ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	612121
ToLID	**iyMesWasm6**
Species	Messor wasmanni
Class	Insecta
Order	Hymenoptera

Genome Traits	Expected	Observed
Haploid size (bp)	298,579,337	326,142,122
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: 5.7.Q60

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

. Observed Ploidy is different from Expected

. More than 1000 gaps/Gbp for collapsed

Curator notes

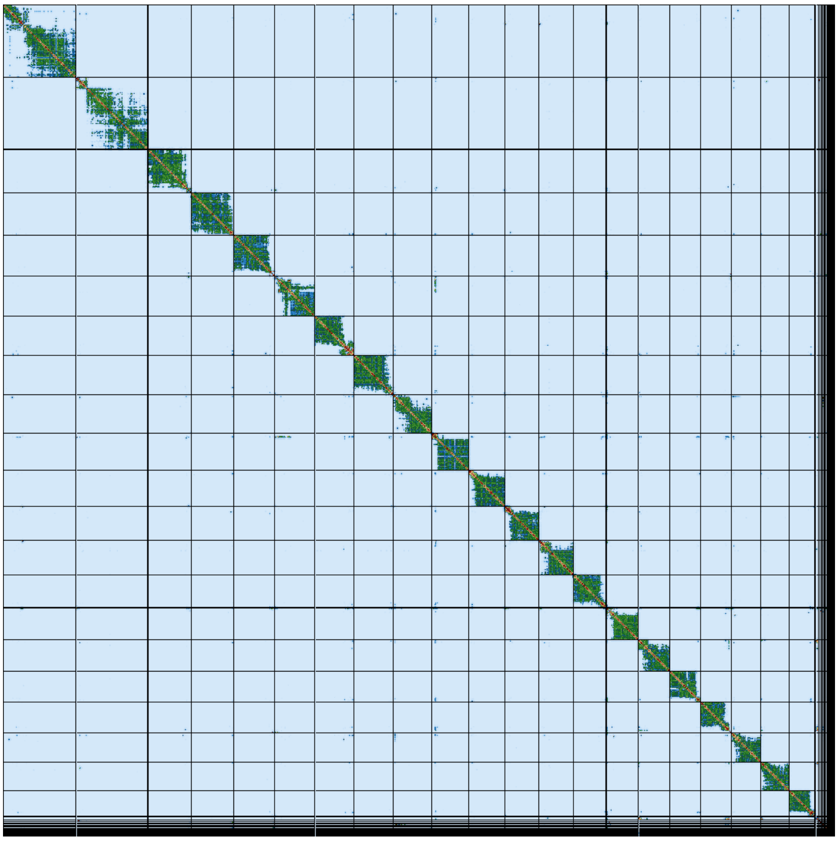
. Interventions/Gb: 138
. Contamination notes: ""
. Other observations: ""The assembly of Messor wasmanni (iyMesWasm6) is based on 41X PacBio data and 234X Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA, https://www.erga-biodiversity.eu/) via the Biodiversity Genomics 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, 2 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.062 Mb (with the largest being 0.045 Mb). Additionally, 144 regions totaling 6.464 Mb (with the largest being 0.254 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, 4 haplotypic regions and 5 contaminant sequences were removed, totaling 0.478 Mb and 0.37 Mb respectively, (with the largest being 0.2 Mb and 0.284 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	327,658,177	326,142,122
GC %	36.28	36.28
Gaps/Gbp	1,663.32	1,707.84
Total gap bp	54,800	58,800
Scaffolds	109	78
Scaffold N50	13,841,405	14,968,348
Scaffold L50	10	9
Scaffold L90	20	19
Contigs	651	635
Contig N50	924,285	929,390
Contig L50	109	108
Contig L90	361	360
QV	60.1421	60.2357
Kmer compl.	97.2601	97.234
BUSCO sing.	96.2%	99.0%
BUSCO dupl.	0.3%	0.4%
BUSCO frag.	0.6%	0.2%
BUSCO miss.	2.9%	0.4%

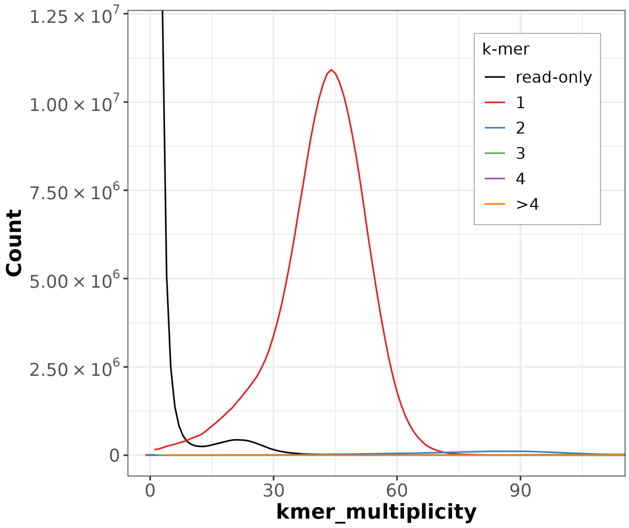
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: formicidae_odb12 (genomes:24, BUSCOs:7266)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: formicidae_odb12 (genomes:24, BUSCOs:7266)

HiC contact map of curated assembly

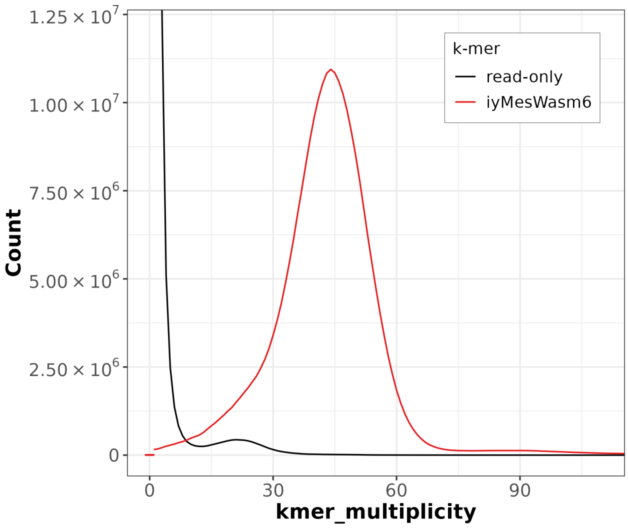


collapsed [LINK]

K-mer spectra of curated assembly

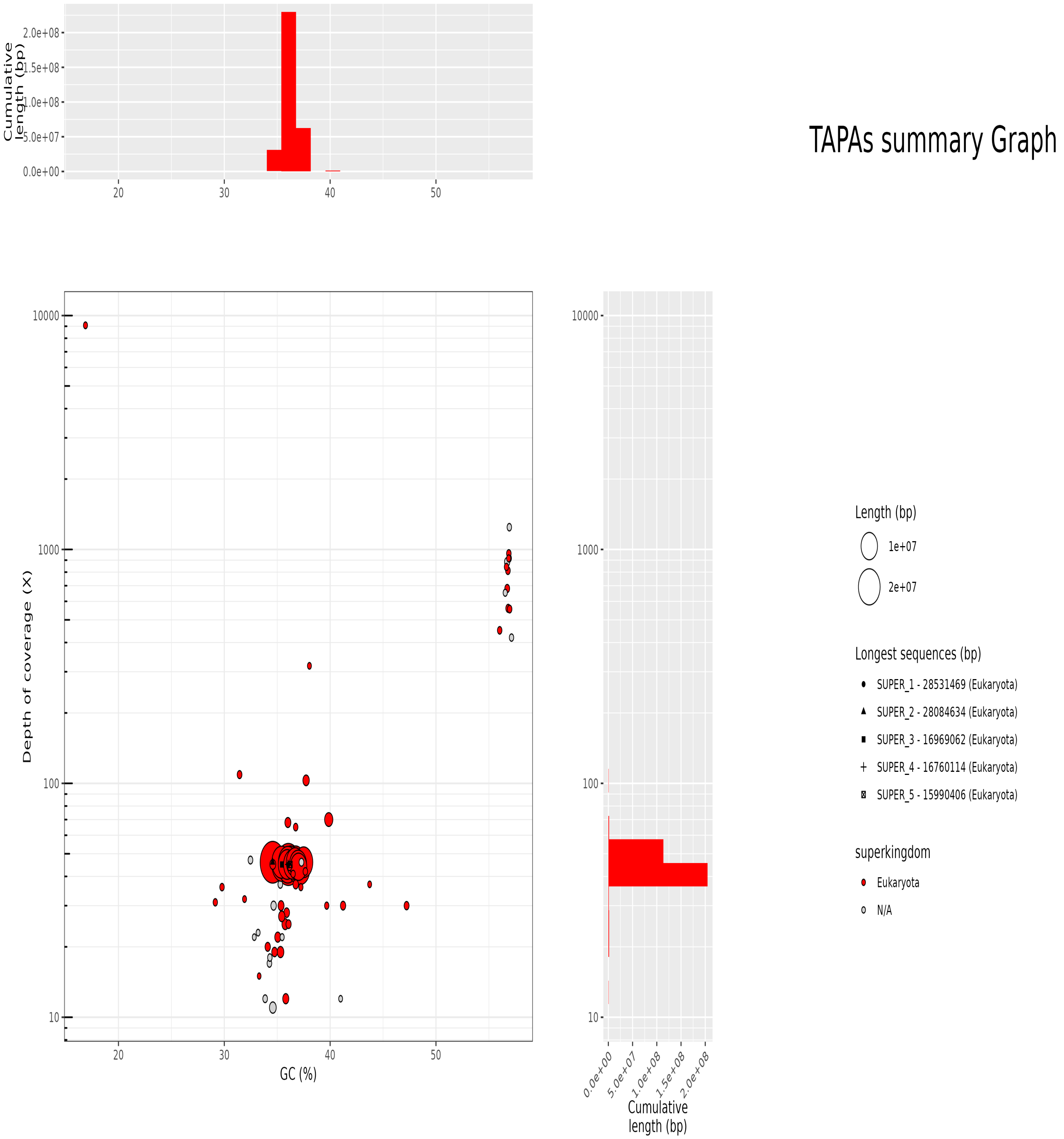


Distribution of k-mer counts per copy numbers found in asm



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	Long reads	Arima
Coverage	41	234

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: Sophie Mangenot
Affiliation: Genoscope

Date and time: 2025-11-20 04:48:18 CET