

ERGA Assembly Report

v24.09.10

Tags: ERGA-BGE

TxID	412090
ToLID	mPipHan1
Species	Pipistrellus hanaki
Class	Mammalia
Order	Chiroptera

Genome Traits	Expected	Observed
Haploid size (bp)	1,711,701,524	1,892,220,384
Haploid Number	22 (source: ancestor)	24
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Y

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q59

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

- . Observed Haploid Number is different from Expected
- . Observed sex is different from Sample sex
- . BUSCO duplicated value is more than 5% for collapsed

Curator notes

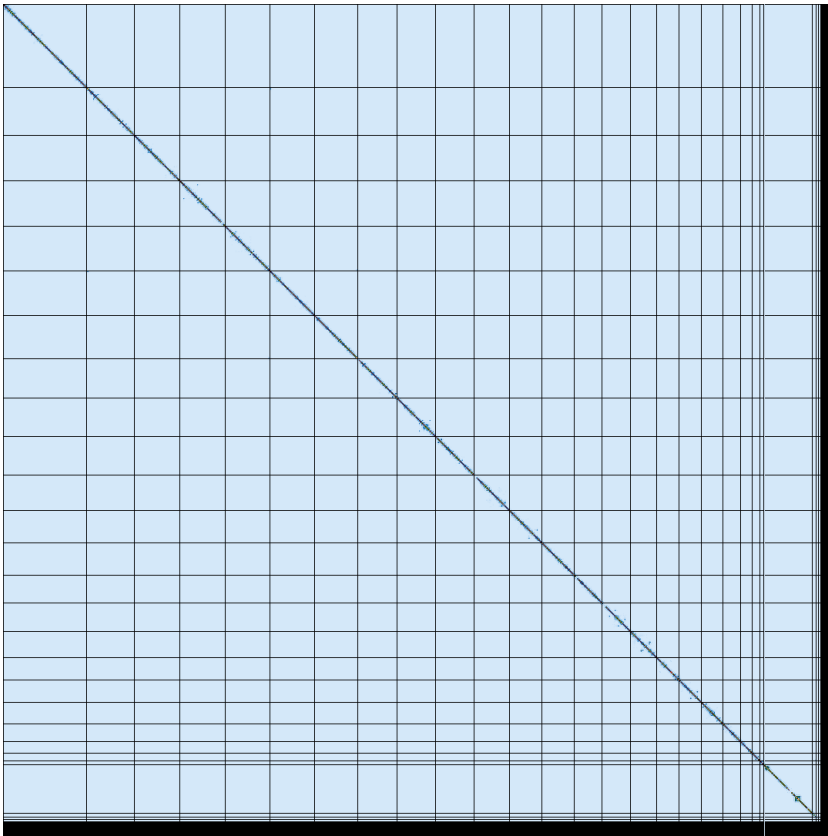
- . Interventions/Gb: 40
- . Contamination notes: ""
- . Other observations: ""

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,912,253,451	1,892,220,384
GC %	42.55	42.55
Gaps/Gbp	149.04	158.02
Total gap bp	28,500	32,500
Scaffolds	167	147
Scaffold N50	91,412,170	89,115,682
Scaffold L50	7	9
Scaffold L90	18	20
Contigs	452	446
Contig N50	45,232,785	49,626,680
Contig L50	16	15
Contig L90	70	70
QV	59.4962	59.4871
Kmer compl.	94.9997	94.6588
BUSCO sing.	90.8%	90.4%
BUSCO dupl.	7.9%	7.9%
BUSCO frag.	0.8%	0.8%
BUSCO miss.	0.5%	0.9%

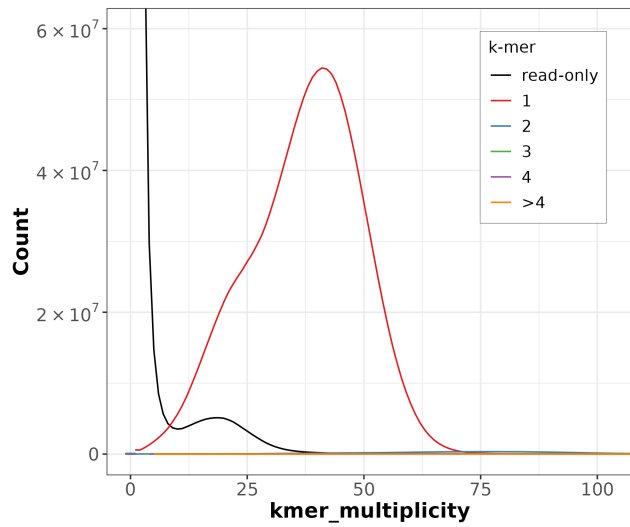
BUSCO 5.4.3 Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

HiC contact map of curated assembly

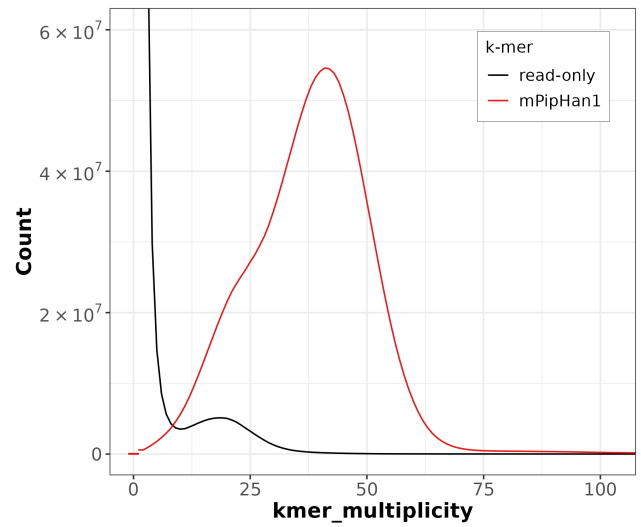


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

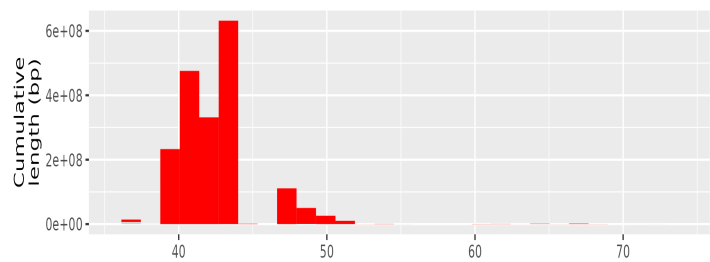


Distribution of k-mer counts per copy numbers found in asm

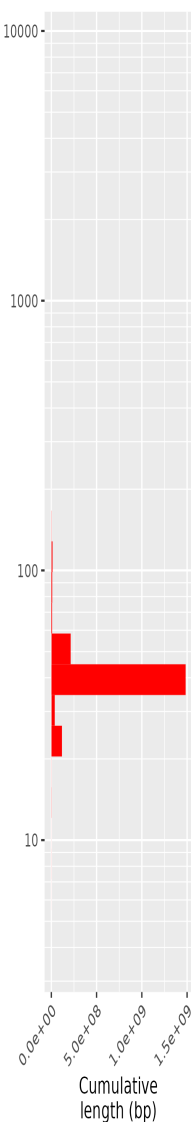
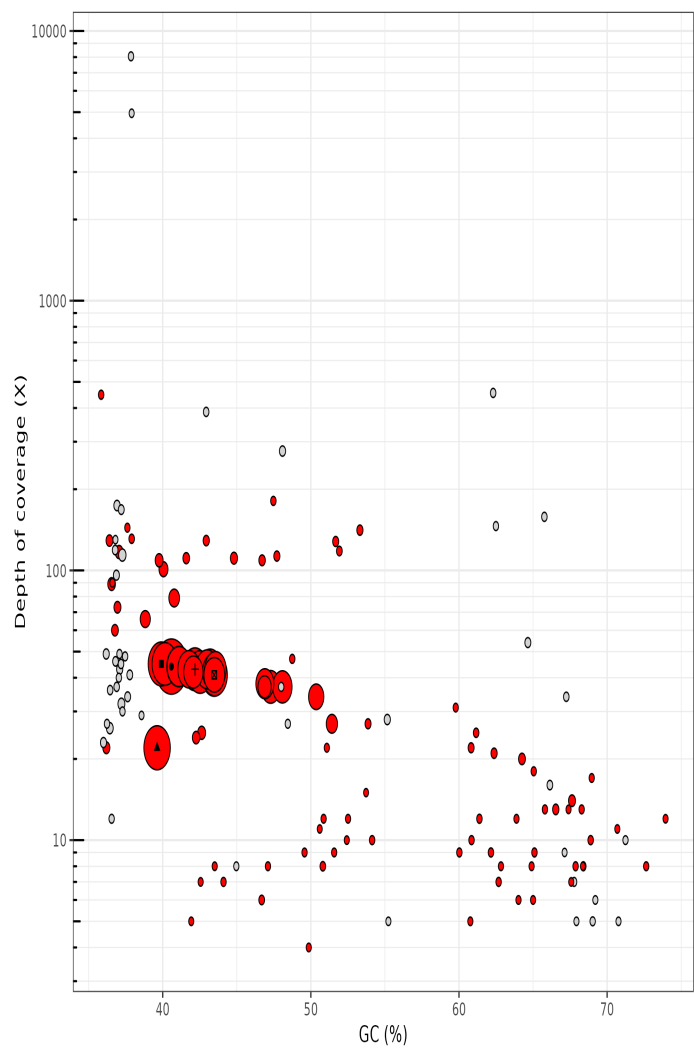


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



- Longest sequences (bp)
- SUPER_1 - 189558047 (Eukaryota)
 - ▲ Y - 110550902 (Eukaryota)
 - SUPER_2 - 109715185 (Eukaryota)
 - + SUPER_3 - 103029890 (Eukaryota)
 - ▣ SUPER_4 - 102958539 (Eukaryota)
- Length (bp)
- 5.0e+07
 - 1.0e+08
 - 1.5e+08
- superkingdom
- Eukaryota
 - N/A

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 (4-enz)
Coverage	36	19

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: Sophie Mangenot

Affiliation: Genoscope

Date and time: 2024-10-03 07:59:52 CEST