ERGA Assembly Report

v24.02.09_beta

Tags: ERGA-BGE

ToLID	**mApoAgr2**
Species	**Apodemus agrarius**
Class	Mammalia
Order	Rodentia

Genome Traits	Expected	Observed
Haploid size (bp)	2,822,679,104	2,613,523,018
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: 7.8.Q47

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

Curator notes

. Interventions/Gb: 9
. 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, now reflected in the pretext map which corresponds to the uncontaminated sequence."
. Other observations: "This is the version of the assembly after going through the ERGA reviewing process. Stats, merqury file and busco were obtained from the assembly without the contaminated short unloc scaffold. The post-curation contamination screening plot corresponds to the curated assembly before removal of the trypanosoma sp. sequence. WARNING: The shared contact map (.pretext) of the curated assembly shows all scaffolds sorted by length. Thus most of the unloc are towards the end and separated from their corresponding SUPER. This will be fixed in future versions of CLAWS"

Quality metrics table

| Metrics	Pre-curation collapsed	Curated collapsed
Total bp	2,614,536,552	2,613,523,018
GC %	42.15	42.15
Gaps/Gbp	64.64	69.64
Total gap bp	33,800	36,400
Scaffolds	75	60
Scaffold N50	113,335,371	119,125,257
Scaffold L50	9	9
Scaffold L90	21	20
Contigs	244	242
Contig N50	35,188,814	35,188,814
Contig L50	22	22
Contig L90	85	85
QV	47.2298	47.2292
Kmer compl.	90.9325	90.9176
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)
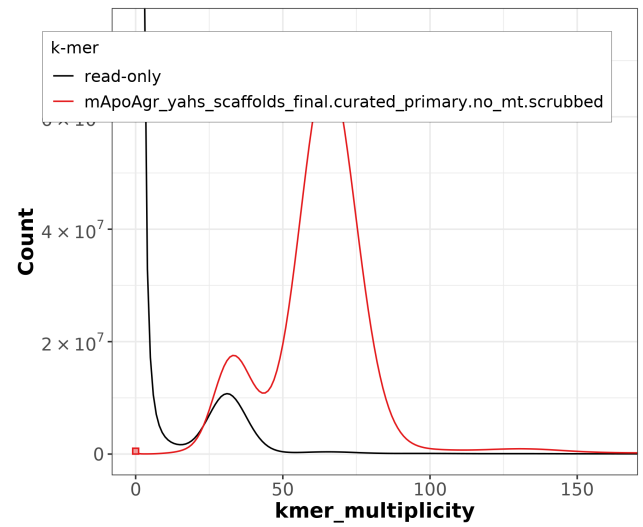
HiC contact map of curated assembly
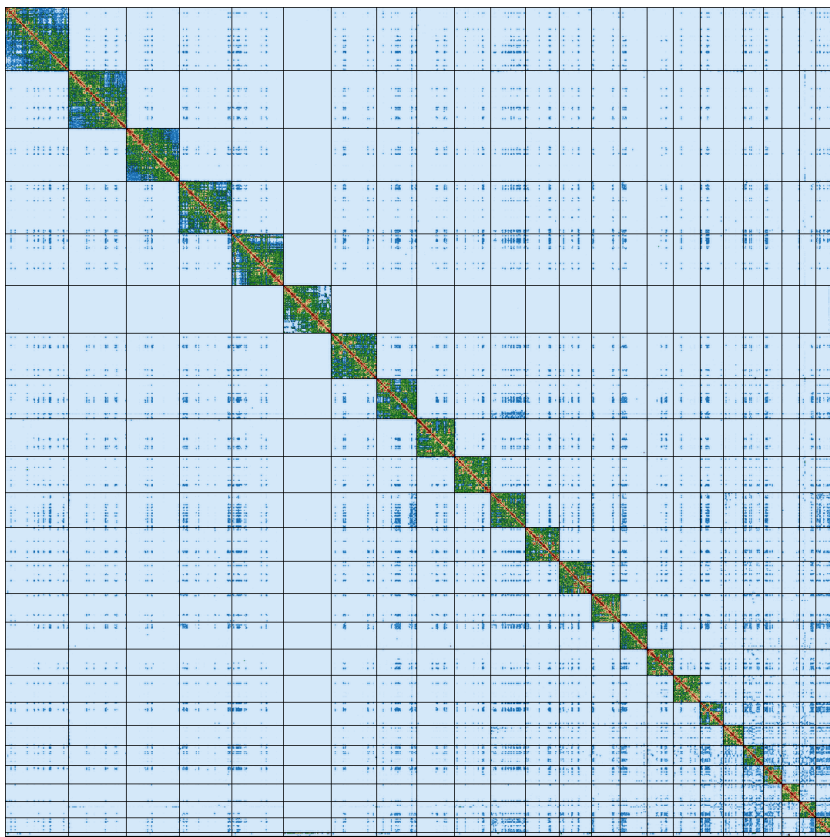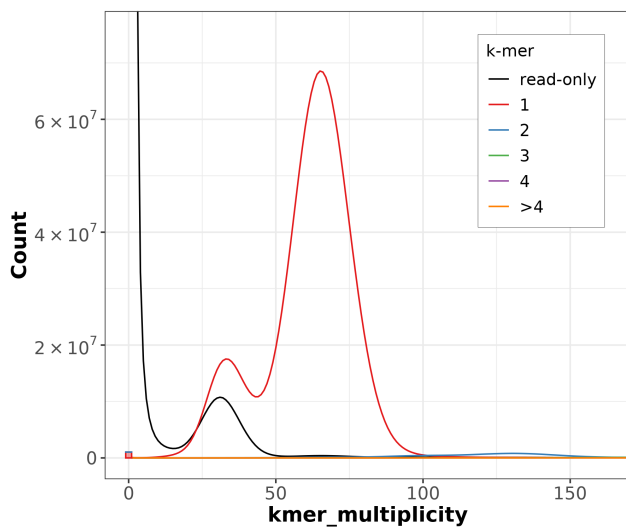

collapsed [LINK]

K-mer spectra of curated assembly
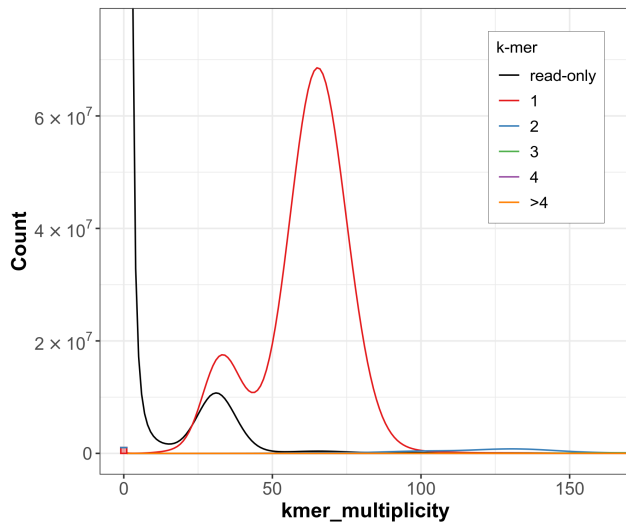
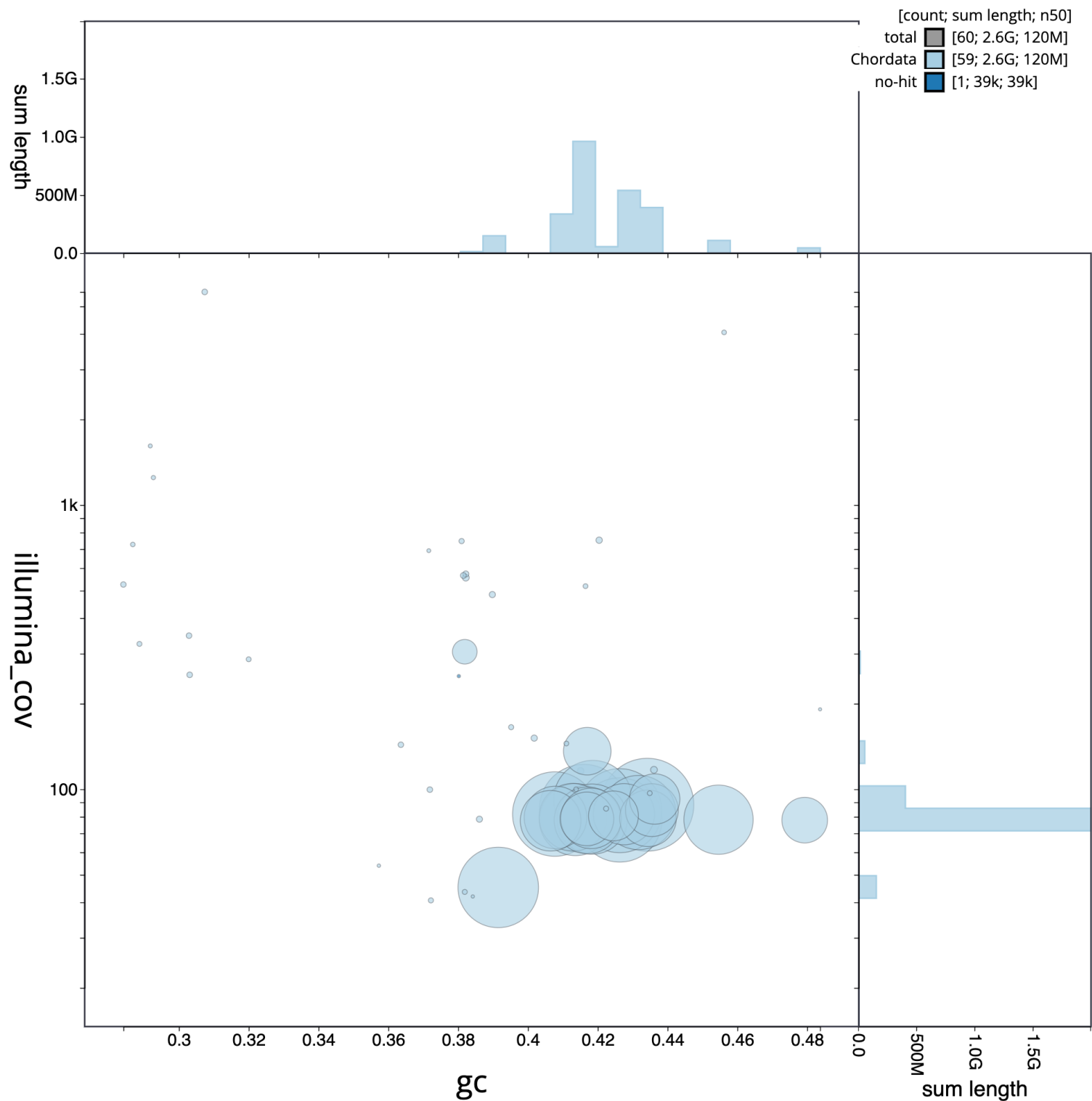Distribution of k-mer counts per copy
numbers found in asm

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: 2024-03-26 11:52:20 CET