

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

TxID	110205
ToLID	rMacScw1
Species	Macrovipera schweizeri
Class	Lepidosauria
Order	Squamata

Genome Traits	Expected	Observed
Haploid size (bp)	1,563,548,624	1,635,265,452
Haploid Number	18 (source: ancestor)	18
Ploidy	3 (source: ancestor)	2
Sample Sex	M	M

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.8.Q54

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

- . Observed Ploidy is different from Expected

Curator notes

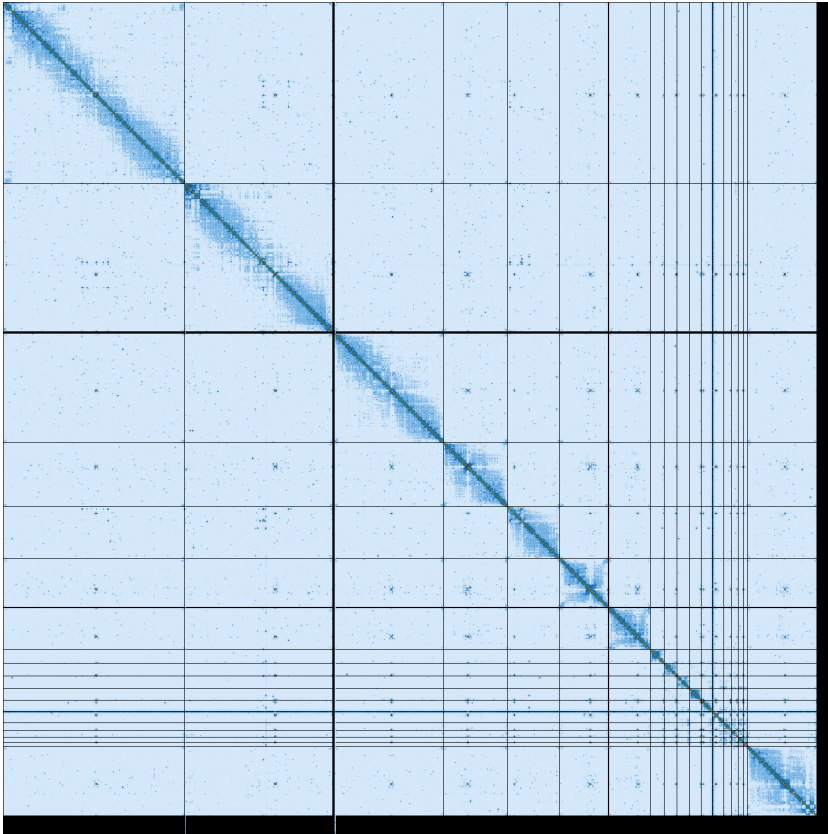
- . Interventions/Gb: 117
- . Contamination notes: "Contamination report for assembly labelled hap1; Total length of scaffolds removed: 4,310,242 (0.3 %); Scaffolds removed: 152 (20.4 %); Largest scaffold removed: (60,417); FCS-GX contaminant species (number of scaffolds; total length of scaffolds): Nanorana parkeri, amphibians (1; 4,000); Mitochondrion (151; 4,306,242)"
- . Other observations: "This genome has been assembled using PacBio and HiC data and phased. 17 autosomes and 1 ZZ sex chromosome identified. Z chromosome assigned based on synteny to Vipera berus (GCA_964194415.1)."

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	1,641,802,861	1,635,265,452
GC %	40.64	40.63
Gaps/Gbp	797.29	834.12
Total gap bp	130,900	143,600
Scaffolds	746	534
Scaffold N50	139,972,625	214,481,692
Scaffold L50	5	3
Scaffold L90	14	11
Contigs	2,055	1,898
Contig N50	3,155,423	3,162,383
Contig L50	144	143
Contig L90	640	629
QV	54.8	54.9
Kmer compl.	99.48	99.48
BUSCO sing.	92.0%	92.0%
BUSCO dupl.	1.4%	1.4%
BUSCO frag.	0.8%	0.8%
BUSCO miss.	5.8%	5.8%

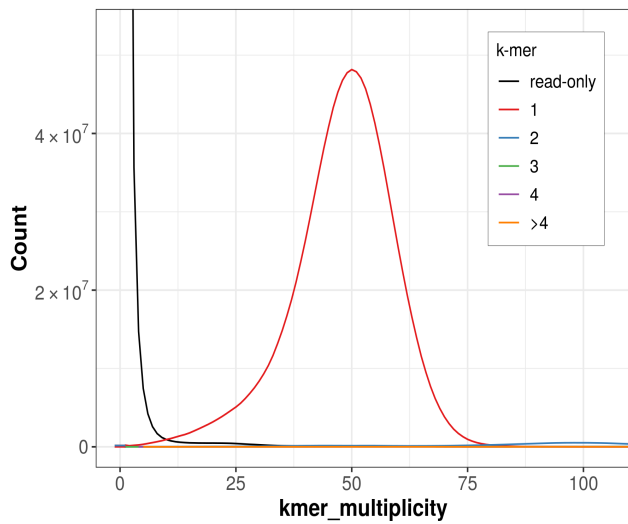
BUSCO 5.3.0 Lineage: sauropsida_odb10 (genomes:76, BUSCOs:7480)

HiC contact map of curated assembly

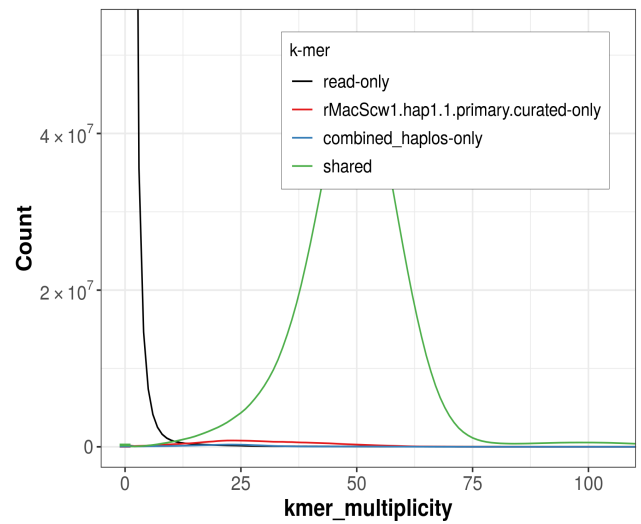


pri [\[LINK\]](#)

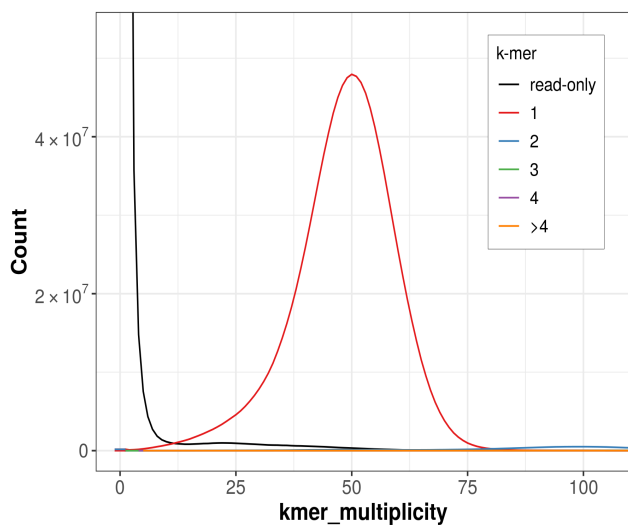
K-mer spectra of curated assembly



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

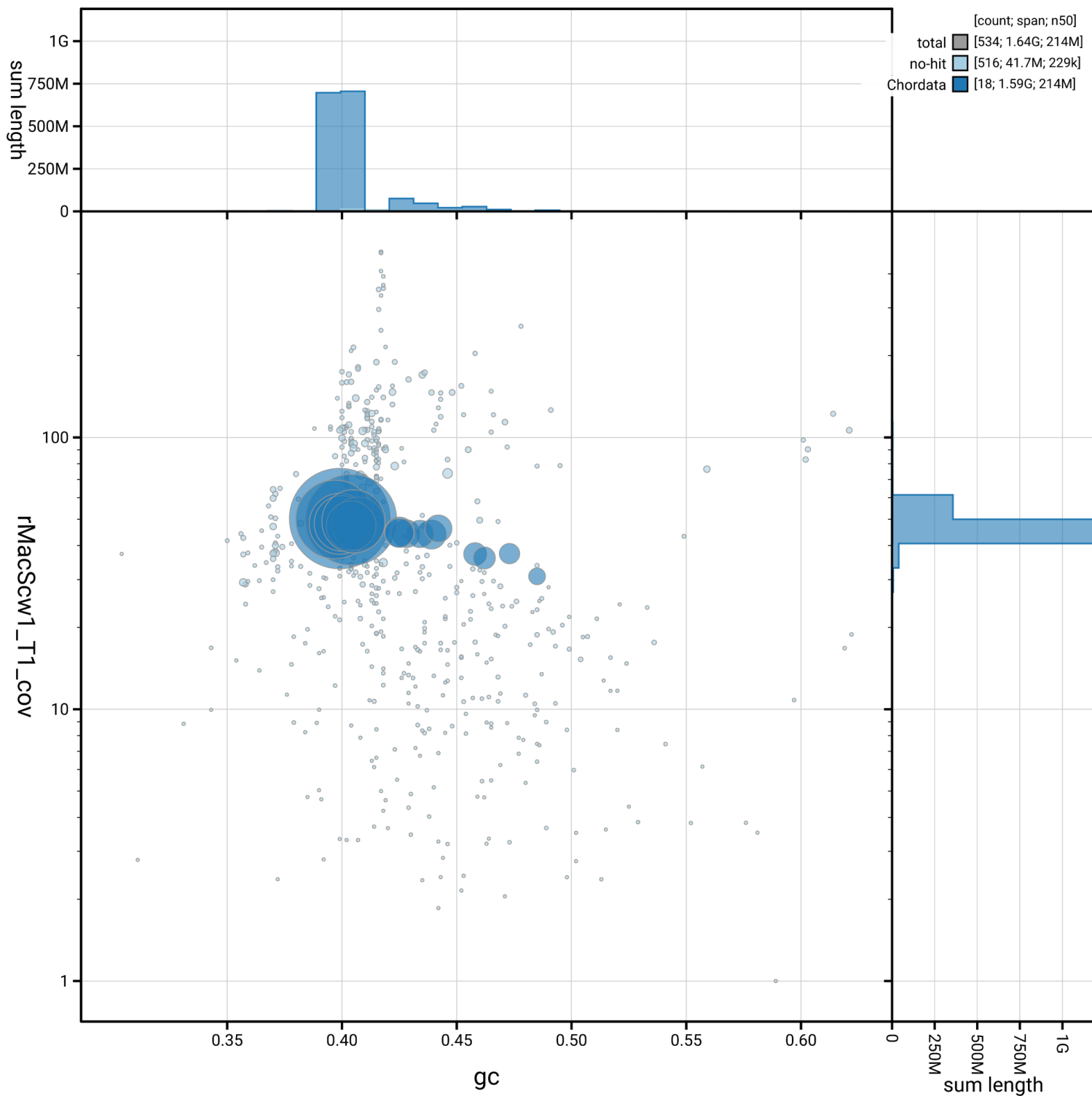


Distribution of k-mer counts coloured by their presence in reads/assemblies



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

Post-curation contamination screening



pri. 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 v2
Coverage	50x	100x

Assembly pipeline

- **hifiasm-hic**
 - |_ *ver*: 0.19.8-r603
 - |_ *key param*: --h1/--h2
- **yahs**
 - |_ *ver*: 1.2.2
 - |_ *key param*: NA

Curation pipeline

- **hifiasm-hic**
 - |_ *ver*: 0.19.8-r603/--h1/--h2
 - |_ *key param*: NA
- **yahs**
 - |_ *ver*: 1.2.2
 - |_ *key param*: NA
- **TreeVal**
 - |_ *ver*: 1.2.1
 - |_ *key param*: NA

Submitter: Jo Collins

Affiliation: WSI

Date and time: 2025-02-26 12:47:05 CET