ERGA Assembly Report
v24.10.15

Tags: ERGA-BGE

TxID	618066
ToLID	**icLucTetr**
Species	Lucanus tetraodon
Class	Insecta
Order	Coleoptera

Genome Traits	Expected	Observed
Haploid size (bp)	360,638,005	339,386,420
Haploid Number	11 (source: ancestor)	21
Ploidy	2 (source: ancestor)	2
Sample Sex	XY	XY

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q39

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

. Observed Haploid Number is different from Expected

. Assembly length loss > 3% for collapsed

Curator notes

. Interventions/Gb: 273
. Contamination notes: "No contaminants were detected."
. Other observations: "The assembly of species Lucanus tetraodon (icLucTetr) is based on 150X long read ONT data and 85X Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA, https://www.erga-biodiversity.eu/) via the Biodiversity Genomics Europe project (BGE, https://biodiversitygenomics.eu/). The assembly process included the following steps: ONT reads shorter than 3 kb were filtered out, thus the remaining reads were assembled using Necat with polishing performed using NextPolish. Contigs were subsequently filtered with purge_dups to remove haplotigs and obtain a more accurate, non-redundant assembly. Scaffolding was performed by aligning Hi-C reads to the purged contigs using the Omni-C mapping pipeline, followed by YaHS for scaffold construction. Contamination was checked using BlobTools, and no contaminants were detected. The scaffolds were processed through sanger_tol/curationpretext pipeline to generate the contact map, which was manually improved using PretextView. The presence of a mitochondrial genome contig/scaffold was not assessed. 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	362,404,194	339,386,420
GC %	35.67	35.66
Gaps/Gbp	711.91	860.38
Total gap bp	25,800	36,900
Scaffolds	356	299
Scaffold N50	18,996,966	19,915,458
Scaffold L50	8	7
Scaffold L90	28	19
Contigs	614	591
Contig N50	1,875,229	2,042,896
Contig L50	51	47
Contig L90	211	192
QV	39.8955	39.6768
Kmer compl.		
BUSCO sing.	92.5%	96.7%
BUSCO dupl.	5.3%	0.9%
BUSCO frag.	0.7%	0.5%
BUSCO miss.	1.6%	1.9%

BUSCO: 6.0.0 (euk_genome_aug, augustus) / Lineage: endopterygota_odb10 (genomes:56, BUSCOs:2124)
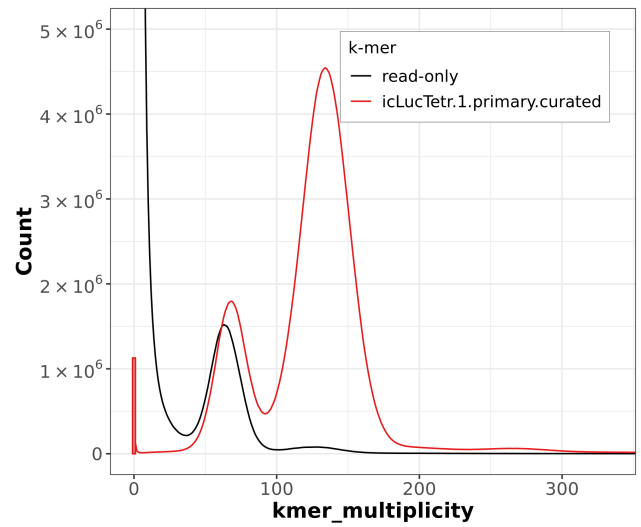
HiC contact map of curated assembly
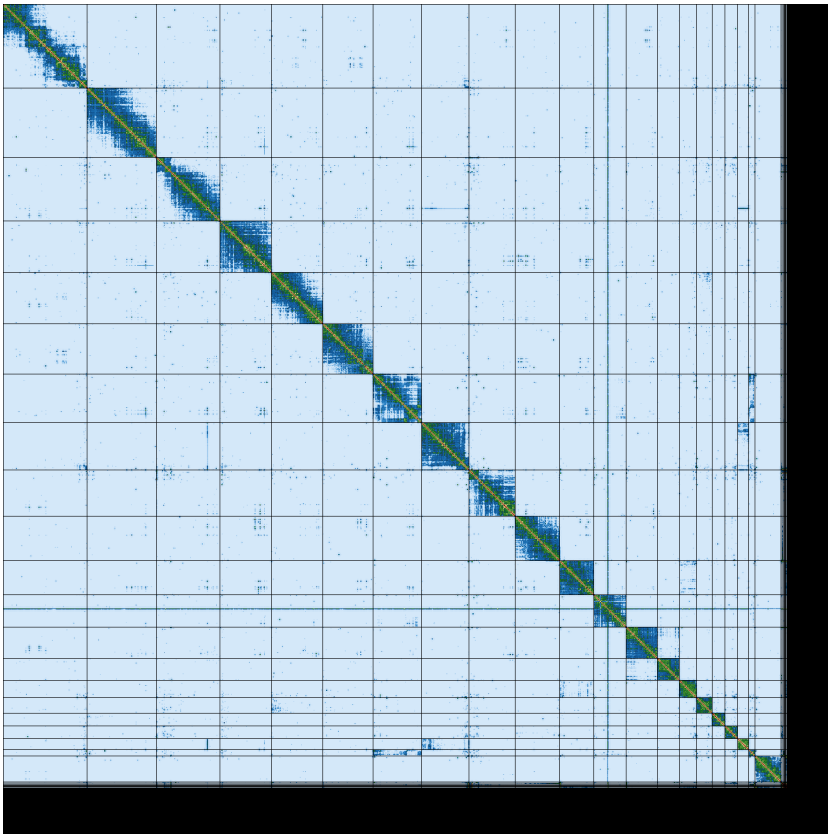
collapsed [LINK]
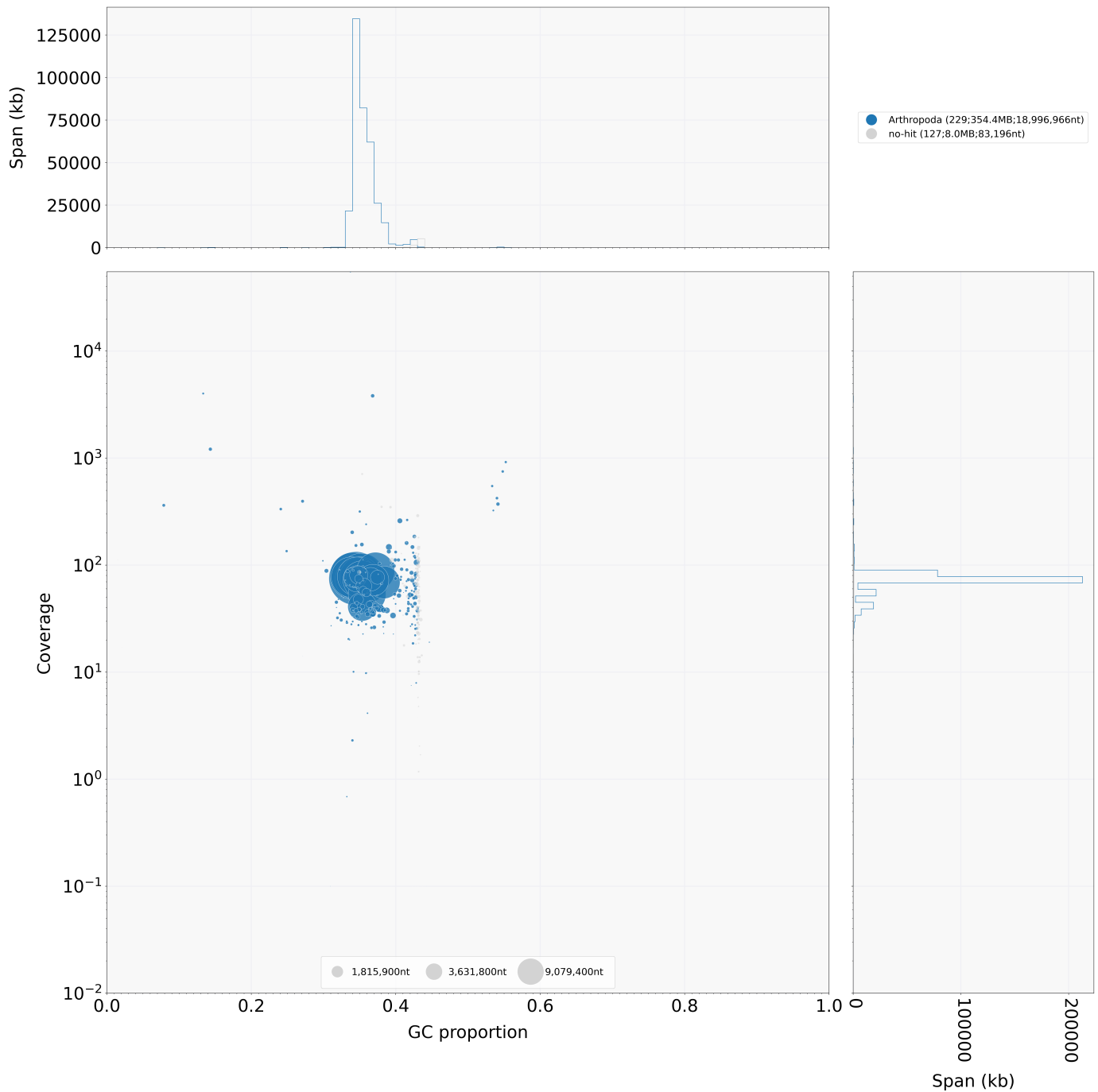
K-mer spectra of curated assembly
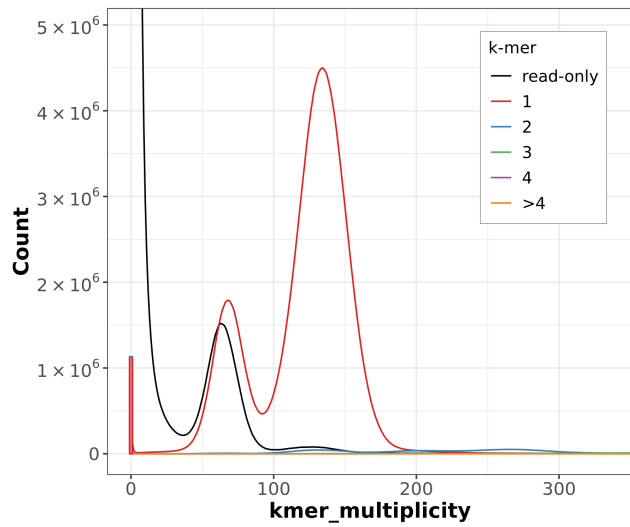
Distribution of k-mer counts per copy
numbers found in asm

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

Post-curation contamination screening

filename.blobDB.json.bestsum.phylum.p8.span.100.blobplot.bam0

Arthropoda (229;354.4MB;18,996,966nt)
no-hit (127;8.0MB;83,196nt)

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	ONT	Arima HiC
Coverage	150	85

Assembly pipeline

- **Necat**
    |_ *ver:* 0.0.1
    |_ *key param:* NA
- **purge_dups**
    |_ *ver:* 1.2.5
    |_ *key param:* NA
- **YaHS**
    |_ *ver:* 1.2.2
    |_ *key param:* NA

Curation pipeline

- **sanger-tol/curationpretext**
    |_ *ver:* 1.5.0
    |_ *key param:* NA
- **PretextView**
    |_ *ver:* 1.0.5
    |_ *key param:* NA

Submitter: Melania Spinelli, Ilenia Urso
Affiliation: UNIBA

Date and time: 2025-12-17 10:59:27 CET