

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	3229266
ToLID	iyEucMavr2
Species	Eucera mavromoustakisi
Class	Insecta
Order	Hymenoptera

Genome Traits	Expected	Observed
Haploid size (bp)	458,179,693	620,345,852
Haploid Number	21 (source: ancestor)	22
Ploidy	1 (source: ancestor)	1
Sample Sex	unknown	unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.7.Q58

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

- . Observed Haploid size (bp) has >20% difference with Expected
- . Observed Haploid Number is different from Expected
- . Not 90% of assembly in chromosomes for pri

Curator notes

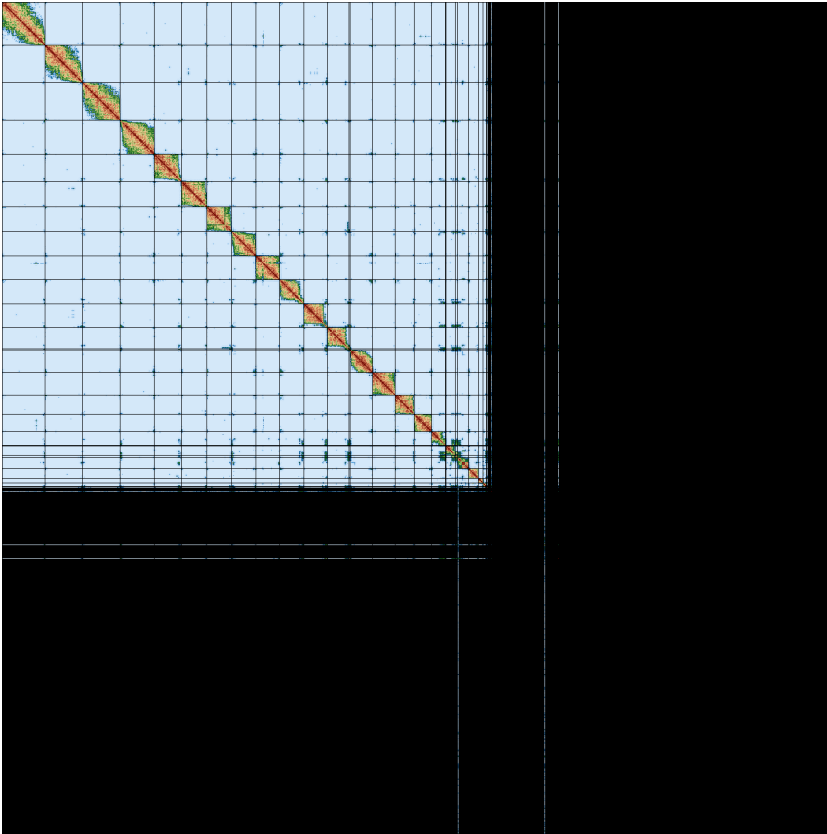
. Interventions/Gb: 6
. Contamination notes: "No contaminations were detected with FCS-GX."
. Other observations: "PacBio reads were subsampled to 50X. Reads were assembled with hifiiasm (yield: 1003M, N50: 209Kb, 12561 contigs). FCS-GX detected no contaminations. Purge_dups removed 7301 contigs (379.4Mb). HiC data is of a decent quality with 14M cis-reads pairs above 40Kb. Yahs was NOT used to join contigs into scaffolds as it oversplit many contigs (removed telomere motifs) and overjoined many shrapnel contigs. The current map still contains many shrapnel contigs that I could not unambiguously assign to any of the chromosomes. Super_18 looks still a bit odd also due to the large unlocalized contig. Some of the chromosomes e.g. SUPER_8 and SUPER_11 have a weaker signal towards the end. I assigned the contigs based on a map without shrapnels shorter than 100K (4843 contigs) and at that time the signal was much clearer. The final map now has all shrapnels and therefore the signal is much ore diluted. "

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	624,165,344	620,345,852
GC %	36.83	36.8
Gaps/Gbp	0	80.6
Total gap bp	0	10,000
Scaffolds	5,346	5,297
Scaffold N50	6,484,053	13,027,790
Scaffold L50	19	16
Scaffold L90	2,738	2,701
Contigs	5,346	5,347
Contig N50	6,484,053	5,915,780
Contig L50	19	20
Contig L90	2,738	2,751
QV	58.9556	58.9288
Kmer compl.	91.5536	91.4924
BUSCO sing.	95.8%	97.5%
BUSCO dupl.	2.1%	0.3%
BUSCO frag.	0.5%	0.5%
BUSCO miss.	1.6%	1.6%

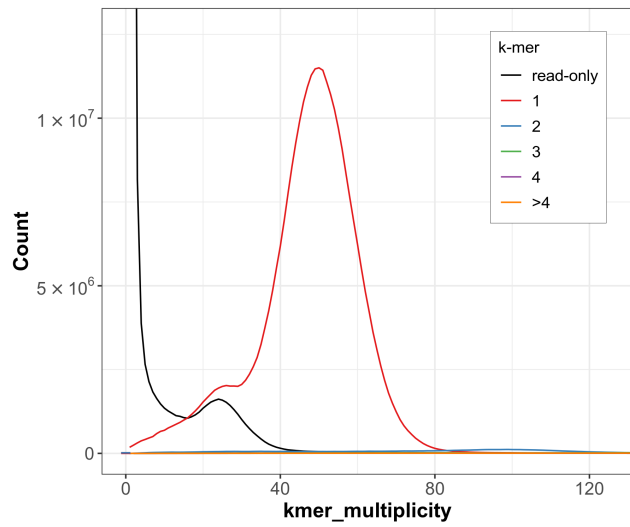
BUSCO: 5.8.3 (euk_genome_min, miniprot) / Lineage: endopterygota_odb12 (genomes:76, BUSCOs:3754)

HiC contact map of curated assembly

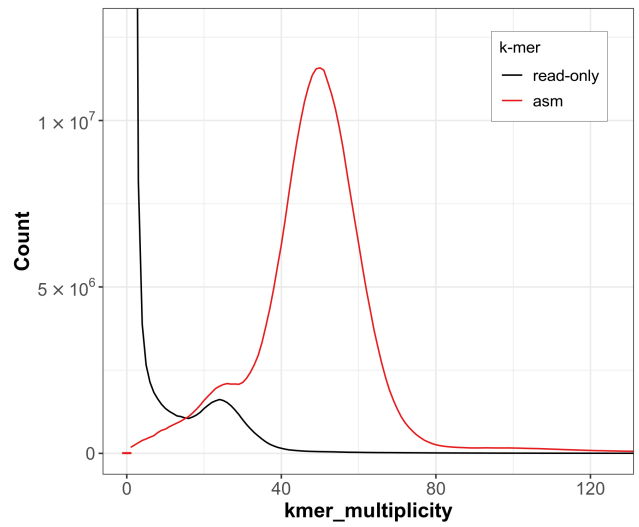


pri [\[LINK\]](#)

K-mer spectra of curated assembly

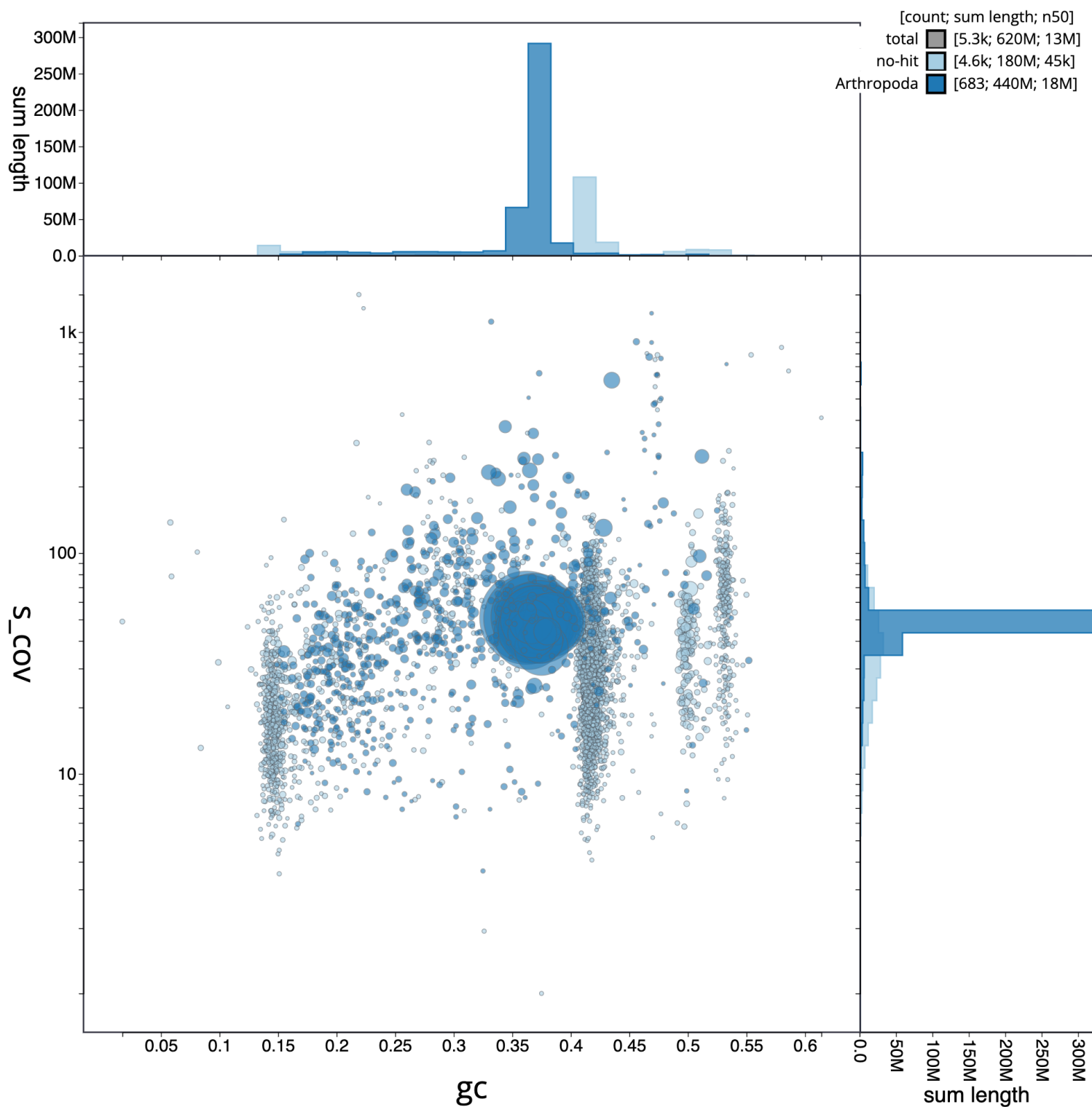


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	123x	231x

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: 2025-10-22 12:49:36 CEST