

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

TxID	204053
ToLID	qmScyLatul
Species	Scyllarides latus
Class	Malacostraca
Order	Decapoda

Genome Traits	Expected	Observed
Haploid size (bp)	4,658,004,081	4,733,579,267
Haploid Number	70 (source: ancestor)	70
Ploidy	2 (source: ancestor)	2
Sample Sex	U	U

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.7.Q56

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

. Assembly length loss > 3% for pri

Curator notes

. Interventions/Gb: 278
. Contamination notes: "Contamination report for assembly labelled hap1; Total length of scaffolds removed: 329,932 (0.0 %); Scaffolds removed: 6 (0.1 %); Largest scaffold removed: (130,169); FCS-GX contaminant species (number of scaffolds; total length of scaffolds): Pseudomonas coleopterorum, g-proteobacteria (2; 195,759); Mariniblastus fucicola, bacteria (1; 36,546); Mitochondrion (3; 97,627)"
. Other observations: "Assembly was Hi-C phased; The exact order and orientation of the contigs on chromosome 1 (66,839 - 90,907 Kbp) are unknown; The exact order and orientation of the contigs on chromosome 6 (84,960 - 70,000 & 107,580 - 110,190 Kbp) are unknown; The exact order and orientation of the contigs on chromosome 7 (40,500 - 45,500 & 57,850 - 73,500 Kbp) are unknown; The exact order and orientation of the contigs on chromosome 9 (61,000 - 80,000 Kbp) are unknown; The exact order and orientation of the contigs on chromosome 11 (0 - 15,000 Kbp) are unknown; The exact order and orientation of the contigs on chromosome 12 (2,000 - 18,800 Kbp) are unknown; The exact order and orientation of the contigs on chromosome 14 (54,300 - 62,000 Kbp) are unknown; The exact order and orientation of the contigs on chromosome 17 (35,000 - 50,000 Kbp) are unknown; The exact order and orientation of the contigs on chromosome 18 (0 - 5,000 Kbp) are unknown; The exact order and orientation of the contigs on chromosome 21 (28,700 - 41,300 Kbp) are unknown; The exact order and

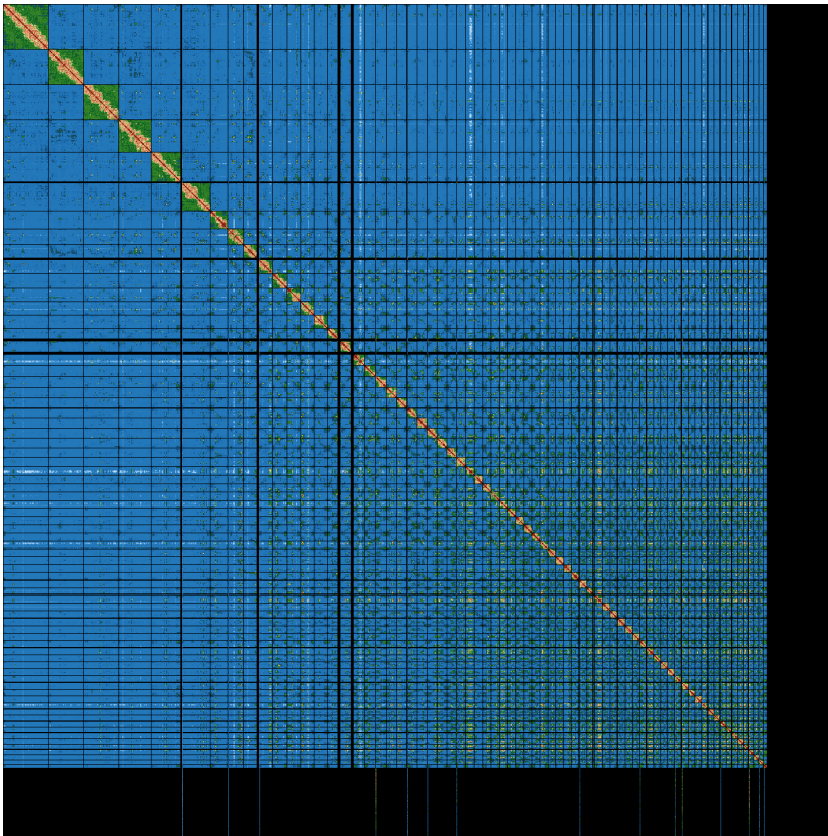
orientation of the contigs on chromosome 23 (0 - 16,300 Kbp) are unknown; The exact order and orientation of the contigs on chromosome 27 (13,000 - 24,000 Kbp) are unknown; The exact order and orientation of the contigs on chromosome 33 (17,500 - 39,500 Kbp) are unknown; The exact order and orientation of the contigs on chromosome 34 (10,700 - 20,700 Kbp) are unknown; The exact order and orientation of the contigs on chromosome 36 (5,800 - 16,000 Kbp) are unknown; The exact order and orientation of the contigs on chromosome 41 (2,700 - 25,800 Kbp) are unknown; The exact order and orientation of the contigs on chromosome 45 (0 - 10,800 Kbp) are unknown; The exact order and orientation of the contigs on chromosome 48 (21,300 - 35,300 Kbp) are unknown; The exact order and orientation of the contigs on chromosome 60 (29,000 - 8,000 & 19,500 - 36,246 Kbp) are unknown; The exact order and orientation of the contigs on chromosome 64 (22,000 - 32,500 Kbp) are unknown; The exact order and orientation of the contigs on chromosome 67 (2,300 - 8,000 & 19,500 - 26,000 Kbp) are unknown"

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	5,323,139,245	4,733,579,267
GC %	42.84	42.99
Gaps/Gbp	622	852.42
Total gap bp	662,200	806,990
Scaffolds	10,129	4,456
Scaffold N50	40,182,599	56,908,864
Scaffold L50	38	23
Scaffold L90	946	70
Contigs	13,440	8,491
Contig N50	1,599,401	1,680,176
Contig L50	920	821
Contig L90	4,030	3,148
QV	56.7	56.8
Kmer compl.	98.92	98.89
BUSCO sing.	92.9%	92.8%
BUSCO dupl.	1.6%	1.4%
BUSCO frag.	3.6%	3.6%
BUSCO miss.	1.9%	2.2%

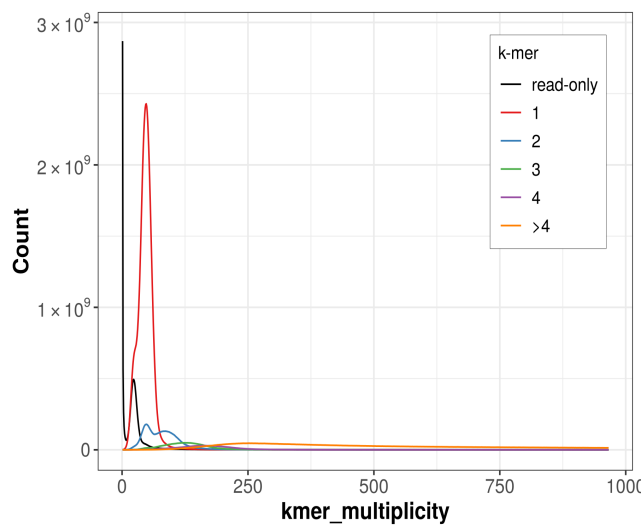
BUSCO 5.3.0 Lineage: arthropoda_odb10 (genomes:90, BUSCOs:1013)

HiC contact map of curated assembly

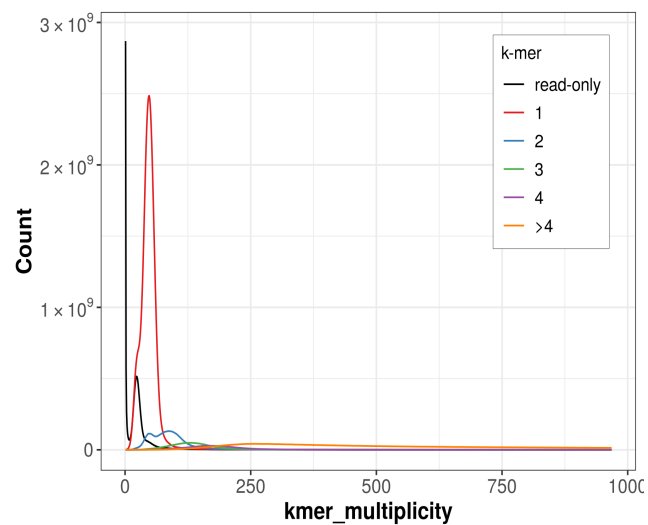


pri [\[LINK\]](#)

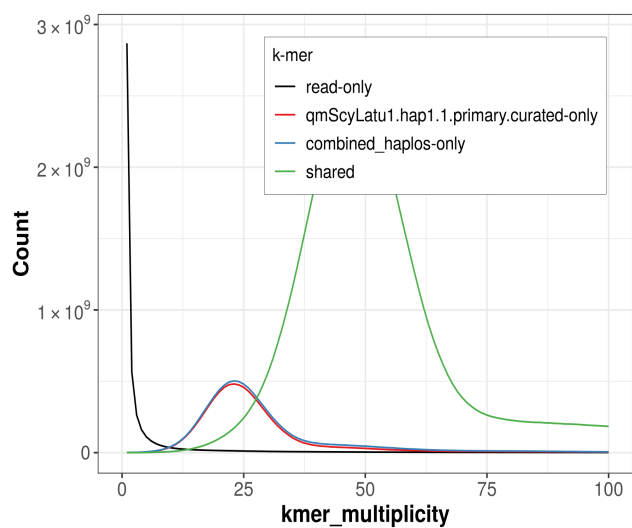
K-mer spectra of curated assembly



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

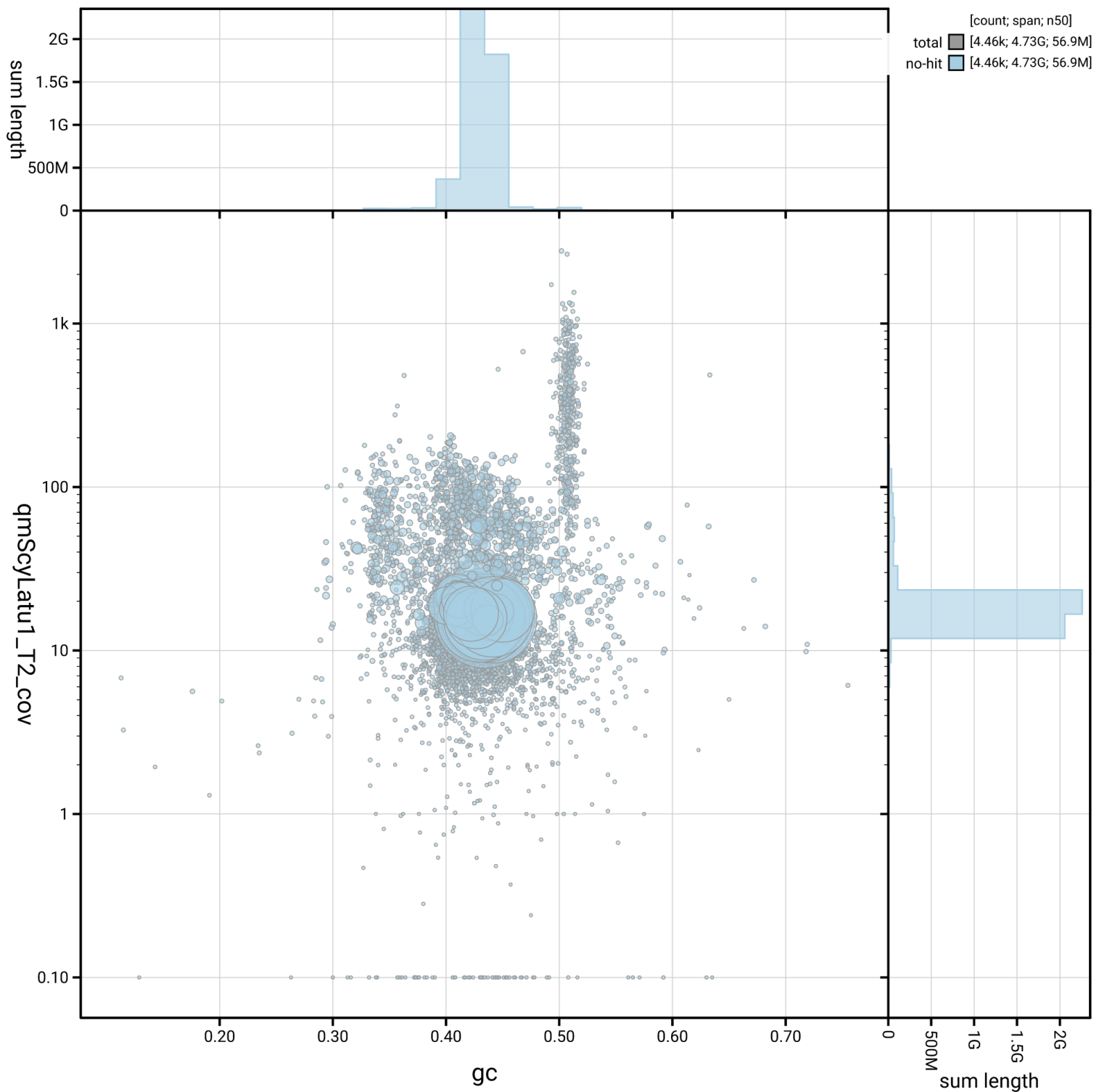


Distribution of k-mer counts per copy numbers found in asm



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

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	47x	175x

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

Submitter: Karen Brooks

Affiliation: WSI

Date and time: 2025-05-09 11:54:41 CEST