

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	2828945
ToLID	<b>ilHipCypr1</b>
Species	Hipparchia cypriensis
Class	Insecta
Order	Lepidoptera

Genome Traits	Expected	Observed
Haploid size (bp)	382,535,333	386,736,412
Haploid Number	29 (source: ancestor)	28
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	ZZ

## EBP metrics summary and curation notes

Obtained EBP quality metric for hap1: 7.7.Q65

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
- . Kmer completeness value is less than 90 for hap1

### Curator notes

. Interventions/Gb: 5  
. Contamination notes: "33 contaminants (pseudomonadota, bacillota, nematoda) detected with blobtools and removed."  
. Other observations: "Curation made 0 cuts in contigs, 1 break at a gap and 1 join. SUPER\_5 identified as Z postcuration -- will tag and fix for final EAR and ENA submission."

# Quality metrics table

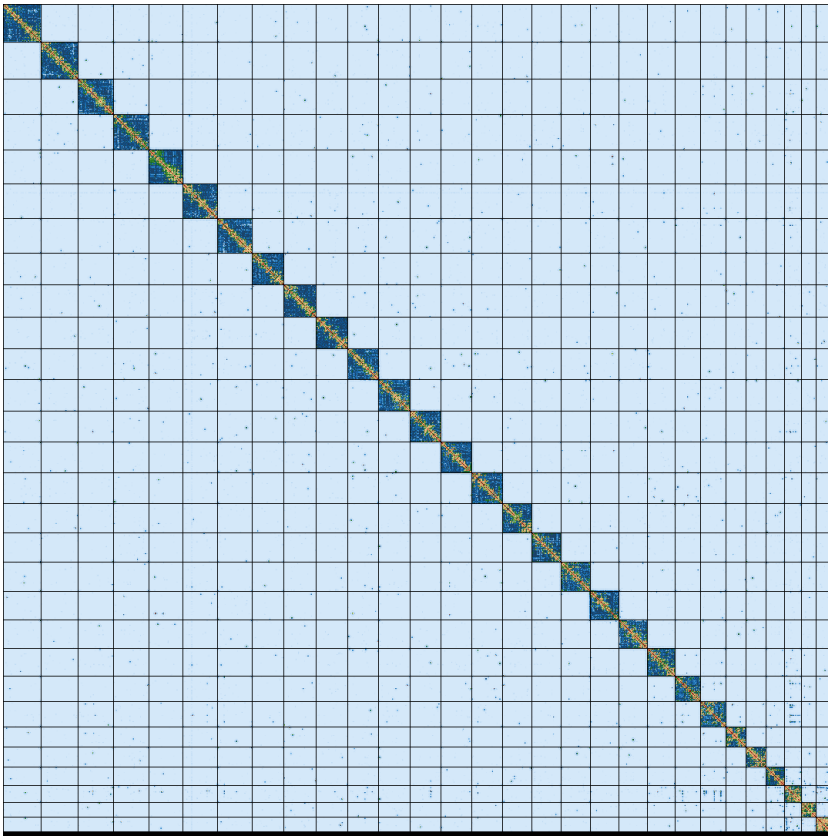
Metrics	Pre-curation hap1	Curated hap1
Total bp	391,457,550	386,736,412
GC %	37.09	36.97
Gaps/Gbp	5.11	5.17
Total gap bp	400	400
Scaffolds	110	77
Scaffold N50	14,412,884	14,412,884
Scaffold L50	13	13
Scaffold L90	25	25
Contigs	112	79
Contig N50	14,412,884	14,412,884
Contig L50	13	13
Contig L90	25	25
QV	62.7713	65.2577
Kmer compl.	67.8475	67.8229
BUSCO sing.	98.2%	98.6%
BUSCO dupl.	0.3%	0.3%
BUSCO frag.	0.3%	0.1%
BUSCO miss.	1.2%	1.0%

Warning! BUSCO versions or lineage datasets are not the same across results:

BUSCO: 5.5.0 (euk\_genome\_met, metaeuk) / Lineage: lepidoptera\_odb10 (genomes:16, BUSCOs:5286)

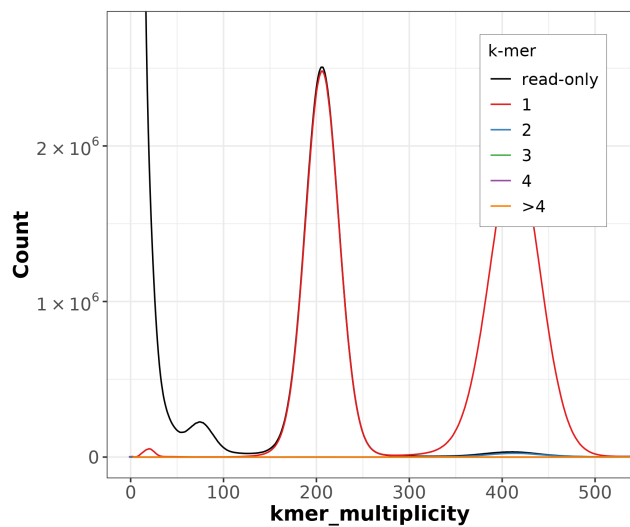
BUSCO: 6.0.0 (euk\_genome\_min, miniprot) / Lineage: lepidoptera\_odb12 (genomes:79, BUSCOs:5760)

# HiC contact map of curated assembly

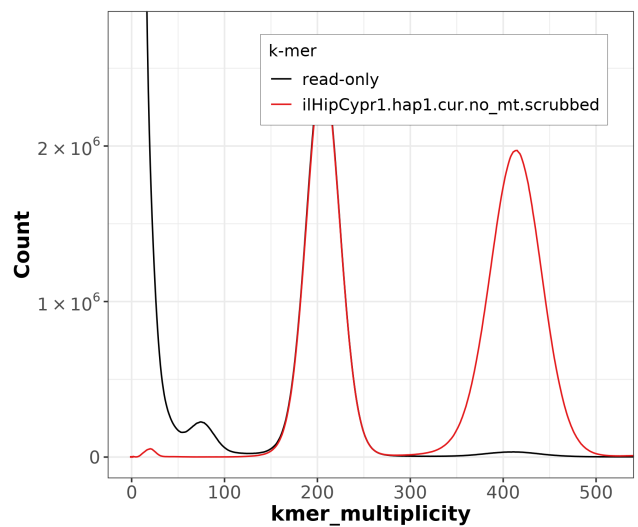


hap1 [\[LINK\]](#)

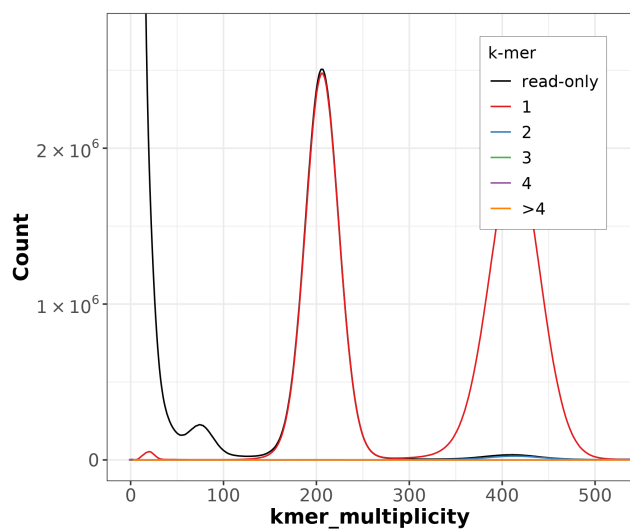
# K-mer spectra of curated assembly



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

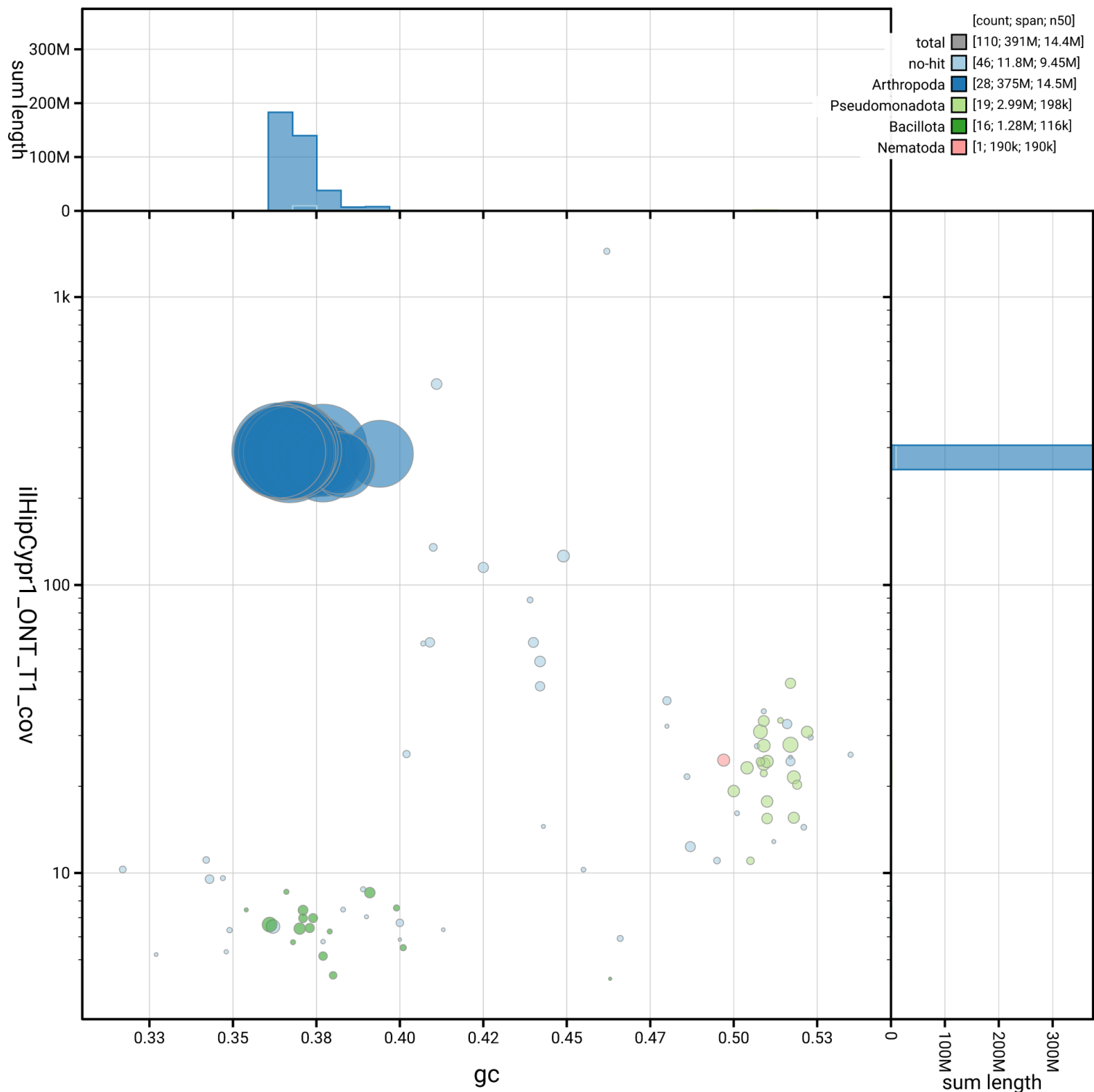


Distribution of k-mer counts coloured by their presence in reads/assemblies



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

# Post-curation contamination screening



**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	ONT	Illumina	Omni-C
Coverage	300x	198x	137x

## Assembly pipeline

- **Filtlong**
  - |\_ *ver*: 0.2.1
  - |\_ *key param*: --min\_length 1000 --min\_mean\_q 80
- **Hifiasm**
  - |\_ *ver*: 0.24.0
  - |\_ *key param*: --telo-m AACCT"
- **YaHS**
  - |\_ *ver*: 1.2a
  - |\_ *key param*: --no-contig-ec
- **CLAWS pipeline**
  - |\_ *ver*: 2.3.0
  - |\_ *key param*: NA
- **Blobtoolkit Nextflow pipeline (latest)**
  - |\_ *ver*: 0.6
  - |\_ *key param*: NA
- **FOAM pipeline**
  - |\_ *ver*: 0.5
  - |\_ *key param*: NA
- **mitos**
  - |\_ *ver*: 2.1.3
  - |\_ *key param*: NA

## Curation pipeline

- **GRIT\_Rapid**
  - |\_ *ver*: 2.0
  - |\_ *key param*: NA
- **PretextViewAI**
  - |\_ *ver*: 1.0.4
  - |\_ *key param*: NA

Submitter: Tyler Alioto

Affiliation: CNAG

Date and time: 2025-10-24 17:55:48 CEST