

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	1208769
ToLID	ipEmbTyrr
Species	<i>Embia tyrrhenica</i>
Class	Insecta
Order	Embioptera

Genome Traits	Expected	Observed
Haploid size (bp)	2,837,334,911	2,563,847,327
Haploid Number	11 (source: direct)	11
Ploidy	2 (source: ancestor)	2
Sample Sex	NA	NA

EBP metrics summary and curation notes

Obtained EBP quality metric for hap1: 7.8.Q66

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

- . Kmer completeness value is less than 90 for hap1

Curator notes

- . Interventions/Gb: 66
- . Contamination notes: "Proteobacteria contaminants were found in two scaffolds and removed with the pipeline NCBI FCS GX v Galaxy Version 0.5.5+galaxy1. Scaffolds that are flagged as contaminants by BlobToolKit were confirmed with FCS output table and with contact in the map as non-contaminants. The big pseudomonata in the blob view is SUPER_11 which is confirmed to be part of our genome based on FCS output and HiC contact in the map."
- . Other observations: "None"

Quality metrics table

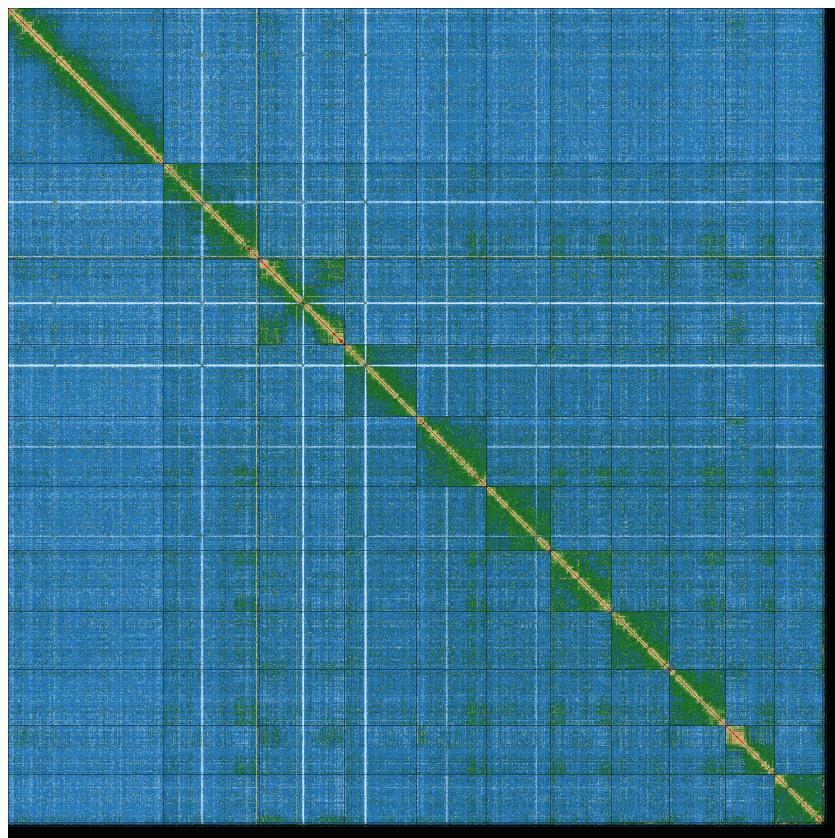
Metrics	Pre-curation hap1	Curated hap1
Total bp	2,577,151,225	2,563,847,327
GC %	33.88	33.88
Gaps/Gbp	190.13	207.11
Total gap bp	98,000	106,200
Scaffolds	708	674
Scaffold N50	589,349,641	216,130,847
Scaffold L50	2	5
Scaffold L90	7	10
Contigs	1,198	1,205
Contig N50	12,054,559	10,602,285
Contig L50	60	67
Contig L90	242	262
QV	66.3173	66.443
Kmer compl.	73.1091	72.8772
BUSCO sing.	92.0%	92.6%
BUSCO dupl.	3.6%	3.0%
BUSCO frag.	0.7%	1.0%
BUSCO miss.	3.7%	3.4%

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

BUSCO: 5.8.0 (euk_genome_met, metaeuk) / Lineage: insecta_odb10 (genomes:75, BUSCOs:1367)

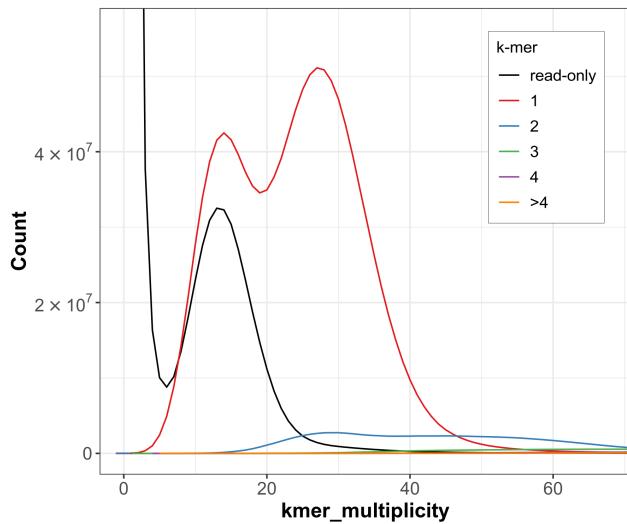
BUSCO: 5.8.0 (euk_genome_min, miniprot) / Lineage: insecta_odb10 (genomes:75, BUSCOs:1367)

HiC contact map of curated assembly

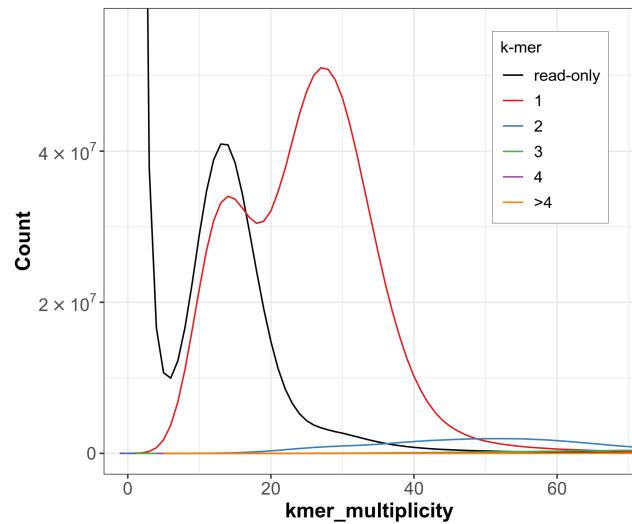


hap1 [\[LINK\]](#)

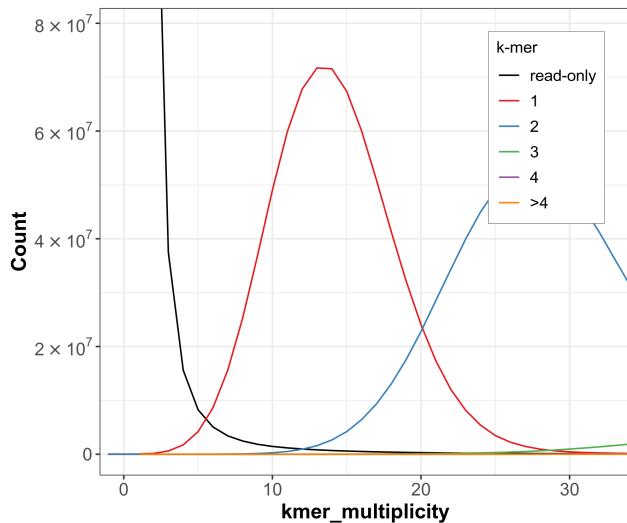
K-mer spectra of curated assembly



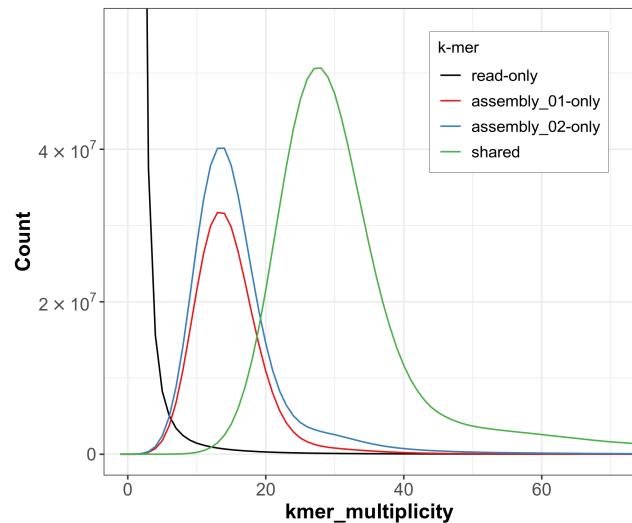
Distribution of k-mer counts per copy numbers found in **assembly_02** (hapl.)



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

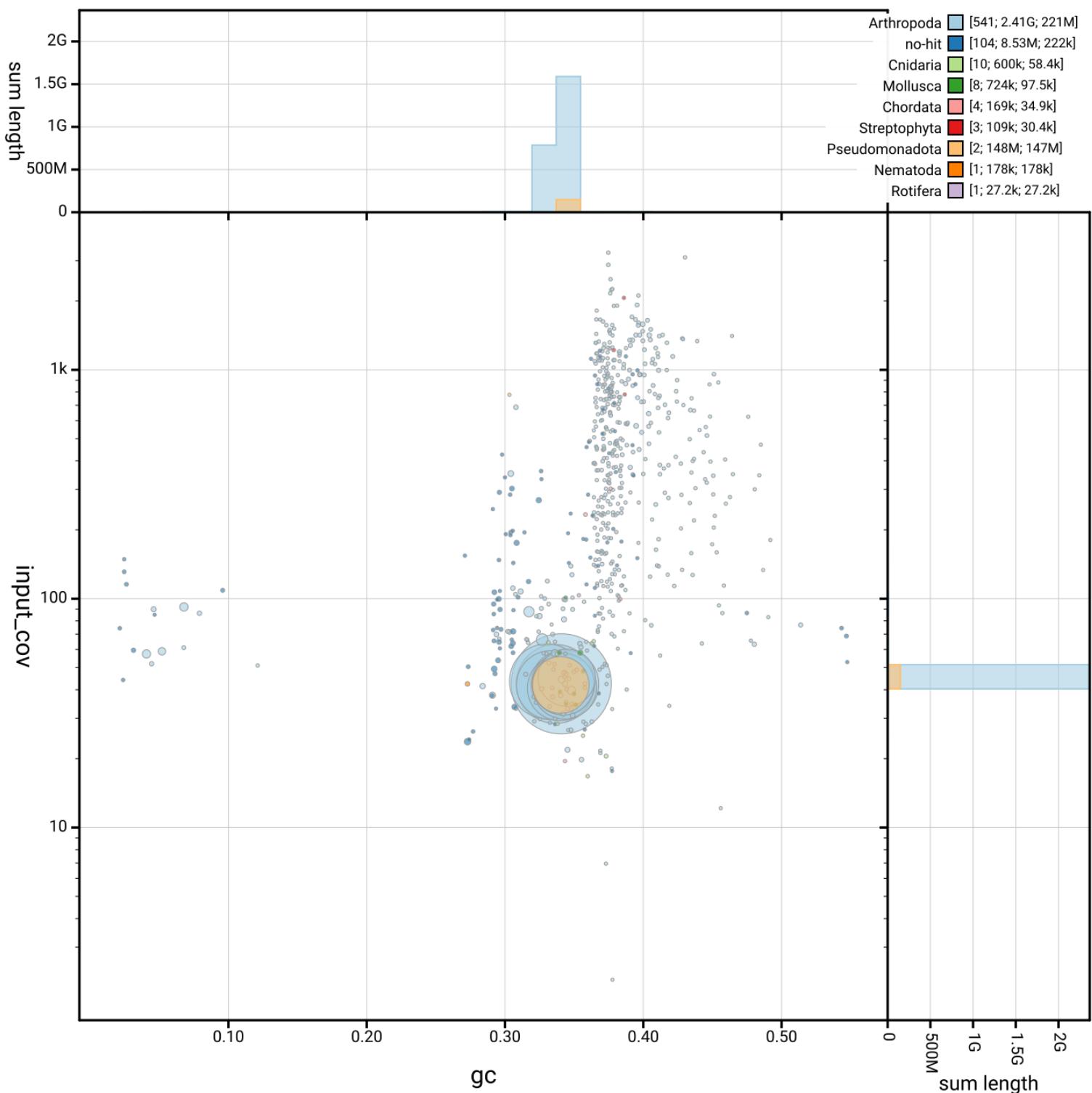


Distribution of k-mer counts per copy numbers found in **asm** (dipl.)



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	29.77x	254x	

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
- PretextView
|_ ver: 1.0.0
|_ key param: NA
```

Submitter: Marcella Sozzoni
Affiliation: UNIFI

Date and time: 2025-07-10 11:43:50 CEST