

# ERGA Assembly Report

v24.04.03\_beta

Tags: ERGA-BGE

TxID	420550
ToLID	<b>idCulPerx1</b>
Species	Culex perexiguus
Class	Insecta
Order	Diptera

Genome Traits	Expected	Observed
Haploid size (bp)	353,856,597	437,724,337
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 Haploid size (bp) has >20% difference with Expected
- . Observed Ploidy is different from Expected
- . Assembly length loss > 3% for pri
- . More than 1000 gaps/Gbp for pri

### Curator notes

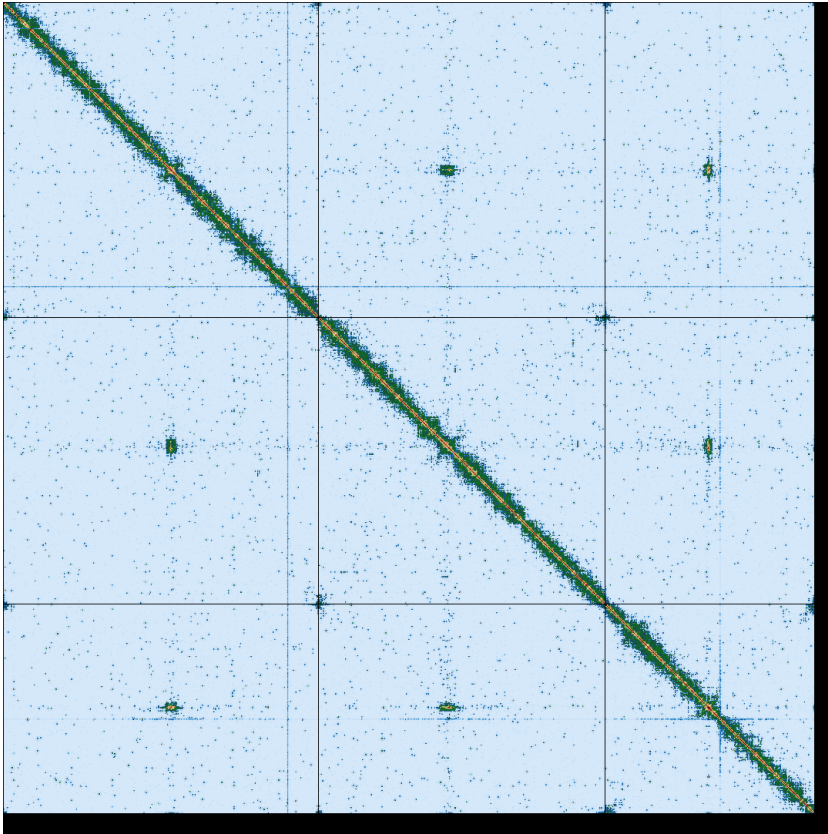
- . Interventions/Gb: 651
- . Contamination notes: "Total length of scaffolds removed: 84,166 (0.0 %); Scaffolds removed: 3 (0.4 %); Largest scaffold removed: (42,305); FCS-GX contaminant species (number of scaffolds; total length of scaffolds): None; Mitochondrion (3; 84,166)"
- . Other observations: "Chromosomes named by size; PacBio library was ULI hence lower contig N50; Hi-C from a different individual, so assembly is not Hi-C phased"

# Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	477,427,365	437,724,337
GC %	38.57	38.6
Gaps/Gbp	3,248.66	3,310.3
Total gap bp	310,200	289,800
Scaffolds	764	494
Scaffold N50	114,754,111	150,776,469
Scaffold L50	2	2
Scaffold L90	4	3
Contigs	2,315	1,943
Contig N50	448,257	475,086
Contig L50	325	289
Contig L90	1,159	934
QV	55.1	55.2
Kmer compl.	99.00	99.17
BUSCO sing.	91.6%	95.6%
BUSCO dupl.	7.7%	3.1%
BUSCO frag.	0.4%	0.4%
BUSCO miss.	0.3%	0.9%

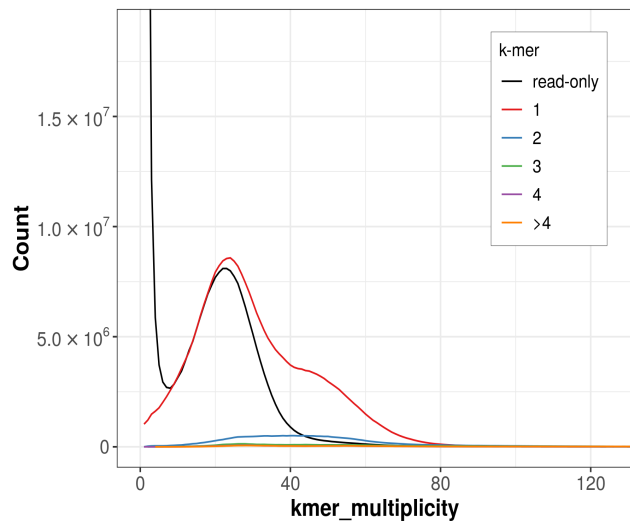
BUSCO 5.3.0 Lineage: insecta\_odb10 (genomes:75, BUSCOs:1367)

# HiC contact map of curated assembly

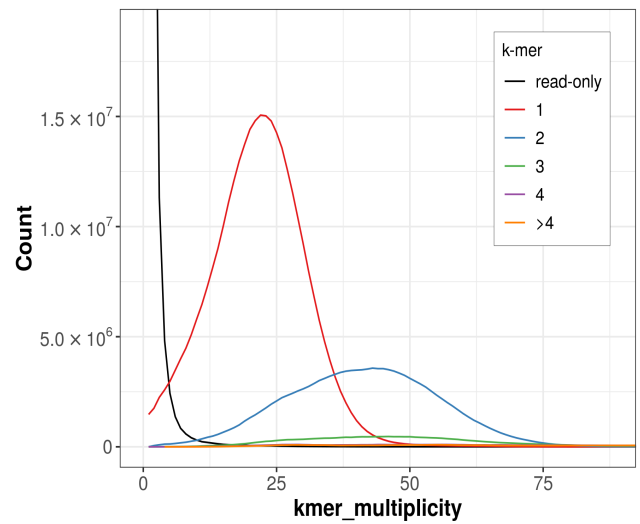


pri [\[LINK\]](#)

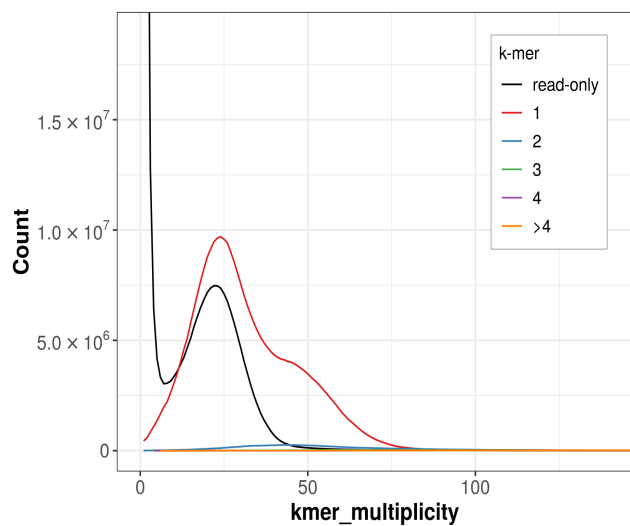
# K-mer spectra of curated assembly



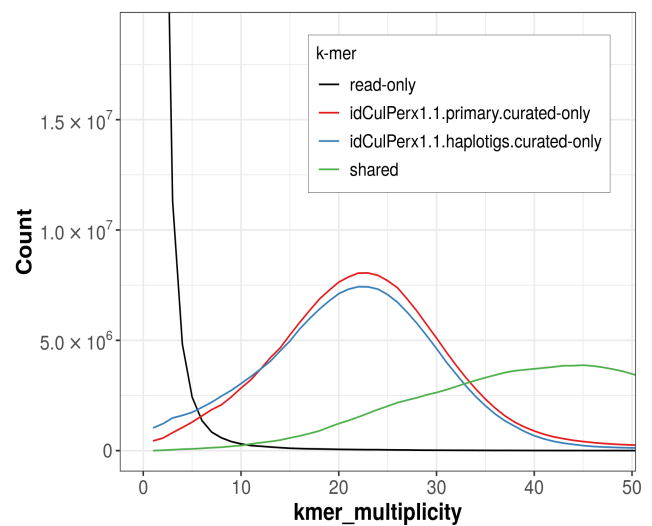
Distribution of k-mer counts per copy numbers found in **haplotigs** (hapl.)



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

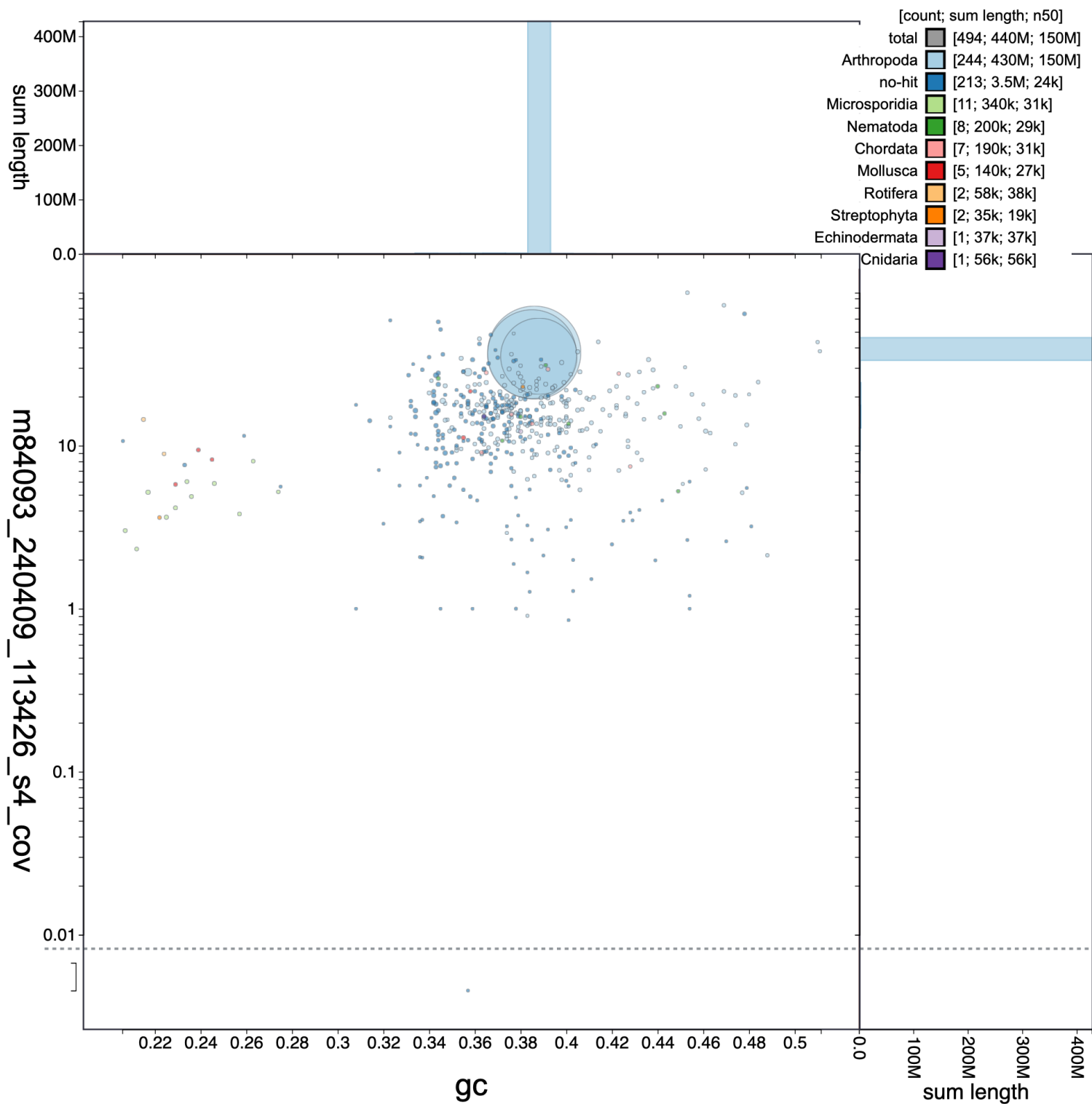


Distribution of k-mer counts per copy numbers found in (hapl.)



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	PacBio HiFi	Arima v2
Coverage	49x	187x

# 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: Thomas Mathers

Affiliation: WSI

Date and time: 2024-06-17 15:42:05 CEST