ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	446453
ToLID	**inOsmFulv5**
Species	Osmylus fulvicephalus
Class	Insecta
Order	Neuroptera

Genome Traits	Expected	Observed
Haploid size (bp)	625,527,197	674,675,484
Haploid Number	8 (source: ancestor)	8
Ploidy	2 (source: ancestor)	2
Sample Sex	XY	XY

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 7.8.Q68

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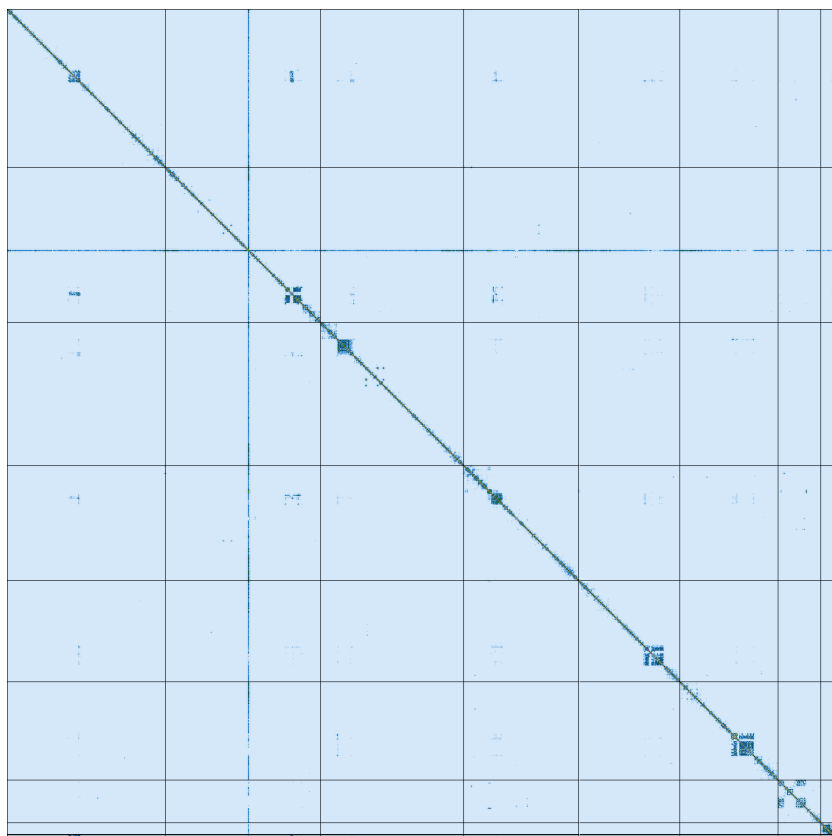
Curator notes

. Interventions/Gb: 9
. Contamination notes: "No contaminations were detected with FCS-GX."
. Other observations: "PacBio reads were subsampled to 60X. Reads were assembled with hifiasm (yield: 488M, N50: 49.1Mb, 176 contigs). FCS-GX detected no contaminations. Purge_dups was run on the contigs but only the the dups.bed file was used in the manual curation step, as it removed to much sequence and telomere motifs. Due to the low HiC quality (2.8M cis-read pairs >40Kb) the scaffolding was manually done in the first manual curation round. In this step the information from tiara, mitohifi, blobtools and purge_dups was used and 110 contigs (15M) were removed. SUPER_8 and SUPER_11 had a very low coverage profile - roughly half coverage. A synteny analysis with GCA_958496175, GCA_958496155, GCF_905475395.1 and GCA_020423425.1 suggests that SUPER_8 is the X chromosome. Super_11 is the Y chromosomes. Furthermore, the chromosome structure seems to be very stable within these insects (6 "closely" related species). All of them suggest the following joins: SUPER_5 AND SUPER_7 (both telomere motifs available), SUPER_2 AND SUPER_10, SUPER_6 and SUPER_9 (one telomere motif). Then these weird repeats (centromers?) are always within the chromosomes. "

Quality metrics table

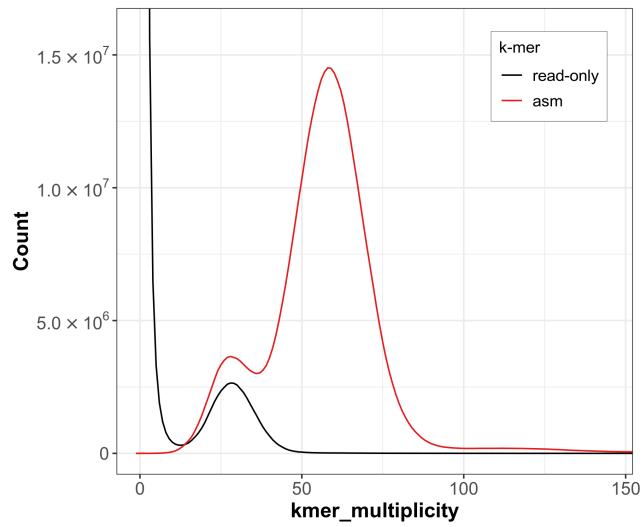| Metrics	Pre-curation pri	Curated pri
Total bp	666,151,682	674,675,484
GC %	36.31	36.29
Gaps/Gbp	90.07	53.36
Total gap bp	12,000	7,200
Scaffolds	123	24
Scaffold N50	104,575,600	116,168,846
Scaffold L50	3	3
Scaffold L90	16	6
Contigs	183	60
Contig N50	22,977,000	51,542,635
Contig L50	8	5
Contig L90	43	16
QV	70.6914	68.5927
Kmer compl.	90.8782	90.8268
BUSCO sing.	93.3%	94.4%
BUSCO dupl.	1.6%	0.4%
BUSCO frag.	1.4%	1.4%
BUSCO miss.	3.7%	3.8%

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

HiC contact map of curated assembly

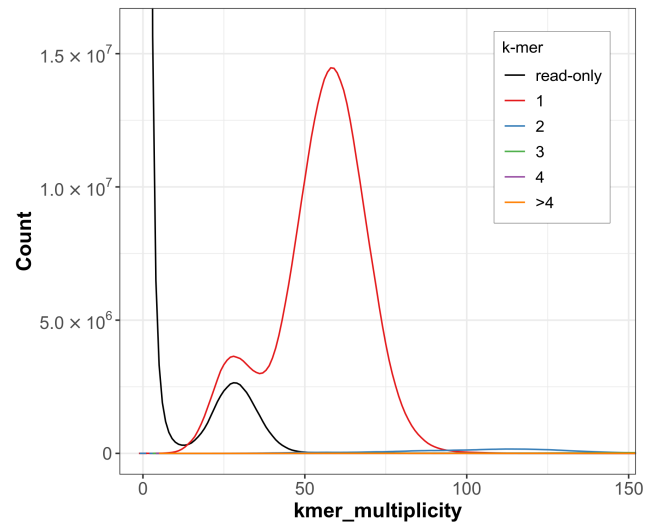**pri** [LINK]
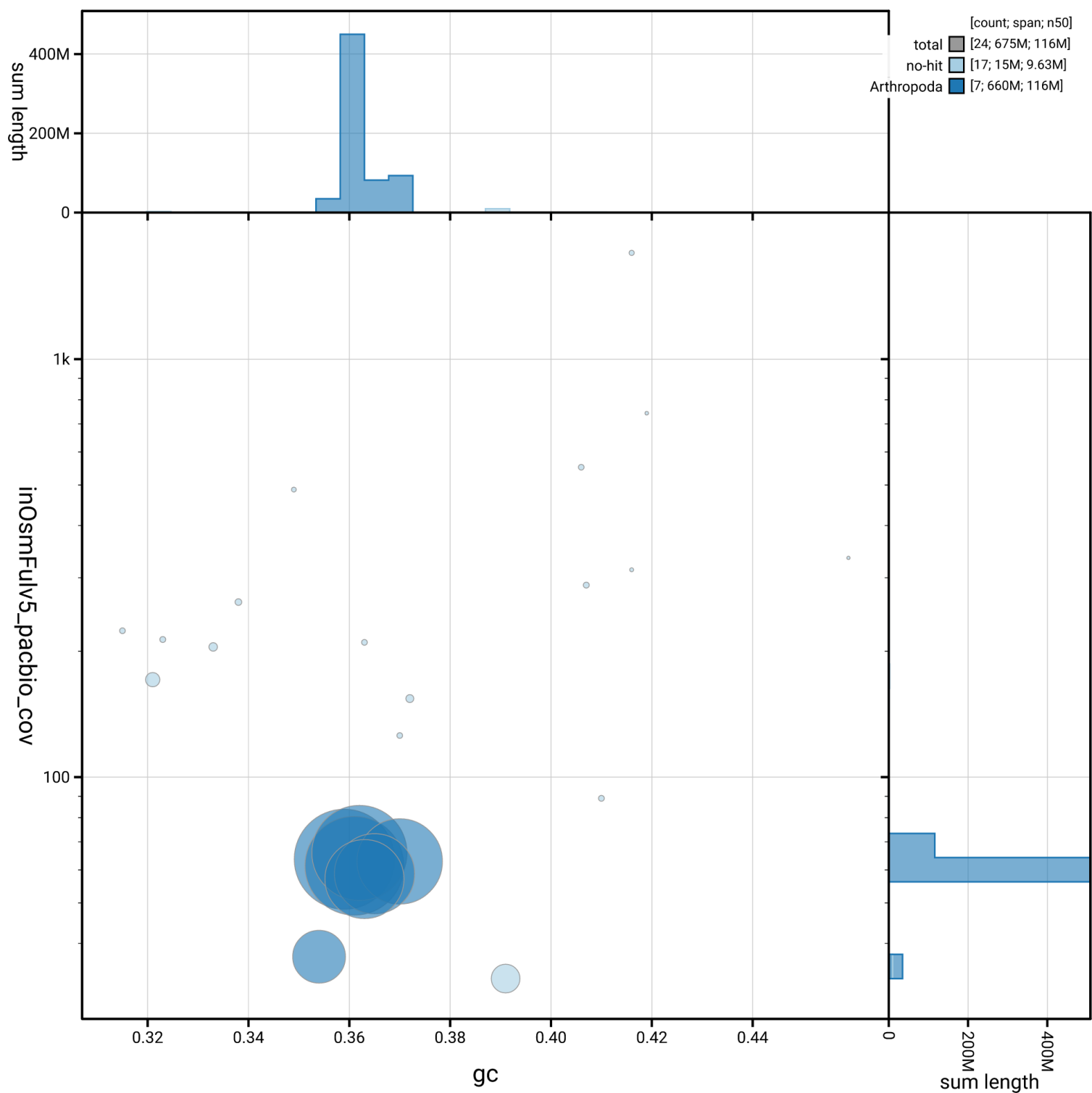
K-mer spectra of curated assembly

Distribution of k-mer counts 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	164x	49x

Assembly pipeline

- **Hifiasm**
  |_ *ver:* 0.25.0-r726
  |_ *key param:* l3
- **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: 2025-11-24 21:17:25 CET