

ERGA Assembly Report

v24.10.15

Tags: ERGA-Pilot

TxID	1091357
ToLID	icCryBari
Species	<i>Cryptocephalus barii</i>
Class	Insecta
Order	Coleoptera

Genome Traits	Expected	Observed
Haploid size (bp)	631,517,076	539,065,588
Haploid Number	15 (source: ancestor)	15
Ploidy	2 (source: ancestor)	2
Sample Sex	XY	XY

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.7.Q40

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

- . Kmer completeness value is less than 90 for pri

Curator notes

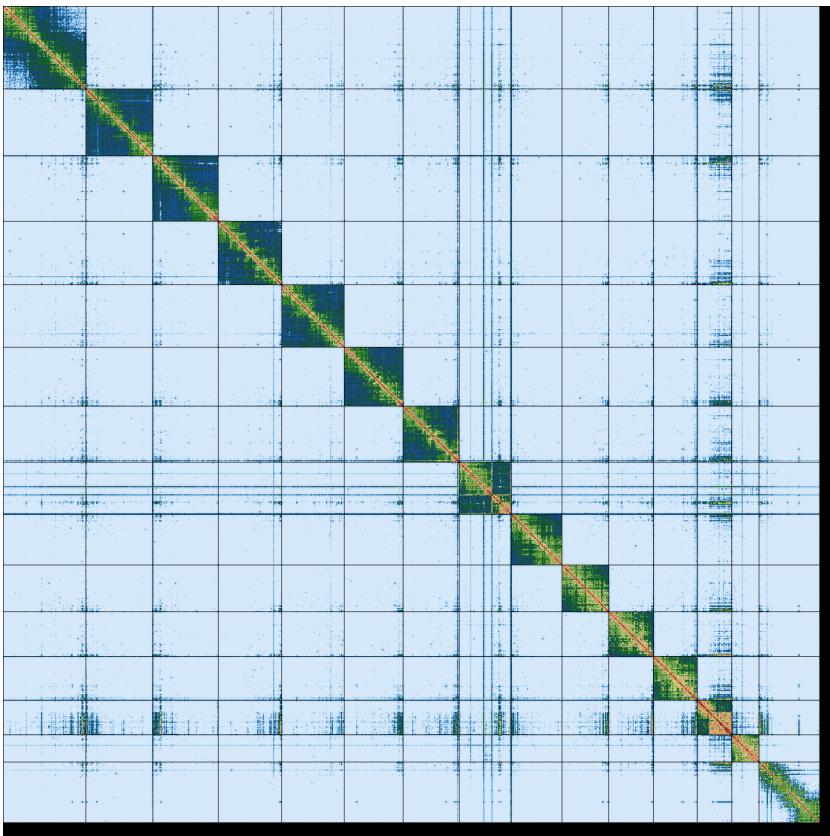
- . Interventions/Gb: 58
- . Contamination notes: "The sequence of a known endosymbiont was identified and excluded from the final assembly."
- . Other observations: "The assembly of species *Cryptocephalus barii* (icCryBari) is based on 189X long read ONT data and 170X Dovetail OmniC data. The assembly process included the following steps: ONT reads shorter than 3 kb were filtered out, thus the remaining reads -for a total of ~102 Gb- 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 Omni-C reads to the purged contigs using the Omni-C mapping pipeline, followed by YaHS for scaffold construction. Contamination was checked using BlobTools. 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 Omni-C data were named in order of size. Chromosome X was identified."

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	541,810,164	539,065,588
GC %	35.53	35.53
Gaps/Gbp	326.68	374.72
Total gap bp	35,400	40,400
Scaffolds	281	254
Scaffold N50	36,393,358	37,896,997
Scaffold L50	7	7
Scaffold L90	15	13
Contigs	458	456
Contig N50	5,575,892	5,575,892
Contig L50	28	28
Contig L90	109	108
QV	40.4262	40.4107
Kmer compl.	85.4392	85.3064
BUSCO sing.	97.1%	97.3%
BUSCO dupl.	1.2%	1.0%
BUSCO frag.	0.3%	0.3%
BUSCO miss.	1.4%	1.4%

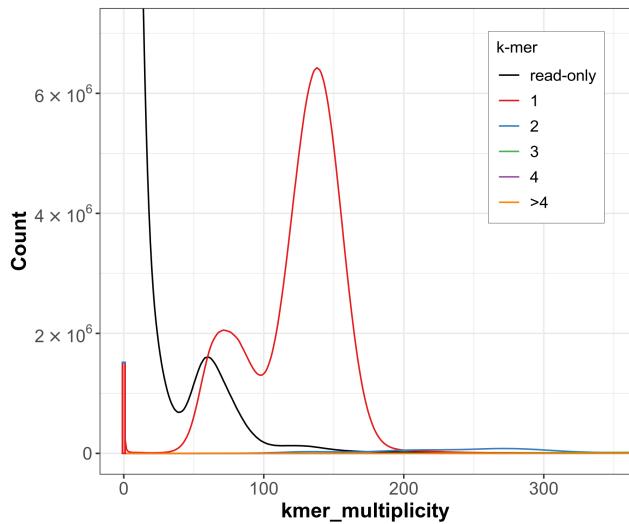
BUSCO: 5.8.2 (euk_genome_aug, augustus) / Lineage: arthropoda_odb10 (genomes:90, BUSCOs:1013)

HiC contact map of curated assembly

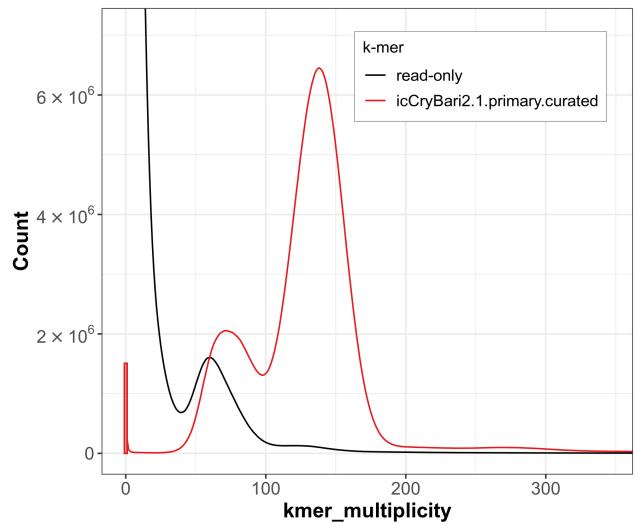


pri [LINK]

K-mer spectra of curated assembly



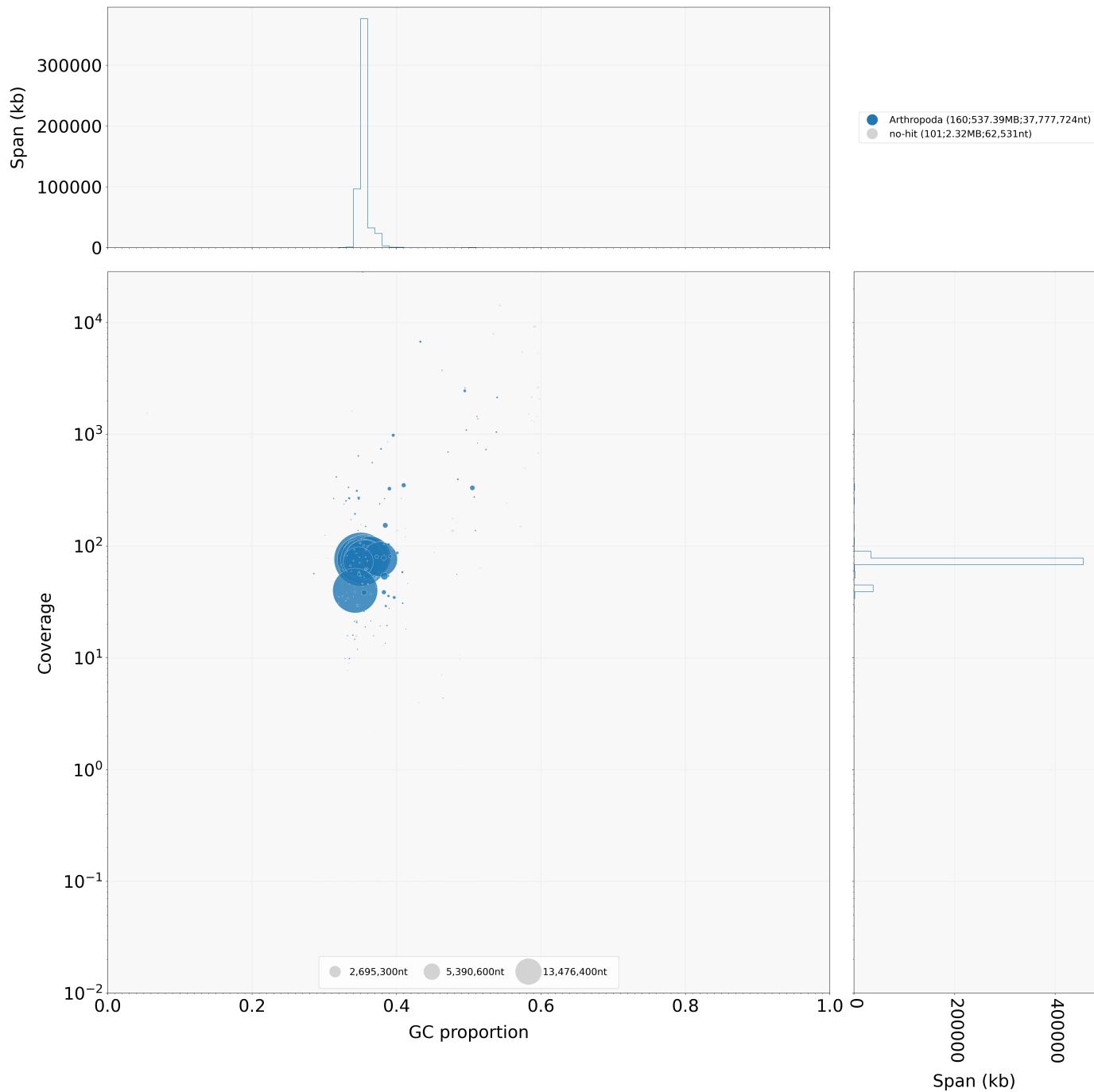
Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening

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



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	ONT	OmniC
Coverage	189x	170x

Assembly pipeline

- **Necat**
 - |_ ver: 0.0.1
 - |_ key param: NA
- **purge_dups**
 - |_ ver: 1.2.5
 - |_ key param: NA
- **YaHS**
 - |_ ver: 1.1
 - |_ key param: NA

Curation pipeline

- **sanger-tol/curationpretext**
 - |_ ver: 1.4.1
 - |_ key param: NA
- **PretextView**
 - |_ ver: 1.0.3
 - |_ key param: NA

Submitter: Ilenia Urso
Affiliation: UNIBA

Date and time: 2025-11-19 16:55:26 CET