

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	612075
ToLID	iyMesOrie12
Species	Messor orientalis
Class	Insecta
Order	Hymenoptera

Genome Traits	Expected	Observed
Haploid size (bp)	339,111,674	366,044,330
Haploid Number	21 (source: ancestor)	21
Ploidy	1 (source: ancestor)	1
Sample Sex	Unknown	unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.7.Q52

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

- . Assembly length loss > 3% for pri
- . Not 90% of assembly in chromosomes for pri

Curator notes

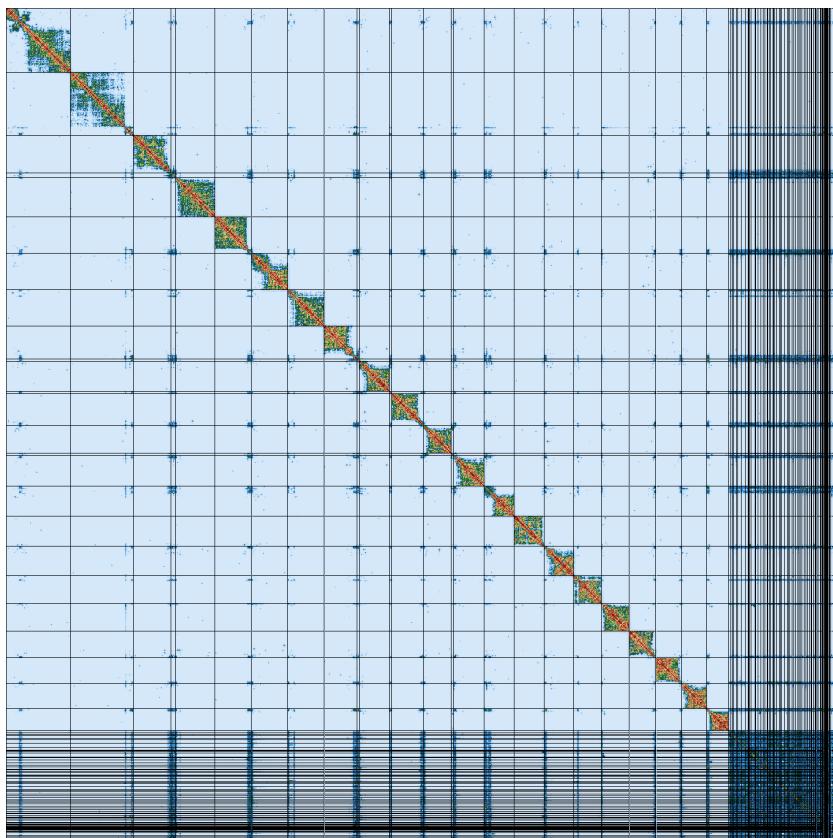
- . Interventions/Gb: 102
- . Contamination notes: "Partial viral hits (Streptomyces phage, 4 contigs, ~130Kb) and proteobacteria (Luteimonas saliphila, 1 contig, 25Kb) were detected with FCS-GX. Mitochondrial genome was removed from the assembly."
- . Other observations: "PacBio reads were assembled with hifiasm (yield: 474M, N50: 1.5Mb, 953 contigs). FCS-GX detected few contaminations. Purge_dups was run to remove haplotypic duplications (purged: 633 sequences, 90Mb remaining assembly: yield: 384M, N50: 2.4Mb, 360 contigs). Yahs was used to scaffold the contigs (--no-contig-ec was used to avoid over-cutting the contigs and the removal of telomere motifs). The HiC quality is suboptimal (just 2M cis read pairs >40Kb). It seems that this species potentially contains heterochromatin regions which show a much lower PacBio coverage profile and repetitive HiC interactions. At first I tried to assign more shrapnels as unlocalized contigs to the corresponding chromosomes, but then I removed those assignments again as the supports are very thin. I did a synteny analysis with iyMesOert5 and iyMesBuce7 and assigned a larger contig (previously chromosome 22) to SUPER_2. So I ended up with the expected 21 chromosomes."

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	384,225,878	366,044,330
GC %	36.92	36.85
Gaps/Gbp	491.9	611.95
Total gap bp	18,900	27,500
Scaffolds	171	108
Scaffold N50	11,360,363	13,350,424
Scaffold L50	13	11
Scaffold L90	59	33
Contigs	360	332
Contig N50	2,414,186	2,309,951
Contig L50	43	42
Contig L90	193	179
QV	48.1265	52.7534
Kmer compl.	95.5682	95.5015
BUSCO sing.	97.1%	97.1%
BUSCO dupl.	0.6%	0.5%
BUSCO frag.	0.8%	0.8%
BUSCO miss.	1.6%	1.6%

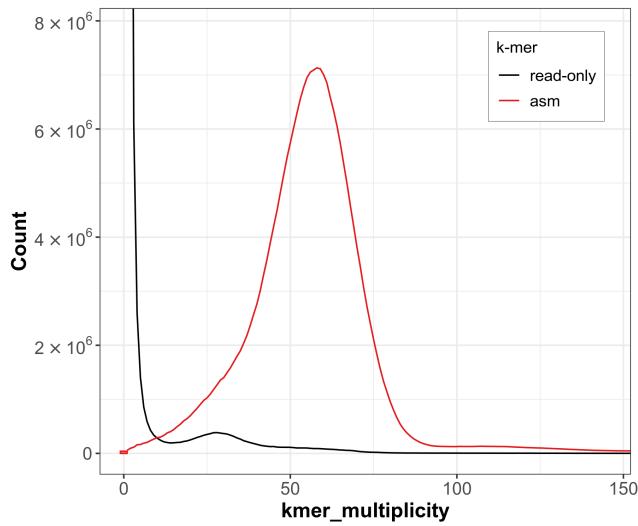
BUSCO: 5.8.3 (euk_genome_min, miniprot) / Lineage: hymenoptera_odb12 (genomes:78, BUSCOs:5920)

HiC contact map of curated assembly

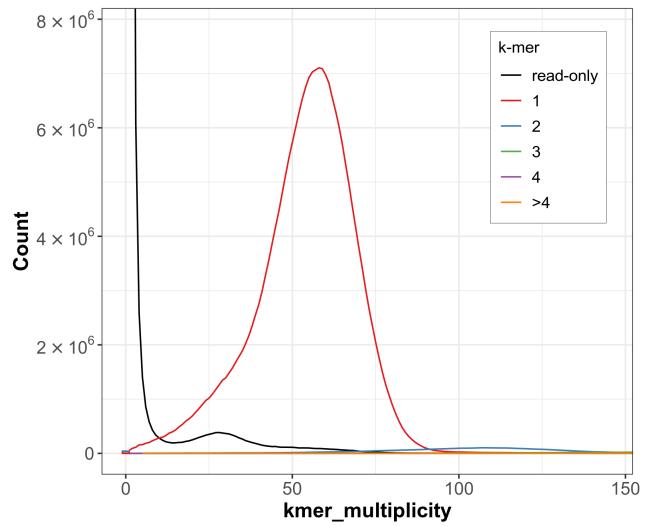


pri [LINK]

K-mer spectra of curated assembly

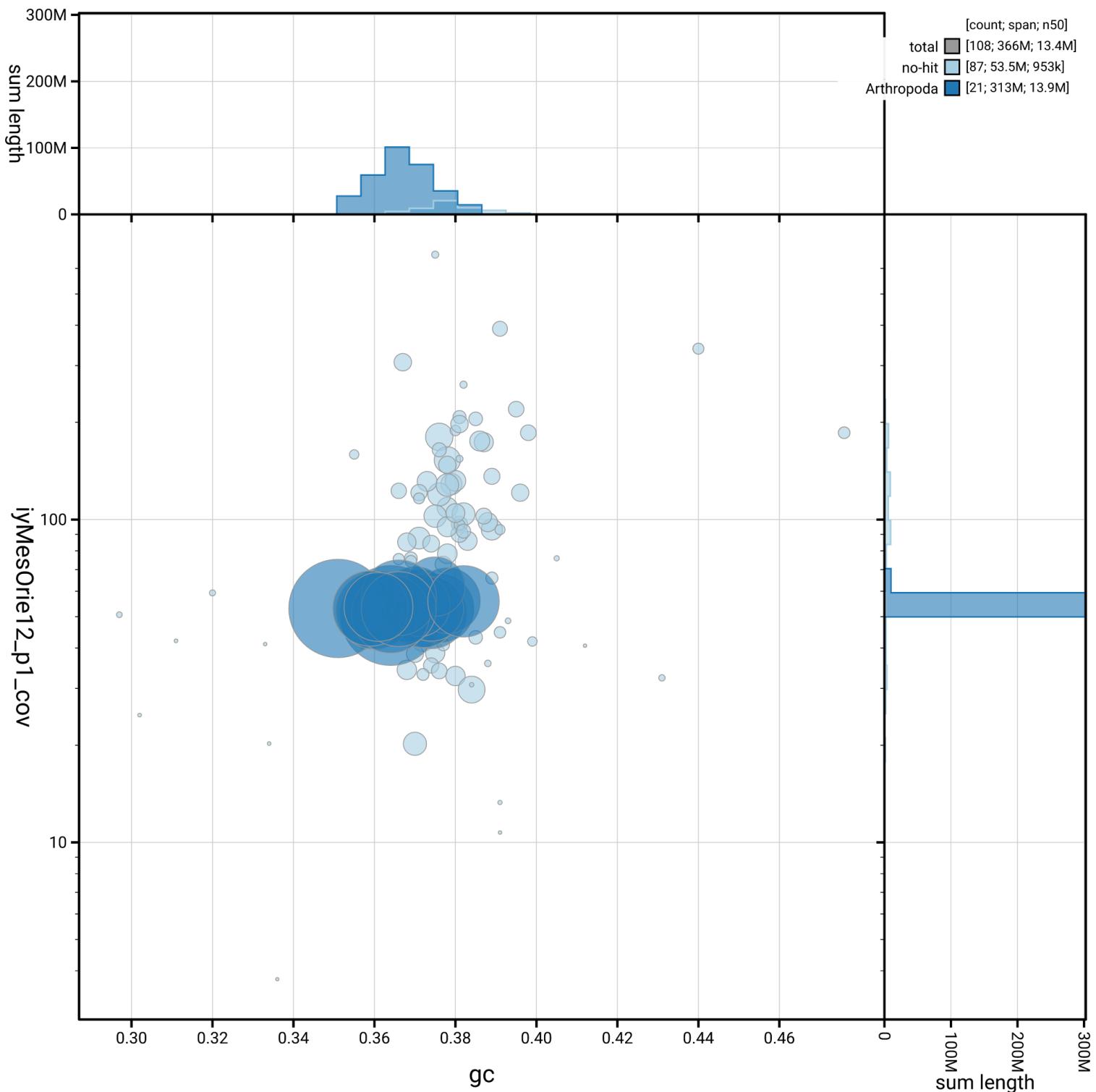


Distribution of k-mer counts coloured by their presence in reads/assemblies



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

Post-curation contamination screening



pri. 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	HiFi	HiC
Coverage	209x	301x

Assembly pipeline

- **Hifiasm**
 - |_ ver: 0.25.0-r726
 - |_ key param: 13
- **purge_dups**
 - |_ ver: 1.2.6
 - |_ key param: NA

Curation pipeline

- **GRIT_Rapid**
 - |_ ver: 1a3d79a8
 - |_ key param: NA
- **HiGlass**
 - |_ ver: 0.10.4
 - |_ key param: NA

Submitter: Martin Pippel

Affiliation: SciLifeLab

Date and time: 2025-12-19 15:51:20 CET