

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	1526741
ToLID	<b>xbXylDors1</b>
Species	Xylophaga dorsalis
Class	Bivalvia
Order	Myida

Genome Traits	Expected	Observed
Haploid size (bp)	453,366,001	451,794,626
Haploid Number	19 (source: ancestor)	18
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	ubnknown

## 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:

- . Observed Haploid Number is different from Expected
- . Observed sex is different from Sample sex
- . Kmer completeness value is less than 90 for collapsed
- . More than 1000 gaps/Gbp for collapsed

### Curator notes

- . Interventions/Gb: 367
- . Contamination notes: "We found contaminants in the yabs scaffolded assembly within six small scaffolds which were removed from the curated assembly submitted for review. The contaminants corresponded mostly to bacterial sequences (Pseudomonadota, Bacillota)"
- . Other observations: "The reviewing process led to several changes in many of the super scaffolds, with six additional haplotigs found and removed. We found it very difficult to localize a number of these scaffolds (mostly centromeric) which were set as unlocs. The sex of the specimen was set to -unknown- in our lms system. We could find the sex chromosome in the assembly. We found a large haplotig (4 Mbases) which was removed before remapping."

## Quality metrics table

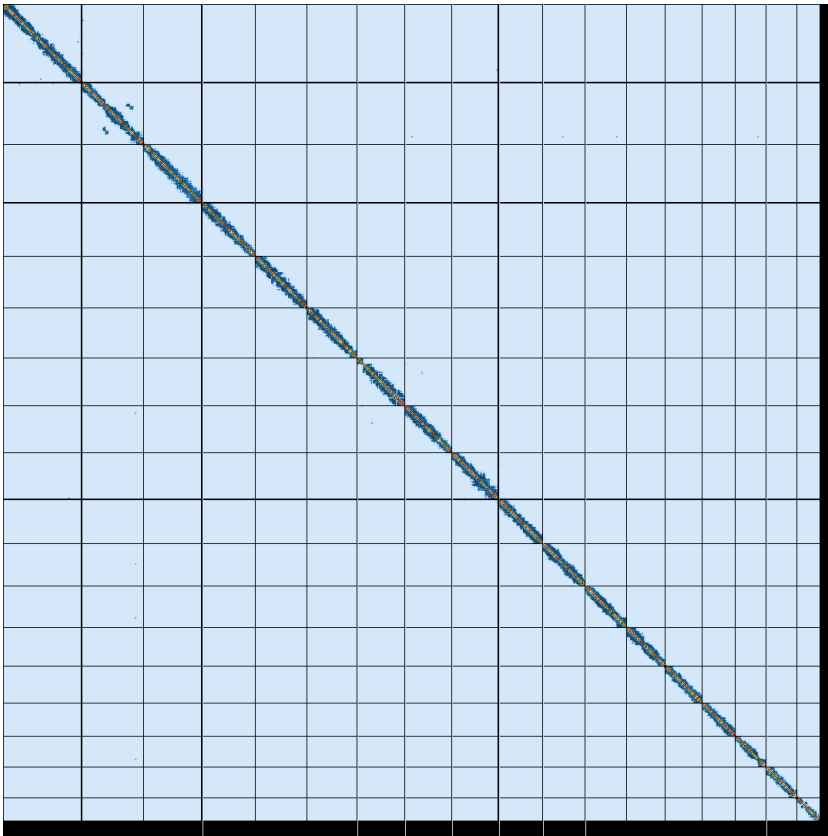
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	456,718,437	451,794,626
GC %	38.08	38.08
Gaps/Gbp	1,907.08	2,078.38
Total gap bp	174,200	187,800
Scaffolds	392	320
Scaffold N50	24,961,909	25,438,142
Scaffold L50	8	8
Scaffold L90	18	16
Contigs	1,263	1,259
Contig N50	1,293,548	1,253,263
Contig L50	83	85
Contig L90	509	510
QV	51.1047	51.104
Kmer compl.	68.8674	68.4412
BUSCO sing.	90.5%	91.2%
BUSCO dupl.	1.5%	0.7%
BUSCO frag.	4.6%	4.6%
BUSCO miss.	3.4%	3.5%

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

BUSCO: 5.4.0 (euk\_genome\_met, metaeuk) / Lineage: metazoa\_odb10 (genomes:65, BUSCOs:954)

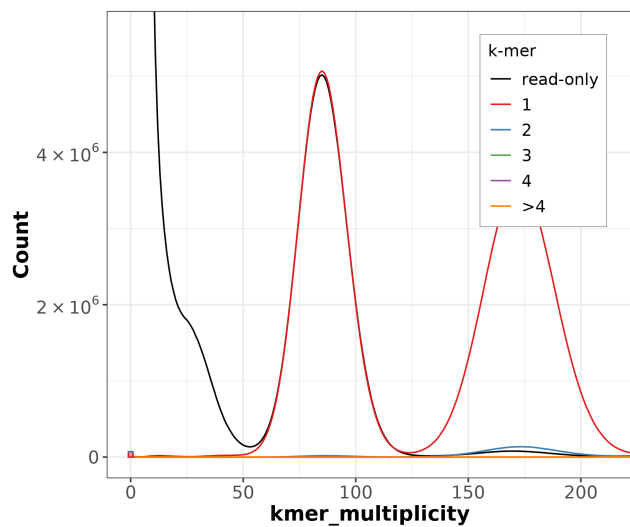
BUSCO: 5.5.0 (euk\_genome\_met, metaeuk) / Lineage: metazoa\_odb10 (genomes:65, BUSCOs:954)

# HiC contact map of curated assembly

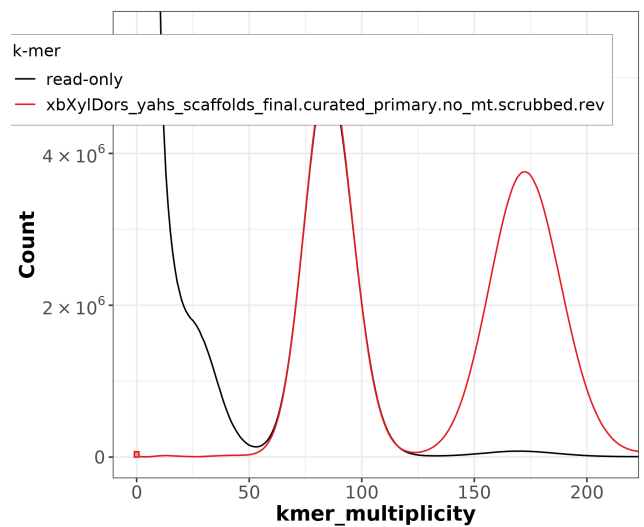


collapsed [\[LINK\]](#)

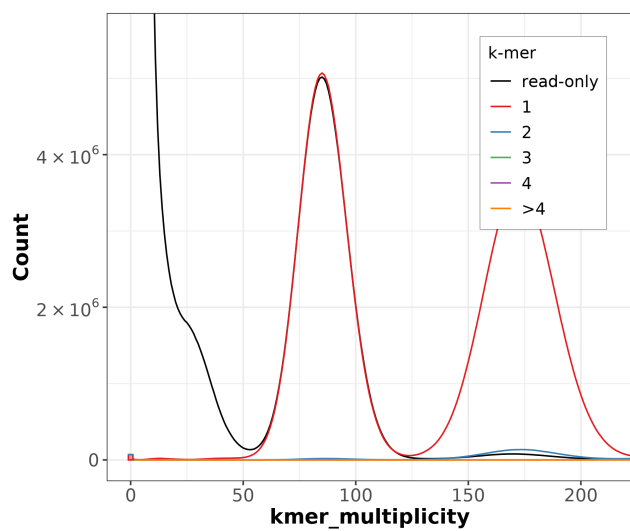
# K-mer spectra of curated assembly



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

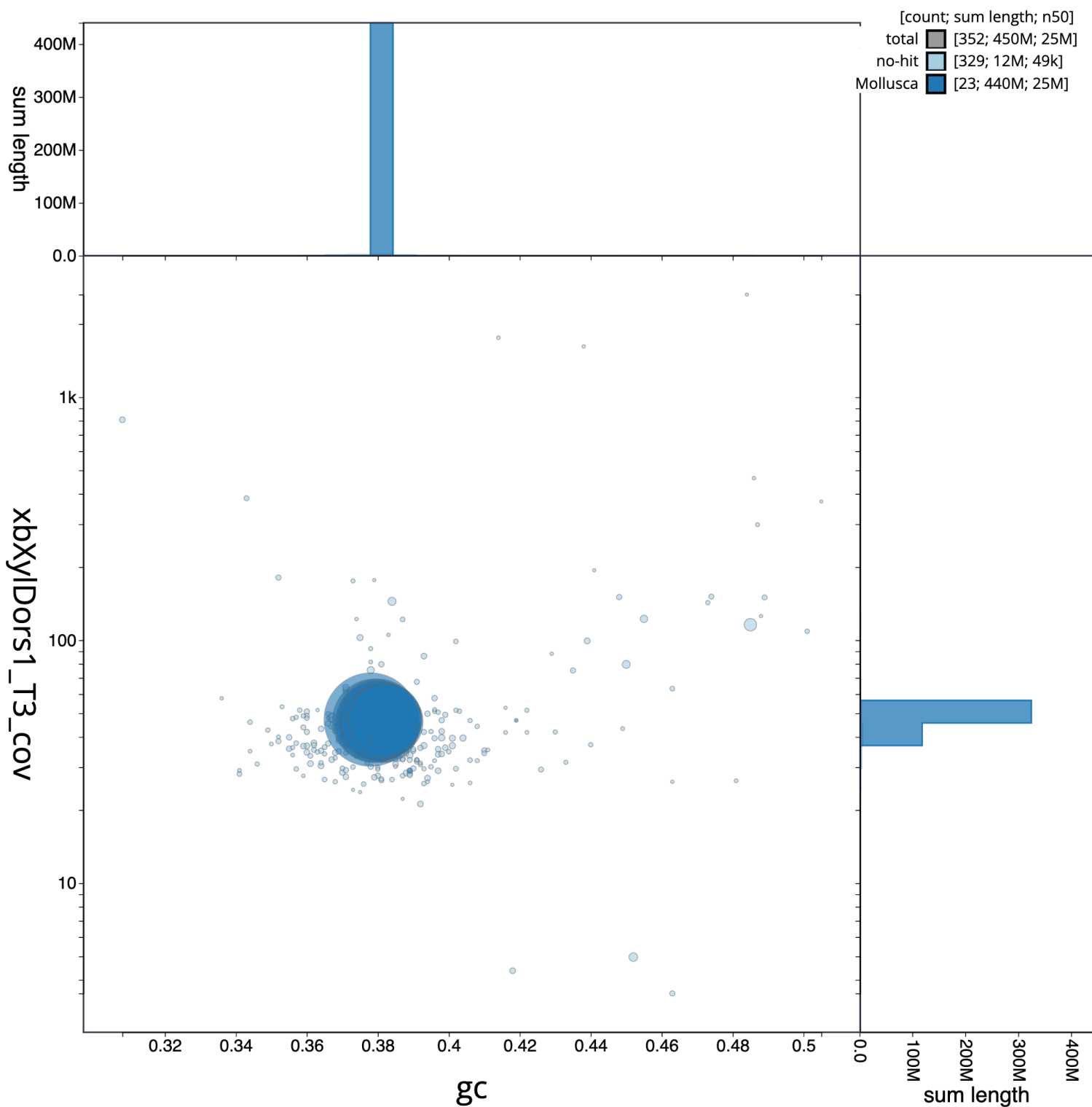


Distribution of k-mer counts coloured by their presence in reads/assemblies



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

# Post-curation contamination screening



**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	ONT	Illumina	OmniC
Coverage	93x	230x	178x

## Assembly pipeline

- **Trim\_galore**
  - |\_ *ver*: 0.6.7
  - |\_ *key param*: --gzip
  - |\_ *key param*: =q 20
  - |\_ *key param*: --paired
  - |\_ *key param*: retain\_unpaired
- **Filtlong**
  - |\_ *ver*: 0.2.1
  - |\_ *key param*: NA
- **Flye**
  - |\_ *ver*: 2.9.1-b1780
  - |\_ *key param*: NA
- **hypo**
  - |\_ *ver*: 1.0.3
  - |\_ *key param*: NA
- **purge\_dups**
  - |\_ *ver*: 1.2.6
  - |\_ *key param*: NA
- **YaHS**
  - |\_ *ver*: 1.2a
  - |\_ *key param*: NA

## Curation pipeline

- **PretextView**
  - |\_ *ver*: 0.2.5
  - |\_ *key param*: NA
- **GRIT\_Rapid**
  - |\_ *ver*: 2.0
  - |\_ *key param*: NA

Submitter: Francisco Camara

Affiliation: CNAG Barcelona

Date and time: 2025-03-21 16:17:25 CET