

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	2033273
ToLID	imEmpPeni1
Species	<i>Empusa pennicornis</i>
Class	Insecta
Order	Mantodea

Genome Traits	Expected	Observed
Haploid size (bp)	1,947,027,834	2,199,553,360
Haploid Number	14 (source: direct)	14
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 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 collapsed

Curator notes

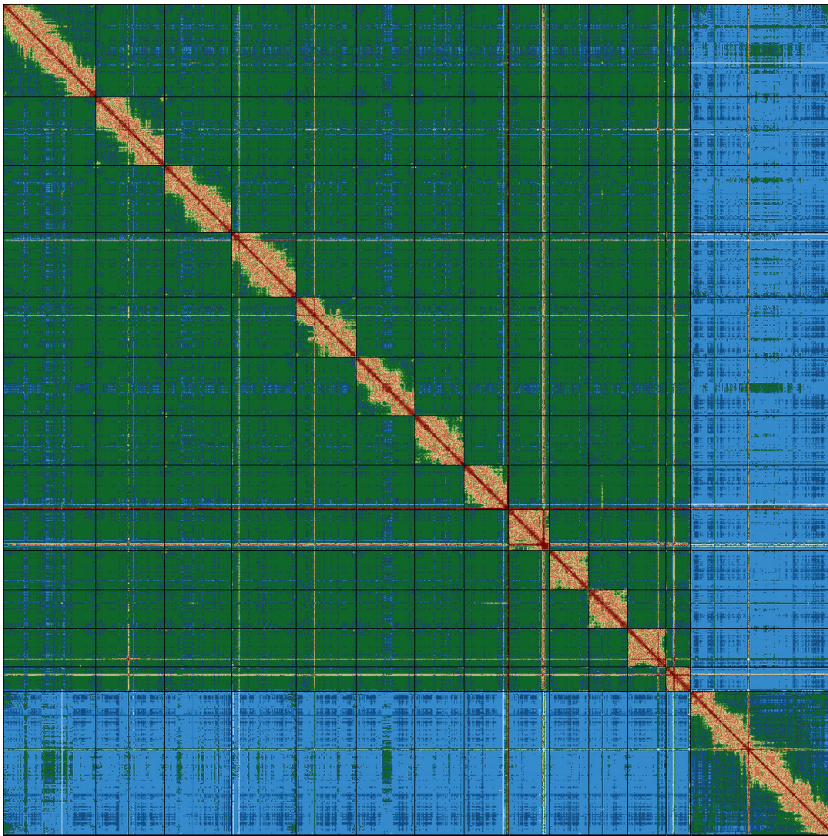
- . Interventions/Gb: 1
- . Contamination notes: ""
- . Other observations: "The assembly of *Empusa pennicornis* (imEmpPeni1.1) is based on 71X PacBio data Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA, <https://www.erga-biodiversity.eu/>) via the Biodiversity Genomics Europe project (BGE, <https://biodiversitygenomics.eu/>). The assembly process included the following steps: initial PacBio assembly generation with Hifiiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. No contigs were found to be contaminants (bacterial, archaeal, or viral). Additionally, 140 regions totaling 30.4 Mb (with the largest being 7.6 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	2,199,512,032	2,199,553,360
GC %	37.43	37.43
Gaps/Gbp	25.91	25.91
Total gap bp	6,100	6,300
Scaffolds	36	36
Scaffold N50	171,005,424	171,005,424
Scaffold L50	5	5
Scaffold L90	12	12
Contigs	92	93
Contig N50	87,323,851	87,323,851
Contig L50	10	10
Contig L90	36	36
QV	66.3878	66.3879
Kmer compl.	81.7549	81.7549
BUSCO sing.	94.3%	94.3%
BUSCO dupl.	1.0%	1.0%
BUSCO frag.	3.1%	3.1%
BUSCO miss.	1.7%	1.7%

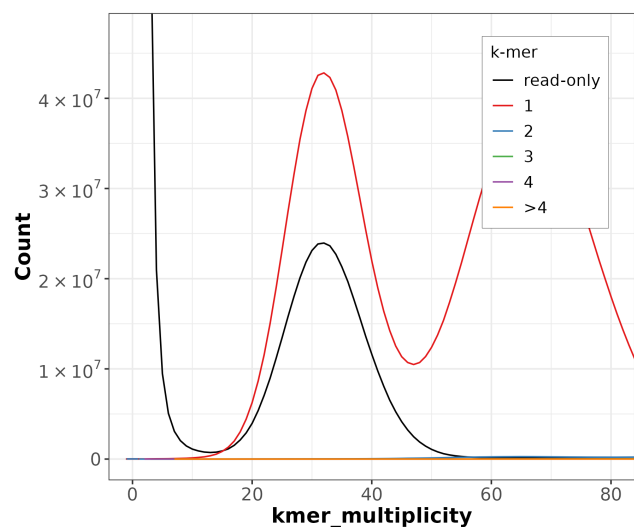
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: insecta_odb12 (genomes:79, BUSCOs:3114)

HiC contact map of curated assembly

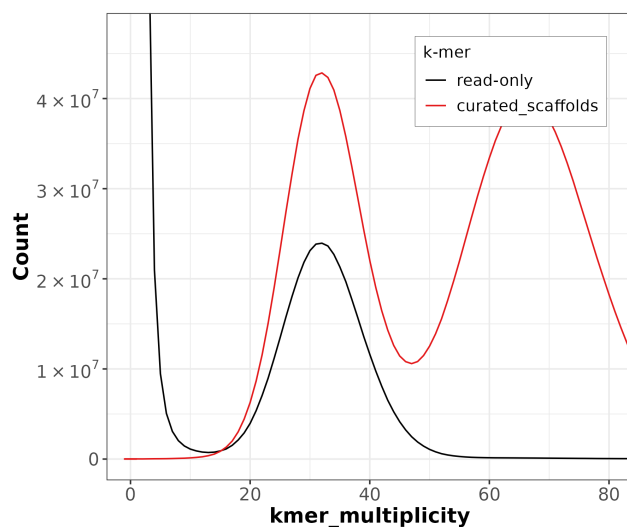


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

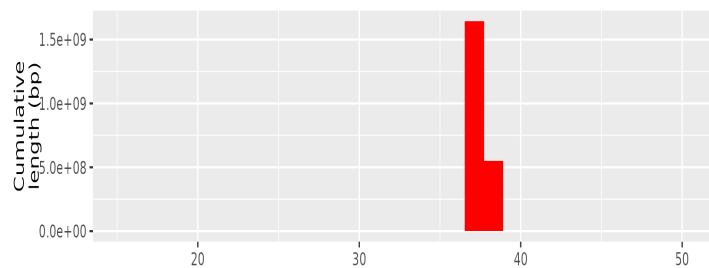


Distribution of k-mer counts per copy numbers found in asm

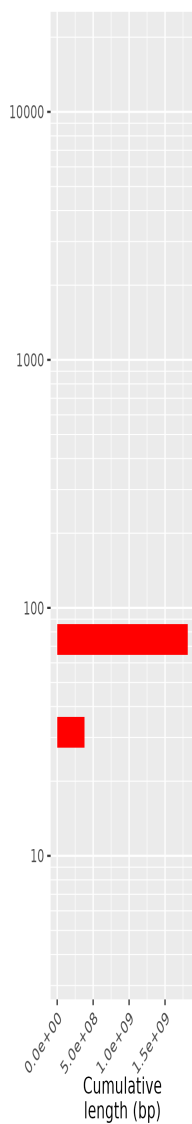
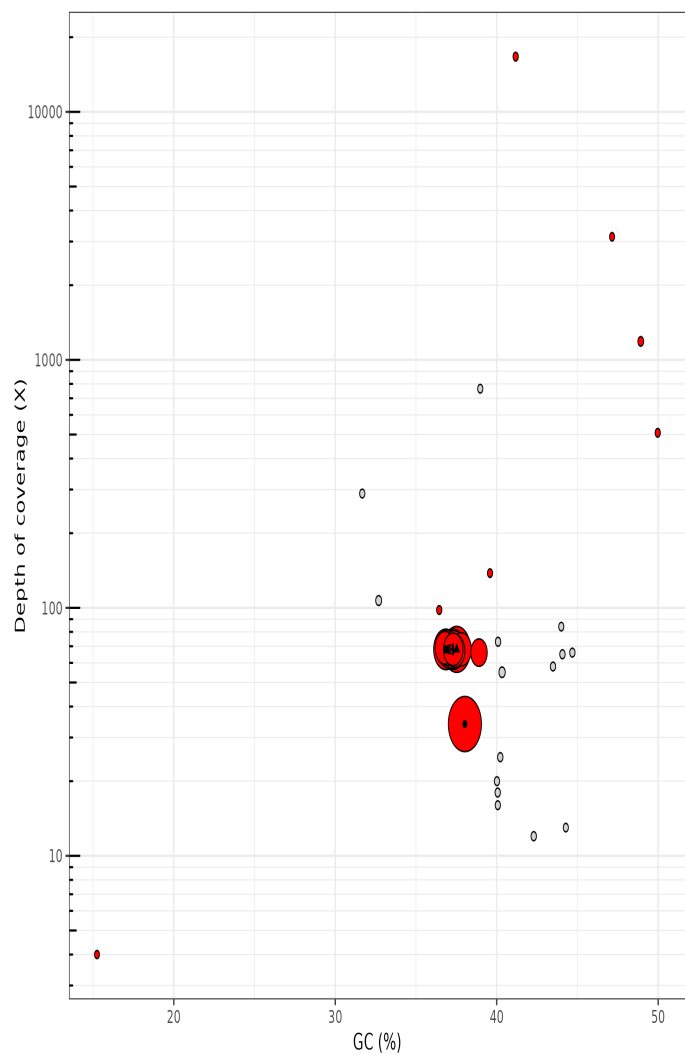


Distribution of k-mer counts coloured by their presence in reads/assemblies

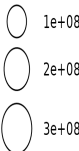
Post-curation contamination screening



TAPAs summary Graph



Length (bp)



Longest sequences (bp)

- SUPER_X - 379952532 (Eukaryota)
- ▲ SUPER_1 - 245530781 (Eukaryota)
- SUPER_2 - 180590110 (Eukaryota)
- + SUPER_3 - 177671489 (Eukaryota)
- ⊠ SUPER_4 - 171005424 (Eukaryota)

superkingdom

- Eukaryota
- N/A

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
Coverage	70	423

Assembly pipeline

- **Hifiasm**
 - |_ *ver*: 0.19.5-r593
 - |_ *key param*: NA
- **purge_dups**
 - |_ *ver*: 1.2.5
 - |_ *key param*: NA
- **YaHS**
 - |_ *ver*: 1.2
 - |_ *key param*: NA

Curation pipeline

- **PretextMap**
 - |_ *ver*: 0.1.9
 - |_ *key param*: NA
- **PretextView**
 - |_ *ver*: 0.2.5
 - |_ *key param*: NA

Submitter: Adama Ndar

Affiliation: Genoscope

Date and time: 2025-08-10 08:56:20 CEST