

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	56557
ToLID	heTriNodul
Species	Triaenophorus nodulosus
Class	Cestoda
Order	Bothriocephalidea

Genome Traits	Expected	Observed
Haploid size (bp)	677,466,258	713,084,067
Haploid Number	7 (source: ancestor)	13
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for hap1: 6.7.Q67

Obtained EBP quality metric for hap2: 6.7.Q67

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 hap1
- . BUSCO single copy value is less than 90% for hap1
- . Kmer completeness value is less than 90 for hap2
- . BUSCO single copy value is less than 90% for hap2
- . Assembly length loss > 3% for hap1
- . Assembly length loss > 3% for hap2

Curator notes

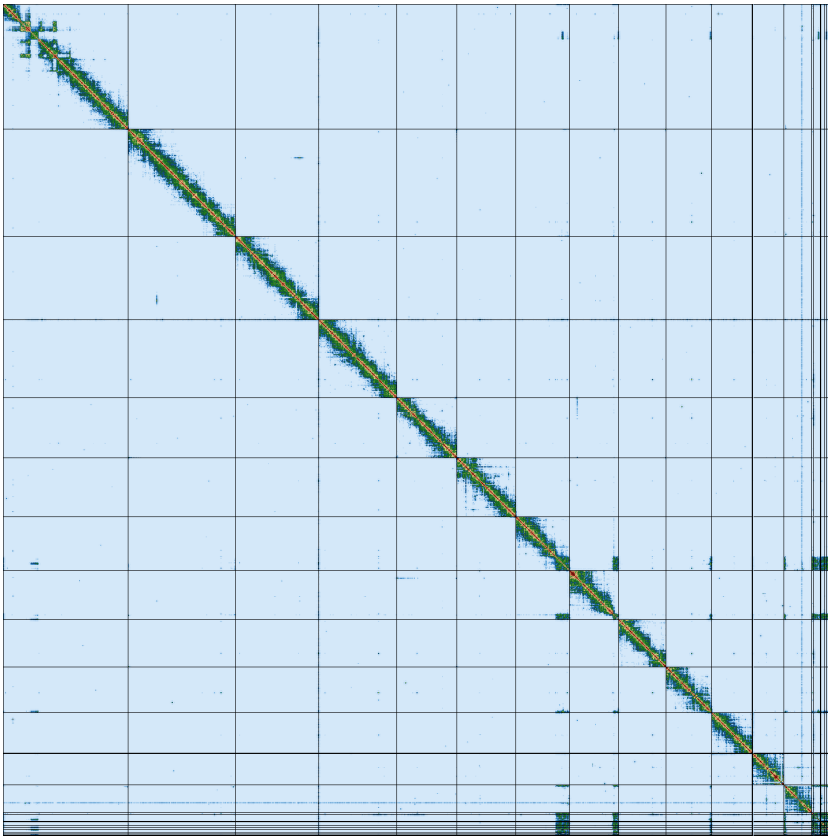
- . Interventions/Gb: 324
- . Contamination notes: "No contaminations were detected with FCS-GX.. Mitochondrial genome was removed from the assembly"
- . Other observations: "Hifiiasm was run in HiC-mode and created two haplotype assemblies (hap1 - contigs: 782, yield 843Mb, N50: 5.5Mb; hap2 - contigs: 532, yield 728Mb, N50: 4.9Mb). A manual contig filtering was done based on the cumulative evidences from: (1) MitoHifi - detect mitochondrial contigs (2) Tiara - deep learning to classify eukaryotic, bacterial, organelle, archaea sequences and (3) BlobtoolKit - detect contaminations. Purge_dups was run to anotate haplotypic duplications. Scaffolding was done with Yahs on both haplotypes. Multiple curation rounds were necessary. The first chromosome still looks a bit odd."

Quality metrics table

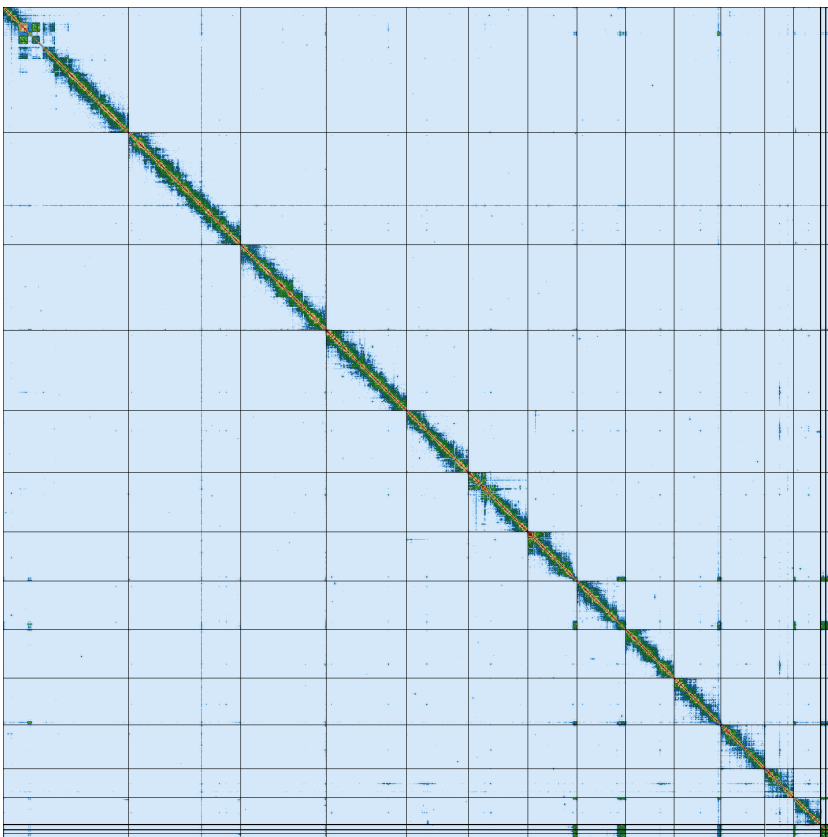
Metrics	Pre-curation hap1	Pre-curation hap2	Curated hap1	Curated hap2
Total bp	843,771,454	728,198,039	713,084,067	682,246,441
GC %	33.81	33.56	33.46	33.39
Gaps/Gbp	237.03	288.38	300.1	357.64
Total gap bp	20,000	21,000	27,100	30,800
Scaffolds	582	322	34	26
Scaffold N50	42,307,306	49,554,147	51,365,156	50,519,038
Scaffold L50	6	6	5	5
Scaffold L90	38	17	11	11
Contigs	782	532	248	270
Contig N50	5,500,272	4,973,056	6,972,067	4,997,201
Contig L50	40	31	31	29
Contig L90	184	164	127	140
QV	66.0261	66.5746	67.6372	67.425
Kmer compl.	88.422	86.4289	87.0801	86.0465
BUSCO sing.	65.5%	66.5%	66.1%	67.0%
BUSCO dupl.	2.1%	1.0%	1.2%	0.7%
BUSCO frag.	13.2%	13.1%	13.8%	13.5%
BUSCO miss.	19.2%	19.3%	18.9%	18.8%

BUSCO: 5.8.3 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

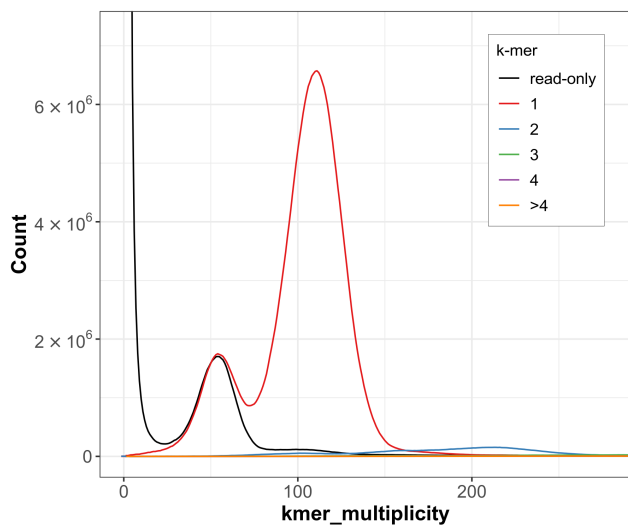


hap1 [\[LINK\]](#)

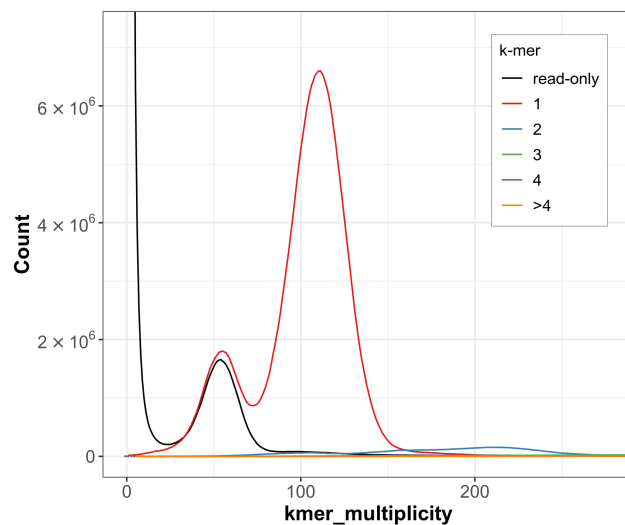


hap2 [\[LINK\]](#)

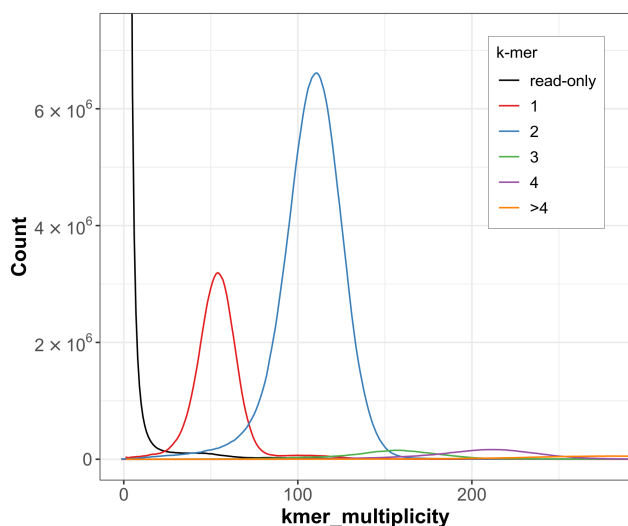
K-mer spectra of curated assembly



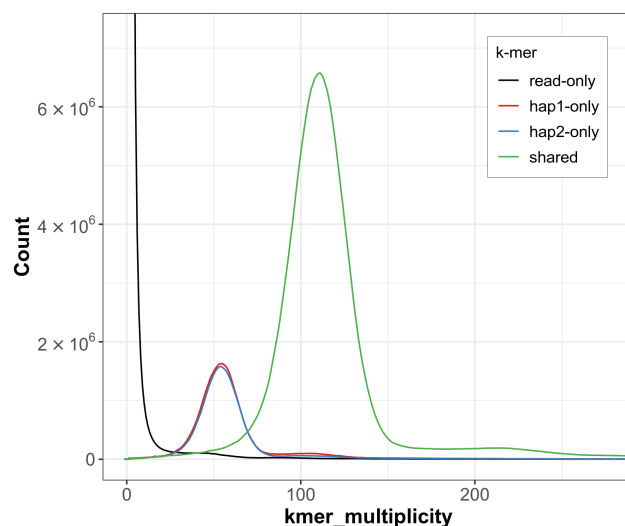
Distribution of k-mer counts per copy numbers found in **hap2** (hap1.)



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

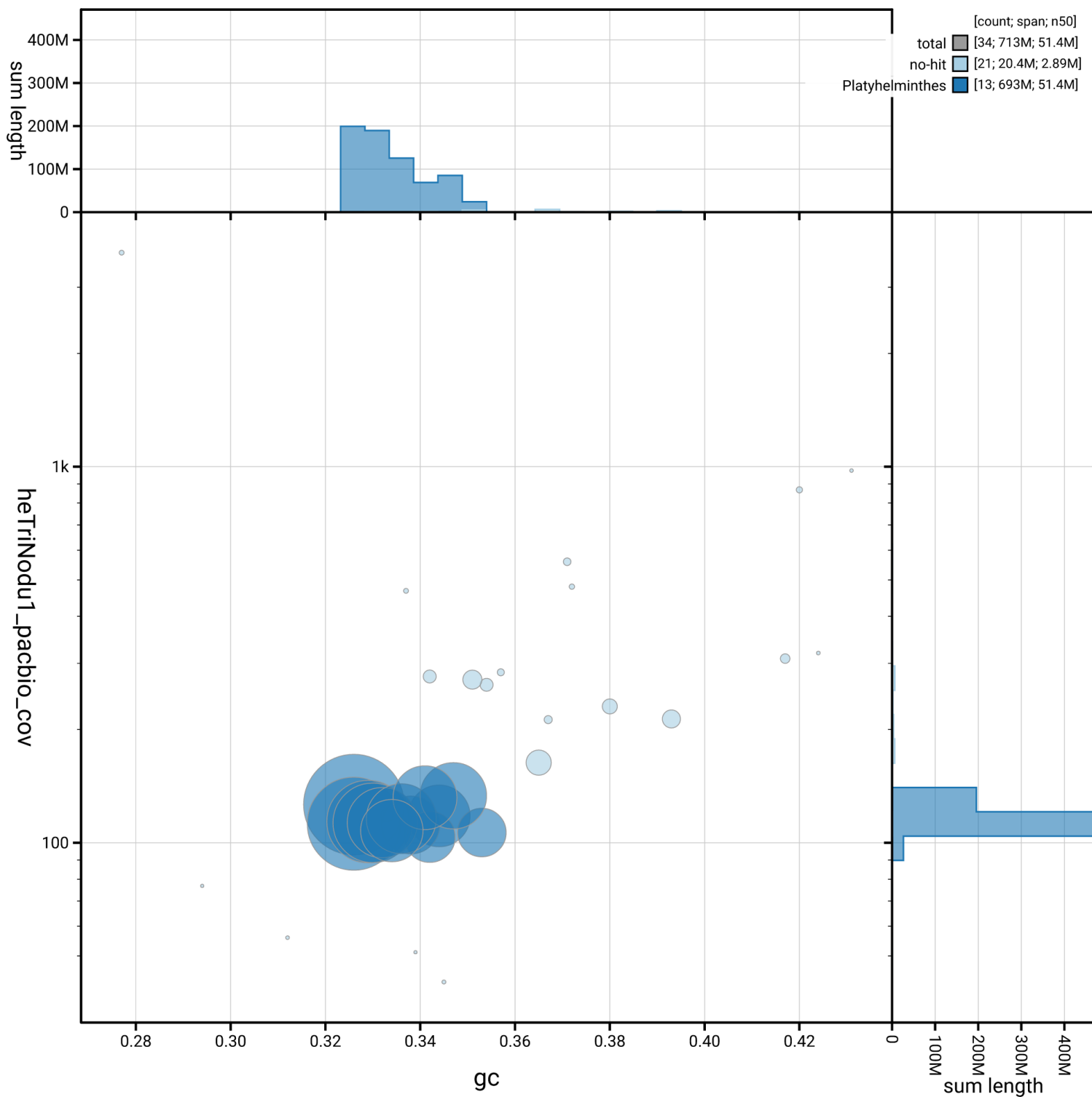


Distribution of k-mer counts per copy numbers found in **asm** (dipl.)

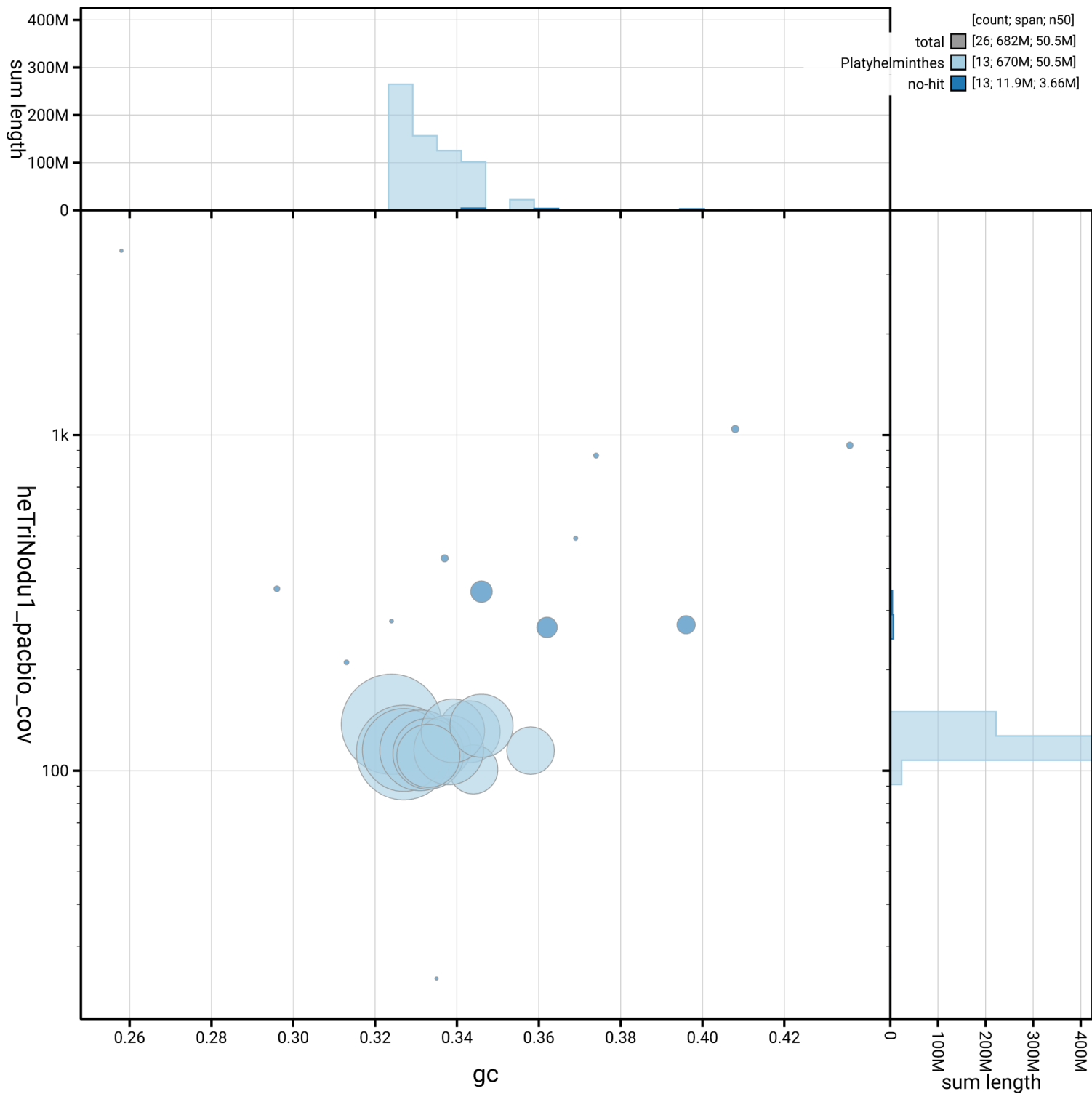


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



hap1. 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.



hap2. 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	HiFi	Arima3
Coverage	106x	35x

Assembly pipeline

- **Hifiasm**
 - |_ *ver*: 0.25.0-r726
 - |_ *key param*: HiC
 - |_ *key param*: 13
- **purge_dups**
 - |_ *ver*: 1.2.6
 - |_ *key param*: NA
- **YaHS**
 - |_ *ver*: 1.2.2
 - |_ *key param*: NA

Curation pipeline

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

Submitter: Martin Pippel

Affiliation: SciLifeLab

Date and time: 2025-11-27 17:48:16 CET