ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	549391
ToLID	**xcAllMedi1**
Species	Alloteuthis media
Class	Cephalopoda
Order	Myopsida

Genome Traits	Expected	Observed
Haploid size (bp)	3,383,992,601	4,129,663,161
Haploid Number	6 (source: ancestor)	48
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q36

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
. Assembly length loss > 3% for collapsed

Curator notes

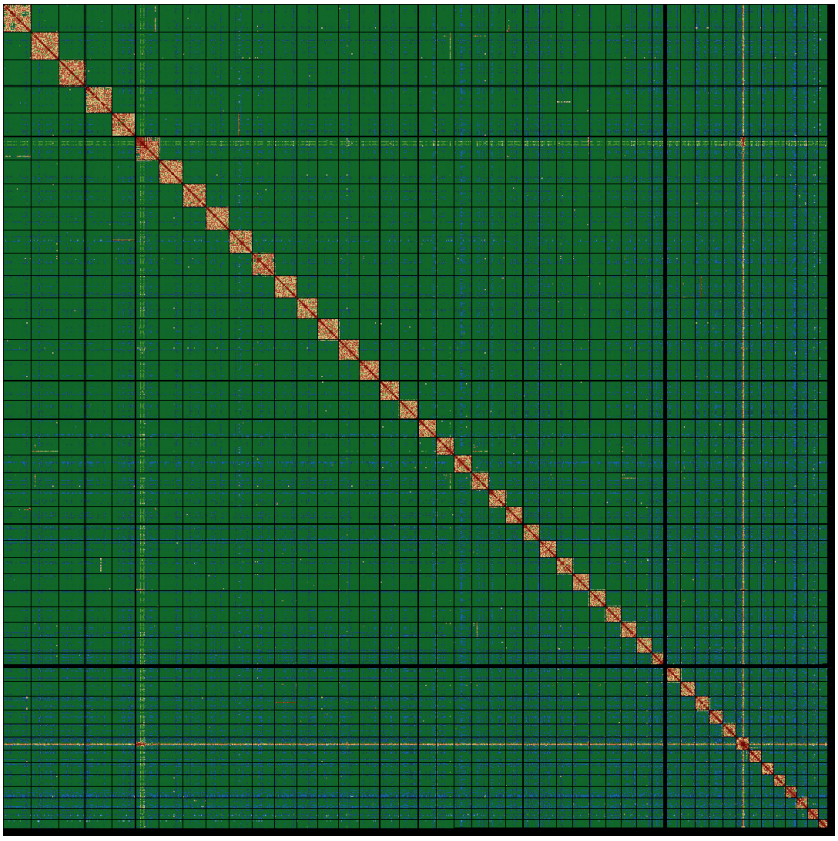
. Interventions/Gb: 45
. Contamination notes: ""
. Other observations: "The assembly of Alloteuthis media (xcAllMedi1) is based on 40X ONT data and 127X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial ONT 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, 6 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.176 Mb (with the largest being 0.08 Mb). Additionally, 1812 regions totaling 363.748 Mb (with the largest being 2.376 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 184 haplotypic regions were removed, totaling 278.975Mb (with the largest being 6.398Mb). 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	4,417,360,938	4,129,663,161
GC %	32.8	32.8
Gaps/Gbp	393.67	407.78
Total gap bp	173,900	178,200
Scaffolds	747	581
Scaffold N50	86,420,237	88,631,987
Scaffold L50	21	19
Scaffold L90	51	40
Contigs	2,486	2,265
Contig N50	4,588,143	4,886,672
Contig L50	280	252
Contig L90	1,073	927
QV	36.327	36.4273
Kmer compl.	72.2036	69.6655
BUSCO sing.	90.7%	91.2%
BUSCO dupl.	2.7%	1.3%
BUSCO frag.	3.0%	3.0%
BUSCO miss.	3.6%	4.4%

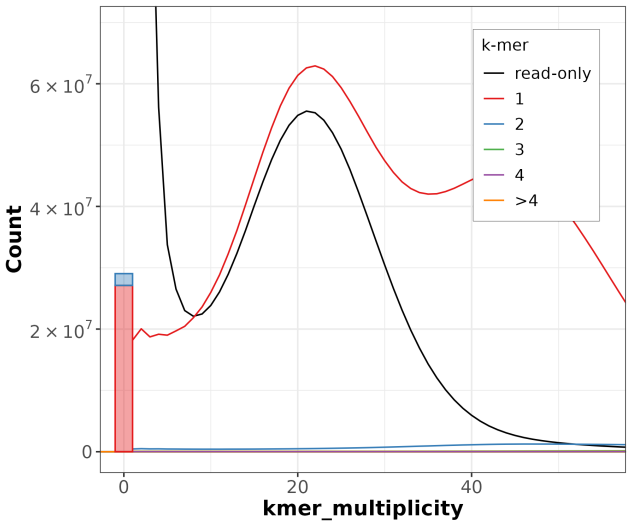
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

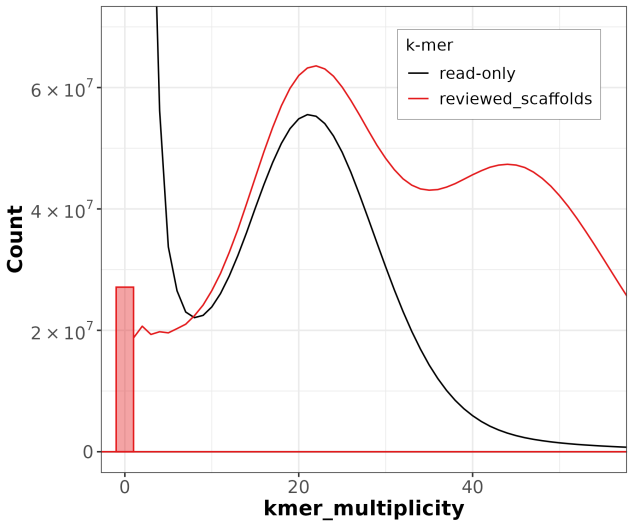


collapsed [LINK]

K-mer spectra of curated assembly

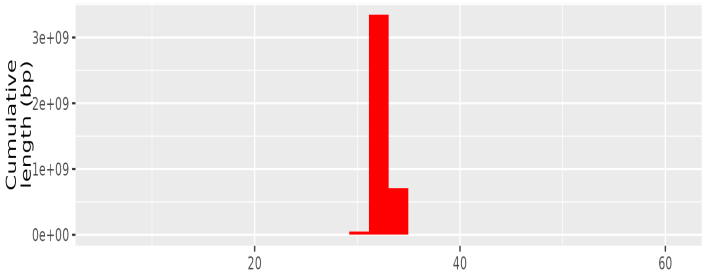


Distribution of k-mer counts per copy numbers found in asm



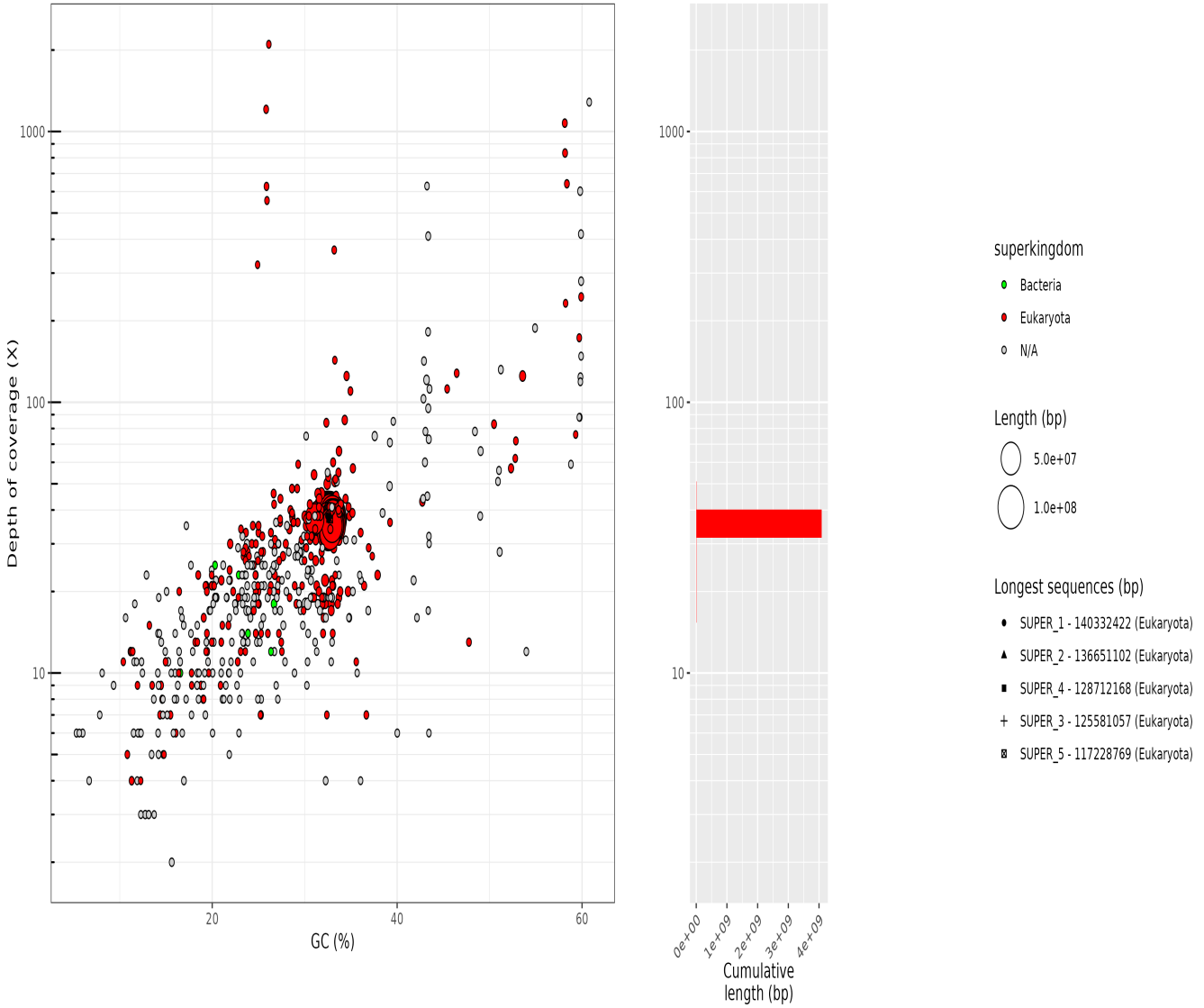
Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph

(6 0X contigs have been hidden)



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	136

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: 2026-06-29 19:19:48 CEST