

# ERGA Assembly Report

v24.04.03\_beta

Tags: ERGA-BGE

TxID	1902203
ToLID	<b>ihSopOriel1</b>
Species	Sophonia orientalis
Class	Insecta
Order	Hemiptera

Genome Traits	Expected	Observed
Haploid size (bp)	1,526,116,243	1,418,181,991
Haploid Number	9 (source: ancestor)	9
Ploidy	2 (source: ancestor)	2
Sample Sex	F	F

## EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.8.Q63

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

- . Assembly length loss > 3% for pri

## Curator notes

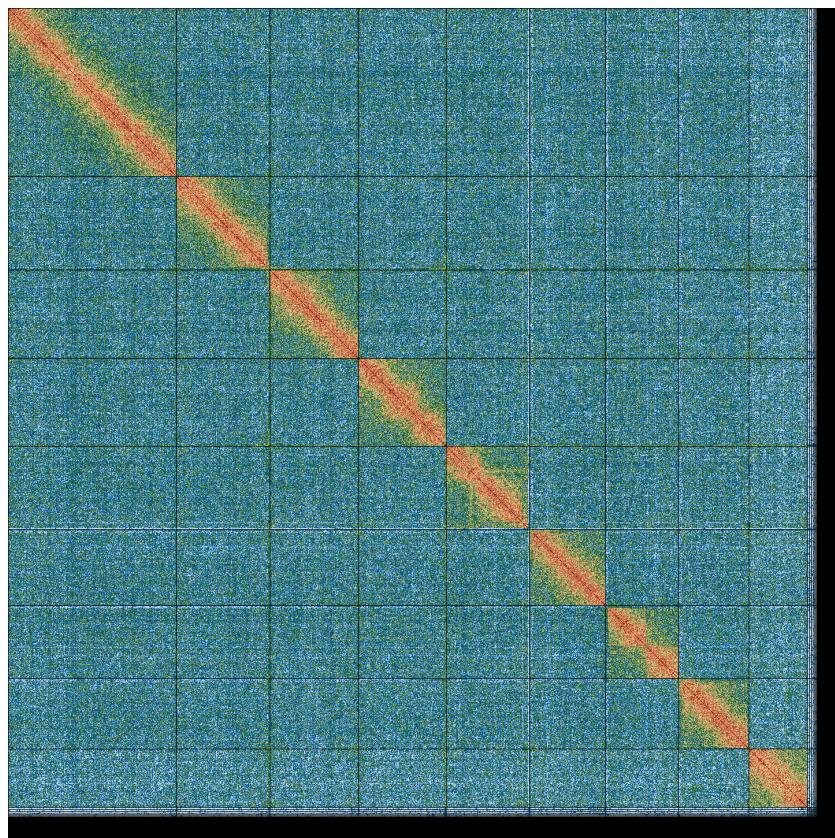
- . Interventions/Gb: 45
- . Contamination notes: "Contamination report for assembly labelled hap1;Total length of scaffolds removed: 10,411,515 (0.7 %);Scaffolds removed: 305 (29.0 %);Largest scaffold removed: (381,165);FCS-GX contaminant species (number of scaffolds; total length of scaffolds):Candidatus Nasuia deltocephalinicola, b-proteobacteria (147; 4,754,856);Candidatus Karelsulcia muelleri, CFB group bacteria (74; 2,408,645);Delftia acidovorans, b-proteobacteria (37; 933,796);Delftia tsuruhatensis, b-proteobacteria (4; 46,586);Cupriavidus basilensis, b-proteobacteria (1; 43,812);Pseudomonas sp. EGD-AKN5, g-proteobacteria (1; 34,674);Curvibacter lanceolatus, b-proteobacteria (1; 23,228);Comamonas odontotermitis, b-proteobacteria (1; 14,537);Mitochondrion (39; 2,151,381)"
- . Other observations: "Hifiasm assembly run in Hi-C phasing mode; X chromosomes were assigned based on synteny to GCA\_964468525."

## Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	1,505,526,806	1,418,181,991
GC %	34.41	34.94
Gaps/Gbp	169.38	180.51
Total gap bp	25,500	28,300
Scaffolds	1,053	395
Scaffold N50	135,243,140	149,258,771
Scaffold L50	5	4
Scaffold L90	13	9
Contigs	1,308	651
Contig N50	8,377,673	9,331,582
Contig L50	49	44
Contig L90	201	164
QV	61.8	63.6
Kmer compl.	99.36	99.33
BUSCO sing.	98.5%	98.5%
BUSCO dupl.	1.3%	1.3%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	0.1%	0.1%

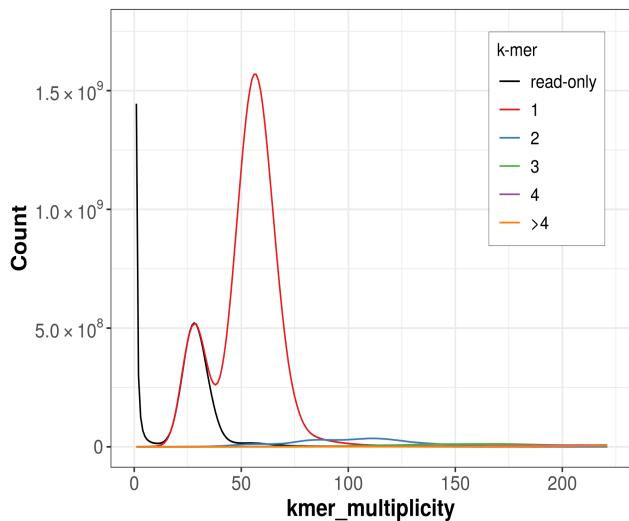
BUSCO 6.0.0 Lineage: insecta\_odb10 (genomes:75, BUSCOs:1367)

# HiC contact map of curated assembly

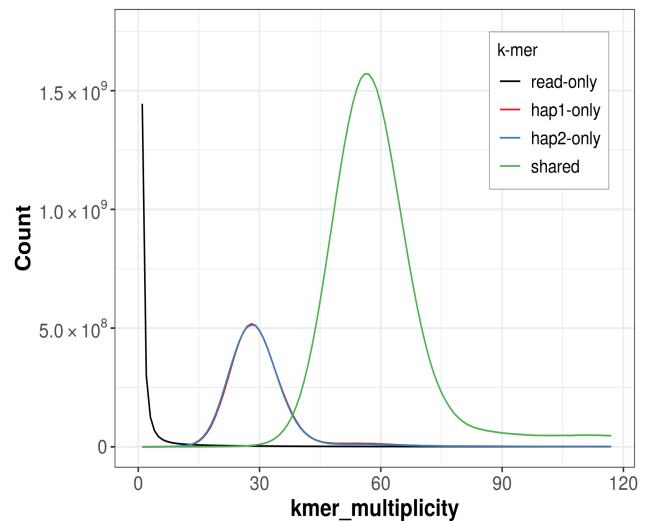


pri [LINK]

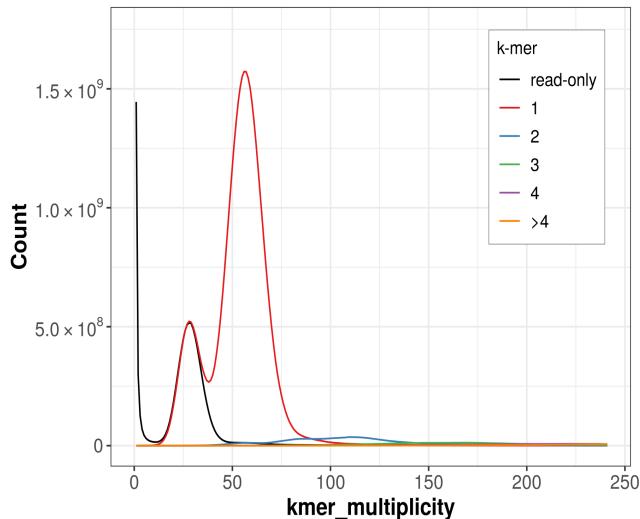
# K-mer spectra of curated assembly



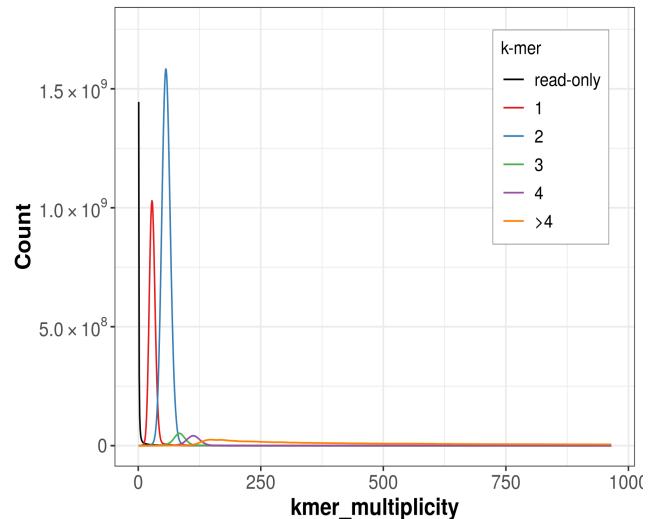
Distribution of k-mer counts per copy numbers found in **hap2** (hapl.)



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

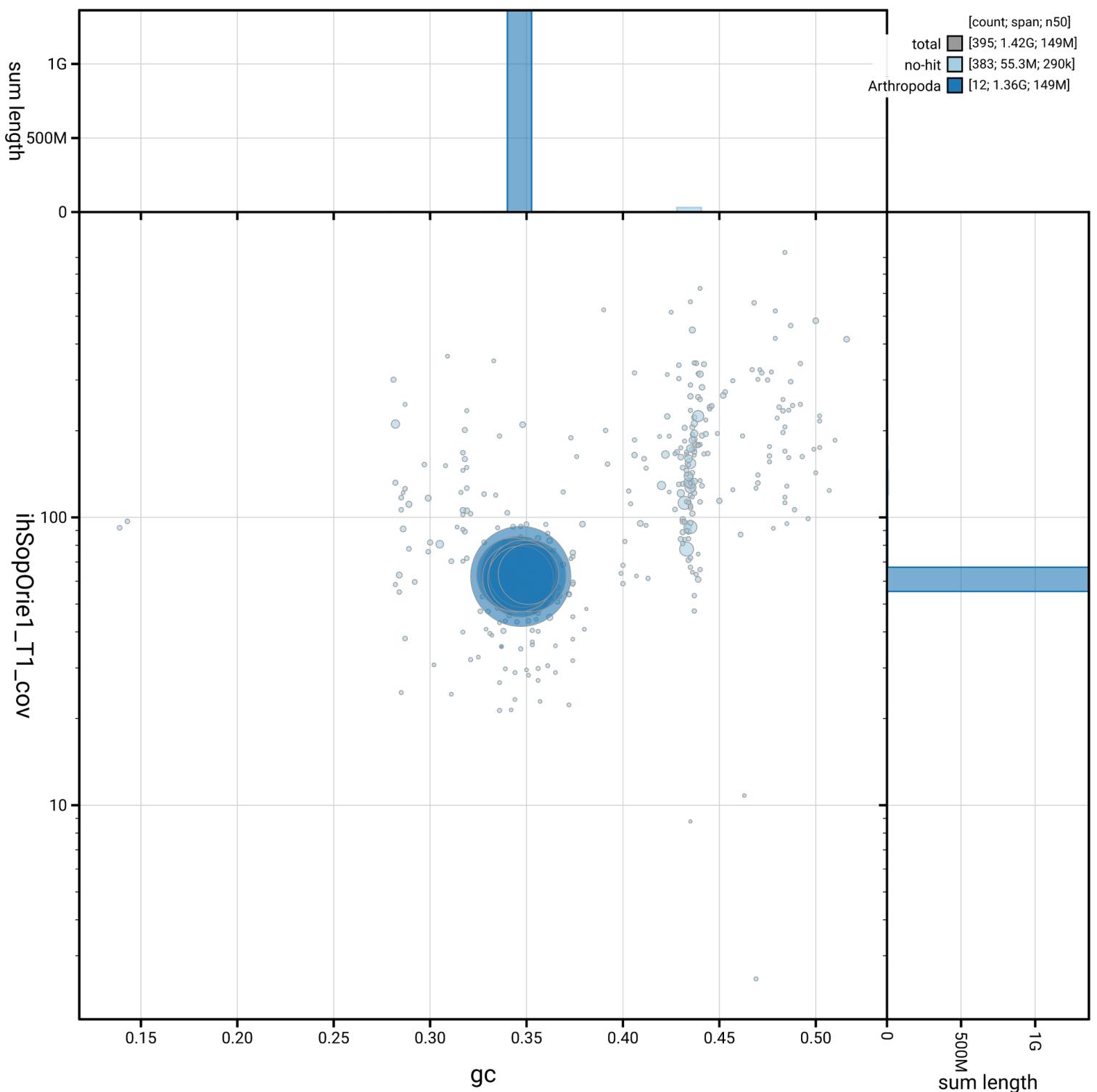


Distribution of k-mer counts per copy numbers found in **hap1** (hapl.)



Distribution of k-mer counts per copy numbers found in **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	56x	848x

## Assembly pipeline

```
- hifiasm
  |_ ver: 0.19.8-r603
  |_ key param: --h1/--h2
- yahs
  |_ ver: 1.2.2
  |_ key param: NA
```

## Curation pipeline

```
- hifiasm
  |_ ver: 0.19.8-r603
  |_ key param: --h1/--h2
- yahs
  |_ ver: 1.2.2
  |_ key param: NA
- TreeVal
  |_ ver: 1.2.1
  |_ key param: NA
```

Submitter: Danil Zilov

Affiliation: WSI

Date and time: 2025-12-03 15:26:47 CET