ERGA Assembly Report
v24.10.15

Tags: ERGA-BGE

TxID	3229104
ToLID	**qqAegCypr1**
Species	Aegaeobuthus cyprius
Class	Arachnida
Order	Scorpiones

Genome Traits	Expected	Observed
Haploid size (bp)	1,139,951,727	1,147,323,490
Haploid Number	11 (source: ancestor)	10
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q76

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

. Observed Haploid Number is different from Expected

. BUSCO duplicated value is more than 5% for collapsed

Curator notes
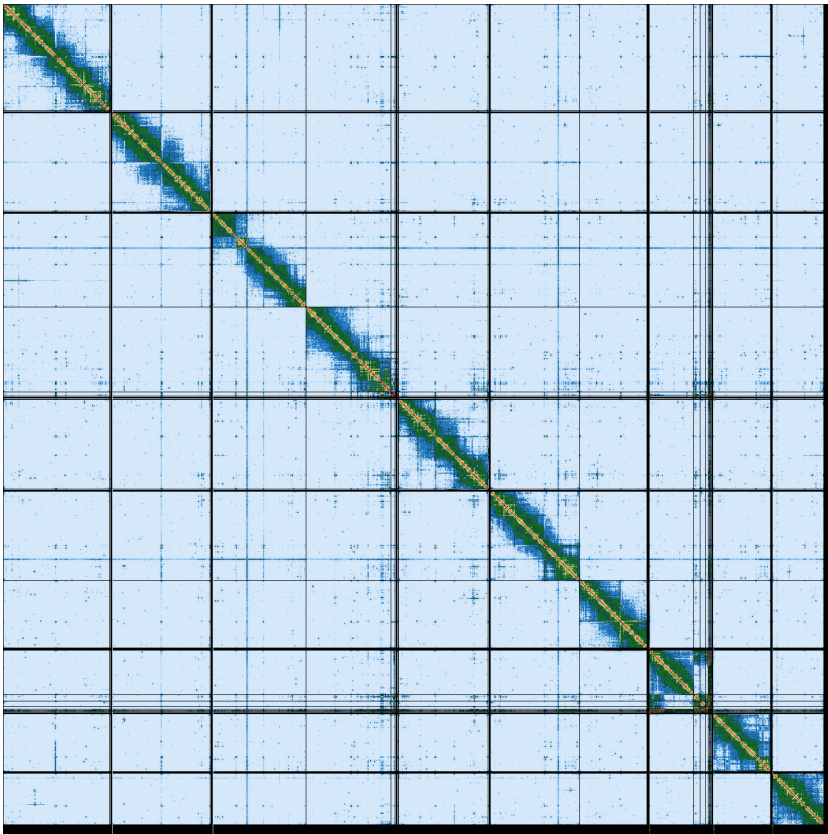
. Interventions/Gb:
. Contamination notes: "FCS-GX detected four small bacterial contigs matching Knoellia remsis and Knoellia locipacati, and totalling 125,318 bp. These were removed from the assembly. Additionally, a 579,681 bp long contig was removed from the assembly during curation on the basis of skewed GC content, high coverage, very little Hi-C contact with the rest of the genome, and tiara classification as bacterial."
. Other observations: "The assembly was produced with hifiasm and the *p_ctg.fa was selected. The assembly was decontaminated with FCS-GX, and haplotigs purged with purge_dups. Scaffolding was performed with YaHS. Mitochondria was assembled with Oatk."

Quality metrics table

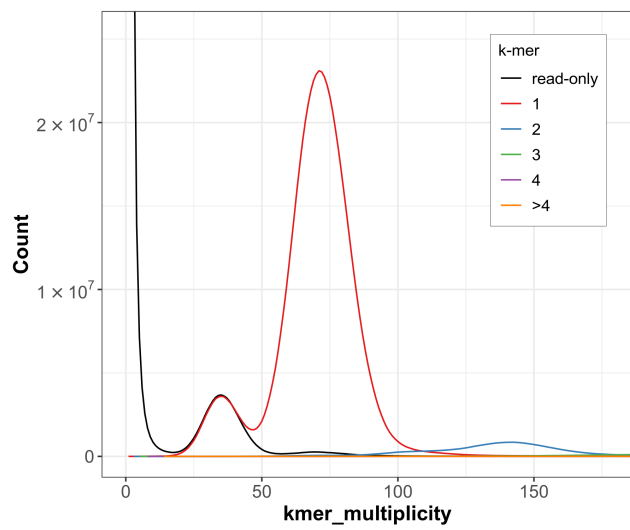| Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,152,164,342	1,147,323,490
GC %	29.5	29.51
Gaps/Gbp	48.6	49.68
Total gap bp	11,200	11,400
Scaffolds	155	151
Scaffold N50	116,441,941	122,189,364
Scaffold L50	5	5
Scaffold L90	10	10
Contigs	211	208
Contig N50	26,129,882	26,129,882
Contig L50	13	13
Contig L90	47	46
QV	76.1731	76.2985
Kmer compl.	91.3086	91.2129
BUSCO sing.	90.6%	90.9%
BUSCO dupl.	7.8%	7.5%
BUSCO frag.	0.4%	0.4%
BUSCO miss.	1.3%	1.3%

BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: arachnida_odb10 (genomes:10, BUSCOs:2934)
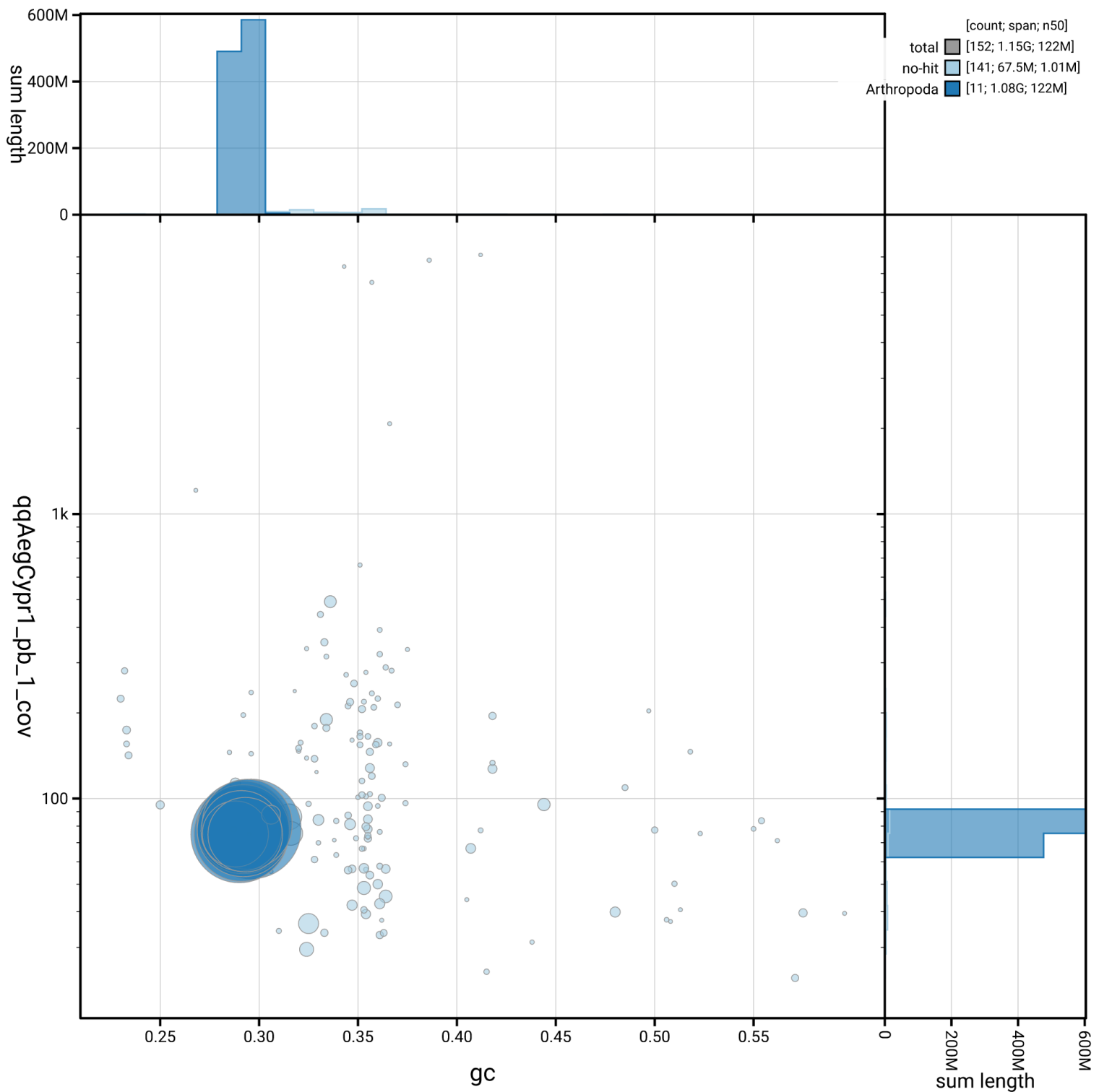
HiC contact map of curated assembly

collapsed [LINK]

K-mer spectra of curated assembly

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

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	PacBio HiFi	Arima HiC v2
Coverage	70x	28x

Assembly pipeline

- **hifiasm-hic**
    |_ *ver:* 0.25.0-r726
    |_ *key param:* NA
- **purge_dups**
    |_ *ver:* 1.2.5
    |_ *key param:* NA
- **YaHS**
    |_ *ver:* 1.2a.2
    |_ *key param:* NA
- **NBIS/Earth-Biogenome-Project-pilot**
    |_ *ver:* 99c8937
    |_ *key param:* NA
- **sanger-tol/blobtoolkit**
    |_ *ver:* 0.8.0
    |_ *key param:* NA
- **Oatk**
    |_ *ver:* 1.0
    |_ *key param:* -m aranae.fam
    |_ *key param:* -c 100

Curation pipeline

- **sanger-tol/curationpretext**
    |_ *ver:* 1.4.1
    |_ *key param:* NA
- **sanger-tol/agp-tpf-utils**
    |_ *ver:* 1.2.3
    |_ *key param:* NA

Submitter: Guilherme B. Dias
Affiliation: SciLifeLab

Date and time: 2026-01-13 18:58:41 CET