ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	612071
ToLID	**iyMesOert5**
Species	Messor oertzeni
Class	Insecta
Order	Hymenoptera

Genome Traits	Expected	Observed
Haploid size (bp)	313,803,573	321,605,597
Haploid Number	21 (source: ancestor)	21
Ploidy	1 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q61

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

. Observed Ploidy is different from Expected

Curator notes

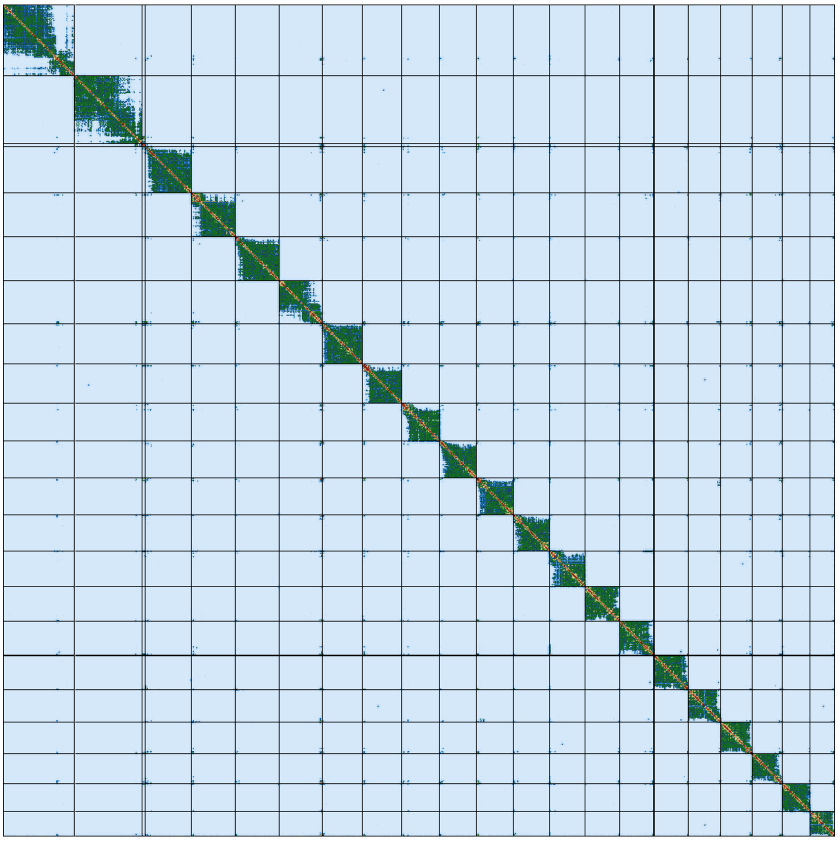
. Interventions/Gb: 21
. Contamination notes: ""
. Other observations: "The assembly of Messor oertzeni (iyMesOert5.1) is based on 42X PacBio data and 242X of Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA, https://www.erga-biodiversity.eu/) via the Biodiversity Genomics Europe project (BGE, https://biodiversitygenomics.eu/). The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 7 contigs were identified as contaminants (bacterial), totaling 0.410 Mb (with the largest being 0.103 Mb). Additionally, 66 regions totaling 6.311 Mb (with the largest being 0.524 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. During manual curation, 7 haplotypic regions were removed, totaling 3 Mb (with the largest being 0.914 Mb). 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	324,621,205	321,605,597
GC %	36.65	36.63
Gaps/Gbp	877.95	873.74
Total gap bp	28,700	28,900
Scaffolds	39	32
Scaffold N50	14,872,299	14,831,354
Scaffold L50	9	9
Scaffold L90	19	19
Contigs	321	313
Contig N50	1,630,400	1,636,165
Contig L50	58	57
Contig L90	188	186
QV	61.6045	61.579
Kmer compl.	98.5159	98.4277
BUSCO sing.	96.1%	96.1%
BUSCO dupl.	0.3%	0.3%
BUSCO frag.	0.6%	0.6%
BUSCO miss.	3.0%	3.0%

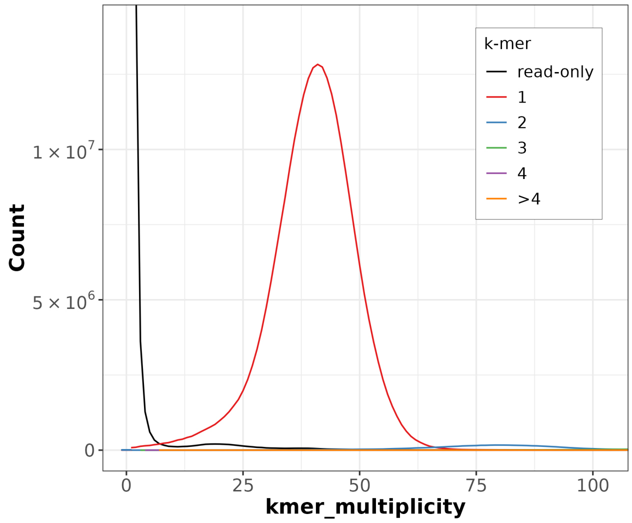
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: formicidae_odb12 (genomes:24, BUSCOs:7266)

HiC contact map of curated assembly

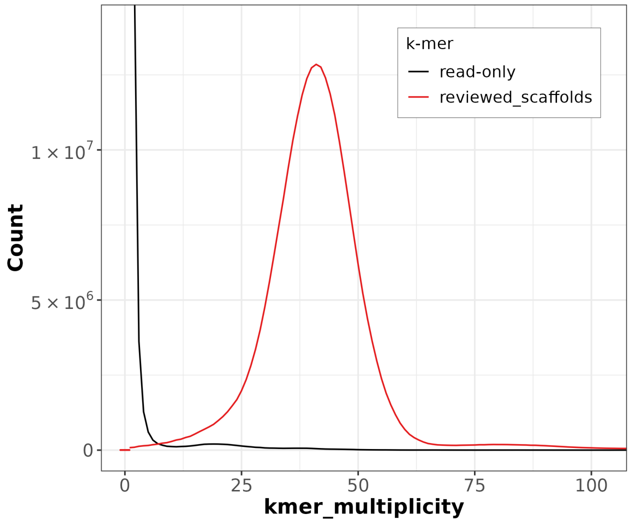


collapsed [LINK]

K-mer spectra of curated assembly

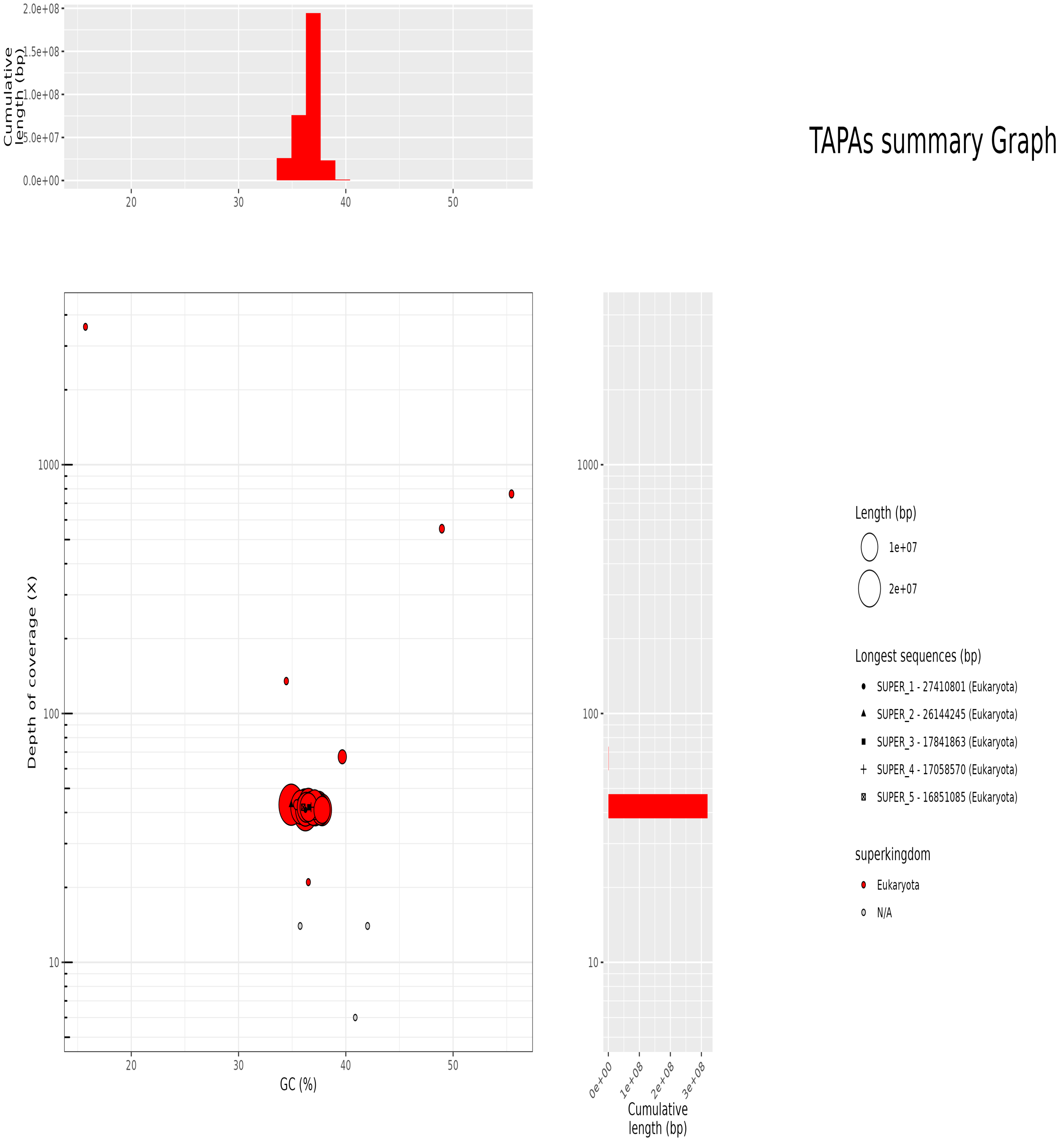


Distribution of k-mer counts per copy numbers found in asm



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	42	242

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: 2025-09-10 23:49:17 CEST