

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

TxID	545656
ToLID	idCulModel
Species	Culex modestus
Class	Insecta
Order	Diptera

Genome Traits	Expected	Observed
Haploid size (bp)	578,241,120	697,666,902
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.Q54

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

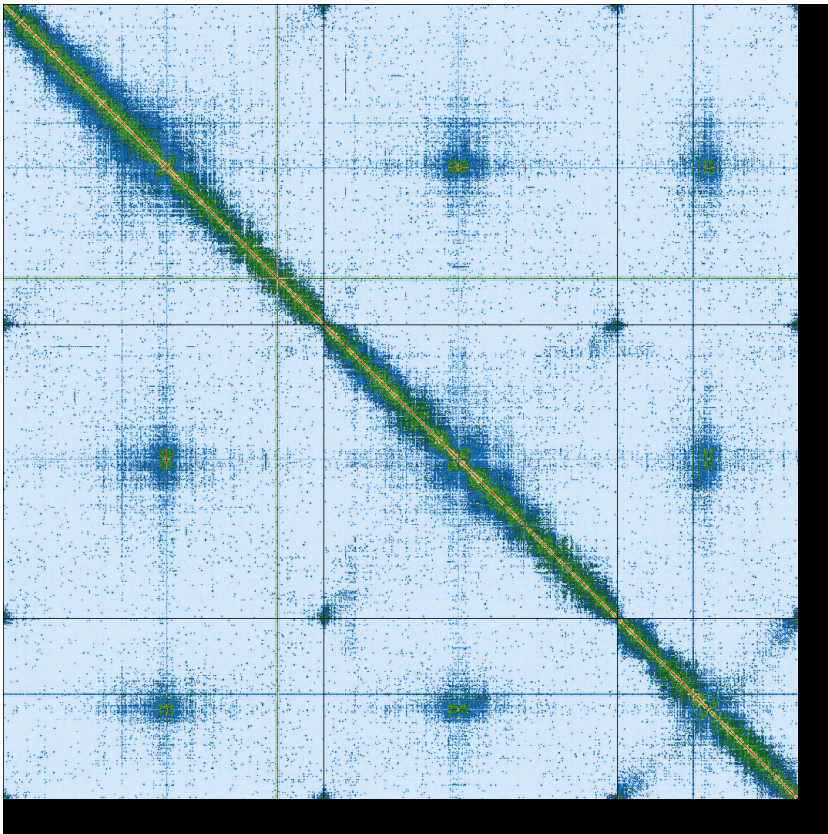
- . Interventions/Gb: 1241
- . Contamination notes: "No contamination detected or removed"
- . Other observations: "PacBio data was from a ULI library, hence a fragmented assembly; Hi-C was also from a different individual (idCulMode3); Chromosomes named by synteny to GCA_963924435.1; This is a version 2 EAR after feedback from the initial version"

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	771,890,934	697,666,902
GC %	36.04	36.15
Gaps/Gbp	4,556.34	4,645.48
Total gap bp	703,400	648,200
Scaffolds	1,301	978
Scaffold N50	515,923,348	246,156,793
Scaffold L50	1	2
Scaffold L90	17	3
Contigs	4,818	4,219
Contig N50	287,270	312,882
Contig L50	787	683
Contig L90	2,755	2,338
QV	54.1	54.1
Kmer compl.	98.93	99.02
BUSCO sing.	90.1%	94.3%
BUSCO dupl.	8.8%	3.7%
BUSCO frag.	0.4%	0.4%
BUSCO miss.	0.7%	1.6%

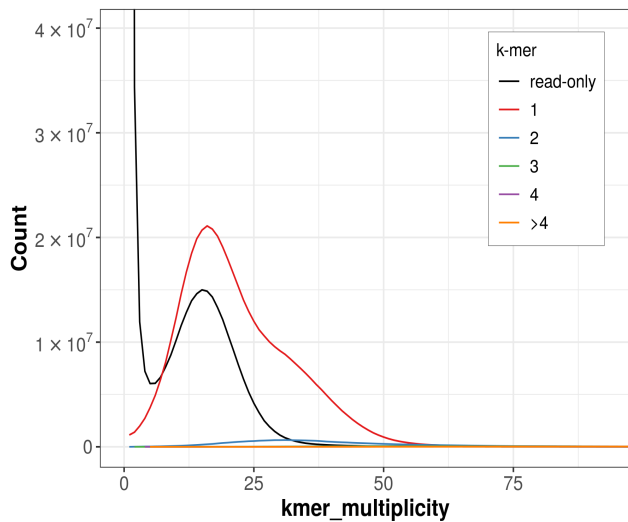
BUSCO 5.3.0 Lineage: insecta_odb10 (genomes:75, BUSCOs:1367)

HiC contact map of curated assembly

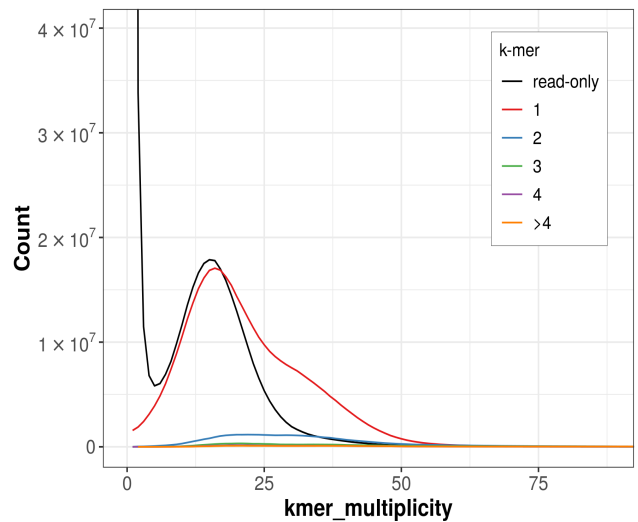


pri [\[LINK\]](#)

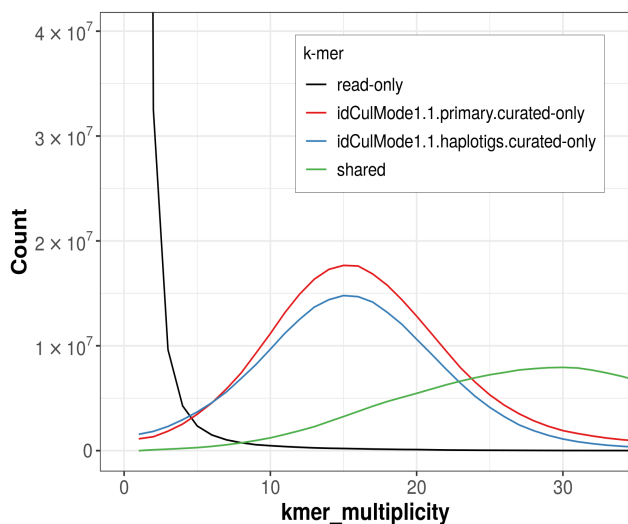
K-mer spectra of curated assembly



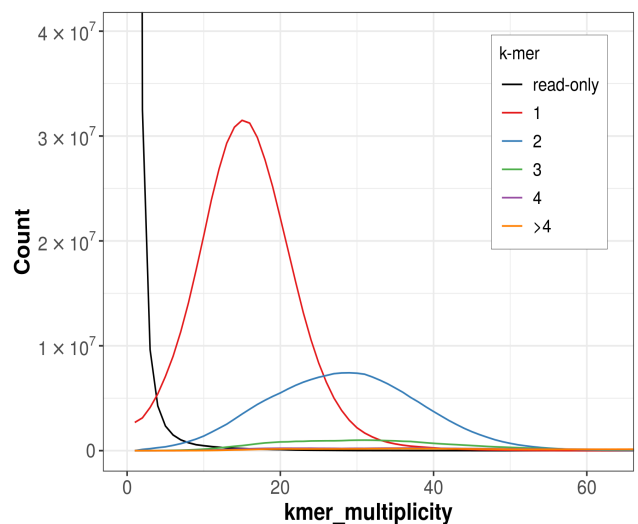
Distribution of k-mer counts per copy numbers found in (hapl.)



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

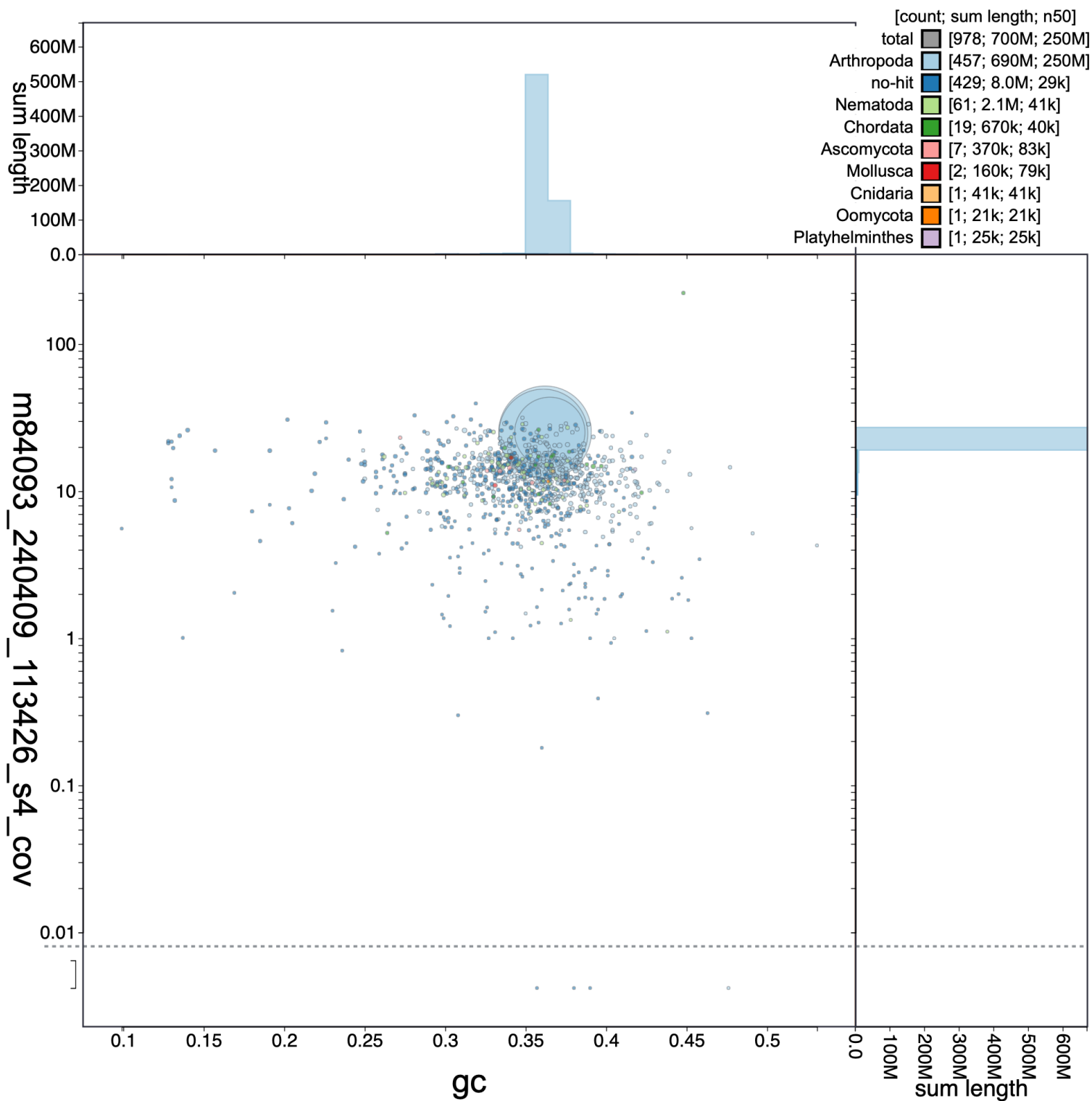


Distribution of k-mer counts coloured by their presence in reads/assemblies



Distribution of k-mer counts per copy numbers found in 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	34x	97x

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: 2024-07-11 11:24:48 CEST