

# ERGA Assembly Report

v24.09.10

Tags: ERGA-BGE

TxID	83528
ToLID	<b>mAcoMin</b>
Species	Acomys minous
Class	Mammalia
Order	Rodentia

Genome Traits	Expected	Observed
Haploid size (bp)	2,168,472,368	2,350,516,084
Haploid Number	18 (source: ancestor)	20
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q57

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

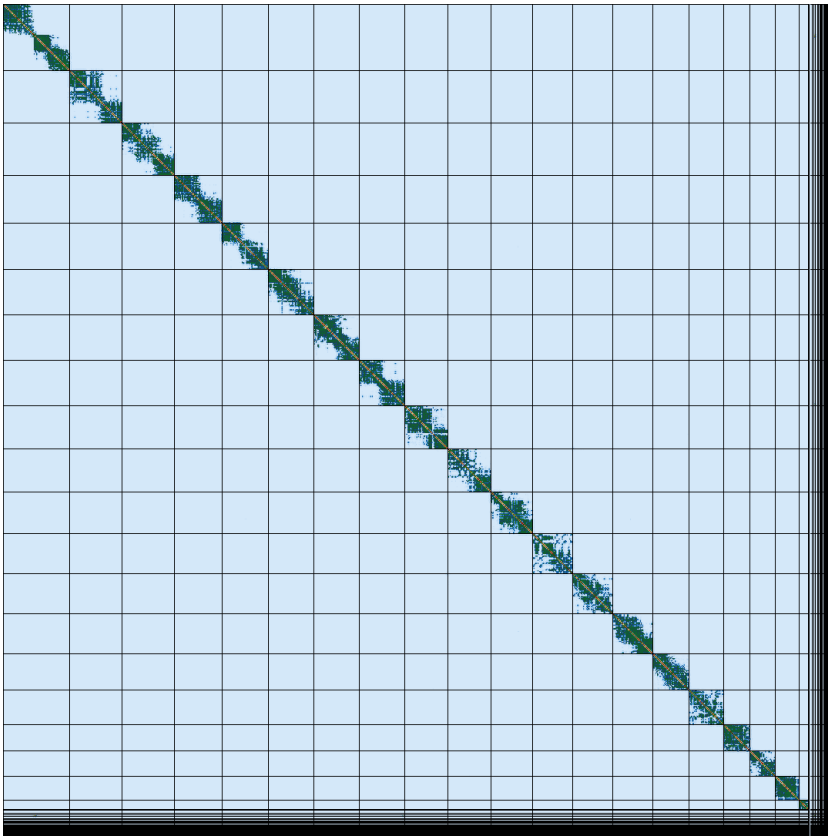
- . Interventions/Gb: 19
- . Contamination notes: ""
- . Other observations: "Some areas are ambiguous. They might look like haplotypic duplications, but there\'s not necessarily a drop in coverage to validate it. "

# Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	2,355,108,853	2,350,516,084
GC %	42.87	42.87
Gaps/Gbp	79.83	78.71
Total gap bp	20,400	20,100
Scaffolds	112	113
Scaffold N50	122,851,966	122,851,966
Scaffold L50	9	9
Scaffold L90	18	18
Contigs	300	298
Contig N50	29,281,624	29,281,624
Contig L50	24	23
Contig L90	106	103
QV	57.3217	57.3117
Kmer compl.	89.6434	89.5961
BUSCO sing.	94.9%	94.9%
BUSCO dupl.	1.3%	1.3%
BUSCO frag.	0.5%	0.5%
BUSCO miss.	3.3%	3.3%

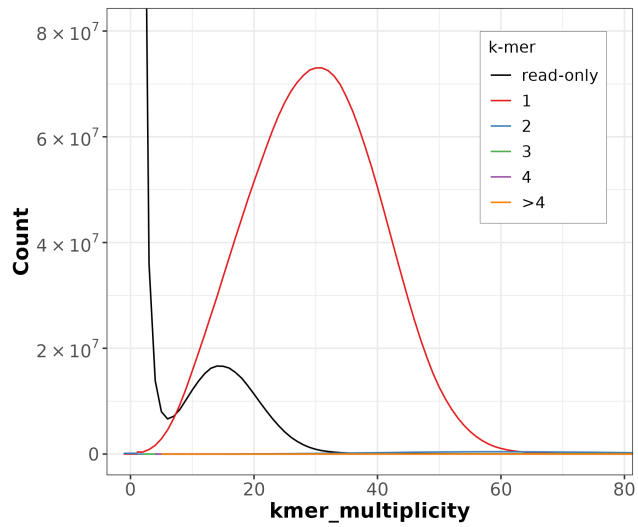
BUSCO 5.4.3 Lineage: glires\_odb10 (genomes:24, BUSCOs:13798)

# HiC contact map of curated assembly

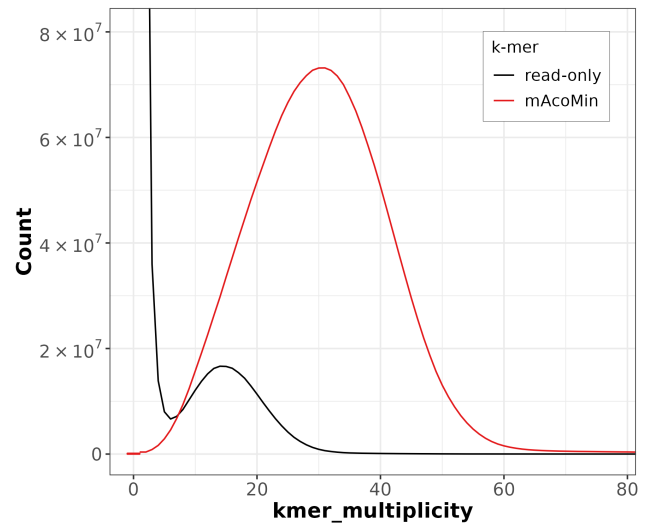


collapsed [\[LINK\]](#)

# K-mer spectra of curated assembly

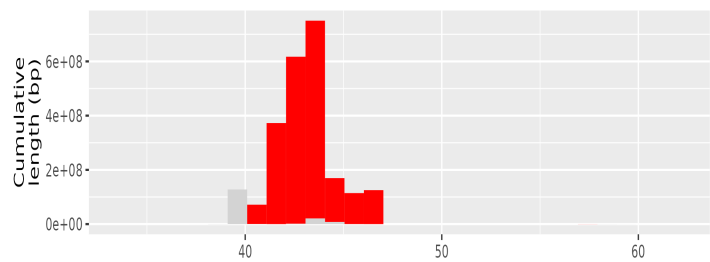


Distribution of k-mer counts per copy numbers found in asm

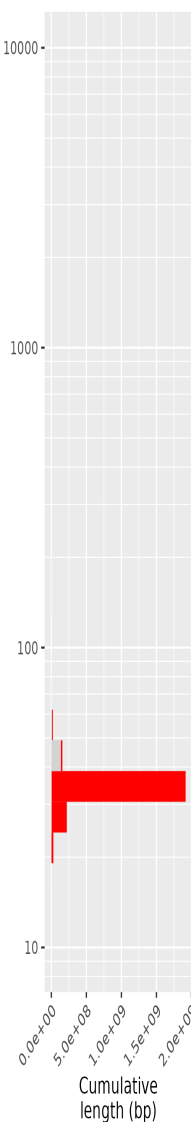
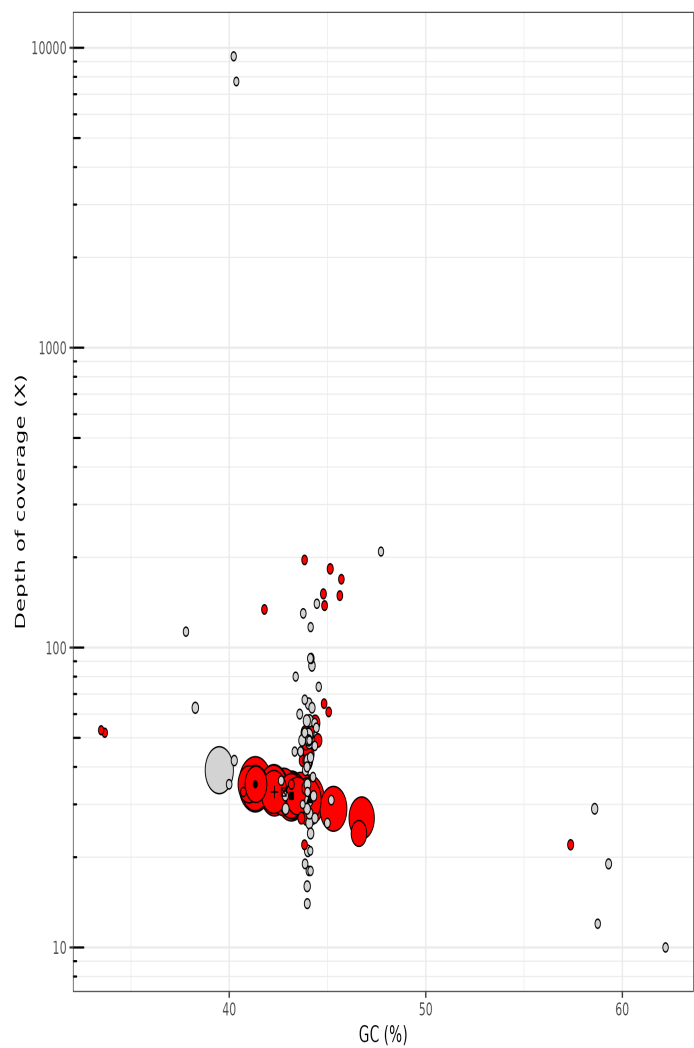


Distribution of k-mer counts coloured by their presence in reads/assemblies

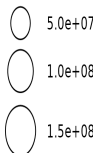
# Post-curation contamination screening



TAPAs summary Graph



Length (bp)



superkingdom



Longest sequences (bp)

- SUPER\_1 - 187704340 (Eukaryota)
- ▲ SUPER\_2 - 149085160 (Eukaryota)
- SUPER\_3 - 147818001 (Eukaryota)
- + SUPER\_4 - 133882899 (Eukaryota)
- ▣ SUPER\_5 - 130827132 (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	33	19

## 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: Sophie Mangenot

Affiliation: Genoscope

Date and time: 2024-09-19 14:04:59 CEST