

# ERGA Assembly Report

v24.04.03\_beta

Tags: ERGA-BGE

TxID	3034287
ToLID	<b>xgDerLasi</b>
Species	Deroceras lasithionense
Class	Gastropoda
Order	Stylommatophora

Genome Traits	Expected	Observed
Haploid size (bp)	1,014,304,119	1,235,896,202
Haploid Number	16 (source: ancestor)	31
Ploidy	2 (source: ancestor)	2
Sample Sex	HERMAPHRODITE_MONOECIOUS	unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q55

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
- . Observed sex is different from Sample sex
- . Kmer completeness value is less than 90 for collapsed
- . Assembly length loss > 3% for collapsed

### Curator notes

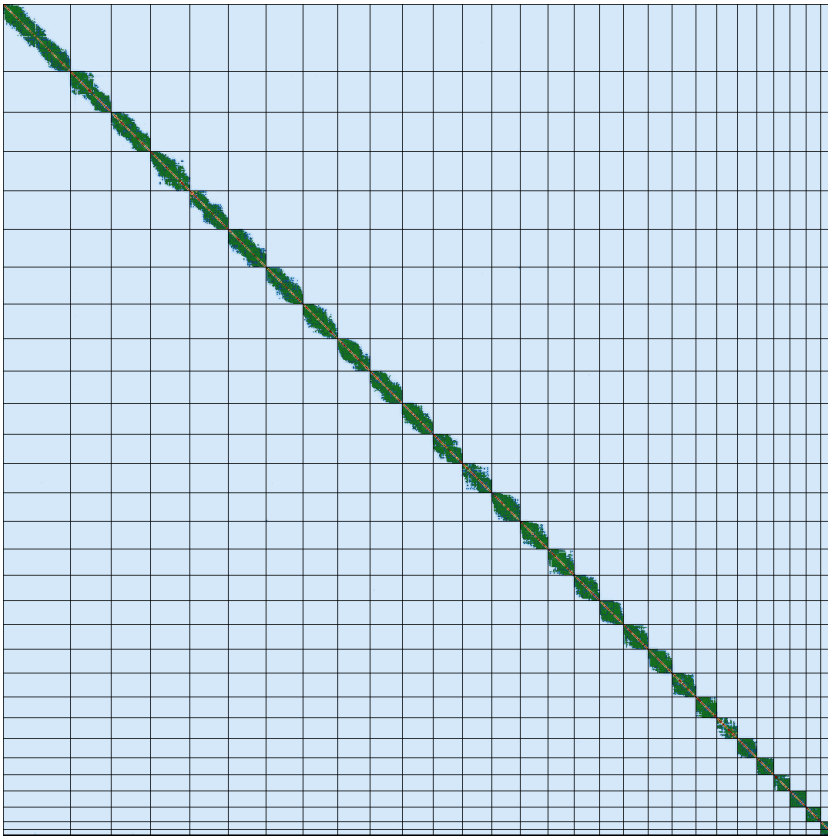
- . Interventions/Gb: 35
- . Contamination notes: "some bacterial contigs were removed"
- . Other observations: "several contigs were tagged as haplotigs"

## Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,330,276,266	1,235,896,202
GC %	41.44	41.45
Gaps/Gbp	798.33	811.56
Total gap bp	106,200	101,200
Scaffolds	90	58
Scaffold N50	45,757,421	45,757,421
Scaffold L50	12	11
Scaffold L90	26	24
Contigs	1,152	1,061
Contig N50	2,037,934	2,037,218
Contig L50	190	180
Contig L90	630	591
QV	55.0595	55.0208
Kmer compl.	77.1708	71.958
BUSCO sing.	98.0%	92.9%
BUSCO dupl.	1.6%	1.6%
BUSCO frag.	0.0%	0.4%
BUSCO miss.	0.4%	5.1%

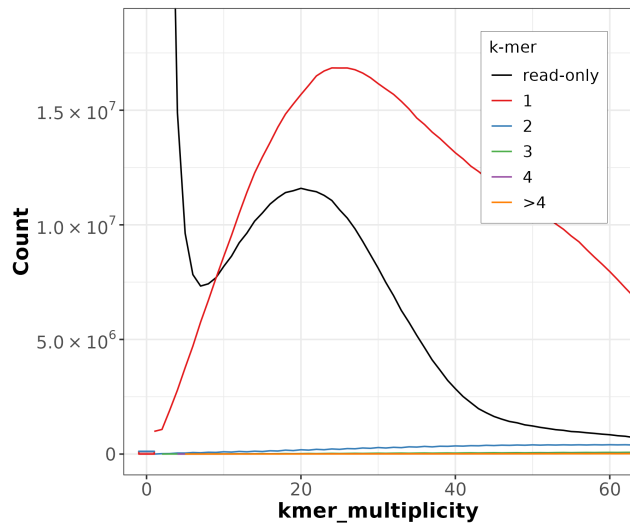
BUSCO 5.4.3 Lineage: eukaryota\_odb10 (genomes:70, BUSCOs:255)

# HiC contact map of curated assembly

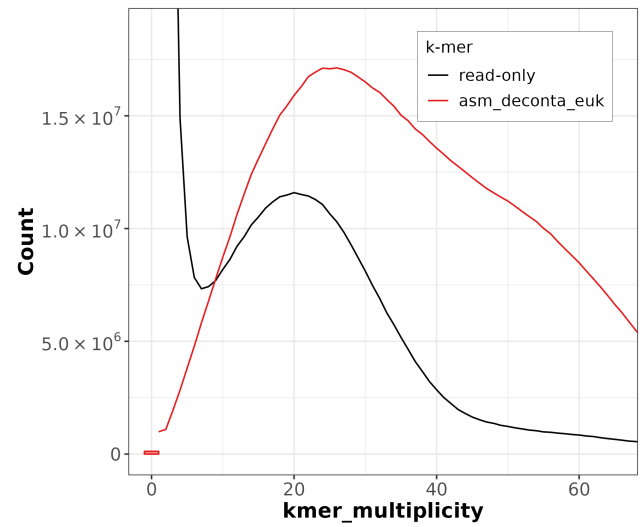


collapsed [\[LINK\]](#)

# K-mer spectra of curated assembly

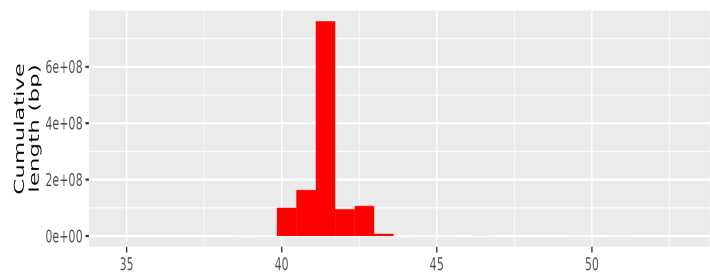


Distribution of k-mer counts per copy numbers found in asm

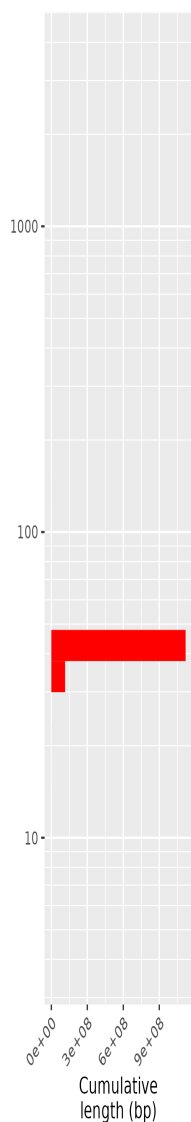
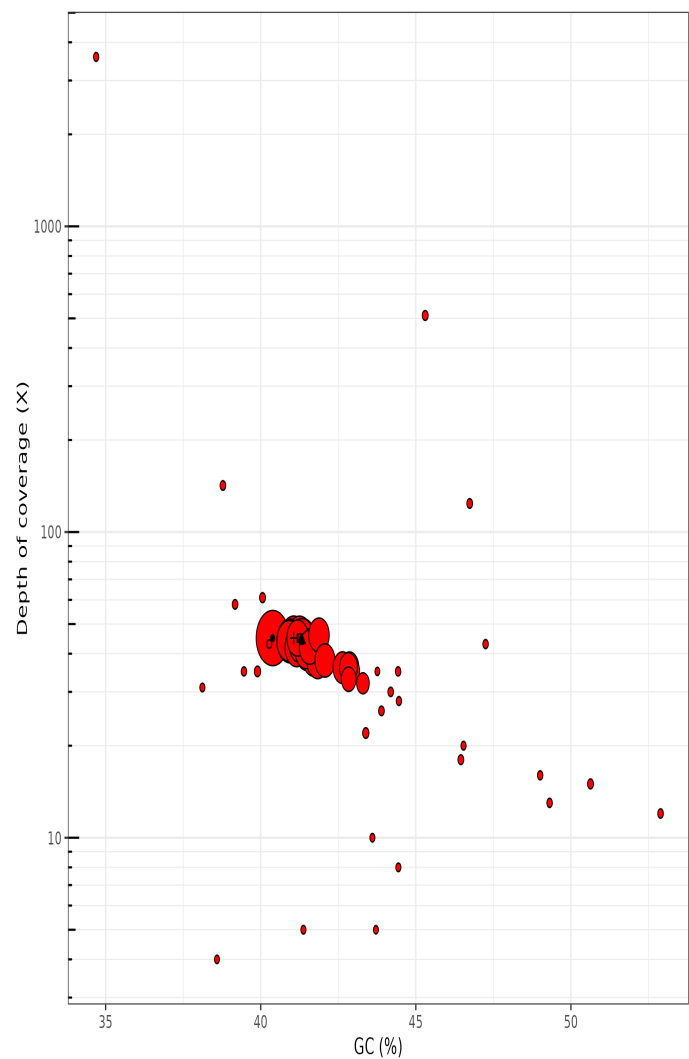


Distribution of k-mer counts coloured by their presence in reads/assemblies

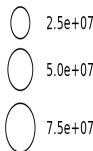
# Post-curation contamination screening



TAPAs summary Graph



Length (bp)



superkingdom

Eukaryota

Longest sequences (bp)

- xgDerLasi1 - 99714049 (Eukaryota)
- xgDerLasi2 - 61325393 (Eukaryota)
- xgDerLasi3 - 58383161 (Eukaryota)
- xgDerLasi4 - 57976078 (Eukaryota)
- xgDerLasi5 - 57252941 (Eukaryota)

**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	PACBIO Hifi	Arima (4-enz)
Coverage	49	45

## Assembly pipeline

- **Hifiasm**
  - |\_ *ver*: 0.19.5-r593
  - |\_ *key param*: NA
- **purge\_dups**
  - |\_ *ver*: 1.2.5
  - |\_ *key param*: NA

## Curation 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

Submitter: Caroline Belser

Affiliation: Genoscope

Date and time: 2024-07-25 14:22:32 CEST