

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

TxID	537450
ToLID	ihGraItal9
Species	Graphosoma italicum
Class	Insecta
Order	Hemiptera

Genome Traits	Expected	Observed
Haploid size (bp)	1,291,444,399	1,491,838,610
Haploid Number	7 (source: direct)	8
Ploidy	2 (source: ancestor)	2
Sample Sex	U	M

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.8.Q59

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

Curator notes

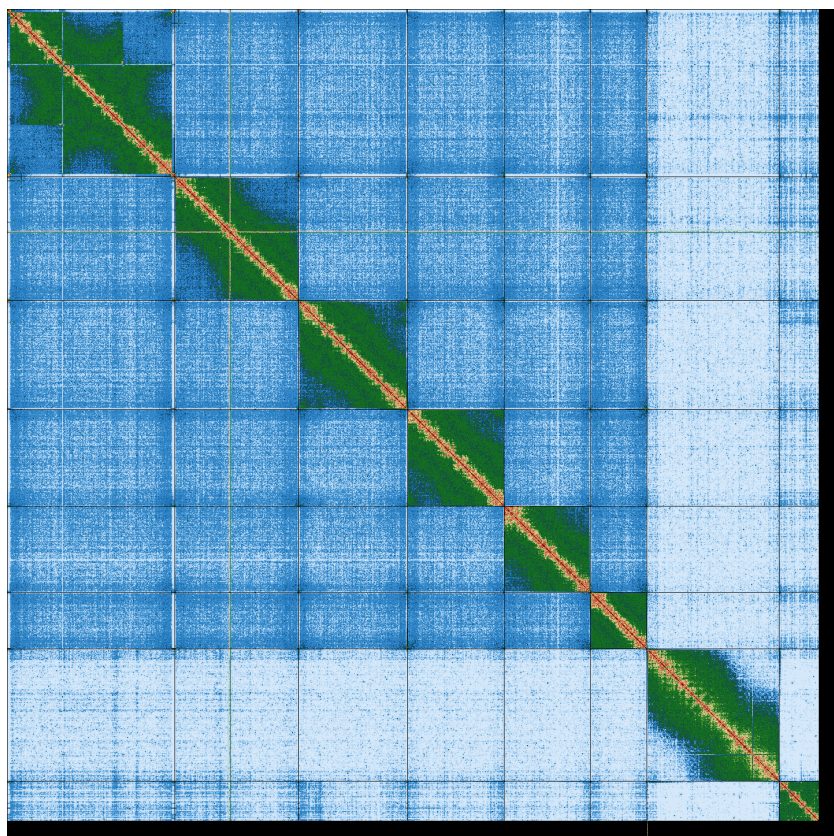
. Interventions/Gb: 92
. Contamination notes: "Contamination report for assembly labelled hap1; Total length of scaffolds removed: 5,649,647 (0.4 %); Scaffolds removed: 57 (14.1 %); Largest scaffold removed: (3,629,078); FCS-GX contaminant species (number of scaffolds; total length of scaffolds): Pantoea brenneri, g-proteobacteria (4; 4,294,550); Erwinia tracheiphila, g-proteobacteria (2; 56,671); Salmonella enterica subsp. enterica serovar Kottbus, g-proteobacteria (1; 50,898); Mitochondrion (50; 1,247,528)"
. Other observations: "Assembly was Hi-C phased; The sex chromosomes X and Y were identified by read coverage and copy number in the diploid assembly. A haplotypic inversion was observed in the region on chromosome 1 (98.4-207.5 Mbp)."

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	1,501,839,401	1,491,838,610
GC %	32.73	32.65
Gaps/Gbp	220.4	248.69
Total gap bp	33,100	42,000
Scaffolds	404	297
Scaffold N50	191,845,183	221,852,459
Scaffold L50	4	3
Scaffold L90	7	7
Contigs	735	668
Contig N50	6,190,740	6,234,386
Contig L50	72	71
Contig L90	239	238
QV	59.6	59.6
Kmer compl.	99.74	99.43
BUSCO sing.	98.3%	
BUSCO dupl.	1.4%	
BUSCO frag.	0.1%	
BUSCO miss.	0.2%	

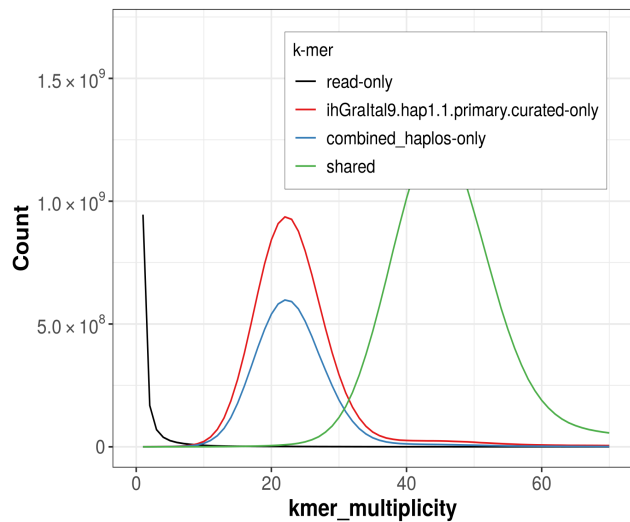
BUSCO 5.3.0 Lineage: insecta_odb10 (genomes:75, BUSCOs:1367)

HiC contact map of curated assembly

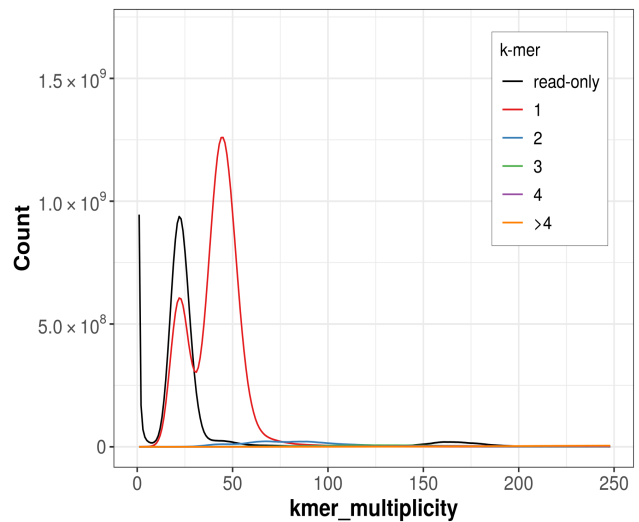


pri [\[LINK\]](#)

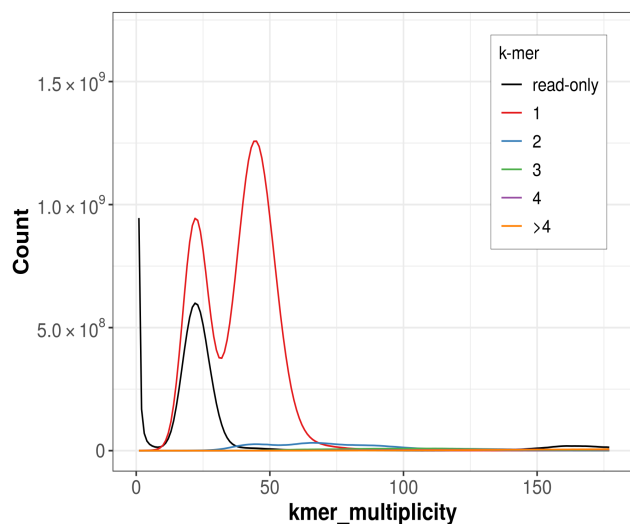
K-mer spectra of curated assembly



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

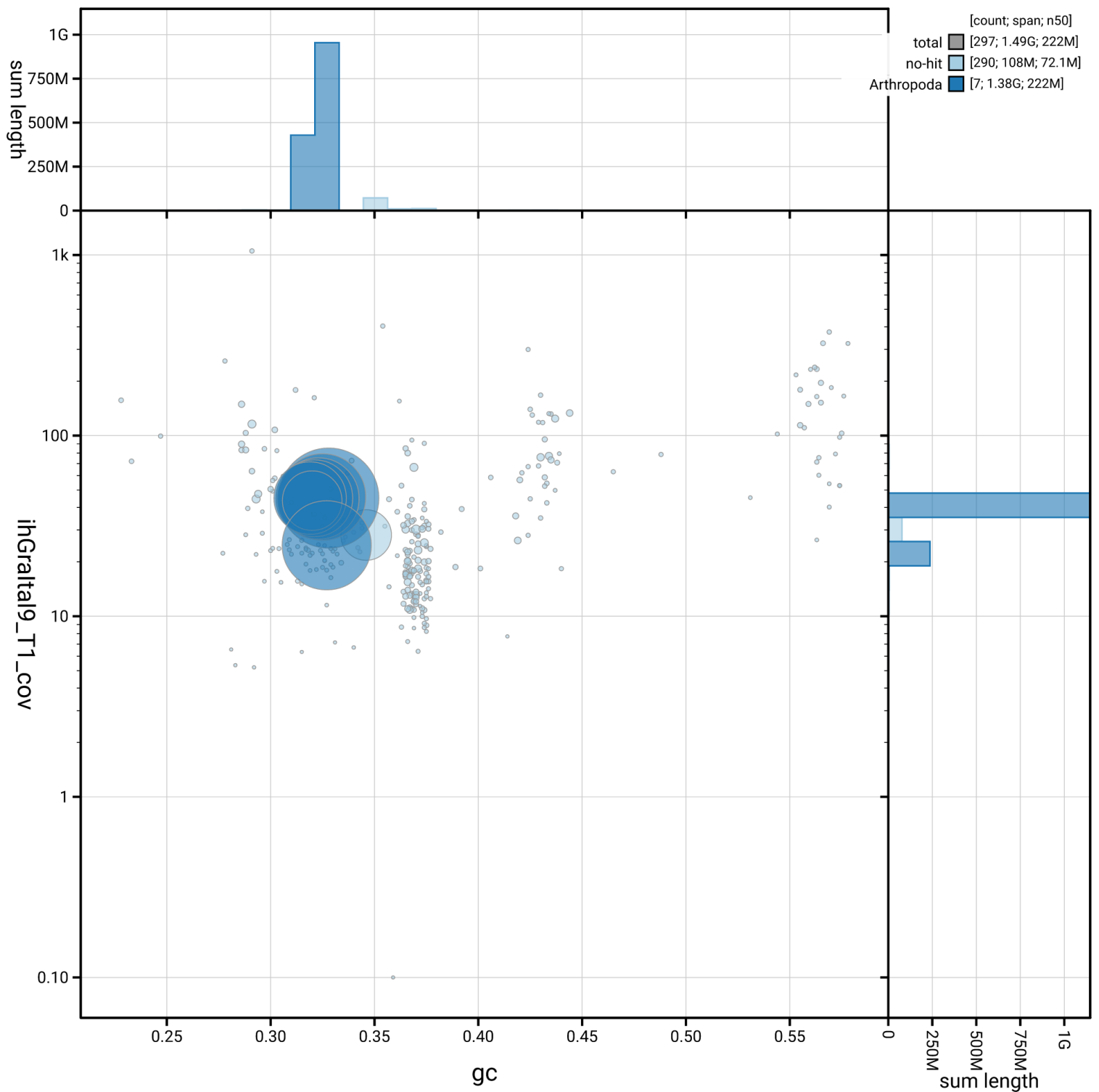


Distribution of k-mer counts per copy numbers found in asm



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	44x	396x

Assembly pipeline

```
- hifiasm-hic
  |_ ver: 0.19.8-r603
  |_ key param: --primary
- yahs
  |_ ver: 1.2.2
  |_ key param: NA
```

Curation pipeline

```
- hifiasm-hic
  |_ ver: 0.19.8-r603
  |_ key param: --primary
- yahs
  |_ ver: 1.2.2
  |_ key param: NA
- TreeVal
  |_ ver: 1.2.1
  |_ key param: NA
```

Submitter: Michael Paulini

Affiliation: WSI

Date and time: 2025-04-23 13:52:42 CEST