

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	1795686
ToLID	<b>qqChaOliv1</b>
Species	Chaetopelma olivaceum
Class	Arachnida
Order	Araneae

Genome Traits	Expected	Observed
Haploid size (bp)	4,899,563,537	5,076,309,119
Haploid Number	9 (source: ancestor)	33
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	XXX

## EBP metrics summary and curation notes

Obtained EBP quality metric for hap1: 8.8.Q70

Obtained EBP quality metric for hap2: 8.8.Q70

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

- . Observed Haploid Number is different from Expected
- . Observed sex is different from Sample sex
- . BUSCO duplicated value is more than 5% for hap1
- . Kmer completeness value is less than 90 for hap2
- . BUSCO single copy value is less than 90% for hap2

## Curator notes

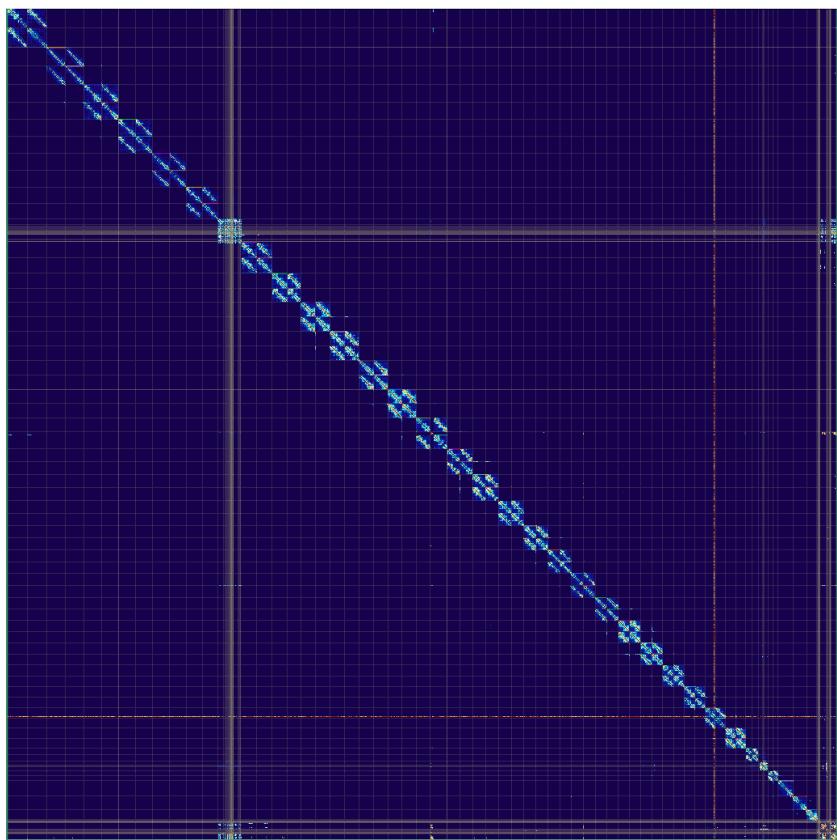
- . Interventions/Gb:
- . Contamination notes: "5,851,380 bp contained in 75 sequences were considered contaminants by FCS-GX and removed from hap1. Additionally, 2,128,985 contained in 32 sequences were also considered contaminants and removed from hap2. The longest removed contigs for hap1 and hap2 were 263,942 and 160,176 bp, respectively."
- . Other observations: "This genome assembly was produced by assembling ONT data with Hifiasm and then scaffolding with Yahs. Given the good quality of both haplotypes, curation was made in diploid mode and we are planning to submit both haplotypes to the ENA. Curation made 2 cuts in contigs, 4 breaks at gaps and 8 joins. 3 sex chromosomes were identified during curation based on half coverage, given their structure and based on some of the existant knowledge in spiders, we believe that they correspond to 3 X chromosomes. The mitogenome was successfully assembled with FOAM."

# Quality metrics table

Metrics	Pre-curation hap1	Pre-curation hap2	Curated hap1	Curated hap2
Total bp	5,070,272,142	4,434,489,659	5,076,309,119	4,420,473,121
GC %	40.97	40.79	40.95	40.76
Gaps/Gbp	2.56	6.31	3.15	6.56
Total gap bp	2,600	5,600	3,200	5,800
Scaffolds	898	266	822	233
Scaffold N50	158,663,373	154,201,259	158,663,373	154,201,259
Scaffold L50	14	13	14	13
Scaffold L90	30	26	30	26
Contigs	911	294	838	262
Contig N50	144,858,072	131,742,126	152,099,163	127,187,526
Contig L50	15	14	15	14
Contig L90	35	34	32	37
QV	68.111	68.896	70.2458	70.2341
Kmer compl.	93.1278	85.1184	93.1071	85.0945
BUSCO sing.	92.3%	89.6%	92.3%	89.5%
BUSCO dupl.	5.6%	3.9%	5.5%	4.0%
BUSCO frag.	1.2%	1.6%	1.2%	1.5%
BUSCO miss.	1.0%	4.9%	1.0%	5.0%

BUSCO: 6.0.0 (euk\_genome\_min, miniprot) / Lineage: arachnida\_odb12 (genomes:34, BUSCOs:1123)

# HiC contact map of curated assembly

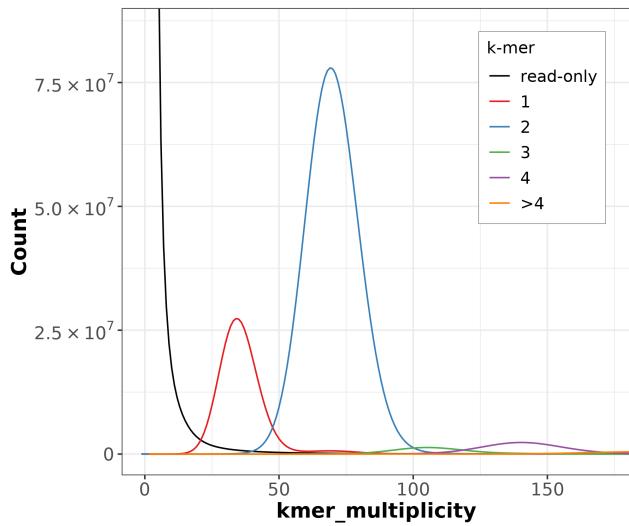


**hap1** [\[LINK\]](#)

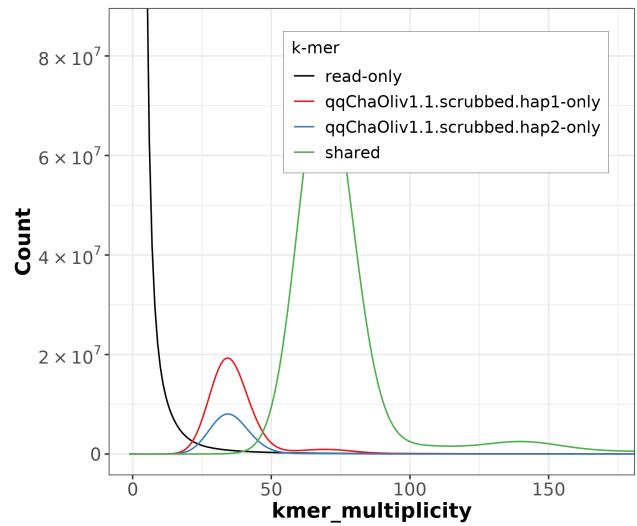
**hap2** HiC PNG is missing!

**hap2** File link is missing!

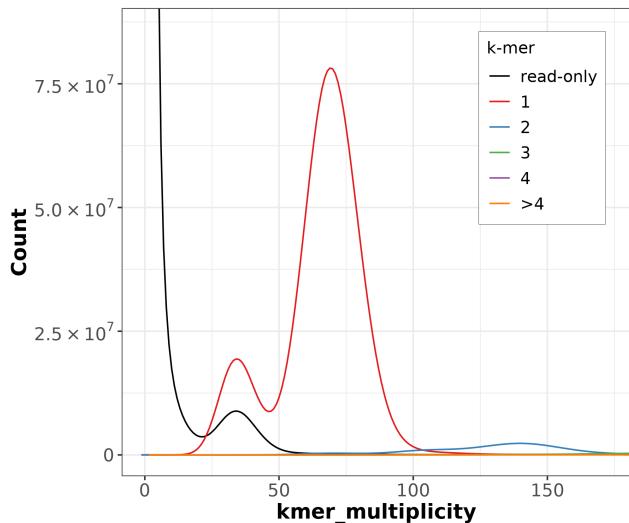
# K-mer spectra of curated assembly



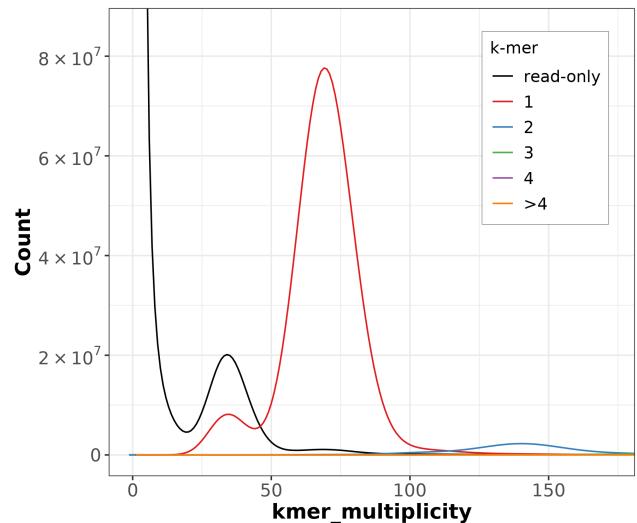
Distribution of k-mer counts per copy numbers found in asm (dipl.)



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



Distribution of k-mer counts per copy numbers found in **hap1** (hapl.)



Distribution of k-mer counts per copy numbers found in **hap2** (hapl.)

# Post-curation contamination screening

## Data profile

Data	ONT	Illumina	Omni-C
Coverage	64	17	32

## Assembly pipeline

```
- CLAWS
  |_ ver: v3.0
  |_ key param: NA
- Filtlong
  |_ ver: v0.2.1
  |_ key param: --minlen 1000
  |_ key param: --min_mean_q 80
  |_ key param: --target_bases 300000000000
- Hifiasm
  |_ ver: 0.24.0
  |_ key param: --ont
  |_ key param: --telo-m ACCCCG
  |_ key param: -h1 -h2
- Yahs
  |_ ver: 1.2a
  |_ key param: -mq 10
  |_ key param: --no-contig-ec
- FOAM
  |_ ver: 0.5
  |_ key param: NA
```

## Curation pipeline

```
- FCS-GX
  |_ ver: None
  |_ key param: NA
- PretextViewAI
  |_ ver: 1.0.5
  |_ key param: NA
```

Submitter: Jessica Gomez-Garrido

Affiliation: CNAG

Date and time: 2026-01-28 10:40:30 CET