

# ERGA Assembly Report

v24.04.03\_beta

Tags: ERGA-BGE

TxID	488418
ToLID	<b>wsMarCava1</b>
Species	Marifugia cavatica
Class	Polychaeta
Order	Sabellida

Genome Traits	Expected	Observed
Haploid size (bp)	1,215,561,629	1,321,948,386
Haploid Number	14 (source: ancestor)	11
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 5.8.Q55

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

- . Observed Haploid Number is different from Expected
- . BUSCO single copy value is less than 90% for pri
- . BUSCO duplicated value is more than 5% for pri
- . Assembly length loss > 3% for pri
- . More than 1000 gaps/Gbp for pri

### Curator notes

. Interventions/Gb: 557  
. Contamination notes: "Total length of scaffolds removed: 2,641,141 (0.2 %);Scaffolds removed: 24 (3.3 %);Largest scaffold removed: (1,646,899);FCS-GX contaminant species (number of scaffolds; total length of scaffolds)::Terrimonas sp., CFB group bacteria (1; 1,646,899);Sediminibacterium salmoneum, CFB group bacteria (1; 188,069);Candidatus Methylophosphatis roskildensis, b-proteobacteria (1; 121,725);Pedobacter quisquiliarum, CFB group bacteria (1; 89,660);Reyranella sp., a-proteobacteria (1; 74,734);Solimonas fluminis, g-proteobacteria (1; 72,394);Paludibacter sp. UBA5753, CFB group bacteria (1; 64,682);Chitinophaga chungangae, CFB group bacteria (1; 38,422);Pedobacter ureilyticus, CFB group bacteria (1; 34,861);Lewinella cohaerens, CFB group bacteria (1; 32,076);Limnohabitans parvus, b-proteobacteria (1; 31,546);Iodobacter ciconiae, b-proteobacteria (1; 30,257);Gelidibacter maritimus, CFB group bacteria (1; 28,085);Cupriavidus lacunae, b-proteobacteria (1; 24,928);Dyadobacter beijingensis, CFB group bacteria (1; 21,585);Hoeflea marina, a-proteobacteria (1; 20,549);Methyloversatilis sp.,

b-proteobacteria (1; 19,482);Rhodoferax sediminis, b-proteobacteria (1; 18,802);Haliscomenobacter hydrossis, CFB group bacteria (1; 18,768);Limnovirga soli, CFB group bacteria (1; 18,687);Ottowia beijingensis, b-proteobacteria (1; 13,239);Dyadobacter jejuensis, CFB group bacteria (1; 2,000);Niastella soli, CFB group bacteria (1; 1,000)"

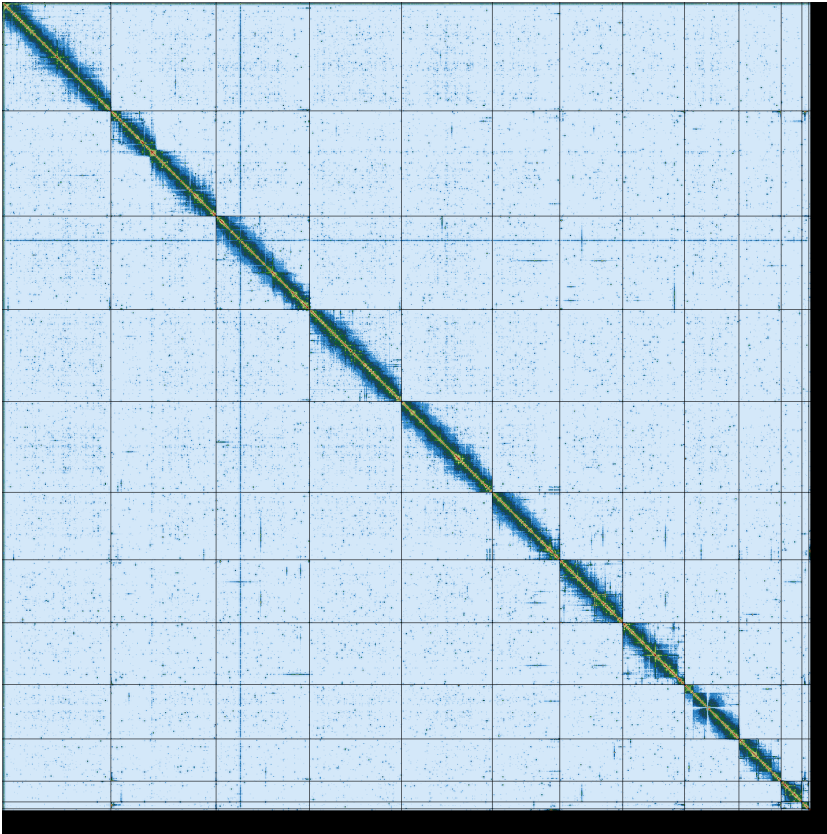
. Other observations: "PacBio is from a ULI library, so lower contig N50 is expected; Hi-C from a different individual (wsMarCava2), so assembly is not Hi-C phased; Expected haploid number from GoaT is an ancestral estimate, likely incorrect in this case."

## Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	1,364,684,544	1,321,948,386
GC %	36.41	36.41
Gaps/Gbp	2,720.04	2,832.18
Total gap bp	742,400	748,800
Scaffolds	728	581
Scaffold N50	91,216,656	144,721,733
Scaffold L50	6	5
Scaffold L90	14	10
Contigs	4,440	4,325
Contig N50	569,563	563,238
Contig L50	707	700
Contig L90	2,476	2,431
QV	55.2	55.2
Kmer compl.	97.46	98.63
BUSCO sing.	81.2%	80.7%
BUSCO dupl.	7.3%	8.1%
BUSCO frag.	6.6%	6.2%
BUSCO miss.	4.9%	5.0%

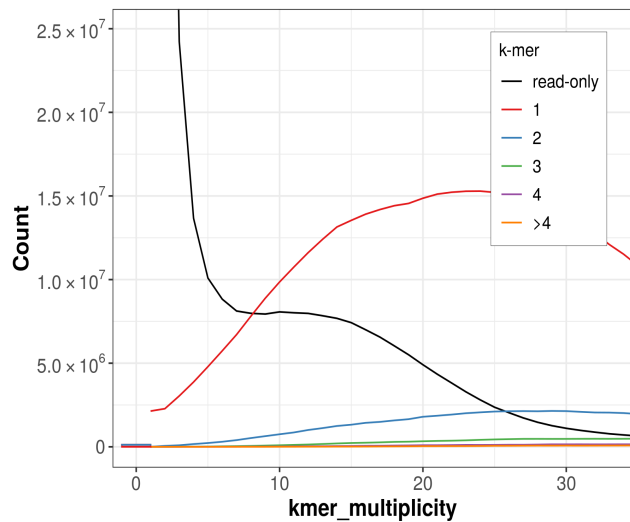
Warning: BUSCO versions or lineage datasets are not the same across results

# HiC contact map of curated assembly

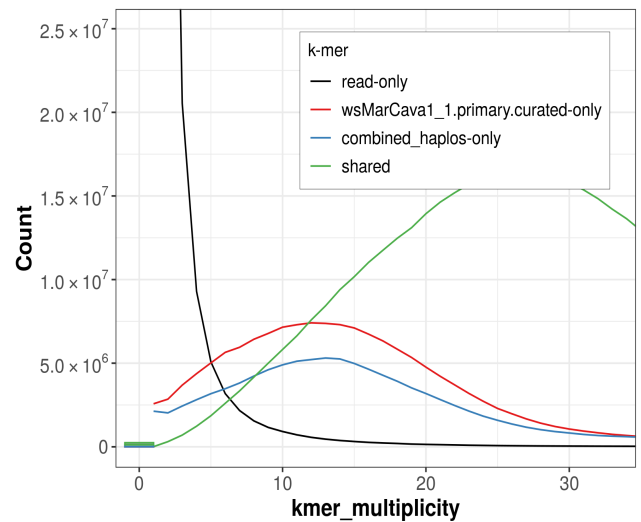


pri [\[LINK\]](#)

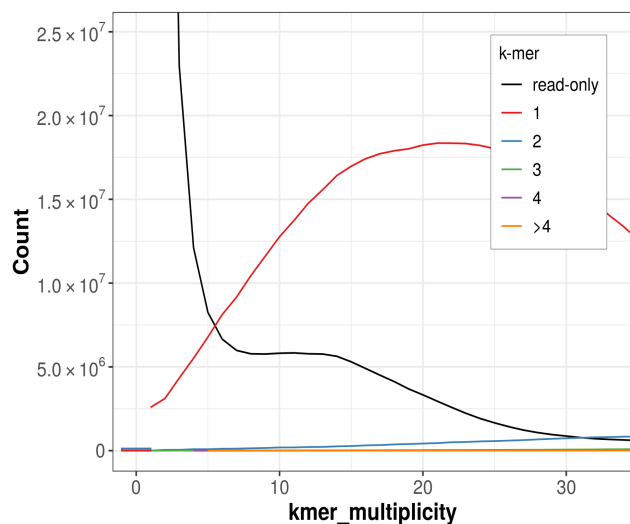
# K-mer spectra of curated assembly



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

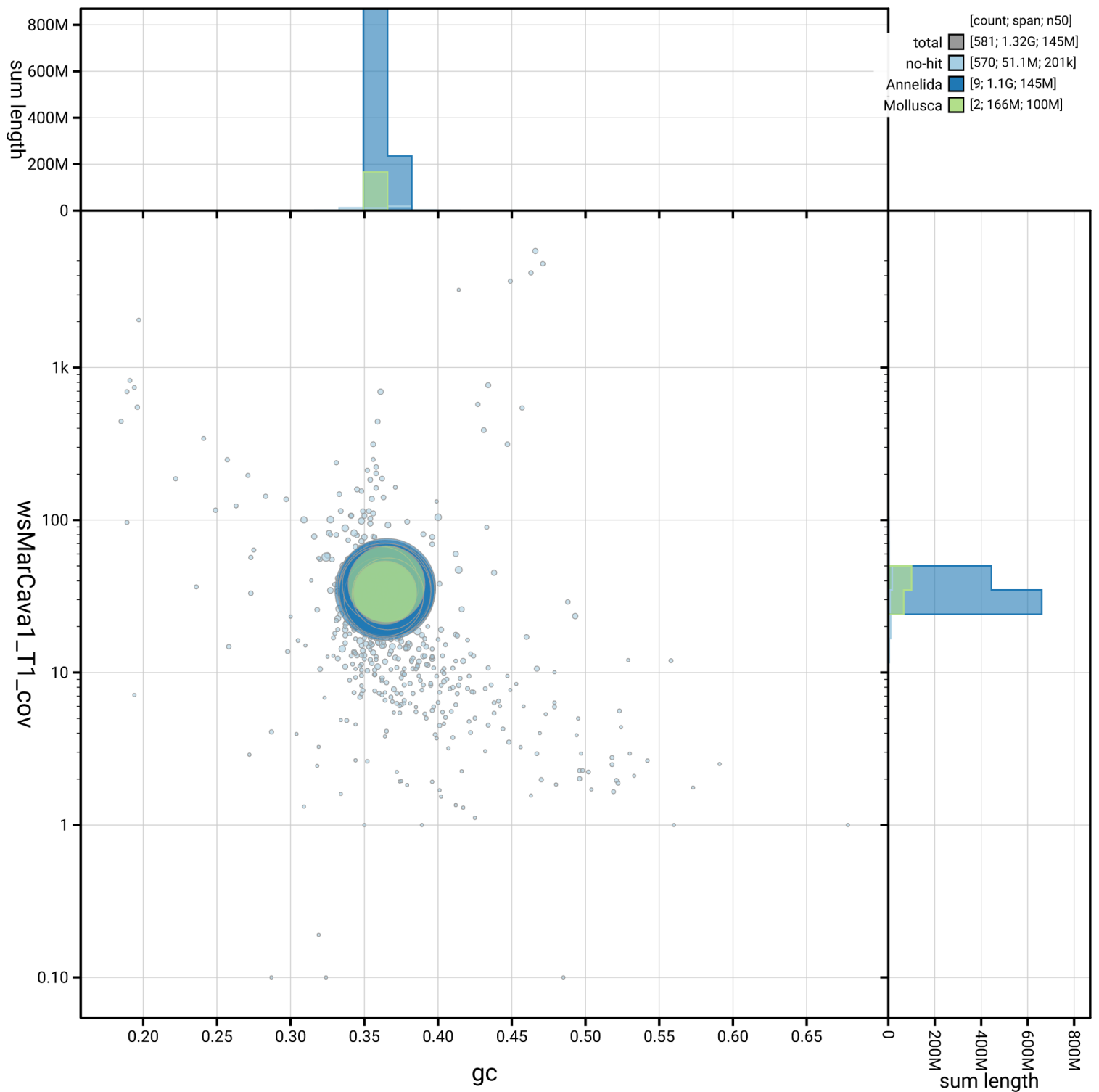


Distribution of k-mer counts coloured by their presence in reads/assemblies



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

# Post-curation contamination screening



**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	32x	83x

# 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*: 1.1.1
  - |\_ *key param*: NA

Submitter: Thomas Mathers

Affiliation: WSI

Date and time: 2024-10-22 13:54:30 CEST