

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

TxID	307205
ToLID	idFanCani4
Species	<i>Fannia canicularis</i>
Class	Insecta
Order	Diptera

Genome Traits	Expected	Observed
Haploid size (bp)	1,930,715,007	1,876,390,951
Haploid Number	6 (source: direct)	6
Ploidy	3 (source: ancestor)	2
Sample Sex	M	M

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 7.8.Q64

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

- . Observed Ploidy is different from Expected

Curator notes

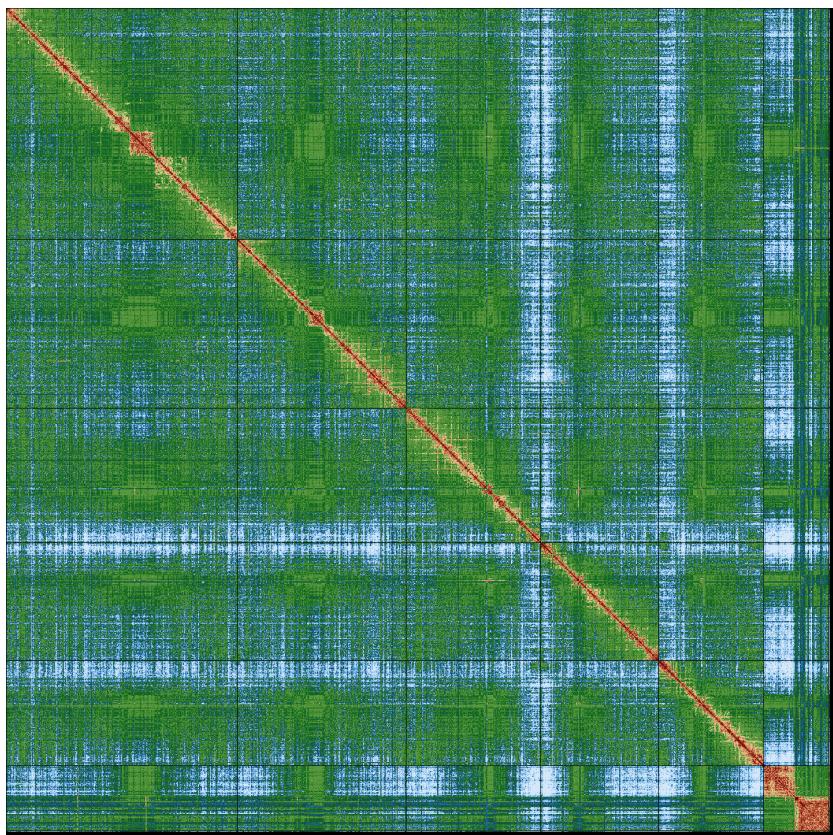
- . Interventions/Gb: 404
- . Contamination notes: "Contamination report for assembly labelled primary; Total length of scaffolds removed: 0 (0.0 %); Scaffolds removed: 0 (0.0 %); Largest scaffold removed: (0); FCS-GX contaminant species (number of scaffolds; total length of scaffolds):"
- . Other observations: "Hi-C was from a different individual (idFanCanil); 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>). The collector confirmed the sample as male, but in *Musca domestica* sex determination is very complex. Maybe something similar occurs in *Fannia canicularis* with XX males carrying male determinating factor on X or autosome (<https://pmc.ncbi.nlm.nih.gov/articles/PMC4524491/pdf/evv128.pdf> [pmc.ncbi.nlm.nih.gov])."

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	1,775,741,293	1,876,390,951
GC %	34.11	34.13
Gaps/Gbp	142.48	195.06
Total gap bp	25,300	48,200
Scaffolds	184	118
Scaffold N50	271,967,115	303,294,413
Scaffold L50	3	3
Scaffold L90	6	5
Contigs	437	484
Contig N50	13,322,456	13,138,360
Contig L50	37	41
Contig L90	138	165
QV	65.1	64.8
Kmer compl.	94.67	97.27
BUSCO sing.	92.8%	97.5%
BUSCO dupl.	4.9%	1.5%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	2.3%	0.9%

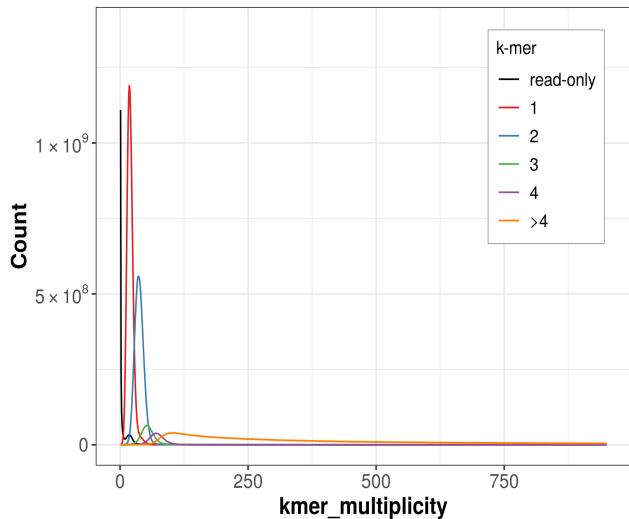
BUSCO 6.0.0 Lineage: insecta_odb10 (genomes:75, BUSCOs:1367)

HiC contact map of curated assembly

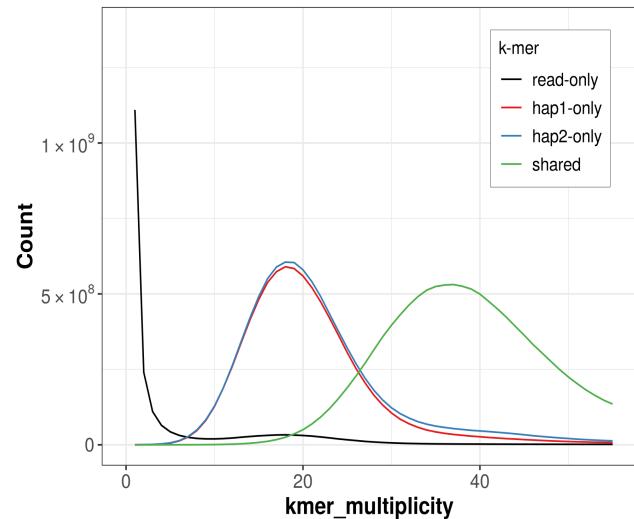


pri [LINK]

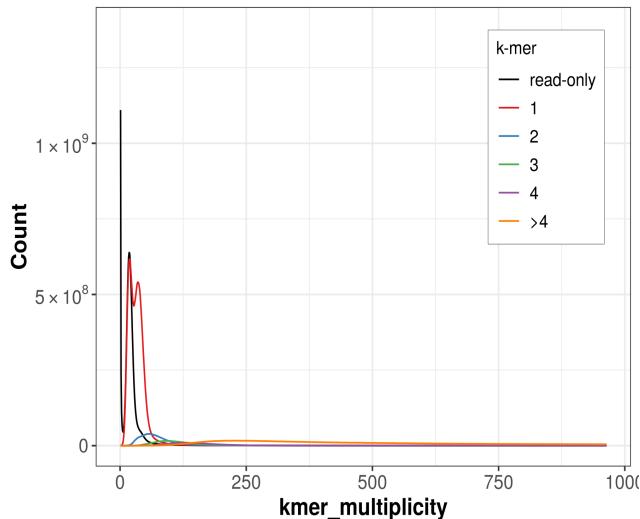
K-mer spectra of curated assembly



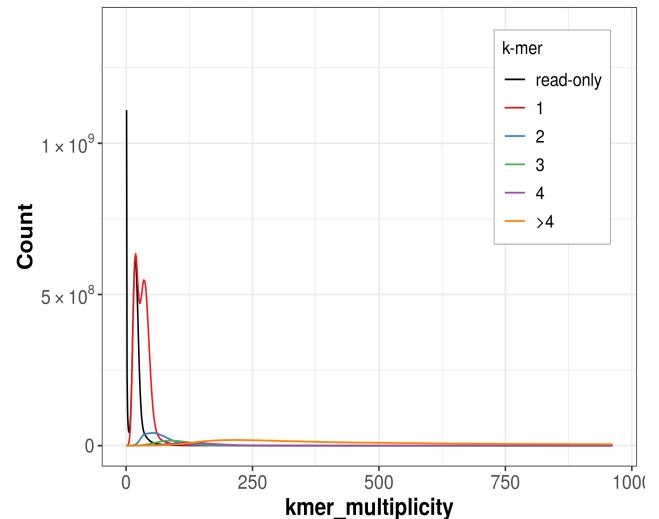
Distribution of k-mer counts per copy numbers found in asm (dipl.)



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

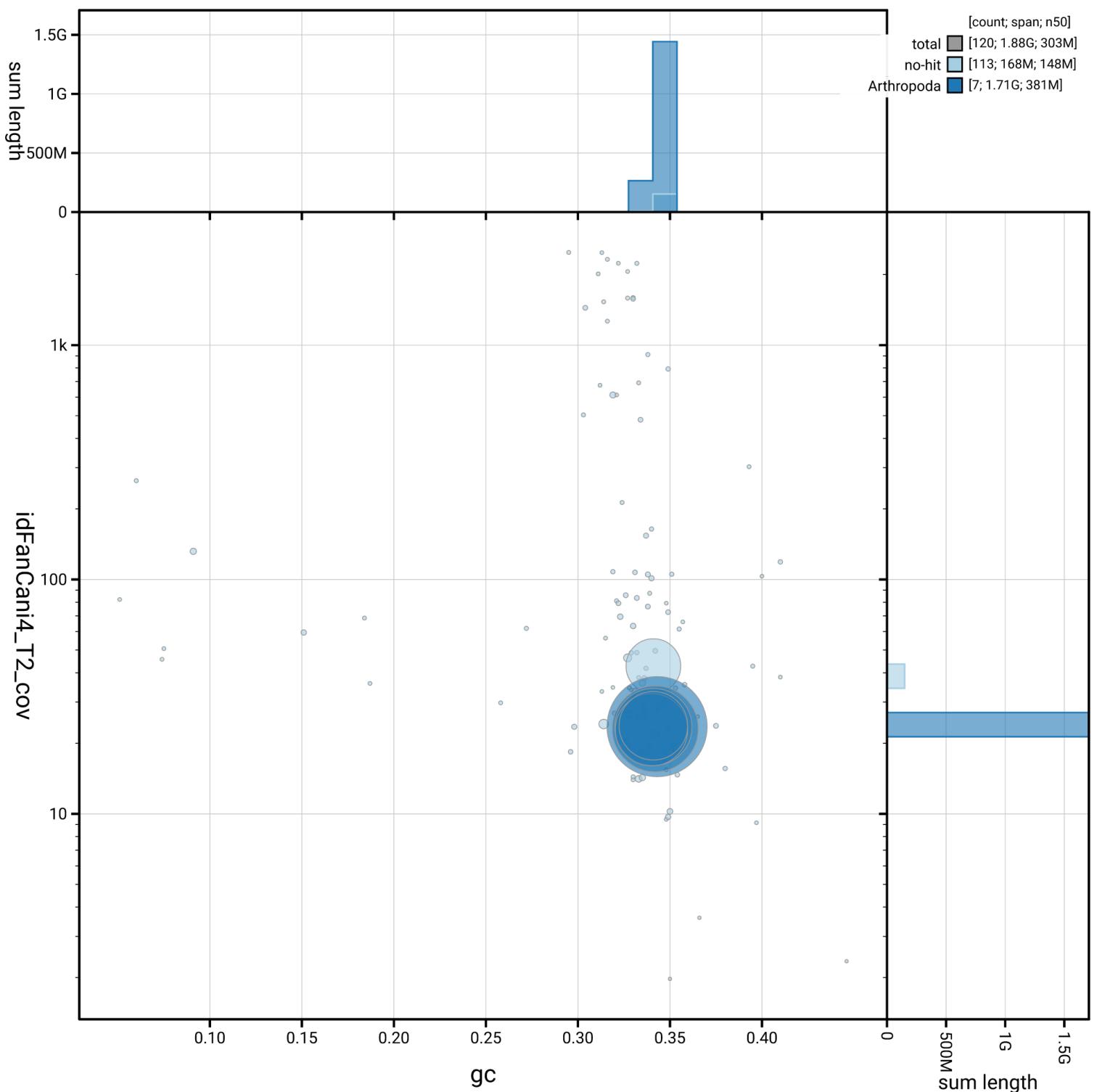


Distribution of k-mer counts per copy numbers found in **hap1** (hapl.)



Distribution of k-mer counts per copy numbers found in **hap2** (hapl.)

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 v3
Coverage	36x	70x

Assembly pipeline

```
- hifiasm
  |_ ver: 0.19.8-r603
  |_ key param: --primary
- yahs
  |_ ver: 1.2.2
  |_ key param: NA
```

Curation pipeline

```
- hifiasm
  |_ ver: 0.19.8-r603
  |_ key param: --primary
- yahs
  |_ ver: 1.2.2
  |_ key param: NA
- TreeVal
  |_ ver: 1.2.1
  |_ key param: NA
```

Submitter: Michael Paulini

Affiliation: WSI

Date and time: 2025-10-13 15:23:13 CEST