

# ERGA Assembly Report

v24.10.15

Tags: non-curated[ INVALID TAG ]

TxID	3229130
ToLID	<b>qqButKunt1</b>
Species	Buthus kunti
Class	Arachnida
Order	Scorpiones

Genome Traits	Expected	Observed
Haploid size (bp)	1,142,780,423	1,162,206,769
Haploid Number	25 (source: ancestor)	20
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 7.7.Q66

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

- . Observed Haploid Number is different from Expected
- . Not 90% of assembly in chromosomes for pri

## Curator notes

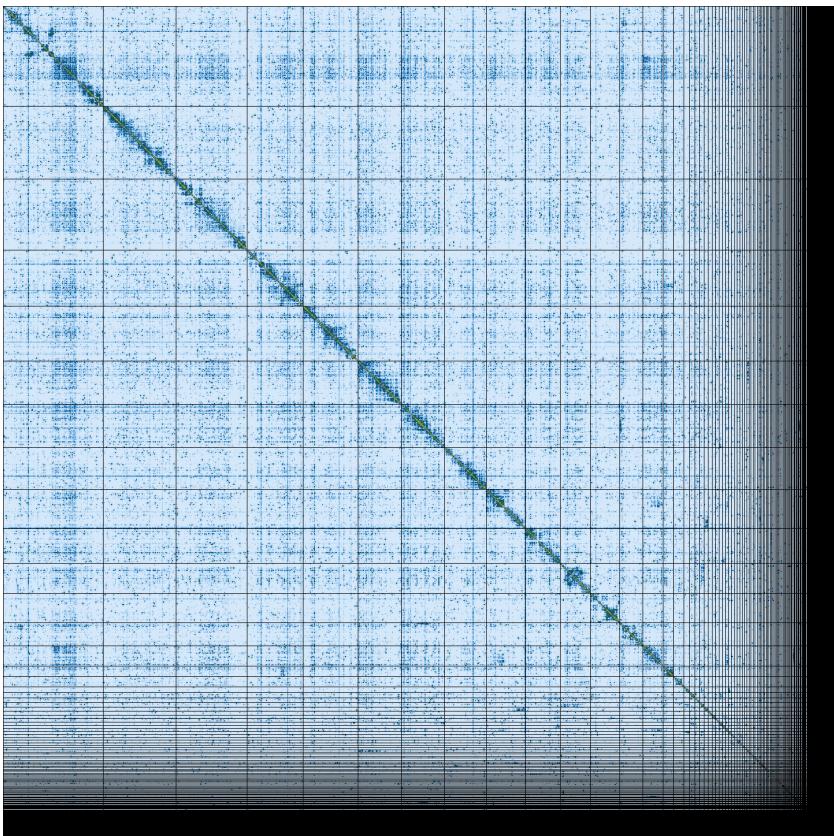
- . Interventions/Gb: 0
- . Contamination notes: "No contaminations were detected with FCS-GX."
- . Other observations: "This is a draft genome, not a manually curated assembly. The sequencing platform sequenced multiple Hi-C libraries, but the number of reads mapping back to the assembly was small, with very little long range information. The pre-curated genome looks nice, with high contiguity, no signs of contamination. This assembly could definitely be annotated and be used by the stakeholder, yet it's not a chromosome level curated genome assembly. I am providing the Hi-C map image and Pretext files anyway. You can see how long the scaffolds are, some are likely chromosome level at this point, but it's hard to prove that, since there are no closely related species for synteny plots and no reliable telomers."

## Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	1,162,206,769	1,162,206,769
GC %	29.51	29.51
Gaps/Gbp	0	0
Total gap bp	0	0
Scaffolds	113	113
Scaffold N50	51,777,195	51,777,195
Scaffold L50	8	8
Scaffold L90	24	24
Contigs	113	113
Contig N50	51,777,195	51,777,195
Contig L50	8	8
Contig L90	24	24
QV	66.1518	66.1518
Kmer compl.		
BUSCO sing.	96.2%	96.2%
BUSCO dupl.	3.4%	3.4%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	0.3%	0.3%

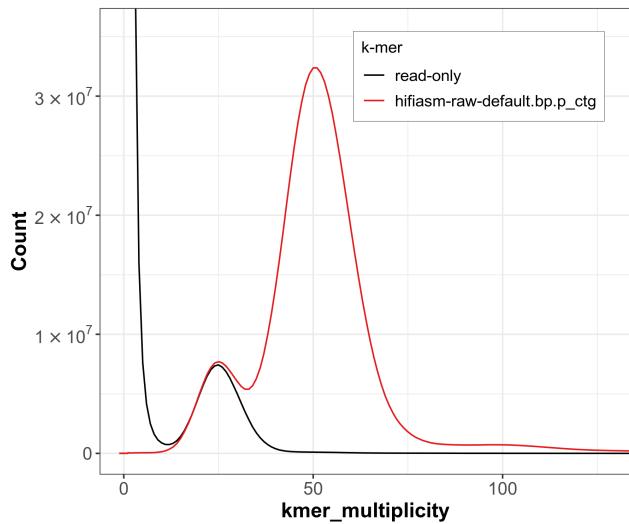
BUSCO: 6.0.0 (euk\_genome\_min, miniprot) / Lineage: metazoa\_odb10 (genomes:65, BUSCOs:954)

# HiC contact map of curated assembly

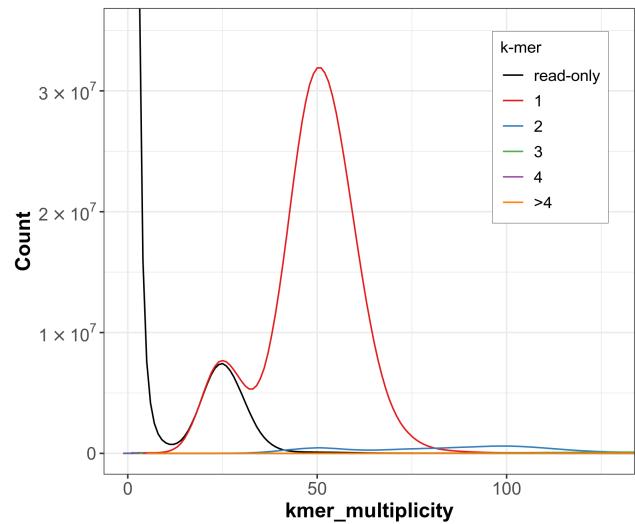


pri [LINK]

# K-mer spectra of curated assembly

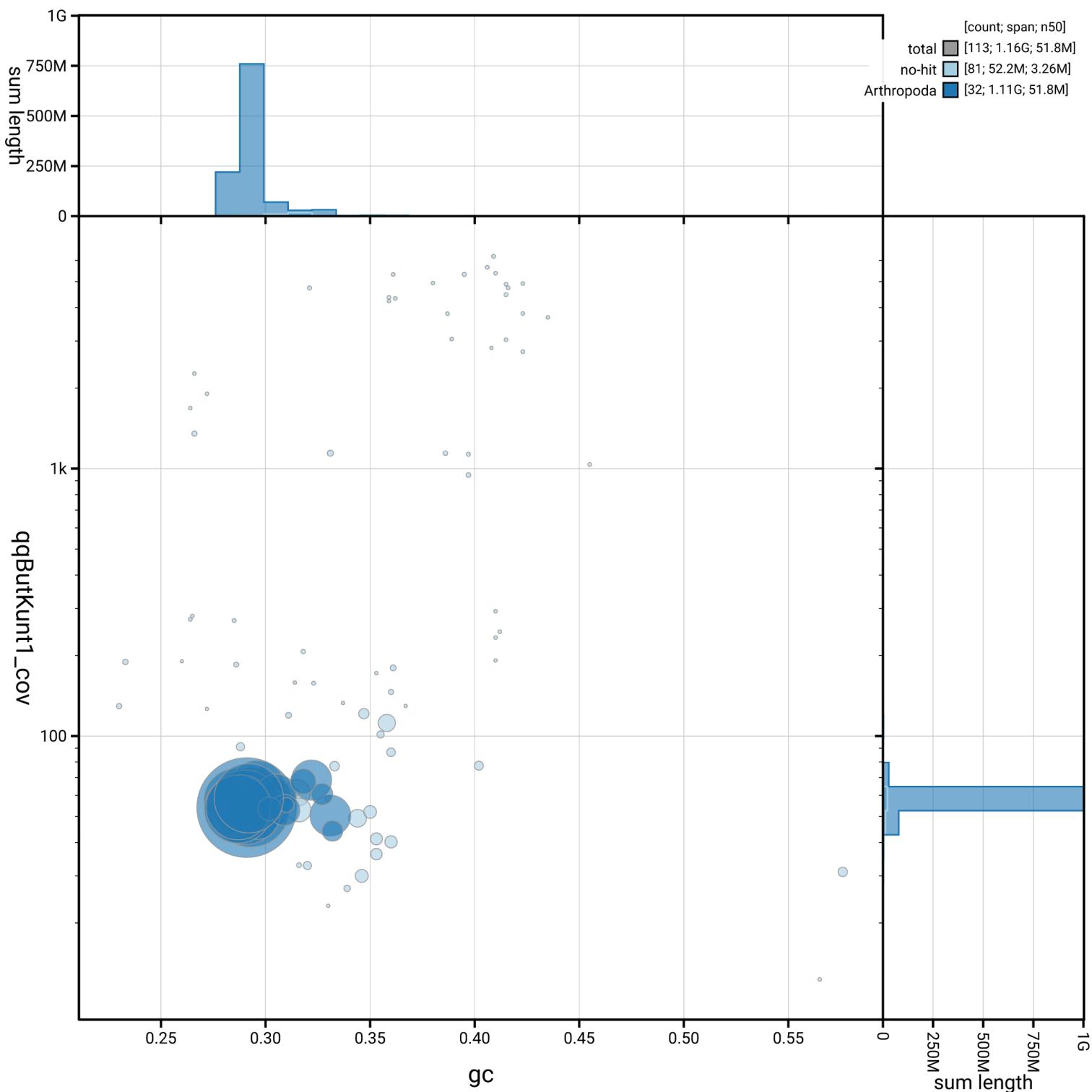


Distribution of k-mer counts coloured by their presence in reads/assemblies



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

# 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	HiFi	HiC
Coverage	52x	209x

## Assembly pipeline

```
- Hifiasm
  |_ ver: 0.25.0-r726
  |_ key param: 13
- purge_dups
  |_ ver: 1.2.6
  |_ key param: NA
```

## Curation pipeline

```
- GRIT_Rapid
  |_ ver: 0
  |_ key param: NA
- HiGlass
  |_ ver: 0
  |_ key param: NA
```

Submitter: André Soares

Affiliation: SciLifeLab

Date and time: 2026-02-09 11:24:54 CET