ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1844856
ToLID	**qmCalOcel1**
Species	Calappa ocellata
Class	Malacostraca
Order	Decapoda

Genome Traits	Expected	Observed
Haploid size (bp)	983,207,047	1,091,462,369
Haploid Number	47 (source: ancestor)	37
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 Number is different from Expected

. Kmer completeness value is less than 90 for collapsed
. Not 90% of assembly in chromosomes for collapsed

Curator notes

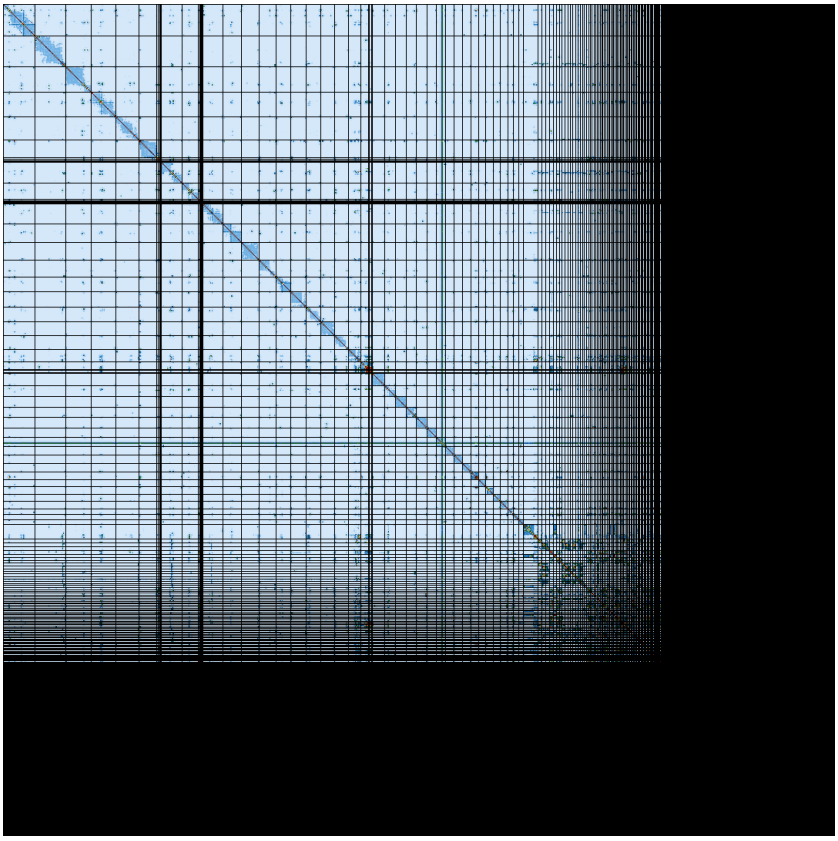
. Interventions/Gb: 107
. Contamination notes: ""
. Other observations: "The assembly of Calappa ocellata (qmCalOcel1) is based on 44.17X ONT data and 159.45X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 4 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.328 Mb (with the largest being 0.218 Mb). Additionally, 567 regions totaling 106.814 Mb (with the largest being 0.56 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using ptGAUL. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 29 haplotypic regions were removed, totaling 8.015Mb (with the largest being 0.759Mb). The telomeric pattern AACCT was identified with TelFinder and used to generate the Pretext Telomeres tracks. 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	1,099,549,814	1,091,462,369
GC %	42.88	42.87
Gaps/Gbp	273.75	295.93
Total gap bp	30,100	37,200
Scaffolds	927	916
Scaffold N50	11,508,002	13,314,000
Scaffold L50	26	24
Scaffold L90	293	285
Contigs	1,228	1,239
Contig N50	4,847,168	4,847,168
Contig L50	62	62
Contig L90	471	475
QV	43.7362	43.7329
Kmer compl.	87.8426	87.6666
BUSCO sing.	97.9%	98.2%
BUSCO dupl.	1.4%	1.1%
BUSCO frag.	0.4%	0.4%
BUSCO miss.	0.2%	0.2%

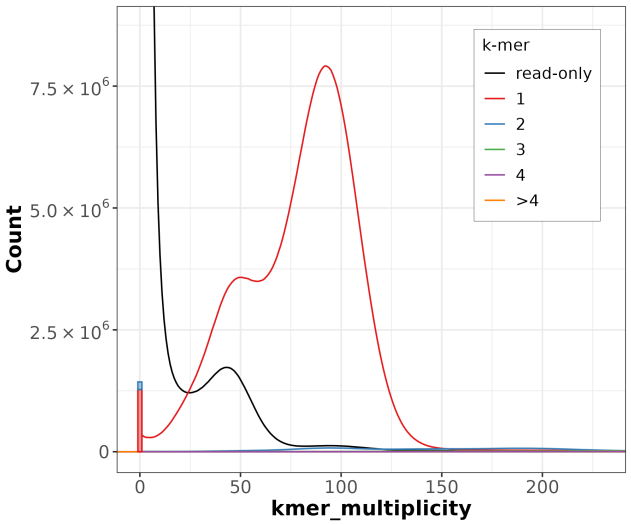
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: arthropoda_odb12 (genomes:76, BUSCOs:1667)

HiC contact map of curated assembly

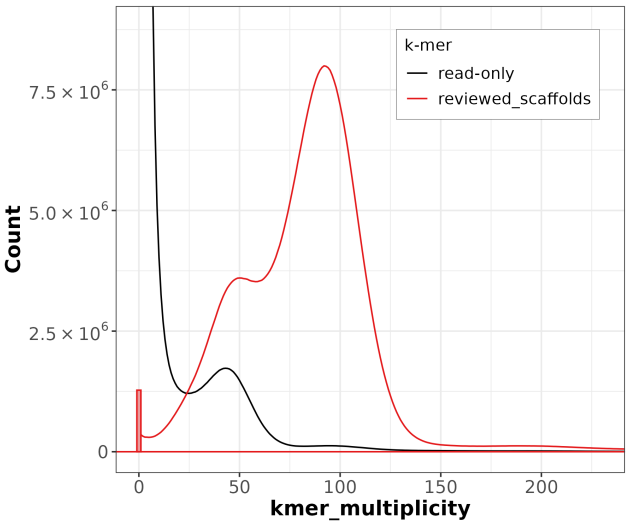


collapsed [LINK]

K-mer spectra of curated assembly

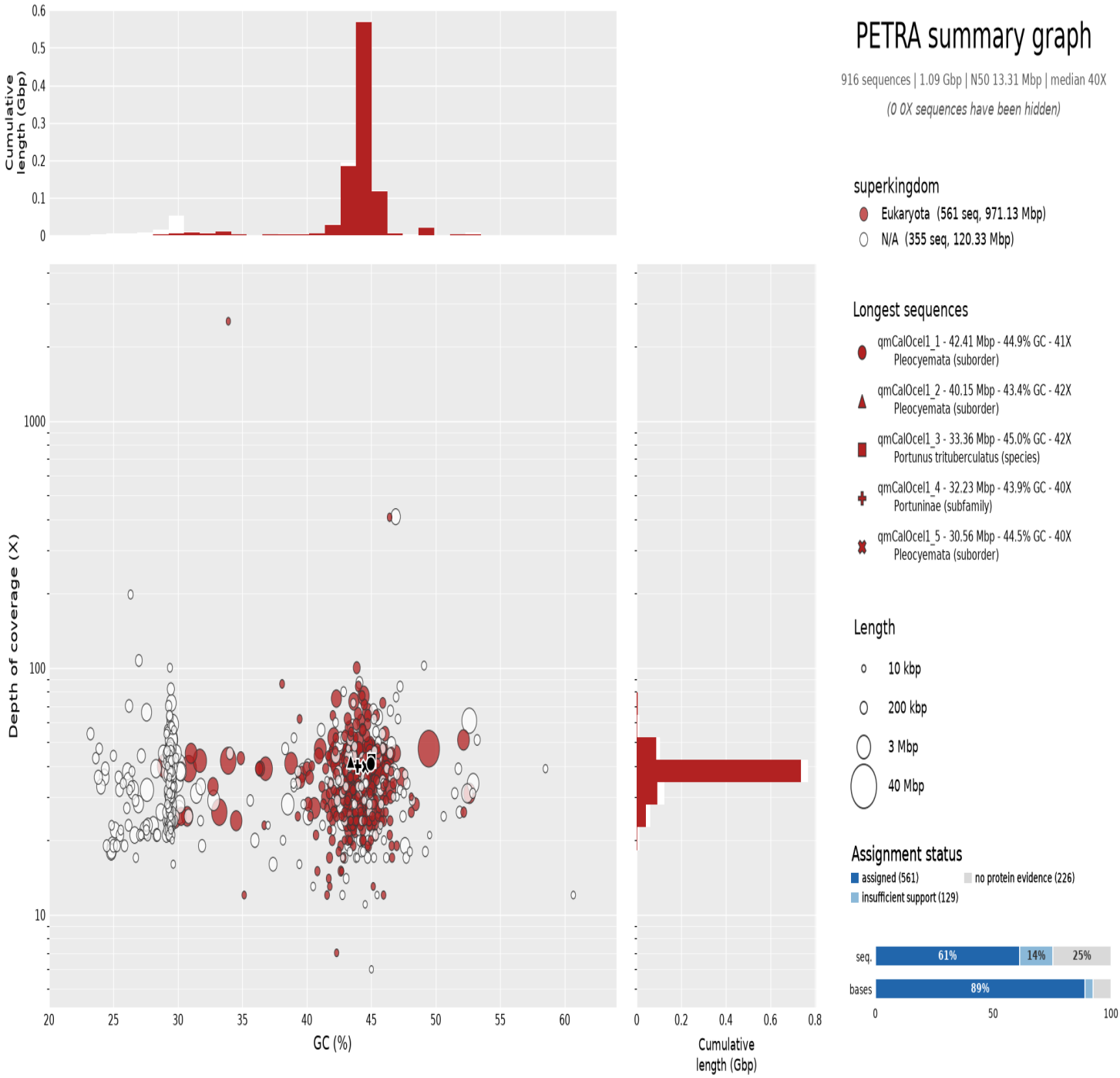


Distribution of k-mer counts per copy numbers found in asm



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	44	160

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:* 0.2.4
  |_ *key param:* NA
- **PretextView**
  |_ *ver:* 1.0.7
  |_ *key param:* NA

Submitter: Eleanore Lacoste
Affiliation: Genoscope

Date and time: 2026-09-29 10:45:06 CEST