CET