

# ERGA Assembly Report

v24.04.03\_beta

Tags: ERGA-BGE

TxID	124803
ToLID	<b>icTarCanal</b>
Species	Tarphius canariensis
Class	Insecta
Order	Coleoptera

Genome Traits	Expected	Observed
Haploid size (bp)	310,233,475	310,723,302
Haploid Number	9 (source: ancestor)	0
Ploidy	2 (source: ancestor)	2
Sample Sex	U	U

## EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 7.7.Q66

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

- . Observed Haploid Number is different from Expected
- . Not 90% of assembly in chromosomes for pri

### Curator notes

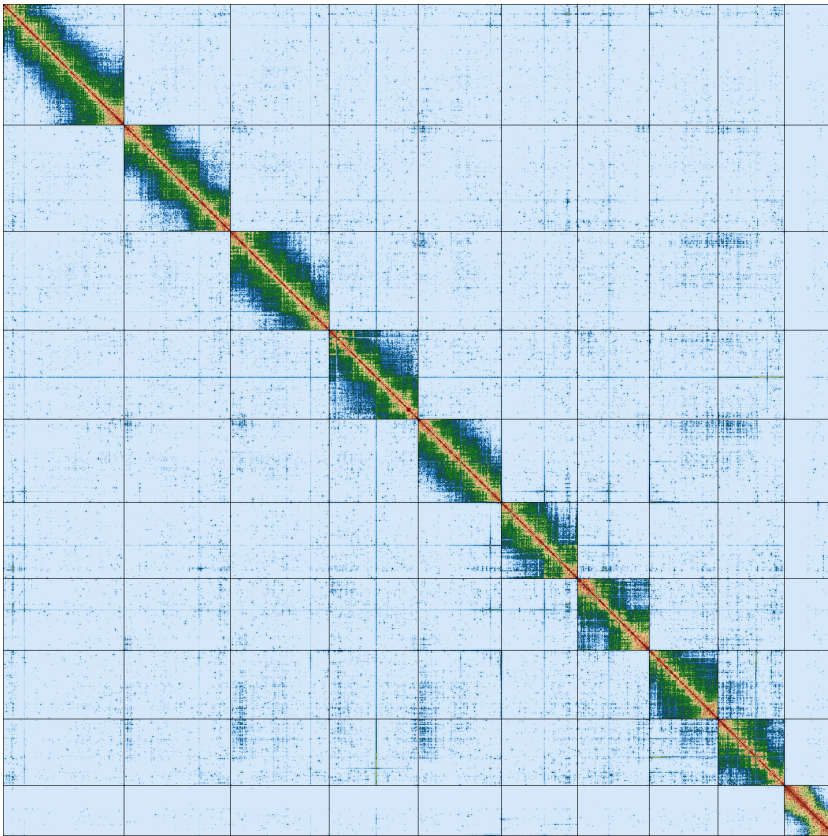
. Interventions/Gb: 0  
. Contamination notes: "Total length of scaffolds removed: 2,074,311 (0.7 %);Scaffolds removed: 59 (72.8 %);Largest scaffold removed: (160,449);FCS-GX contaminant species (number of scaffolds; total length of scaffolds):Erwinia typographi, g-proteobacteria (15; 677,499);Allobranchiibius huperziae, high GC Gram+ (9; 252,647);Erwinia billingiae, g-proteobacteria (5; 190,020);Arachidicoccus soli, CFB group bacteria (2; 185,866);Rhodanobacter denitrificans, g-proteobacteria (1; 86,367);Silvibacterium dinghuense, bacteria (1; 57,248);Filimonas sp., CFB group bacteria (1; 55,454);Acidicapsa acidisoli, bacteria (1; 43,005);Acidobacterium capsulatum, bacteria (1; 42,976);Granulicella mallensis, bacteria (1; 39,590);Terriglobus roseus, bacteria (1; 36,970);Nakamurella endophytica, high GC Gram+ (1; 36,876);Niabella beijingensis, CFB group bacteria (1; 36,705);Erwinia psidii, g-proteobacteria (1; 16,466);Conexibacter sp., actinobacteria (1; 16,280);Metallococcus carri, high GC Gram+ (1; 2,000)"  
. Other observations: "Hi-C from a different individual (icTarCana2), so assembly is not Hi-C phased"

## Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	314,484,547	310,723,302
GC %	34.86	34.73
Gaps/Gbp	92.21	119.08
Total gap bp	5,800	7,400
Scaffolds	81	12
Scaffold N50	31,089,936	31,130,618
Scaffold L50	5	5
Scaffold L90	9	9
Contigs	110	49
Contig N50	14,492,433	14,492,433
Contig L50	8	8
Contig L90	24	25
QV	63.8	66.4
Kmer compl.	99.55	99.20
BUSCO sing.	99.3%	99.4%
BUSCO dupl.	0.4%	0.4%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	0.2%	0.1%

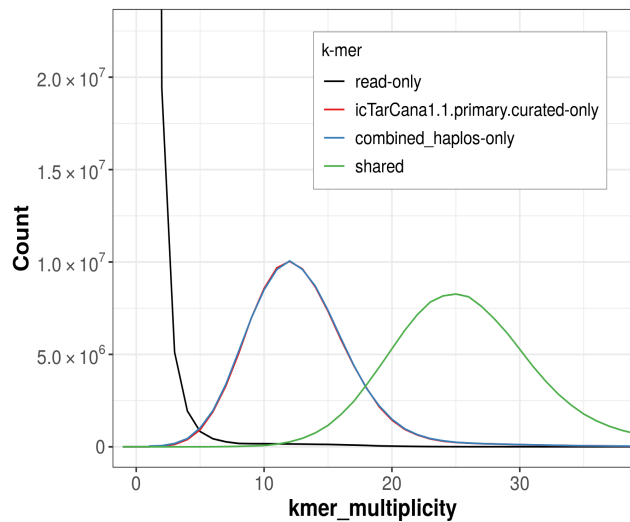
Warning: BUSCO versions or lineage datasets are not the same across results

# HiC contact map of curated assembly

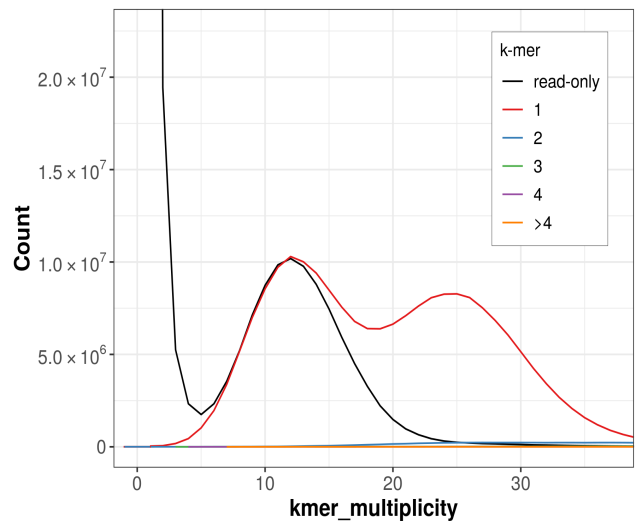


pri [\[LINK\]](#)

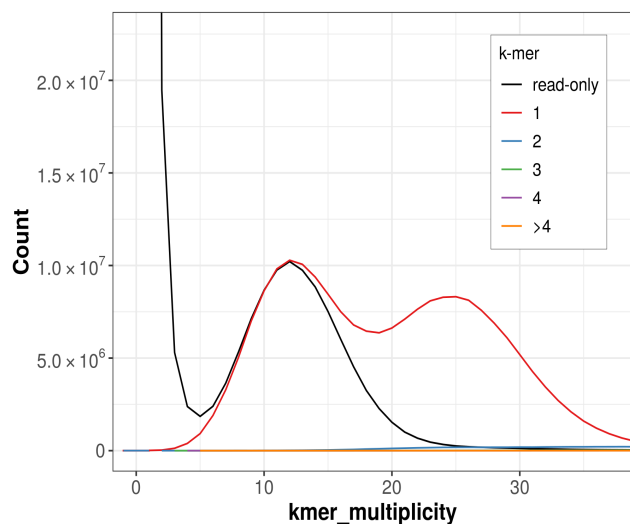
# K-mer spectra of curated assembly



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

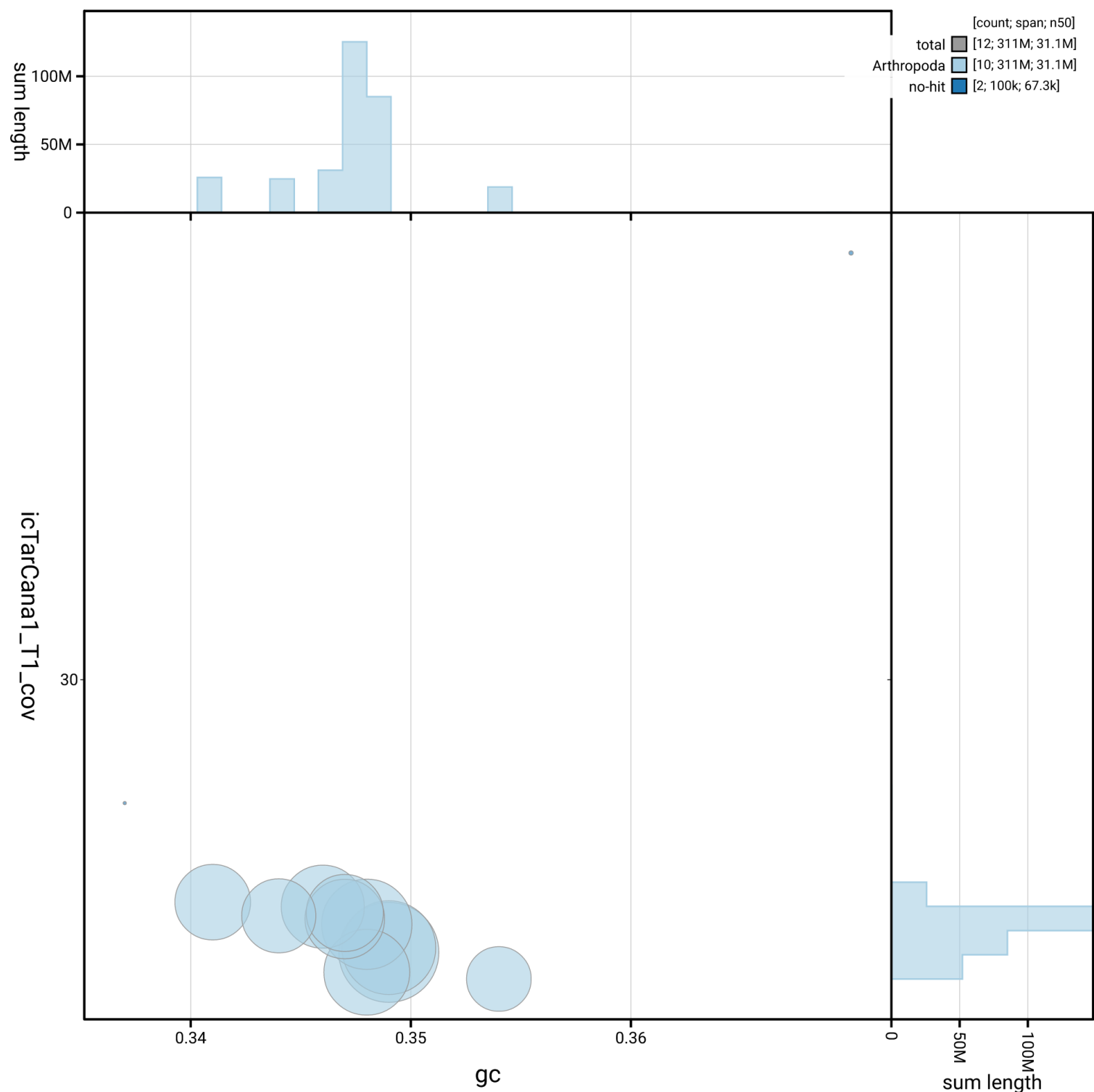


Distribution of k-mer counts per copy numbers found in asm



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 v2
Coverage	25x	291x

# Assembly pipeline

- **hifiasm**
  - |\_ *ver*: 0.19.8-r603
  - |\_ *key param*: --primary
- **purge\_dups**
  - |\_ *ver*: 1.2.5
  - |\_ *key param*: -e
- **yahs**
  - |\_ *ver*: 1.2a.2
  - |\_ *key param*: NA

# Curation pipeline

- **hifiasm**
  - |\_ *ver*: 0.19.8-r603
  - |\_ *key param*: --primary
- **purge\_dups**
  - |\_ *ver*: 1.2.5
  - |\_ *key param*: -e
- **yahs**
  - |\_ *ver*: 1.2a.2
  - |\_ *key param*: NA
- **TreeVal**
  - |\_ *ver*: 1.1
  - |\_ *key param*: NA

Submitter: Joanna Collins

Affiliation: WSI

Date and time: 2024-09-18 14:16:44 CEST