

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	581057
ToLID	<b>kaBotViol</b>
Species	<i>Botrylloides violaceus</i>
Class	Asciidiacea
Order	Stolidobranchia

Genome Traits	Expected	Observed
Haploid size (bp)	139,513,824	129,131,147
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: 6.6.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
- . Assembly length loss > 3% for pri

## Curator notes

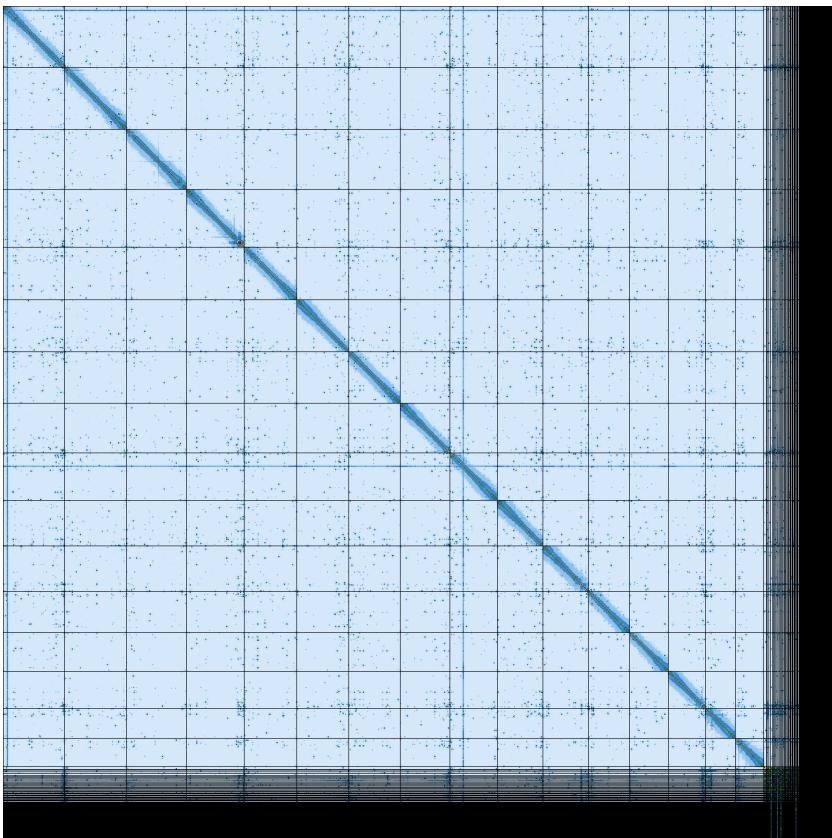
- . Interventions/Gb: 107
- . Contamination notes: "34 sequences were removed as contaminants. In addition, 2 large scaffolds (SUPER\_6, SUPER\_7) were classified as Mollusca. However, they were not removed because their coverage and GC content were consistent with the rest of the genome, and they showed no Hi-C contacts with other scaffolds."
- . Other observations: "The assembly of species *Botrylloides violaceus* is based on 1493x long read ONT data and 1085x 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 ~215 Gb were assembled using Necat with polishing performed using NextPolish. Contigs were subsequently filtered with purge\_dups. Scaffolding was performed by aligning HiC 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 HiC data were named in order of size."

## Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	141,960,073	129,131,147
GC %	41.24	41.16
Gaps/Gbp	331.08	495.62
Total gap bp	4,700	10,500
Scaffolds	258	159
Scaffold N50	7,416,636	7,692,581
Scaffold L50	9	8
Scaffold L90	57	16
Contigs	305	223
Contig N50	3,327,369	3,355,943
Contig L50	16	14
Contig L90	101	58
QV	39.8281	40.5143
Kmer compl.	80.5841	78.6535
BUSCO sing.	89.8%	91.9%
BUSCO dupl.	2.8%	0.3%
BUSCO frag.	1.7%	1.9%
BUSCO miss.	5.7%	5.9%

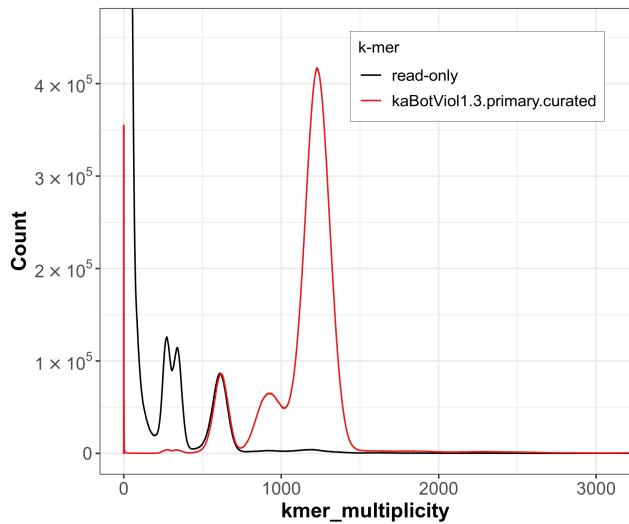
BUSCO: 5.8.2 (euk\_genome\_aug, augustus) / Lineage: metazoa\_odb10 (genomes:65, BUSCOs:954)

# HiC contact map of curated assembly

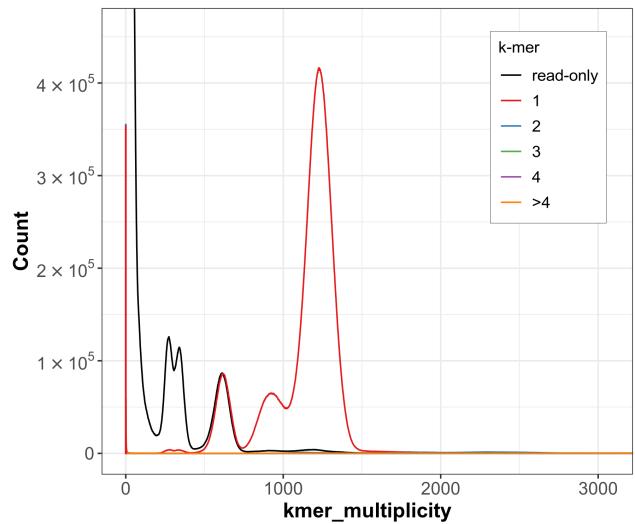


pri [LINK]

# K-mer spectra of curated assembly



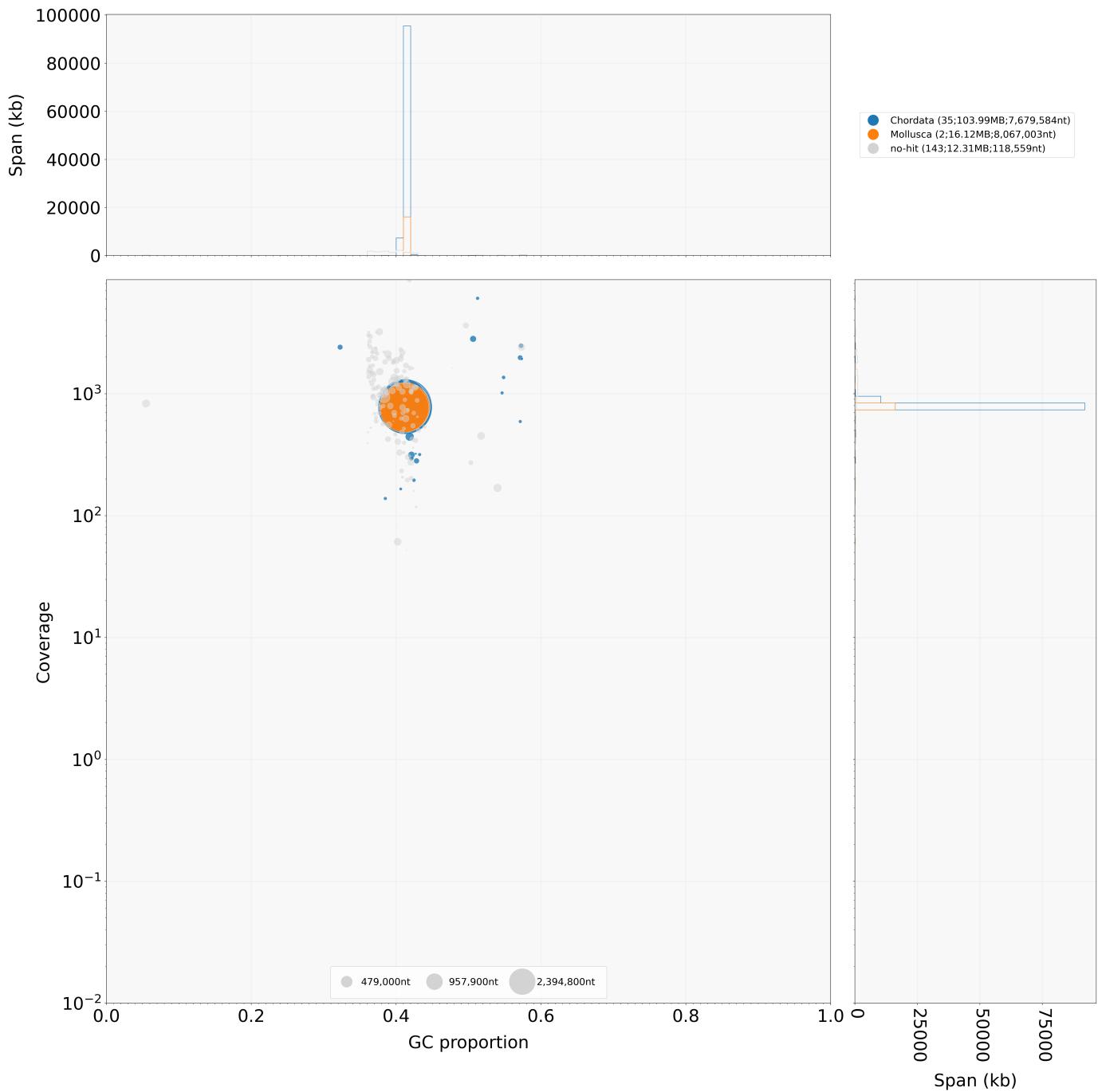
Distribution of k-mer counts coloured by their presence in reads/assemblies



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	1493x	1085x

## 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: Ilenia Urso  
Affiliation: UNIBA

Date and time: 2026-01-20 15:06:28 CET