

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	3065877
ToLID	<b>icHerPorc</b>
Species	Herpes porcellus
Class	Insecta
Order	Coleoptera

Genome Traits	Expected	Observed
Haploid size (bp)	860,627,567	775,828,464
Haploid Number	11 (source: ancestor)	13
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.7.Q49

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 pri
- . BUSCO single copy value is less than 90% for pri

### Curator notes

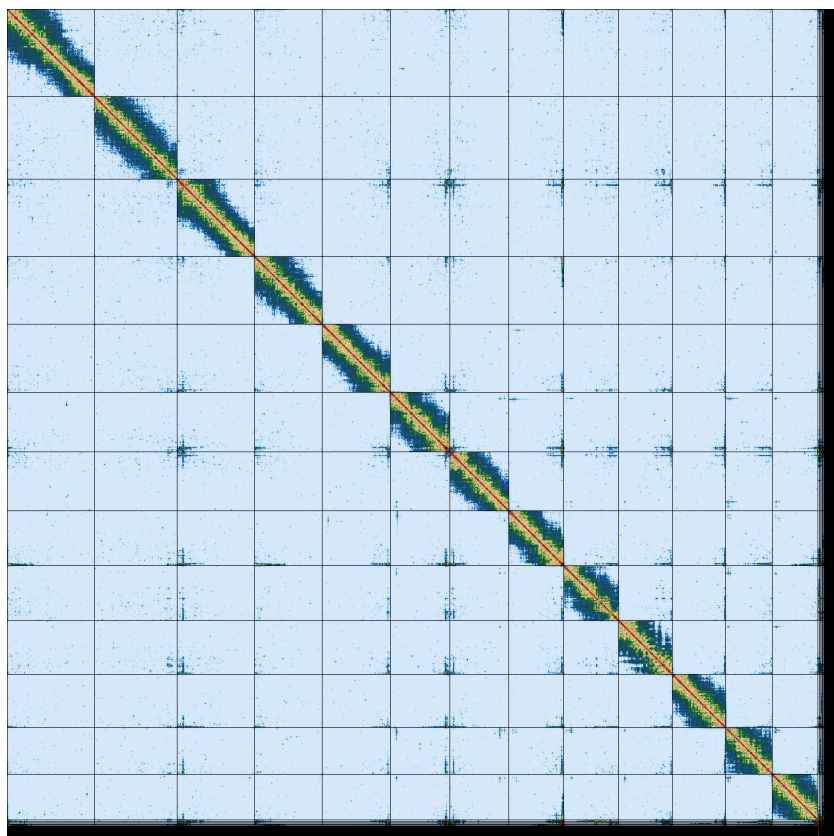
- . Interventions/Gb: 26
- . Contamination notes: "No contaminants were detected."
- . Other observations: "The assembly of species *Herpes porcellus* (icHerPorc) is based on 174X long read ONT data and 48X Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA) via the Biodiversity Genomics Europe project (BGE). The assembly process included the following steps: ONT reads shorter than 3 kb were filtered out, thus the remaining reads -for a total of ~152 Gb- were assembled using Necat with polishing performed using NextPolish. Contigs were subsequently filtered with purge\_dups to remove haplotigs and obtain a more accurate, non-redundant assembly. Scaffolding was performed by aligning Hi-C reads to the purged contigs using the Omni-C mapping pipeline, followed by YaHS for scaffold construction. Contamination was checked using BlobTools. The scaffolds were processed through sanger\_tol/curationpretext pipeline to generate the contact map, which was manually improved using PretextView. The presence of a mitochondrial genome contig/scaffold was not assessed. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size."

# Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	776,183,321	775,828,464
GC %	33.36	33.36
Gaps/Gbp	182.95	194.63
Total gap bp	14,200	17,100
Scaffolds	159	150
Scaffold N50	54,968,996	55,530,766
Scaffold L50	6	6
Scaffold L90	12	12
Contigs	301	301
Contig N50	8,297,486	8,297,486
Contig L50	29	29
Contig L90	98	98
QV	49.1387	49.1367
Kmer compl.	86.6404	86.6333
BUSCO sing.	89.7%	89.8%
BUSCO dupl.	0.6%	0.5%
BUSCO frag.	1.0%	1.0%
BUSCO miss.	8.7%	8.7%

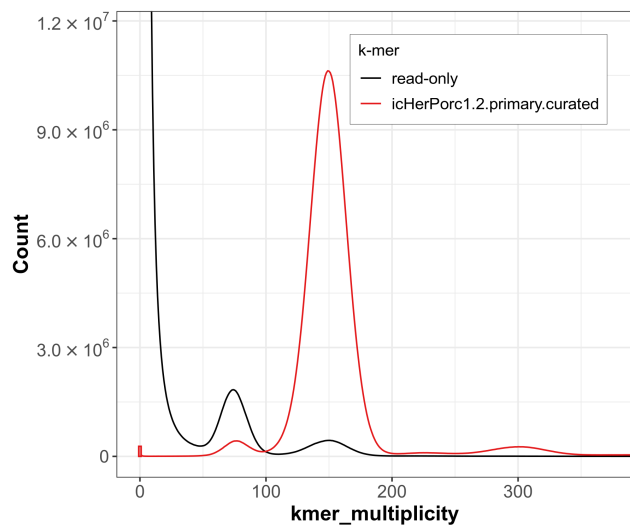
BUSCO: 5.8.2 (euk\_genome\_aug, augustus) / Lineage: arthropoda\_odb10 (genomes:90, BUSCOs:1013)

# HiC contact map of curated assembly

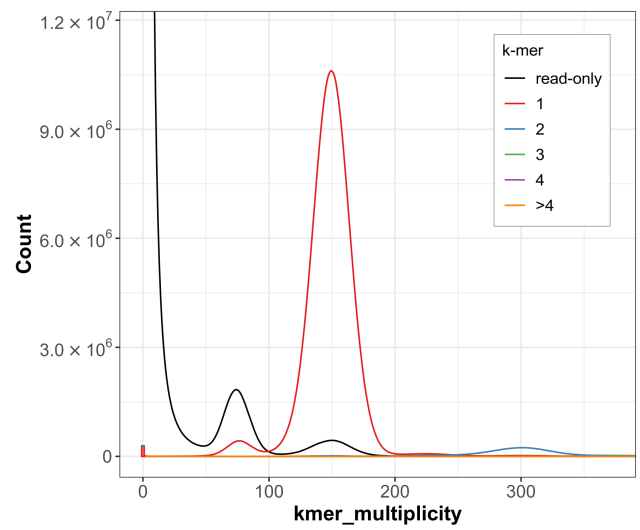


pri [\[LINK\]](#)

# K-mer spectra of curated assembly



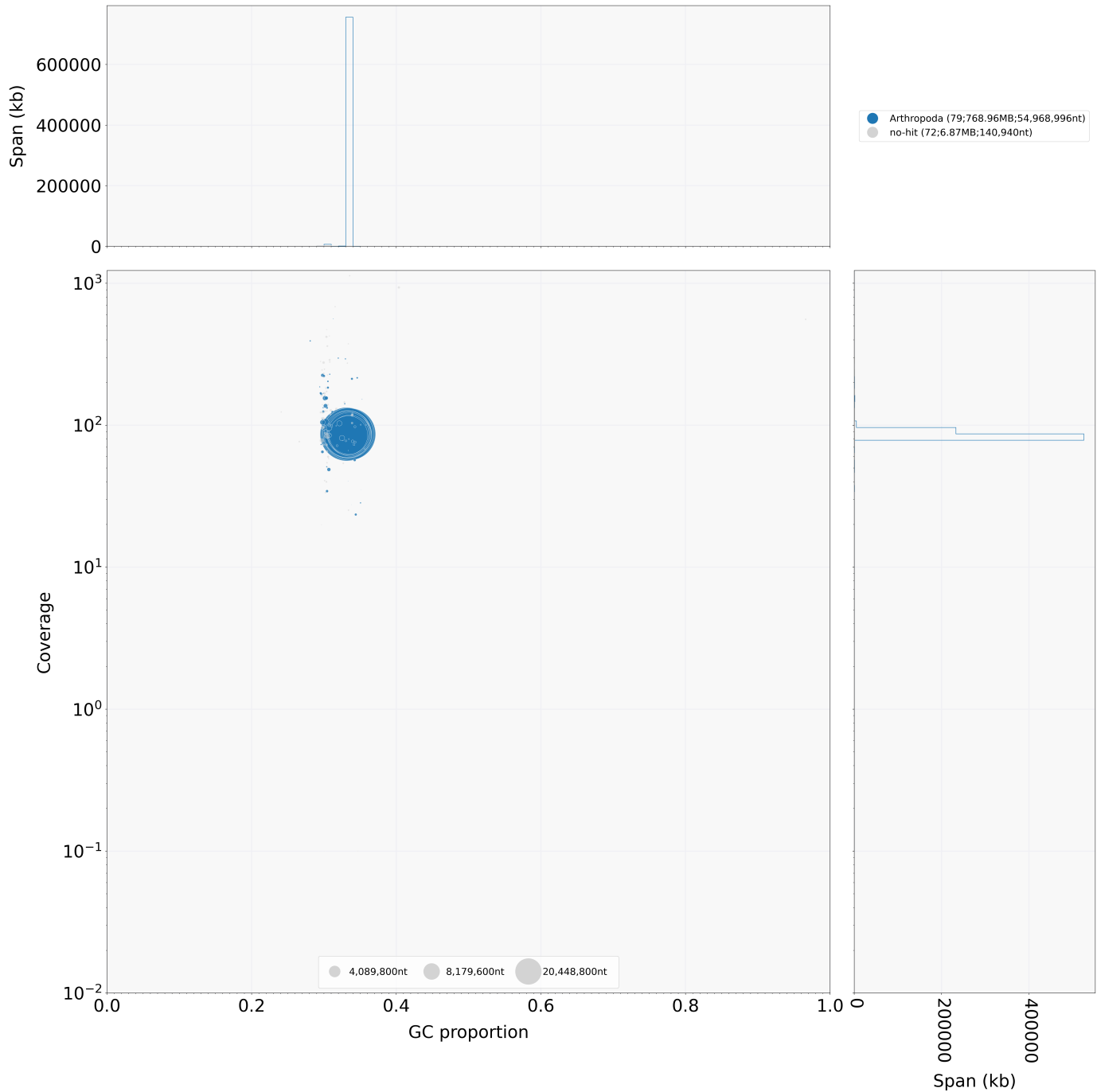
Distribution of k-mer counts coloured by their presence in reads/assemblies



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

# Post-curation contamination screening

filename.blobDB.json.bestsum.phylum.p8.span.100.blobplot.bam0



**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	ONT	Arima HiC
Coverage	174x	48x

## Assembly pipeline

- **Necat**
  - |\_ *ver*: 0.0.1
  - |\_ *key param*: NA
- **purge\_dups**
  - |\_ *ver*: 1.2.5
  - |\_ *key param*: NA
- **YaHS**
  - |\_ *ver*: 1.1
  - |\_ *key param*: NA

## Curation pipeline

- **sanger-tol/curationpretext**
  - |\_ *ver*: 1.4.1
  - |\_ *key param*: NA
- **PretextView**
  - |\_ *ver*: 1.0.3
  - |\_ *key param*: NA

Submitter: Ilenia Urso

Affiliation: UNIBA

Date and time: 2025-11-19 16:45:21 CET