ERGA Assembly Report

v24.10.15

Tags: Other[INVALID TAG]

TxID	40151
ToLID	**mMonMon2**
Species	Monodon monoceros
Class	Mammalia
Order	Artiodactyla

Genome Traits	Expected	Observed
Haploid size (bp)	34,818,102	2,341,957,608
Haploid Number	18 (source: ancestor)	22
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q18

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

. Observed Haploid size (bp) has >20% difference with Expected
. Observed Haploid Number is different from Expected

. QV value is less than 40 for collapsed
. Kmer completeness value is less than 90 for collapsed

Curator notes

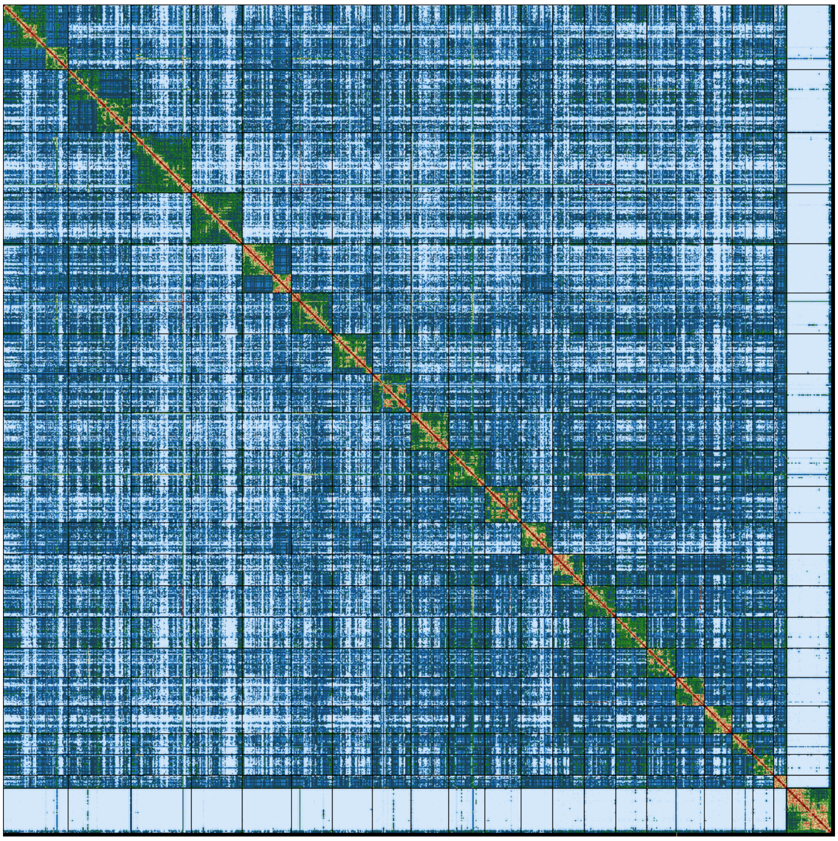
. Interventions/Gb: 14
. Contamination notes: ""
. Other observations: "The assembly of Monodon monoceros (mMonMon2) was generated as part of the Vertebrate Genomes Project (https://vertebrategenomesproject.org/). Chromosome 6 has half coverage, which makes it a good candidate for being the X chromosome, but no trace of a Y chromosome was found, even though the end of chromosome X aligns with the Y chromosome of Delphinus delphis. However, the terminal region of Chromosome 6 (118-125 Mb) has normal coverage compared to the rest of the chromosome, which may indicate the PAR region. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	2,341,938,226	2,341,957,608
GC %	41.26	41.26
Gaps/Gbp	133.65	137.92
Total gap bp	31,300	34,300
Scaffolds	100	91
Scaffold N50	108,465,051	108,465,051
Scaffold L50	9	9
Scaffold L90	19	19
Contigs	413	414
Contig N50	22,031,227	22,031,227
Contig L50	36	36
Contig L90	112	112
QV	18.1954	18.1954
Kmer compl.	32.9761	32.9761
BUSCO sing.	91.5%	91.5%
BUSCO dupl.	1.2%	1.2%
BUSCO frag.	2.0%	2.1%
BUSCO miss.	5.2%	5.2%

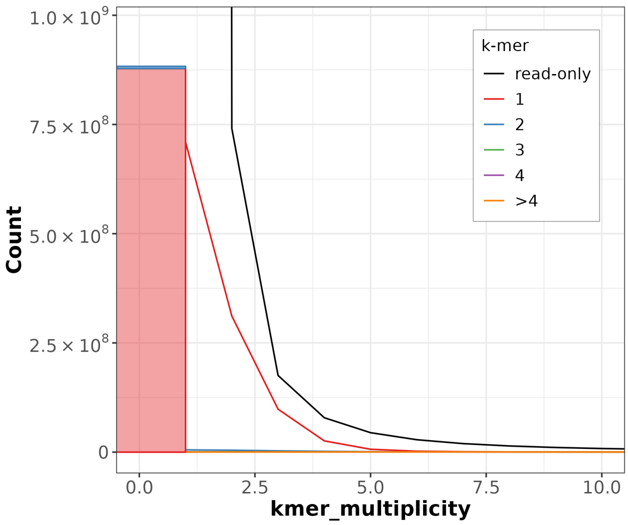
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: mammalia_odb12 (genomes:77, BUSCOs:12277)

HiC contact map of curated assembly

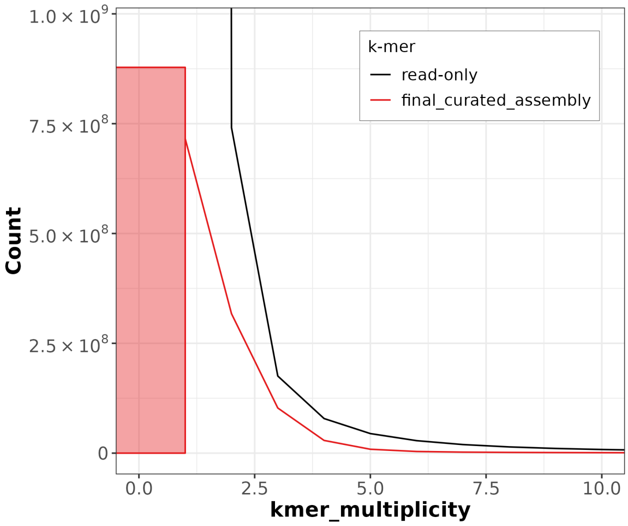


collapsed [LINK]

K-mer spectra of curated assembly

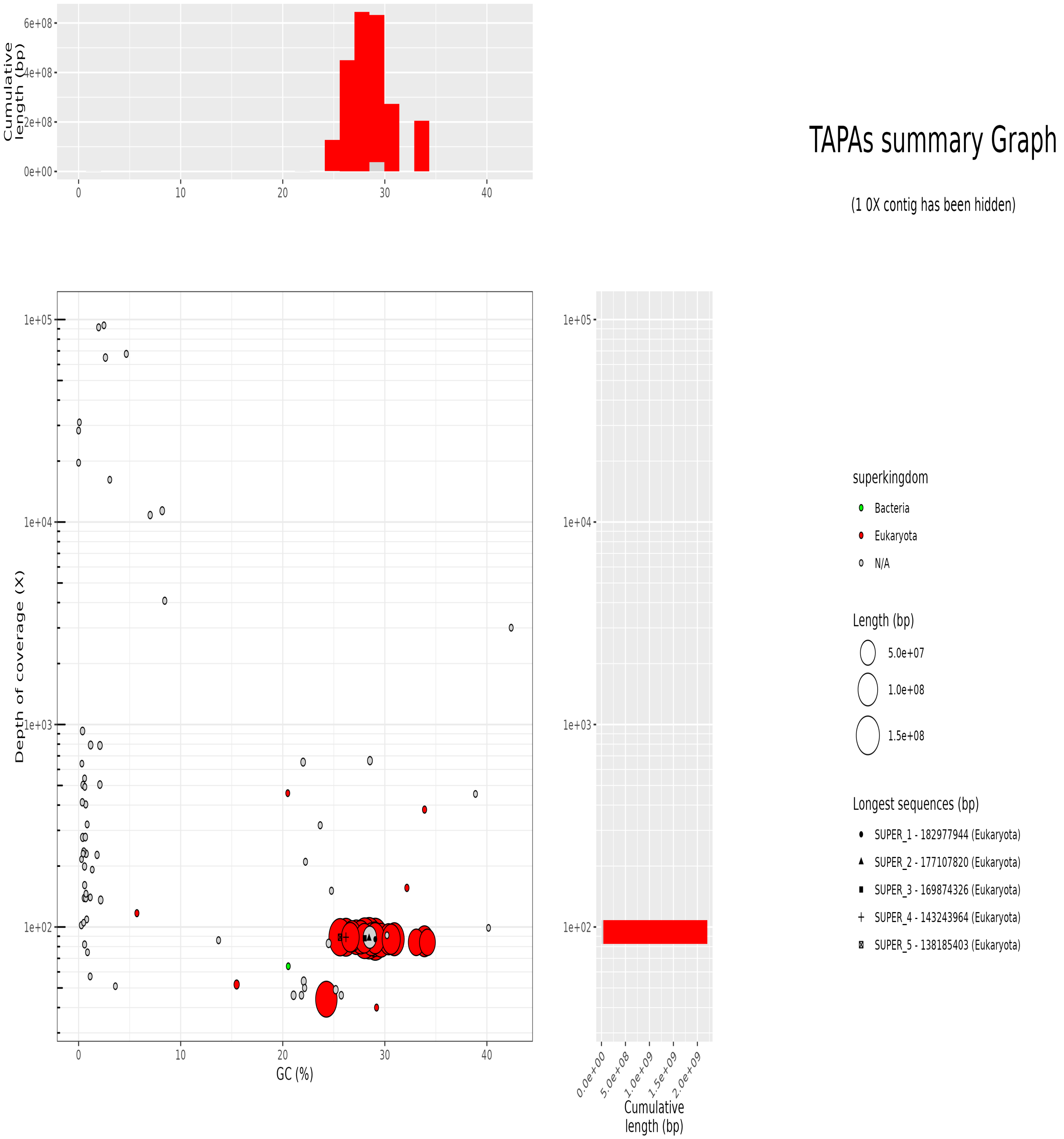


Distribution of k-mer counts per copy numbers found in asm



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

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	Long reads	Arima
Coverage	NA	NA

Assembly pipeline

- **Hifiasm**
  - |_ *ver:* 0.19.5-r593
  - |_ *key param:* NA
- **purge_dups**
  - |_ *ver:* 1.2.5
  - |_ *key param:* NA
- **YaHS**
  - |_ *ver:* 1.2
  - |_ *key param:* NA

Curation pipeline

- **PretextMap**
  - |_ *ver:* 0.1.9
  - |_ *key param:* NA
- **PretextView**
  - |_ *ver:* 0.2.5
  - |_ *key param:* NA

Submitter: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2025-10-30 09:30:29 CET