

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	173051
ToLID	qcDicCarn1
Species	Dicellogophilus carniolensis
Class	Chilopoda
Order	Geophilomorpha

Genome Traits	Expected	Observed
Haploid size (bp)	324,282,458	316,843,155
Haploid Number	8 (source: ancestor)	17
Ploidy	2 (source: ancestor)	2
Sample Sex	NA	NA

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q68

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

- . Observed Haploid Number is different from Expected
- . Kmer completeness value is less than 90 for collapsed

Curator notes

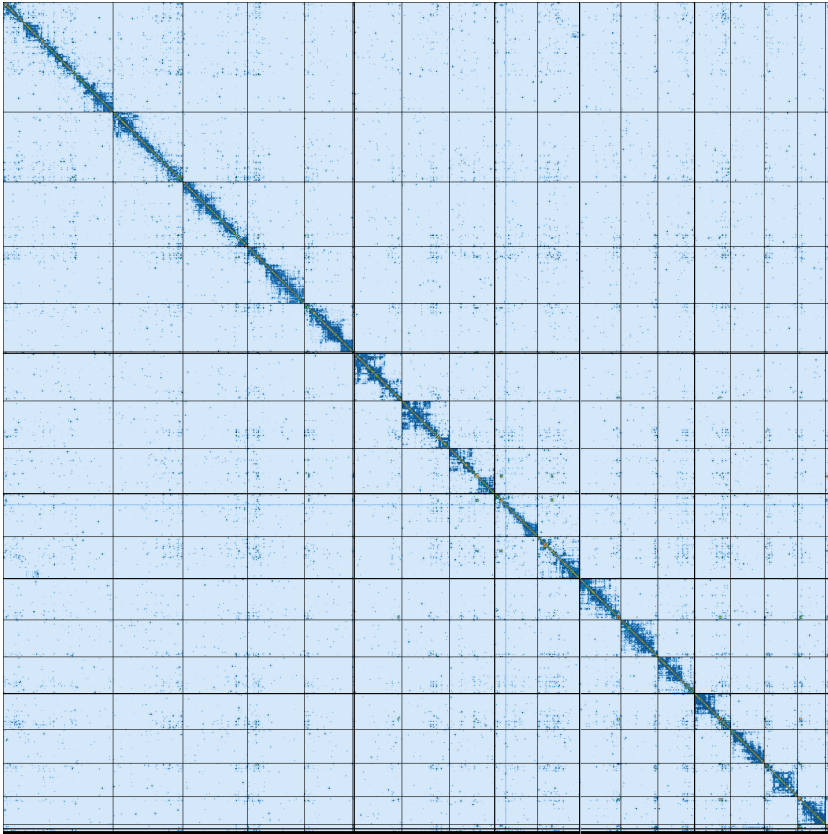
. Interventions/Gb:
. Contamination notes: "FCS-GX detected some bacterial fragments in the hifi.asm p_ctgs matching Wolbachia and indicative of a complete genome but without circularized contigs. We collected the fragments and used these as reference for read mapping with minimap2, then extracted mapped reads with samtools (-F 2052) and re-assembled them with hifi.asm with default settings. This process yielded a circular contig of 1.3Mb. Allele typing using PubMLST (https://pubmlst.org/bigdb?db=pubmlst_rmlst_seqdef_kiosk) revealed a close match to Candidatus Mesenot endosymbiont of Agriotes lineatus, which is a proposed new genus in the Anaplasmataceae family."
. Other observations: "The assembly was produced with hifi.asm and the *p_ctg.fa was selected. The assembly was decontaminated with FCS-GX, and haplotigs purged with purge_dups. Scaffolding was performed with YaHS. Mitochondria was assembled with Oatk. HiFi and Hi-C data come from distinct individuals."

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	320,158,379	316,843,155
GC %	37.32	37.34
Gaps/Gbp	165.54	173.59
Total gap bp	10,600	11,000
Scaffolds	47	38
Scaffold N50	17,904,240	17,904,240
Scaffold L50	6	7
Scaffold L90	15	15
Contigs	100	93
Contig N50	6,899,197	7,036,659
Contig L50	15	14
Contig L90	47	45
QV	67.8974	68.1445
Kmer compl.	89.278	88.641
BUSCO sing.	92.7%	92.7%
BUSCO dupl.	4.6%	4.6%
BUSCO frag.	0.4%	0.4%
BUSCO miss.	2.3%	2.3%

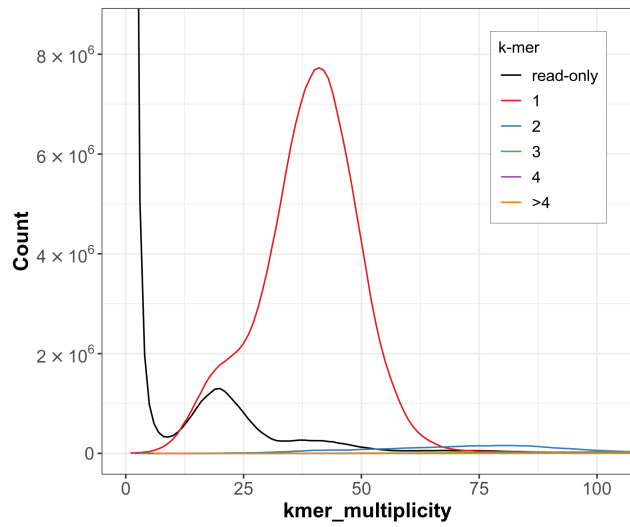
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: arthropoda_odb10 (genomes:90, BUSCOs:1013)

HiC contact map of curated assembly

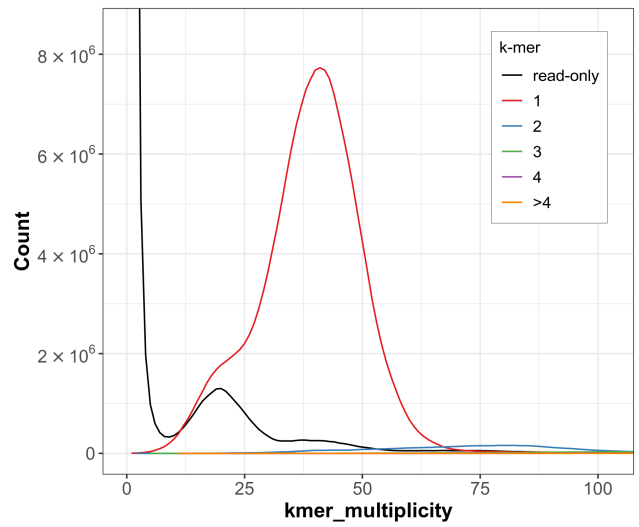


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

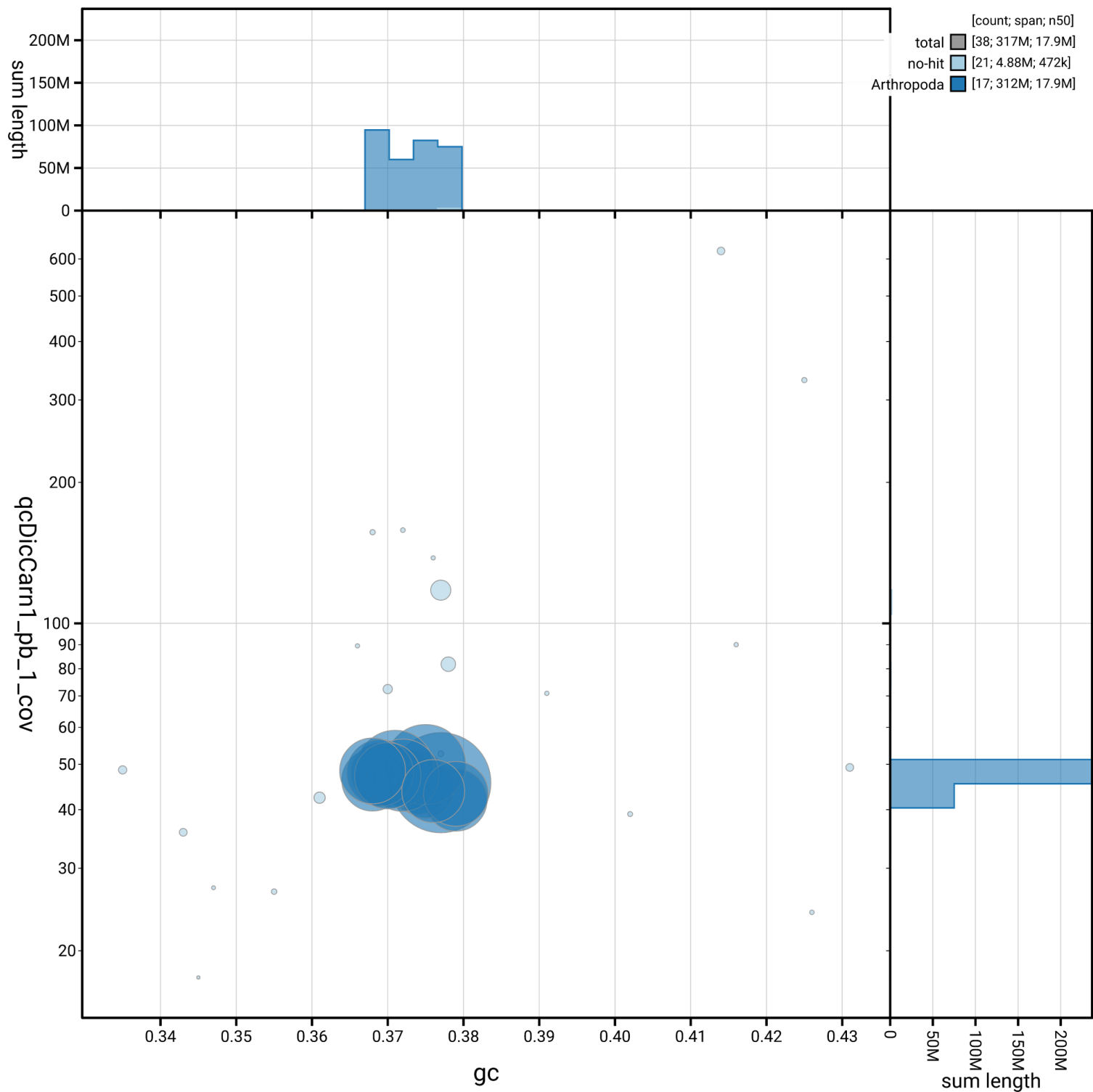


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



collapsed. 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 HiC v2
Coverage	42x	80x

Assembly pipeline

```
- hifiasm-hic
  |_ ver: 0.25.0-r726
  |_ key param: NA
- purge_dups
  |_ ver: 1.2.5
  |_ key param: NA
- YaHS
  |_ ver: 1.2a.2
  |_ key param: NA
- NBIS/Earth-Biogenome-Project-pilot
  |_ ver: 99c8937
  |_ key param: NA
- sanger-tol/blobtoolkit
  |_ ver: 0.8.0
  |_ key param: NA
- Oatk
  |_ ver: 1.0
  |_ key param: -m insecta.fam
  |_ key param: -c 100
```

Curation pipeline

```
- sanger-tol/curationpretext
  |_ ver: 1.4.2
  |_ key param: NA
- sanger-tol/agp-tpf-utils
  |_ ver: 1.2.3
  |_ key param: NA
```

Submitter: Guilherme B. Dias

Affiliation: SciLifeLab

Date and time: 2025-11-10 13:12:14 CET