

ERGA Assembly Report

v24.09.10

Tags: ERGA-BGE

TxID	3034272
ToLID	qqChaLymb1.1
Species	Chaetopelma lymerakisi
Class	Arachnida
Order	Araneae

Genome Traits	Expected	Observed
Haploid size (bp)	5,204,012,179	5,029,523,872
Haploid Number	23 (source: ancestor)	33
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q60

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 collapsed
- . BUSCO duplicated value is more than 5% for collapsed

Curator notes

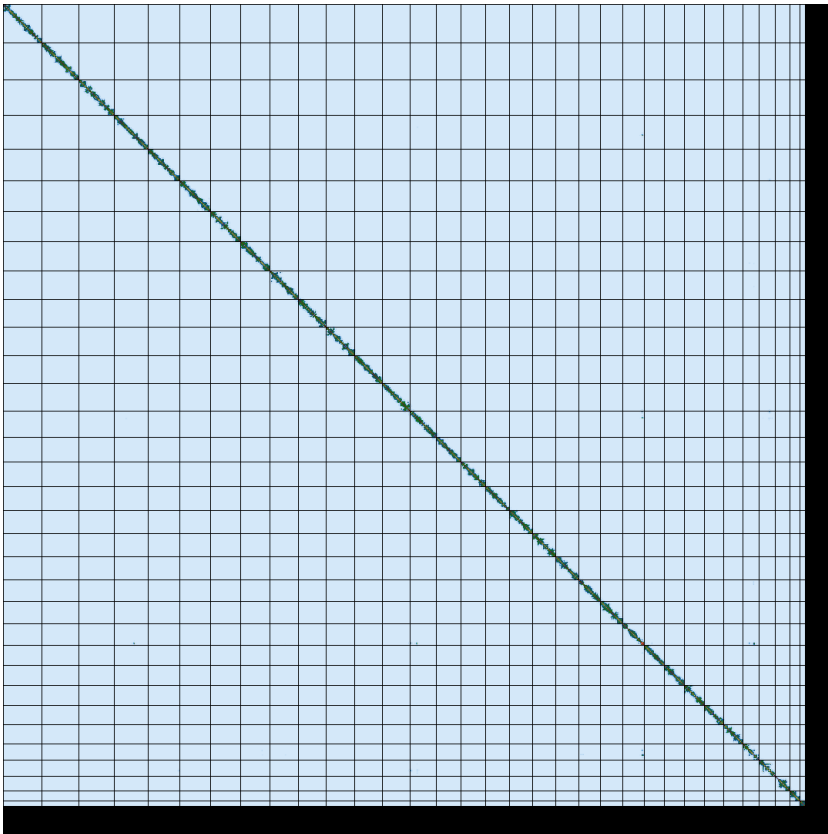
- . Interventions/Gb: 60
- . Contamination notes: "72 contigs were identified as bacterial and removed, totaling 3.9 Mb in size, with the longest being 239 Kb in length "
- . Other observations: "80 allelic duplications were removed, totaling 47 Mb in size. The expected size was 1.2 Gb, but we ended up with an assembly of 5 Gb. This explains the relatively low coverage of the PacBio data (20x). However, we stopped at this point since the contig N50 was well above 1 Mb. As for the Hi-C, we sequenced two libraries, but again, the coverage is low. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	5,076,762,340	5,029,523,872
GC %	41.45	41.45
Gaps/Gbp	179.05	203.4
Total gap bp	90,900	116,900
Scaffolds	2,152	2,068
Scaffold N50	159,851,774	161,780,521
Scaffold L50	14	14
Scaffold L90	30	29
Contigs	3,061	3,091
Contig N50	11,532,000	10,395,000
Contig L50	132	145
Contig L90	469	508
QV	60.2668	60.266
Kmer compl.	97.8331	97.1094
BUSCO sing.	89.9%	88.9%
BUSCO dupl.	6.9%	6.5%
BUSCO frag.	1.3%	1.4%
BUSCO miss.	1.9%	3.2%

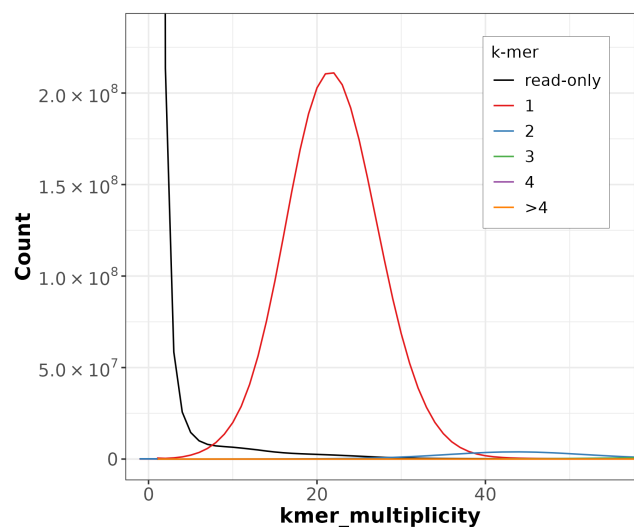
BUSCO 5.4.3 Lineage: arachnida_odb10 (genomes:10, BUSCOs:2934)

HiC contact map of curated assembly

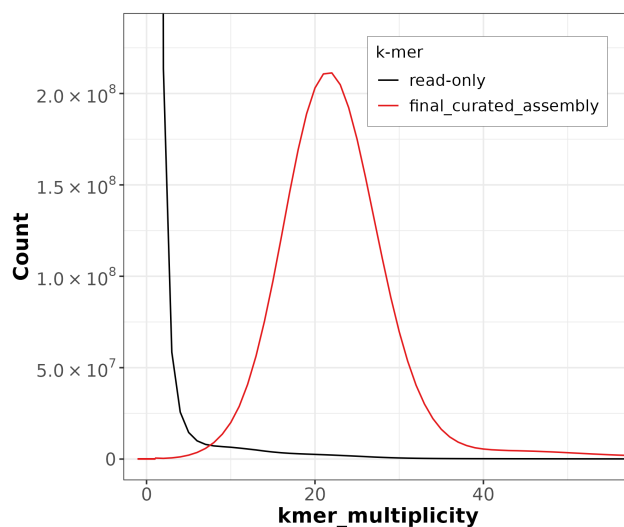


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

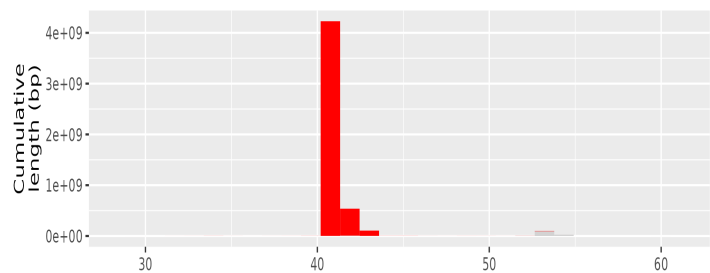


Distribution of k-mer counts per copy numbers found in asm



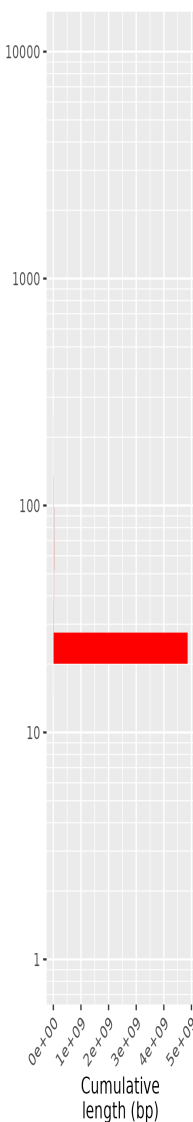
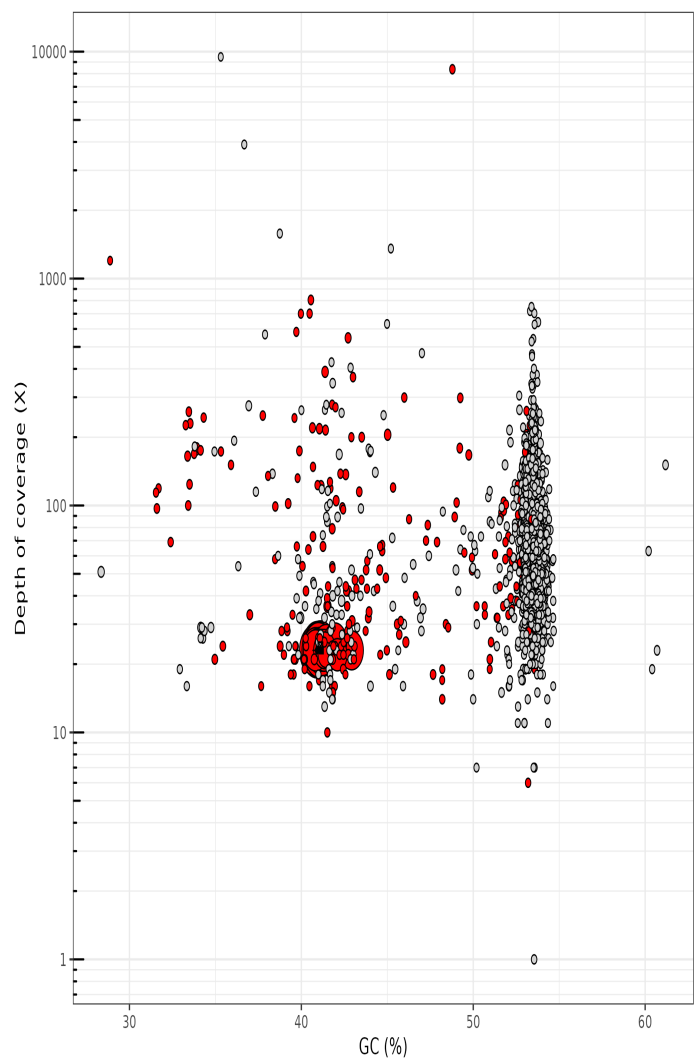
Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph

(1 0X contig has been hidden)



Longest sequences (bp)

- qqChalymb1_1 - 233910238 (Eukaryota)
- ▲ qqChalymb1_2 - 224799449 (Eukaryota)
- qqChalymb1_3 - 213397768 (Eukaryota)
- + qqChalymb1_4 - 206859435 (Eukaryota)
- ▣ qqChalymb1_5 - 189381632 (Eukaryota)

superkingdom

- Eukaryota
- N/A

Length (bp)

- 5.0e+07
- 1.0e+08
- 1.5e+08
- 2.0e+08

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	PACBIO Hifi	Arima (4-enz)
Coverage	20	17

Assembly pipeline

- **Hifiasm**
 - |_ *ver*: 0.19.5-r593
 - |_ *key param*: NA
- **purge_dups**
 - |_ *ver*: 1.2.5
 - |_ *key param*: NA
- **YaHS**
 - |_ *ver*: 1.2
 - |_ *key param*: NA

Curation pipeline

- **PretextMap**
 - |_ *ver*: 0.1.9
 - |_ *key param*: NA
- **PretextView**
 - |_ *ver*: 0.2.5
 - |_ *key param*: NA

Submitter: Jean-Marc Aury

Affiliation: Genoscope

Date and time: 2024-10-20 13:37:04 CEST