

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	947578
ToLID	tnRamLact8
Species	Lineus lacteus
Class	Pilidiophora
Order	Heteronemertea

Genome Traits	Expected	Observed
Haploid size (bp)	370,936,233	370,829,418
Haploid Number	16 (source: direct)	19
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q42

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 collapsed

Curator notes

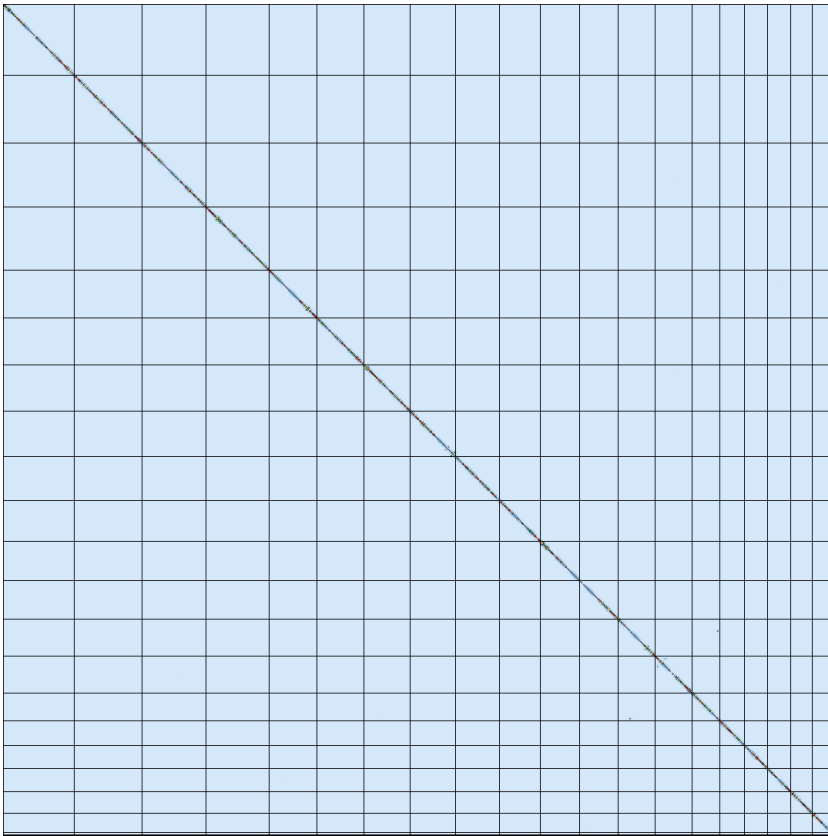
- . Interventions/Gb: 38
- . Contamination notes: "BTK pipeline detected just false positives with short hits to various metazoan phylum. This is expectable as Nemertea is an undersequenced phylum and the hits found are short and divergent."
- . Other observations: "The HiC data was not very good, we ended up having 7.3 million valid pairs for scaffolding and we dropped the mq to 10 with YaHS. However the scaffolding worked pretty well for such as small genome and we ended up with 19 super-scaffolds after curation and review. During review, we ran whole genome alignments against tnLinLong1.2 (n=19) to confirm the join of some scaffolds into a single SUPER (same chromosome). Note that SUPER_8 contains a true segmental duplication as confirmed by coverage, ONT and self-alignments. Curation and review involved 4 cuts in contigs, 3 breaks at gaps and 7 joins. The pretext contact map and final savestate (tnRamLact8.curated_primary.no_mt.unscrubbed_mq0.extensions.pretext.savestate_3) are located into the multimapping folder following the shared link. Regarding the plastids, we did not find reads to assemble the mitogenome properly."

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	371,087,788	370,829,418
GC %	41.34	41.35
Gaps/Gbp	107.79	118.65
Total gap bp	8,000	8,800
Scaffolds	31	27
Scaffold N50	20,380,758	20,380,758
Scaffold L50	8	8
Scaffold L90	17	16
Contigs	71	71
Contig N50	8,945,886	8,945,886
Contig L50	15	14
Contig L90	39	38
QV	42.9662	42.9629
Kmer compl.	77.4608	77.4551
BUSCO sing.	94.1%	94.1%
BUSCO dupl.	0.4%	0.4%
BUSCO frag.	2.7%	2.7%
BUSCO miss.	2.8%	2.8%

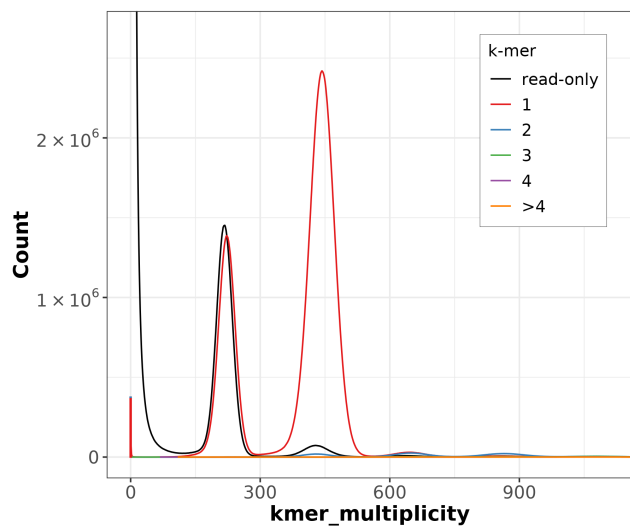
BUSCO: 5.4.0 (euk_genome_met, metaeuk) / Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

HiC contact map of curated assembly

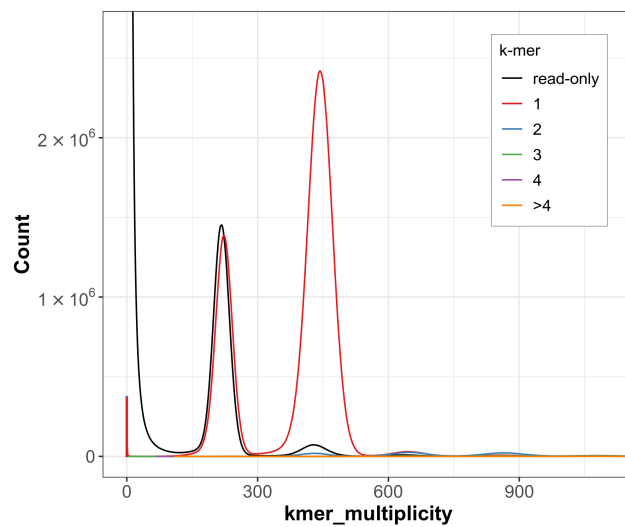


collapsed [\[LINK\]](#)

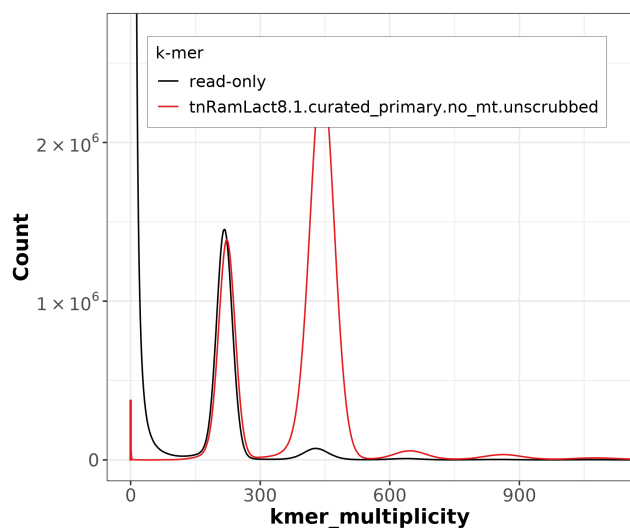
K-mer spectra of curated assembly



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

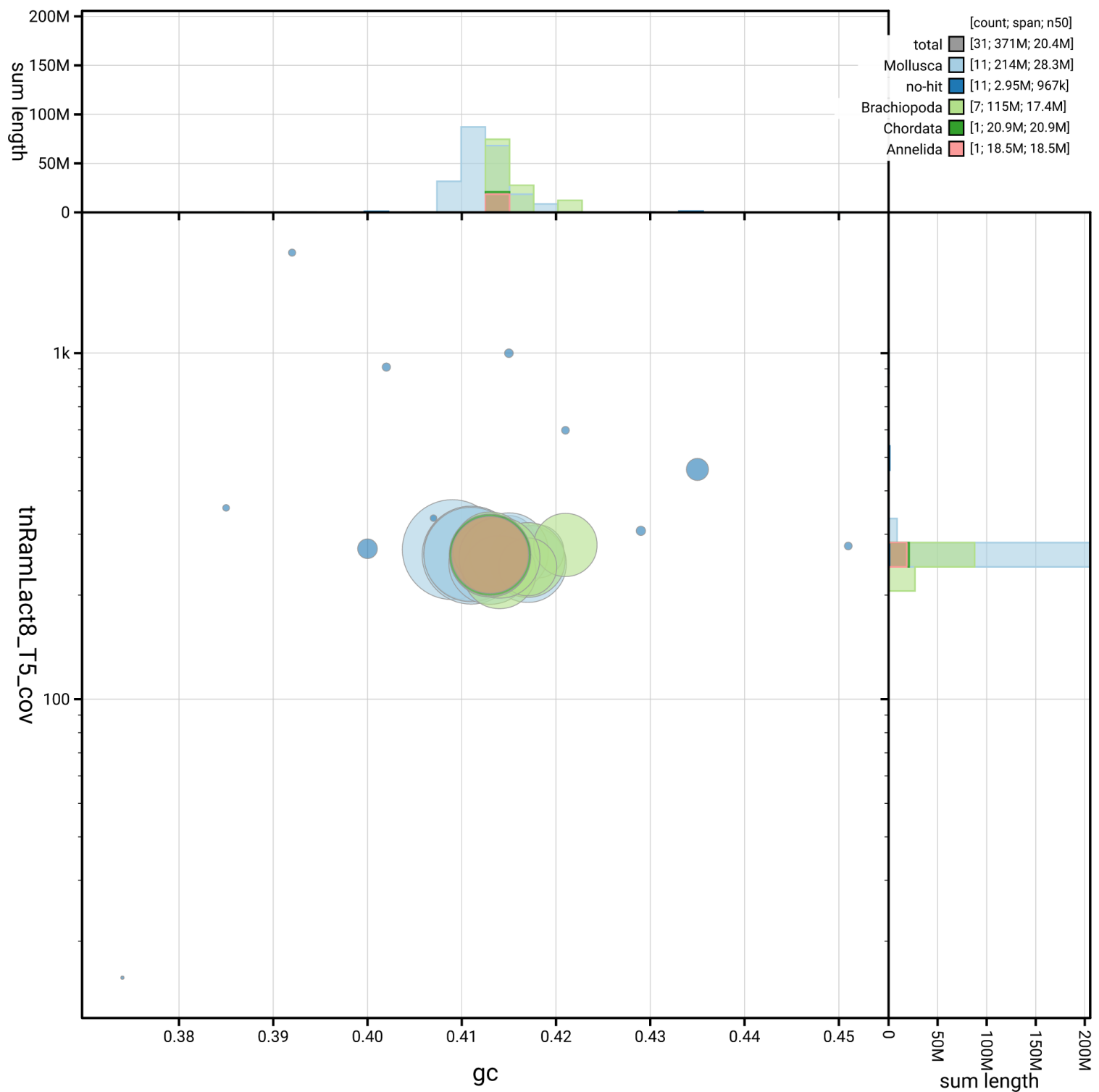


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



collapsed. 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	Illumina	OmniC
Coverage	339x	274x	64.6x

Assembly pipeline

- **CLAWS pipeline**
 - |_ *ver*: 2.2.0
 - |_ *key param*: NA
- **Trim_galore**
 - |_ *ver*: 0.6.7
 - |_ *key param*: NA
- **Filtlong**
 - |_ *ver*: 0.2.1
 - |_ *key param*: NA
- **flye**
 - |_ *ver*: 2.9.1
 - |_ *key param*: NA
- **nextdenovo**
 - |_ *ver*: 2.5.0
 - |_ *key param*: NA
- **hypo**
 - |_ *ver*: 1.0.3
 - |_ *key param*: NA
- **purge_dups**
 - |_ *ver*: 1.2.6
 - |_ *key param*: NA
- **YaHS**
 - |_ *ver*: 1.2a
 - |_ *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

Affiliation: CNAG

Date and time: 2025-02-14 11:17:46 CET