ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1129383
ToLID	**qmAcaLunu1**
Species	Acanthonyx lunulatus
Class	Malacostraca
Order	Decapoda

Genome Traits	Expected	Observed
Haploid size (bp)	1,518,924,091	1,894,749,031
Haploid Number	47 (source: ancestor)	51
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q43

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

. Kmer completeness value is less than 90 for collapsed
. Assembly length loss > 3% for collapsed
. More than 1000 gaps/Gbp for collapsed
. Not 90% of assembly in chromosomes for collapsed

Curator notes

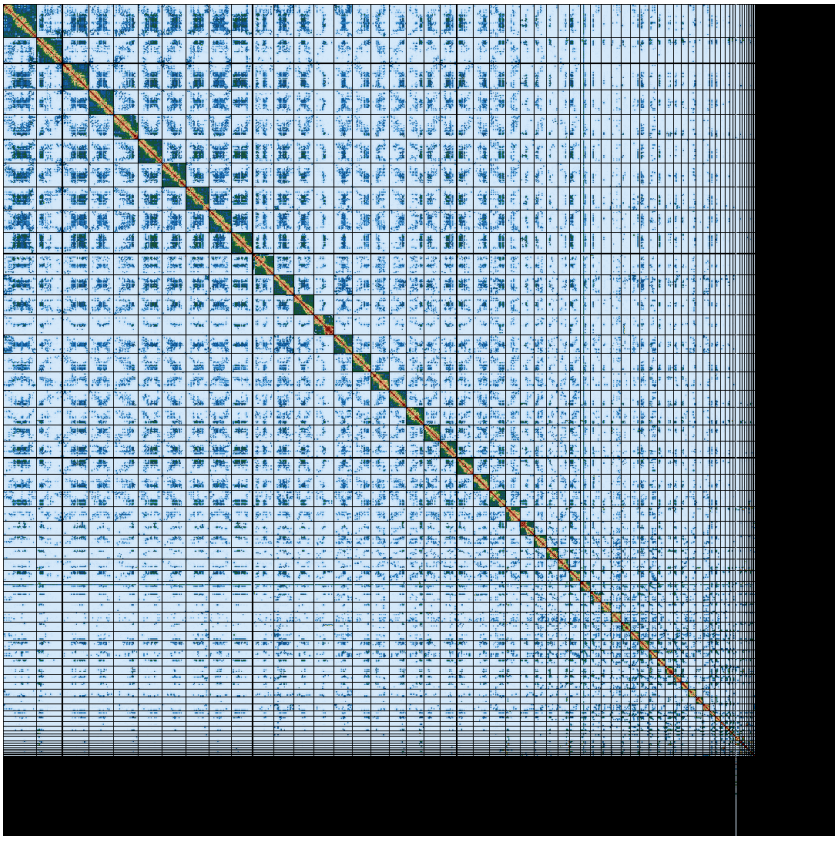
. Interventions/Gb: 55
. Contamination notes: ""
. Other observations: "The assembly of Acanthonyx lunulatus (qmAcaLunu1) is based on 33X ONT data and 124X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial ONT+HiC assembly generation with Hifiasm, removal of contaminant sequences using Context, and Hi-C-based scaffolding with YaHS of each haplotype. In total, 72 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 2.908 Mb (with the largest being 0.173 Mb). Mitochondrial genome was assembled using ptGAUL. Finally, Both haplotypes were analyzed and manually improved using Pretext. During manual curation, 6 haplotypic regions and 6 contaminant sequences were removed, totaling 4.786Mb and 0.415Mb, respectively (with the largest being 1.682Mb and 0.134Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	2,248,316,312	1,894,749,031
GC %	40.8	40.97
Gaps/Gbp	647.15	1,219.69
Total gap bp	145,500	237,900
Scaffolds	658	1,195
Scaffold N50	42,317,519	38,999,931
Scaffold L50	21	19
Scaffold L90	90	60
Contigs	2,113	3,506
Contig N50	2,135,964	1,281,470
Contig L50	283	410
Contig L90	1,071	1,646
QV	43.4704	43.0339
Kmer compl.	79.5856	67.6136
BUSCO sing.	62.8%	96.4%
BUSCO dupl.	37.1%	3.3%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	0.1%	0.2%

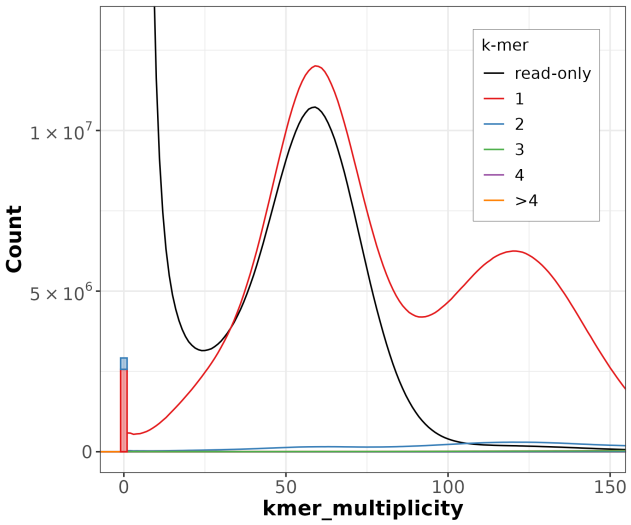
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: crustacea_odb12 (genomes:25, BUSCOs:1536)

HiC contact map of curated assembly

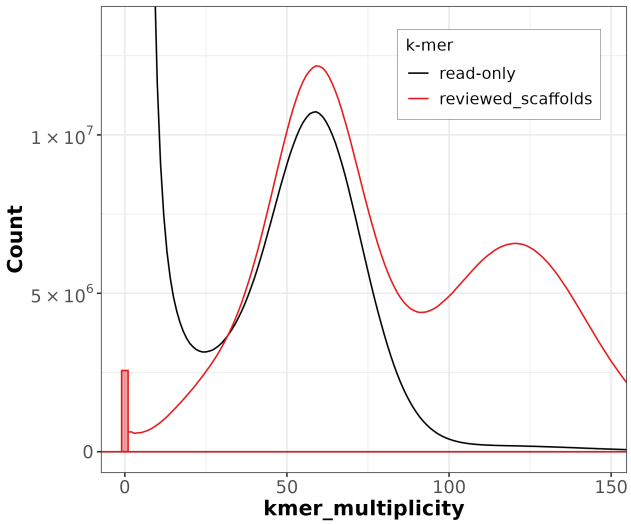


collapsed [LINK]

K-mer spectra of curated assembly

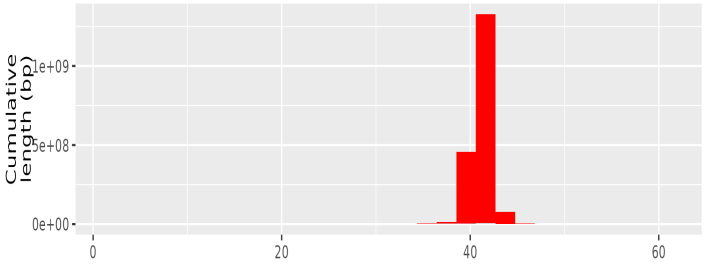


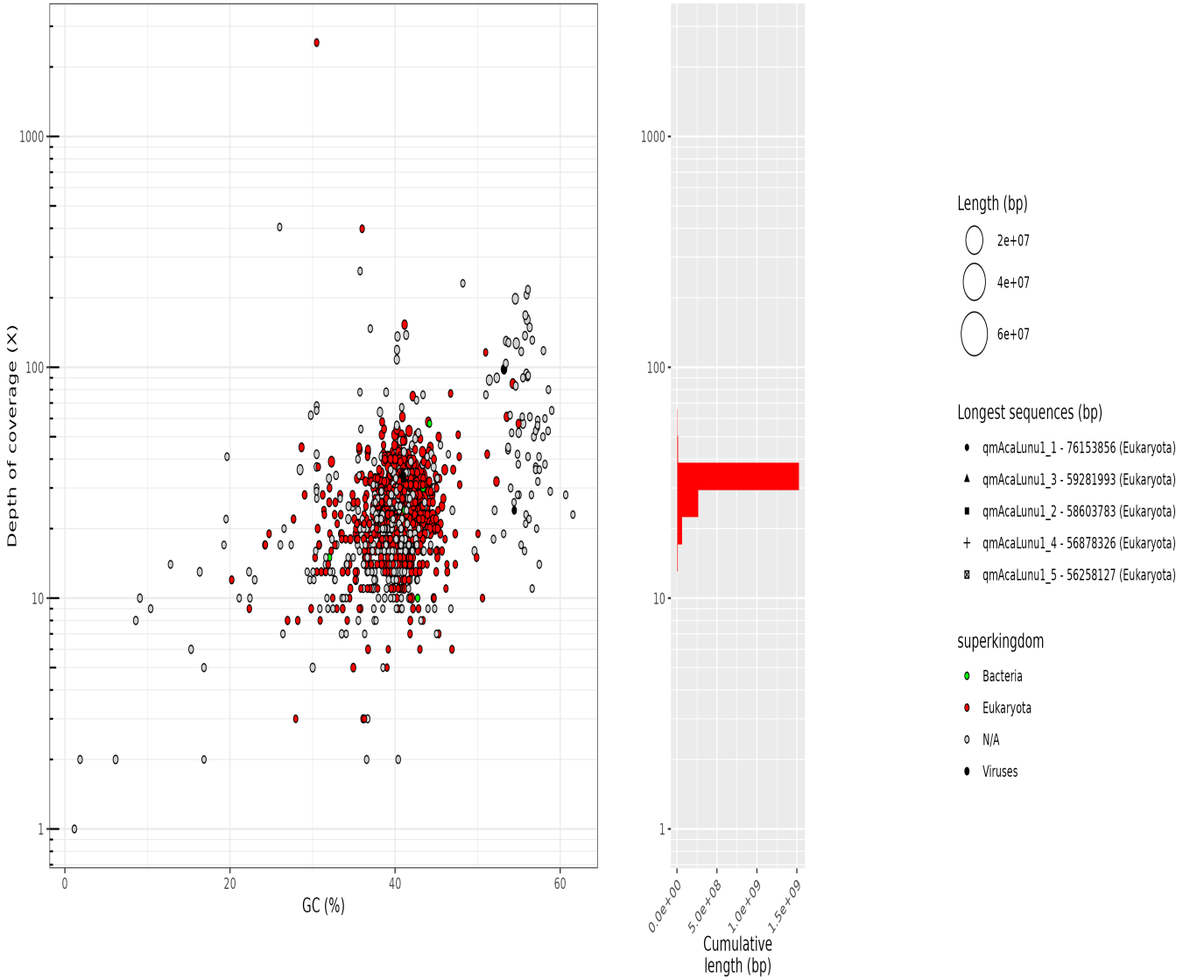
Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening





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	Long reads	Arima
Coverage	39	147

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: Emilie Teodori
Affiliation: Genoscope

Date and time: 2026-06-11 06:53:35 CEST