ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	3198454
ToLID	**xbLimCari2**
Species	Lima caribaea
Class	Bivalvia
Order	Limoida

Genome Traits	Expected	Observed
Haploid size (bp)	1,574,497,015	1,636,194,742
Haploid Number	14 (source: ancestor)	20
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q50

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

. Observed Haploid Number is different from Expected

. Kmer completeness value is less than 90 for collapsed
. Assembly length loss > 3% for collapsed

Curator notes

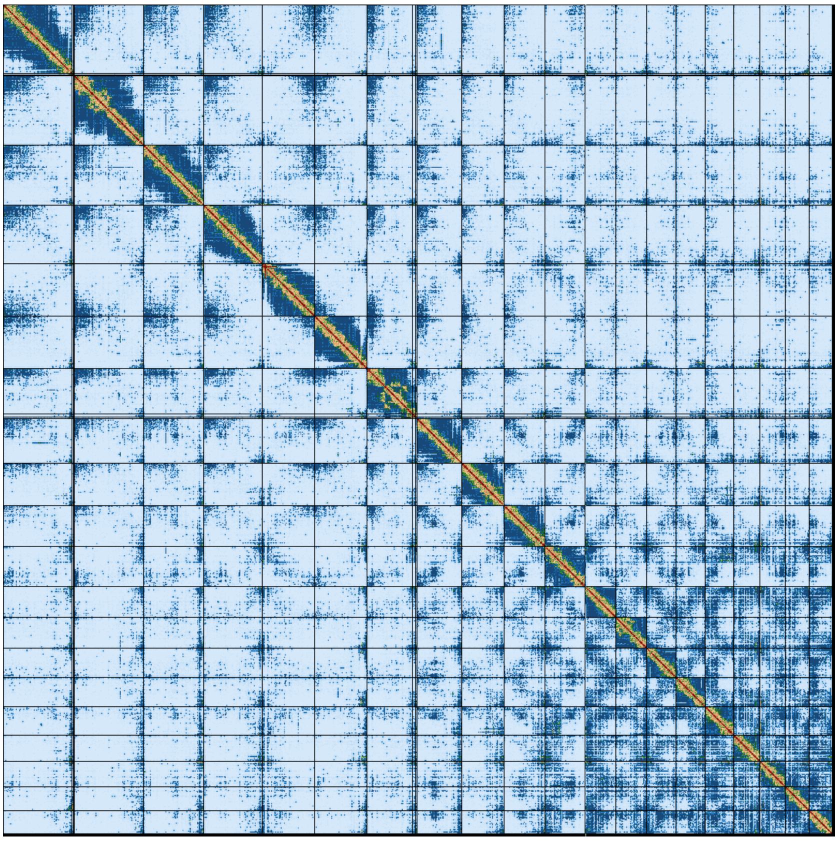
. Interventions/Gb: 37
. Contamination notes: ""
. Other observations: "The assembly of Lima caribaea (xbLimCari2) is based on 58X PacBio data and 202X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 2 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.063 Mb (with the largest being 0.041 Mb). Additionally, 467 regions totaling 256.132 Mb (with the largest being 11.429 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. During manual curation, 15 haplotypic regions were removed, totaling 101,625,18 pb (with the largest being 26,019,980 pb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,737,852,810	1,636,194,742
GC %	37.49	37.49
Gaps/Gbp	140.4	135.07
Total gap bp	25,700	24,900
Scaffolds	66	56
Scaffold N50	88,802,729	88,922,029
Scaffold L50	8	8
Scaffold L90	19	17
Contigs	302	277
Contig N50	17,359,956	17,215,279
Contig L50	26	24
Contig L90	112	104
QV	50.0085	50.0269
Kmer compl.	66.6045	64.1208
BUSCO sing.	91.1%	97.0%
BUSCO dupl.	7.0%	1.1%
BUSCO frag.	0.5%	0.5%
BUSCO miss.	1.4%	1.4%

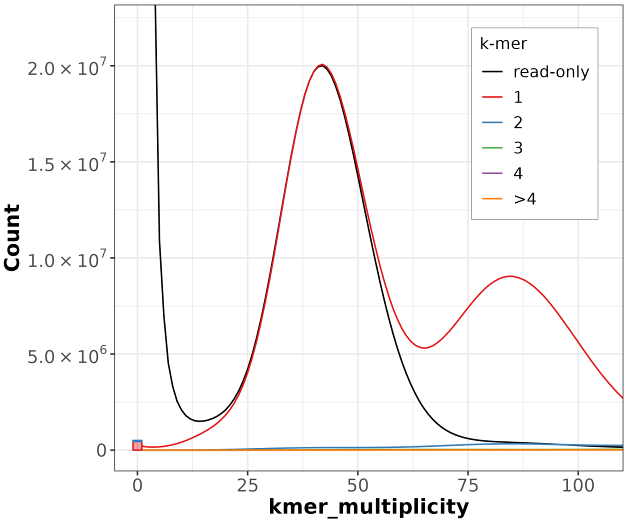
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

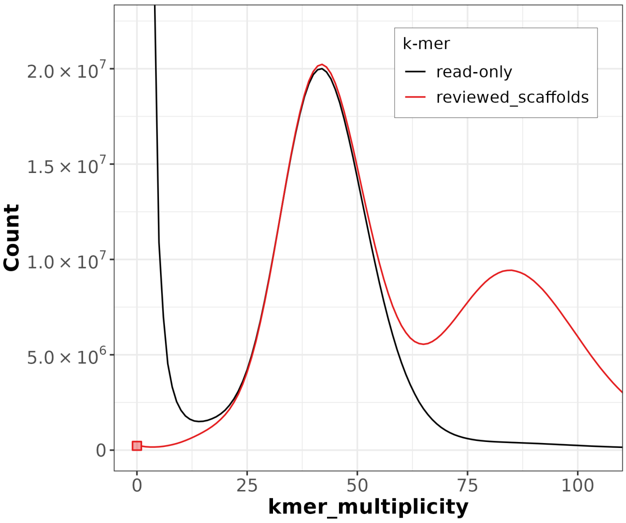


collapsed [LINK]

K-mer spectra of curated assembly

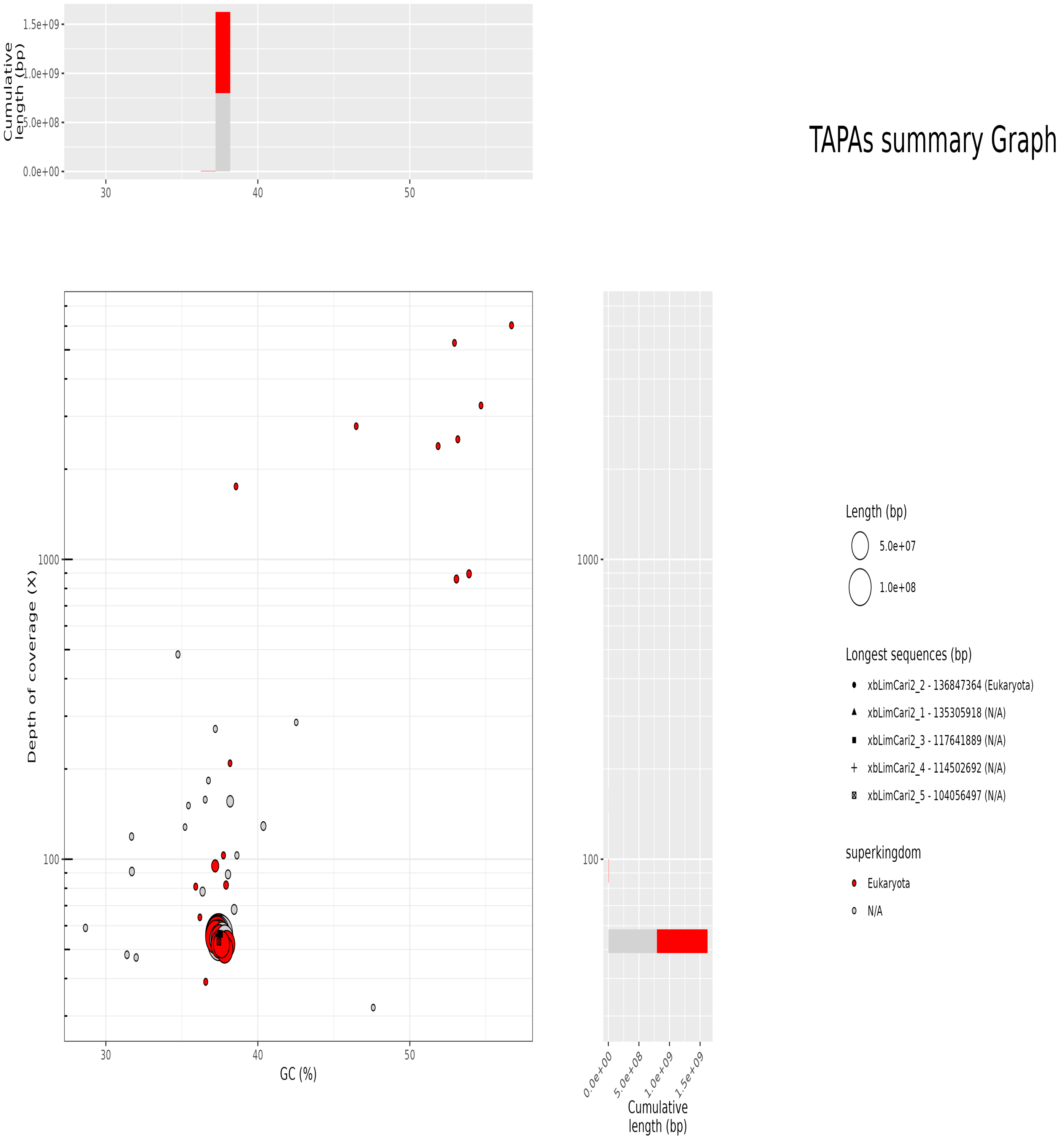


Distribution of k-mer counts per copy numbers found in asm



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	58	202

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: Lola Demirdjian
Affiliation: Genoscope

Date and time: 2025-12-07 11:16:28 CET