

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	3229133
ToLID	<b>iyCatAphr12</b>
Species	Cataglyphis aphrodite
Class	Insecta
Order	Hymenoptera

Genome Traits	Expected	Observed
Haploid size (bp)	206,071,442	209,631,910
Haploid Number	26 (source: ancestor)	33
Ploidy	1 (source: ancestor)	1
Sample Sex	unknown	unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.6.Q61

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

- . Observed Haploid Number is different from Expected
- . Observed sex is different from Sample sex

### Curator notes

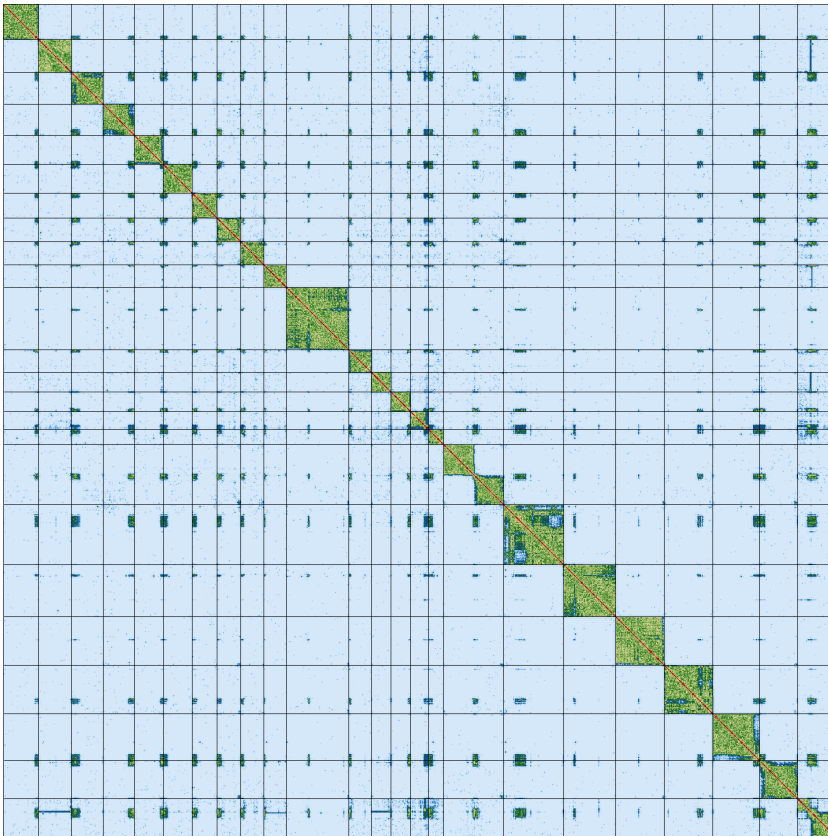
- . Interventions/Gb: 5
- . Contamination notes: "FCS-GX detected 2 contigs: Candidatus Nanopelagicus limnes and Tsukamurella tyrosinosolvens"
- . Other observations: "Straightforward assembly. PacBio HiFi reads were subsampled to 60x coverage. HiFiasm (yield: 219Mb, N50: 2.5M) + FCS-GX (removed 2 contigs, 25Kb) + purge\_dups (yield: 210Mb, N50: 2.5M) + yahs + curation. No major issues. I could not identify any sex chromosome. The PacBio coverage is uniformly distributed over all chromosomes and I could not find any other ant assembly with known sex chromosomes."

# Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	210,149,118	209,631,910
GC %	34.64	34.64
Gaps/Gbp	590.06	610.59
Total gap bp	24,800	25,600
Scaffolds	41	26
Scaffold N50	9,432,304	9,569,503
Scaffold L50	9	8
Scaffold L90	20	20
Contigs	165	154
Contig N50	2,598,554	2,598,554
Contig L50	26	26
Contig L90	83	82
QV	61.394	61.4242
Kmer compl.	98.1756	98.1153
BUSCO sing.	98.6%	98.6%
BUSCO dupl.	0.1%	0.1%
BUSCO frag.	0.7%	0.7%
BUSCO miss.	0.6%	0.6%

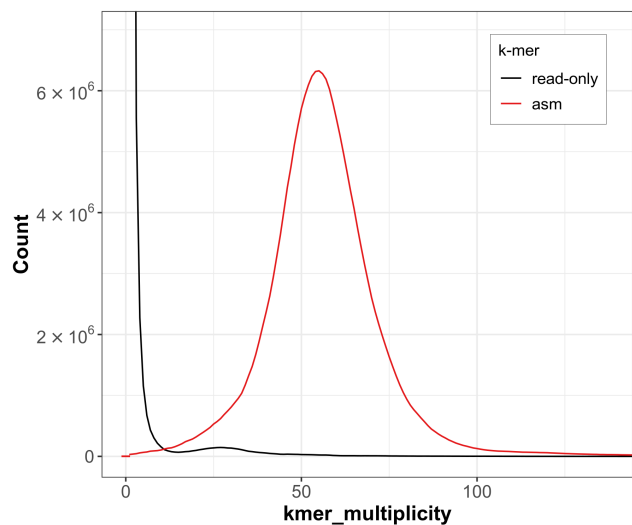
BUSCO: 5.8.3 (euk\_genome\_min, miniprot) / Lineage: formicidae\_odb12 (genomes:24, BUSCOs:7266)

# HiC contact map of curated assembly

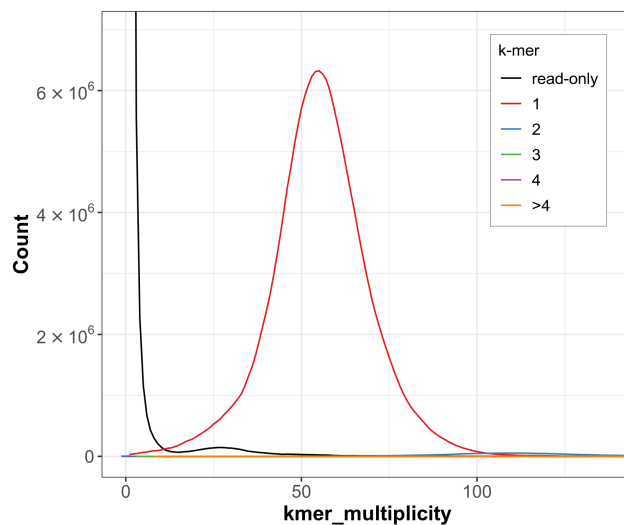


pri [\[LINK\]](#)

# K-mer spectra of curated assembly

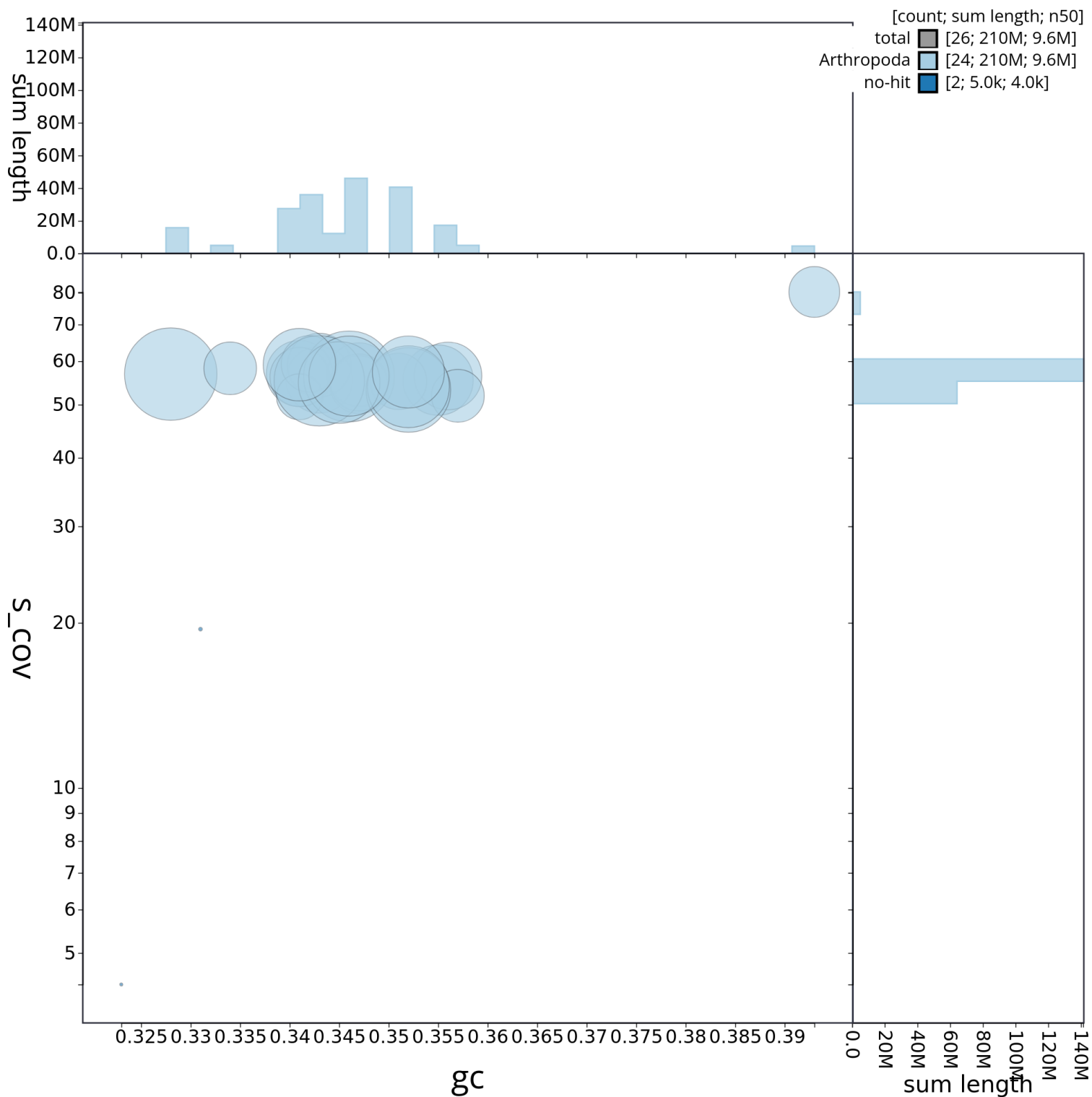


Distribution of k-mer counts coloured by their presence in reads/assemblies



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

# 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	HiFi	HiC
Coverage	276x	816x

## Assembly pipeline

- **Hifiasm**
  - |\_ *ver*: 0.25.0-r726
  - |\_ *key param*: 13
- **purge\_dups**
  - |\_ *ver*: 1.2.6
  - |\_ *key param*: NA

## Curation pipeline

- **GRIT\_Rapid**
  - |\_ *ver*: 1a3d79a8
  - |\_ *key param*: NA
- **HiGlass**
  - |\_ *ver*: 0.10.4
  - |\_ *key param*: NA

Submitter: Martin Pippel

Affiliation: SciLifeLab

Date and time: 2025-10-08 13:37:04 CEST