

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

TxID	647718
ToLID	whApomoll1
Species	<i>Allolobophora molleri</i>
Class	Clitellata
Order	Crassiclitellata

Genome Traits	Expected	Observed
Haploid size (bp)	1,378,169,206	1,374,945,028
Haploid Number	16 (source: ancestor)	16
Ploidy	2 (source: ancestor)	4
Sample Sex	H	H

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 5.7.Q55

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

- . Observed Ploidy 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: 1627
- . Contamination notes: "Contamination report for assembly: Total length of scaffolds removed: 6,073,400 (0.2 %); Scaffolds removed: 79 (3.0 %); Largest scaffold removed: (944,258); FCS-GX contaminant species (number of scaffolds; total length of scaffolds):: Ohtaekwangia sp., CFB group bacteria (10; 1,567,218); Chryseotalea sp. WA131a, CFB group bacteria (9; 824,648); Chryseolinea soli, CFB group bacteria (8; 700,420); Parachryseolinea silvisoli, CFB group bacteria (4; 509,091); Chryseolinea serpens, CFB group bacteria (4; 158,193); Verminephrobacter aporrectodeae, b-proteobacteria (3; 606,208); Verminephrobacter eiseniae, b-proteobacteria (3; 264,809); Chryseosolibacter histidini, CFB group bacteria (3; 232,020); Ohtaekwangia koreensis, CFB group bacteria (3; 89,069); Ginsengibacter hankyongi, CFB group bacteria (2; 149,731); Hydrotalea sp. AMD, CFB group bacteria (2; 89,940); Chitinophaga niabensis, CFB group bacteria (2; 46,774); Flavobacterium cutihirudinis, CFB group bacteria (1; 162,664); Pecten maximus, molluscs (1; 71,752); Sediminibacterium salmoneum, CFB group bacteria (1; 59,000); Dawidia cretensis, CFB

group bacteria (1; 57,000); Chitinophaga rhizosphaerae, CFB group bacteria (1; 56,048); Mucilaginibacter sp., CFB group bacteria (1; 46,382); Sediminibacterium soli, CFB group bacteria (1; 39,177); Candidatus Vicinibacter affinis, CFB group bacteria (1; 35,999); Mucilaginibacter rigui, CFB group bacteria (1; 34,774); Solitalea koreensis, CFB group bacteria (1; 33,272); Sediminicola luteus, CFB group bacteria (1; 22,441); Chitinophaga sp., CFB group bacteria (1; 20,160); Haliscomenobacter hydrossis, CFB group bacteria (1; 14,028); Flavisolibacter ginsengisoli, CFB group bacteria (1; 9,752); Niabella sp., CFB group bacteria (1; 7,000); Paraflavitalea devenefica, CFB group bacteria (1; 1,000); Foetidibacter luteolus, CFB group bacteria (1; 14,740); Dreissena polymorpha, molluscs (1; 14,383); Niastella vici, CFB group bacteria (1; 11,747); Segetibacter aerophilus, CFB group bacteria (1; 10,115); Segetibacter koreensis, CFB group bacteria (1; 7,200); Mitochondrion (5; 106,645)"

. Other observations: "Getting an assembly that was curatable was troublesome for this genome. In the end we have used an experimental tool in development that uses FastGA based alignments to look for syntenic units across the 4 haplotypes. This gave a version of the assembly that was curatable, although many phasing errors remained to be dealt with during curation. The sample is tetraploid (2n=32) and seems to be highly polymorphic as many rearrangements could be detected during curation.

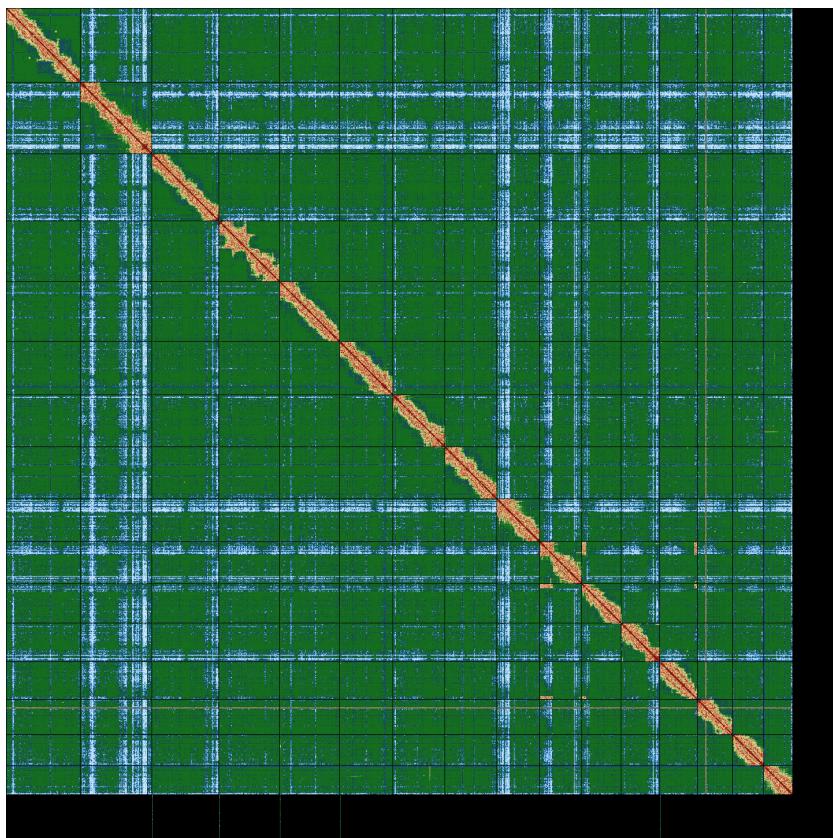
Haplotypic inversions were detected on: Chromosome 1 from 33.40 to 84.84 Mbp and from 50.98 to 84.84 Mbp; Chromosome 2 from 22.18 to 45.74 Mbp; Chromosome 6 from 80.33 Mbp to the end; Chromosome 9 from 7.65 to 28.41 Mbp; Chromosome 10 from 545 Kbp to 3.67 Mbp; Chromosome 12 from 55.32 to 61.14 Mbp; Chromosome 14 from 31.51 to 55.14 Mbp; The HAP1 file has the curated chromosome-level assembly and the other 3 haplotypes are in HAP2 file as scaffold-level assemblies"

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	2,546,329,153	1,374,945,028
GC %	41.01	40.93
Gaps/Gbp	2,044.51	1,743.34
Total gap bp	520,600	313,000
Scaffolds	2,995	2,120
Scaffold N50	74,114,021	86,106,293
Scaffold L50	13	7
Scaffold L90	37	15
Contigs	8,201	4,517
Contig N50	869,876	937,760
Contig L50	743	380
Contig L90	3,537	1,644
QV	54.9	55.0
Kmer compl.	98.83	98.10
BUSCO sing.	27.5%	79.6%
BUSCO dupl.	69.0%	6.5%
BUSCO frag.	0.4%	0.8%
BUSCO miss.	3.1%	13.1%

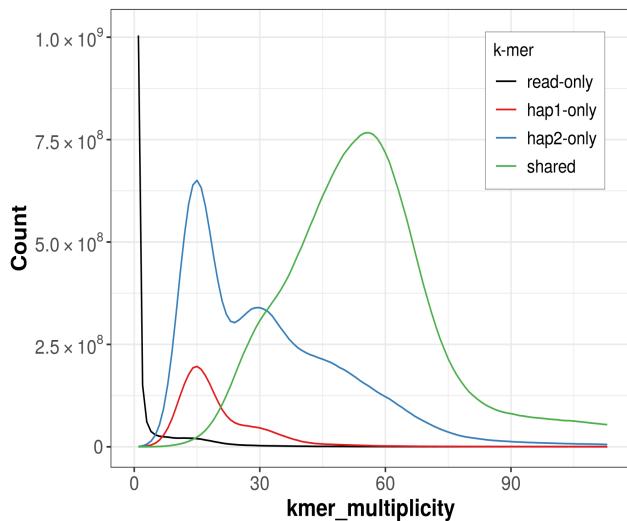
BUSCO 6.0.0 Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

HiC contact map of curated assembly

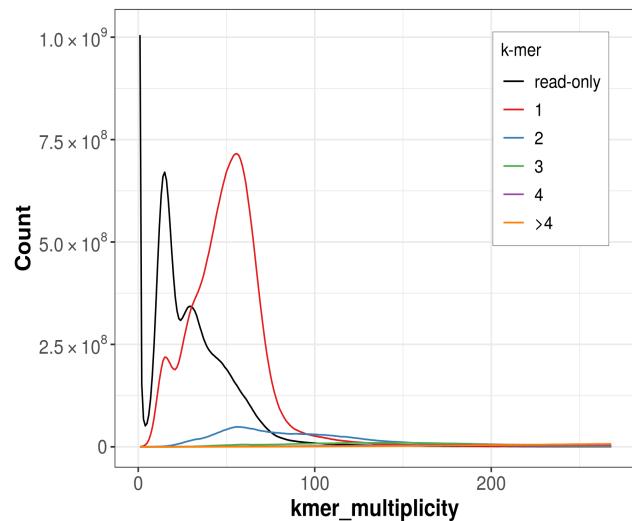


pri [LINK]

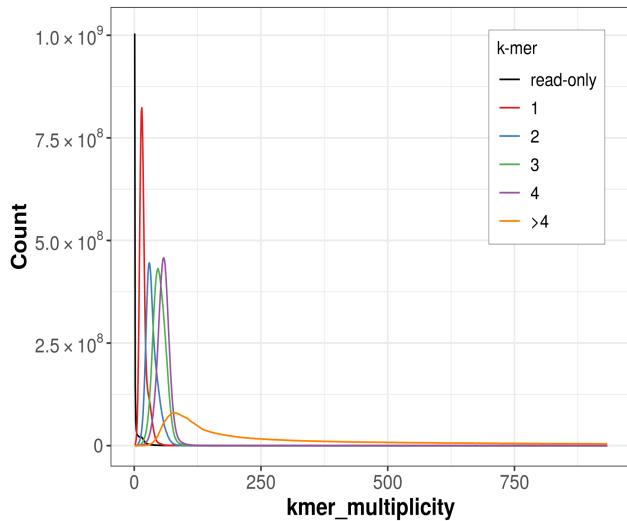
K-mer spectra of curated assembly



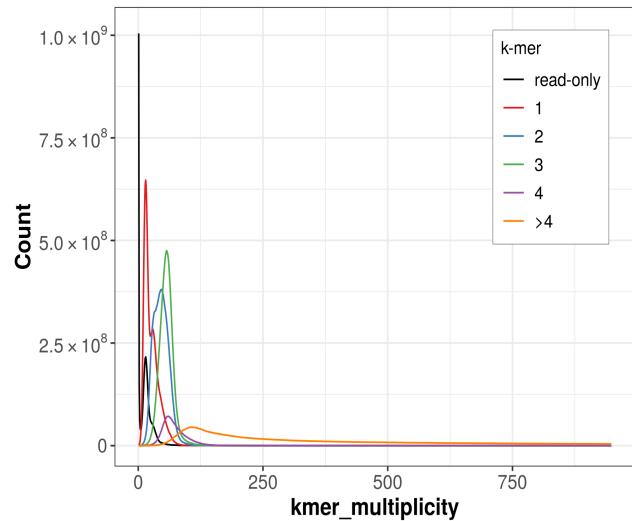
Distribution of k-mer counts coloured by their presence in reads/assemblies



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

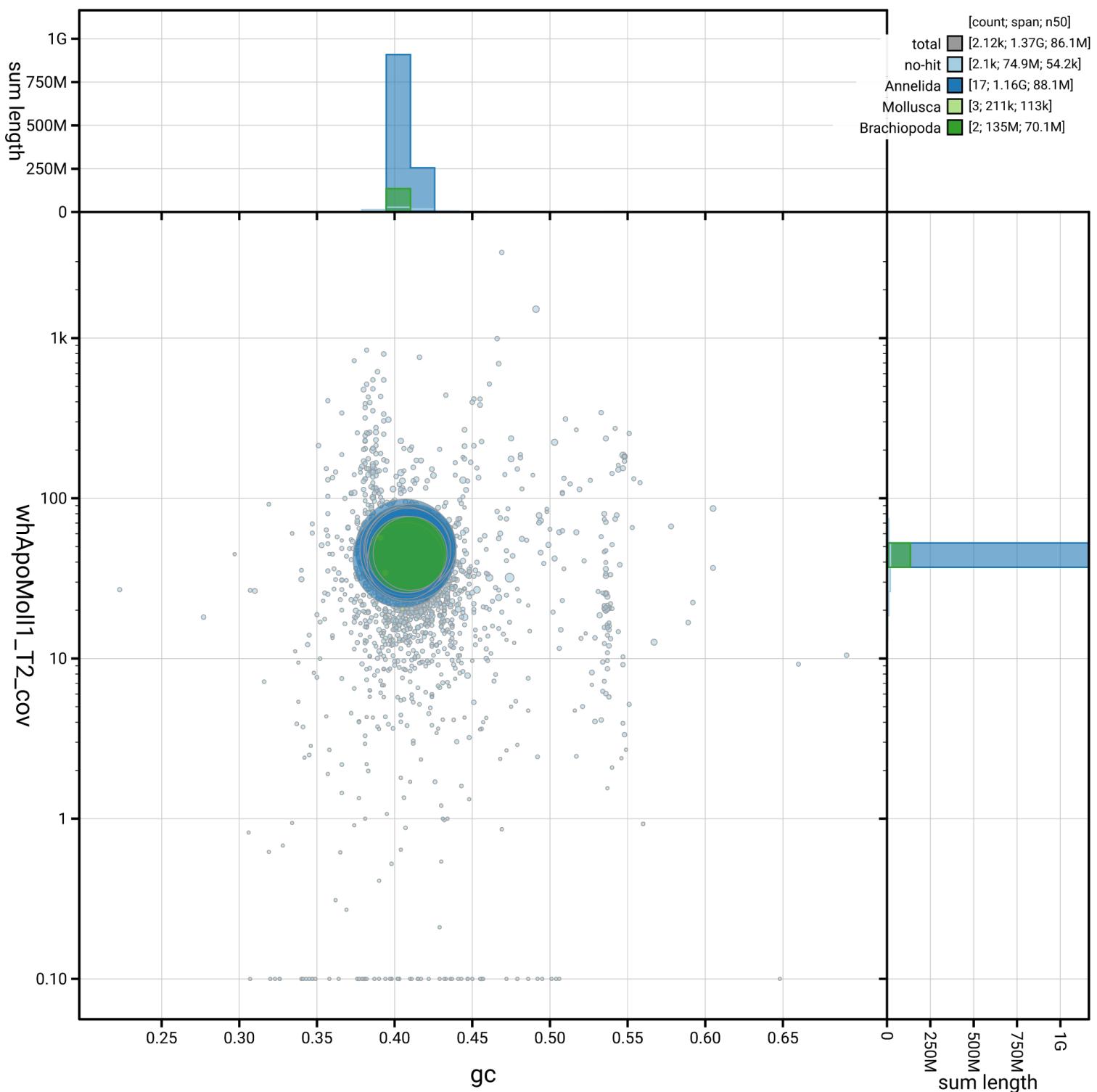


Distribution of k-mer counts per copy numbers found in **asm** (dippl.)



Distribution of k-mer counts per copy numbers found in **hap2** (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	60x	80x

Assembly pipeline

```
- hifiasm
  |_ ver: 0.19.8-r603
  |_ key param: --h1/--h2
- FastGA synteny based experimental phasing for polyploids
  |_ ver: NA
  |_ key param: NA
- yahs
  |_ ver: 1.2.2
  |_ key param: NA
```

Curation pipeline

```
- hifiasm
  |_ ver: 0.19.8-r603
  |_ key param: --h1/--h2/--hom-cov 58
- FastGA synteny based experimental phasing for polyploids
  |_ ver: NA
  |_ key param: NA
- yahs
  |_ ver: 1.2.2
  |_ key param: NA
- TreeVal
  |_ ver: 1.2.2
  |_ key param: NA
```

Submitter: Camilla Santos

Affiliation: WSI

Date and time: 2025-11-21 16:08:32 CET