

ERGA Assembly Report

v24.02.02_beta

Tags: ERGA-BGE

ToLID	mApoAgr2
Species	Apodemus agrarius
Class	Mammalia
Order	Rodentia

Genome Traits	Expected	Observed
Haploid size (bp)	2,822,679,104	2,614,354,400
Haploid Number	24 (source: direct)	24
Ploidy	2 (source: direct)	2
Sample Sex	XY	XY

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 8.8.Q47

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

Curator notes

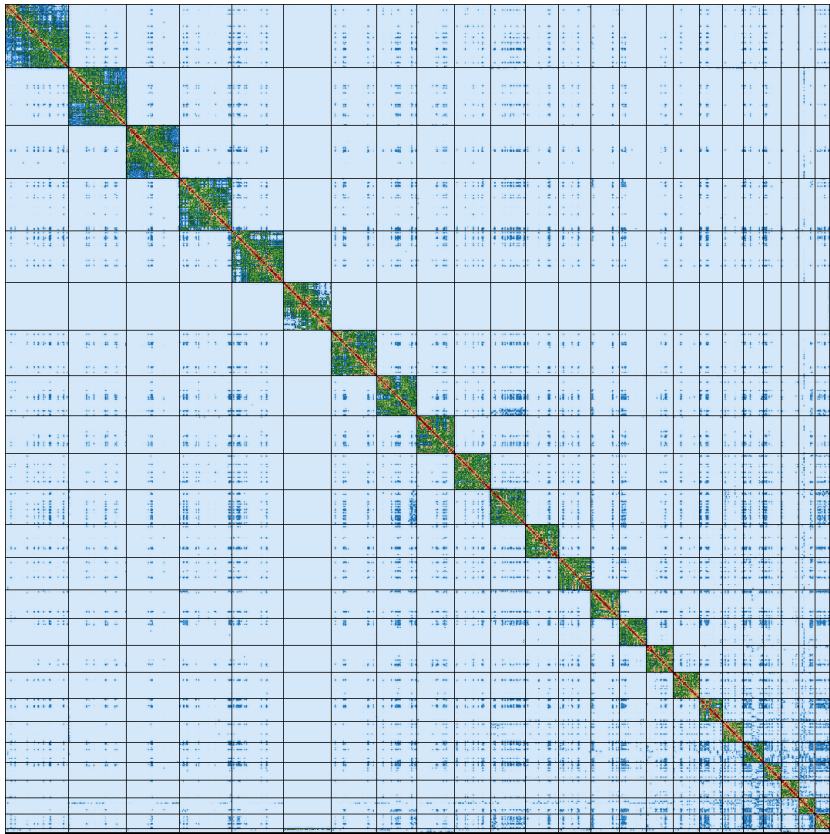
. Interventions/Gb: 5
. Contamination notes: "found fragments of a trypanosoma sp. corresponding to a small unloc of SUPER_1 (SUPER_1_unloc_3) of length 184,352 bases, which was removed from the final assembly "
. Other observations: "Yahs did a good initial job. Manual curation was minimal. Stats, mercury file and busco were obtained from the assembly without the contaminated short unloc scaff & with the added mitochondria sequence for this organism. The post-curation contamination screening plot correspond the the curated assembly before removal of the trypanosoma sp. sequence"

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	2,614,536,552	2,614,354,400
GC %	42.15	42.15
Gaps/Gbp	64.64	68.85
Total gap bp	33,800	36,000
Scaffolds	75	63
Scaffold N50	113,335,371	117,961,188
Scaffold L50	9	9
Scaffold L90	21	20
Contigs	244	243
Contig N50	35,188,814	35,188,814
Contig L50	22	22
Contig L90	85	85
QV	47.2298	47.2305
Kmer compl.	90.9325	90.9314
BUSCO sing.	96.4%	96.4%
BUSCO dupl.	1.9%	1.9%
BUSCO frag.	0.5%	0.5%
BUSCO miss.	1.2%	1.2%

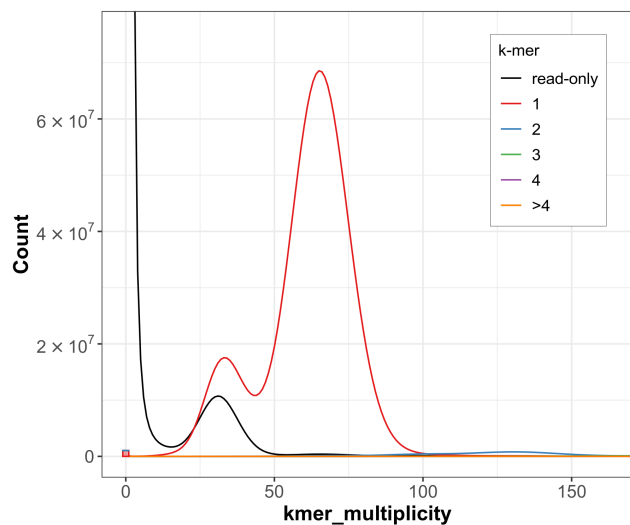
BUSCO 5.4.0 Lineage: vertebrata_odb10 (genomes:67, BUSCOs:3354)

HiC contact map of curated assembly

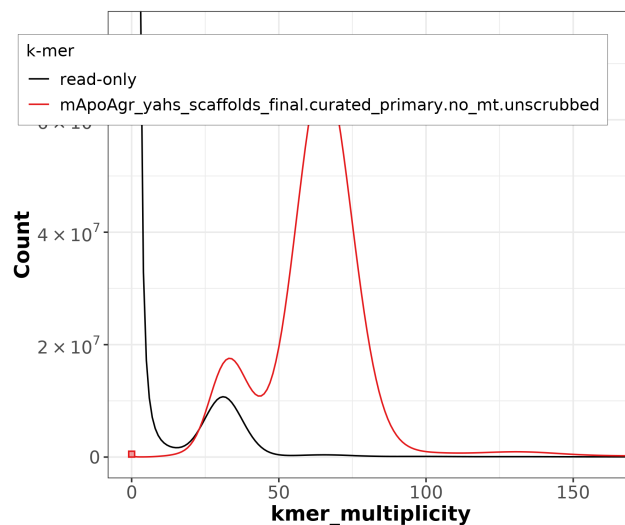


collapsed [\[LINK\]](#)

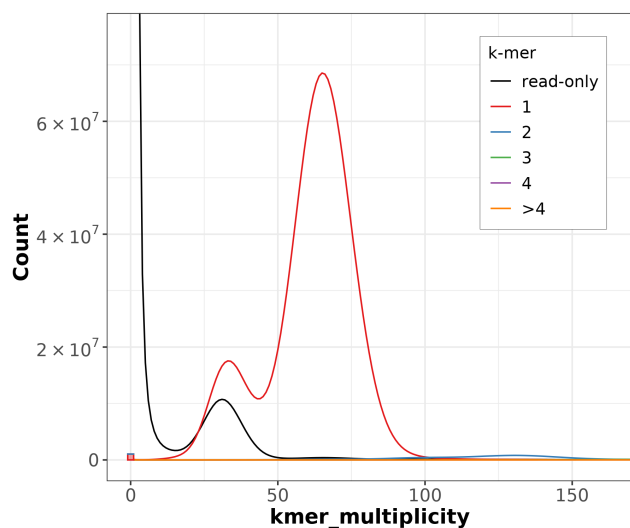
K-mer spectra of curated assembly



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

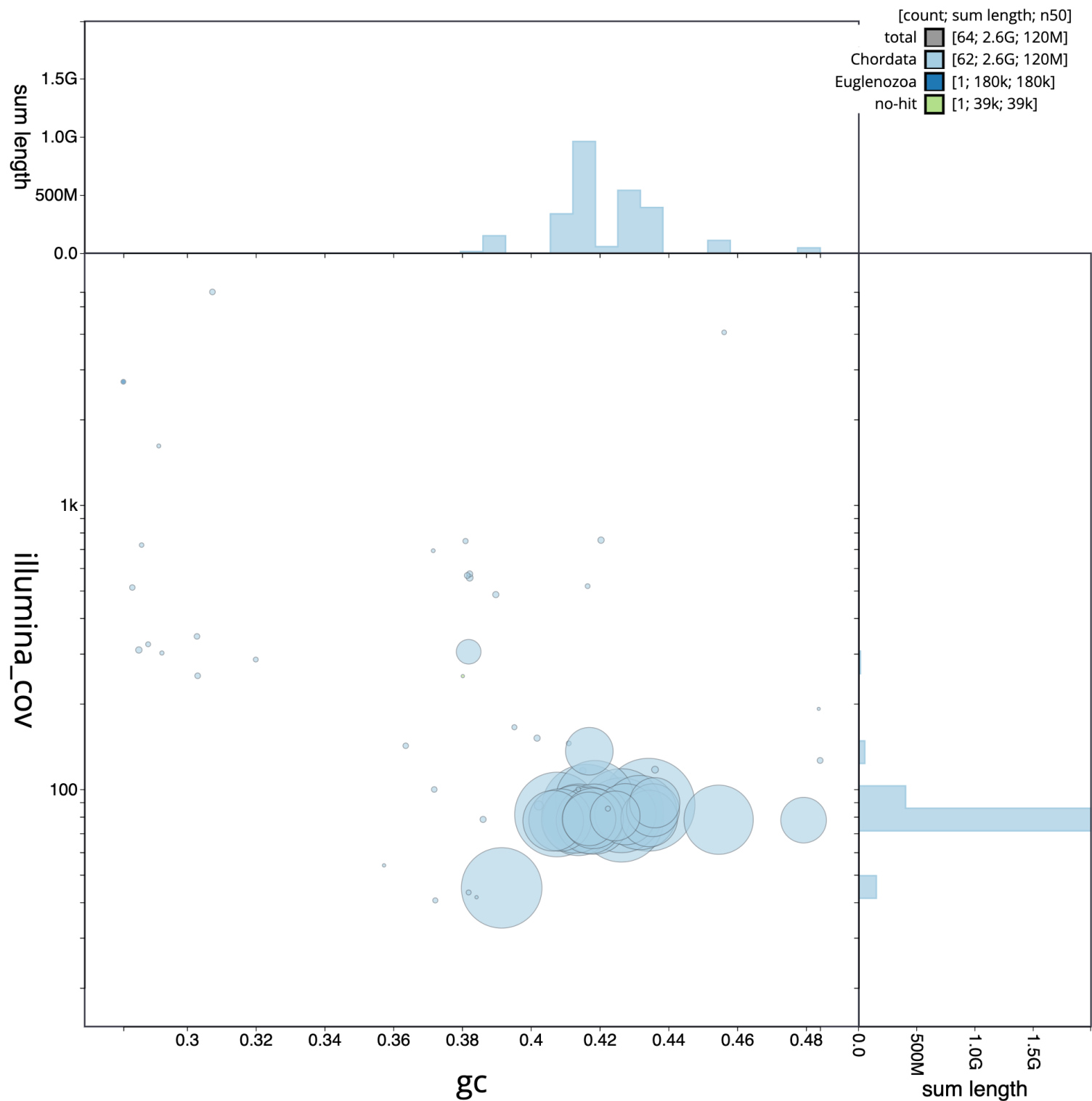


Distribution of k-mer counts coloured by their presence in reads/assemblies



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

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	ONT	Illumina	OmniC
Coverage	91x	87x	64x

Assembly pipeline

- **Trim_Galore**
 - |_ *ver*: 0.6.7
 - |_ *key param*: "--gzip -q 20"
 - |_ *key param*: "--paired"
 - |_ *key param*: "--retain_unpaired"
- **Filtlong**
 - |_ *ver*: 0.2.1
 - |_ *key param*: NA
- **nextdenovo**
 - |_ *ver*: 2.5.0
 - |_ *key param*: NA
- **hypo**
 - |_ *ver*: 1.0.3
 - |_ *key param*: NA
- **purge_dups**
 - |_ *ver*: 1.2.6
 - |_ *key param*: NA
- **YaHS**
 - |_ *ver*: 1.2a
 - |_ *key param*: NA

Curation pipeline

- **PretextView**
 - |_ *ver*: 0.2.5
 - |_ *key param*: NA

Submitter: Francisco Camara

Affiliation: CNAG Barcelona

Date and time: 2024-02-05 14:37:30 CET