# ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

| | |
|---|---|
| TxID | 3199284 |
| ToLID | **xgLotLoto1** |
| Species | Lotoria lotoria |
| Class | Gastropoda |
| Order | Littorinimorpha |

| Genome Traits | Expected | Observed |
|---|---|---|
| Haploid size (bp) | 1,280,986,079 | 2,996,862,105 |
| Haploid Number | 17 (source: ancestor) | 35 |
| Ploidy | 2 (source: ancestor) | 2 |
| Sample Sex | Unknown | Unknown |

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q37

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
. BUSCO single copy value is less than 90% for collapsed
. BUSCO duplicated value is more than 5% for collapsed
. More than 1000 gaps/Gbp for collapsed

### Curator notes

. Interventions/Gb: 39
. Contamination notes: ""
. Other observations: "The assembly of Lotoria lotoria (xgLotLoto1.1) is based on 77X ONT data and 193X 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, 125 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 12.499 Mb (with the largest being 2.159 Mb). Additionally, 3595 regions totaling 169.543 Mb (with the largest being 0.389 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using ptGAUL. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 23 haplotypic regions were removed, totaling 9.472

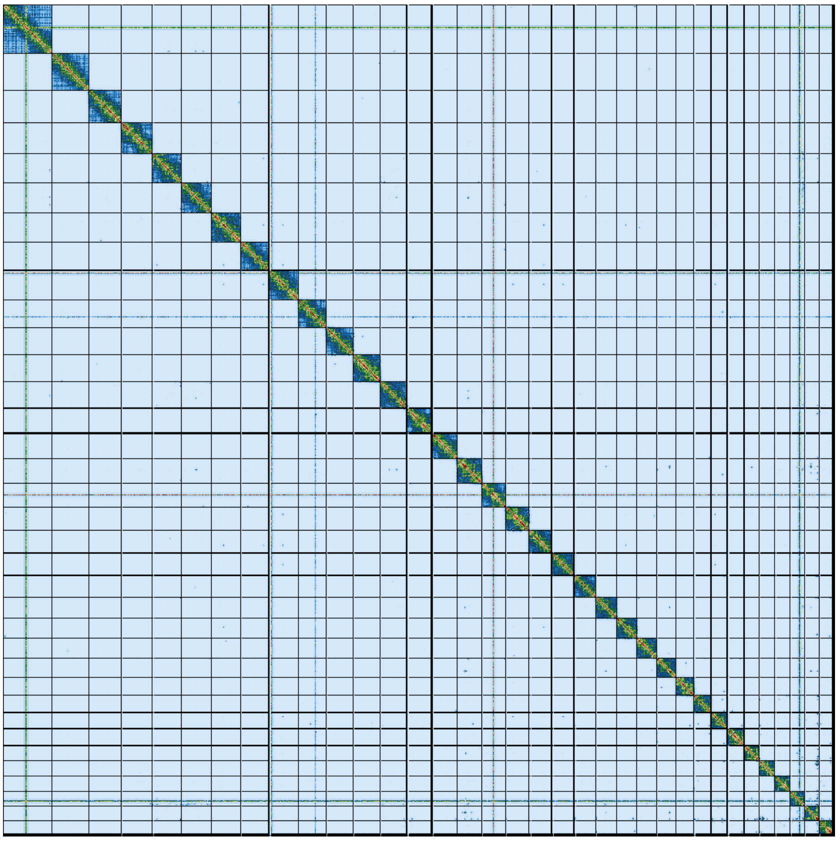
Mb (with the largest being 1.027 Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

# Quality metrics table

| Metrics | Pre-curation collapsed | Curated collapsed |
|---|---|---|
| Total bp | 3,006,440,675 | 2,996,862,105 |
| GC % | 40.99 | 41.02 |
| Gaps/Gbp | 1,234.35 | 1,249.97 |
| Total gap bp | 371,100 | 382,900 |
| Scaffolds | 565 | 443 |
| Scaffold N50 | 89,157,620 | 89,534,283 |
| Scaffold L50 | 14 | 14 |
| Scaffold L90 | 31 | 30 |
| Contigs | 4,276 | 4,189 |
| Contig N50 | 1,551,082 | 1,553,312 |
| Contig L50 | 599 | 596 |
| Contig L90 | 2,109 | 2,091 |
| QV | 37.1244 | 37.1174 |
| Kmer compl. | 84.637 | 84.5827 |
| BUSCO sing. | 76.4% | 76.4% |
| BUSCO dupl. | 21.4% | 21.3% |
| BUSCO frag. | 0.7% | 0.7% |
| BUSCO miss. | 1.5% | 1.6% |

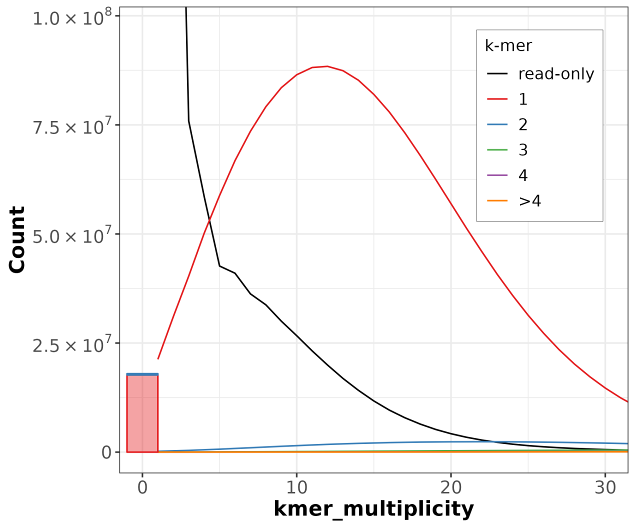
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

# HiC contact map of curated assembly

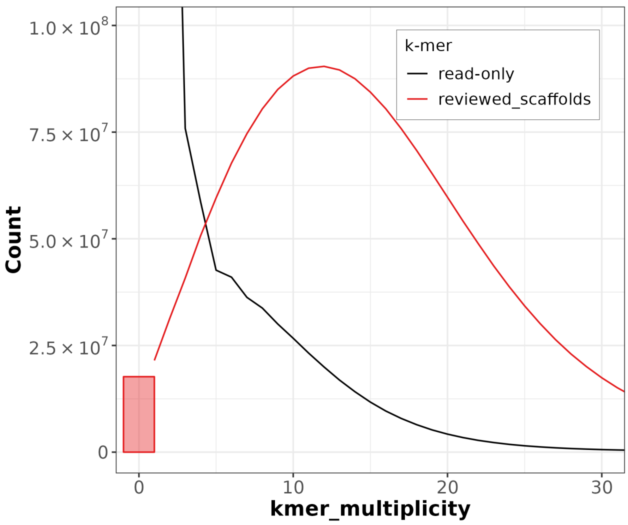


**collapsed** [LINK]

# K-mer spectra of curated assembly

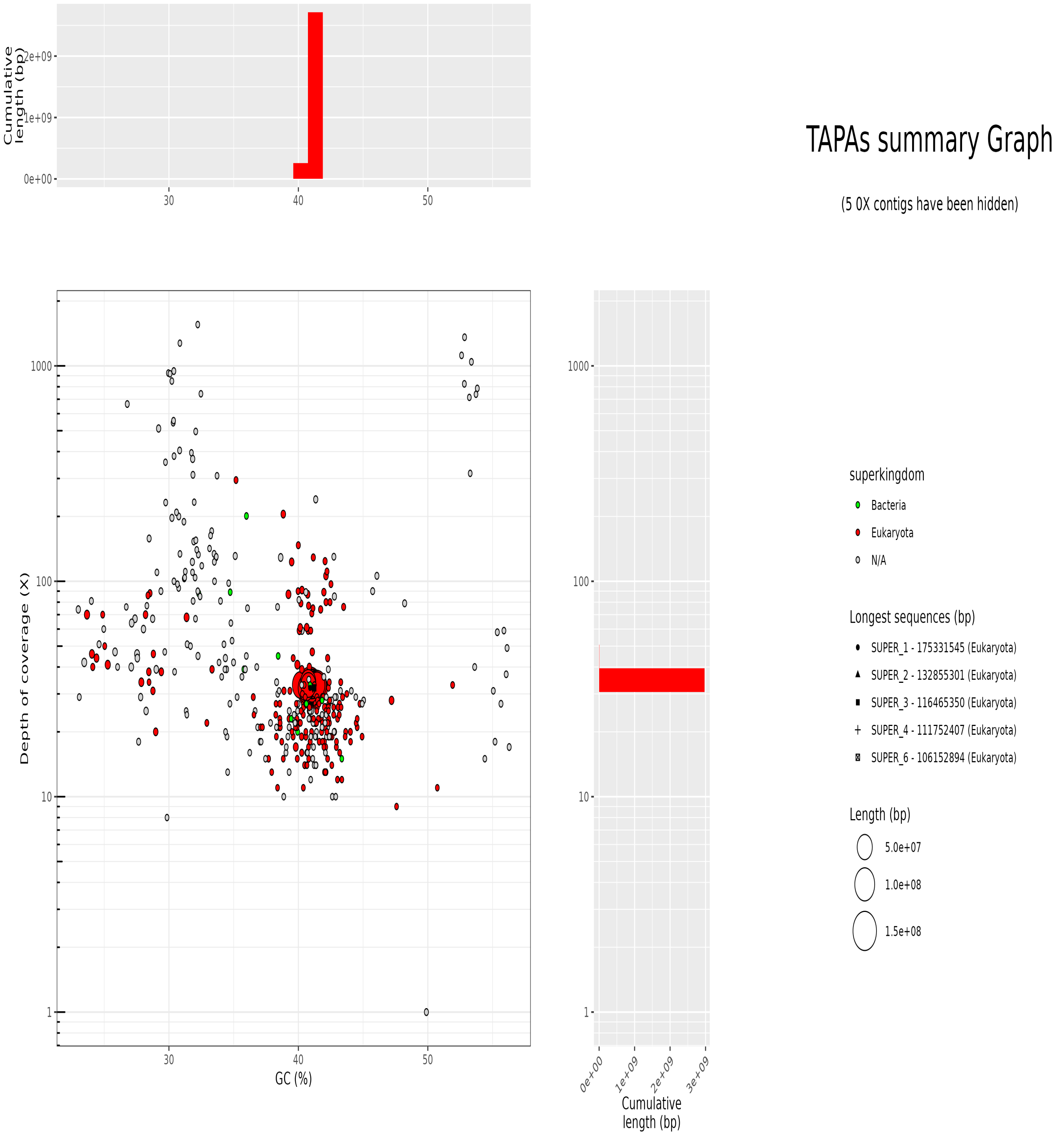


Distribution of k-mer counts per copy numbers found in asm



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

# Post-curation contamination screening



**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 | 77 | 193 |

# 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: 2025-11-29 01:30:58 CET