

# ERGA Assembly Report

v24.04.03\_beta

Tags: ERGA-BGE

TxID	3071549
ToLID	<b>xgAniVori1</b>
Species	Anisus vorticulus
Class	Gastropoda
Order	NA

Genome Traits	Expected	Observed
Haploid size (bp)	1,029,141,755	1,043,635,919
Haploid Number	18 (source: ancestor)	18
Ploidy	2 (source: ancestor)	2
Sample Sex	H	H

## EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.7.Q54

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

### Curator notes

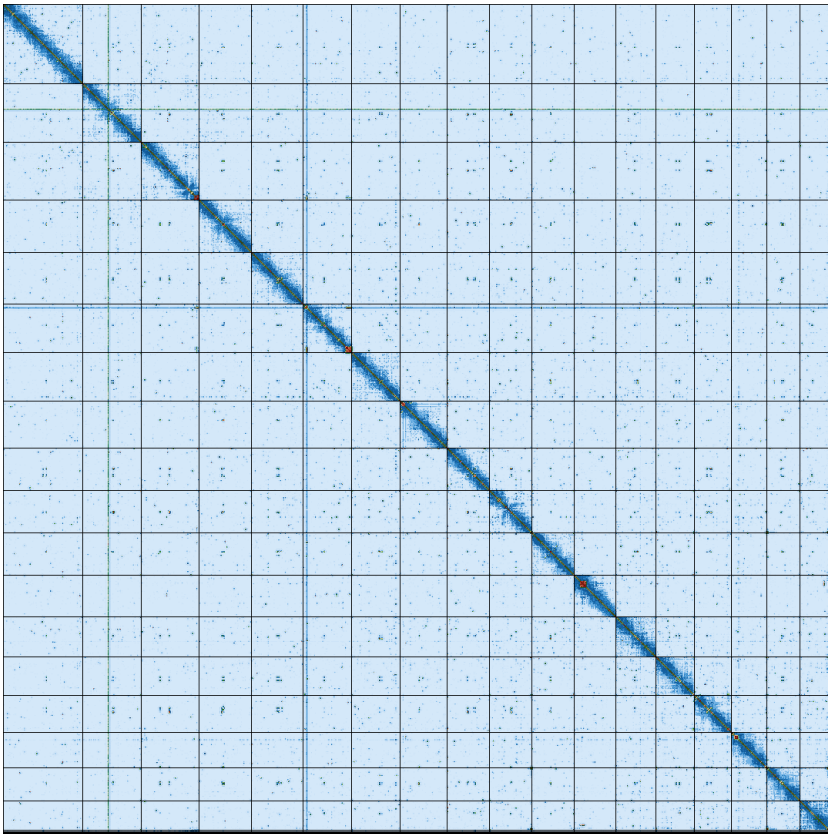
. Interventions/Gb: 123  
. Contamination notes: "Total length of scaffolds removed: 1,456,649 (0.1 %);Scaffolds removed: 53 (20.0 %);Largest scaffold removed: (57,806);FCS-GX contaminant species (number of scaffolds; total length of scaffolds):Pararheinheimera texasensis, g-proteobacteria (17; 358,477);Rheinheimera riviphila, g-proteobacteria (11; 369,938);Flavobacterium difficile, CFB group bacteria (5; 193,175);Pseudomonas campi, g-proteobacteria (3; 93,379);Acinetobacter johnsonii, g-proteobacteria (2; 84,971);Hydrogenophaga aromaticivorans, b-proteobacteria (2; 78,989);Flavobacterium amnicola, CFB group bacteria (2; 36,842);Flavobacterium succinicans, CFB group bacteria (1; 41,361);Aquabacterium pictum, b-proteobacteria (1; 37,989);Lacihabitans soyangensis, CFB group bacteria (1; 34,303);Rhodoferax sediminis, b-proteobacteria (1; 32,174);Flavobacterium nackdongense, CFB group bacteria (1; 30,736);Pelomonas sp., b-proteobacteria (1; 29,770);Rubrivivax albus, b-proteobacteria (1; 22,812);Rubrivivax sp., b-proteobacteria (1; 2,000);Flavobacterium bernardetii, CFB group bacteria (1; 2,000);Ideonella sp. WA131b, b-proteobacteria (1; 1,000);Mitochondrion (1; 6,733)"  
. Other observations: "Chromosomes named by size; Hi-C from a different individual (xgAniVori2), so assembly is not Hi-C phased"

# Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	1,049,506,174	1,043,635,919
GC %	35.29	35.28
Gaps/Gbp	662.22	725.35
Total gap bp	139,000	151,400
Scaffolds	265	79
Scaffold N50	55,257,980	59,244,224
Scaffold L50	8	8
Scaffold L90	17	16
Contigs	960	836
Contig N50	2,384,665	2,352,869
Contig L50	129	130
Contig L90	438	443
QV	54.5	54.5
Kmer compl.	98.89	99.27
BUSCO sing.	92.8%	92.8%
BUSCO dupl.	0.9%	0.9%
BUSCO frag.	2.3%	2.3%
BUSCO miss.	4.0%	4.0%

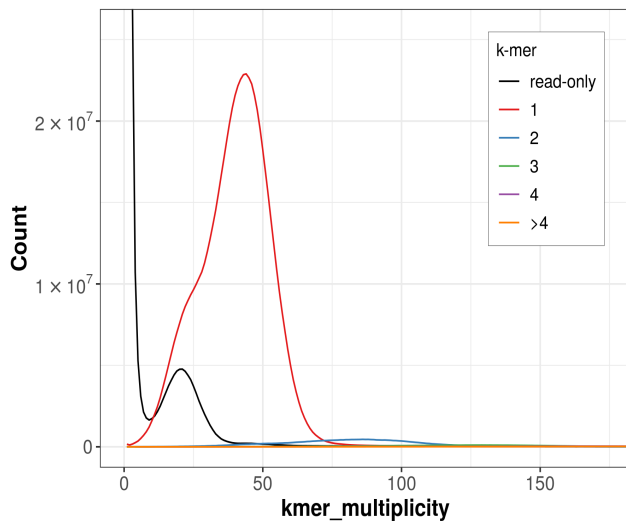
BUSCO 5.3.0 Lineage: mollusca\_odb10 (genomes:7, BUSCOs:5295)

# HiC contact map of curated assembly

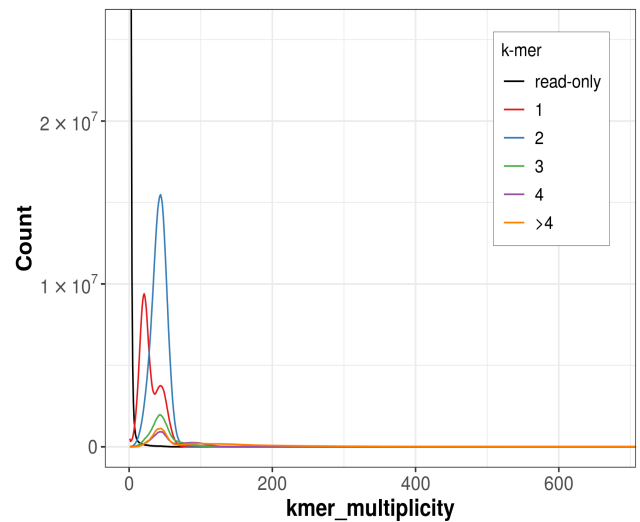


pri [\[LINK\]](#)

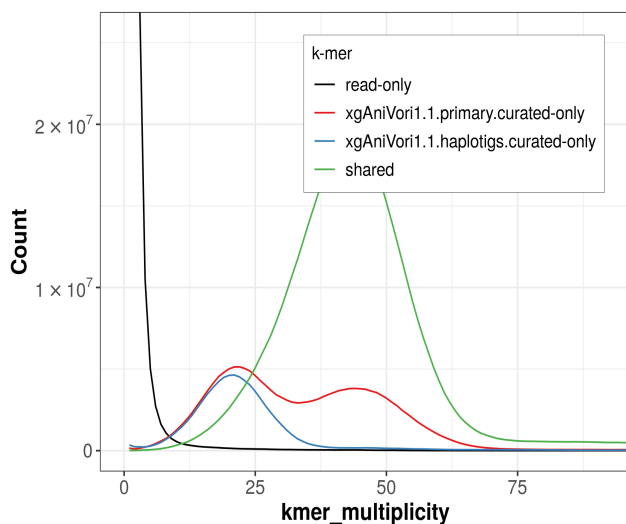
# K-mer spectra of curated assembly



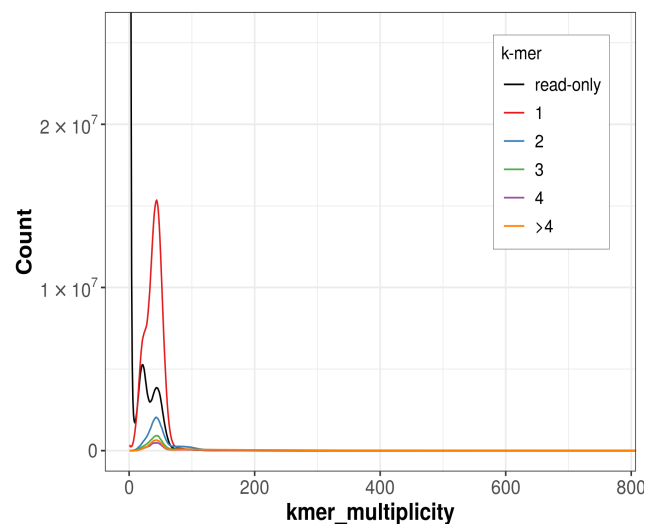
Distribution of k-mer counts per copy numbers found in (hapl.)



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

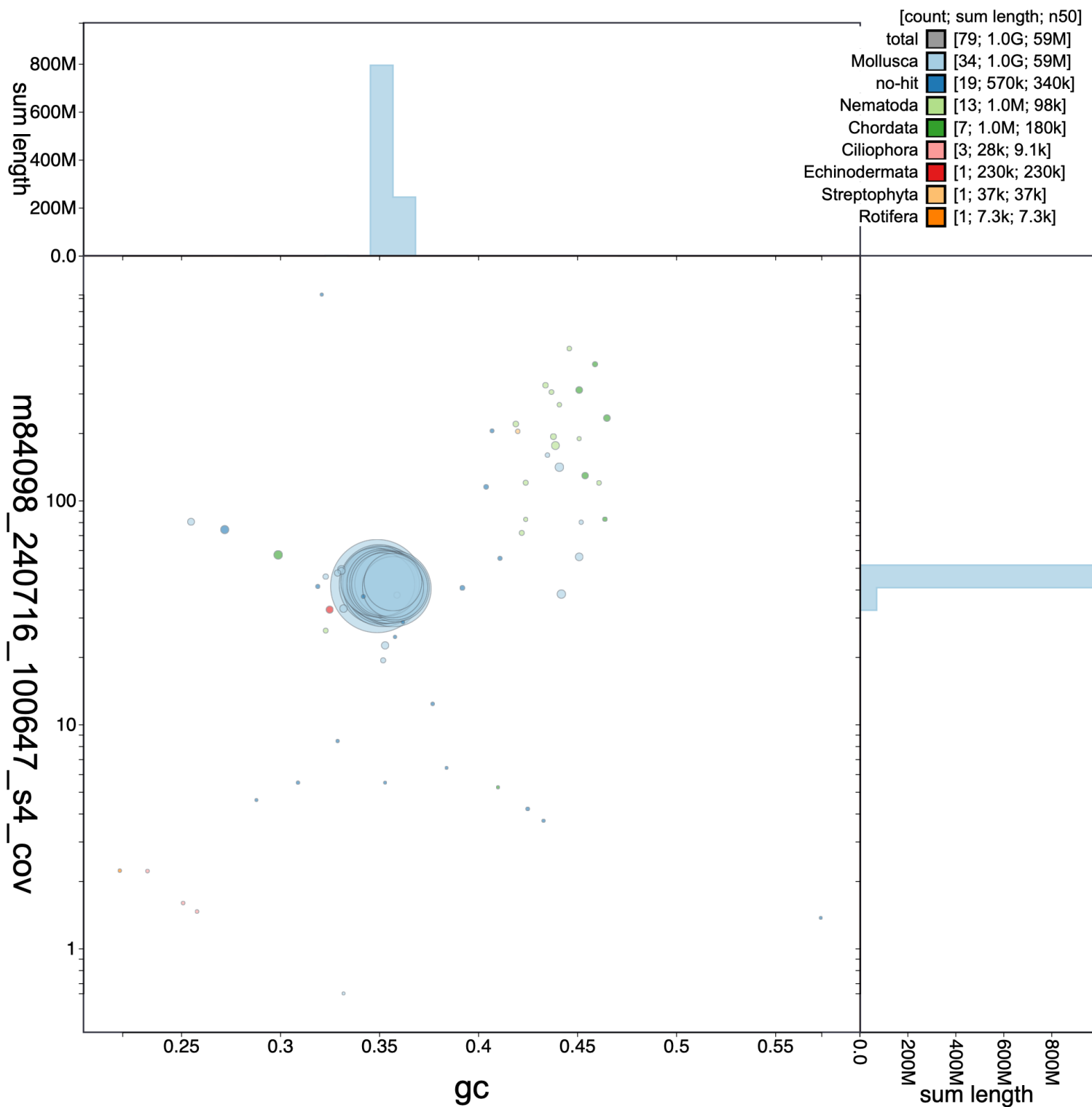


Distribution of k-mer counts coloured by their presence in reads/assemblies



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

# Post-curation contamination screening



**pri.** 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	PacBio HiFi	Arima v2
Coverage	45x	142x

# Assembly pipeline

- **hifiasm**
  - |\_ *ver*: 0.19.8-r603
  - |\_ *key param*: --primary
- **purge\_dups**
  - |\_ *ver*: 1.2.5
  - |\_ *key param*: -e
- **yahs**
  - |\_ *ver*: 1.2a.2
  - |\_ *key param*: NA

# Curation pipeline

- **hifiasm**
  - |\_ *ver*: 0.19.8-r603
  - |\_ *key param*: --primary
- **purge\_dups**
  - |\_ *ver*: 1.2.5
  - |\_ *key param*: -e
- **yahs**
  - |\_ *ver*: 1.2a.2
  - |\_ *key param*: NA
- **TreeVal**
  - |\_ *ver*: 1.1
  - |\_ *key param*: NA

Submitter: Michael Paulini

Affiliation: WSI

Date and time: 2024-08-16 16:05:56 CEST