

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	3073148
ToLID	<b>icBatTris</b>
Species	Bathyscidius tristiculus
Class	Insecta
Order	Coleoptera

Genome Traits	Expected	Observed
Haploid size (bp)	434,092,953	488,115,944
Haploid Number	11 (source: ancestor)	13
Ploidy	2 (source: ancestor)	2
Sample Sex	NA	X0

## EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.7.Q66

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 pri
- . Assembly length loss > 3% for pri
- . More than 1000 gaps/Gbp for pri
- . Not 90% of assembly in chromosomes for pri

### Curator notes

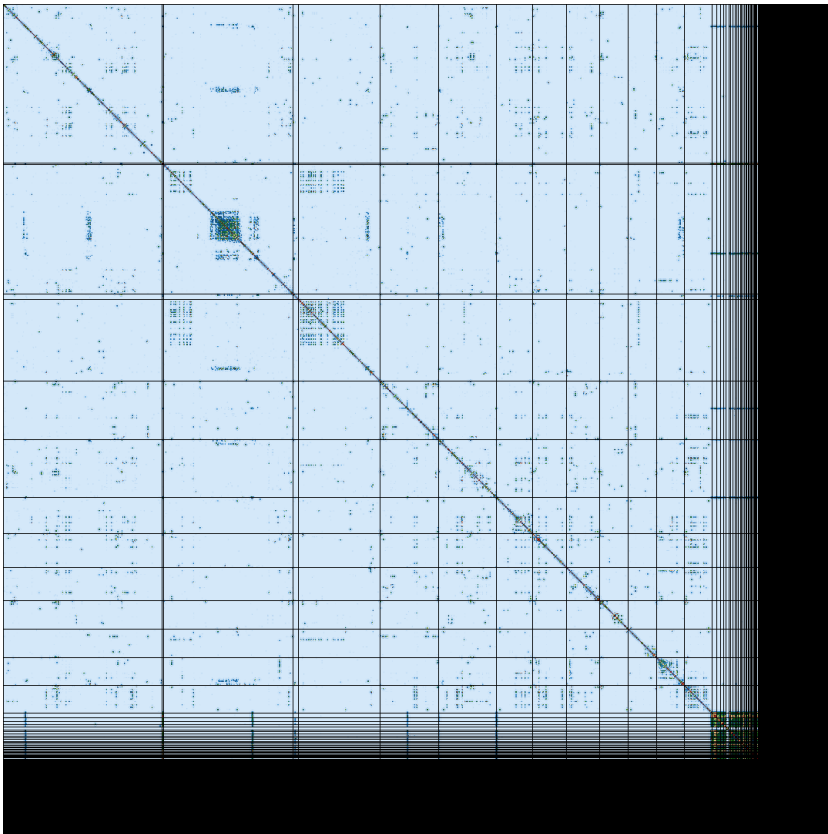
- . Interventions/Gb: None
- . Contamination notes: "No contamination was detected."
- . Other observations: "This genome assembly was produced by assembling PacBio HiFi data from two different individuals with Hifiiasm, purging with purge\_dups and then scaffolding with Yahs. Curation made 27 cuts in contigs, 46 breaks at gaps and 47 joins. 44 haplotypic sequences, composing a total of 28,686,280 bp were identified and removed during curation. Mitogenome was successfully assembled with FOAM."

# Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	516,802,241	488,115,944
GC %	32.89	32.92
Gaps/Gbp	1,079.72	1,143.17
Total gap bp	111,600	111,600
Scaffolds	331	313
Scaffold N50	33,037,159	34,462,195
Scaffold L50	6	4
Scaffold L90	45	35
Contigs	889	871
Contig N50	1,128,672	1,117,445
Contig L50	121	122
Contig L90	468	462
QV	66.3313	66.2884
Kmer compl.	79.3841	78.5194
BUSCO sing.	92.5%	94.5%
BUSCO dupl.	3.3%	0.9%
BUSCO frag.	0.7%	0.7%
BUSCO miss.	3.5%	3.9%

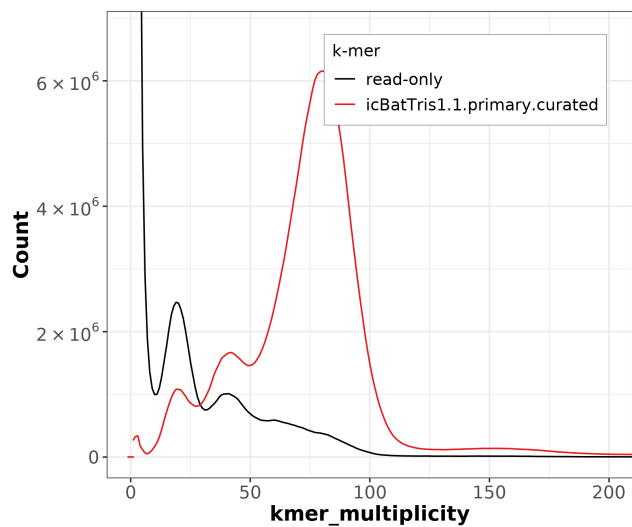
BUSCO: 6.0.0 (euk\_genome\_min, miniprot) / Lineage: insecta\_odb12 (genomes:79, BUSCOs:3114)

# HiC contact map of curated assembly

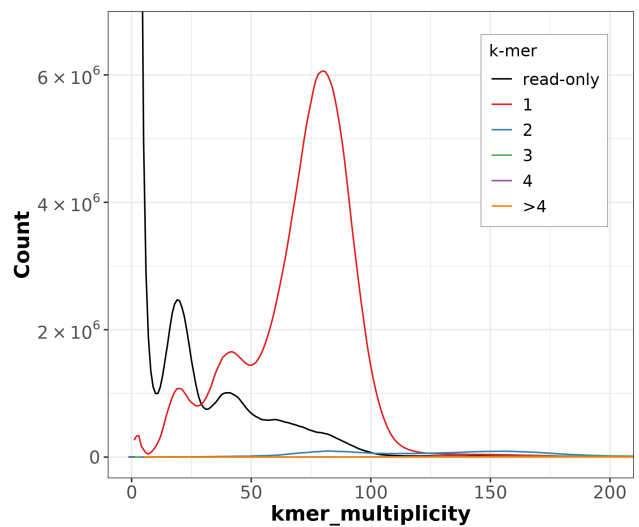


pri [\[LINK\]](#)

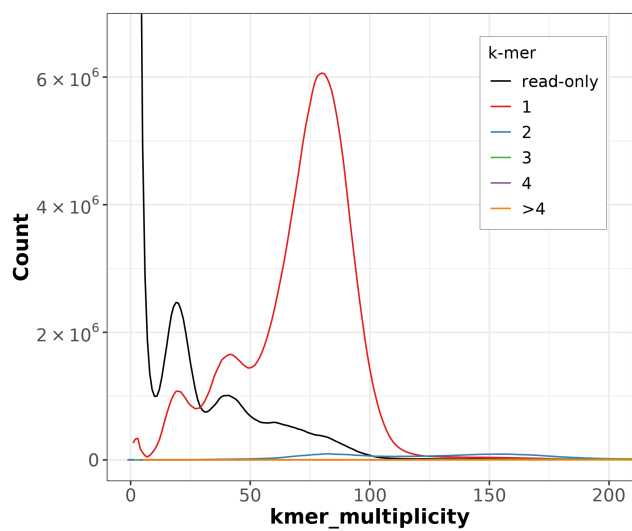
# K-mer spectra of curated assembly



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

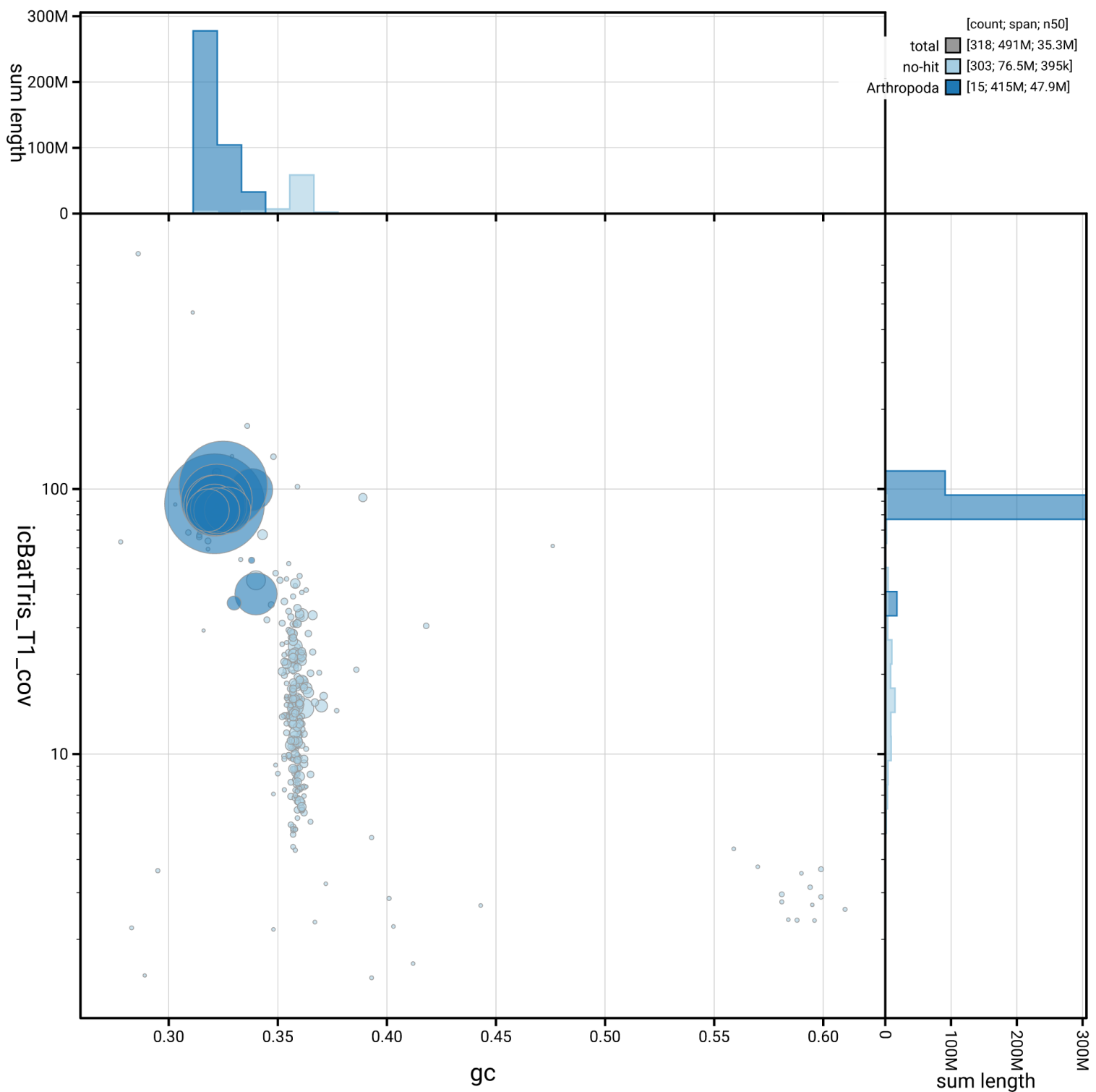


Distribution of k-mer counts per copy numbers found in asm



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	PacBio HiFi	Arima Hi-C
Coverage	74	171

# Assembly pipeline

- **CLAWS**
  - |\_ *ver*: v3.0
  - |\_ *key param*: NA
- **Hifiasm**
  - |\_ *ver*: 0.24.0
  - |\_ *key param*: --telo-m ACACAC
  - |\_ *key param*: --n-perturb 30000
  - |\_ *key param*: --f-perturb 0.3
  - |\_ *key param*: --min-hist-cnt 5
  - |\_ *key param*: -N 150
- **Purge\_dups**
  - |\_ *ver*: 1.2.6
  - |\_ *key param*: NA
- **Yahs**
  - |\_ *ver*: 1.2a
  - |\_ *key param*: -mq 10
  - |\_ *key param*: --no-contig-ec
- **FOAM**
  - |\_ *ver*: 0.5
  - |\_ *key param*: NA

# Curation pipeline

- **Blobtoolkit Nextflow pipeline**
  - |\_ *ver*: 0.6
  - |\_ *key param*: NA
- **PretextViewAI**
  - |\_ *ver*: 1.0.5
  - |\_ *key param*: NA

Submitter: Jessica Gomez-Garrido  
Affiliation: CNAG

Date and time: 2026-02-10 15:29:18 CET