

# ERGA Assembly Report

v24.04.03\_beta

Tags: ERGA-BGE

TxID	134416
ToLID	<b>ihCicOrni1</b>
Species	Cicada orni
Class	Insecta
Order	Hemiptera

Genome Traits	Expected	Observed
Haploid size (bp)	4,668,429,022	4,991,566,508
Haploid Number	9 (source: ancestor)	10
Ploidy	2 (source: ancestor)	2
Sample Sex	M	M

## EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.8.Q62

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

- . Observed Haploid Number is different from Expected
- . Not 90% of assembly in chromosomes for pri

## Curator notes

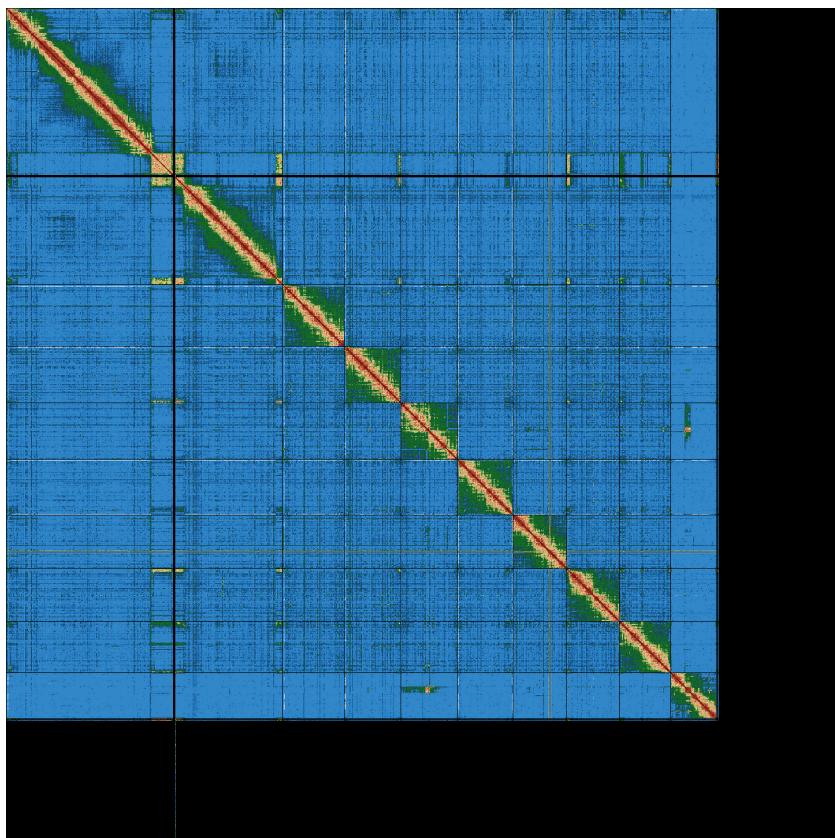
- . Interventions/Gb: 292
- . Contamination notes: "Contamination report for assembly labelled hap1; Total length of scaffolds removed: 10,646,772 (0.2 %); Scaffolds removed: 202 (6.4 %); Largest scaffold removed: (4,261,779); FCS-GX contaminant species (number of scaffolds; total length of scaffolds): Candidatus Karelsulcia muelleri, CFB group bacteria (176; 5,743,823); Arthrobacter cavernae, high GC Gram+ (1; 4,261,779); Hypocreales sp. Mo6-1, ascomycetes (1; 110,855); Mitochondrion (24; 530,315); FCS-Adaptor (1; 30)"
- . Other observations: "Hifiasm assembly run in Hi-C phasing mode; X chromosome identified based on read coverage and copy number in the diploid assembly."

## Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	4,966,805,591	4,991,566,508
GC %	34.57	34.55
Gaps/Gbp	159.66	187.12
Total gap bp	79,300	109,330
Scaffolds	3,177	2,773
Scaffold N50	315,173,651	339,203,014
Scaffold L50	6	5
Scaffold L90	297	131
Contigs	3,970	3,707
Contig N50	6,396,890	6,598,048
Contig L50	217	215
Contig L90	1,012	978
QV	61.8	62.1
Kmer compl.	99.13	98.97
BUSCO sing.	95.9%	97.3%
BUSCO dupl.	2.0%	2.0%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	2.0%	0.7%

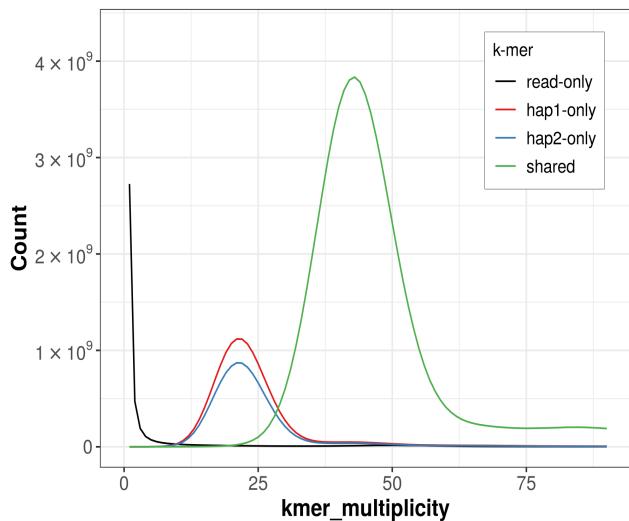
BUSCO 6.0.0 Lineage: insecta\_odb10 (genomes:75, BUSCOs:1367)

# HiC contact map of curated assembly

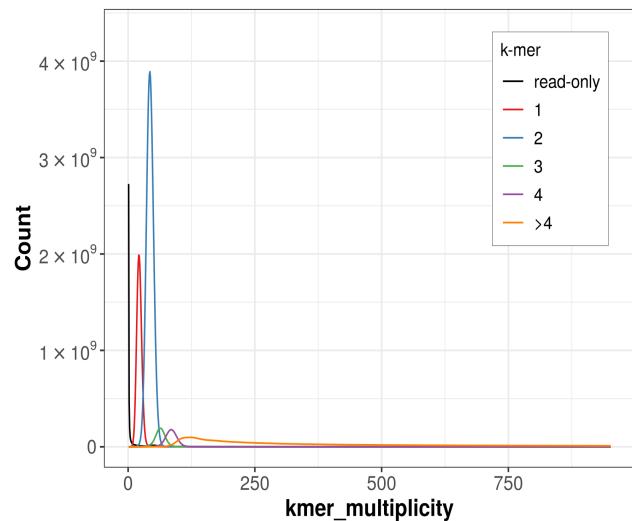


pri [LINK]

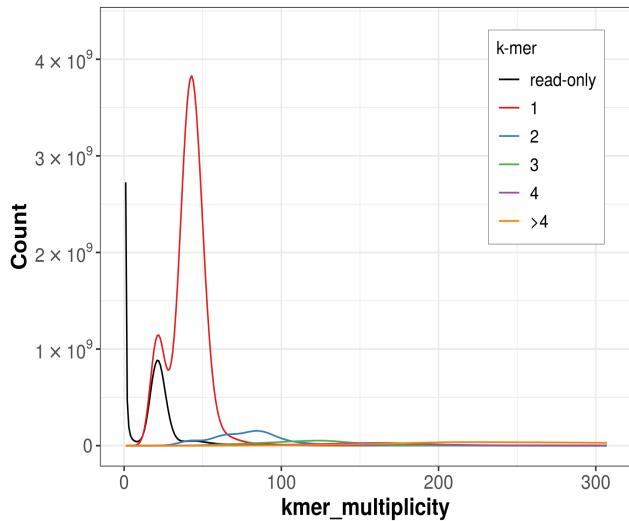
# K-mer spectra of curated assembly



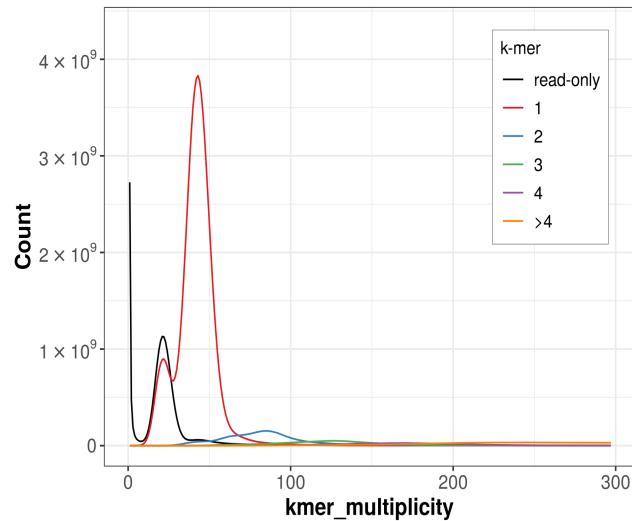
Distribution of k-mer counts coloured by their presence in reads/assemblies



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

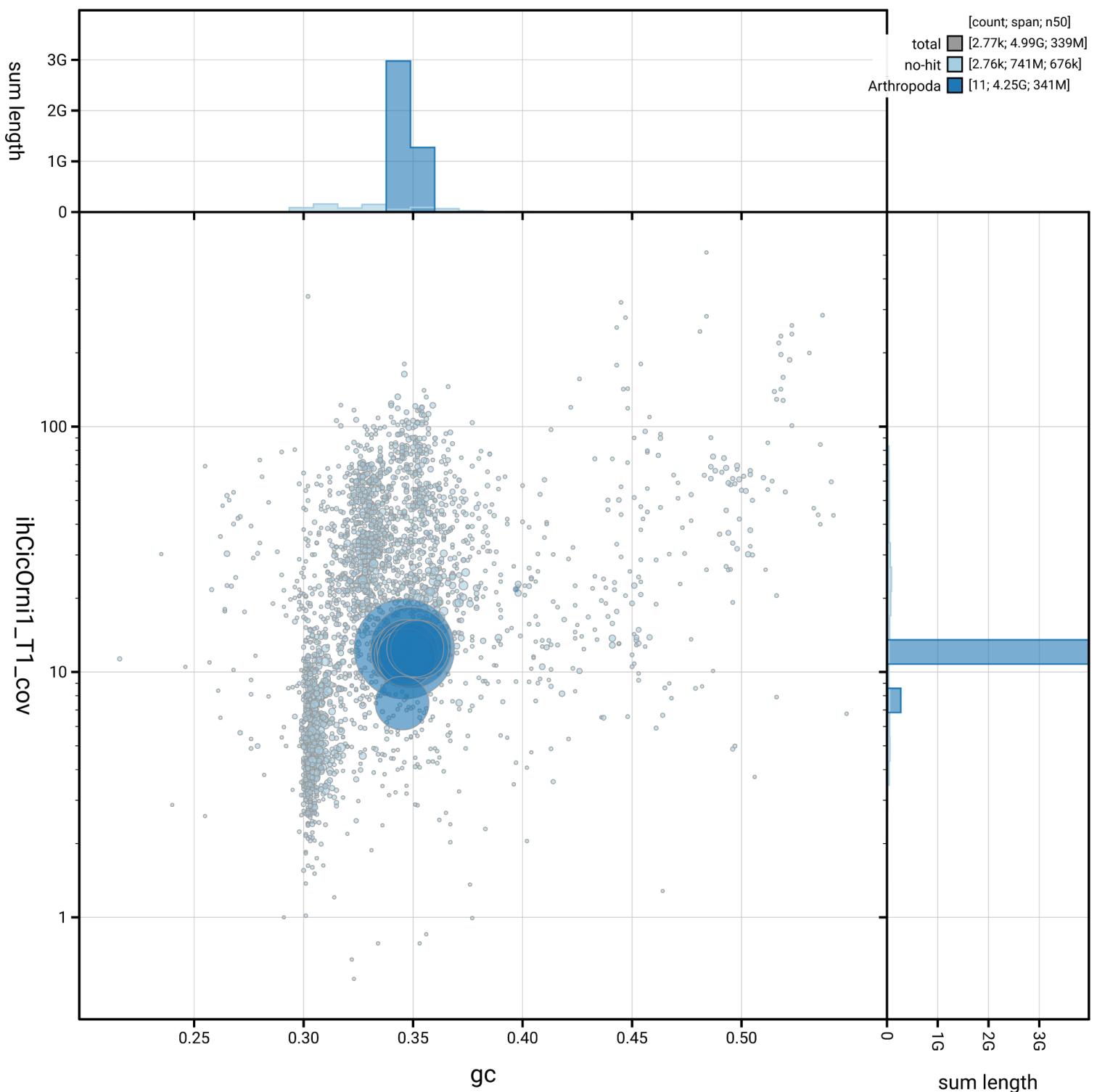


Distribution of k-mer counts per copy numbers found in **hap1** (hapl.)



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 v2
Coverage	42x	251x

## 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: Thomas Mathers

Affiliation: WSI

Date and time: 2026-01-27 15:00:43 CET