

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	3086104
ToLID	<b>qqLeiSuba9</b>
Species	Leiobunum subalpinum
Class	Arachnida
Order	Opiliones

Genome Traits	Expected	Observed
Haploid size (bp)	674,105,690	578,300,461
Haploid Number	10 (source: ancestor)	11
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q46

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

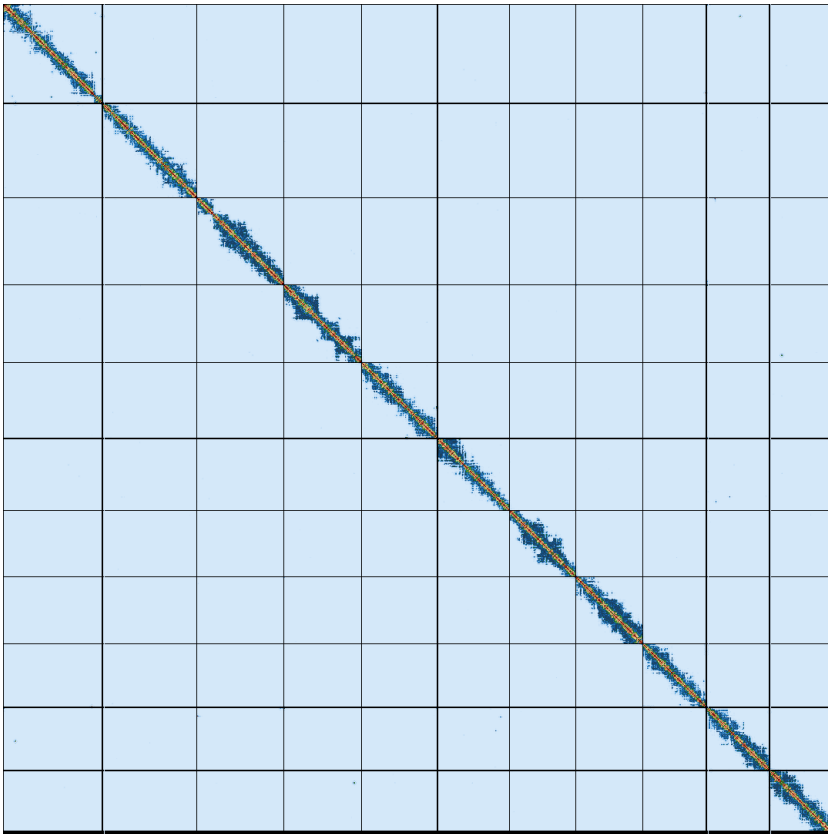
- . Interventions/Gb: 40
- . Contamination notes: "Sequences of Pseudomonadota (which include symbionts of many arthropod species) were present in 14 scaffolds of this organism. Scaffold\_13 and scaffold\_14 and especially long regions of similarity to to sequences of bacteria from this phylum. We also found eight other small scaffold with strong similarity to either chordata or mollusca which were also filtered out without detriment to the BUSCO or merquy QV. All sequences were removed from the yahs scaffolded assembly and are not present in the curated pretext and assembly."
- . Other observations: "This is the second version of the EAR report for this species which resulted in a new assembly after addressing the changes and suggestions provided by the reviewers. The curation was generally straightforward and the HiC signal along the diagonal quite strong as we produced a lot of HiC data assuming the genome size to be larger than what it turned out to be. The sex of the specimen was set to unknown in our lims system. We were not able to detect any lower coverage super\_scaffolds and therefore were not able to detect the sex chromosome. The curated assembly stats (gfastas, busco and merquy), the pretext file and the blobplot were generated from the assembly after filtering out for the scaffold-containing contaminants."

# Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	584,641,086	578,300,461
GC %	36.54	36.37
Gaps/Gbp	354.06	378.7
Total gap bp	41,400	43,800
Scaffolds	74	44
Scaffold N50	52,294,751	52,575,036
Scaffold L50	5	5
Scaffold L90	11	10
Contigs	281	263
Contig N50	5,187,333	5,243,763
Contig L50	35	34
Contig L90	128	123
QV	46.3203	46.5446
Kmer compl.	88.4393	88.4034
BUSCO sing.	95.5%	95.8%
BUSCO dupl.	0.6%	0.3%
BUSCO frag.	1.8%	1.8%
BUSCO miss.	2.1%	2.1%

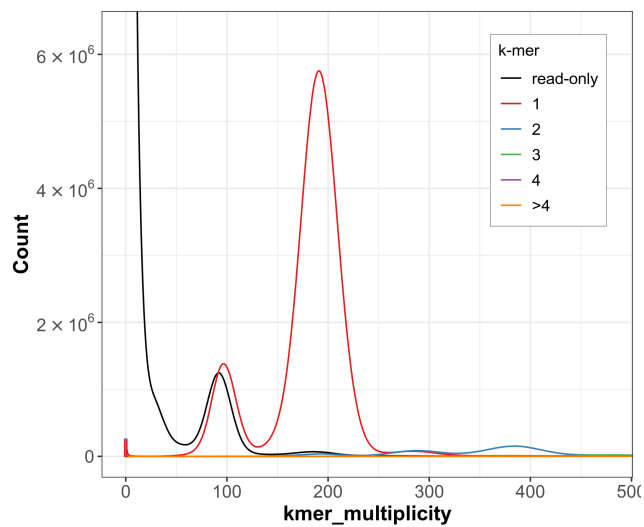
BUSCO: 5.4.0 (euk\_genome\_met, metaeuk) / Lineage: metazoa\_odb10 (genomes:65, BUSCOs:954)

# HiC contact map of curated assembly

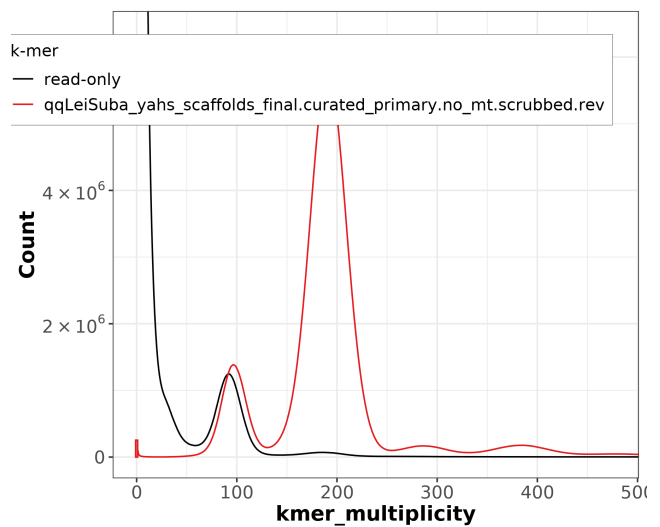


collapsed [\[LINK\]](#)

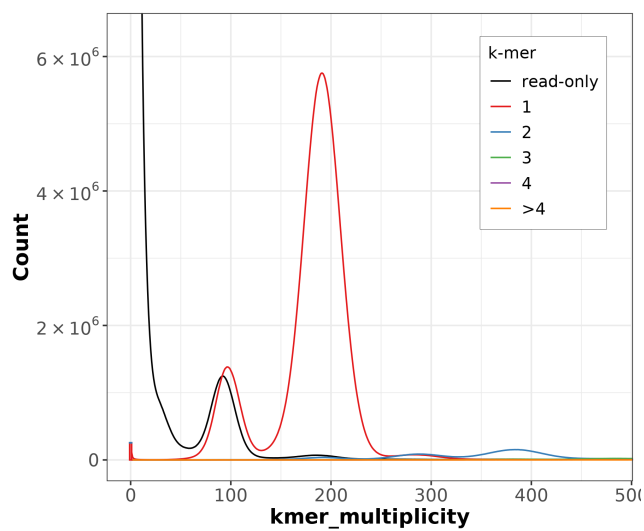
# K-mer spectra of curated assembly



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

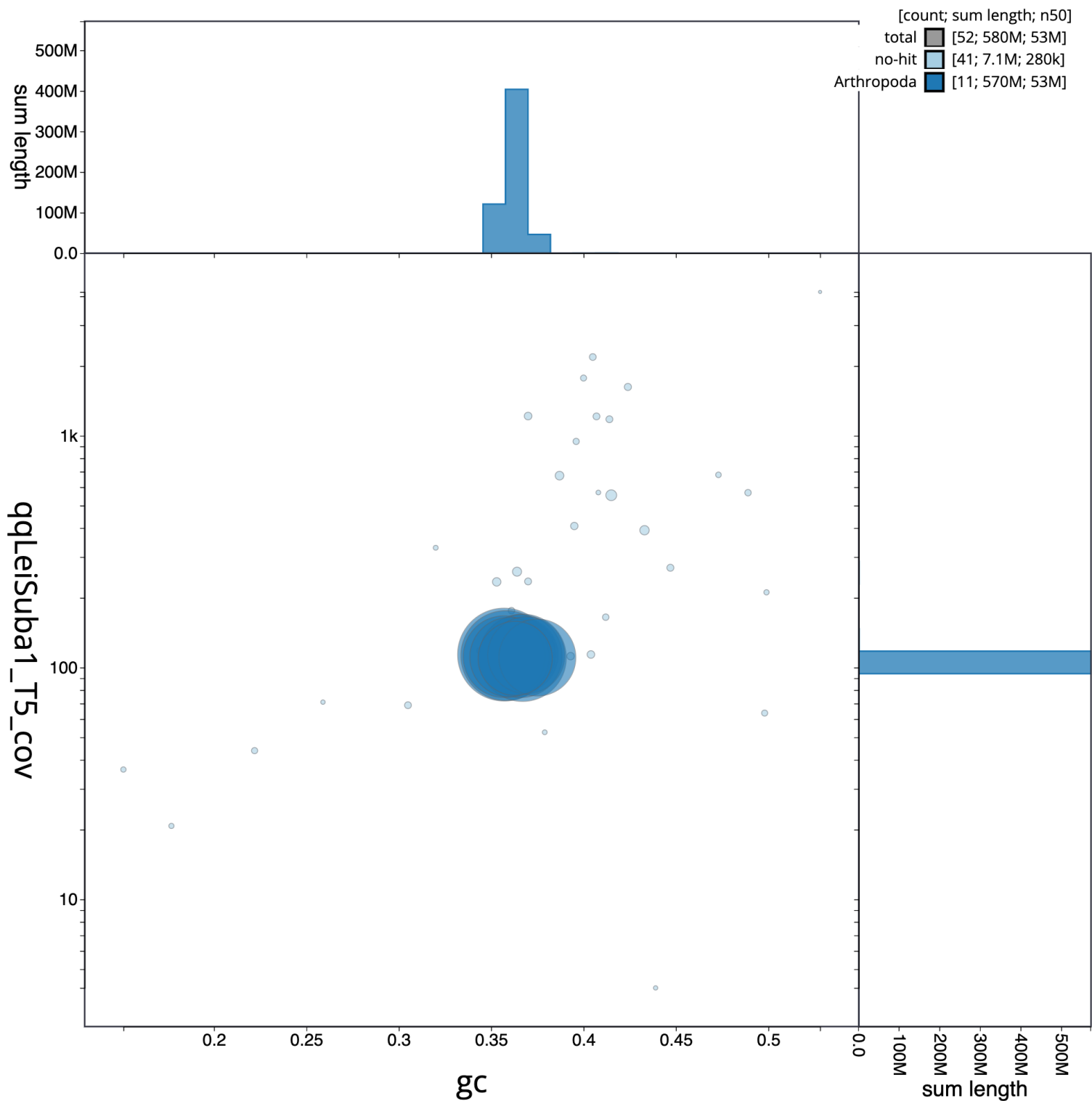


Distribution of k-mer counts coloured by their presence in reads/assemblies



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

# 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	115x	140x	210x

## 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: 2025-02-25 14:14:04 CET