

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	3229145
ToLID	<b>iyColCypr3</b>
Species	<i>Colletes cyprius</i>
Class	Insecta
Order	Hymenoptera

Genome Traits	Expected	Observed
Haploid size (bp)	347,125,941	289,597,081
Haploid Number	11 (source: ancestor)	14
Ploidy	1 (source: ancestor)	2
Sample Sex	Female	NA

## EBP metrics summary and curation notes

Obtained EBP quality metric for hap1: 7.7.Q69

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

- . Observed Haploid Number is different from Expected
- . Observed Ploidy is different from Expected
- . Observed sex is different from Sample sex
- . Kmer completeness value is less than 90 for hap1
- . Assembly length loss > 3% for hap1

## Curator notes

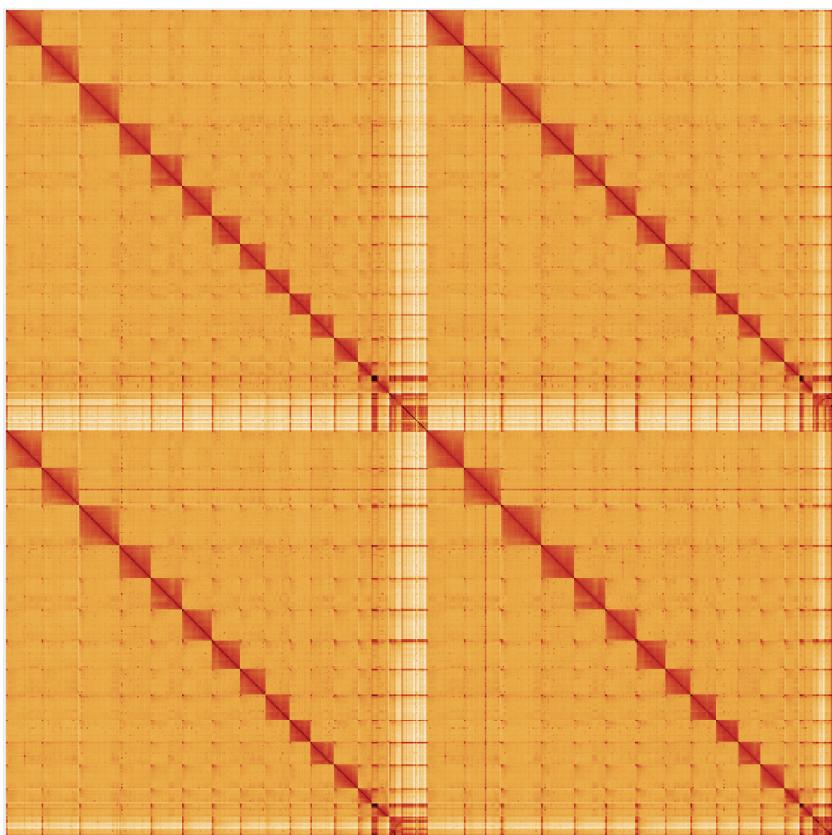
- . Interventions/Gb: 1
- . Contamination notes: "Only one contig was split to make an unloc, otherwise all other contigs were left as chromosomes. Slight uncertainty about whether there are 13 or 14 chromosomes in this species, but I left the two smallest scaffolds as two separate chromosomes without strong evidence for joining them. Approx 35Mb of contaminant sequence were removed"
- . Other observations: "None"

## Quality metrics table

Metrics	Pre-curation hap1	Curated hap1
Total bp	304,177,643	289,597,081
GC %	40.39	40.89
Gaps/Gbp	9.86	10.36
Total gap bp	300	300
Scaffolds	320	302
Scaffold N50	19,470,507	19,287,579
Scaffold L50	7	7
Scaffold L90	17	14
Contigs	323	305
Contig N50	17,189,520	17,189,520
Contig L50	7	7
Contig L90	20	16
QV	69.7979	69.8373
Kmer compl.	91.0673	86.8134
BUSCO sing.	99.3%	99.3%
BUSCO dupl.	0.2%	0.2%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	0.5%	0.5%

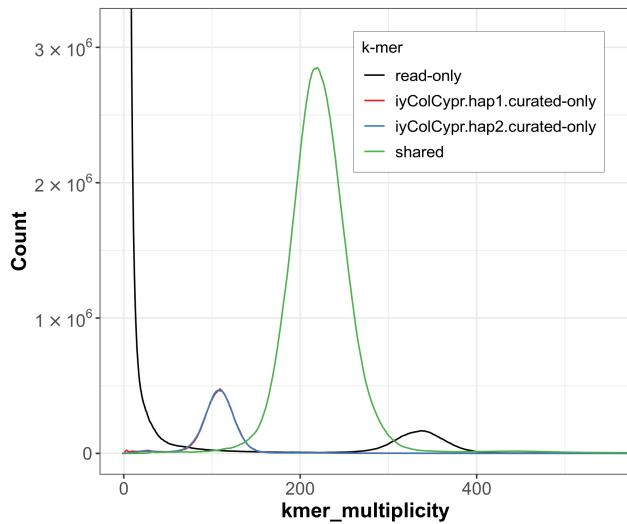
BUSCO: 6.0.0 (euk\_genome\_min, miniprot) / Lineage: hymenoptera\_odb12 (genomes:78, BUSCOs:5920)

# HiC contact map of curated assembly

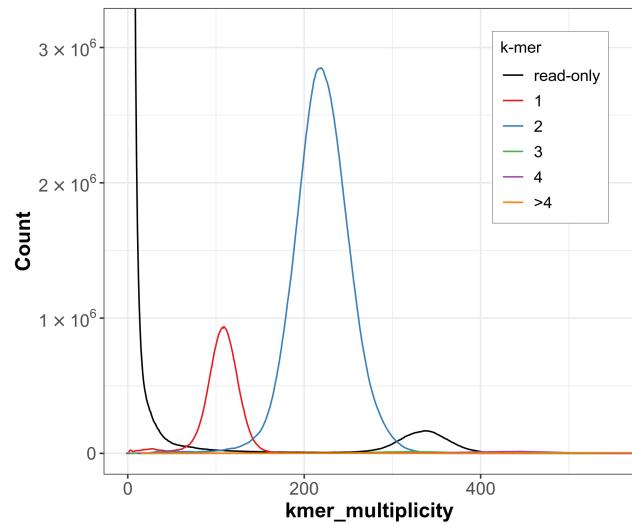


hap1 [\[LINK\]](#)

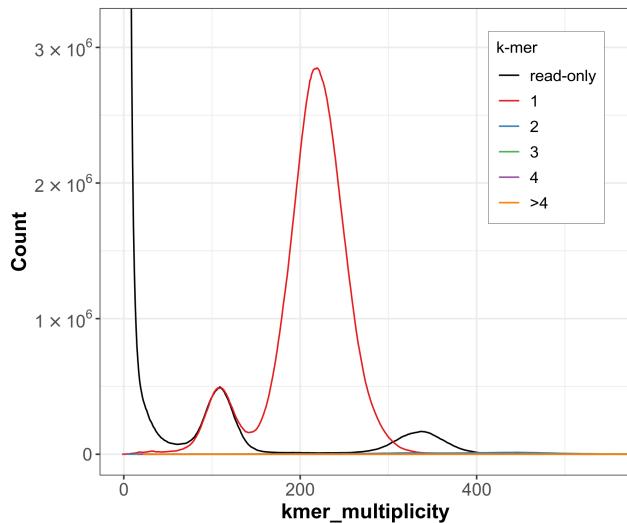
# K-mer spectra of curated assembly



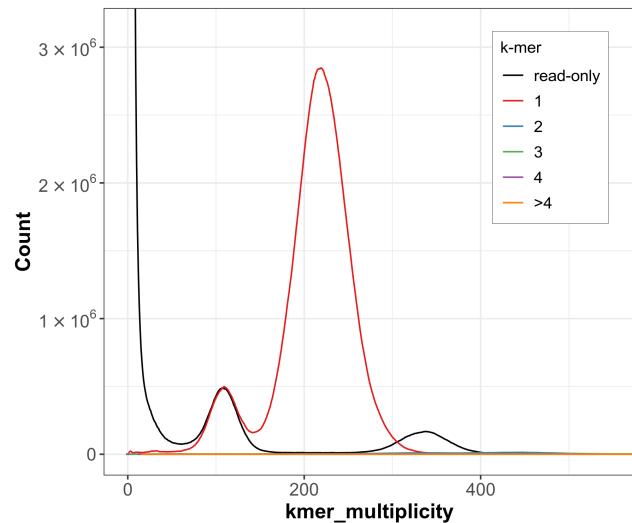
Distribution of k-mer counts coloured by their presence in reads/assemblies



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

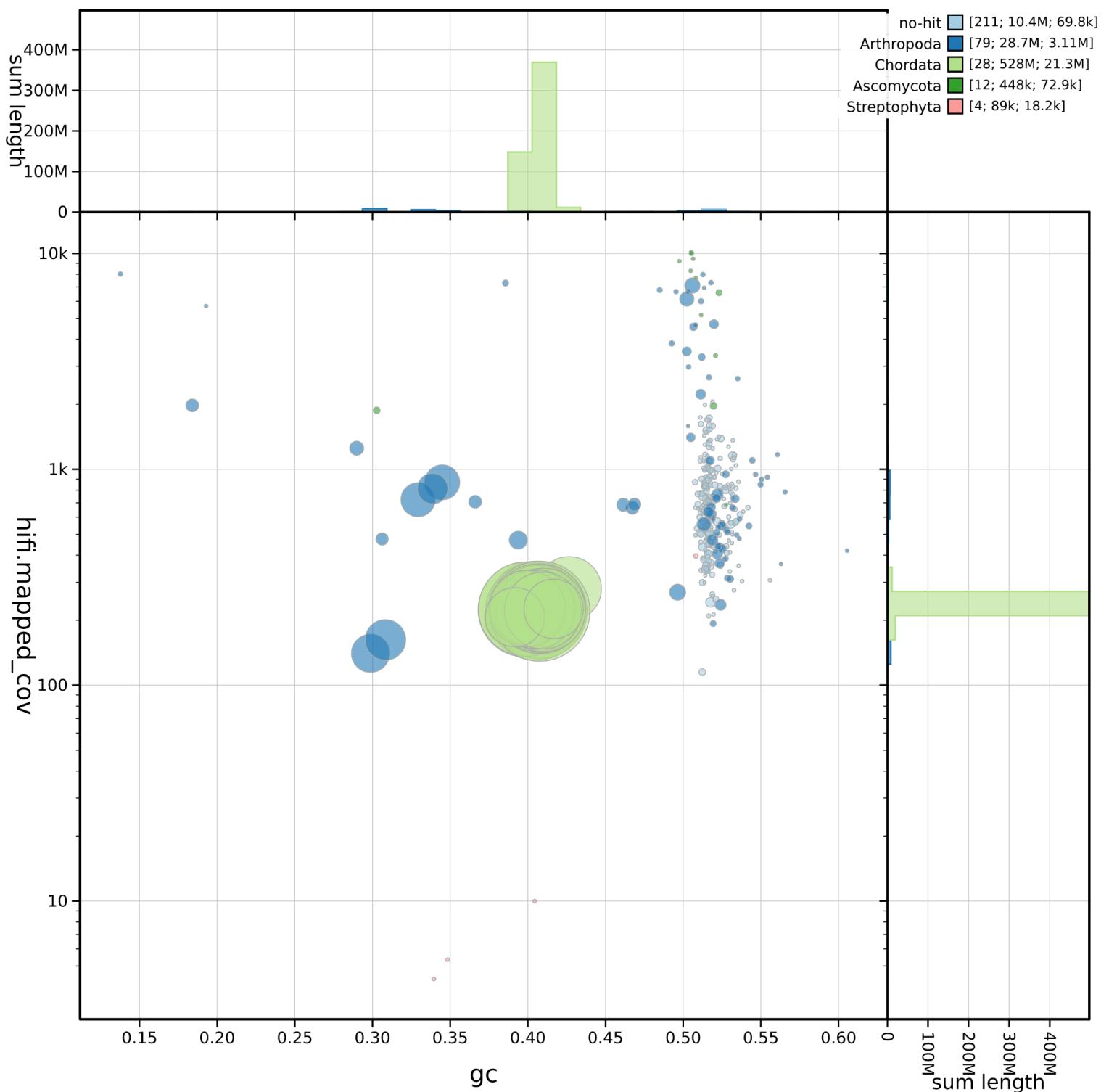


Distribution of k-mer counts per copy numbers found in **hap2** (hapl.)



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

# Post-curation contamination screening



**hap1.** 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	HiC
Coverage	280X	1000X

## Assembly pipeline

```
- hifiasm
  |_ ver: 0.25.0
  |_ key param: --h1 --h2
- purge-dups
  |_ ver: 1.2.6
  |_ key param: NA
- yahs
  |_ ver: 1.2a
  |_ key param: NA
- fcs-gx
  |_ ver: 0.5.4
  |_ key param: NA
- blob toolkit
  |_ ver: 4..40
  |_ key param: NA
```

## Curation pipeline

```
- PretextViewAI
  |_ ver: 1.0
  |_ key param: NA
```

Submitter: Tom Brown

Affiliation: IZW

Date and time: 2025-12-02 14:35:18 CET