ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	121182
ToLID	**xgClaCruc1**
Species	Clanculus cruciatus
Class	Gastropoda
Order	Trochida

Genome Traits	Expected	Observed
Haploid size (bp)	1,442,321,658	1,488,842,268
Haploid Number	9 (source: ancestor)	18
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q38

The following metrics were automatically flagged as below EBP recommended standards or different from 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
. Assembly length loss > 3% for collapsed

Curator notes

. Interventions/Gb: 63
. Contamination notes: ""
. Other observations: "The assembly of Clanculus cruciatus (xgClaCruc1) is based on 60X PacBio data and 180X 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, 238 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 12.548 Mb (with the largest being 0.329 Mb). Additionally, 827 regions totaling 120.322 Mb (with the largest being 7.005 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 33 haplotypic regions were removed, totaling 99.716Mb (with the largest being 11.058Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. Chromosomes 2 and 9 might have required a dual curation, they contain more gaps than the rest of the chromosomes and

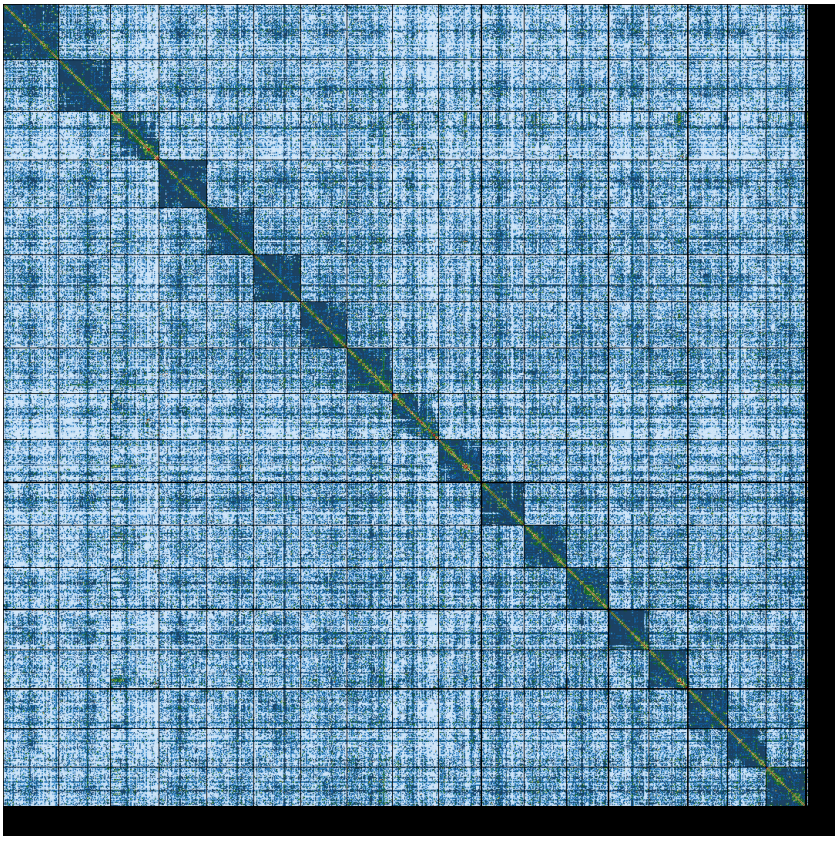
uncertainties remain regarding the presence of haplotypic duplications for those two chromosomes. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,608,235,278	1,488,842,268
GC %	37.55	37.49
Gaps/Gbp	240.64	113.51
Total gap bp	38,700	20,600
Scaffolds	1,797	1,136
Scaffold N50	82,298,841	82,025,990
Scaffold L50	10	9
Scaffold L90	18	17
Contigs	2,184	1,305
Contig N50	19,688,000	21,928,000
Contig L50	22	20
Contig L90	106	82
QV	36.1478	38.6626
Kmer compl.	64.8988	62.2077
BUSCO sing.	94.1%	98.3%
BUSCO dupl.	5.2%	1.0%
BUSCO frag.	0.2%	0.2%
BUSCO miss.	0.5%	0.5%

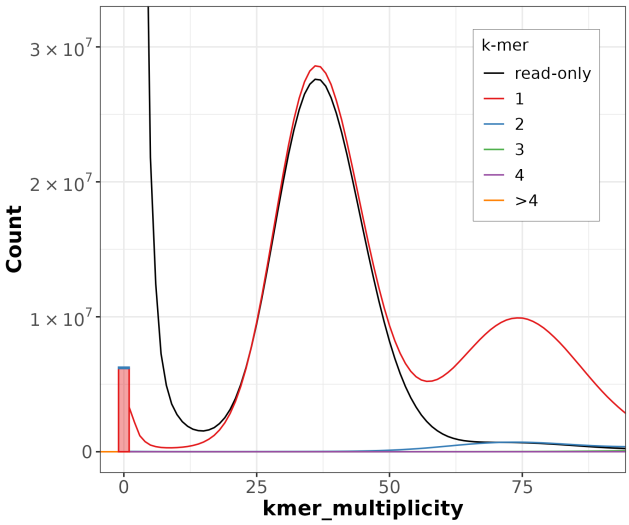
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

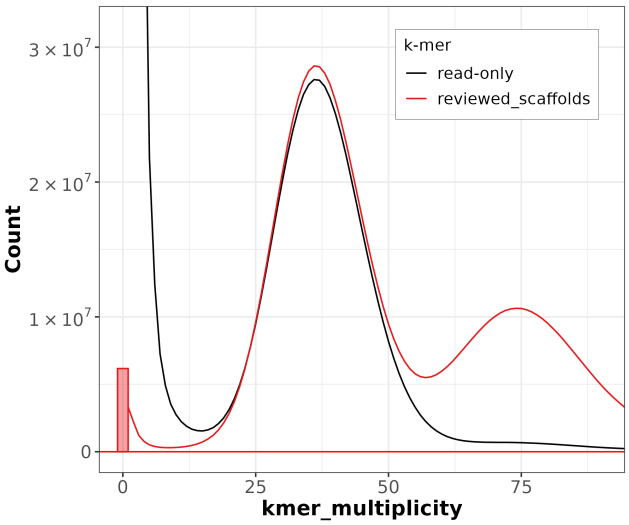


collapsed [LINK]

K-mer spectra of curated assembly

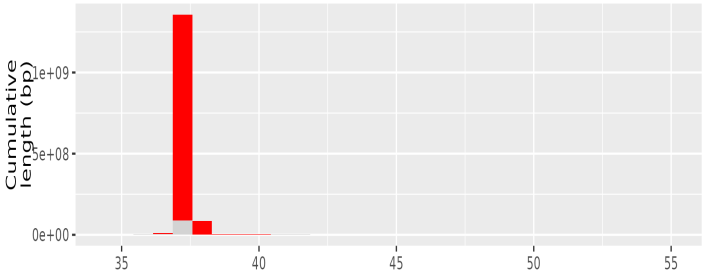


Distribution of k-mer counts per copy numbers found in asm

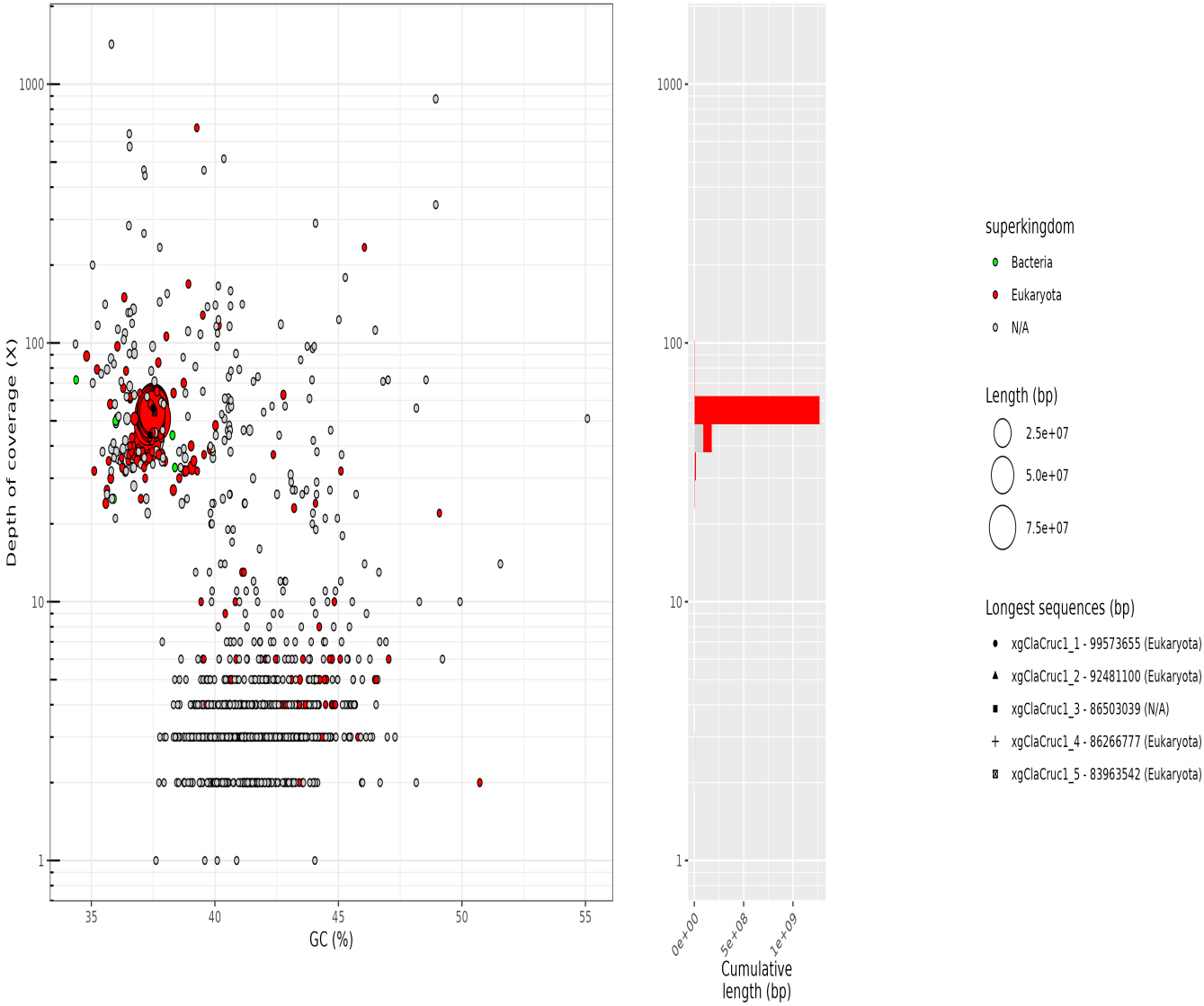


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	65	194

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: 2026-06-23 03:35:00 CEST