ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	248254
ToLID	**mMonMoa1**
Species	Monachus monachus
Class	Mammalia
Order	Carnivora

Genome Traits	Expected	Observed
Haploid size (bp)	2,268,833,171	2,437,305,367
Haploid Number	17 (source: ancestor)	17
Ploidy	2 (source: ancestor)	1
Sample Sex	XX	XX

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q68

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

. Observed Ploidy is different from Expected

Curator notes

. Interventions/Gb: None
. Contamination notes: "FCS-GX and tiara detected no contamination."
. Other observations: "The Hi-C library for this species failed and it was not used in assembly or scaffolding. The hifiasm contigs were ordered and oriented through sequence alignments relative to the closest species available, Neomonachus schauinslandi (Hawaiian Monk Seal), using RagTag in scaffold mode. The contig-level assembly had very good stats already. We are reporting stats for the contig-level assembly (raw hifiasm contigs, pre-curation), and the curated RagTag scaffolds. Curation involved checking telomeres and coverage, RagTag scores for grouping, location and orientation, as well as visualizing synteny through protein alignments with N. schauinslandi. We have also prepared and checked Pretext maps with the Hi-C data but these are not very informative as you can see. In the final M. monachus scaffolds, 16 out of 17 putative chromosomes contain telomeres at both ends in proper orientation, the exception being scaf4. The 5' telomere for scaf2 is not at the tip but present in window 30-40kb. All other telomeres are at the very end of scaffolds. Mitochondria was assembled with mitohifi. This individual has extremely low heterozygosity and GenomeScope2 with ploidy=2 was giving a genome size estimate of 1.2Gb, I think because of the apparent lack of a heterozygous peak, so we changed ploidy to 1 to get a better estimate."

Quality metrics table

| Metrics	Pre-curation collapsed	Curated collapsed
Total bp	2,452,791,602	2,437,305,367
GC %	41.67	41.63
Gaps/Gbp	0	16.82
Total gap bp	0	4,100
Scaffolds	457	275
Scaffold N50	90,875,510	157,001,452
Scaffold L50	10	7
Scaffold L90	28	15
Contigs	457	316
Contig N50	90,875,510	90,875,510
Contig L50	10	10
Contig L90	28	27
QV	67.2327	68.6446
Kmer compl.	99.4876	99.4716
BUSCO sing.	97.7%	97.7%
BUSCO dupl.	1.6%	1.5%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	0.6%	0.7%

BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: carnivora_odb10 (genomes:12, BUSCOs:14502)
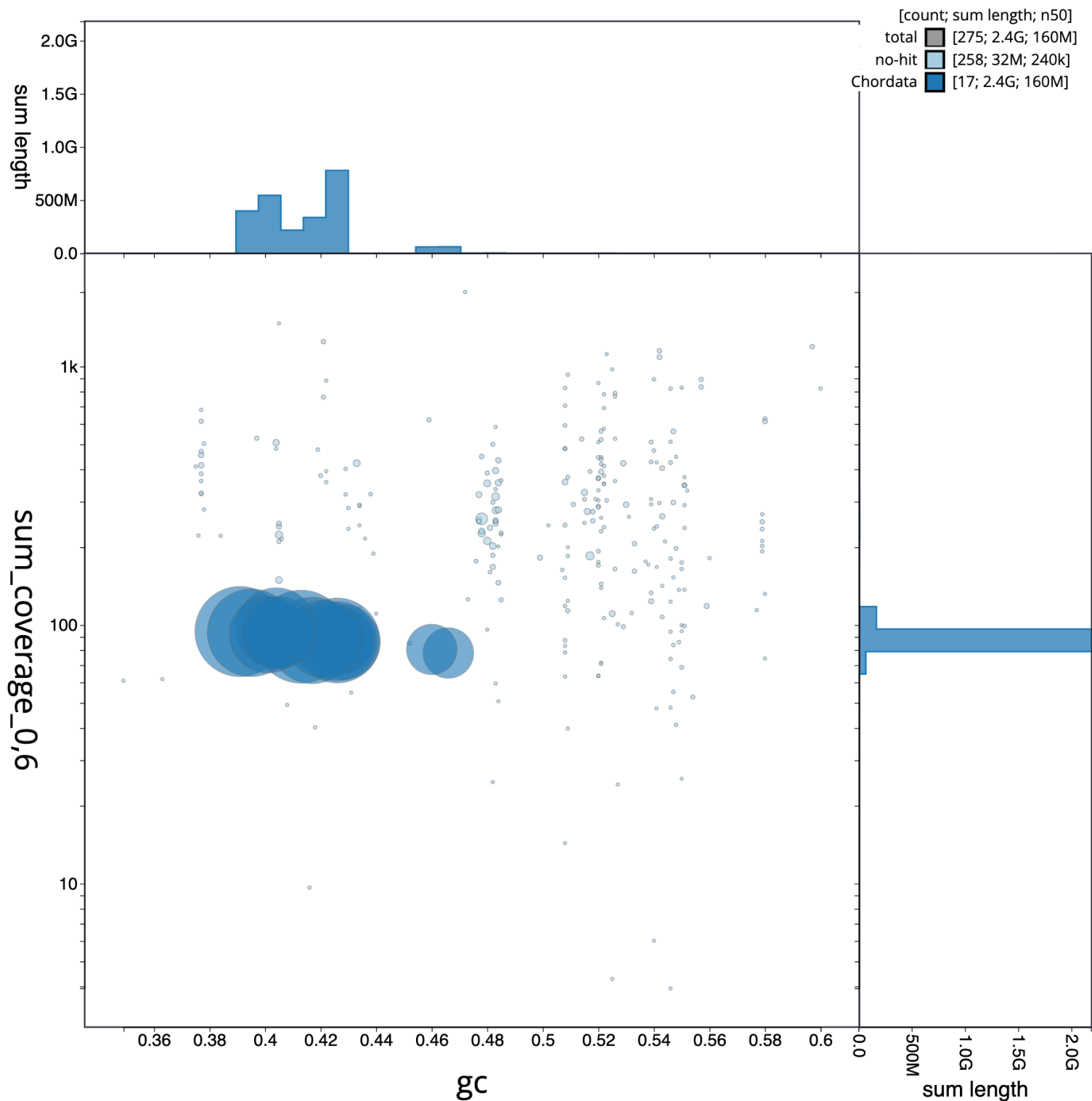
HiC contact map of curated assembly
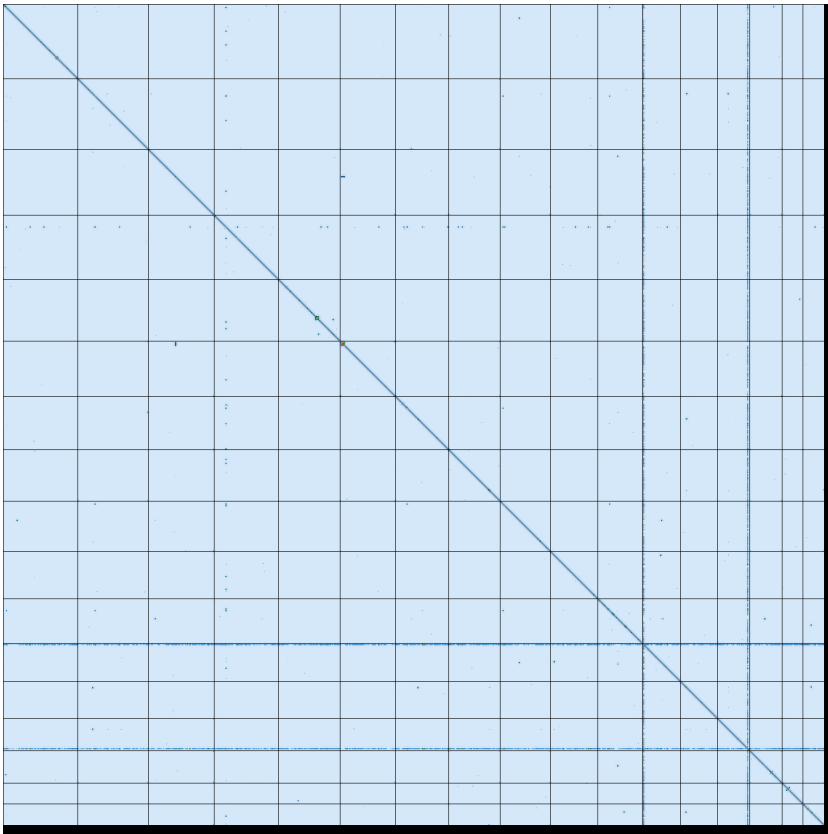
collapsed [LINK]

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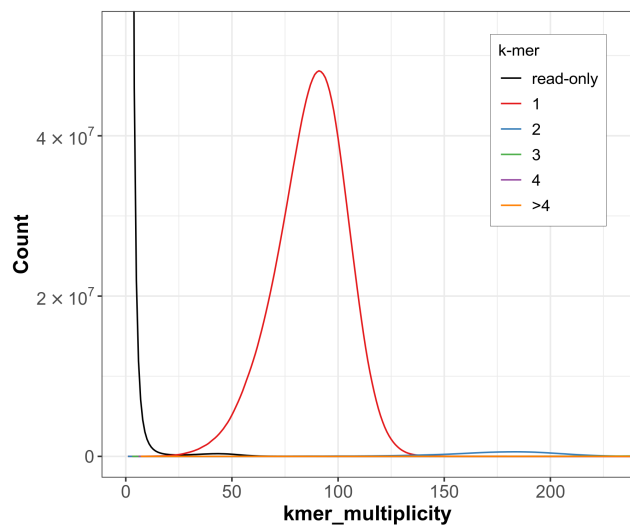
Distribution of k-mer counts per copy
numbers found in asm

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 HiC v2
Coverage	91x	2x

Assembly pipeline

- **hifiasm-hic**
    |_ *ver:* 0.25.0-r726
    |_ *key param:* NA
- **tiara**
    |_ *ver:* 1.0.3
    |_ *key param:* NA
- **RagTag**
    |_ *ver:* 2.1.0
    |_ *key param:* NA
- **NBIS/Earth-Biogenome-Project-pilot**
    |_ *ver:* 1d2a1e4
    |_ *key param:* NA
- **sanger-tol/blobtoolkit**
    |_ *ver:* 0.8.0
    |_ *key param:* NA

Curation pipeline

- **sanger-tol/curationpretext**
    |_ *ver:* 1.4.2
    |_ *key param:* NA

Submitter: Guilherme B. Dias
Affiliation: SciLifeLab

Date and time: 2025-09-21 15:55:12 CEST