

# ERGA Assembly Report

v24.04.03\_beta

Tags: ERGA-BGE

TxID	311461
ToLID	<b>wcHirVerb1</b>
Species	Hirudo verbana
Class	Clitellata
Order	Hirudinida

Genome Traits	Expected	Observed
Haploid size (bp)	198,057,365	177,750,590
Haploid Number	13 (source: ancestor)	14
Ploidy	2 (source: ancestor)	2
Sample Sex	H	H

## EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.7.Q55

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

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

### Curator notes

- . Interventions/Gb: 237
- . Contamination notes: "Contamination report for assembly labelled hap1; Total length of scaffolds removed: 31,600,116 (14.6 %); Scaffolds removed: 112 (31.2 %); Largest scaffold removed: (3,411,796); FCS-GX contaminant species (number of scaffolds; total length of scaffolds):Bdellovibrio sp. UBA4095, proteobacteria (12; 614,916); Myxococcus sp., d-proteobacteria (8; 289,269); Archangium sp., d-proteobacteria (5; 3,328,070); Pedobacter planticola, CFB group bacteria (5; 3,603,178); Pedobacter frigiditerrae, CFB group bacteria (4; 779,107); Verminephrobacter aporrectodeae, b-proteobacteria (3; 5,446,586); Xenophilus azovorans, b-proteobacteria (2; 2,934,109); Hydrogenophaga crocea, b-proteobacteria (2; 4,125,851); Hydrogenophaga flava, b-proteobacteria (2; 272,515); Bdellovibrio bacteriovorus, proteobacteria (2; 173,700); Pedobacter nanyangensis, CFB group bacteria (2; 133,137); Acidovorax citrulli, b-proteobacteria (2; 97,107); Aminobacter carboxidus, a-proteobacteria (2; 83,352); Myxococcus virescens, d-proteobacteria (2; 75,242); Giesbergeria sp., b-proteobacteria (2; 71,504); Pedobacter nanyangjuensis, CFB group bacteria (2;

36,528); *Nitrosomonas* sp., b-proteobacteria (2; 60,276); *Niabella ginsengisoli*, CFB group bacteria (2; 32,242); *Aminobacter aganoensis*, a-proteobacteria (2; 46,536); *Rikenella microfus*, CFB group bacteria (1; 3,411,796); *Comamonas granuli*, b-proteobacteria (1; 2,105,643); *Rhodoferax bucti*, b-proteobacteria (1; 1,487,571); *Pedobacter polaris*, CFB group bacteria (1; 243,386); *Rhodoferax lacus*, b-proteobacteria (1; 179,524); *Pedobacter chitinilyticus*, CFB group bacteria (1; 157,701); *Bacteroides neonati*, CFB group bacteria (1; 149,537); *Rhodospirillum rubrum*, a-proteobacteria (1; 133,336); *Stigmatella erecta*, d-proteobacteria (1; 118,242); *Pseudobdellovibrio exovorus*, proteobacteria (1; 77,394); *Polaromonas naphthalenivorans*, b-proteobacteria (1; 77,280); *Candidatus Methylophosphatis roskildensis*, b-proteobacteria (1; 72,446); *Comamonas aquatica*, b-proteobacteria (1; 66,664); *Corallococcus soli*, d-proteobacteria (1; 65,410); *Xanthobacter oligotrophicus*, a-proteobacteria (1; 58,146); *Anaeromyxobacter paludicola*, d-proteobacteria (1; 55,828); *Aminobacter ciceronei*, a-proteobacteria (1; 49,286); *Paludibacter* sp. 47-17, CFB group bacteria (1; 46,723); *Piscinibacter* sp., b-proteobacteria (1; 41,004); *Archangium violaceum*, d-proteobacteria (1; 39,712); *Candidatus Dechloromonas phosphorivorans*, b-proteobacteria (1; 39,057); *Diaphorobacter caeni*, b-proteobacteria (1; 39,008); *Myxococcus dinghuensis*, d-proteobacteria (1; 37,771); *Sphingobium jiangsuense*, a-proteobacteria (1; 37,293); *Brucella pseudogrignonensis*, a-proteobacteria (1; 36,081); *Niabella hibiscisoli*, CFB group bacteria (1; 33,595); *Ralstonia solanacearum*, b-proteobacteria (1; 32,147); *Bacteroides uniformis*, CFB group bacteria (1; 32,020); *Ottowia thiooxydans*, b-proteobacteria (1; 29,486); *Tepidicella xavieri*, b-proteobacteria (1; 26,292); *Leptothrix cholodnii*, b-proteobacteria (1; 25,479); *Candidatus Accumulibacter* sp. UBA704, b-proteobacteria (1; 24,995); *Hyalangium minutum*, d-proteobacteria (1; 21,851); *Brucella anthropi*, a-proteobacteria (1; 21,521); *Pseudodonghicola xiamenensis*, a-proteobacteria (1; 16,901); *Crenobacter luteus*, b-proteobacteria (1; 13,877); *Bacteroides* sp. UBA939, CFB group bacteria (1; 11,107); *Vitreoscilla* sp., b-proteobacteria (1; 10,000); *Tepidicella baoligensis*, b-proteobacteria (1; 8,909); *Gemmatirosa kalamazoonensis*, bacteria (1; 3,000); *Corallococcus macrosporus*, d-proteobacteria (1; 2,000); *Anaeromyxobacter dehalogenans*, d-proteobacteria (1; 2,000); *Polaromonas glacialis*, b-proteobacteria (1; 1,000); *Mitochondrion* (6; 254,872)"

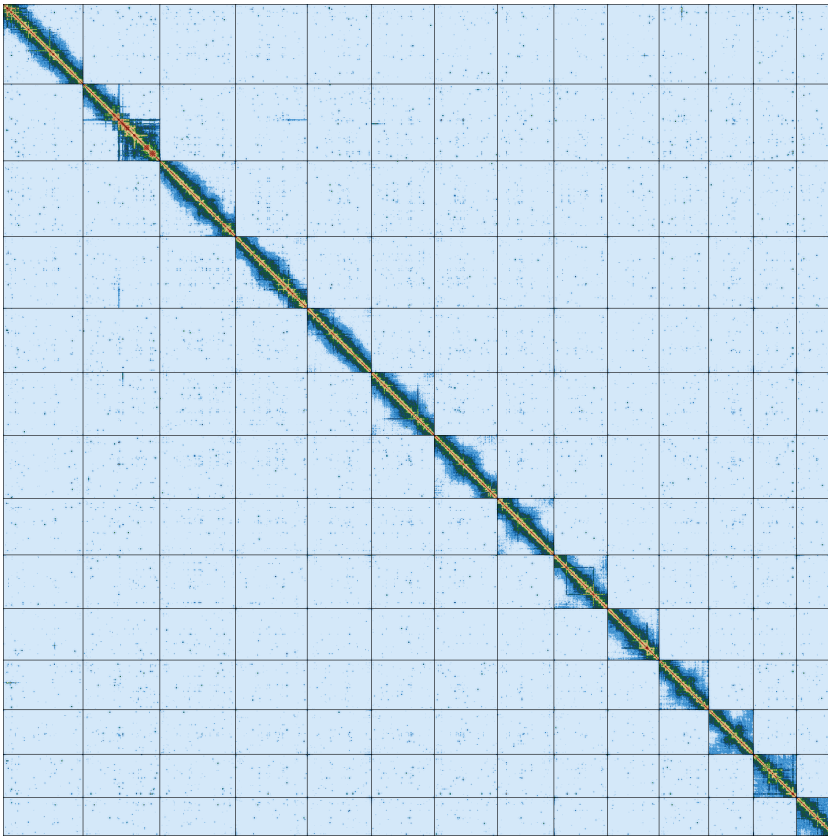
. Other observations: "This genome has been assembled using PacBio and HiC data and phased; We believe the drop in kmer completeness is due to the large amount of bacterial contamination that was removed;"

# Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	216,164,912	177,750,590
GC %	38.61	35.97
Gaps/Gbp	846.58	1,406.47
Total gap bp	18,300	33,500
Scaffolds	359	27
Scaffold N50	11,399,698	13,422,010
Scaffold L50	8	6
Scaffold L90	22	12
Contigs	542	277
Contig N50	1,402,608	1,338,446
Contig L50	42	36
Contig L90	161	132
QV	55.2	55.1
Kmer compl.	99.65	87.29
BUSCO sing.	79.7%	80.7%
BUSCO dupl.	5.7%	4.1%
BUSCO frag.	7.8%	7.7%
BUSCO miss.	6.8%	7.5%

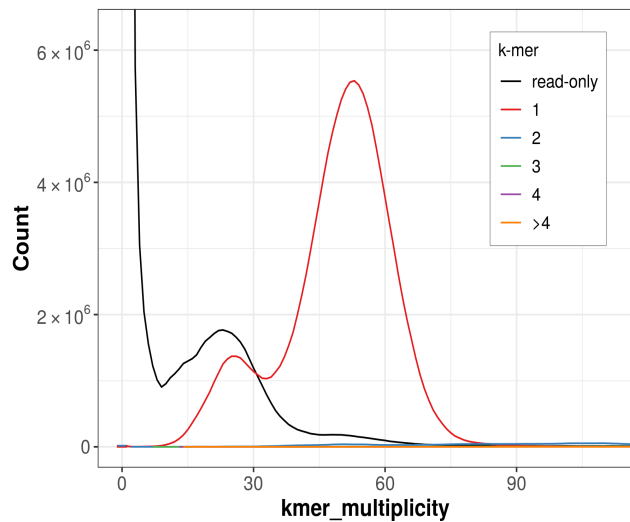
BUSCO 5.3.0 Lineage: metazoa\_odb10 (genomes:65, BUSCOs:954)

# HiC contact map of curated assembly

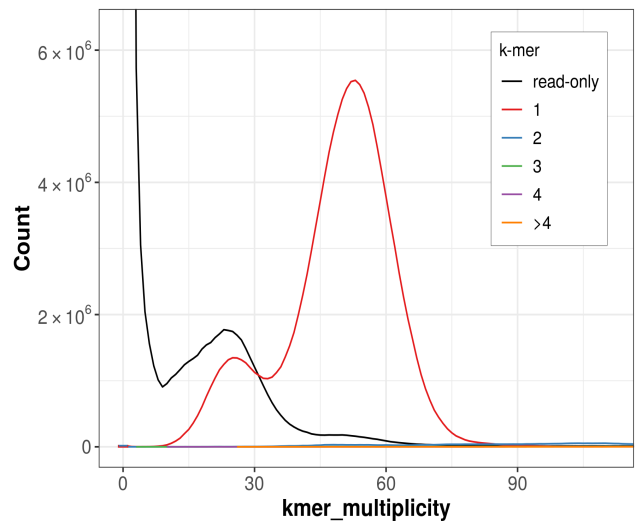


pri [\[LINK\]](#)

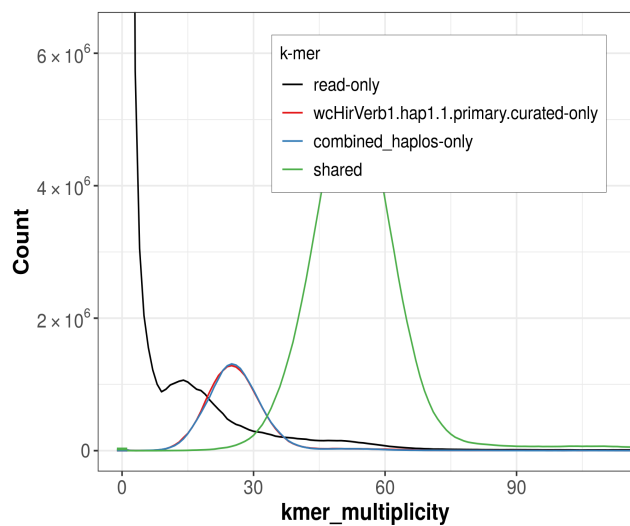
# K-mer spectra of curated assembly



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

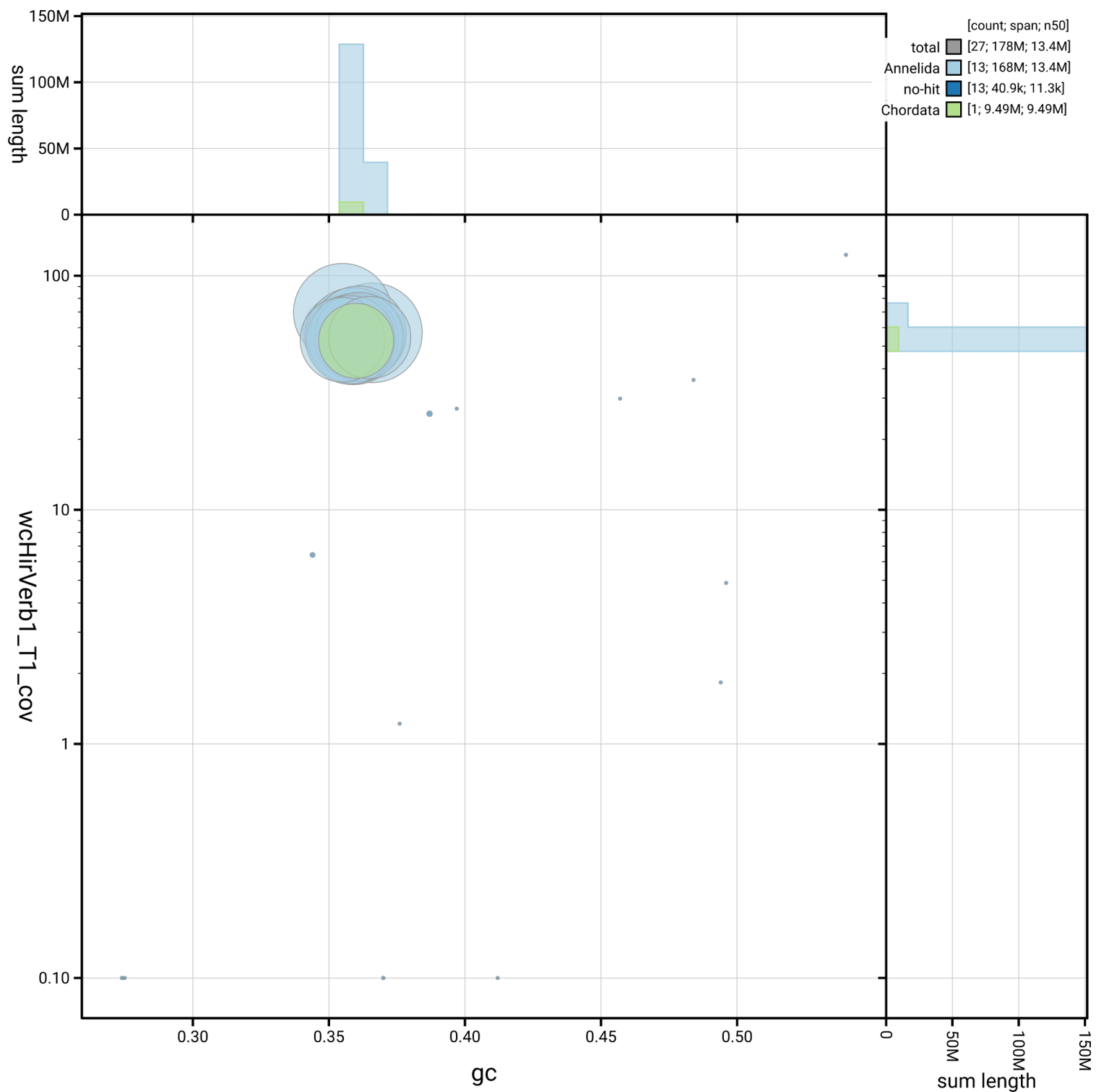


Distribution of k-mer counts per copy numbers found in asm



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

# 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	49x	536x

## Assembly pipeline

- **hifiasm**
  - |\_ *ver*: 0.19.8-r603
  - |\_ *key param*: --h1/--h2
- **yahs**
  - |\_ *ver*: 1.2.2
  - |\_ *key param*: NA

## Curation pipeline

- **hifiasm**
  - |\_ *ver*: 0.19.8-r603
  - |\_ *key param*: --h1/--h2
- **yahs**
  - |\_ *ver*: 1.2.2
  - |\_ *key param*: NA
- **TreeVal**
  - |\_ *ver*: 1.2.1
  - |\_ *key param*: NA

Submitter: Michael Paulini

Affiliation: WSI

Date and time: 2025-02-12 11:29:12 CET