ERGA Assembly Report
v24.10.15

Tags: ERGA-BGE

TxID	1662264
ToLID	**qqLevThor3**
Species	Leviellus thorelli
Class	Arachnida
Order	Araneae

Genome Traits	Expected	Observed
Haploid size (bp)	2,048,663,629	2,190,932,037
Haploid Number	12 (source: ancestor)	13
Ploidy	2 (source: ancestor)	2
Sample Sex	female	female

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q51

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

. Observed Haploid Number is different from Expected

Curator notes

. Interventions/Gb: 26
. Contamination notes: "Scaffold_22 (1.5 MBases) was removed from the yahs scaffolded assembly because it was identitfied as pseudomonadota (which include symbionts of many arthropod species). The contaminant was removed from the yahs scaffolded assembly and is not present in the curated pretext which was shared."
. Other observations: "The HiC signal along the diagonal is quite strong. We were able to place several scaffolds to many of the super scaffolds but we found it very difficult to localize most of them. The sex of the specimen was set to -female- in our lims system. We could find the sex chromosome in the assembly"

Quality metrics table

| Metrics	Pre-curation collapsed	Curated collapsed
Total bp	2,192,434,344	2,190,932,037
GC %	34.31	34.31
Gaps/Gbp	356.68	367.42
Total gap bp	156,400	161,000
Scaffolds	149	130
Scaffold N50	162,679,147	167,117,025
Scaffold L50	6	7
Scaffold L90	12	12
Contigs	931	935
Contig N50	5,371,768	5,371,768
Contig L50	119	119
Contig L90	422	422
QV	51.5219	51.5224
Kmer compl.	95.3695	95.2763
BUSCO sing.	93.5%	93.8%
BUSCO dupl.	2.7%	2.4%
BUSCO frag.	1.0%	0.9%
BUSCO miss.	2.8%	2.9%

BUSCO: 5.4.0 (euk_genome_met, metaeuk) / Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)
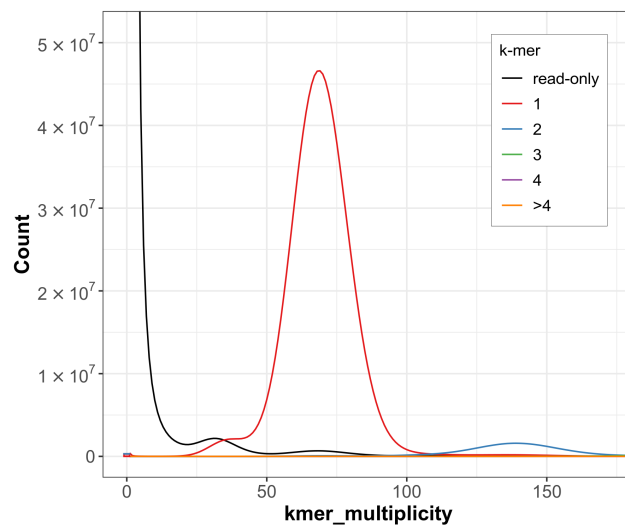
HiC contact map of curated assembly
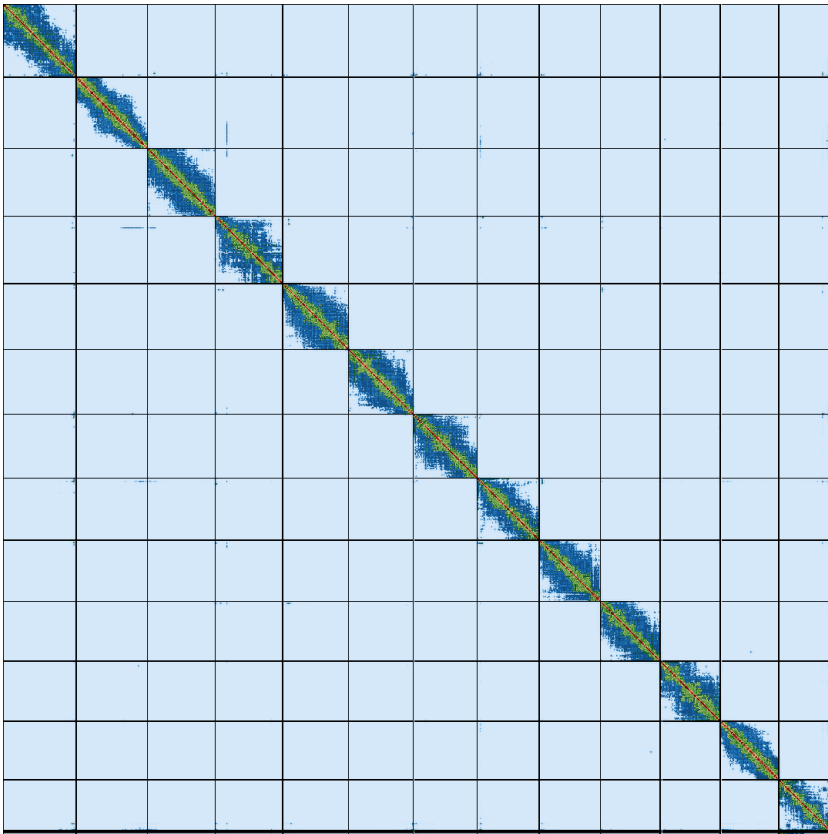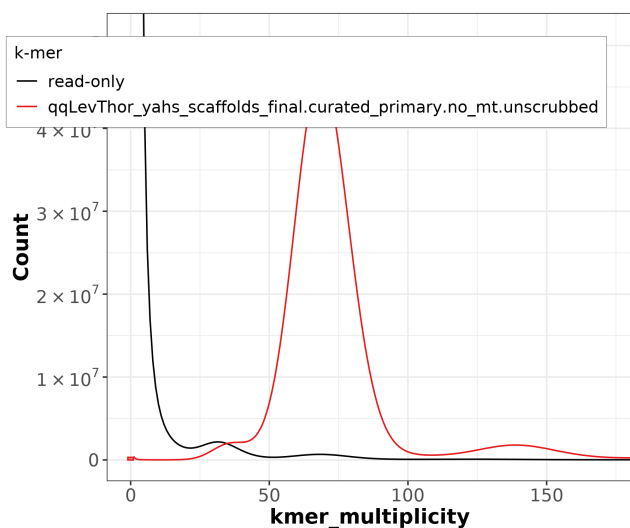

collapsed [LINK]

K-mer spectra of curated assembly
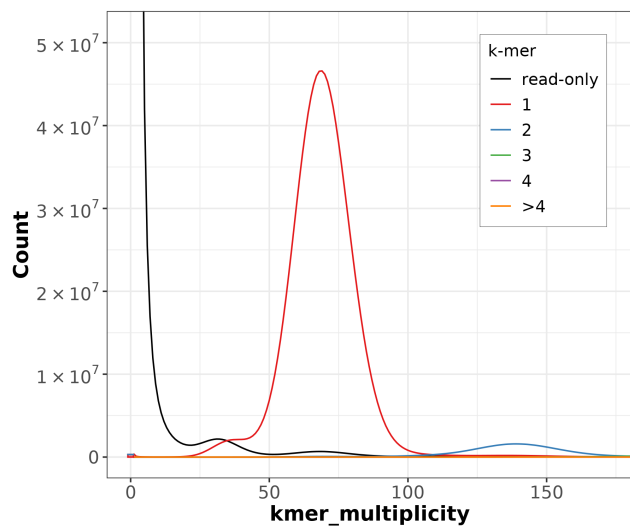
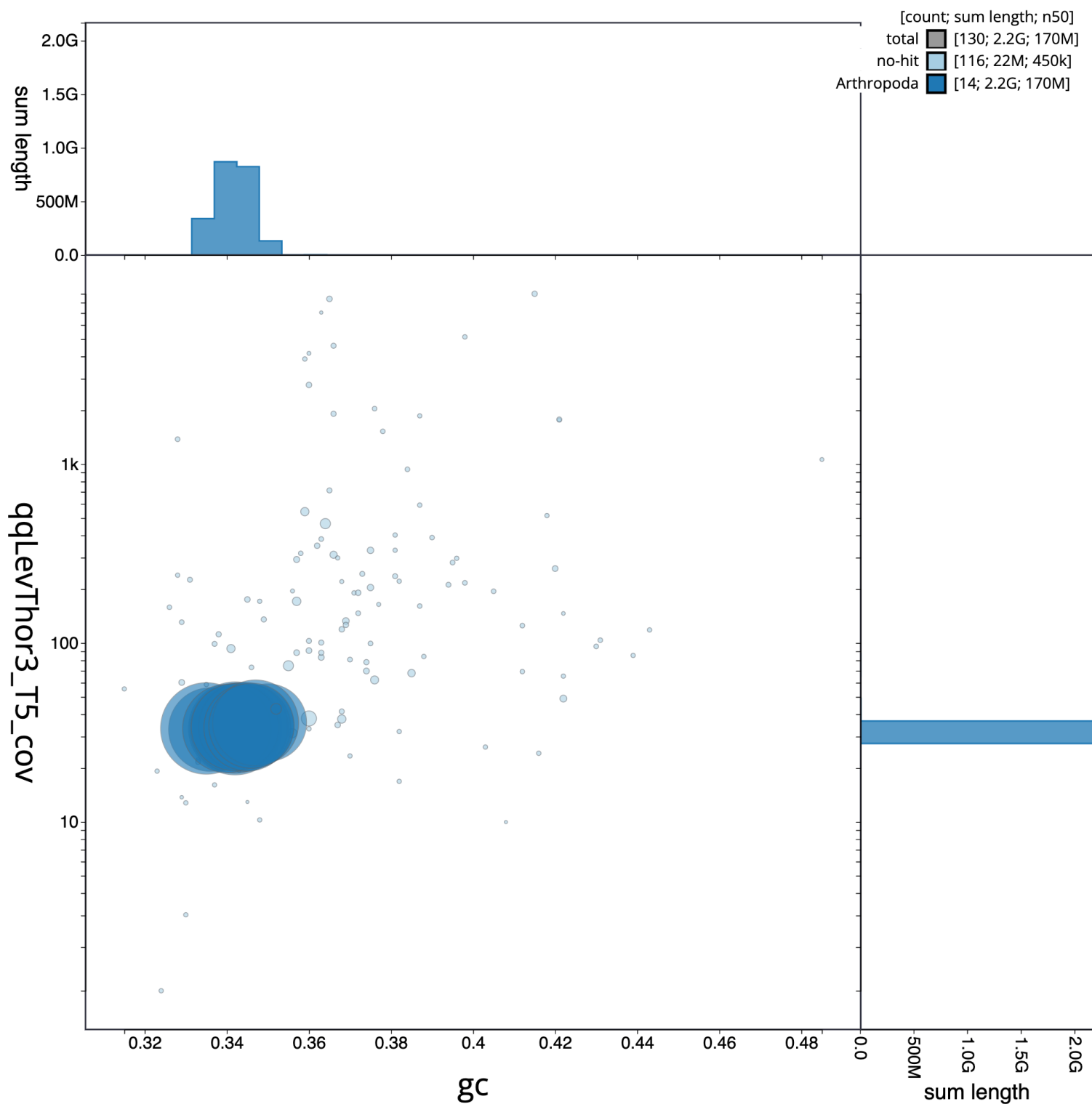Distribution of k-mer counts per copy numbers found in asm

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	40x	62x	125x

Assembly pipeline

- **Trim_galore**
  |_ *ver:* 0.6.7
  |_ *key param:* --gzip
  |_ *key param:* =q 20
  |_ *key param:* --paired
  |_ *key param:* retain_unpaired
- **Filtlong**
  |_ *ver:* 0.2.1
  |_ *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

Curation pipeline

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

Submitter: Francisco Camara
Affiliation: CNAG Barcelona

Date and time: 2025-02-18 15:48:54 CET