

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	3229175
ToLID	iyMegPost4
Species	Megachile posti
Class	Insecta
Order	Hymenoptera

Genome Traits	Expected	Observed
Haploid size (bp)	884,321,423	860,368,396
Haploid Number	16 (source: ancestor)	16
Ploidy	1 (source: ancestor)	2
Sample Sex	NA	NA

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.6.Q69

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

- . Observed Ploidy is different from Expected
- . Not 90% of assembly in chromosomes for collapsed

Curator notes

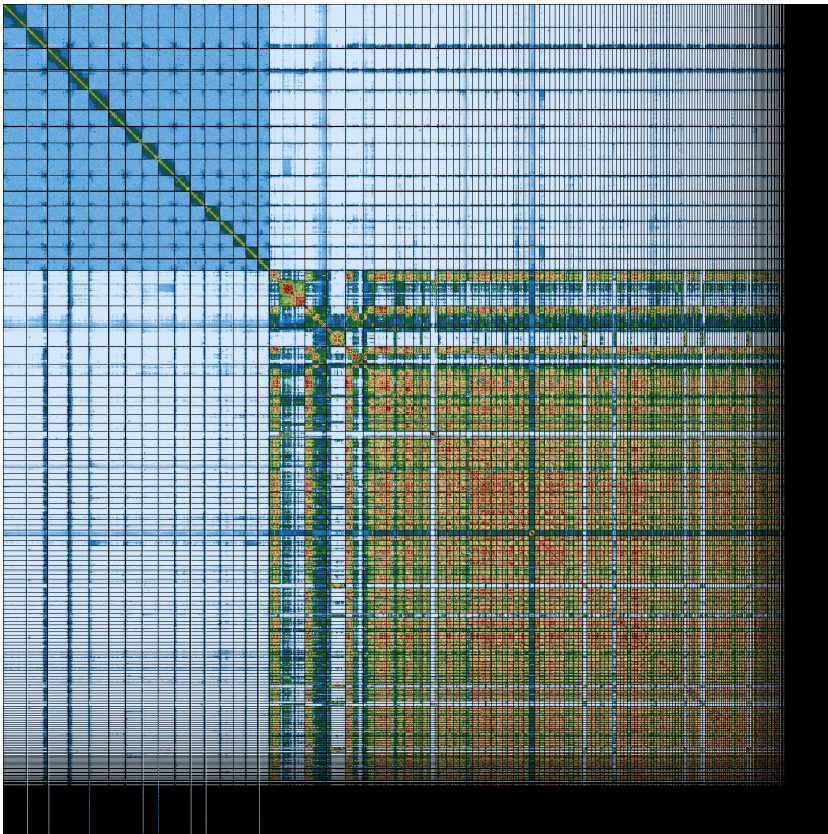
- . Interventions/Gb:
- . Contamination notes: "FCS-GX detected 13 small (< 100kb) contigs matching *Apilactobacillus micheneri* and *timberlakei*, but no complete genome could be recovered for those bacteria."
- . Other observations: "The assembly was produced with hifiasm in Hi-C mode, decontaminated with FCS-GX, and haplotigs purged with `purge_dups`. Scaffolding was performed with YaHS. The resulting scaffolds are mostly gapless, with 16 scaffolds matching the known karyotype for this genus (n=16), and containing all of the identified BUSCOs (99.1%, `hymenoptera_odb10`). These likely represent the euchromatic arms of each chromosome, and accordingly most contain telomeric sequences on one end only. The assembly also contains a large number of scaffolds that couldn't be assigned to chromosomes, likely representing the heterochromatic and highly repetitive fraction of this genome. Assembly size is roughly consistent with the k-mer genome size estimate. Mitochondria was assembled with Oatk."

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	860,368,196	860,368,396
GC %	31.13	31.13
Gaps/Gbp	0	1.16
Total gap bp	0	200
Scaffolds	738	737
Scaffold N50	8,649,748	8,649,748
Scaffold L50	27	27
Scaffold L90	118	118
Contigs	738	738
Contig N50	8,649,748	8,649,748
Contig L50	27	27
Contig L90	118	118
QV	69.3758	69.3758
Kmer compl.	94.8476	94.8476
BUSCO sing.	99.1%	99.1%
BUSCO dupl.	0.1%	0.1%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.8%	0.8%

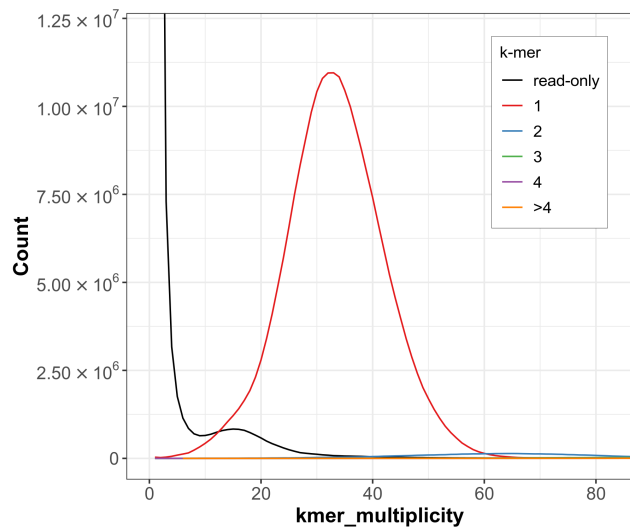
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: hymenoptera_odb10 (genomes:40, BUSCOs:5991)

HiC contact map of curated assembly



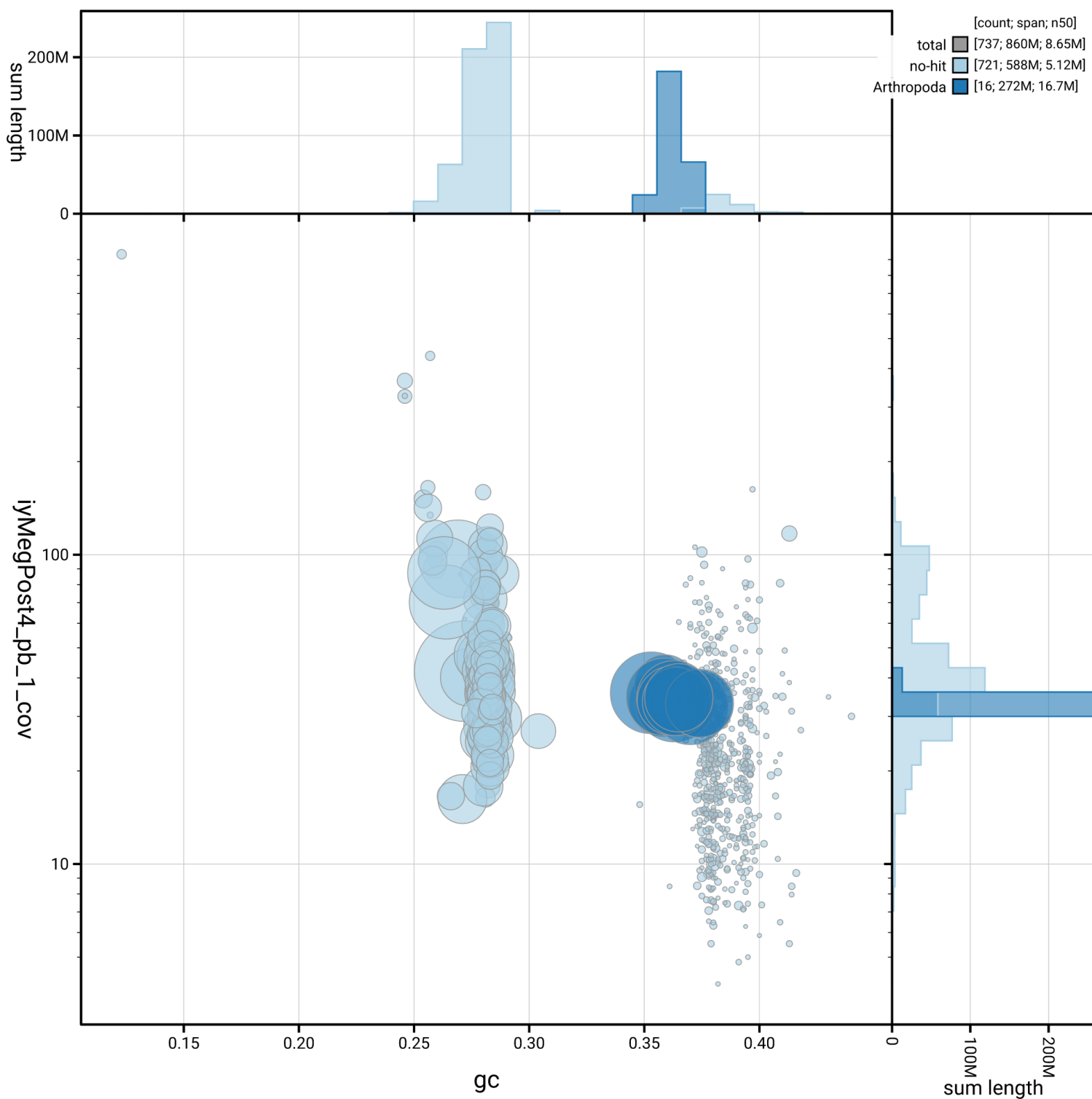
collapsed [\[LINK\]](#)

K-mer spectra of curated assembly



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

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 HiC v2
Coverage	33x	67x

Assembly pipeline

- **hifiasm-hic**
 - |_ *ver*: 0.25.0-r726
 - |_ *key param*: --h1
 - |_ *key param*: --h2
- **purge_dups**
 - |_ *ver*: 1.2.5
 - |_ *key param*: NA
- **YaHS**
 - |_ *ver*: 1.2a.2
 - |_ *key param*: --no-contig-ec
- **NBIS/Earth-Biogenome-Project-pilot**
 - |_ *ver*: 1d2a1e4
 - |_ *key param*: NA
- **sanger-tol/blobtoolkit**
 - |_ *ver*: 0.8.0
 - |_ *key param*: NA
- **Oatk**
 - |_ *ver*: 1.0
 - |_ *key param*: -m hymenoptera.fam

Curation pipeline

- **sanger-tol/curationpretext**
 - |_ *ver*: 1.4.2
 - |_ *key param*: NA
- **sanger-tol/agp-tpf-utils**
 - |_ *ver*: 1.2.3
 - |_ *key param*: NA

Submitter: Guilherme B. Dias

Affiliation: SciLifeLab

Date and time: 2025-09-23 10:19:47 CEST