

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	1656920
ToLID	kaSymBrak
Species	Symplegma brakenhielmi
Class	Ascidacea
Order	Stolidobranchia

Genome Traits	Expected	Observed
Haploid size (bp)	639,728,327	562,276,130
Haploid Number	16 (source: ancestor)	16
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.7.Q41

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

- . Kmer completeness value is less than 90 for pri
- . Assembly length loss > 3% for pri
- . More than 1000 gaps/Gbp for pri

Curator notes

. Interventions/Gb: 219

. Contamination notes: "315 small sequences were removed as contaminants; in addition, 2 large scaffolds (SUPER_12, SUPER_16) were classified as Arthropoda. However, they were not removed because their coverage and GC content were consistent with the rest of the genome, and they showed no Hi-C contacts with other scaffolds."

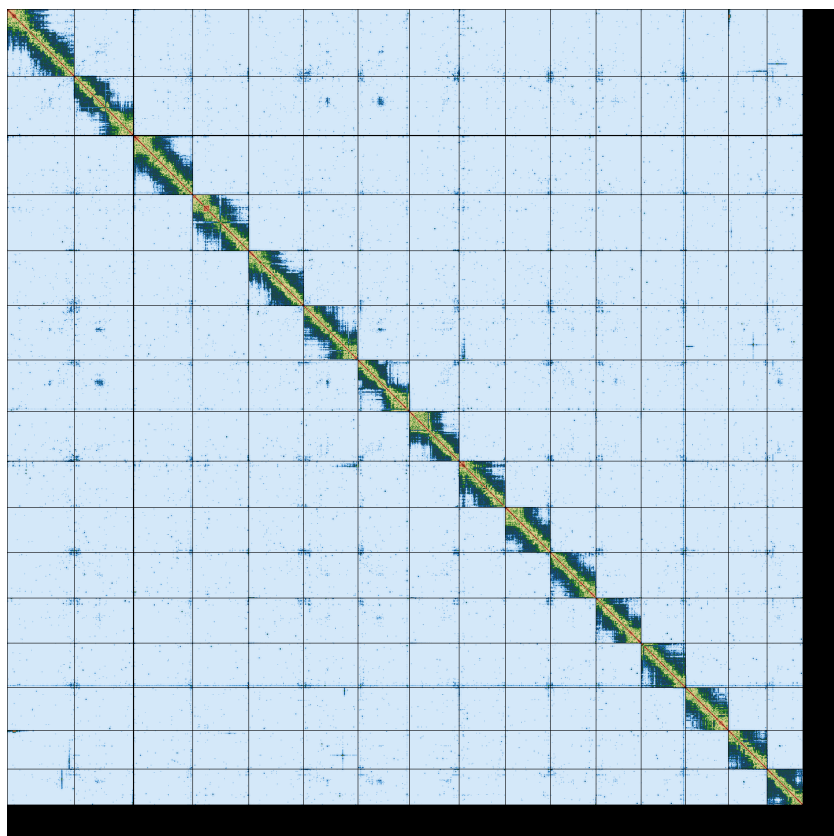
. Other observations: "The assembly of species *Symplegma brakenhielmi* (kaSymBrak) is based on 169X long read ONT data and 104X Arima HiC data generated as part of the ERGA initiative via the BGE project. The assembly process included the following steps: ONT reads shorter than 3 kb were filtered out, thus the remaining reads -for a total of ~108 Gb- were assembled using Necat with polishing performed using NextPolish. Contigs were subsequently filtered with purge_dups to remove haplotigs. 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. The scaffolds were processed through curationpretext pipeline to generate 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 pri	Curated pri
Total bp	600,884,103	562,276,130
GC %	40.66	40.22
Gaps/Gbp	1,214.88	1,412.12
Total gap bp	73,000	95,600
Scaffolds	1,101	700
Scaffold N50	30,480,455	33,584,967
Scaffold L50	9	8
Scaffold L90	28	15
Contigs	1,831	1,494
Contig N50	1,215,688	1,243,442
Contig L50	135	125
Contig L90	595	508
QV	39.7478	41.8443
Kmer compl.	84.7795	80.5569
BUSCO sing.	88.4%	90.3%
BUSCO dupl.	4.1%	1.9%
BUSCO frag.	2.2%	1.8%
BUSCO miss.	5.3%	6.1%

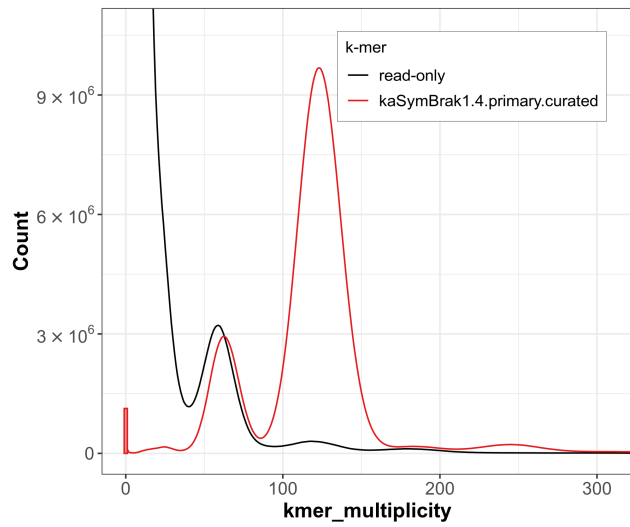
BUSCO: 5.8.2 (euk_genome_aug, augustus) / Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

HiC contact map of curated assembly

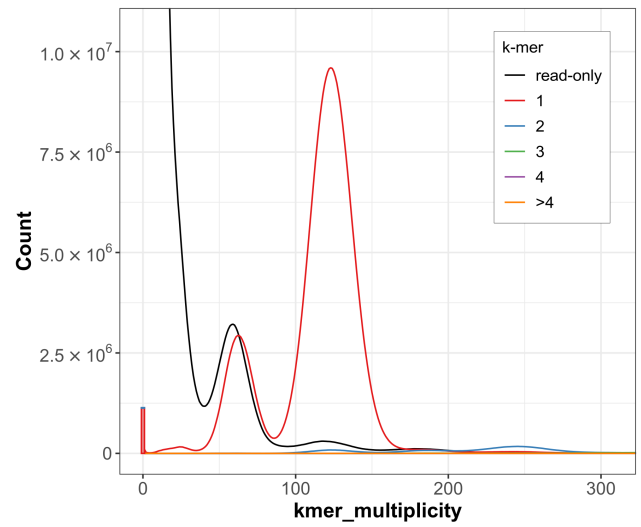


pri [\[LINK\]](#)

K-mer spectra of curated assembly



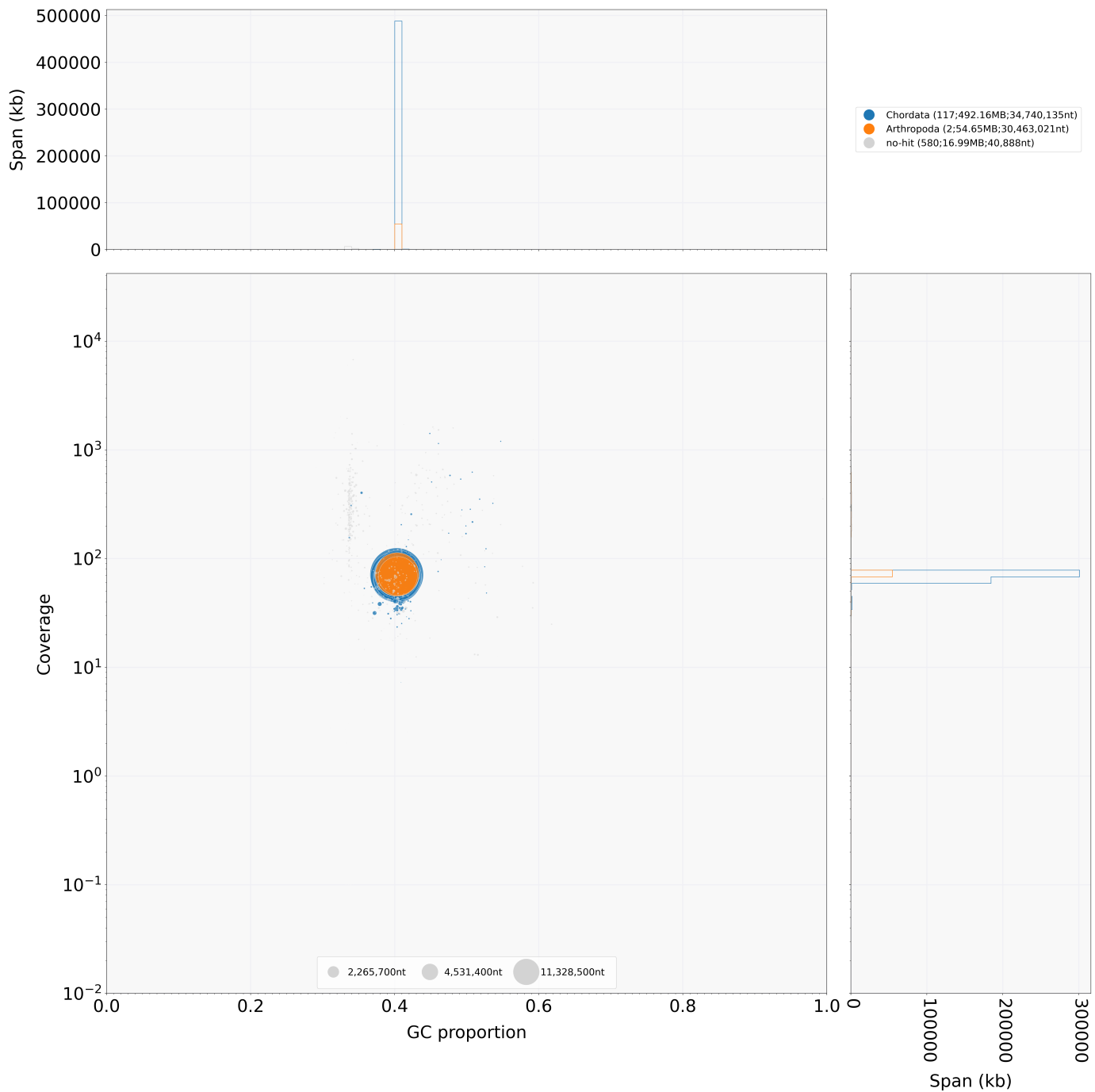
Distribution of k-mer counts coloured by their presence in reads/assemblies



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

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	Arima HiC
Coverage	169x	104x

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*: 1.5.0
 - |_ *key param*: NA
- **PretextView**
 - |_ *ver*: 1.0.3
 - |_ *key param*: NA

Submitter: Ilenia Urso

Affiliation: UNIBA

Date and time: 2025-11-25 15:11:51 CET