

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

TxID	2546182
ToLID	igCheThal5
Species	Chelidura thaleri
Class	Insecta
Order	Dermaptera

Genome Traits	Expected	Observed
Haploid size (bp)	632,131,113	648,254,250
Haploid Number	12 (source: ancestor)	24
Ploidy	2 (source: ancestor)	2
Sample Sex	F	F

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.7.Q61

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

- . Observed Haploid Number is different from Expected

Curator notes

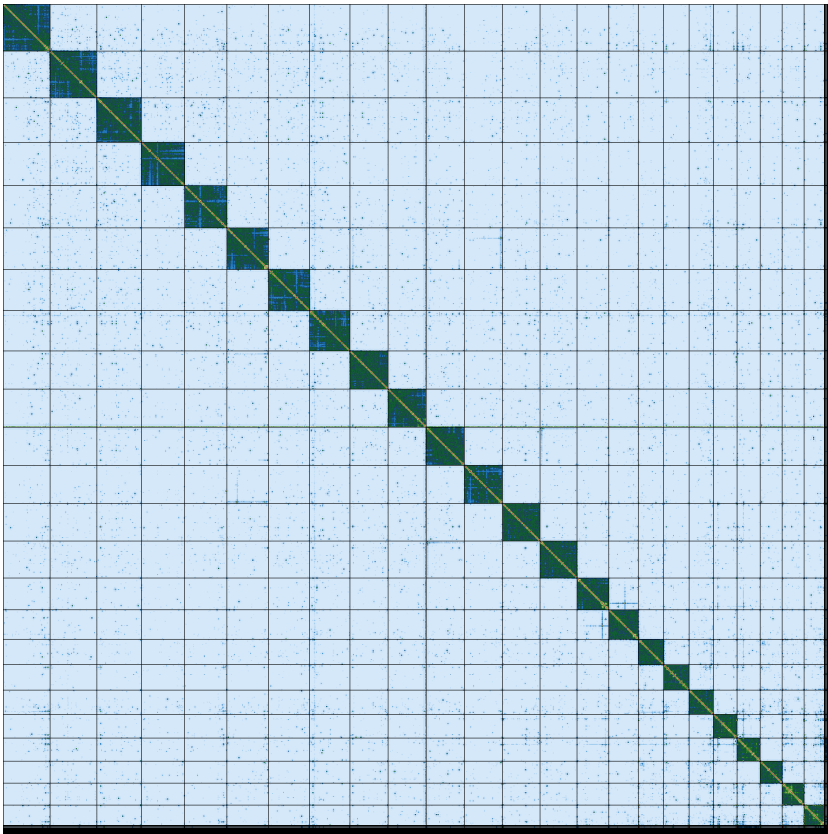
- . Interventions/Gb: 551
- . Contamination notes: "Contamination report for assembly labelled primary; Total length of scaffolds removed: 1,389,335 (0.2 %); Scaffolds removed: 42 (38.9 %); Largest scaffold removed: (89,911); FCS-GX contaminant species (number of scaffolds; total length of scaffolds): Pseudomonas shirazensis, g-proteobacteria (32; 1,101,785); Pseudomonas reidholzensis, g-proteobacteria (2; 68,525); Rhodococcus fascians, high GC Gram+ (2; 30,104); Pseudomonas brassicae, g-proteobacteria (1; 24,215); Rhodococcus kyotonensis, high GC Gram+ (1; 21,219); Mitochondrion (4; 143,487)"
- . Other observations: "Hi-C was from a second individual (igCheThal4), so assembly is not phased. The sex chromosomes were not assigned, as the sample was homogametous and the closest comparators (Labia minor / GCA_963082975.1) sex chromosomes map to multiple ones."

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	648,809,072	648,254,250
GC %	30.33	30.29
Gaps/Gbp	299.01	345.54
Total gap bp	19,400	25,500
Scaffolds	108	61
Scaffold N50	29,721,480	29,769,575
Scaffold L50	10	10
Scaffold L90	21	21
Contigs	302	285
Contig N50	5,209,279	4,844,708
Contig L50	41	44
Contig L90	133	143
QV	61.1	61.5
Kmer compl.	99.66	99.69
BUSCO sing.	94.6%	95.1%
BUSCO dupl.	3.6%	3.4%
BUSCO frag.	1.1%	0.8%
BUSCO miss.	0.7%	0.7%

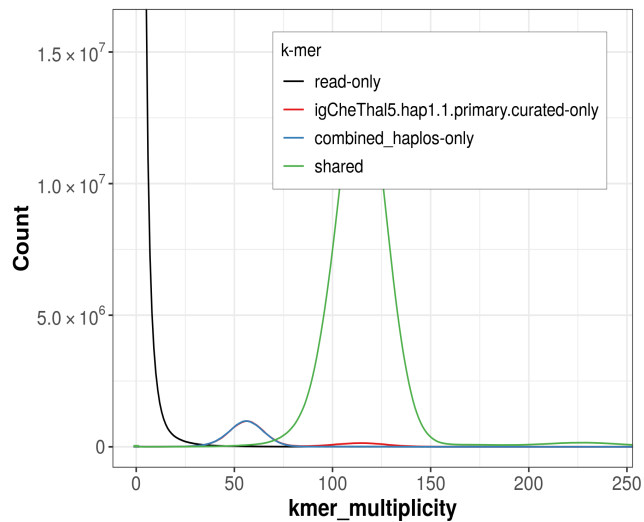
Warning: BUSCO versions or lineage datasets are not the same across results

HiC contact map of curated assembly

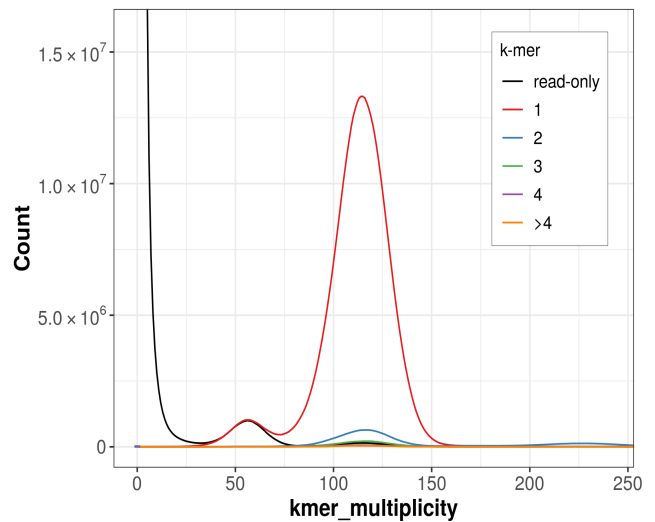


pri [\[LINK\]](#)

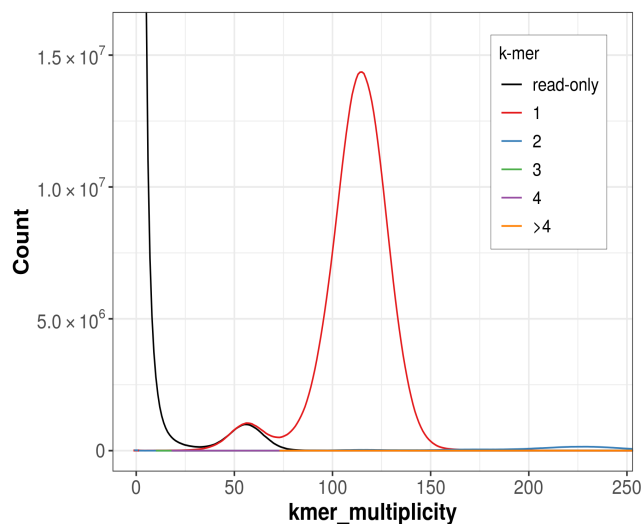
K-mer spectra of curated assembly



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

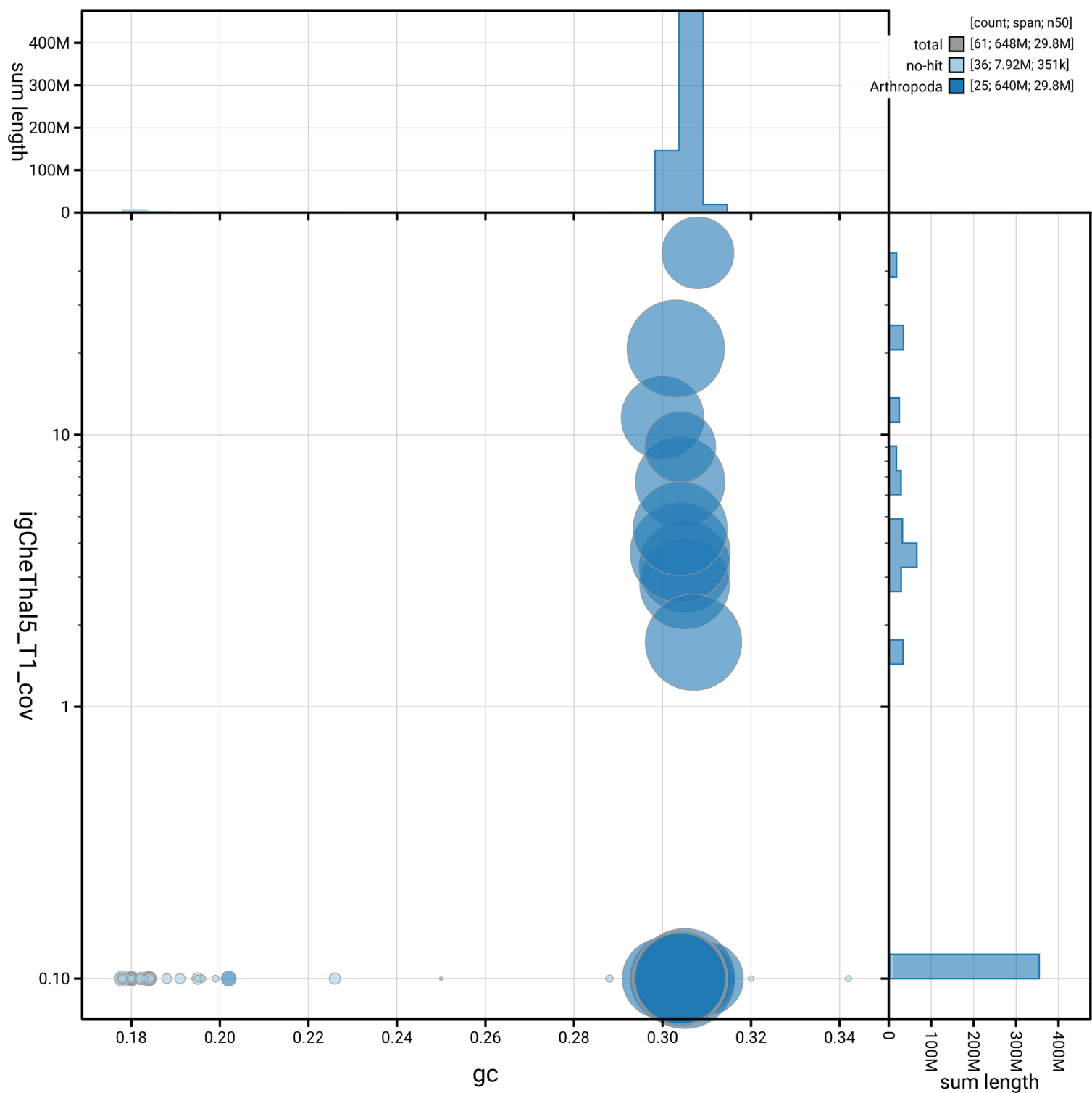


Distribution of k-mer counts per copy numbers found in asm



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	115x	152x

Assembly pipeline

- **hifiasm**
 - |_ *ver*: 0.19.8-r603
 - |_ *key param*: --primary
- **yahs**
 - |_ *ver*: 1.2.2
 - |_ *key param*: NA

Curation pipeline

- **hifiasm**
 - |_ *ver*: 0.19.8-r603
 - |_ *key param*: --primary
- **yahs**
 - |_ *ver*: 1.2.2
 - |_ *key param*: NA
- **TreeVal**
 - |_ *ver*: 1.2.1
 - |_ *key param*: NA

Submitter: Michael Paulini

Affiliation: WSI

Date and time: 2025-02-17 14:29:10 CET