

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	3229181
ToLID	iyMesBuce7
Species	Messor bucephalus
Class	Insecta
Order	Hymenoptera

Genome Traits	Expected	Observed
Haploid size (bp)	369,553,906	397,571,255
Haploid Number	21 (source: ancestor)	21
Ploidy	1 (source: ancestor)	2
Sample Sex	NA	NA

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q66

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

- . Observed Ploidy is different from Expected
- . Assembly length loss > 3% for collapsed
- . Not 90% of assembly in chromosomes for collapsed

Curator notes

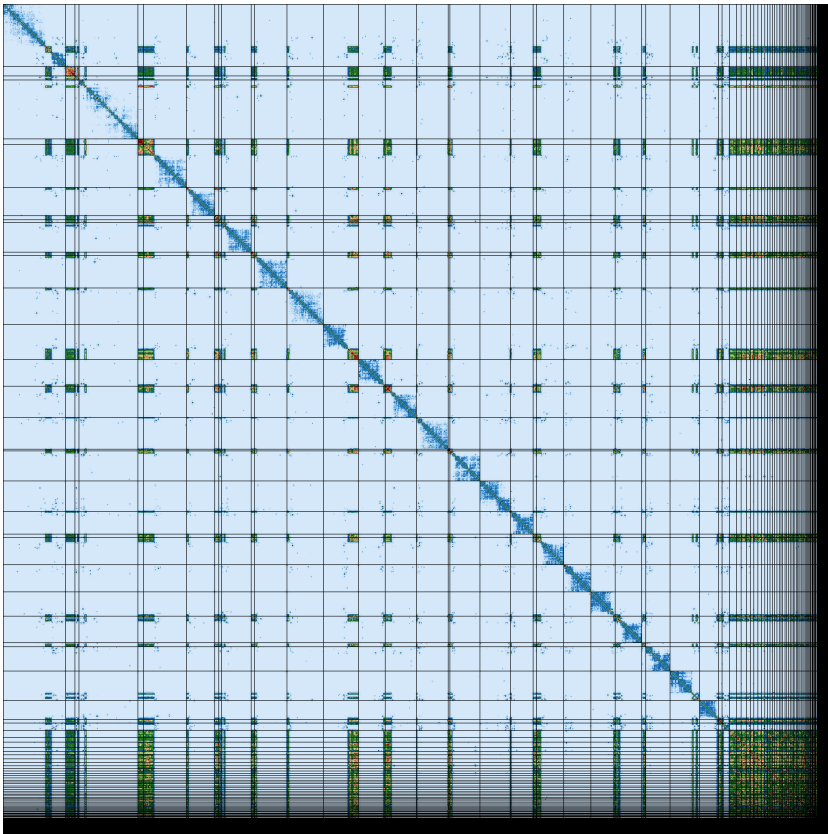
. Interventions/Gb:
. Contamination notes: "FCS-GX detected 3 small (< 100kb) contigs matching Tunavirus ISF002, Phage NBSal001, and Mixta mediterraneensis."
. Other observations: "The assembly was produced with hifiiasm, decontaminated with FCS-GX, and haplotigs purged with purge_dups. Scaffolding was performed with YaHS and the scaffolds were analysed with blobtools for the presence of residual contaminants. Although no contaminants were found, scaffolds with extreme coverages (<50x and >140x) were filtered out of the assembly. The curated assembly has 21 putative chromosomes likely representing the euchromatic arms of each chromosome. The assembly also contains a large number of scaffolds that could not be assigned to chromosomes, likely representing the heterochromatic and highly repetitive fraction of this genome. Mitochondria was assembled with MitoHiFi."

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	587,439,348	397,571,255
GC %	35.82	36.03
Gaps/Gbp	464.73	447.72
Total gap bp	54,600	35,600
Scaffolds	468	143
Scaffold N50	10,345,731	14,167,757
Scaffold L50	20	11
Scaffold L90	156	36
Contigs	741	321
Contig N50	2,283,000	3,175,829
Contig L50	69	37
Contig L90	307	138
QV	67.9468	66.8606
Kmer compl.	94.9617	94.5173
BUSCO sing.	99.3%	99.3%
BUSCO dupl.	0.3%	0.3%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.4%	0.4%

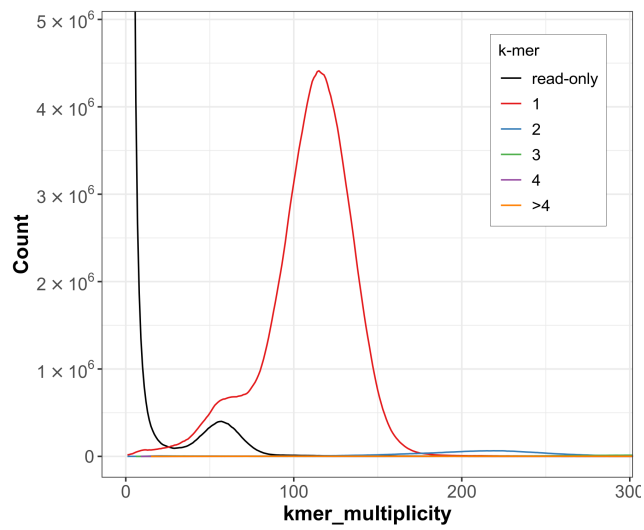
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: hymenoptera_odb10 (genomes:40, BUSCOs:5991)

HiC contact map of curated assembly

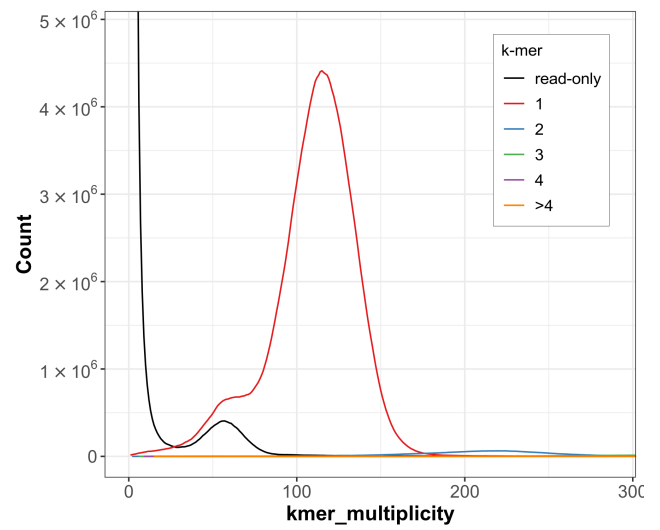


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

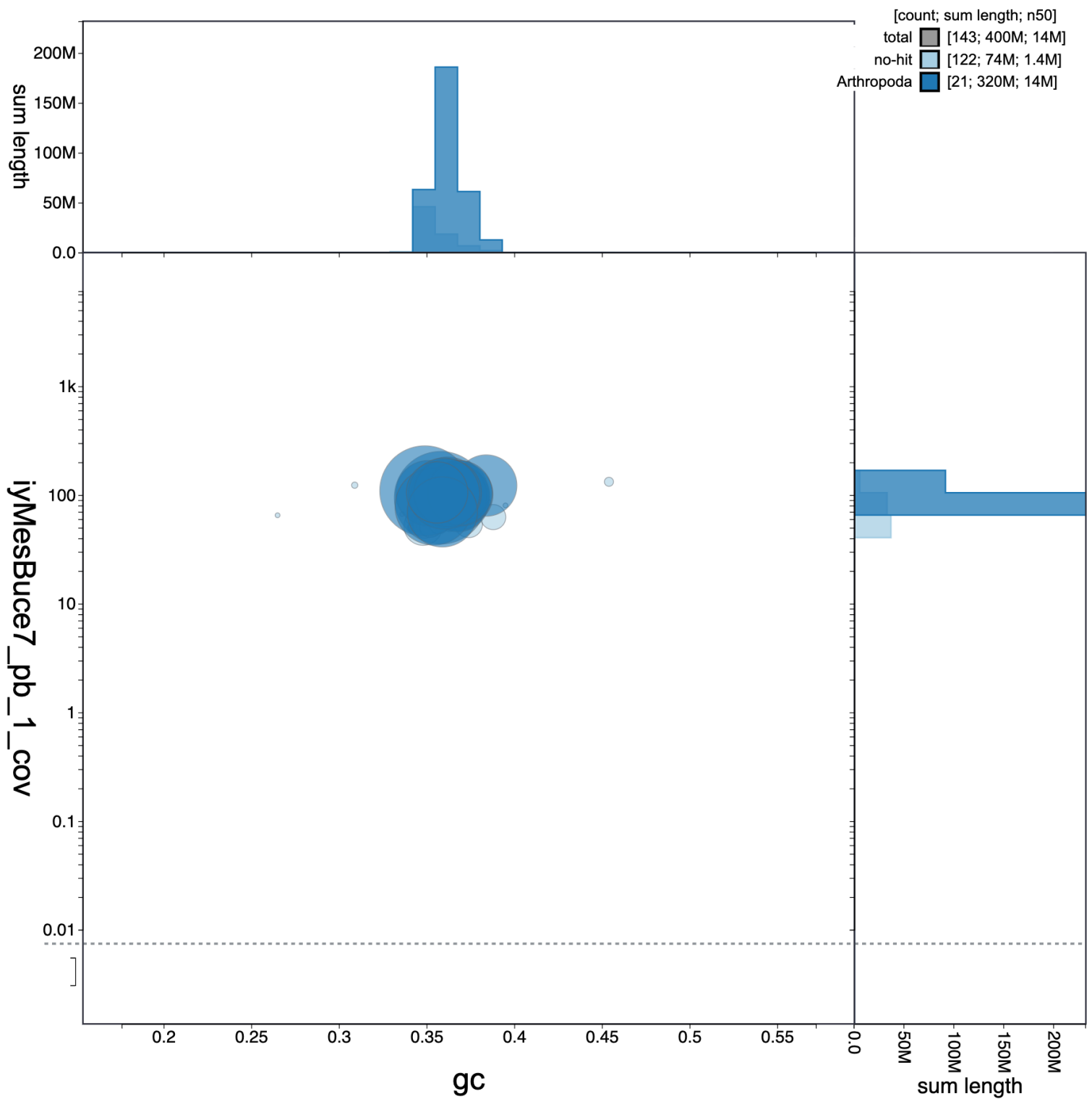


Distribution of k-mer counts per copy numbers found in asm



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	115x	50x

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*: NA
- **NBIS/Earth-Biogenome-Project-pilot**
 - |_ *ver*: 99c8937
 - |_ *key param*: NA
- **sanger-tol/blobtoolkit**
 - |_ *ver*: 0.8.0
 - |_ *key param*: NA

Curation pipeline

- **sanger-tol/curationpretext**
 - |_ *ver*: 1.4.2
 - |_ *key param*: NA
- **sanger-tol/pretextview**
 - |_ *ver*: 1.0.3
 - |_ *key param*: NA
- **sanger-tol/agp-tpf-utils**
 - |_ *ver*: 1.2.3
 - |_ *key param*: NA

Submitter: Guilherme B. Dias

Affiliation: SciLifeLab

Date and time: 2025-10-15 09:50:47 CEST