ERGA Assembly Report
v24.10.15

Tags: ERGA-BGE

TxID	2974020
ToLID	**icMorOrie1**
Species	Morimus orientalis
Class	Insecta
Order	Coleoptera

Genome Traits	Expected	Observed
Haploid size (bp)	889,376,030	900,791,559
Haploid Number	12 (source: direct)	13
Ploidy	2 (source: direct)	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:

. Observed Haploid Number is different from Expected

. Kmer completeness value is less than 90 for collapsed

Curator notes

. Interventions/Gb: 31
. Contamination notes: ""
. Other observations: "The assembly of Morimus orientalis (icMorOrie1) is based on 55X PacBio data and 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, 24 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 3.6 Mb (with the largest being 1.5Mb). Additionally, 358 regions totaling 36 Mb (with the largest being 3.5 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. During manual curation, no haplotypic regions were removed. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. The X chromosome has been identified thanks to comparison with another insect genome assembly and the lower coverage in Long Reads data. But the Y chromosome has not been identified in the assembly even if the specimen collected was

a male. "

Quality metrics table

| Metrics	Pre-curation collapsed	Curated collapsed
Total bp	901,550,623	900,791,559
GC %	34.54	34.54
Gaps/Gbp	137.54	152.09
Total gap bp	15,000	18,500
Scaffolds	138	113
Scaffold N50	67,096,688	67,096,688
Scaffold L50	5	5
Scaffold L90	11	11
Contigs	259	250
Contig N50	14,206,665	14,040,000
Contig L50	21	21
Contig L90	70	71
QV	67.7369	67.7805
Kmer compl.	88.1455	88.138
BUSCO sing.	96.7%	96.7%
BUSCO dupl.	0.5%	0.5%
BUSCO frag.	0.5%	0.5%
BUSCO miss.	2.3%	2.3%

BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: polyphaga_odb12 (genomes:60, BUSCOs:4010)
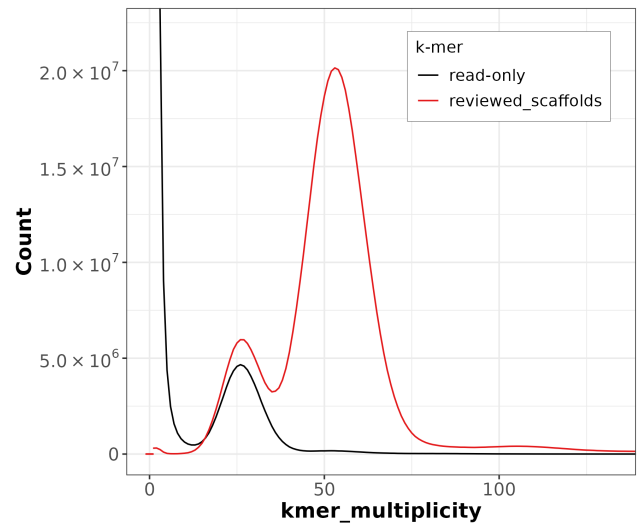
HiC contact map of curated assembly
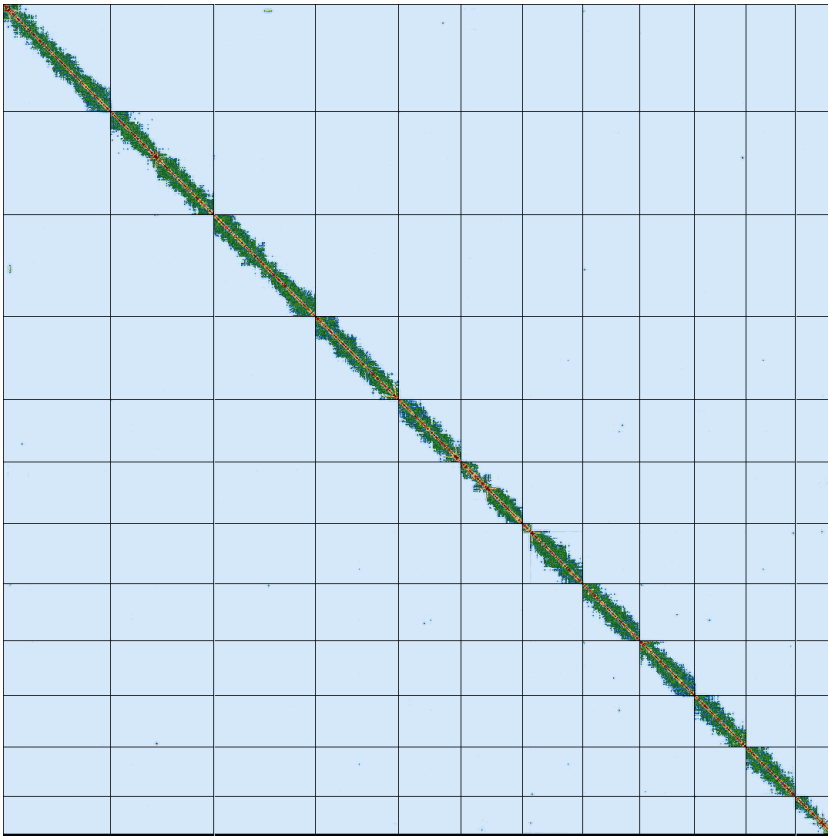
collapsed [LINK]
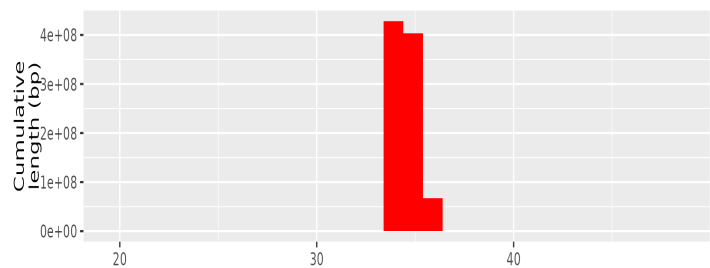
K-mer spectra of curated assembly
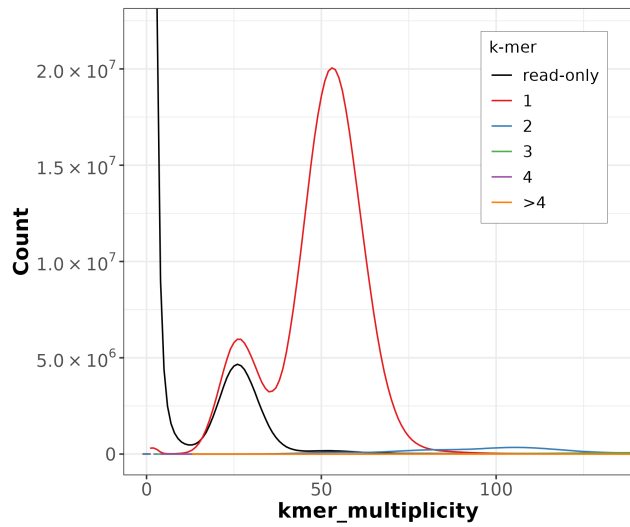
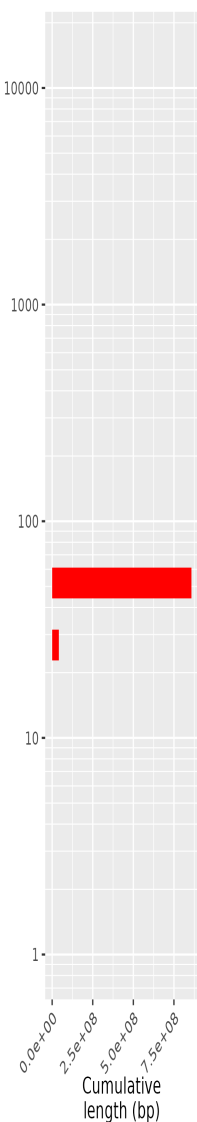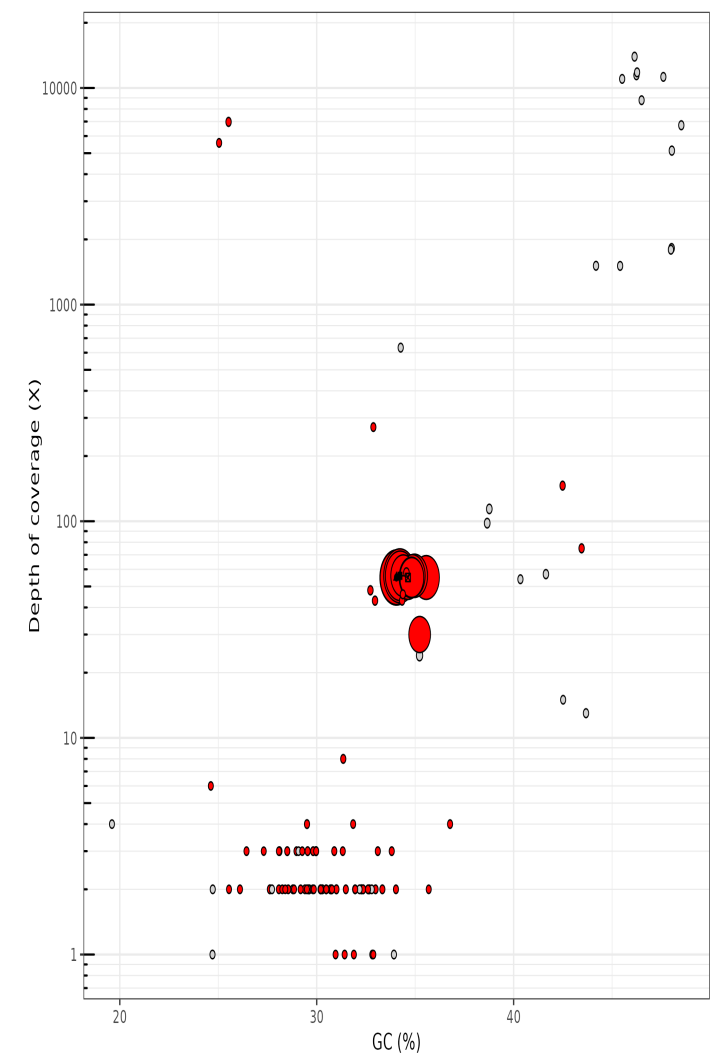
Distribution of k-mer counts per copy numbers found in asm

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	55	84

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: Caroline Menguy
Affiliation: Genoscope

Date and time: 2025-09-21 01:19:15 CEST