ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	3366363
ToLID	**icEndArme8**
Species	Endomychus armeniacus
Class	Insecta
Order	Coleoptera

Genome Traits	Expected	Observed
Haploid size (bp)	696,186,225	577,691,013
Haploid Number	9 (source: ancestor)	9
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q51

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

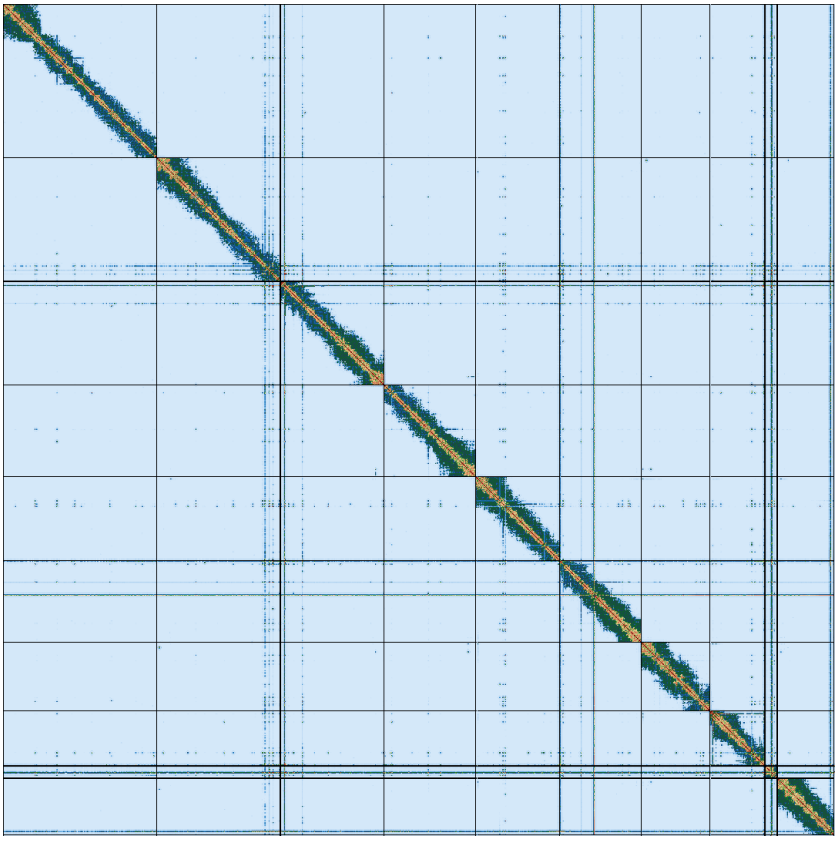
. Interventions/Gb: 176
. Contamination notes: ""
. Other observations: "The assembly of ENDOMYCHUS ARMENIACUS (icEndArme8) is based on 133X PacBio data and 143X Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA, https://www.erga-biodiversity.eu/) via the Biodiversity Genomics Europe project (BGE, https://biodiversitygenomics.eu/). The assembly process included the following steps: initial PacBio assembly generation withHifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups and Hi-C-based scaffolding with YaHS. In total, 5 contigs were identified as contaminants (bacterial), totaling 2,499,279 pb (with the largest being 1,018,873 pb). Additionally, 406 regions totaling 36,877,930 pb (with the largest being 390,577 pb) 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. During manual curation, 1 haplotypic region, totaling 618,187 pb and 19 contaminants, totaling 3,216,312 pb (with the largest being 501,195 pb) were removed. 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	591,296,259	577,691,013
GC %	33.9	33.82
Gaps/Gbp	591.92	586.82
Total gap bp	37,000	38,500
Scaffolds	118	23
Scaffold N50	63,372,648	63,483,907
Scaffold L50	4	4
Scaffold L90	9	8
Contigs	449	362
Contig N50	3,731,000	3,810,340
Contig L50	40	39
Contig L90	192	175
QV	51.1493	51.3006
Kmer compl.	87.6353	86.921
BUSCO sing.	96.1%	98.9%
BUSCO dupl.	0.3%	0.2%
BUSCO frag.	1.2%	0.0%
BUSCO miss.	2.4%	0.9%

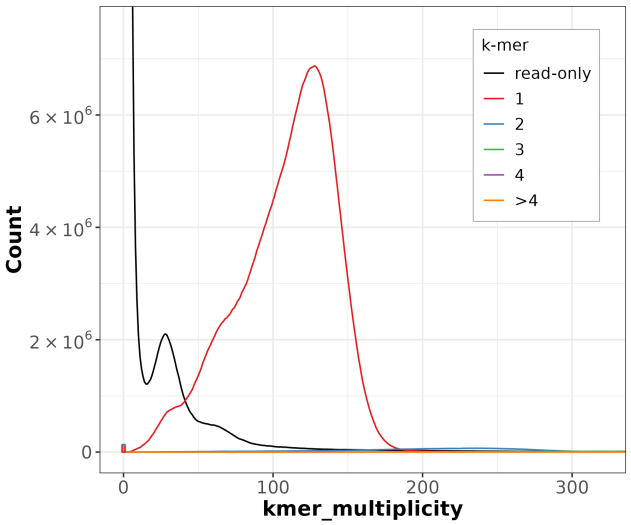
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: polyphaga_odb12 (genomes:60, BUSCOs:4010)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: polyphaga_odb12 (genomes:60, BUSCOs:4010)

HiC contact map of curated assembly

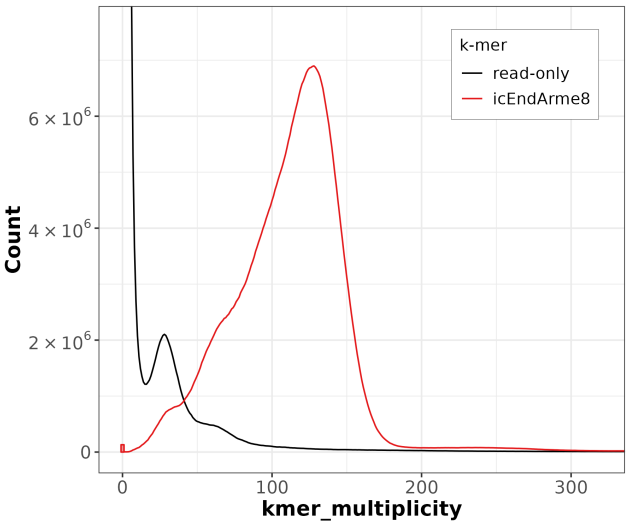


collapsed [LINK]

K-mer spectra of curated assembly

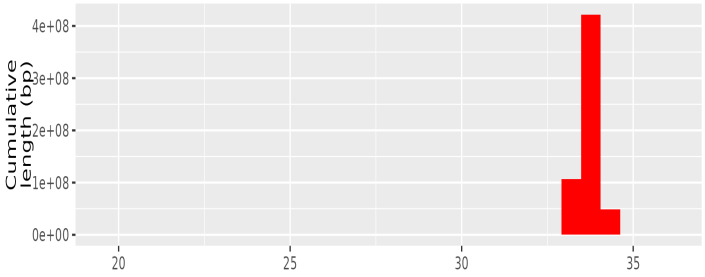


Distribution of k-mer counts per copy numbers found in asm

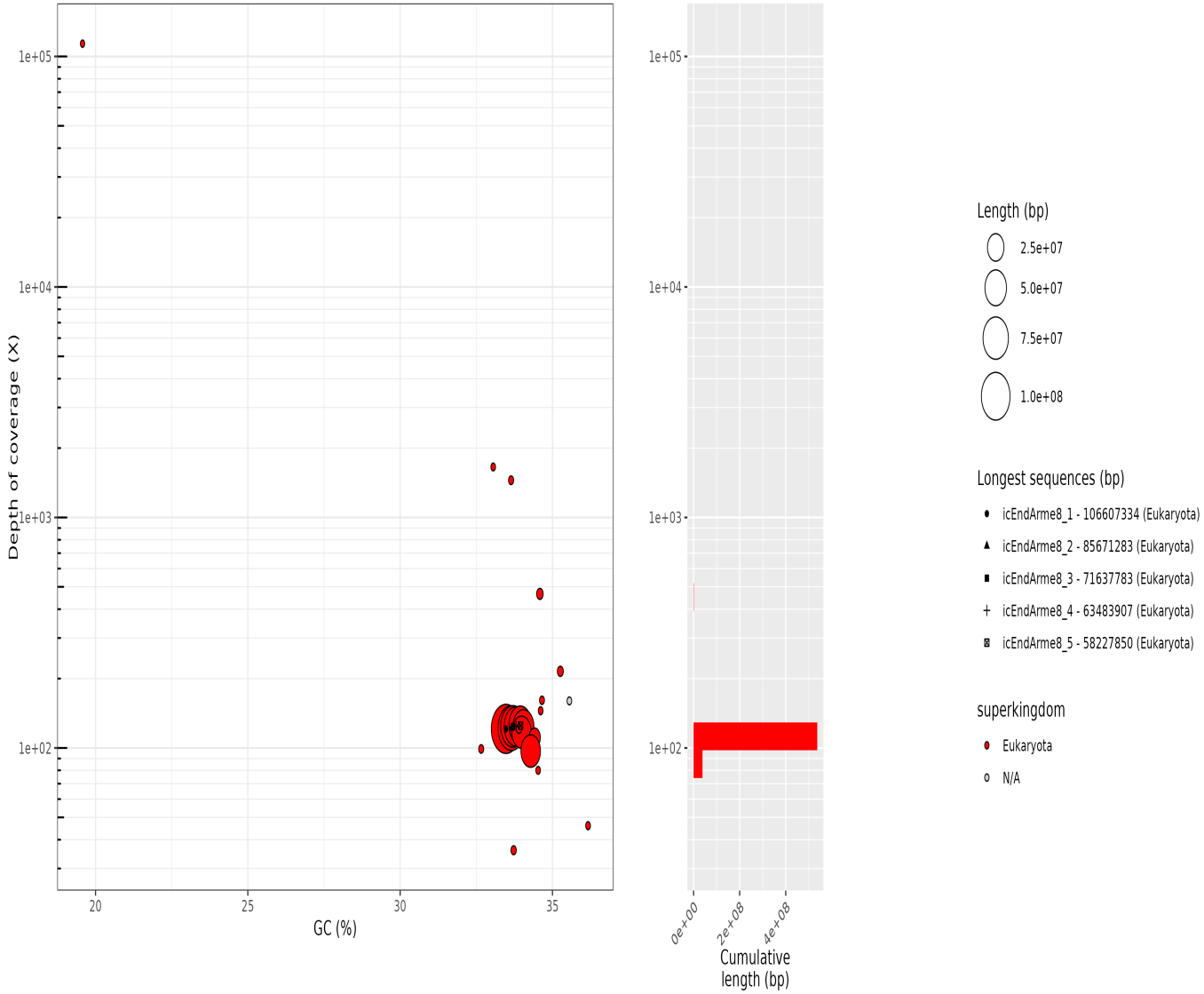


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	Long reads	Arima
Coverage	133	143

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: Lola Demirdjian
Affiliation: Genoscope

Date and time: 2025-10-16 05:45:06 CEST