

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	293813
ToLID	qqAgeOrie5
Species	<i>Agelena orientalis</i>
Class	Arachnida
Order	Araneae

Genome Traits	Expected	Observed
Haploid size (bp)	6,397,309,515	6,488,071,794
Haploid Number	22 (source: ancestor)	22
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	X1X1X2X2

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q66

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

- . Observed sex is different from Sample sex
- . Kmer completeness value is less than 90 for collapsed
- . BUSCO duplicated value is more than 5% for collapsed

Curator notes

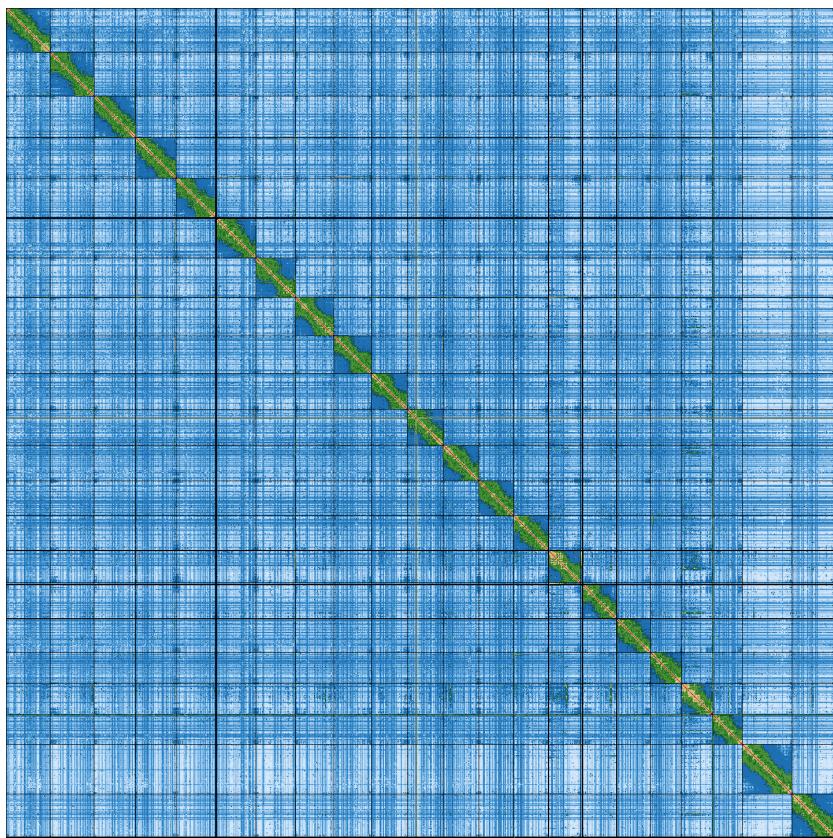
- . Interventions/Gb:
- . Contamination notes: "FCS-GX detected four bacterial contigs matching Cardinium endosymbiont and totalling 1,438,579 bp. These were removed from the assembly."
- . Other observations: "The assembly was produced with hifiasm and the collapsed *p_ctg.fa was selected. The assembly was decontaminated with FCS-GX, and haplotigs purged with purge_dups. Scaffolding was performed with YaHS. Mitochondria was assembled with MitoHiFi."

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	6,499,455,072	6,488,071,794
GC %	31.81	31.81
Gaps/Gbp	188.79	189.58
Total gap bp	245,400	246,000
Scaffolds	157	146
Scaffold N50	292,788,066	298,975,081
Scaffold L50	10	10
Scaffold L90	20	20
Contigs	1,384	1,376
Contig N50	11,280,987	11,280,987
Contig L50	170	170
Contig L90	618	615
QV	66.1059	66.1112
Kmer compl.	88.7927	88.7414
BUSCO sing.	93.7%	93.7%
BUSCO dupl.	5.4%	5.3%
BUSCO frag.	0.2%	0.2%
BUSCO miss.	0.7%	0.7%

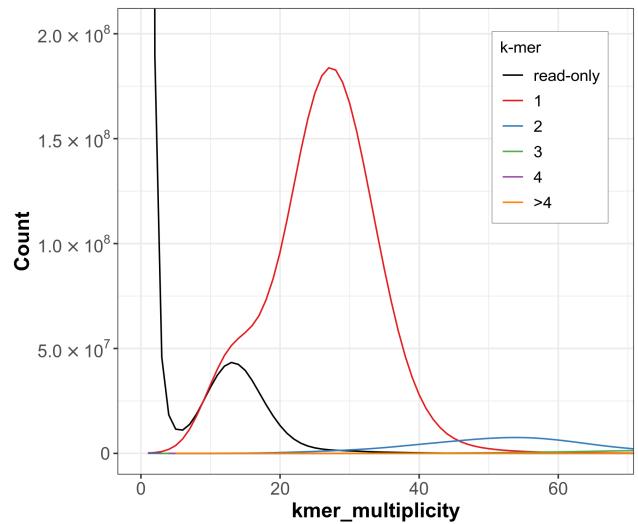
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: arachnida_odb10 (genomes:10, BUSCOs:2934)

HiC contact map of curated assembly



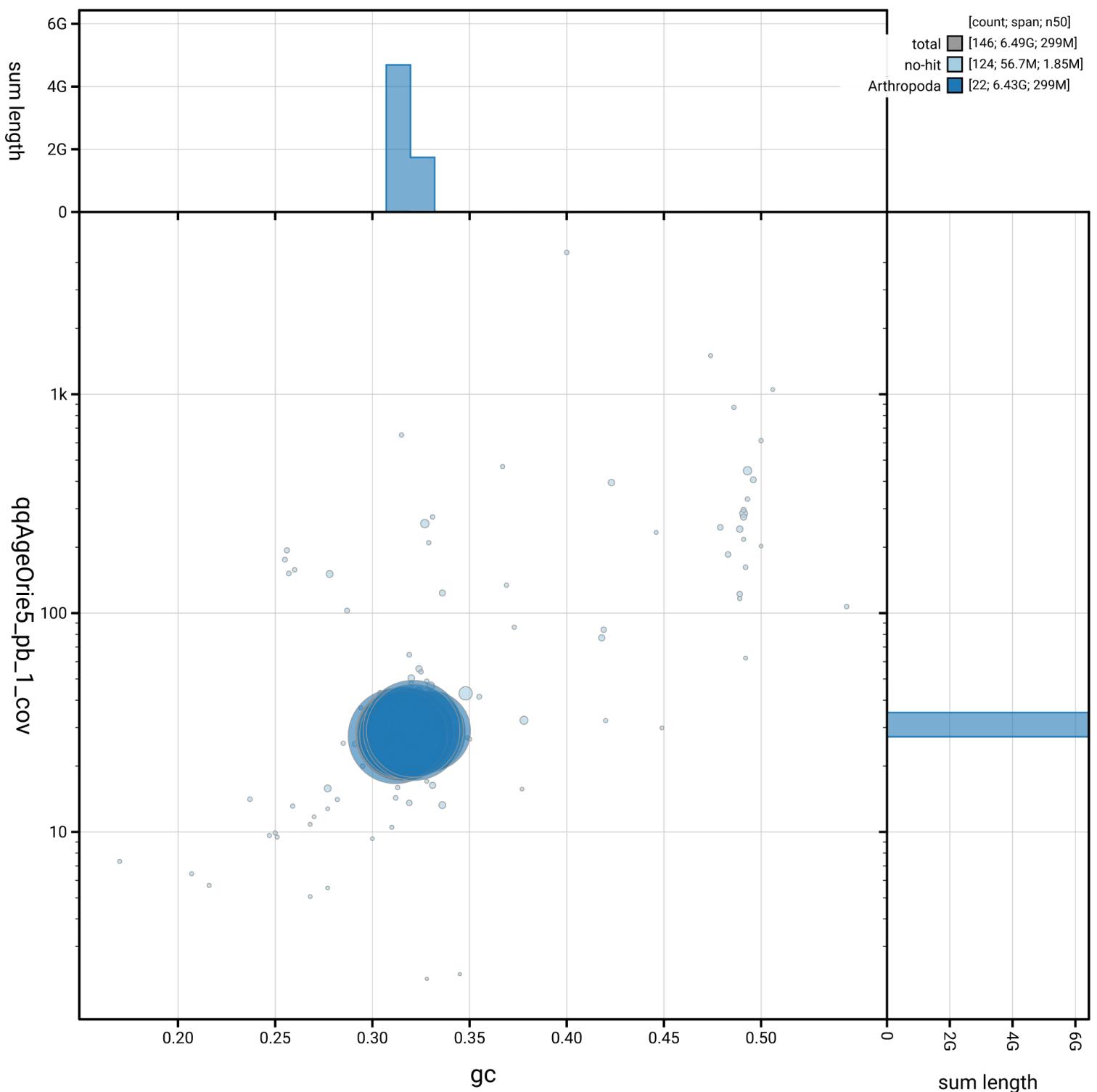
collapsed [\[LINK\]](#)

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	27x	34x

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
- mitchifi
  |_ ver: 3.0.0
  |_ key param: NA
```

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: 2025-12-17 17:31:31 CET