

# ERGA Assembly Report

v24.04.03\_beta

Tags: ERGA-BGE

TxID	270691
ToLID	<b>wcHelEuro2</b>
Species	Helobdella europaea
Class	Clitellata
Order	Rhynchobdellida

Genome Traits	Expected	Observed
Haploid size (bp)	200,414,275	247,076,144
Haploid Number	9 (source: ancestor)	18
Ploidy	2 (source: ancestor)	2
Sample Sex	H	H

## EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 5.7.Q54

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
- . Assembly length loss > 3% for pri
- . More than 1000 gaps/Gbp for pri

### Curator notes

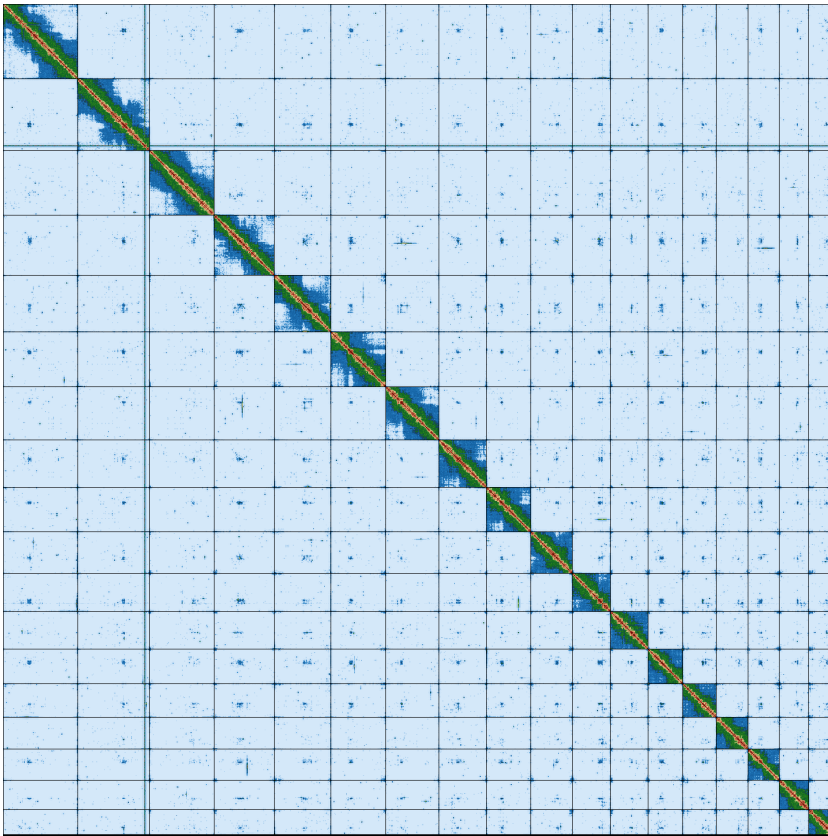
- . Interventions/Gb: 1124
- . Contamination notes: "Contamination report for assembly labelled primary; Total length of scaffolds removed: 9,839,755 (3.8 %); Scaffolds removed: 62 (9.1 %); Largest scaffold removed: (2,264,000); FCS-GX contaminant species (number of scaffolds; total length of scaffolds):; Aeromonas hydrophila, g-proteobacteria (45; 4,449,244); Flavobacterium collinsii, CFB group bacteria (9; 5,224,684); Brevundimonas sp. UBA6550, a-proteobacteria (7; 150,695); Zootermopsis nevadensis, insects (1; 15,132); FCS-Adaptor (8; 577)"
- . Other observations: "Hi-C was from a separate individual (wcHelEuro3);"

# Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	259,830,783	247,076,144
GC %	31.95	31.32
Gaps/Gbp	9,598.55	11,409.44
Total gap bp	249,400	342,800
Scaffolds	682	109
Scaffold N50	13,616,141	15,836,313
Scaffold L50	8	7
Scaffold L90	18	16
Contigs	3,176	2,928
Contig N50	126,265	124,626
Contig L50	589	587
Contig L90	1,993	1,915
QV	54.0	54.1
Kmer compl.	99.71	99.74
BUSCO sing.	92.3%	92.1%
BUSCO dupl.	1.5%	1.2%
BUSCO frag.	1.8%	1.8%
BUSCO miss.	4.4%	4.9%

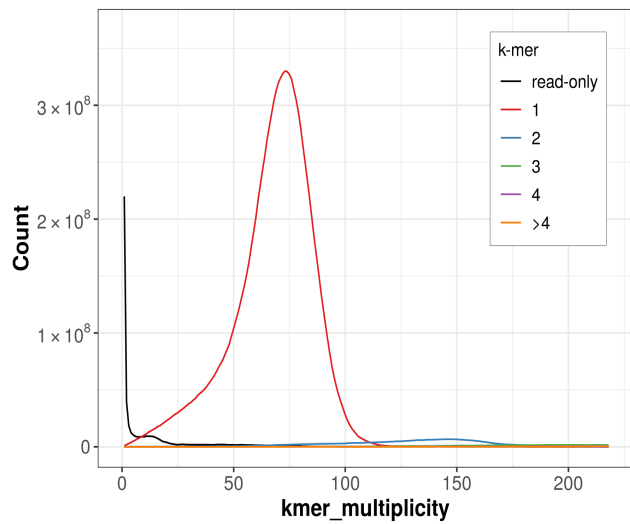
BUSCO 6.0.0 Lineage: metazoa\_odb10 (genomes:65, BUSCOs:954)

# HiC contact map of curated assembly

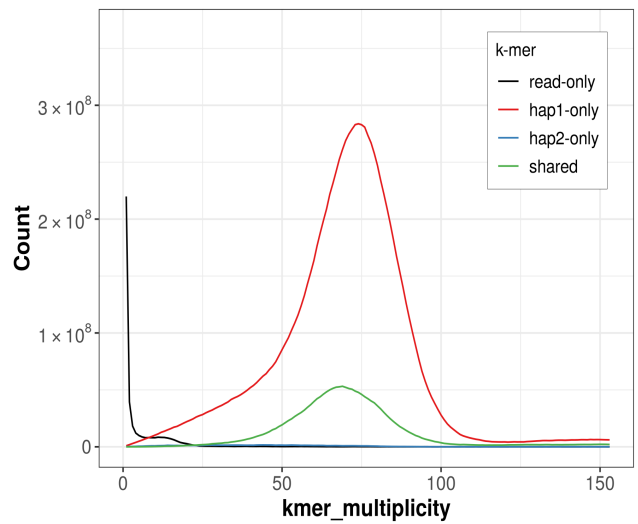


pri [\[LINK\]](#)

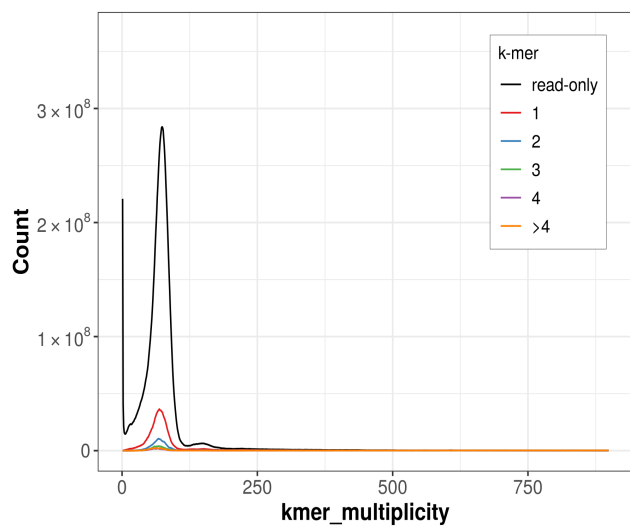
# K-mer spectra of curated assembly



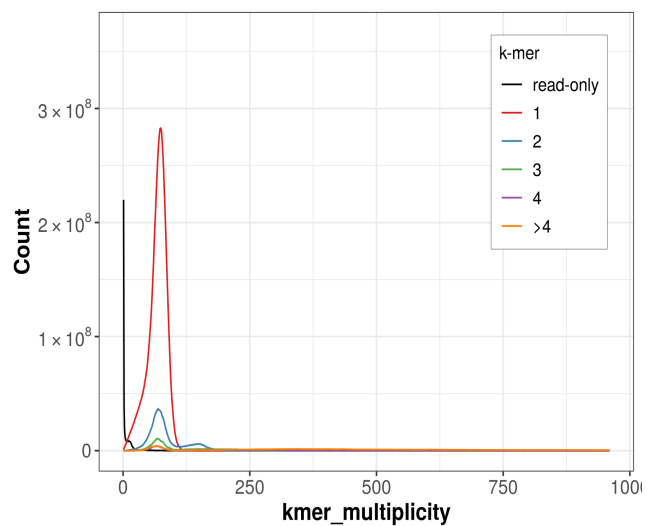
Distribution of k-mer counts per copy numbers found in **hap1** (hap1.)



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

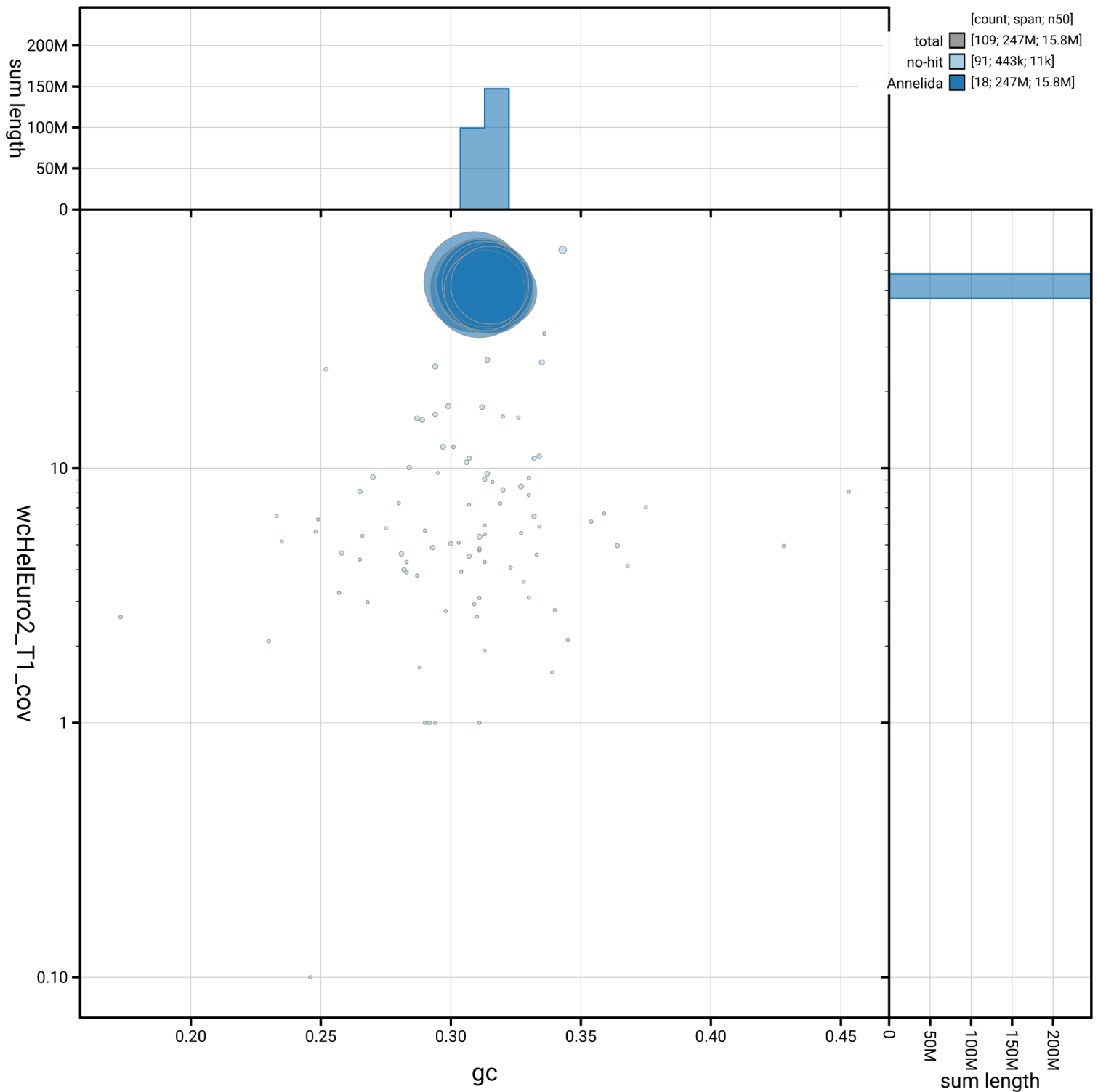


Distribution of k-mer counts per copy numbers found in **hap2** (hap1.)



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

# 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	51x	652x

# Assembly pipeline

- **hifiasm**
  - |\_ *ver*: 0.19.8-r603
  - |\_ *key param*: --primary
- **yahs**
  - |\_ *ver*: 1.2.2
  - |\_ *key param*: NA

# Curation pipeline

- **hifiasm**
  - |\_ *ver*: 0.19.8-r603
  - |\_ *key param*: --primary
- **yahs**
  - |\_ *ver*: 1.2.2
  - |\_ *key param*: NA
- **TreeVal**
  - |\_ *ver*: 1.2.1
  - |\_ *key param*: NA

Submitter: Michael Paulini

Affiliation: WSI

Date and time: 2025-08-26 15:41:55 CEST