

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	218040
ToLID	xgLurLuri1
Species	Luria lurida
Class	Gastropoda
Order	Littorinimorpha

Genome Traits	Expected	Observed
Haploid size (bp)	1,355,810,243	1,481,276,304
Haploid Number	17 (source: ancestor)	36
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q44

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
- . BUSCO single copy value is less than 90% for collapsed
- . BUSCO duplicated value is more than 5% for collapsed

Curator notes

. Interventions/Gb: 2
. Contamination notes: ""
. Other observations: "The assembly of *Luria lurida* (xgLurLuri1) is based on 42X ONT data and 159X 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 Hifiiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 2 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.24 Mb (with the largest being 0.162 Mb). Additionally, 48 regions totaling 17.882 Mb (with the largest being 4.353 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, 1 haplotypic region was removed, totaling 0.513Mb. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

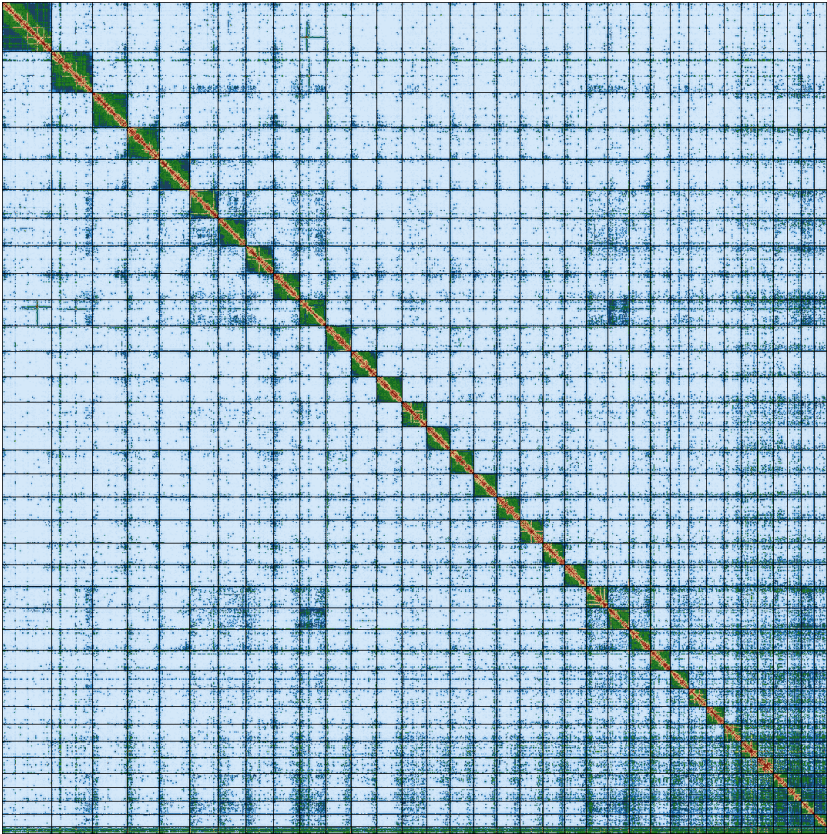
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,481,772,862	1,481,276,304
GC %	42.44	42.44
Gaps/Gbp	6.75	6.75
Total gap bp	1,000	1,100
Scaffolds	38	39
Scaffold N50	43,927,190	43,927,190
Scaffold L50	14	14
Scaffold L90	30	30
Contigs	48	49
Contig N50	41,772,086	41,772,086
Contig L50	15	15
Contig L90	32	32
QV	44.7543	44.7515
Kmer compl.	75.0689	75.061
BUSCO sing.	76.5%	75.3%
BUSCO dupl.	21.3%	19.9%
BUSCO frag.	0.9%	3.2%
BUSCO miss.	1.3%	1.6%

Warning! BUSCO versions or lineage datasets are not the same across results:

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

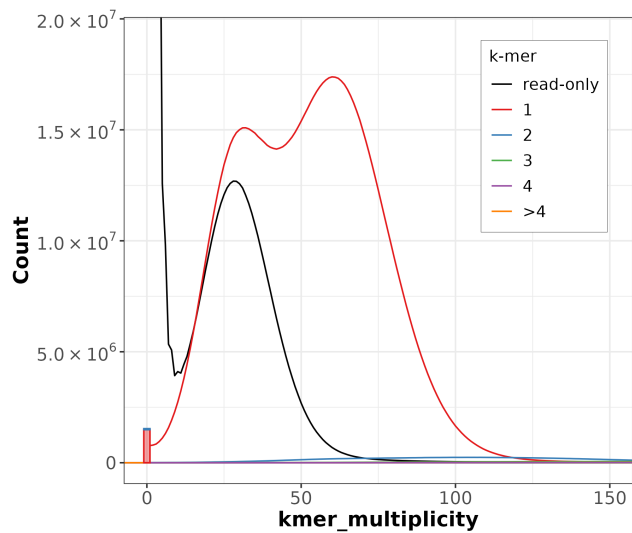
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

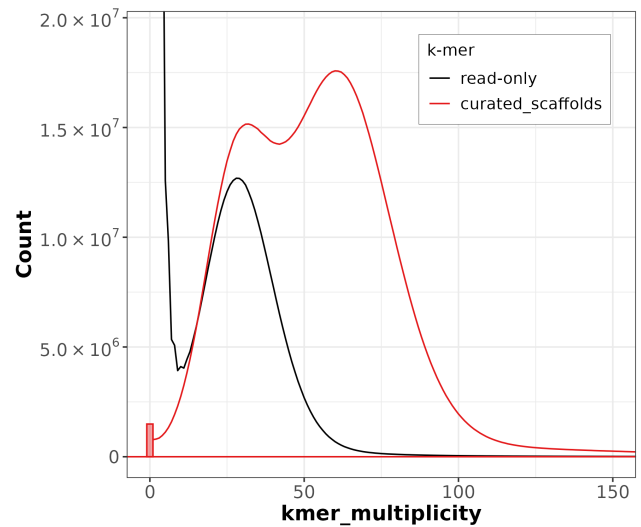


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

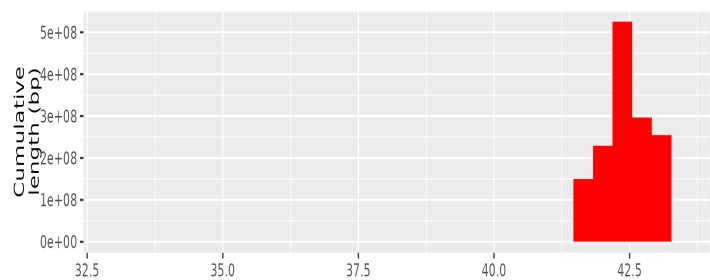


Distribution of k-mer counts per copy numbers found in asm

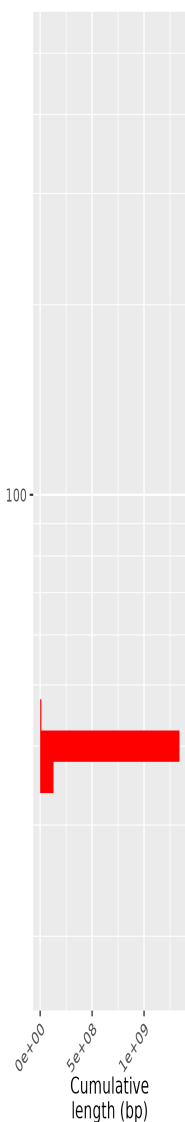
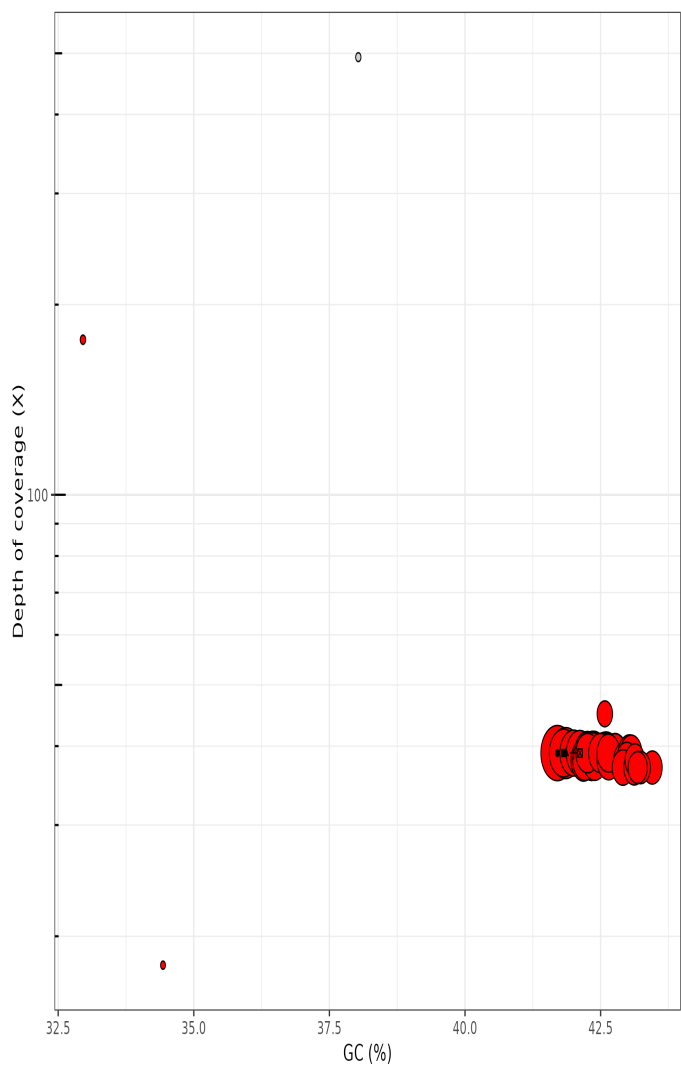


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



- Length (bp)
- 2e+07
 - 4e+07
 - 6e+07
 - 8e+07
- Longest sequences (bp)
- xgLurLuri1_1 - 87885100 (Eukaryota)
 - xgLurLuri1_2 - 72828418 (Eukaryota)
 - xgLurLuri1_3 - 62455121 (Eukaryota)
 - xgLurLuri1_4 - 56730159 (Eukaryota)
 - xgLurLuri1_5 - 54645859 (Eukaryota)
- superkingdom
- Eukaryota
 - N/A

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	159

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: Arnaud Couloux

Affiliation: Genoscope

Date and time: 2026-07-16 16:12:54 CEST