ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	53602
ToLID	**qmCarAest1**
Species	Carcinus aestuarii
Class	Malacostraca
Order	Decapoda

Genome Traits	Expected	Observed
Haploid size (bp)	828,767,381	1,115,306,513
Haploid Number	52 (source: ancestor)	45
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 5.7.Q27

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

. QV value is less than 40 for collapsed
. Kmer completeness value is less than 90 for collapsed
. BUSCO single copy value is less than 90% for collapsed
. BUSCO duplicated value is more than 5% for collapsed
. More than 1000 gaps/Gbp for collapsed
. Not 90% of assembly in chromosomes for collapsed

Curator notes

. Interventions/Gb: 172
. Contamination notes: ""
. Other observations: "The assembly of Carcinus aestuarii (qmCarAest1) is based on 32X ONT data and 312X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial ONT assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 183 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 14.874 Mb (with the largest being 1.738 Mb). Additionally, 2667 regions totaling 100.363 Mb (with the largest being 0.253 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. During manual curation, 23 haplotypic regions were removed, totaling 19.6Mb

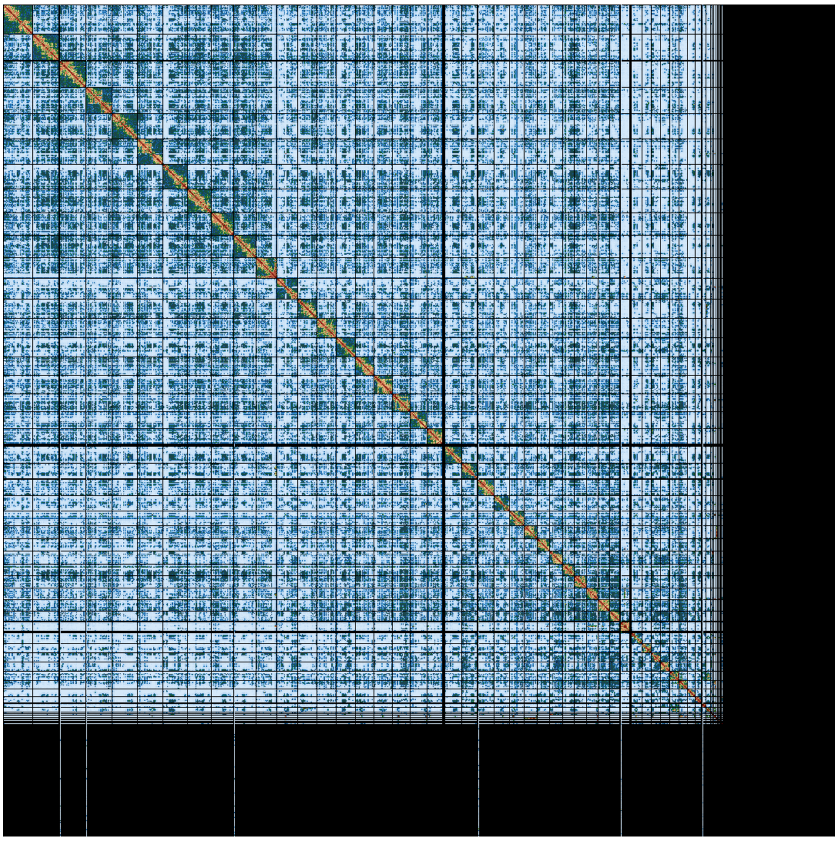
(with the largest being 11.8Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,136,870,084	1,115,306,513
GC %	43.08	43.09
Gaps/Gbp	3,791.99	3,864.41
Total gap bp	431,100	443,600
Scaffolds	1,941	1,607
Scaffold N50	23,015,468	23,281,088
Scaffold L50	19	19
Scaffold L90	210	145
Contigs	6,252	5,917
Contig N50	350,539	347,462
Contig L50	901	887
Contig L90	3,457	3,411
QV	27.3091	27.3208
Kmer compl.	73.7922	72.5458
BUSCO sing.	88.9%	88.2%
BUSCO dupl.	9.1%	10.1%
BUSCO frag.	1.4%	0.3%
BUSCO miss.	0.6%	1.4%

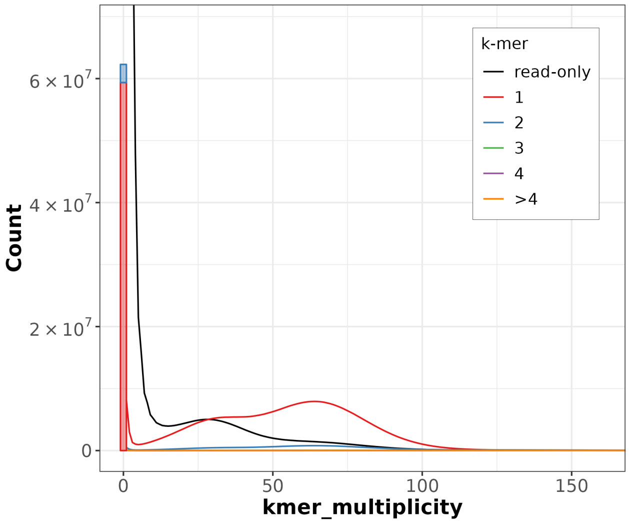
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: crustacea_odb12 (genomes:25, BUSCOs:1536)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: crustacea_odb12 (genomes:25, BUSCOs:1536)

HiC contact map of curated assembly

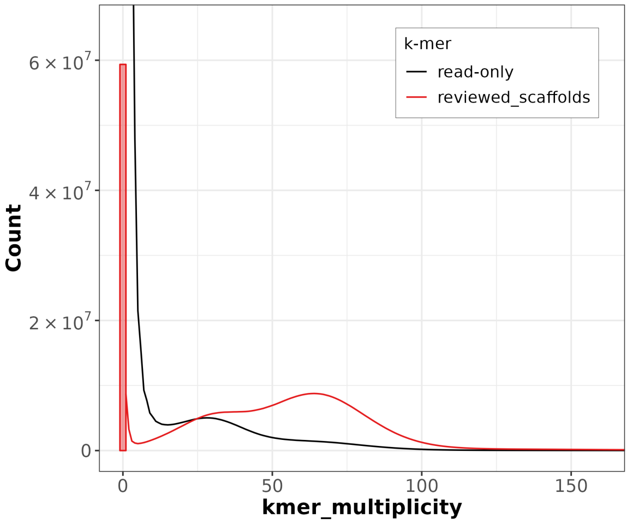


collapsed [LINK]

K-mer spectra of curated assembly

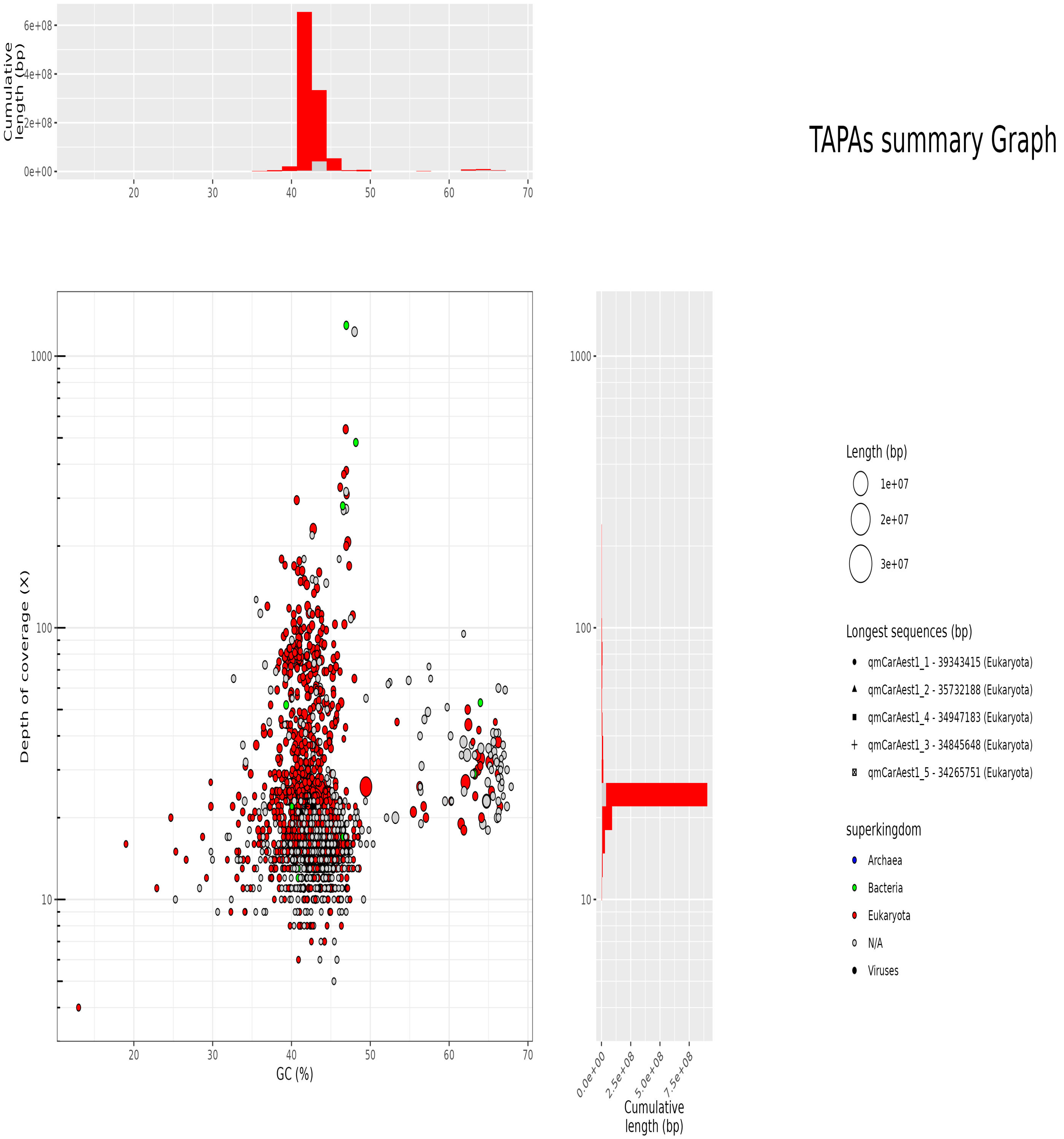


Distribution of k-mer counts per copy numbers found in asm



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	32	312

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: Benjamin Istace
Affiliation: Genoscope

Date and time: 2025-11-28 07:12:28 CET