

# ERGA Assembly Report

v24.04.03\_beta

Tags: ERGA-BGE

TxID	1464561
ToLID	<b>idCullati1</b>
Species	Culex laticinctus
Class	Insecta
Order	Diptera

Genome Traits	Expected	Observed
Haploid size (bp)	726,182,214	833,812,495
Haploid Number	3 (source: ancestor)	3
Ploidy	3 (source: ancestor)	2
Sample Sex	XX	XX

## EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 5.8.Q55

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

- . Observed Ploidy is different from Expected
- . More than 1000 gaps/Gbp for pri

### Curator notes

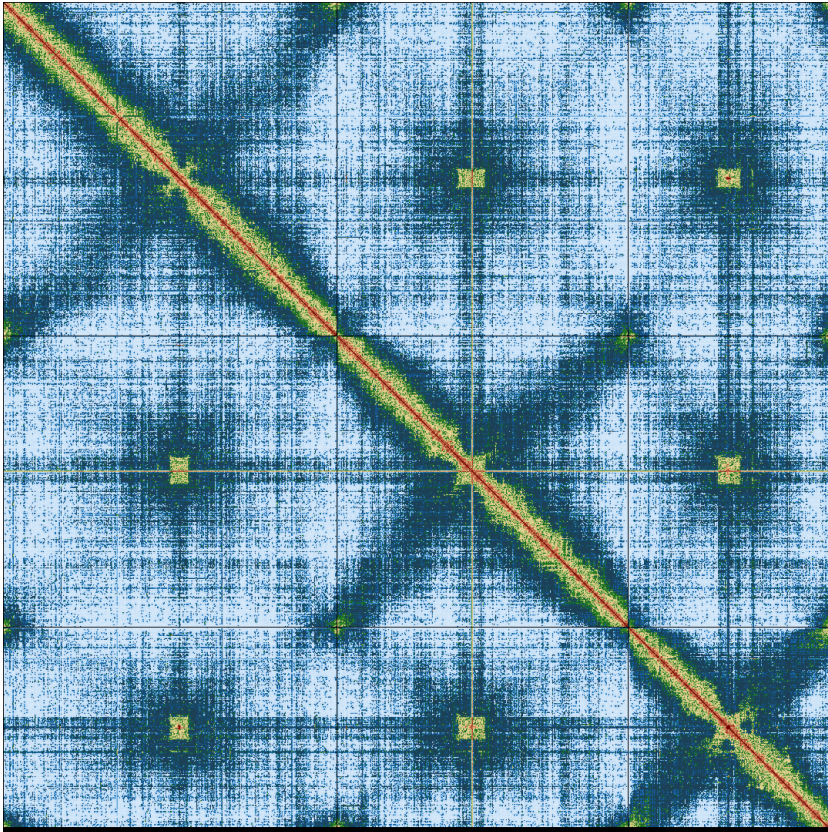
- . Interventions/Gb: 430
- . Contamination notes: "Total length of scaffolds removed: 989,717 (0.1 %) Scaffolds removed: 1 (0.2 %) Largest scaffold removed: (989,717) FCS-GX contaminant species (number of scaffolds; total length of scaffolds): Wolbachia endosymbiont (group B) of Melanostoma mellinum, a-proteobacteria (1; 989,717)"
- . Other observations: "Chromosomes named by size"

# Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	835,263,260	833,812,495
GC %	38.73	38.73
Gaps/Gbp	3,382.17	3,564.35
Total gap bp	565,000	594,400
Scaffolds	458	311
Scaffold N50	287,761,923	291,894,813
Scaffold L50	2	2
Scaffold L90	3	3
Contigs	3,283	3,283
Contig N50	435,834	433,957
Contig L50	593	596
Contig L90	1,913	1,919
QV	55.5	55.5
Kmer compl.	99.13	99.00
BUSCO sing.	97.7%	97.6%
BUSCO dupl.	1.0%	1.0%
BUSCO frag.	0.5%	0.6%
BUSCO miss.	0.8%	0.8%

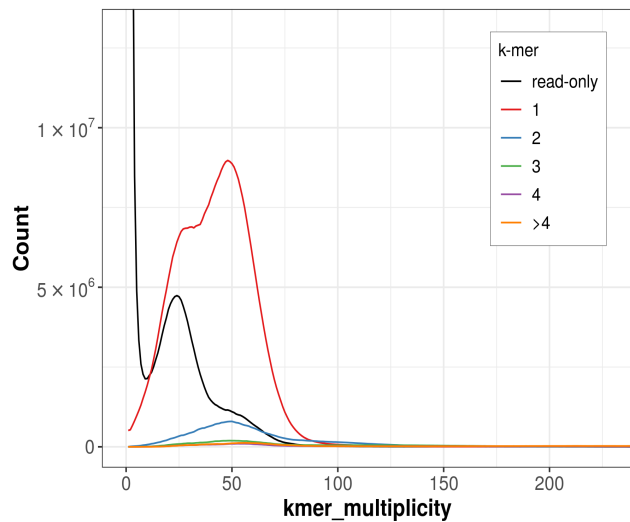
BUSCO 5.3.0 Lineage: insecta\_odb10 (genomes:75, BUSCOs:1367)

# HiC contact map of curated assembly

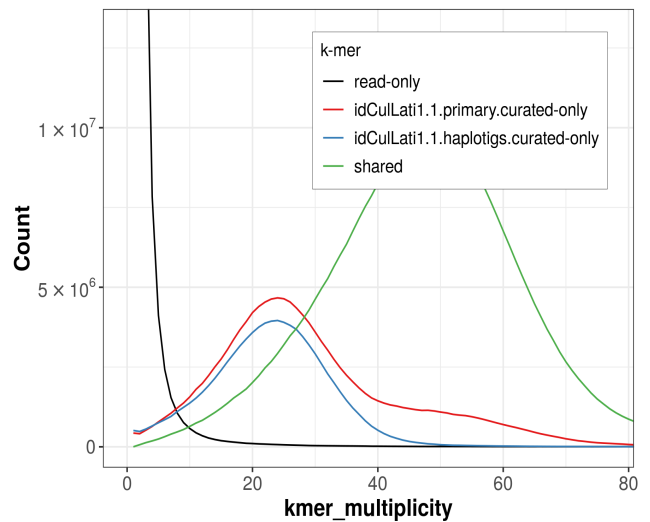


pri [\[LINK\]](#)

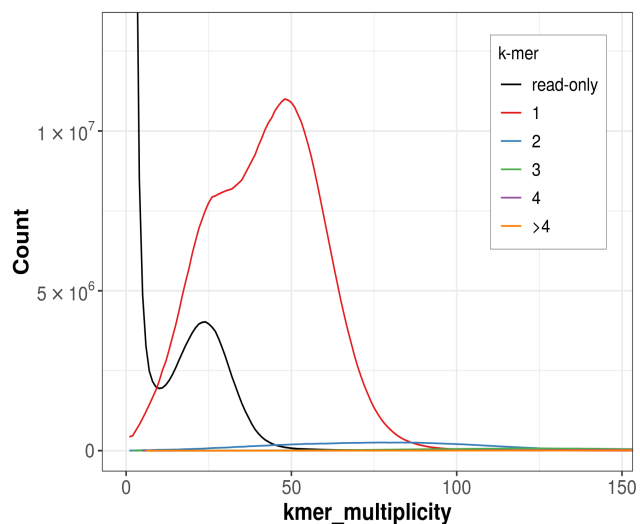
# K-mer spectra of curated assembly



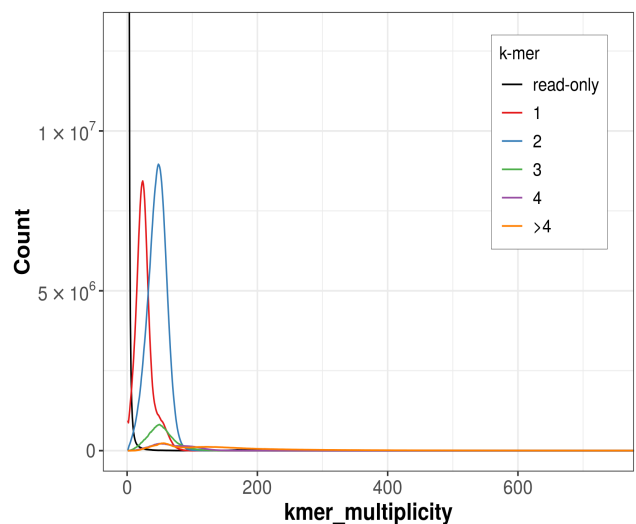
Distribution of k-mer counts per copy numbers found in **haplotigs** (hapl.)



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

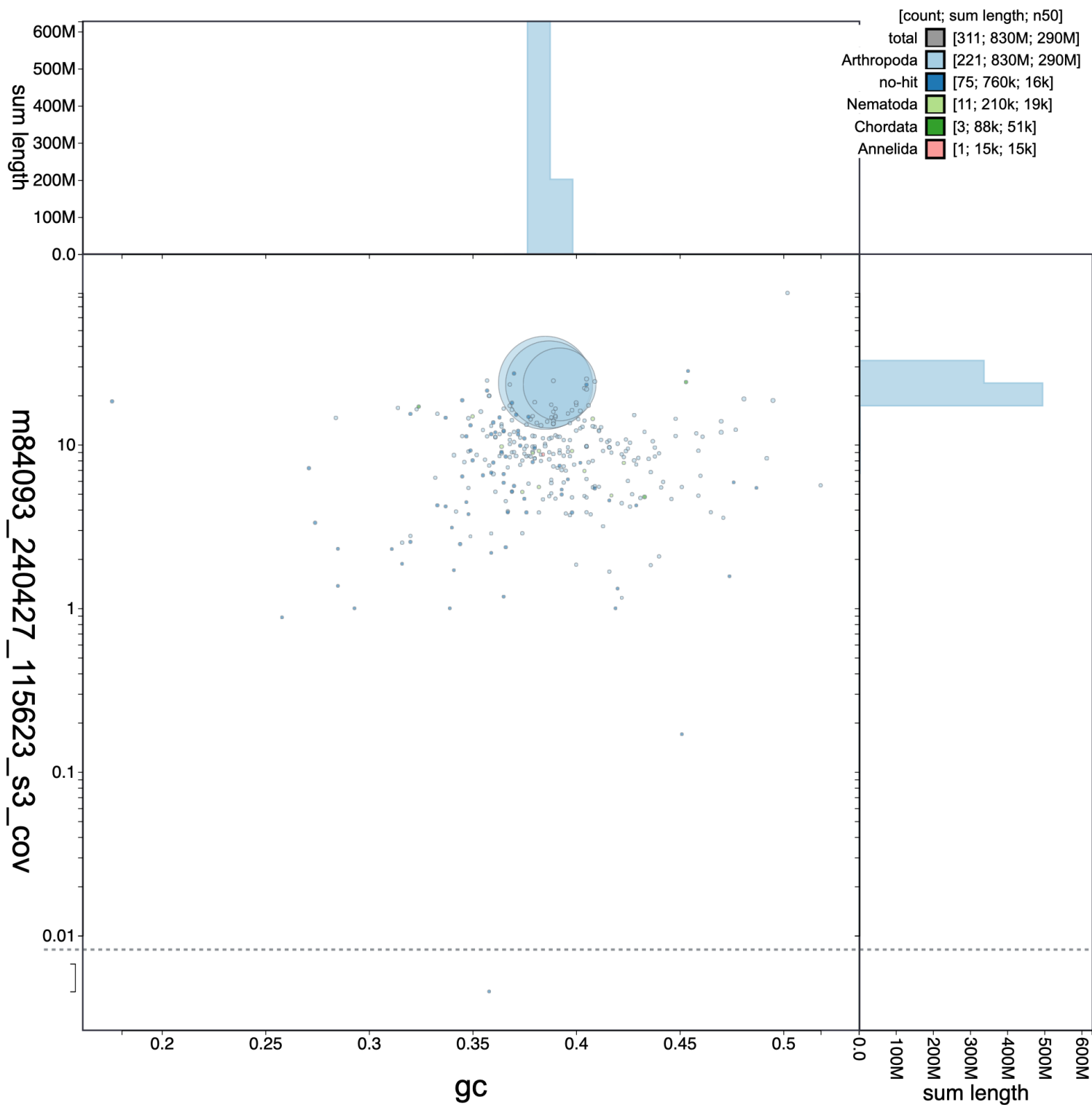


Distribution of k-mer counts per copy numbers found in (hapl.)



Distribution of k-mer counts per copy numbers found in asm (dipl.)

# 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	PacBio HiFi	Arima v2
Coverage	51x	NA

## Assembly pipeline

- **hifiasm**
  - |\_ *ver*: 0.19.8-r603
  - |\_ *key param*: --primary
- **purge\_dups**
  - |\_ *ver*: 1.2.5
  - |\_ *key param*: -e
- **yahs**
  - |\_ *ver*: 1.2a.2
  - |\_ *key param*: NA

## Curation pipeline

- **hifiasm**
  - |\_ *ver*: 0.19.8-r603
  - |\_ *key param*: --primary
- **purge\_dups**
  - |\_ *ver*: 1.2.5
  - |\_ *key param*: -e
- **yahs**
  - |\_ *ver*: 1.2a.2
  - |\_ *key param*: NA
- **TreeVal**
  - |\_ *ver*: 1.1
  - |\_ *key param*: NA

Submitter: Michael Paulini

Affiliation: WSI

Date and time: 2024-05-30 10:56:09 CEST