

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	1046305
ToLID	whAilLact1
Species	Ailoscolex lacteospumousus
Class	Clitellata
Order	Crassiclitellata

Genome Traits	Expected	Observed
Haploid size (bp)	479,185,312	493,041,447
Haploid Number	18 (source: ancestor)	17
Ploidy	2 (source: ancestor)	4
Sample Sex	XX	unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 7.7.Q63

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

- . Observed Haploid Number is different from Expected
- . Observed Ploidy is different from Expected
- . Observed sex is different from Sample sex
- . Kmer completeness value is less than 90 for pri
- . BUSCO single copy value is less than 90% for pri

Curator notes

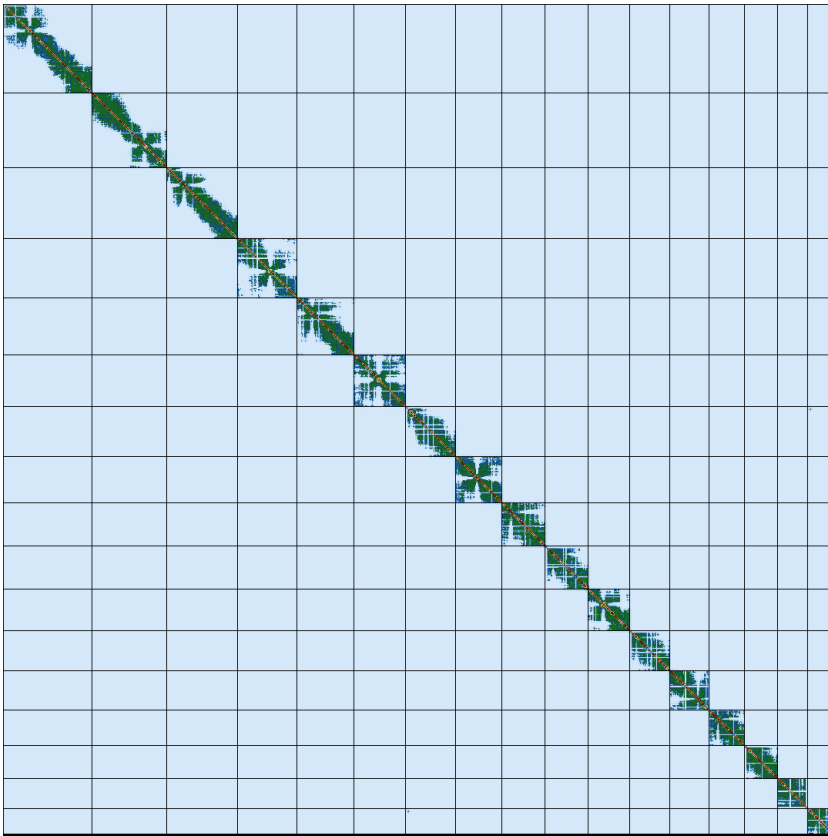
- . Interventions/Gb: 4
- . Contamination notes: "No true contaminants found. All scaffolds had a mix of annelid, mollusk and brachiopod hits. All were considered to be derived from annelid."
- . Other observations: "Assembly with Nextdenovo and Flye did not work very well. Hifiiasm assembly with default settings resulted in two assemblies of approx. 1 Mb each. Purging resulted in four assemblies of 450-480 Mb each. We curated the two purged assemblies and results were satisfactory, but a considerable amount of manual interventions were necessary. We tried again using the --n-hap 4 setting for internal purging in hifiiasm resulting in a primary assembly of the appropriate length, few duplicate buscos and best contiguity of any assembly. We curated this assembly directly. We cut two contigs, removing two haplotigs. Two more were tagged in the shrapnel. One unloc was placed. No other edits were made."

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	497,772,681	493,041,447
GC %	42.52	42.48
Gaps/Gbp	44.2	44.62
Total gap bp	4,400	4,400
Scaffolds	27	24
Scaffold N50	29,876,410	29,876,410
Scaffold L50	7	7
Scaffold L90	15	15
Contigs	49	46
Contig N50	15,094,000	15,094,000
Contig L50	10	10
Contig L90	29	28
QV	63.8379	63.8293
Kmer compl.	57.7849	57.6119
BUSCO sing.	87.9%	88.5%
BUSCO dupl.	3.9%	3.4%
BUSCO frag.	4.4%	4.4%
BUSCO miss.	3.8%	3.7%

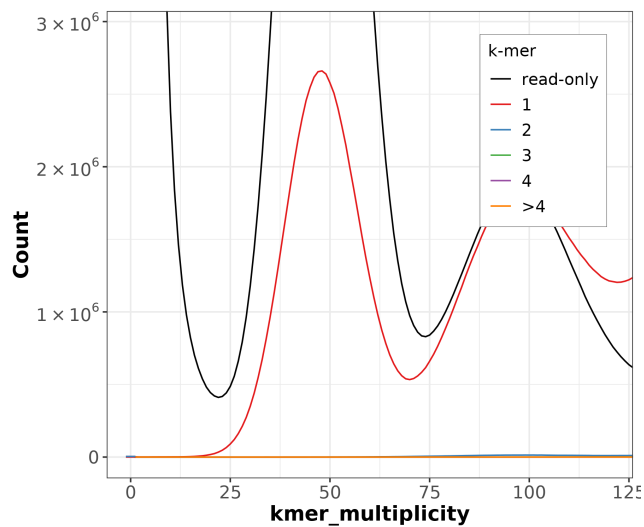
BUSCO: 5.4.0 (euk_genome_met, metaeuk) / Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

HiC contact map of curated assembly

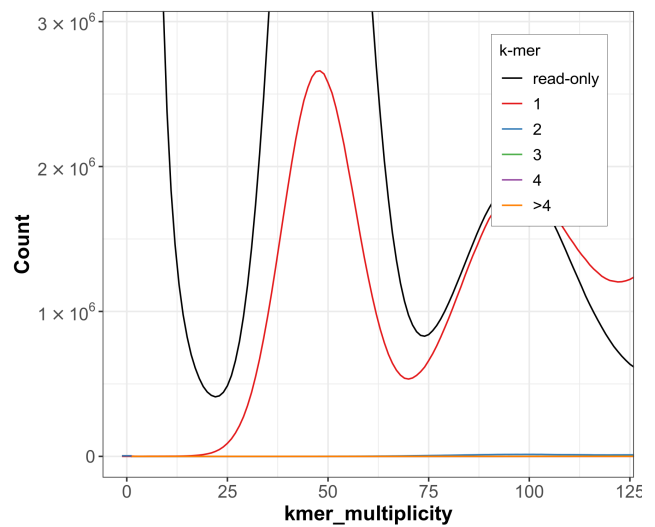


pri [\[LINK\]](#)

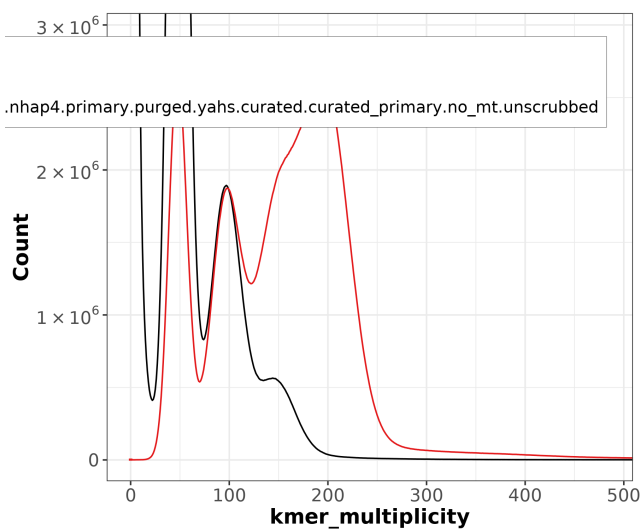
K-mer spectra of curated assembly



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

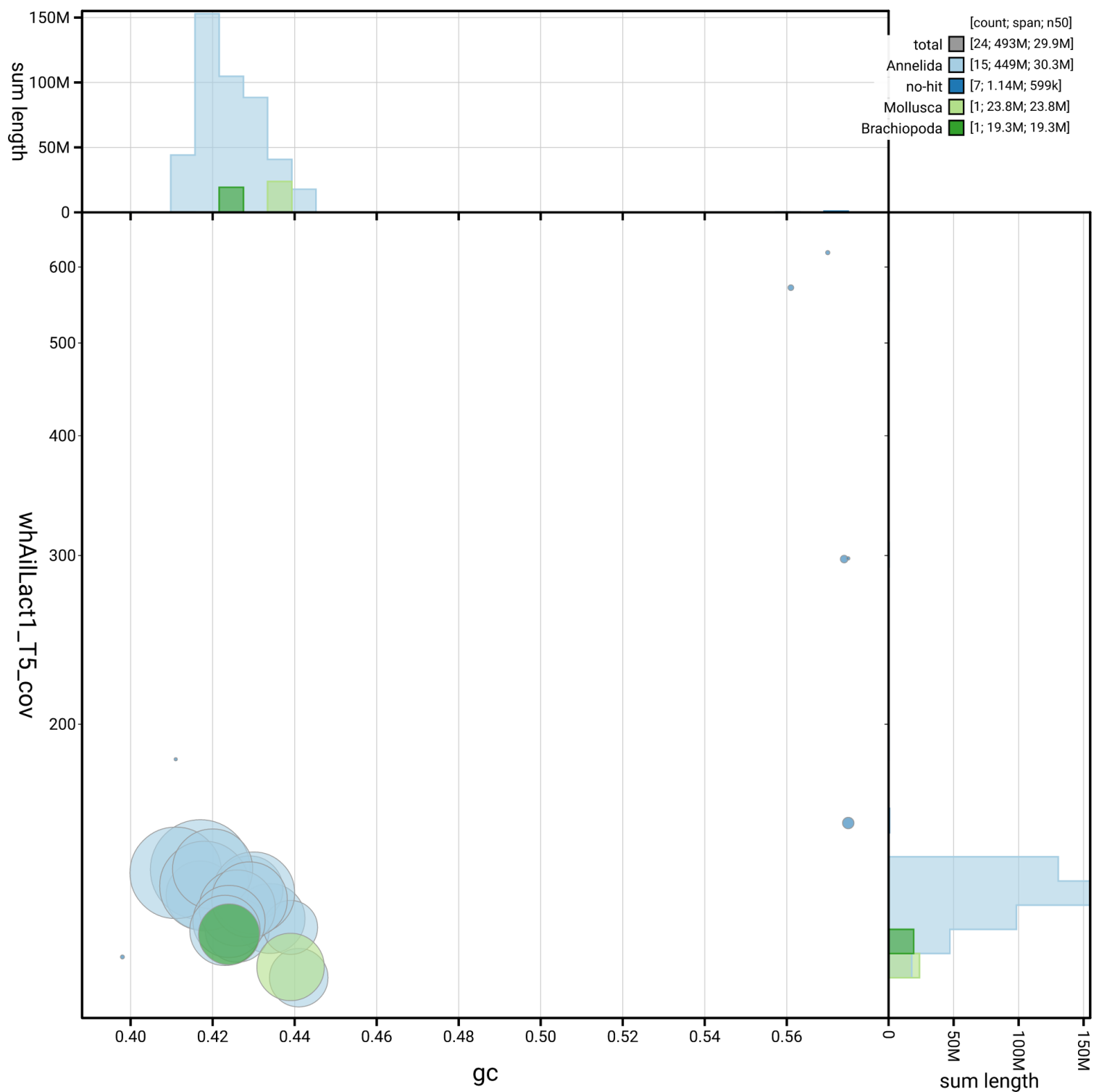


Distribution of k-mer counts per copy numbers found in asm



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	ONT
Coverage	124x

Assembly pipeline

- **Filtlong**
 - |_ *ver*: 0.2.1
 - |_ *key param*: --min_length 1000 --min_mean_q 80 -t 610000000000
- **Hifiasm**
 - |_ *ver*: 0.24.0
 - |_ *key param*: --n-hap 4 --ont
- **YaHS**
 - |_ *ver*: 1.2a
 - |_ *key param*: NA
- **CLAWS pipeline**
 - |_ *ver*: 2.2.0
 - |_ *key param*: NA
- **Blobtoolkit Nextflow pipeline (latest)**
 - |_ *ver*: 0.6
 - |_ *key param*: NA
- **FOAM pipeline**
 - |_ *ver*: 0.5
 - |_ *key param*: NA
- **mitos**
 - |_ *ver*: 2.1.3
 - |_ *key param*: NA

Curation pipeline

- **PretextView**
 - |_ *ver*: 0.2.5
 - |_ *key param*: NA
- **GRIT_Rapid**
 - |_ *ver*: 2.0
 - |_ *key param*: NA

Submitter: Tyler Alioto

Affiliation: CNAG

Date and time: 2025-02-17 01:16:20 CET