ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	83528
ToLID	**mAcoMin1**
Species	Acomys minous
Class	Mammalia
Order	Rodentia

Genome Traits	Expected	Observed
Haploid size (bp)	2,168,472,367	2,350,193,447
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: 25
. Contamination notes: ""
. Other observations: "The assembly of Acomys minous (mAcoMin1) is based on 33X PacBio data and 19X Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA, https://www.erga-biodiversity.eu/) via the Biodiversity Genomics Europe project (BGE, https://biodiversitygenomics.eu/). The assembly process included the following steps: initial PacBio 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, 5 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 273 kb (with the largest being 133 kb). Additionally, 124 regions totaling 13 Mb (with the largest being 762 kb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 8 haplotypic regions totaling 9 Mb (with the largest being 2 Mb).Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size.Some areas are ambiguous. They might look like haplotypic duplications, but there\'s not necessarily a drop in coverage to validate it. A misassembled 44 kb scaffold corresponding to mitochondria was deleted as the

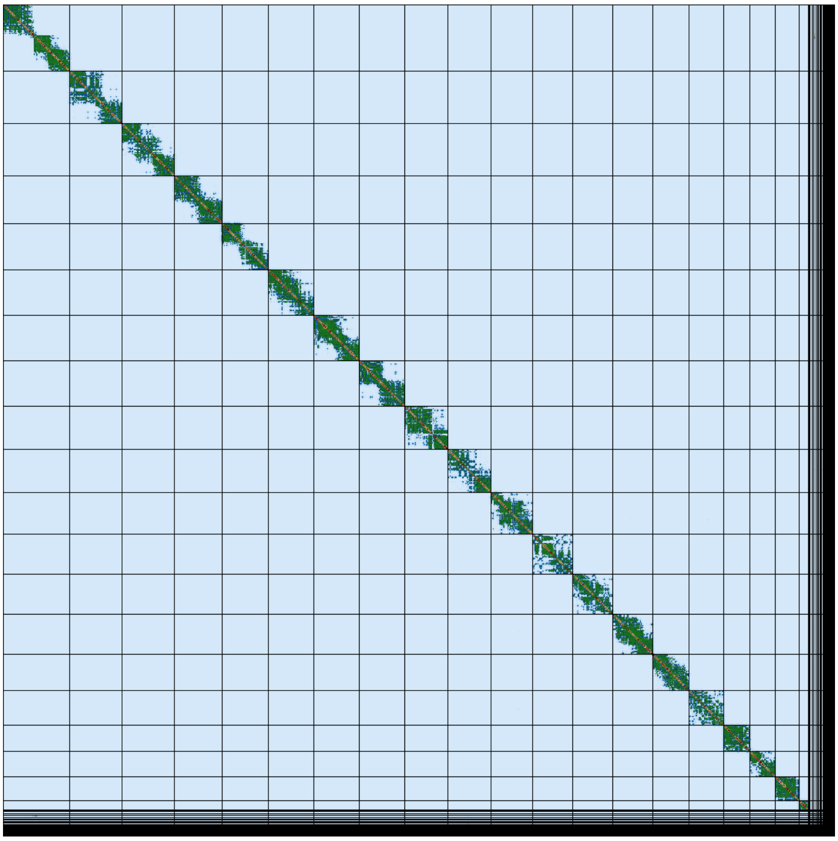
mitochondrial sequence is already integrated into the assembly. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	2,383,455,893	2,350,193,447
GC %	42.88	42.87
Gaps/Gbp	77.2	78.29
Total gap bp	18,400	20,100
Scaffolds	120	112
Scaffold N50	122,327,231	122,851,966
Scaffold L50	9	9
Scaffold L90	18	18
Contigs	304	296
Contig N50	23,975,000	29,281,624
Contig L50	28	23
Contig L90	117	102
QV	42.0629	57.312
Kmer compl.	89.5095	89.5959
BUSCO sing.	96.3%	96.2%
BUSCO dupl.	0.5%	0.6%
BUSCO frag.	1.0%	1.0%
BUSCO miss.	2.2%	2.3%

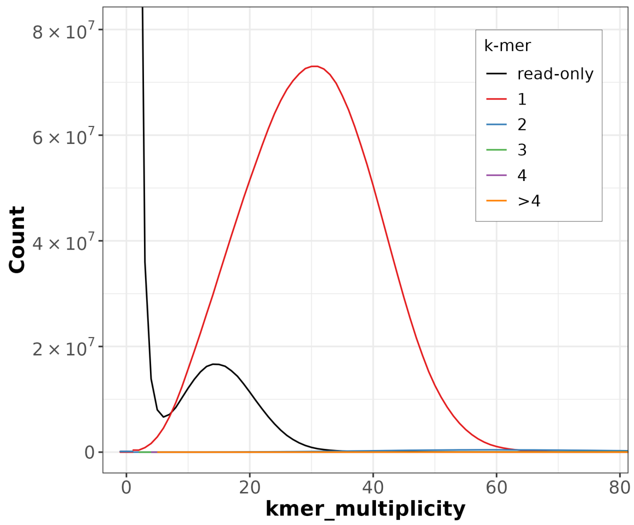
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: rodentia_odb12 (genomes:32, BUSCOs:12639)

HiC contact map of curated assembly

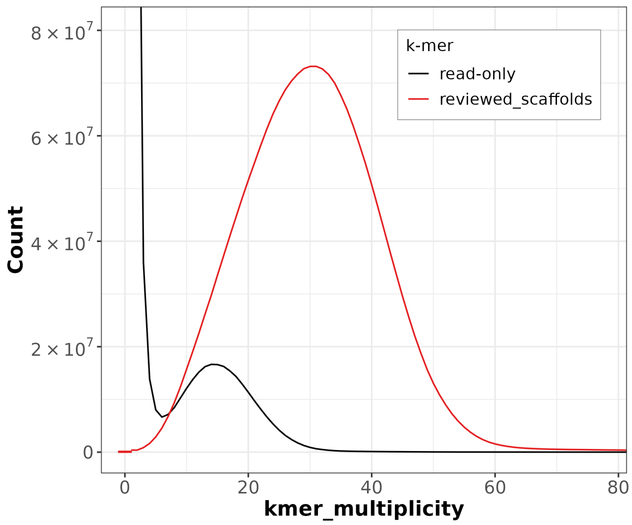


collapsed [LINK]

K-mer spectra of curated assembly

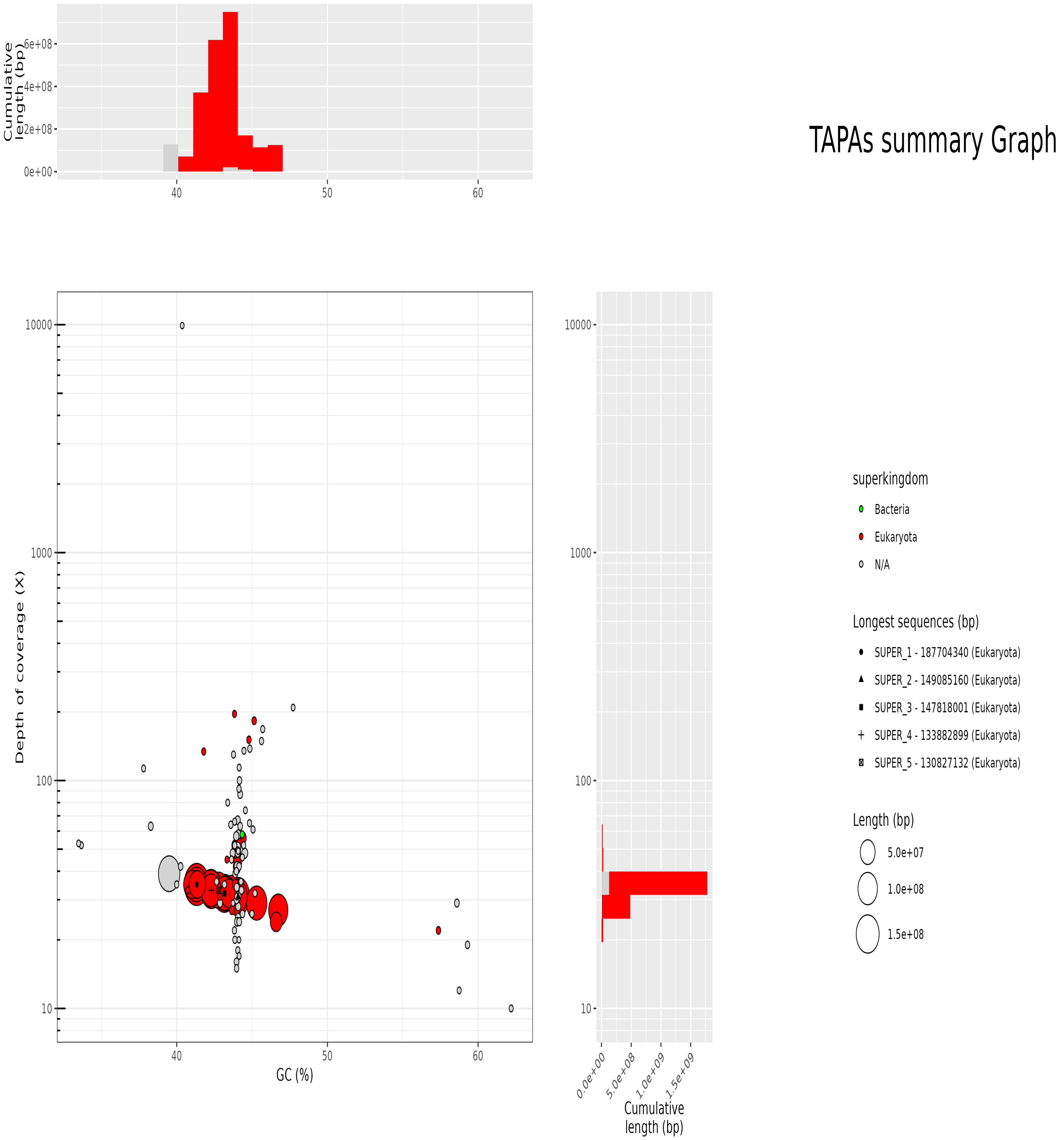


Distribution of k-mer counts per copy numbers found in asm



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	PACBIO Hifi	Arima
Coverage	33	18

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 Layac
Affiliation: Genoscope

Date and time: 2025-04-08 22:58:54 CEST