

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	1046855
ToLID	qqAmiAura12
Species	Amilenus aurantiacus
Class	Arachnida
Order	Opiliones

Genome Traits	Expected	Observed
Haploid size (bp)	493,755,031	787,926,943
Haploid Number	12 (source: ancestor)	10
Ploidy	2 (source: ancestor)	2
Sample Sex	XX	unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.7.Q62

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

- . Observed Haploid size (bp) has >20% difference with Expected
- . Observed Haploid Number is different from Expected
- . Observed sex is different from Sample sex
- . Not 90% of assembly in chromosomes for pri

Curator notes

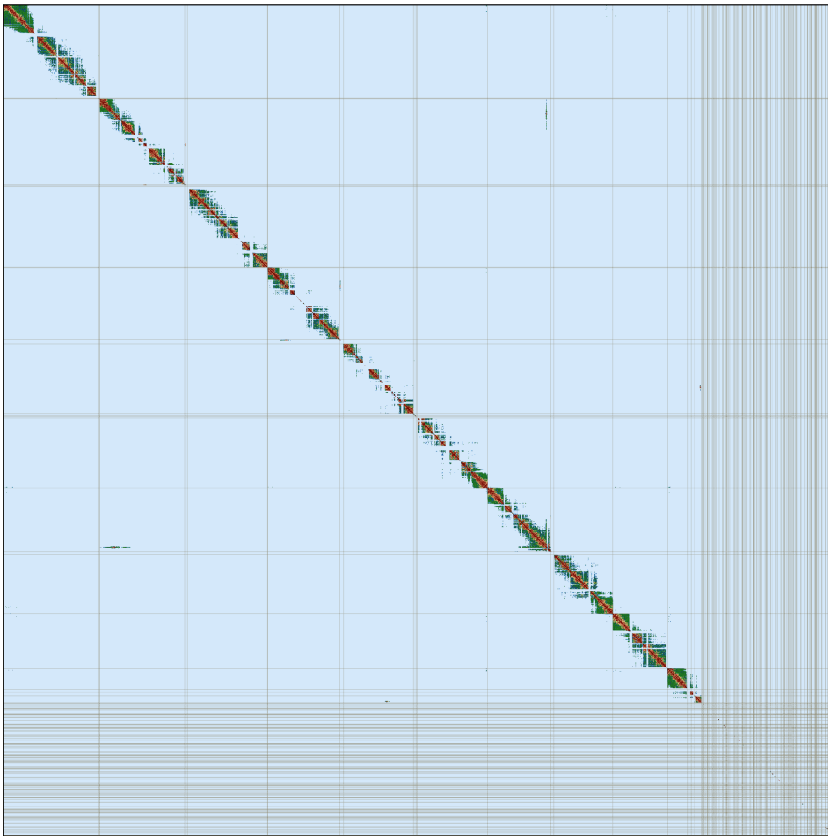
- . Interventions/Gb: 35
- . Contamination notes: "BTK on the primary assembly failed. The run on hap2 is shown. No true contaminants found - some false positives. BTK says Mollusca on a 45M scaffold, but there are more arthropoda hits with higher bitscores than mollusk hits. There are some extreme GC distributions, but related to repeats I think. "
- . Other observations: "Assembly with Nextdenovo leaves out most repetitive regions. QV50. 9Mb contig N50. Hifiiasm assembly with default settings about same contig N50, QV62, more repeats assembled. Curation made one contig break, 4 breaks in gaps and 23 joins. Referred to hap1 and hap2 assemblies as well as MQ0 map to see nature of repeats."

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	792,648,794	787,926,943
GC %	32.78	32.78
Gaps/Gbp	108.5	133.26
Total gap bp	17,200	21,000
Scaffolds	360	340
Scaffold N50	28,513,148	66,091,289
Scaffold L50	9	6
Scaffold L90	56	37
Contigs	446	445
Contig N50	6,101,000	6,101,000
Contig L50	33	33
Contig L90	133	132
QV	62.3124	62.2943
Kmer compl.	96.8274	96.7186
BUSCO sing.		94.8%
BUSCO dupl.		0.5%
BUSCO frag.		3.0%
BUSCO miss.		1.7%

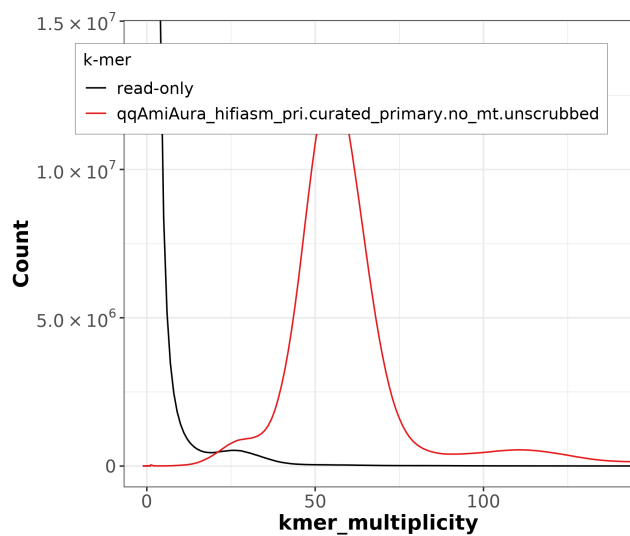
BUSCO: 5.4.0 (euk_genome_met, metaeuk) / Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

HiC contact map of curated assembly

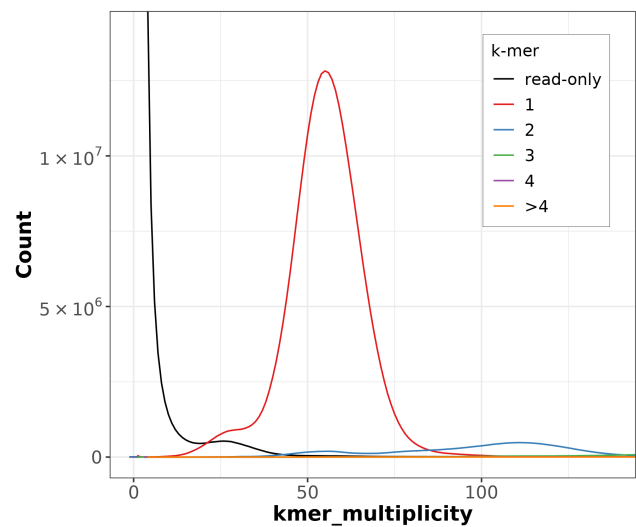


pri [\[LINK\]](#)

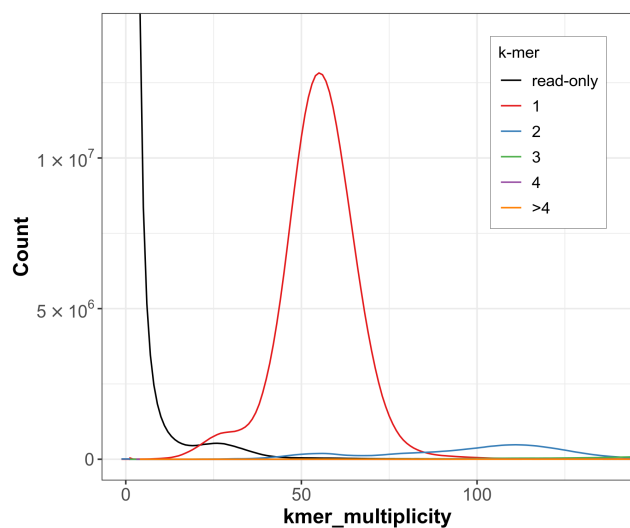
K-mer spectra of curated assembly



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

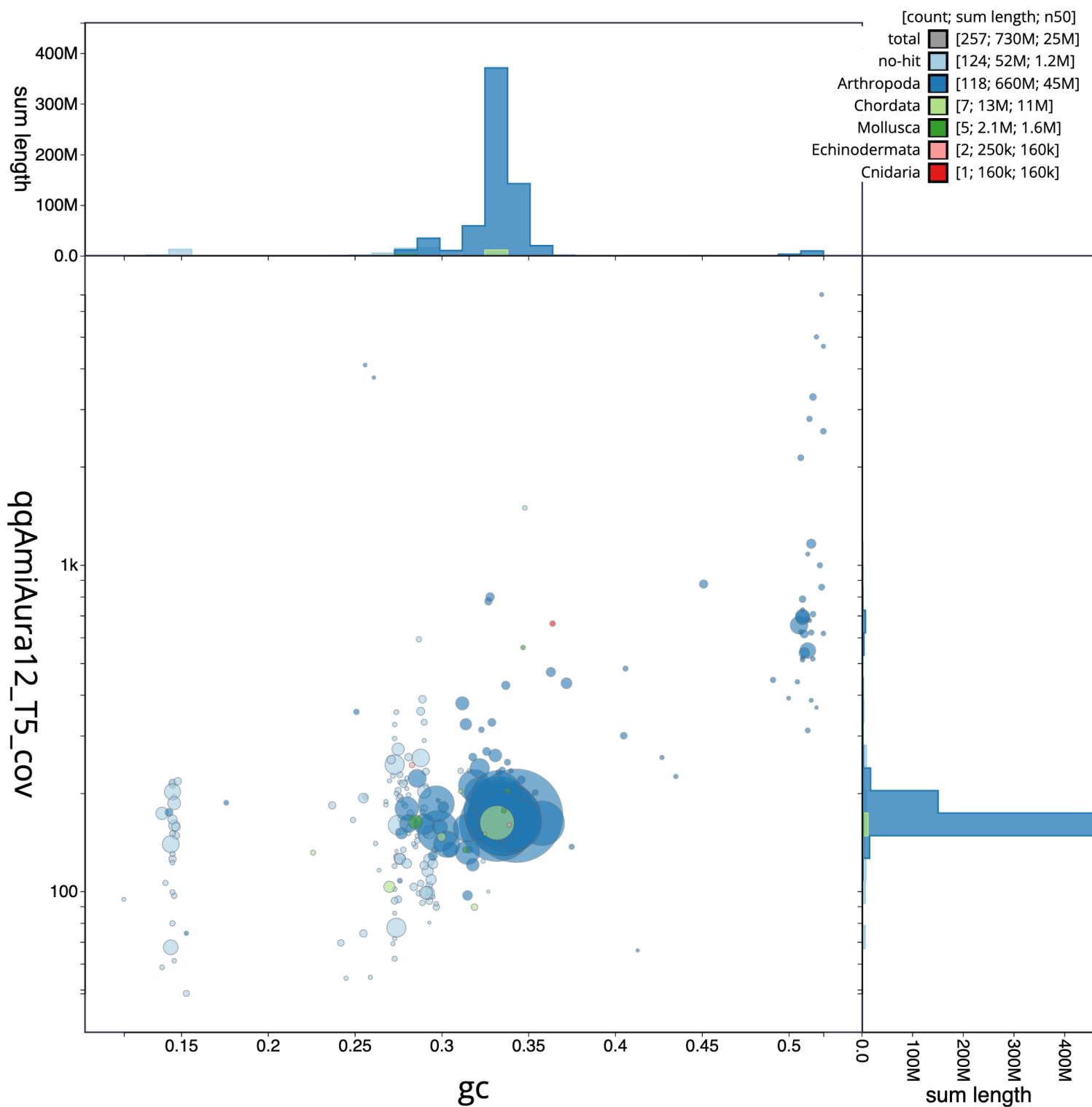


Distribution of k-mer counts per copy numbers found in asm



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

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	Omni-C	Illumina WGS
Coverage	112x	131x	54x

Assembly pipeline

- **Filtlong**
 - |_ *ver*: 0.2.1
 - |_ *key param*: --min_length 1000 --min_mean_q 97 -t 42000000000
- **Hifiasm**
 - |_ *ver*: 0.24.0
 - |_ *key param*: --ont
- **YaHS**
 - |_ *ver*: 1.2a
 - |_ *key param*: NA
- **CLAWS pipeline**
 - |_ *ver*: 2.2.0
 - |_ *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

- **PretextView**
 - |_ *ver*: 0.2.5
 - |_ *key param*: NA
- **GRIT_Rapid**
 - |_ *ver*: 2.0
 - |_ *key param*: NA

Submitter: Tyler Alioto

Affiliation: CNAG

Date and time: 2025-02-21 15:35:23 CET