

# ERGA Assembly Report

v24.04.03\_beta

Tags: ERGA-BGE

TxID	3139654
ToLID	<b>qqHisDubi2</b>
Species	Histopona dubia
Class	Arachnida
Order	Araneae

Genome Traits	Expected	Observed
Haploid size (bp)	4,357,519,084	4,819,036,337
Haploid Number	22 (source: ancestor)	23
Ploidy	2 (source: ancestor)	2
Sample Sex	M	M

## EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 7.8.Q66

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

- . Observed Haploid Number is different from Expected

### Curator notes

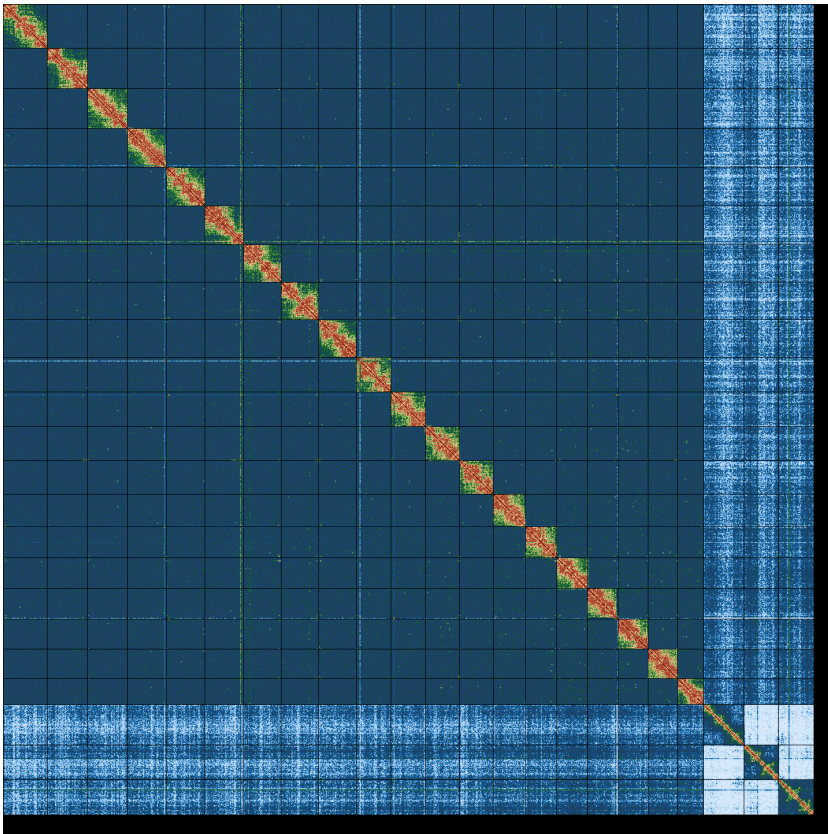
. Interventions/Gb: 116  
. Contamination notes: "Contamination report for assembly labelled hap1; Total length of scaffolds removed: 263,028 (0.0 %); Scaffolds removed: 8 (0.5 %); Largest scaffold removed: (57,504); FCS-GX contaminant species (number of scaffolds; total length of scaffolds):Pectobacterium phage Astalicious, prokaryotic viruses (1; 57,504);Mitochondrion (7; 205,524)"  
. Other observations: "Assembly was Hi-C phased; Chromosomes X1, X2 and X3 were identified by copy number in the diploid assembly and named to match the chromosome naming of GCA\_965122215. Chromosome X3 is a part of ancestral X1 that underwent a fusion/fission event. A haplotypic inversion was observed in the region on chromosome 15 (0-27.2 Mbp). A large number of centromeric repeats is in the unlocalised sequences."

# Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	4,126,461,816	4,819,036,337
GC %	32.73	32.77
Gaps/Gbp	259.54	281.8
Total gap bp	107,100	151,900
Scaffolds	1,647	1,562
Scaffold N50	195,020,524	203,116,814
Scaffold L50	10	11
Scaffold L90	19	21
Contigs	2,718	2,920
Contig N50	13,528,000	12,783,366
Contig L50	79	98
Contig L90	374	450
QV	65.8	66.8
Kmer compl.	99.52	99.52
BUSCO sing.	90.5%	94.1%
BUSCO dupl.	2.9%	4.3%
BUSCO frag.	0.6%	0.5%
BUSCO miss.	6.0%	1.1%

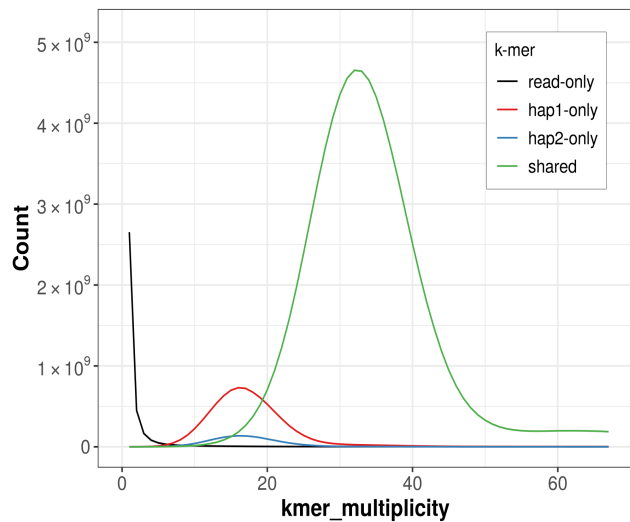
BUSCO 6.0.0 Lineage: arthropoda\_odb10 (genomes:90, BUSCOs:1013)

# HiC contact map of curated assembly

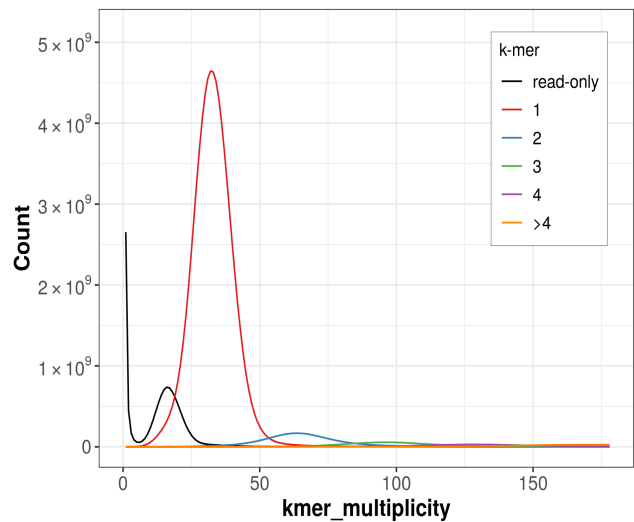


pri [\[LINK\]](#)

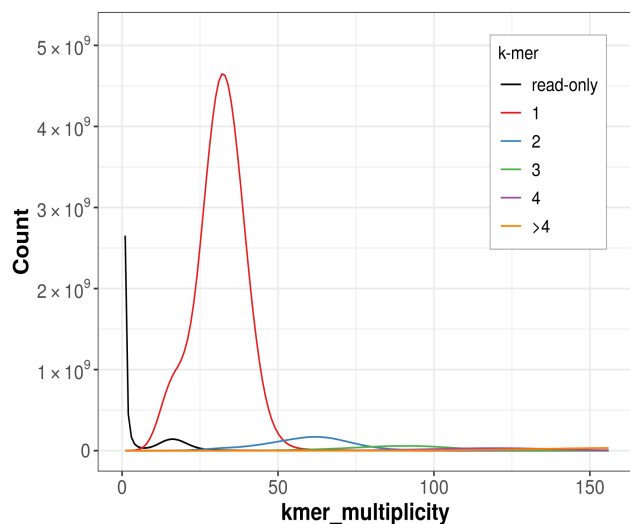
# K-mer spectra of curated assembly



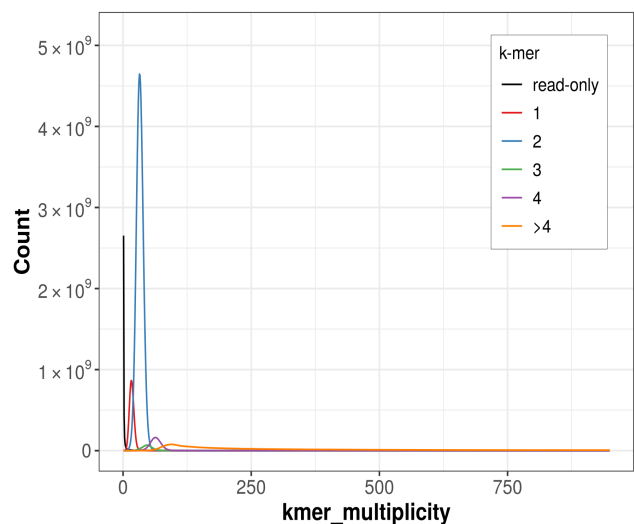
Distribution of k-mer counts coloured by their presence in reads/assemblies



Distribution of k-mer counts per copy numbers found in **hap2** (hap1.)

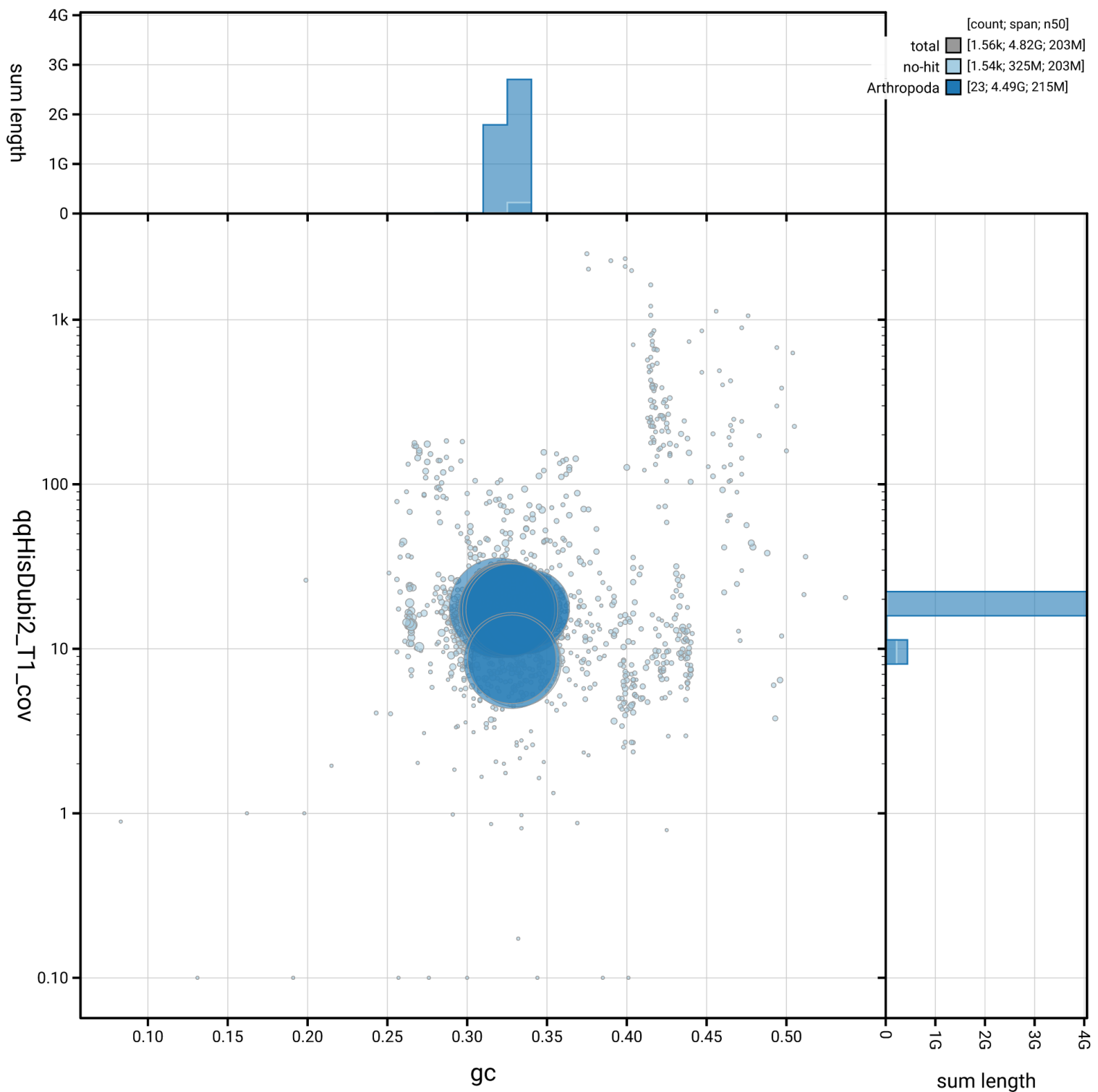


Distribution of k-mer counts per copy numbers found in **hap1** (hap1.)



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

# 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	32x	279x

# Assembly pipeline

- **hifiasm**
  - |\_ *ver*: 0.25.0-r726
  - |\_ *key param*: --h1/--h2
- **yahs**
  - |\_ *ver*: 1.2.2
  - |\_ *key param*: NA

# Curation pipeline

- **hifiasm**
  - |\_ *ver*: 0.25.0-r726
  - |\_ *key param*: --h1/--h2
- **yahs**
  - |\_ *ver*: 1.2.2
  - |\_ *key param*: NA
- **TreeVal**
  - |\_ *ver*: 1.2.2
  - |\_ *key param*: NA

Submitter: Danil Zilov

Affiliation: WSI

Date and time: 2025-12-16 12:01:37 CET