ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	3229133
ToLID	**iyCatAphr12**
Species	Cataglyphis aphrodite
Class	Insecta
Order	Hymenoptera

Genome Traits	Expected	Observed
Haploid size (bp)	206,071,442	209,631,510
Haploid Number	26 (source: ancestor)	26
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:

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. In a second curation round the assembly was mapped to GCF_021464435.1 and GCF_021464445.1 which revealed that scaffold_2 and scaffold_9 needed to be split. Therefore the final chromosome number is now 26."

Quality metrics table

| Metrics	Pre-curation pri	Curated pri
Total bp	210,149,118	209,631,510
GC %	34.64	34.64
Gaps/Gbp	590.06	601.05
Total gap bp	24,800	25,200
Scaffolds	41	28
Scaffold N50	9,432,304	8,279,055
Scaffold L50	9	9
Scaffold L90	20	22
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)
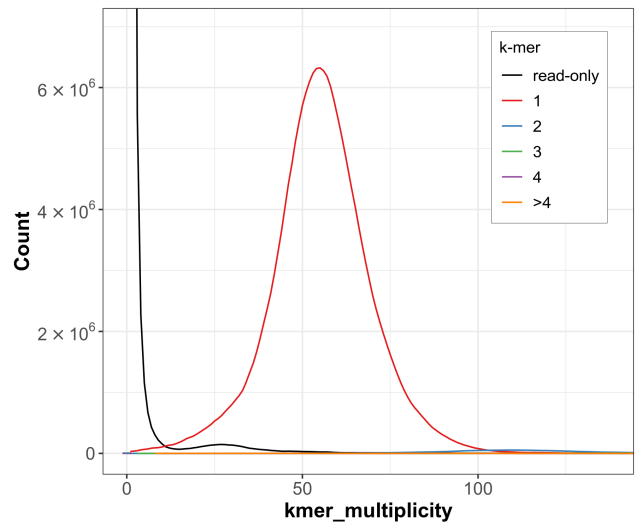
HiC contact map of curated assembly
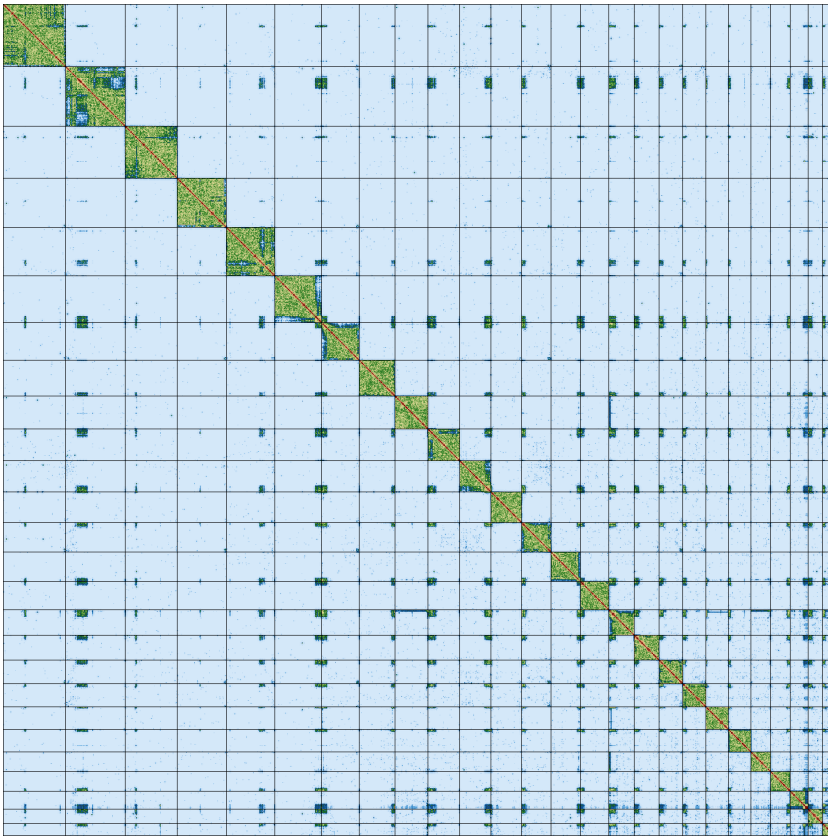
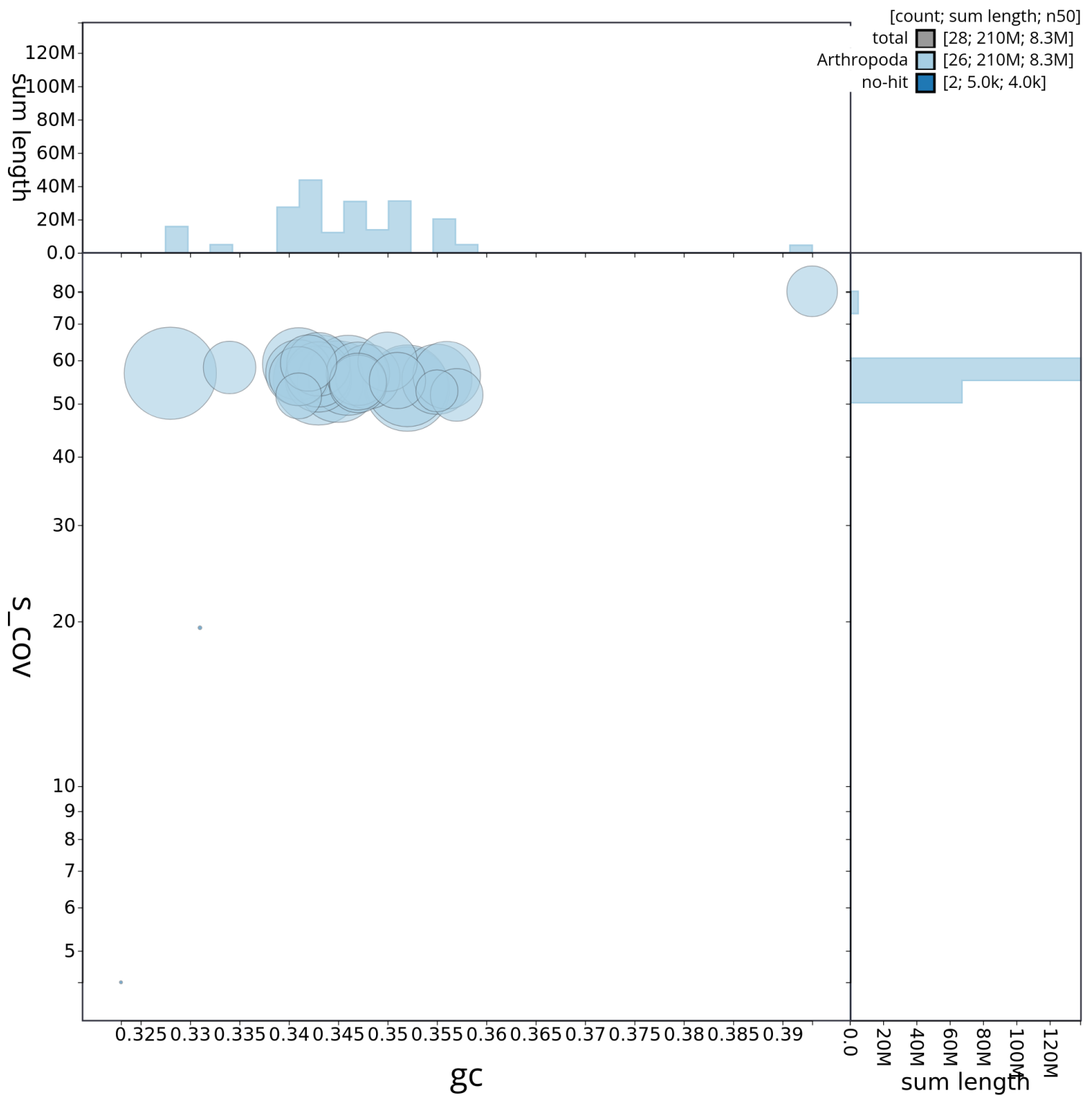**pri** [LINK]

K-mer spectra of curated assembly
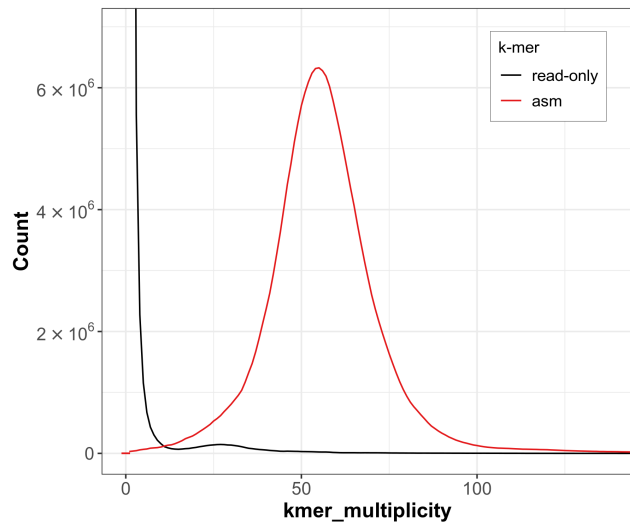
Distribution of k-mer counts coloured by their presence in reads/assemblies

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:* l3
- **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: 2025-10-13 13:56:47 CEST