ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1028565
ToLID	**xbArcCras1**
Species	Arcopagia crassa
Class	Bivalvia
Order	Cardiida

Genome Traits	Expected	Observed
Haploid size (bp)	1,844,185,359	1,959,261,269
Haploid Number	26 (source: ancestor)	19
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q48

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

Curator notes

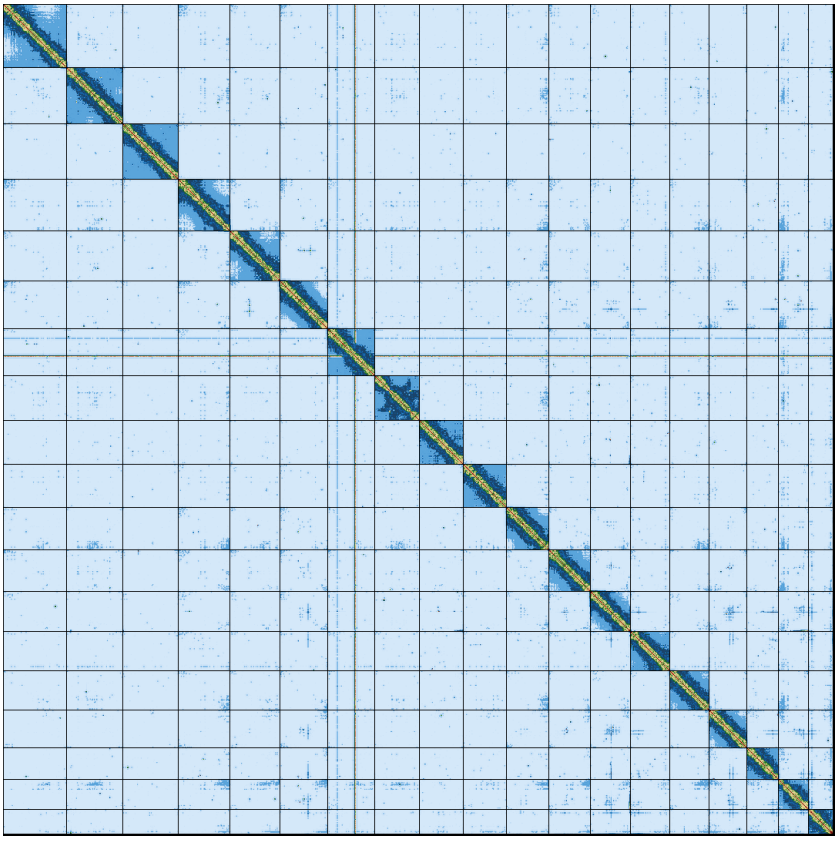
. Interventions/Gb: 14
. Contamination notes: ""
. Other observations: "The assembly of Arcopagia crassa (xbArcCras1.1) is based on 42X PACBIO data and 188X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 7 contigs were identified as contaminants (bacterial), totaling 0.635 Mb (with the largest being 0.312 Mb). Additionally, 152 regions totaling 18.238 Mb (with the largest being 2.306 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 2 haplotypic region and 2 contaminant sequences were removed, totaling 2.1 Mb and 0.218 Mb, respectively (with the largest being 1.256 Mb and 0.122 Mb). 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	1,961,577,604	1,959,261,269
GC %	39.05	39.05
Gaps/Gbp	31.61	33.18
Total gap bp	6,200	7,400
Scaffolds	59	55
Scaffold N50	106,893,526	104,777,630
Scaffold L50	8	8
Scaffold L90	17	17
Contigs	121	120
Contig N50	43,718,000	43,718,000
Contig L50	14	14
Contig L90	43	43
QV	48.5084	48.5098
Kmer compl.	71.2513	71.1984
BUSCO sing.	97.9%	97.8%
BUSCO dupl.	0.9%	0.8%
BUSCO frag.	0.7%	0.7%
BUSCO miss.	0.5%	0.7%

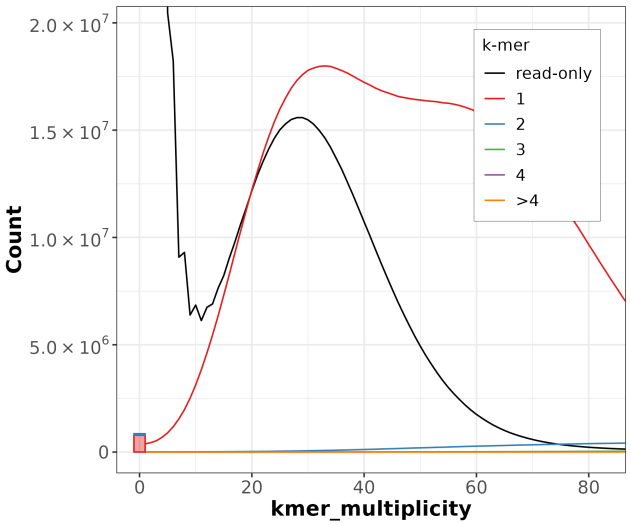
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

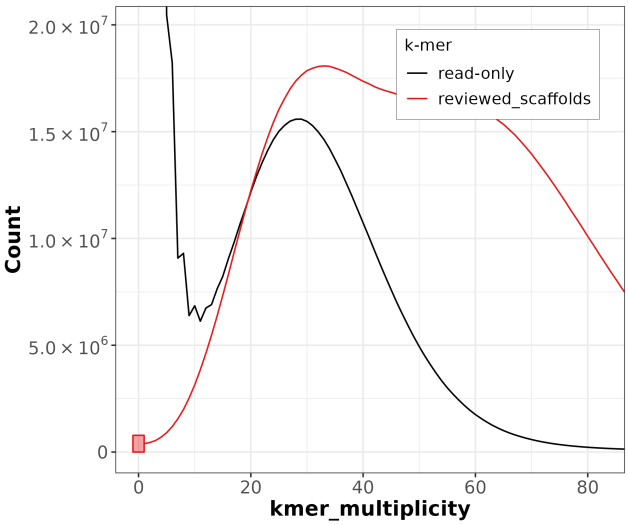


collapsed [LINK]

K-mer spectra of curated assembly

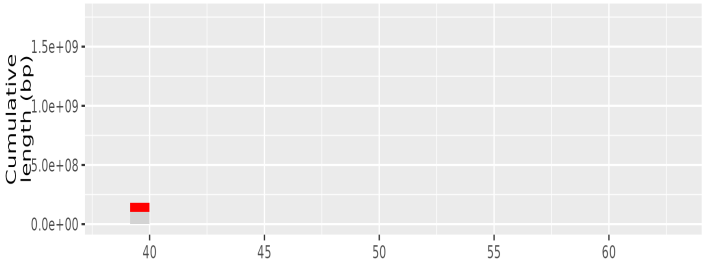


Distribution of k-mer counts per copy numbers found in asm

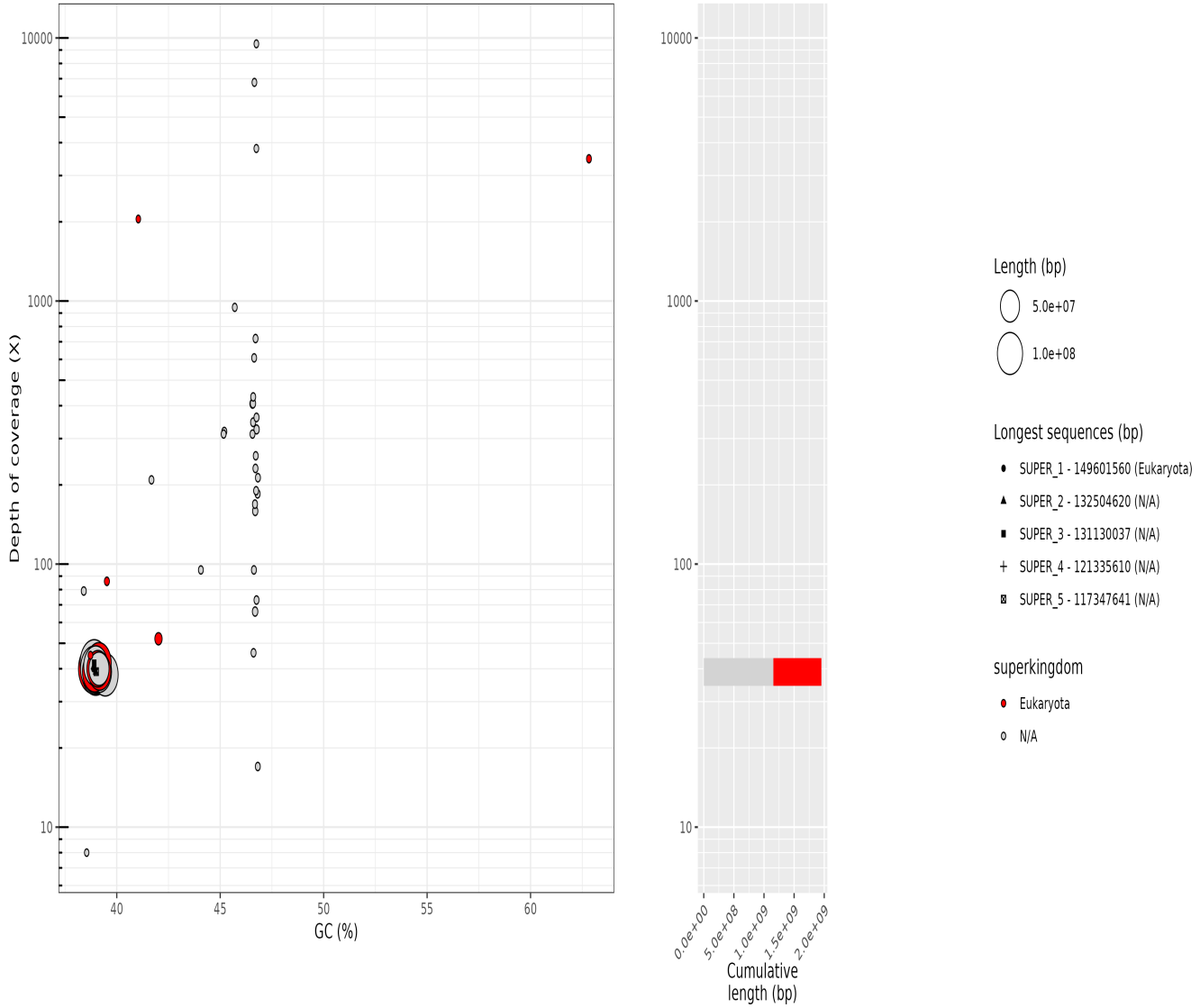


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	42	188

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: Adama Ndar
Affiliation: Genoscope

Date and time: 2025-10-29 13:41:22 CET