ERGA Assembly Report
v24.10.15

Tags: ERGA-BGE

TxID	1849190
ToLID	**icAntGang**
Species	Anthroherpon ganglbaueri
Class	Insecta
Order	Coleoptera

Genome Traits	Expected	Observed
Haploid size (bp)	289,632,189	327,830,440
Haploid Number	11 (source: ancestor)	11
Ploidy	2 (source: ancestor)	2
Sample Sex	NA	NA

EBP metrics summary and curation notes

Obtained EBP quality metric for hap1: 6.7.Q60

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

. More than 1000 gaps/Gbp for hap1
. Not 90% of assembly in chromosomes for hap1

Curator notes

. Interventions/Gb: None
. Contamination notes: "Protobacteria contaminants were found in 2102 scaffolds and removed with the pipeline NCBI FCS GX v Galaxy Version 0.5.5+galaxy1. The first blobtolkit resulted in another 49 scaffolds flagged as contaminant that were removed since there were contact in the HiC maps. In the current graph, scaffolds flagged as contaminants by BlobToolKit were confirmed with FCS output table and with contact in the map as non-contaminants. "
. Other observations: "Chromosomes arms were reconstructed with alignment with a species from the same family (Leptodirus hochenwartii, GCA_947310635.1)."

Quality metrics table

| Metrics	Pre-curation hap1	Curated hap1
Total bp	332,096,506	327,830,440
GC %	32.02	31.9
Gaps/Gbp	1,165.32	1,189.64
Total gap bp	77,400	78,000
Scaffolds	1,498	1,392
Scaffold N50	22,404,328	22,404,528
Scaffold L50	6	6
Scaffold L90	609	558
Contigs	1,885	1,782
Contig N50	1,050,880	1,064,634
Contig L50	89	88
Contig L90	952	905
QV	60.1544	60.392
Kmer compl.	91.1711	91.1154
BUSCO sing.	96.6%	96.6%
BUSCO dupl.	1.0%	1.0%
BUSCO frag.	0.7%	0.7%
BUSCO miss.	1.7%	1.7%

BUSCO: 5.8.0 (euk_genome_met, metaeuk) / Lineage: insecta_odb10 (genomes:75, BUSCOs:1367)
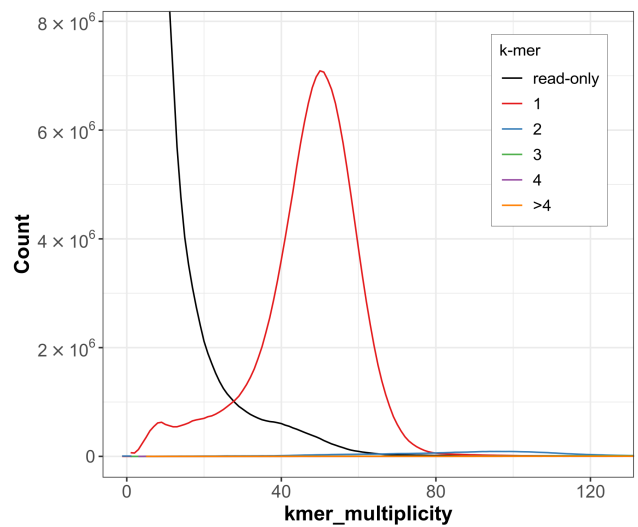
HiC contact map of curated assembly
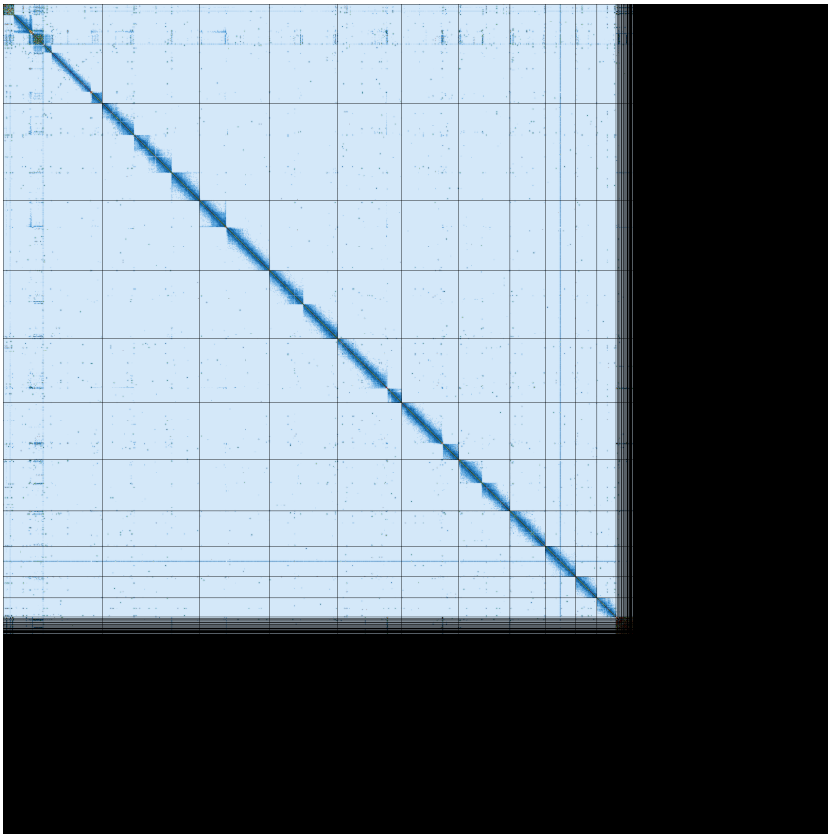
hap1 [LINK]

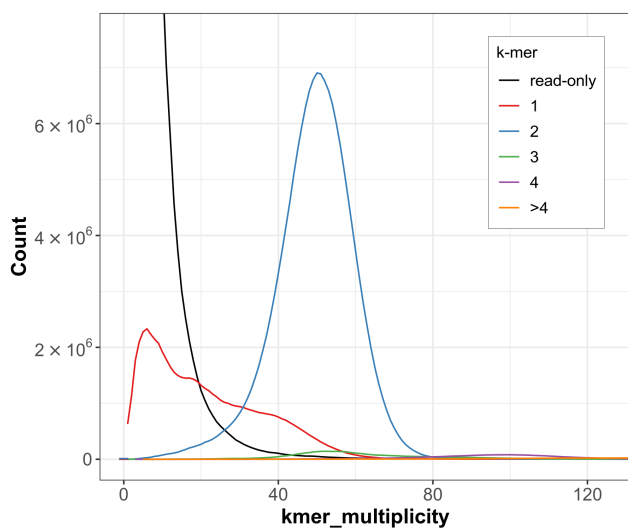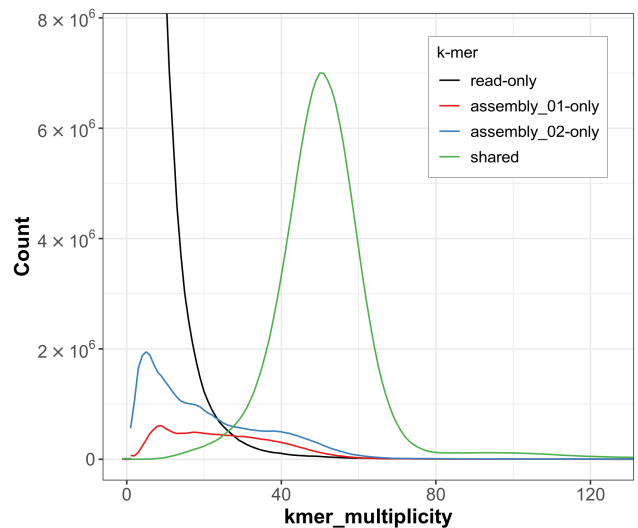# K-mer spectra of curated assembly
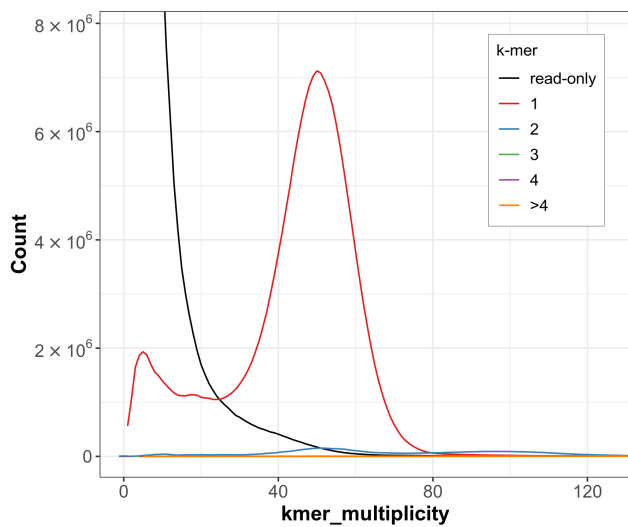
Distribution of k-mer counts per copy
numbers found in **assembly_02** (hapl.)

Distribution of k-mer counts per copy
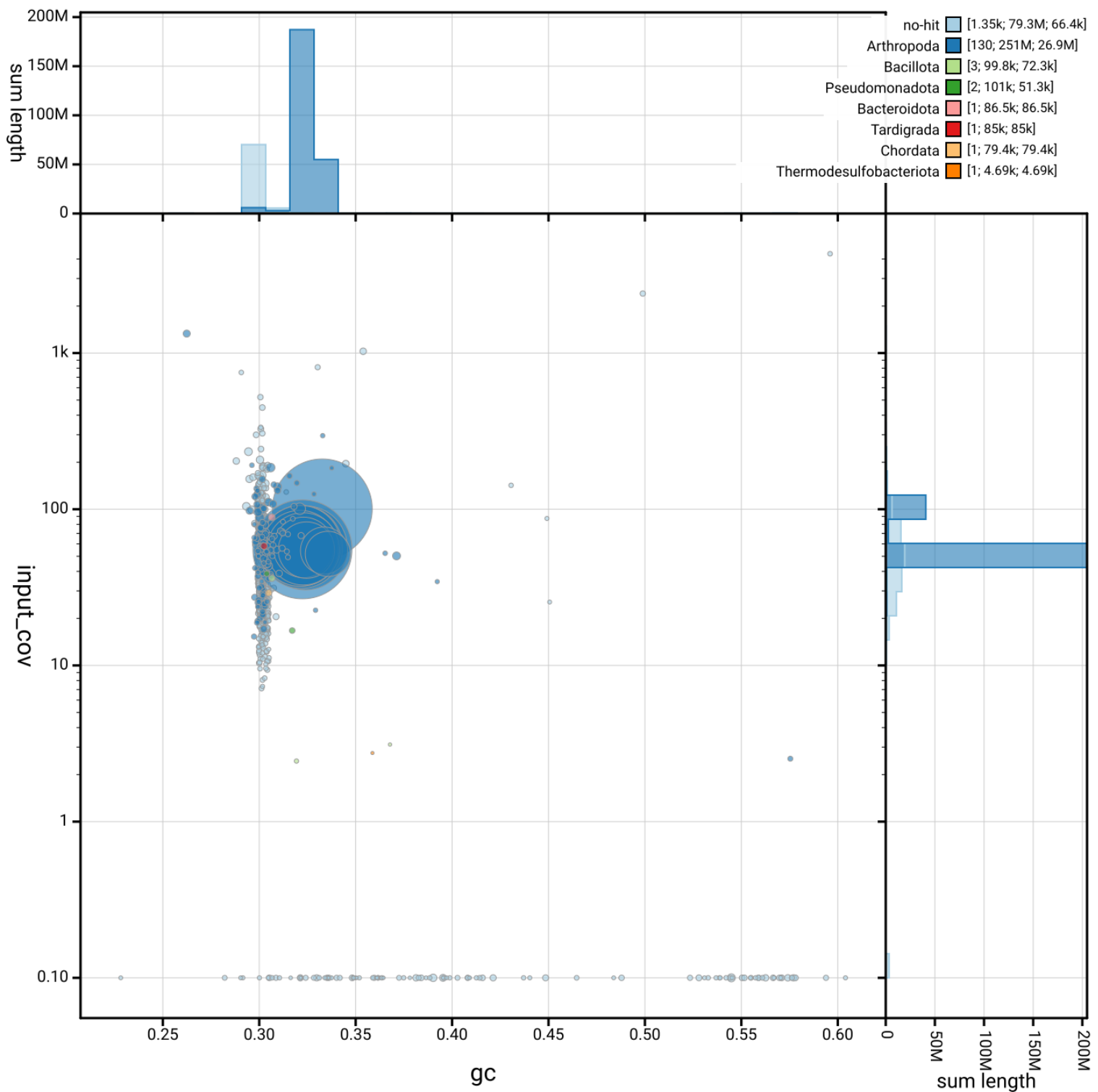numbers found in **assembly_01** (hapl.)

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

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

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	HiFi	HiC
Coverage	34.80x	302x

Assembly pipeline

- **Hifiasm**
    |_ *ver:* Galaxy Version 0.24.0+galaxy0
    |_ *key param:* VGP Galaxy workflow VGP3 v 0.2.8
- **purge_dups**
    |_ *ver:* Galaxy Version 1.2.6+galaxy0
    |_ *key param:* VGP Galaxy workflow VGP6 v 0.5
- **YaHS**
    |_ *ver:* Galaxy Version 1.2a.2+galaxy2
    |_ *key param:* VGP Galaxy workflow VGP8 v 1.4

Curation pipeline

- **PretextMap**
    |_ *ver:* Galaxy Version 0.1.9+galaxy1
    |_ *key param:* VGP workflow VGP PretextMap generation v2
- **PretextViewAI**
    |_ *ver:* 1.0.3
    |_ *key param:* NA

Submitter: Marcella Sozzoni
Affiliation: UNIFI

Date and time: 2025-07-28 10:34:02 CEST