

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	2883604
ToLID	qqEusFeti1
Species	Euscorpius feti
Class	Arachnida
Order	Scorpiones

Genome Traits	Expected	Observed
Haploid size (bp)	4,312,354,347	4,371,788,358
Haploid Number	35 (source: ancestor)	34
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 7.8.Q67

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

- . Observed Haploid Number is different from Expected
- . BUSCO single copy value is less than 90% for pri
- . BUSCO duplicated value is more than 5% for pri

Curator notes

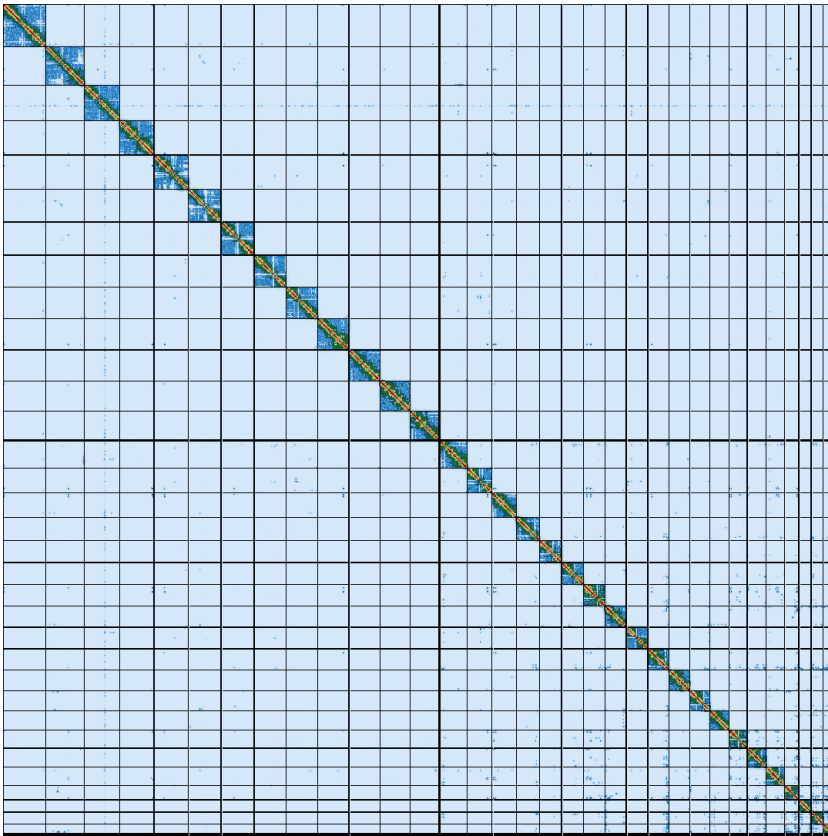
- . Interventions/Gb: 8
- . Contamination notes: "The BTK pipeline detected 1 contaminant scaffold (scaffold_136) in the pre-curated assembly. However this scaffold was kept, as it is a false positive containing 1 arachnida_odb10 Complete BUSCO and contacts strongly with a chromosome (scaffold_31 or SUPER_32 in curated assembly). For this report, we just show the blob plot of the precurated assembly, actually the only difference with the curated is the number of total scaffolds (184 instead of 187 after curation)."
- . Other observations: "Our best genome assembly was obtained running hifiiasm with the Hi-C phasing option and then scaffolding the primary assembly with YaHS. Although the sequenced specimen was a female, during curation we could not detect the sex chromosomes. In fact, the lack of distinguishable sex chromosomes is consistent with previous studies in the Euscorpius genus (10.15671/HJBC.2019.269). Regarding to the mitochondrial genome, this was assembled with FOAM into a single circular contig of 15,340 bp."

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	4,371,786,958	4,371,788,358
GC %	36.59	36.59
Gaps/Gbp	16.7	18.3
Total gap bp	14,600	16,000
Scaffolds	187	184
Scaffold N50	158,507,893	151,694,389
Scaffold L50	12	13
Scaffold L90	28	28
Contigs	260	264
Contig N50	88,203,554	81,262,035
Contig L50	17	18
Contig L90	54	55
QV	67.4071	67.4071
Kmer compl.	97.6951	97.6951
BUSCO sing.	88.7%	88.7%
BUSCO dupl.	8.3%	8.3%
BUSCO frag.	1.3%	1.3%
BUSCO miss.	1.7%	1.7%

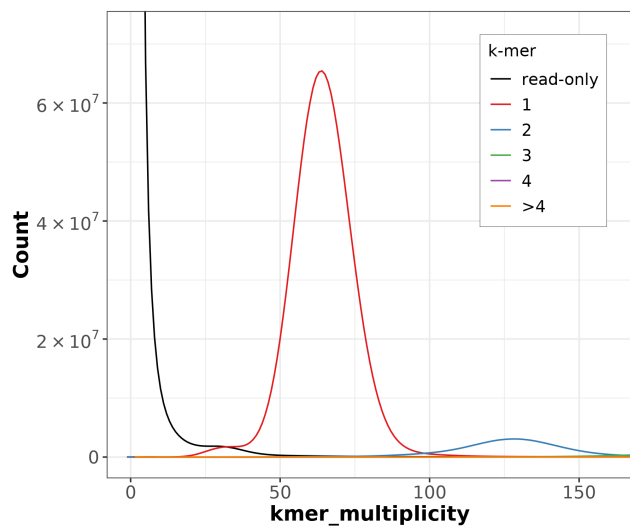
BUSCO: 5.5.0 (euk_genome_met, metaeuk) / Lineage: arachnida_odb10 (genomes:10, BUSCOs:2934)

HiC contact map of curated assembly

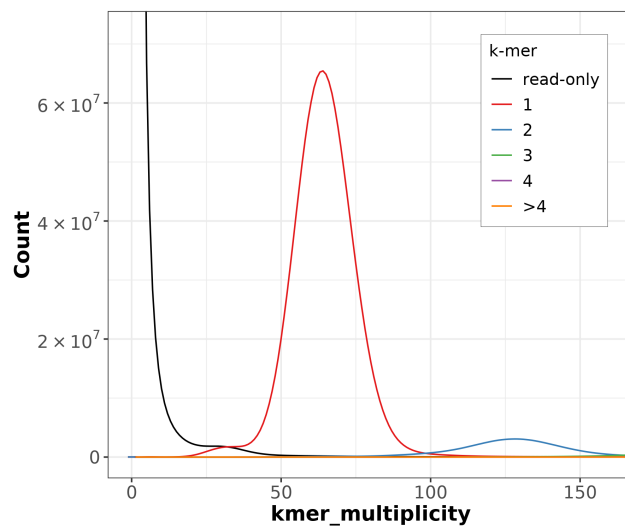


pri [\[LINK\]](#)

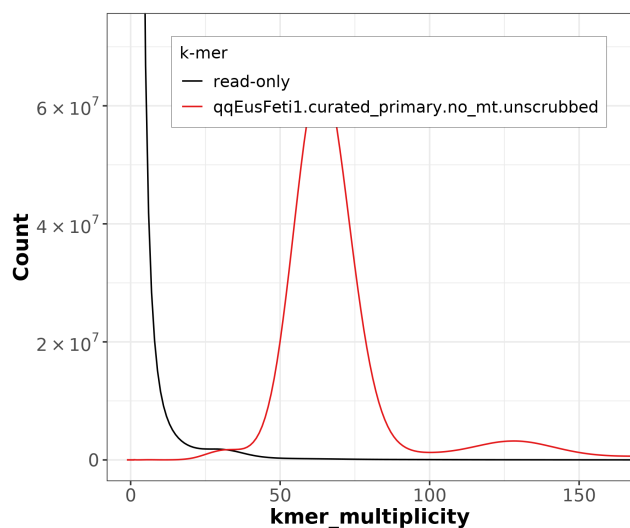
K-mer spectra of curated assembly



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

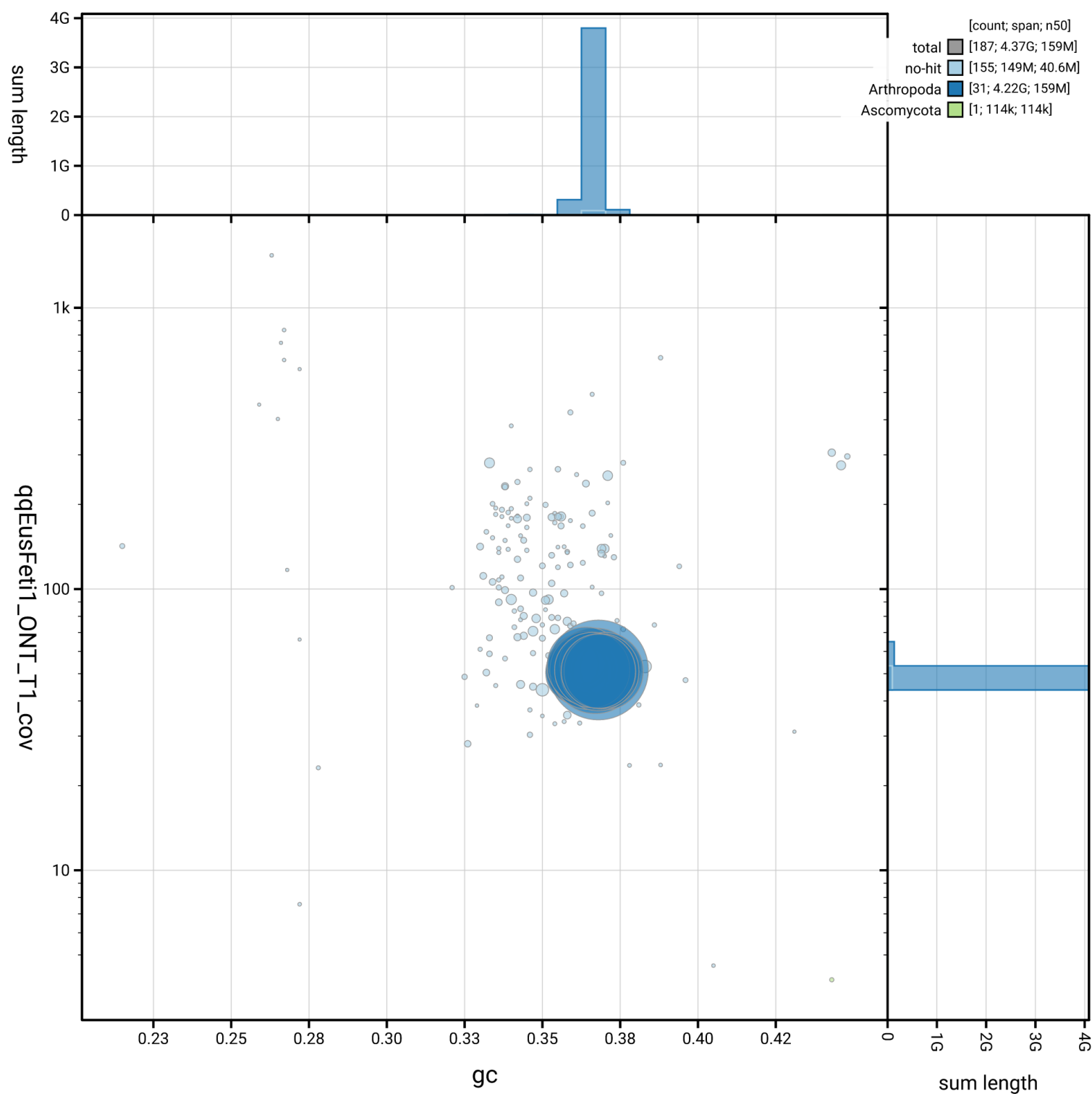


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	ONT	Illumina	OmniC
Coverage	64x	22x	82x

Assembly pipeline

- **CLAWS pipeline**
 - |_ ver: 2.2.0 and 2.3.0
 - |_ 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
- **hifiasm**
 - |_ ver: 0.24.0
 - |_ 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

- **PretextViewAI**
 - |_ ver: 1.0.3
 - |_ key param: NA

Affiliation: CNAG

Date and time: 2025-09-02 16:32:05 CEST