

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

TxID	561267
ToLID	idCulThei1
Species	Culex theileri
Class	Insecta
Order	Diptera

Genome Traits	Expected	Observed
Haploid size (bp)	735,857,783	838,373,201
Haploid Number	3 (source: ancestor)	3
Ploidy	3 (source: ancestor)	2
Sample Sex	F	unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 5.8.Q56

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

- . Observed Ploidy is different from Expected
- . Observed sex is different from Sample sex
- . Assembly length loss > 3% for pri
- . More than 1000 gaps/Gbp for pri

Curator notes

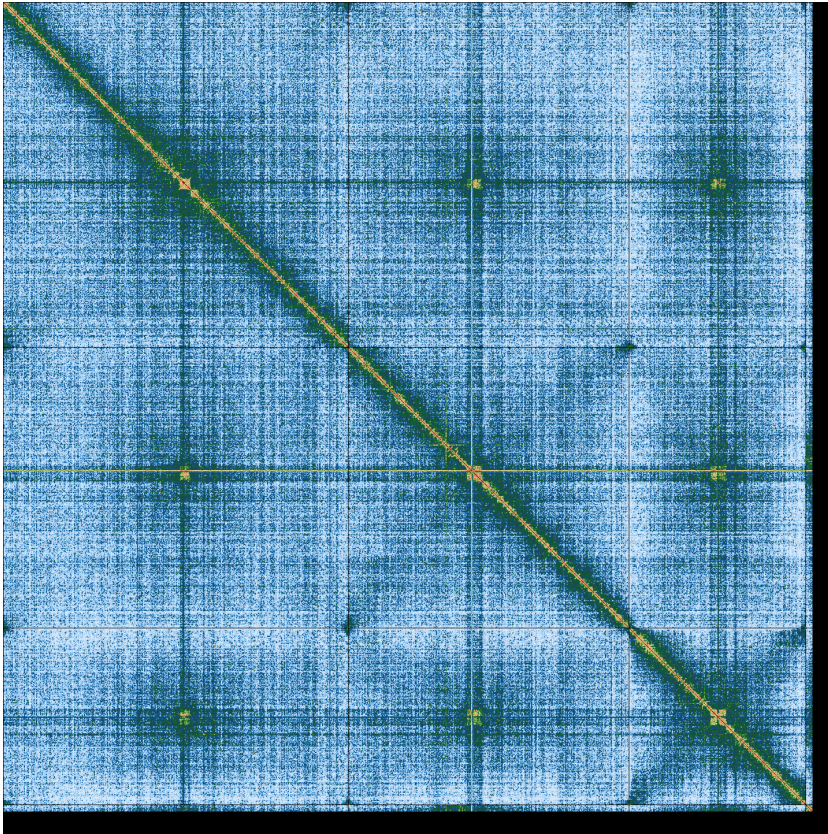
. Interventions/Gb: 440
. Contamination notes: "Total length of scaffolds removed: 434,537 (0.1 %);Scaffolds removed: 16 (2.0 %);Largest scaffold removed: (64,299);FCS-GX contaminant species (number of scaffolds; total length of scaffolds):Acinetobacter guillouiae, g-proteobacteria (15; 419,725);Mitochondrion (1; 14,812)"
. Other observations: "Hi-C mapping + YaHS scaffolding was run twice. PacBio is from a ULI library and Hi-C is from a different individual (idCulThei2). We did not identify the sex chromosome(s) as sequence data from the heterogametic sex was not available and homology is unreliable for sex chromosome identification in Diptera due to frequent sex chromosome turnover (<https://doi.org/10.1371/journal.pbio.1002078>)."

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	865,191,608	838,373,201
GC %	39.12	39.13
Gaps/Gbp	2,115.14	2,449.98
Total gap bp	366,000	411,000
Scaffolds	816	711
Scaffold N50	54,461,784	283,188,383
Scaffold L50	7	2
Scaffold L90	20	3
Contigs	2,646	2,765
Contig N50	788,000	750,539
Contig L50	329	325
Contig L90	1,145	1,157
QV	56.4	56.4
Kmer compl.	99.24	99.38
BUSCO sing.	95.0%	96.9%
BUSCO dupl.	4.0%	2.1%
BUSCO frag.	0.4%	0.4%
BUSCO miss.	0.6%	0.6%

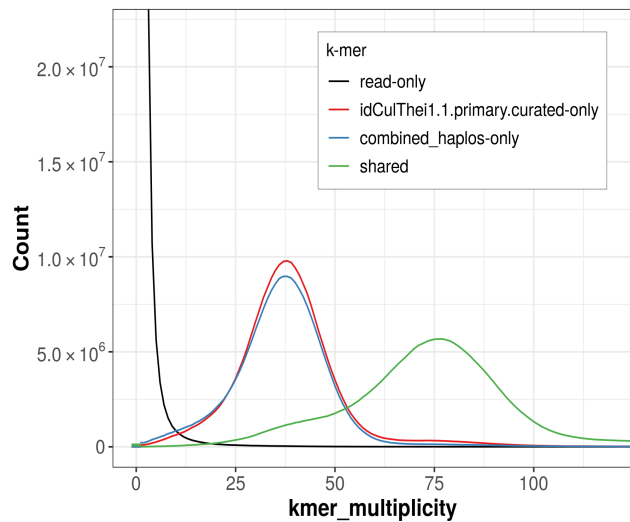
Warning: BUSCO versions or lineage datasets are not the same across results

HiC contact map of curated assembly

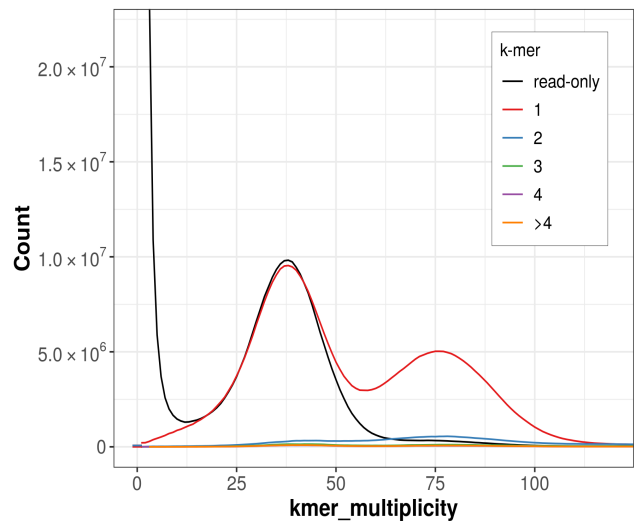


pri [\[LINK\]](#)

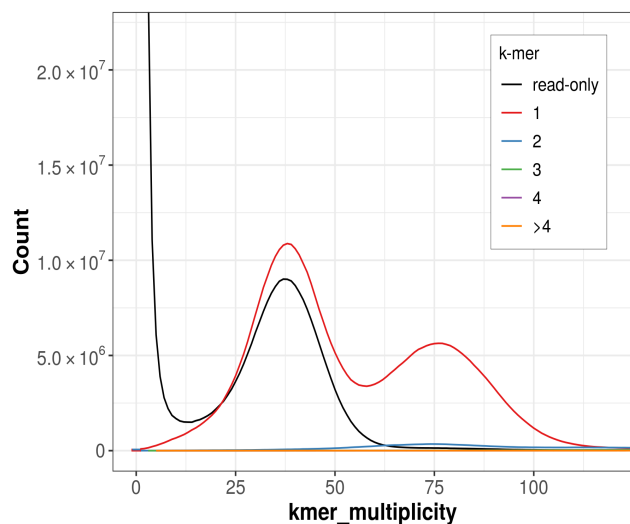
K-mer spectra of curated assembly



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

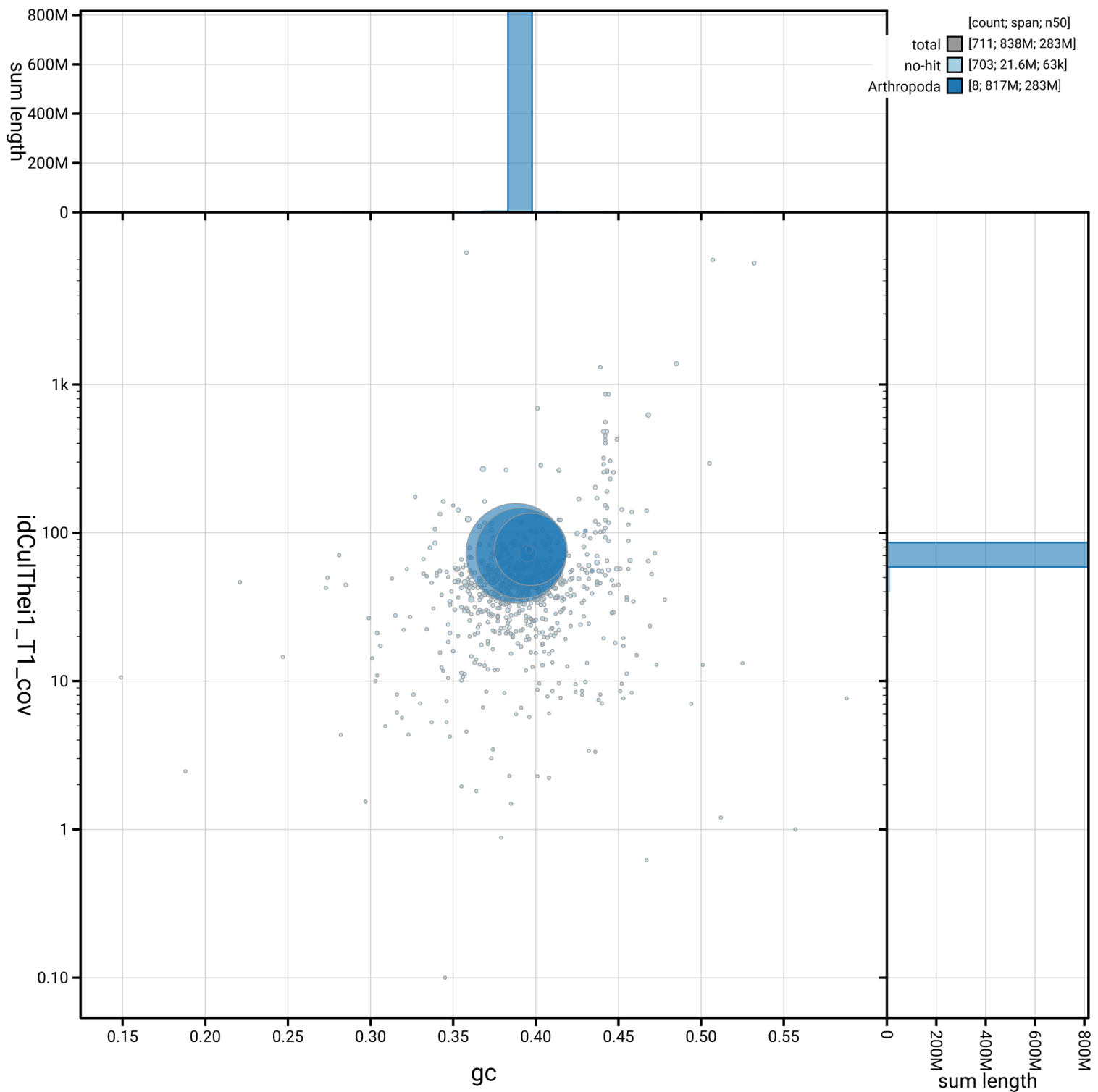


Distribution of k-mer counts per copy numbers found in asm



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	PacBio HiFi	Arima v2
Coverage	78x	140x

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*: 1.1.1
 - |_ *key param*: NA

Submitter: Dominic Absolon

Affiliation: WSI

Date and time: 2024-10-03 22:17:34 CEST