

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	1265421
ToLID	<b>icMorAspel</b>
Species	<i>Morimus asper</i>
Class	Insecta
Order	Coleoptera

Genome Traits	Expected	Observed
Haploid size (bp)	897,237,207	884,731,495
Haploid Number	12 (source: ancestor)	12
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q67

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: 99
- . Contamination notes: ""
- . Other observations: "The assembly of MORIMUS ASPER (icMorAspel) is based on 47X PacBio data and 105X 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 Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge\_dups and Hi-C-based scaffolding with YaHS. In total, 6 contigs were identified as contaminants (bacterial), totaling 1,644,077 pb (with the largest being 1,548,558 pb). Additionally, 388 regions totaling 36 Mb (with the largest being 2,794,973 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, 5 haplotypic region were removed, totaling 3,762,895 pb and 8 scaffolds were identified as contaminants and removed. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

# Quality metrics table

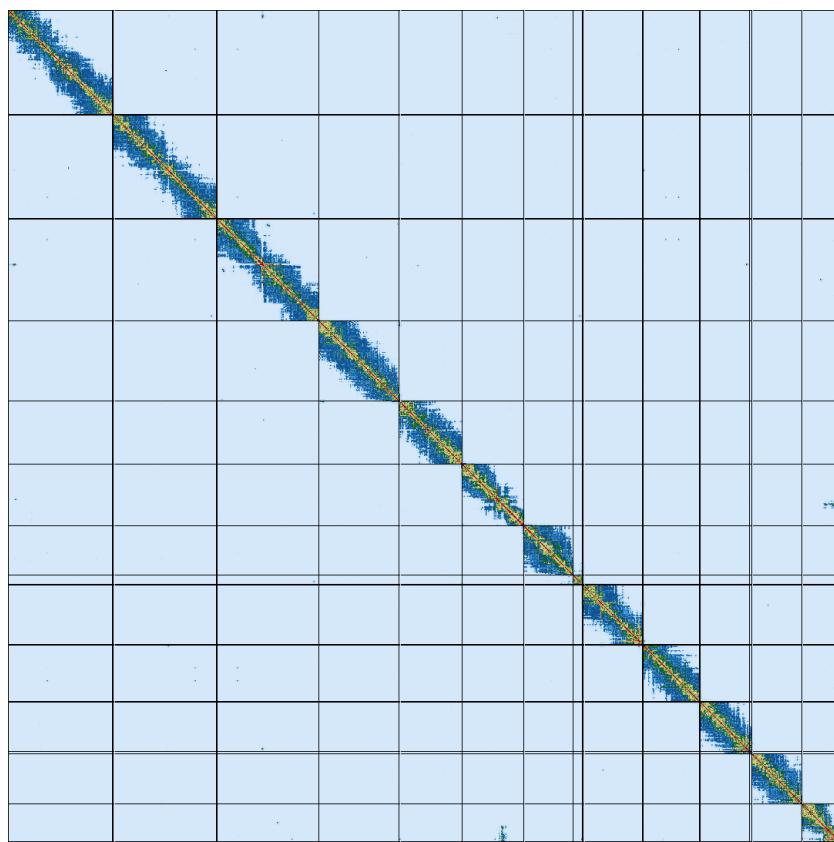
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	888,773,570	884,731,495
GC %	34.54	34.53
Gaps/Gbp	133.89	144.68
Total gap bp	11,900	14,800
Scaffolds	56	41
Scaffold N50	66,556,924	66,557,124
Scaffold L50	5	5
Scaffold L90	11	11
Contigs	175	169
Contig N50	13,036,397	12,620,000
Contig L50	21	21
Contig L90	77	78
QV	67.5746	67.5737
Kmer compl.	85.415	85.3015
BUSCO sing.	96.0%	97.6%
BUSCO dupl.	0.5%	0.7%
BUSCO frag.	0.5%	0.0%
BUSCO miss.	3.0%	1.7%

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)

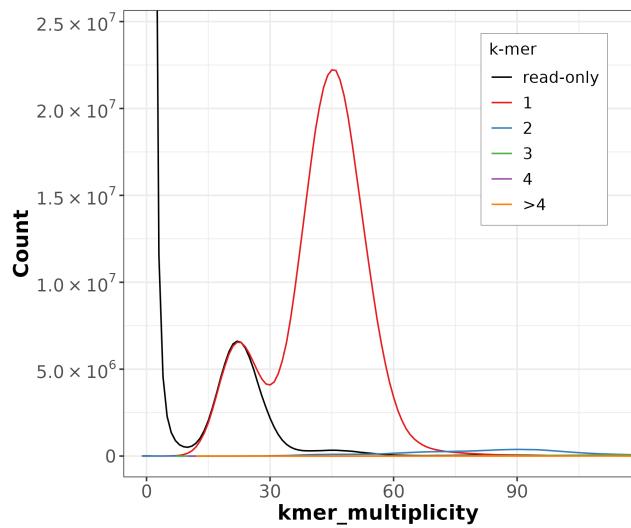
BUSCO: 6.0.0 (euk\_genome\_min, miniprot) / Lineage: polyphaga\_odb12 (genomes:60, BUSCOs:4010)

# HiC contact map of curated assembly

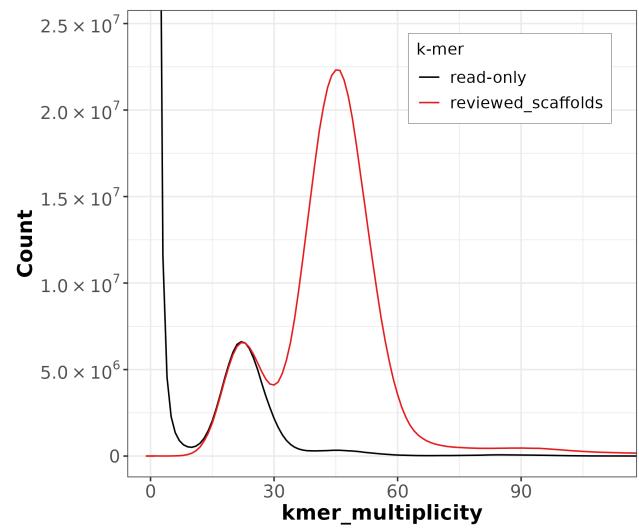


collapsed [\[LINK\]](#)

# K-mer spectra of curated assembly

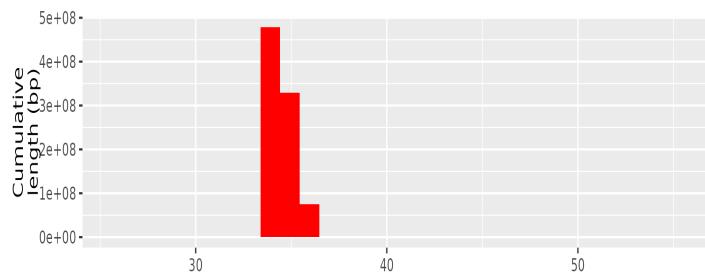


Distribution of k-mer counts per copy numbers found in asm

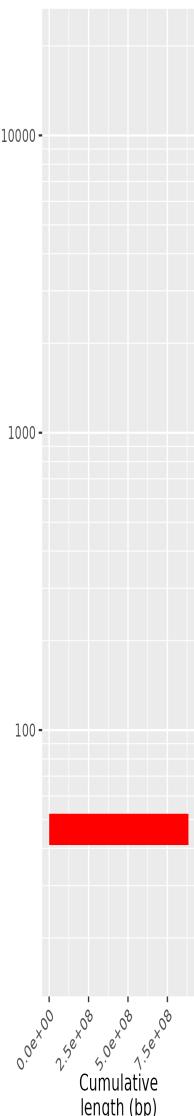
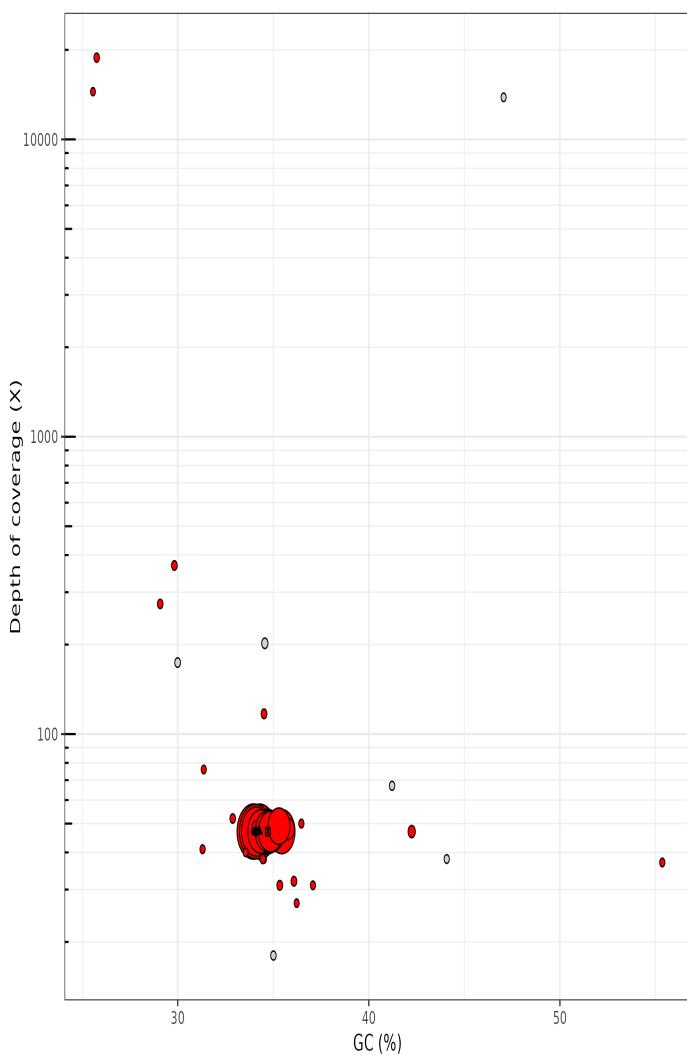


Distribution of k-mer counts coloured by their presence in reads/assemblies

# Post-curation contamination screening



TAPAs summary Graph



Length (bp)

- 3.0e+07
- 6.0e+07
- 9.0e+07

Longest sequences (bp)

- icMorAspel\_1 - 111354115 (Eukaryota)
- icMorAspel\_2 - 110279687 (Eukaryota)
- icMorAspel\_3 - 108535300 (Eukaryota)
- icMorAspel\_4 - 85016283 (Eukaryota)
- icMorAspel\_5 - 66557124 (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	Long reads	Arima
Coverage	46	104

## 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: 2025-10-28 10:47:13 CET