

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	47528
ToLID	<b>xbAnoCygn1</b>
Species	Anodonta cygnea
Class	Bivalvia
Order	Unionida

Genome Traits	Expected	Observed
Haploid size (bp)	2,395,037,852	1,608,328,603
Haploid Number	16 (source: ancestor)	0
Ploidy	2 (source: ancestor)	2
Sample Sex	NA	NA

## EBP metrics summary and curation notes

Obtained EBP quality metric for hap1: 6.6.Q65

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

- . Observed Haploid size (bp) has >20% difference with Expected
- . Observed Haploid Number is different from Expected
- . Kmer completeness value is less than 90 for hap1
- . BUSCO single copy value is less than 90% for hap1
- . Not 90% of assembly in chromosomes for hap1

### Curator notes

- . Interventions/Gb: None
- . Contamination notes: "Decontamination was performed with NCBI FCS were one protoebacteria contaminant was removed. Using Diamond and blobtools others contaminants, mainly Anellida, were removed. "
- . Other observations: "This is a contig level assembly. Despite the scaffolding the N50 was to low to procede with the curation."

# Quality metrics table

Metrics	Pre-curation hap1	Curated hap1
Total bp	1,608,328,603	1,608,328,603
GC %	34.99	34.99
Gaps/Gbp	9.33	9.33
Total gap bp	3,000	3,000
Scaffolds	1,024	1,024
Scaffold N50	3,253,191	3,253,191
Scaffold L50	151	151
Scaffold L90	552	552
Contigs	1,039	1,039
Contig N50	3,195,509	3,195,509
Contig L50	154	154
Contig L90	565	565
QV	65.555	65.555
Kmer compl.	66.2507	66.2507
BUSCO sing.	59.7%	59.7%
BUSCO dupl.	1.2%	1.2%
BUSCO frag.	1.3%	1.3%
BUSCO miss.	37.9%	37.9%

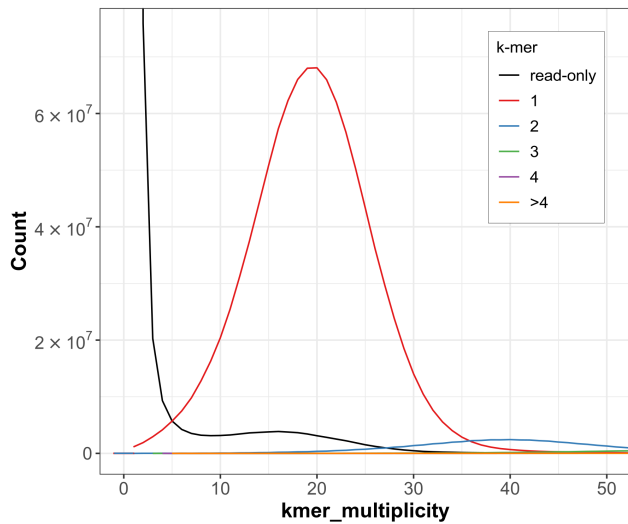
BUSCO: 5.8.0 (euk\_genome\_min, miniprot) / Lineage: mollusca\_odb10 (genomes:7, BUSCOs:5295)

# HiC contact map of curated assembly

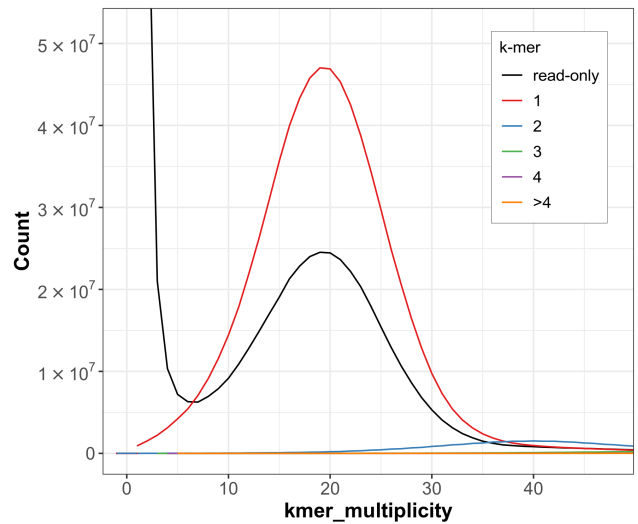
**hap1** HiC PNG is missing!

**hap1** File link is missing!

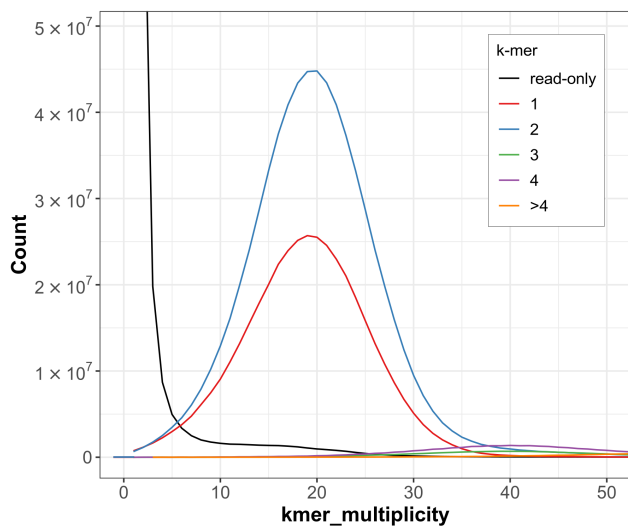
# K-mer spectra of curated assembly



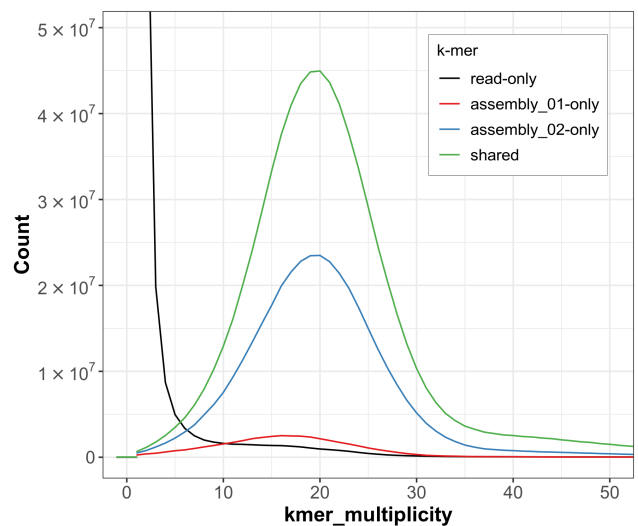
Distribution of k-mer counts per copy numbers found in **assembly\_02** (hapl.)



Distribution of k-mer counts per copy numbers found in **assembly\_01** (hapl.)

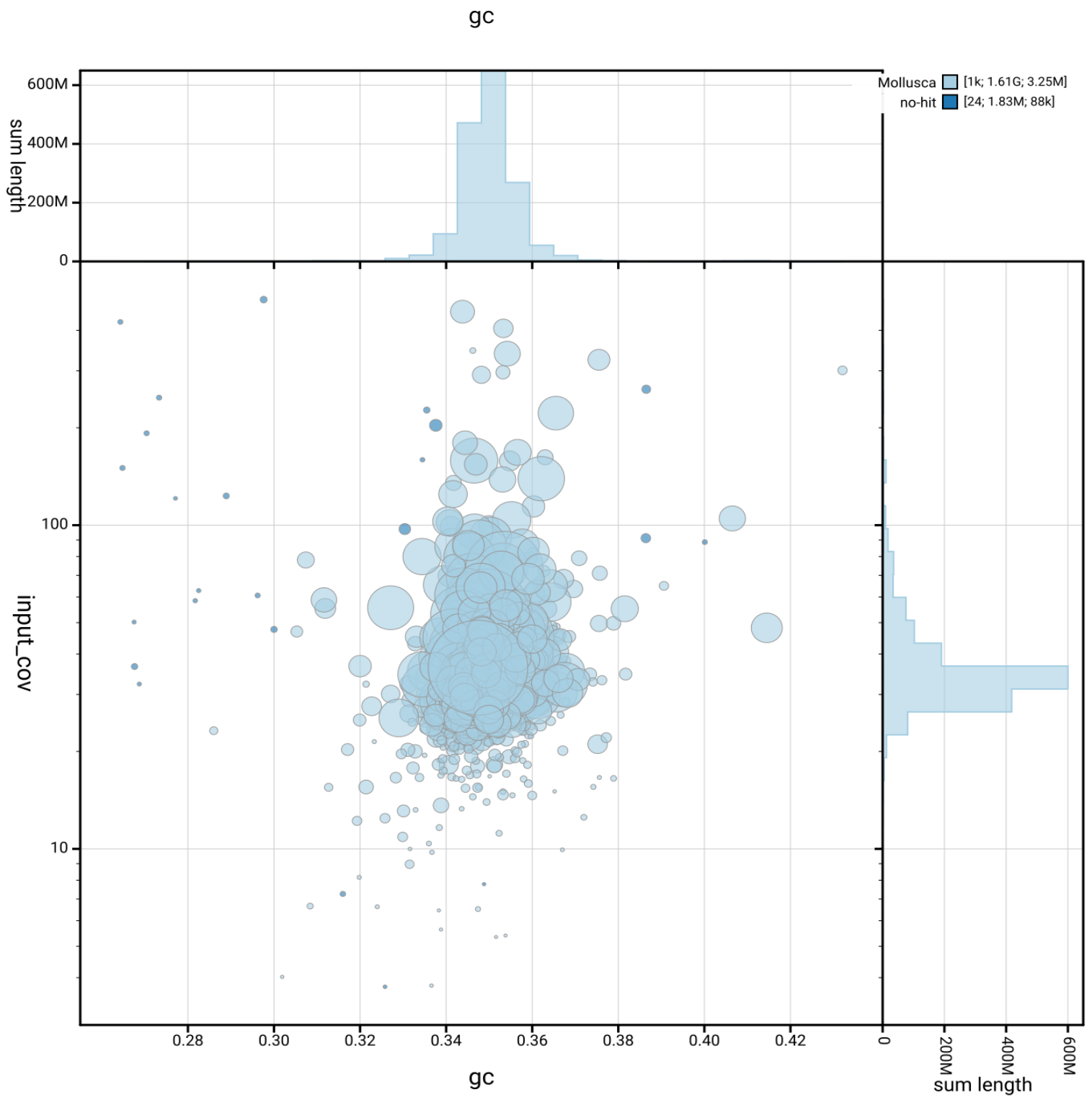


Distribution of k-mer counts per copy numbers found in **asm** (dipl.)



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

# Post-curation contamination screening



**hap1.** 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	HiFi	HiC
Coverage	25.20x	73.55x

## Assembly pipeline

- **Hifiasm**
  - |\_ *ver*: Galaxy Version 0.25.0+galaxy0
  - |\_ *key param*: VGP Galaxy workflow VGP3 v 0.3.4
- **purge\_dups**
  - |\_ *ver*: Galaxy Version 1.2.6+galaxy0
  - |\_ *key param*: VGP Galaxy workflow VGP6 v 0.10.4
- **YaHS**
  - |\_ *ver*: Galaxy Version 1.2a.2+galaxy3
  - |\_ *key param*: VGP Galaxy Workflow VGP8 v 3.3

## Curation pipeline

Invalid pipeline data format

Submitter: Marcella Sozzoni  
Affiliation: UNIFI

Date and time: 2026-02-19 11:32:18 CET