ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

TxID	184164
ToLID	**rMalMon1**
Species	Malpolon monspessulanus
Class	Lepidosauria
Order	Squamata

Genome Traits	Expected	Observed
Haploid size (bp)	1,613,706,619	1,752,149,986
Haploid Number	21 (source: direct)	22
Ploidy	3 (source: ancestor)	2
Sample Sex	ZW	ZW

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q48

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: 6
. Contamination notes: "There seem to be no significant contamination of this assembly"
. Other observations: "Report after trying to address the concerns of expert review of an initial curation effort. For this particular species we decided to scaffold and curate after the hypo polishing step rather than following purge_dups as we do for most species. We adopted this strategy because we noticed that we were missing all of the W chromosome from our final scaffolded assembly after running purge_dups. However upon alignment of the sequences eliminated by purge_dups against NR, we found that they all aligned to the W chromosome of another Sauropsida (Vipera ursinii) from the southern Mediterranean that diverged from the Montpellier snake roughly 38 MYA . We also used a recent publication on this snake from 2023 (https://www.iris.unina.it/ret rieve/19087cf1-7fb4-4df2-8eb2-ff00e72c3809/2023_Mezzasalma%20et%20al_Eur%20Zool%20J_C omparative%20cytogenetics%20Hemorrhois%20and%20Malpolon.pdf) to help guide us on the number of chromosomes we should expect. Expert curation helped us infer that the W chromosome is actually incomplete with loss of signal along the diagonal between many of the scaffolds that make it up. This also made it hard to put these scaffolds in the proper order and orientation. We selected the most contiguous region to be SUPER_W

and all the other SUPER_W-mapping regions being labeled as Unlocs. We share a folder with the pretext file and two saved states. In the second save state we just put W and Z next to one another."

Quality metrics table

| Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,753,287,546	1,752,149,986
GC %	42.26	42.27
Gaps/Gbp	39.92	45.66
Total gap bp	14,000	16,000
Scaffolds	66	60
Scaffold N50	121,923,953	121,923,953
Scaffold L50	5	5
Scaffold L90	16	16
Contigs	136	140
Contig N50	32,743,049	32,743,049
Contig L50	13	13
Contig L90	58	58
QV	48.5441	48.5443
Kmer compl.	95.4836	95.4648
BUSCO sing.	95.0%	95.0%
BUSCO dupl.	0.9%	0.9%
BUSCO frag.	1.1%	1.1%
BUSCO miss.	3.0%	3.0%

BUSCO 5.4.0 Lineage: tetrapoda_odb10 (genomes:38, BUSCOs:5310)
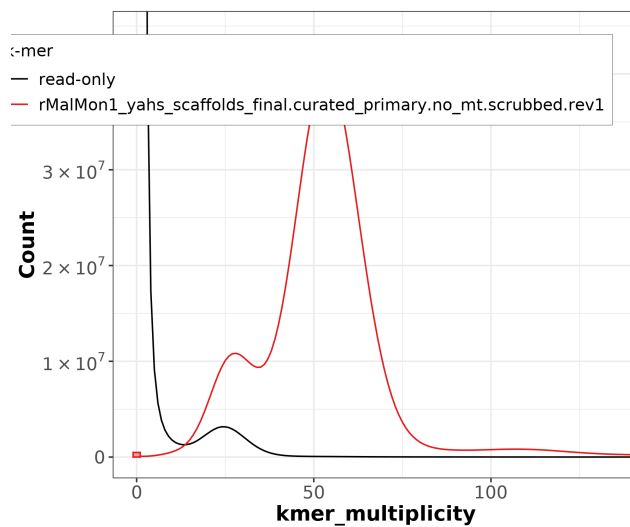
HiC contact map of curated assembly
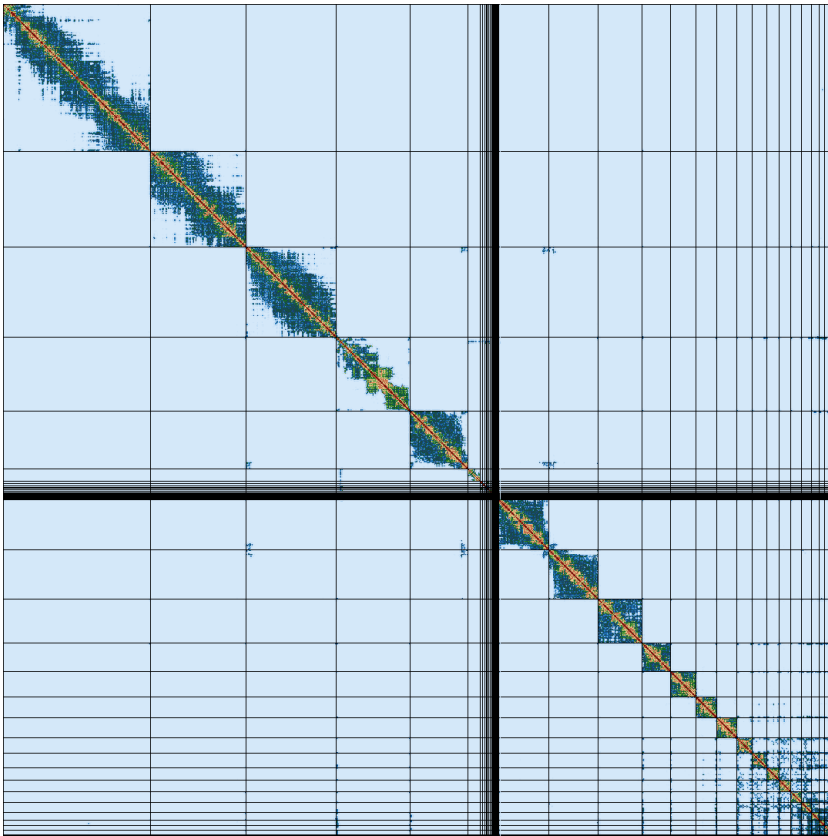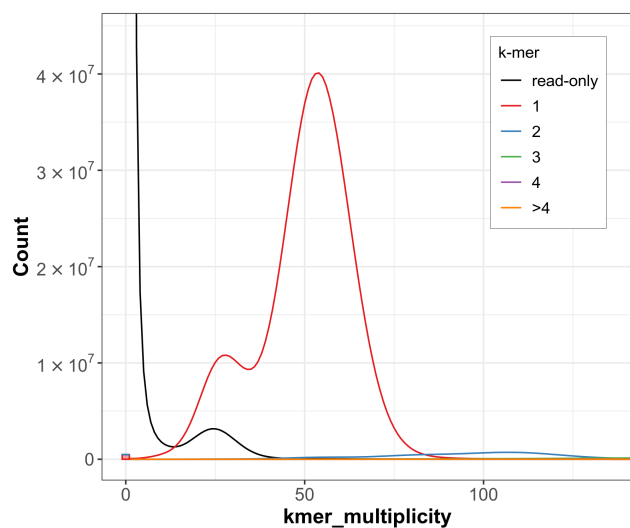

collapsed [LINK]

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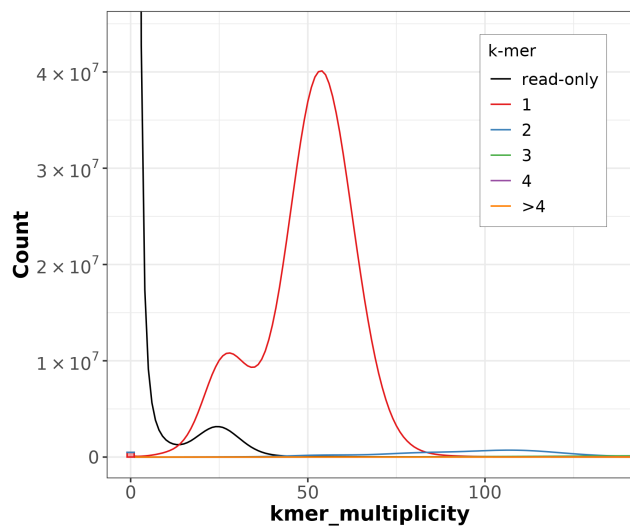
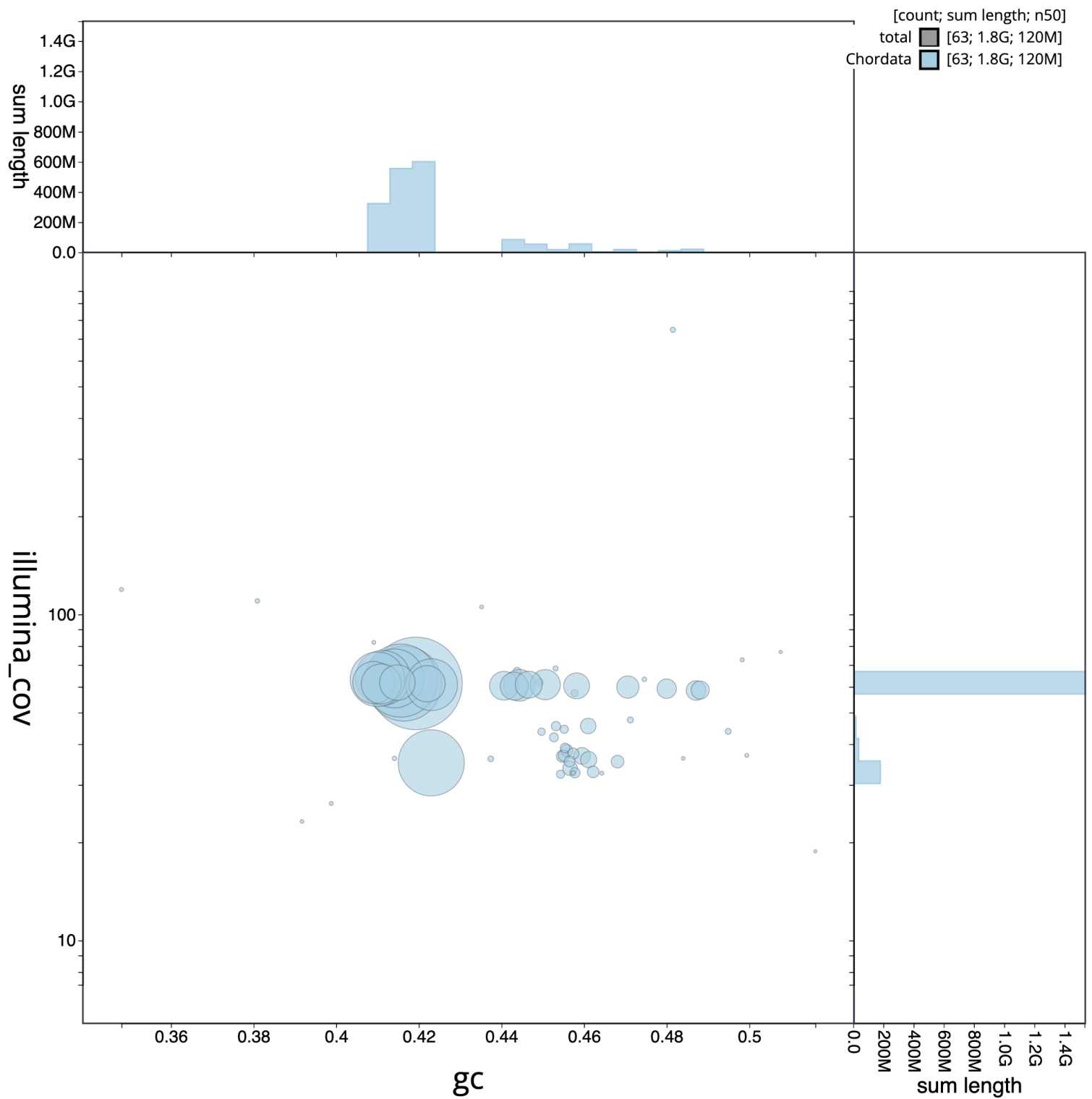Distribution of k-mer counts per copy
numbers found in asm

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	93x	62x	51x

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
- **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: 2024-06-05 12:37:53 CEST