

# ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	196017
ToLID	<b>qmParAntr1</b>
Species	Parribacus antarcticus
Class	Malacostraca
Order	Decapoda

Genome Traits	Expected	Observed
Haploid size (bp)	2,368,924,890	2,723,517,808
Haploid Number	70 (source: ancestor)	41
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q45

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

- . Observed Haploid Number is different from Expected
- . Not 90% of assembly in chromosomes for collapsed

### Curator notes

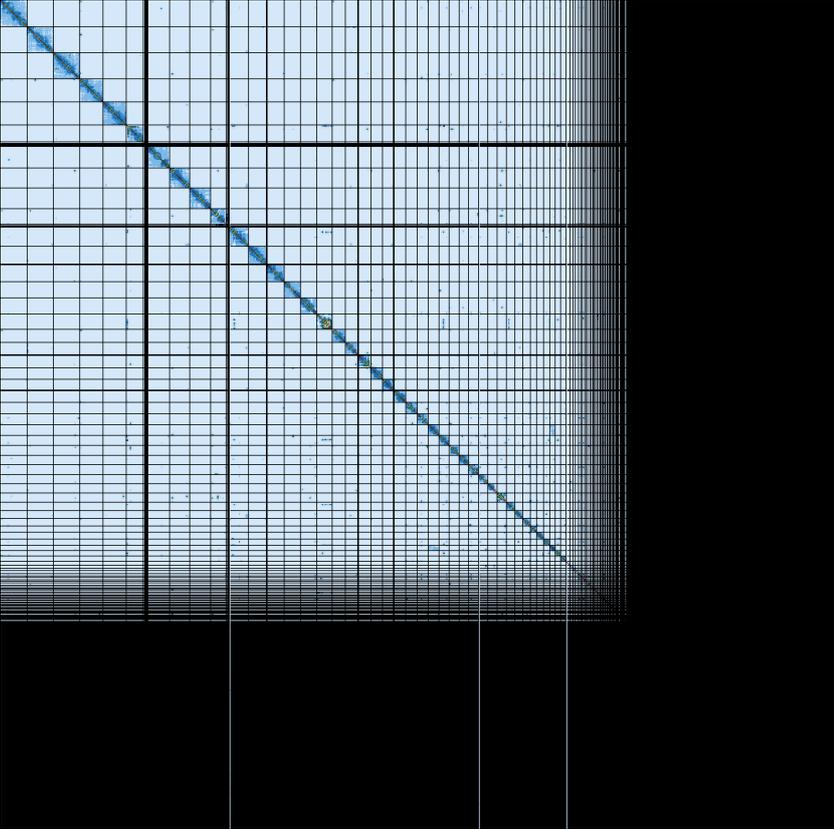
- . Interventions/Gb: 71
- . Contamination notes: ""
- . Other observations: "The assembly of Parribacus antarcticus (qmParAntr1) is based on 30X PacBio data and 103X Arima Hi-C data generated as part of the ATLASea programme (<https://www.atlasea.fr>).The assembly process included the following steps: initial PacBio 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, 110 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 13.6 Mb (with the largest being 3.2Mb). Additionally, 2,079 regions totaling 272 Mb (with the largest being 1.3 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using oatk. The obtained sequence is linear. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 22 supplementary haplotypic regions, totaling 19.2 Mb (with the largest being 2.3 Mb) were removed. Certain regions showing haplotype profiles have been left out because the long reads coverage has an atypical profile. The genome contains large un-scaffolded and un-placed repetitive regions. 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	2,742,685,673	2,723,517,808
GC %	41.31	41.32
Gaps/Gbp	293.51	334.49
Total gap bp	80,500	105,300
Scaffolds	1,849	1,736
Scaffold N50	29,212,696	35,170,851
Scaffold L50	28	24
Scaffold L90	499	397
Contigs	2,654	2,647
Contig N50	2,862,283	2,889,888
Contig L50	230	229
Contig L90	1,053	1,047
QV	45.7704	45.7642
Kmer compl.	90.22	90.121
BUSCO sing.	98.1%	99.2%
BUSCO dupl.	1.8%	0.7%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	0.1%	0.1%

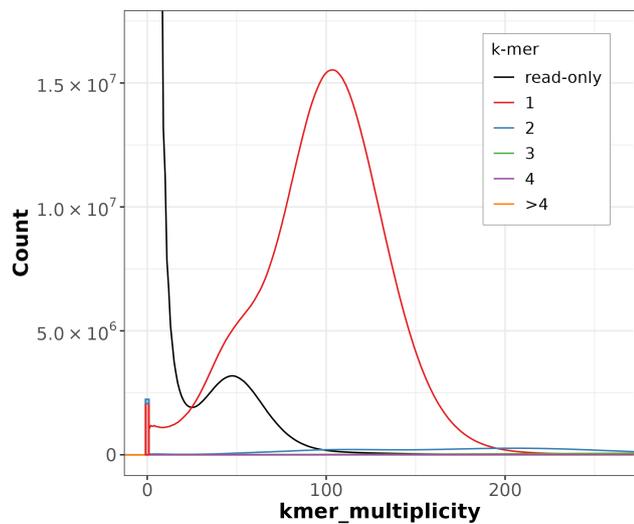
BUSCO: 6.0.0 (euk\_genome\_min, miniprot) / Lineage: crustacea\_odb12 (genomes:25, BUSCOs:1536)

# HiC contact map of curated assembly

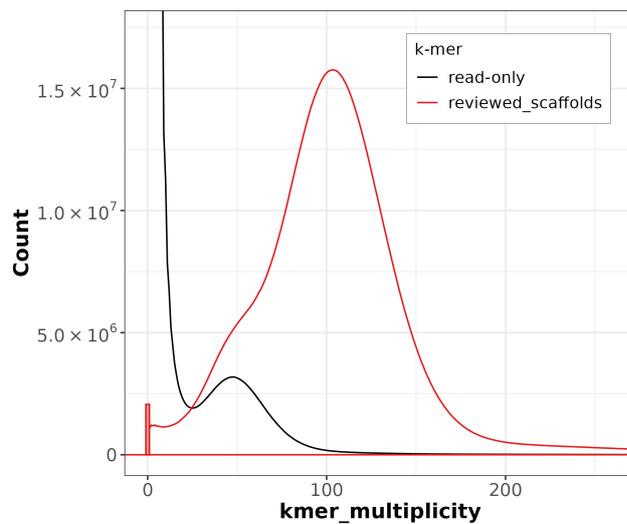


collapsed [\[LINK\]](#)

# K-mer spectra of curated assembly

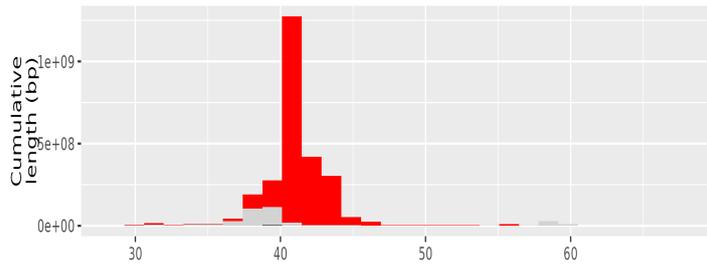


Distribution of k-mer counts per copy numbers found in asm

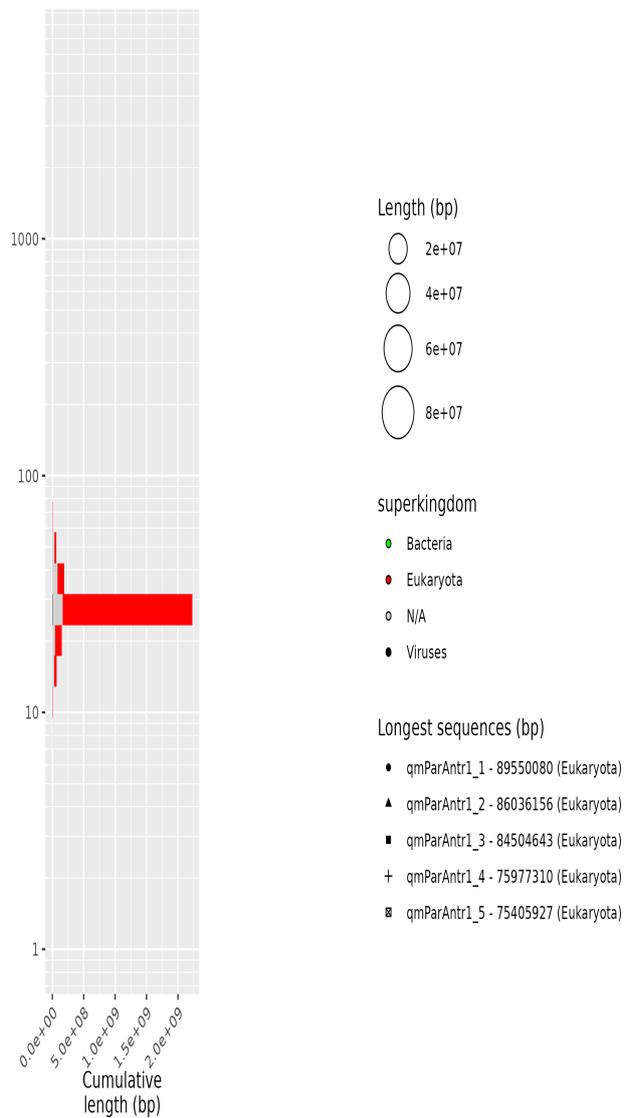
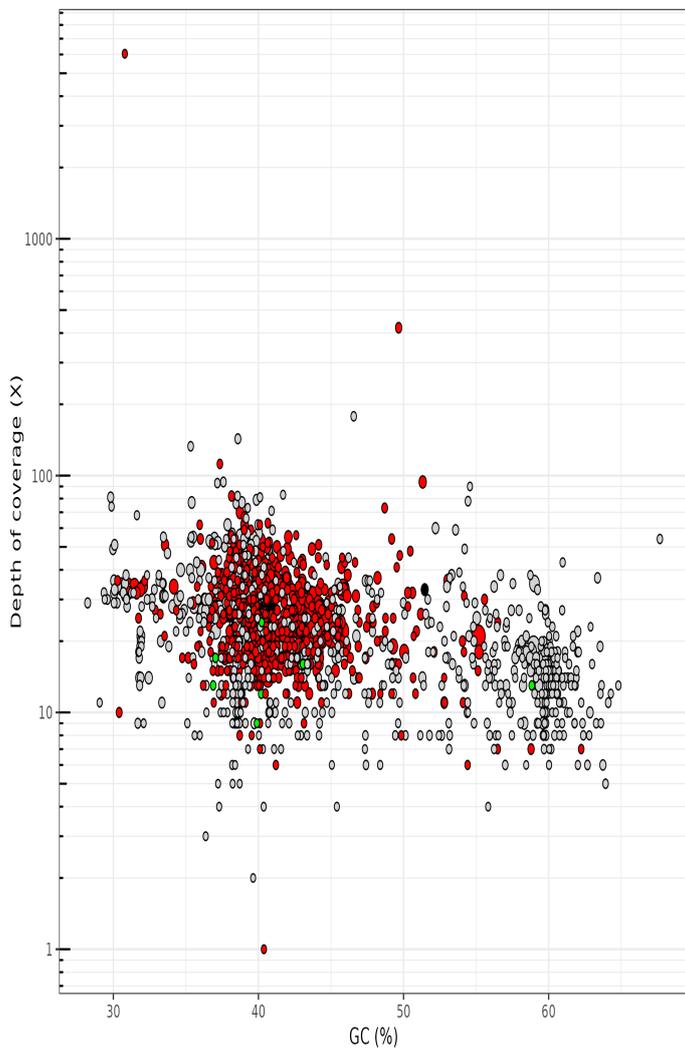


Distribution of k-mer counts coloured by their presence in reads/assemblies

# Post-curation contamination screening



## TAPAs summary Graph



**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	29	103

# 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: Caroline Menguy

Affiliation: Genoscope

Date and time: 2026-03-01 18:17:55 CET