

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	8585
ToLID	rNatMau1
Species	Natrix maura
Class	Lepidosauria
Order	Squamata

Genome Traits	Expected	Observed
Haploid size (bp)	1,604,071,484	1,758,771,001
Haploid Number	18 (source: ancestor)	17
Ploidy	3 (source: ancestor)	2
Sample Sex	ZW	ZW

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q46

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

- . Observed Haploid Number is different from Expected
- . Observed Ploidy is different from Expected

Curator notes

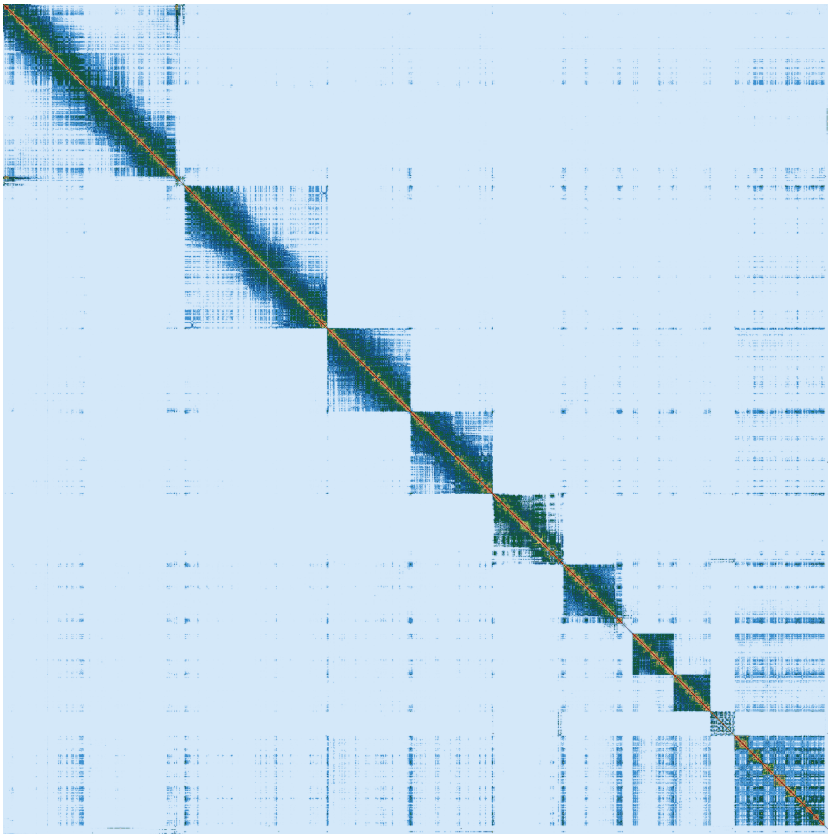
. Interventions/Gb: 5
. Contamination notes: "No contaminants found in the precurated assembly (blob plot shown under the curated assembly as there is need to redo this after curation)"
. Other observations: "First, nextdenovo assembly with no-haplotig sequences matching W were rescued from hap.fa (purge_dups). Second, it was scaffolded with yahs using 439.53 million hic pairs (note: W assembled in 1 scaffold of 5.4Mb and 46Mb of unlocalized W sequences). Third, curation made 0 cuts in contigs, 3 breaks at gaps and 8 join. Finally, the mitogenome was scaffolded into a single circular contig of 21.23 Kb. hic_FullMap_link points to a folder with the nosort pretext map and two savestates: the original (savestate_1) and another placing the SUPER_W scaffolds right after SUPER_Z (savestate_W). Regarding the EAR warnigs, they are unjustified: Chromosome number based on direct estimate in goat is 2n=34 thus the assembly is consistent with n=17. Smudgeplot shows this genome is clearly diploid."

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,758,748,764	1,758,771,001
GC %	41.49	41.49
Gaps/Gbp	35.25	38.09
Total gap bp	12,400	13,400
Scaffolds	243	239
Scaffold N50	173,680,720	173,680,720
Scaffold L50	4	4
Scaffold L90	15	13
Contigs	305	306
Contig N50	35,645,491	35,645,491
Contig L50	14	14
Contig L90	57	57
QV	46.2921	46.2922
Kmer compl.	91.5132	91.5142
BUSCO sing.	95.9%	95.9%
BUSCO dupl.	1.1%	1.1%
BUSCO frag.	0.8%	0.8%
BUSCO miss.	2.2%	2.2%

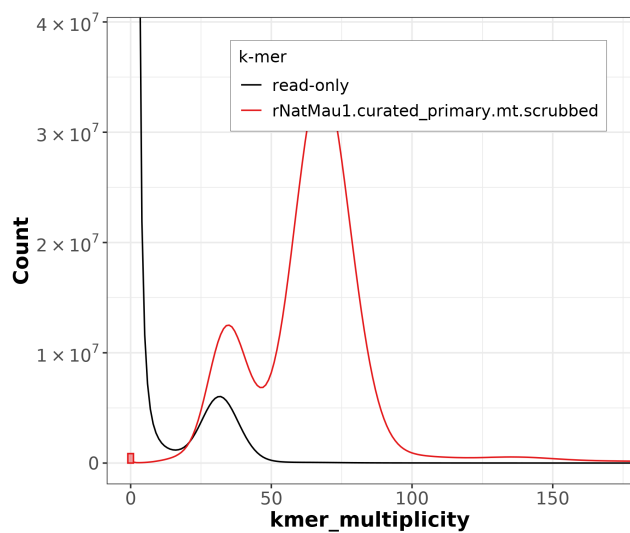
BUSCO: 5.4.0 (euk_genome_met, metaeuk) / Lineage: vertebrata_odb10 (genomes:67, BUSCOs:3354)

HiC contact map of curated assembly

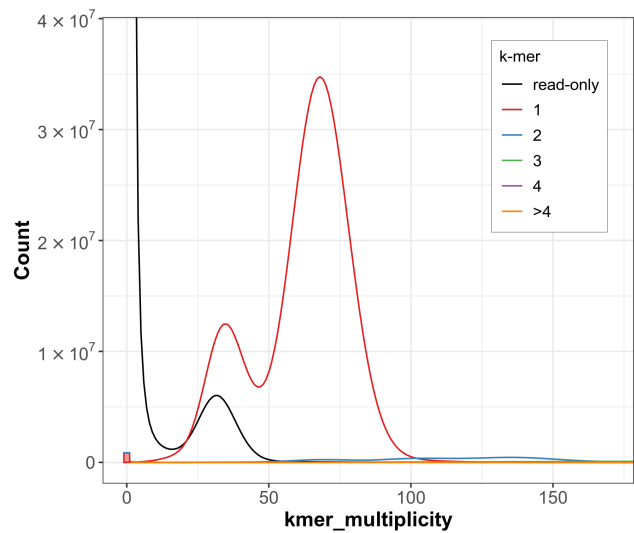


collapsed [\[LINK\]](#)

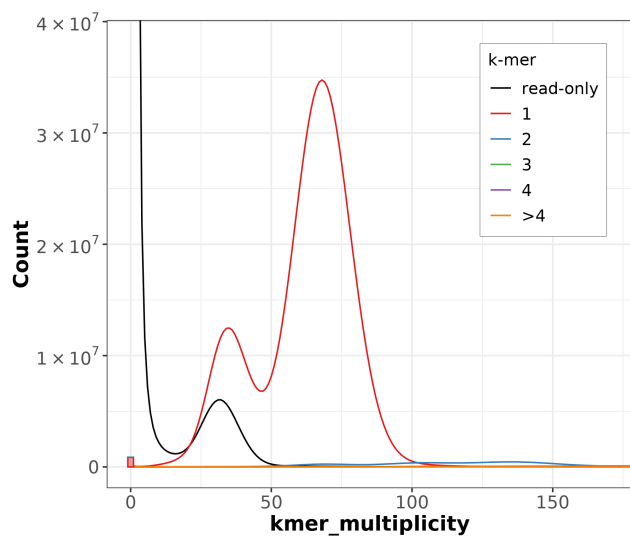
K-mer spectra of curated assembly



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

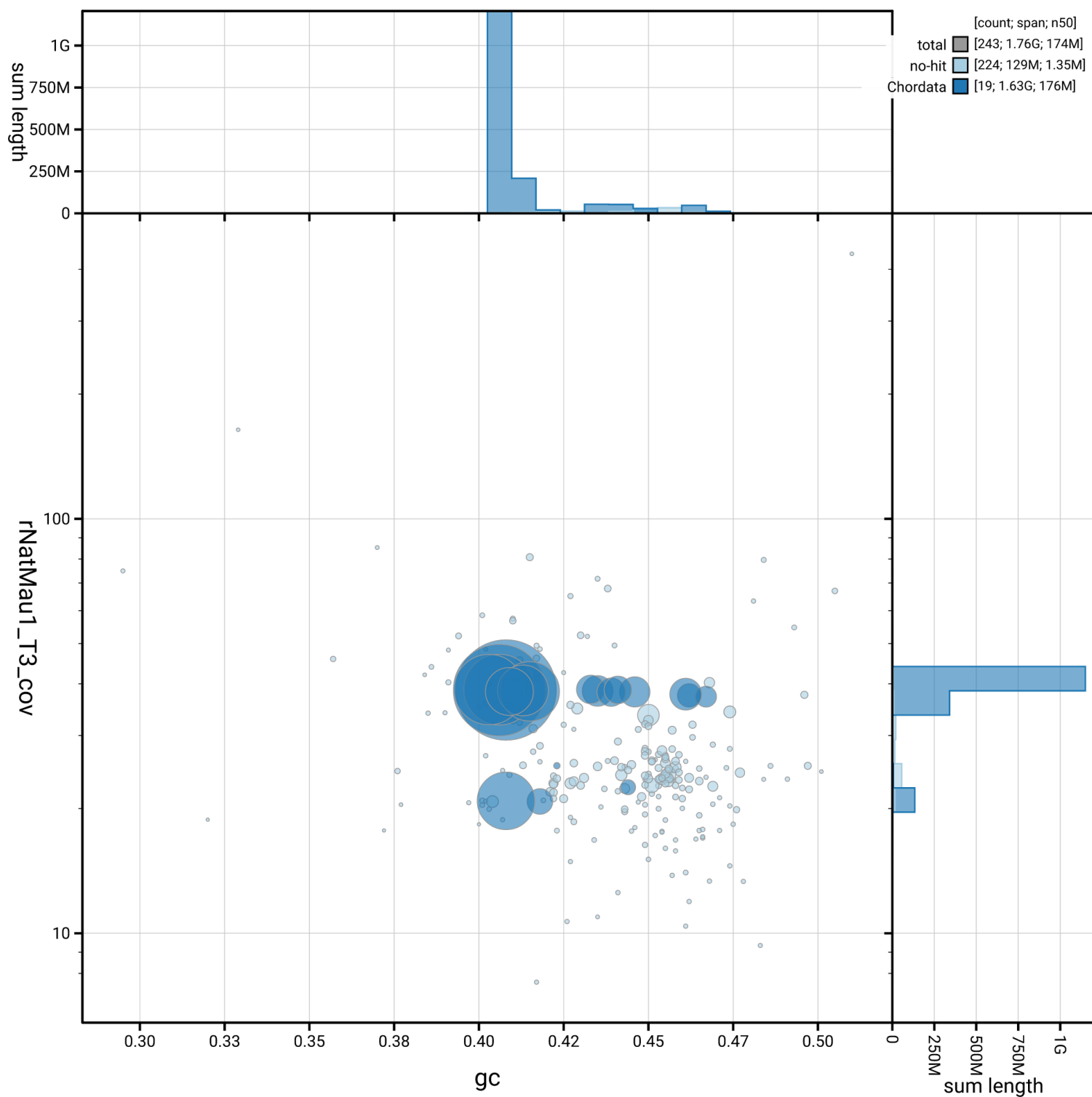


Distribution of k-mer counts per copy numbers found in asm



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	52x	88x	163x

Assembly pipeline

- **CLAWS pipeline**
 - |_ *ver*: 2.2.0
 - |_ *key param*: NA
- **Trim_galore**
 - |_ *ver*: 0.6.7
 - |_ *key param*: NA
- **Filtlong**
 - |_ *ver*: 0.2.1
 - |_ *key param*: NA
- **flye**
 - |_ *ver*: 2.9.1
 - |_ *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
- **Blobtoolkit nexflow pipeline**
 - |_ *ver*: 0.6.0
 - |_ *key param*: NA
- **FOAM pipeline**
 - |_ *ver*: 0.5
 - |_ *key param*: NA
- **mitos**
 - |_ *ver*: 2.1.3
 - |_ *key param*: NA

Curation pipeline

- **PretextView**
 - |_ *ver*: 0.2.5
 - |_ *key param*: NA
- **GRIT_Rapid**
 - |_ *ver*: 2.0
 - |_ *key param*: NA

Affiliation: CNAG, Barcelona

Date and time: 2024-11-26 11:28:05 CET