

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	1917237
ToLID	<b>kaDisBerm</b>
Species	Distaplia bermudensis
Class	Ascidiacea
Order	Aplousobranchia

Genome Traits	Expected	Observed
Haploid size (bp)	467,817,527	446,060,772
Haploid Number	9 (source: ancestor)	11
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:

- . Observed Haploid Number is different from Expected
- . Kmer completeness value is less than 90 for pri
- . BUSCO single copy value is less than 90% for pri
- . Assembly length loss > 3% for pri

### Curator notes

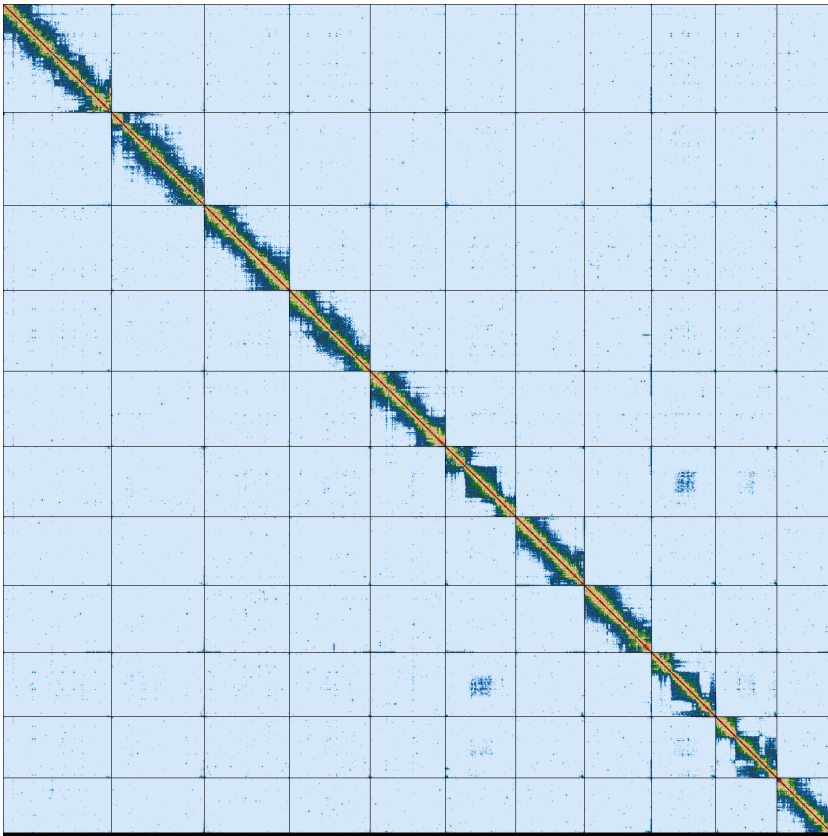
- . Interventions/Gb: 250
- . Contamination notes: "A total of 13 sequences were removed as contaminants."
- . Other observations: "The assembly of species *Distaplia bermudensis* (kaDisBerm) is based on 247X long read ONT data and 105X Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA) via the Biodiversity Genomics Europe project (BGE). The assembly process included the following steps: ONT reads shorter than 3 kb were filtered out, thus the remaining reads -for a total of ~116 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 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 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 pri	Curated pri
Total bp	473,035,248	446,060,772
GC %	34.11	33.93
Gaps/Gbp	600.38	710.67
Total gap bp	28,400	40,600
Scaffolds	159	80
Scaffold N50	38,939,299	40,291,008
Scaffold L50	5	5
Scaffold L90	11	10
Contigs	443	397
Contig N50	2,465,347	2,561,150
Contig L50	58	54
Contig L90	195	175
QV	41.2374	41.0654
Kmer compl.	85.5584	82.6955
BUSCO sing.	87.4%	89.4%
BUSCO dupl.	3.1%	0.9%
BUSCO frag.	1.7%	1.4%
BUSCO miss.	7.8%	8.3%

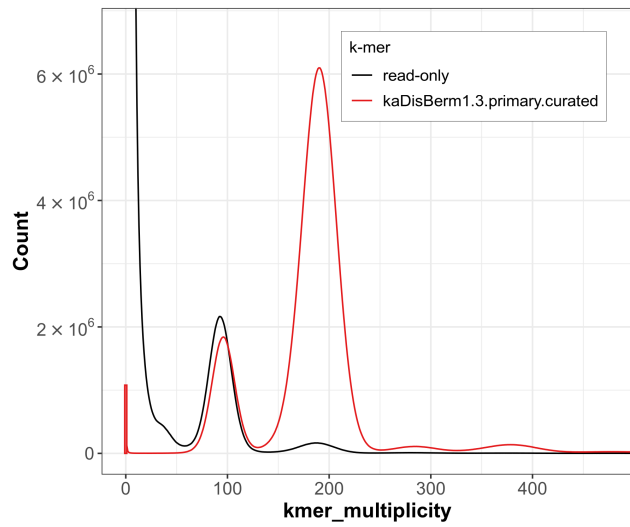
BUSCO: 5.8.2 (euk\_genome\_aug, augustus) / Lineage: metazoa\_odb10 (genomes:65, BUSCOs:954)

# HiC contact map of curated assembly

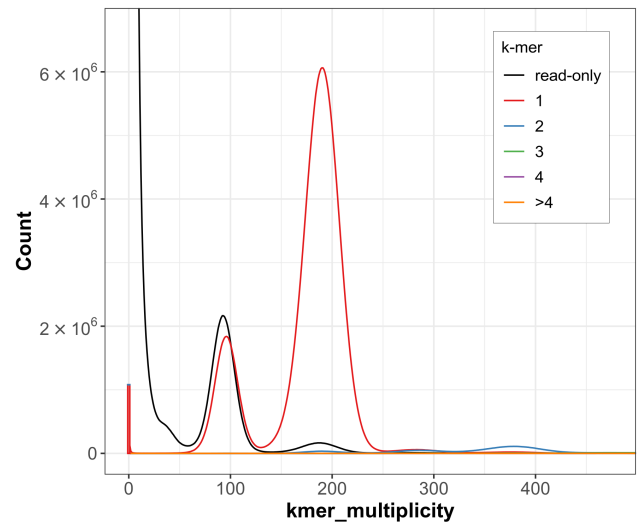


pri [\[LINK\]](#)

# K-mer spectra of curated assembly



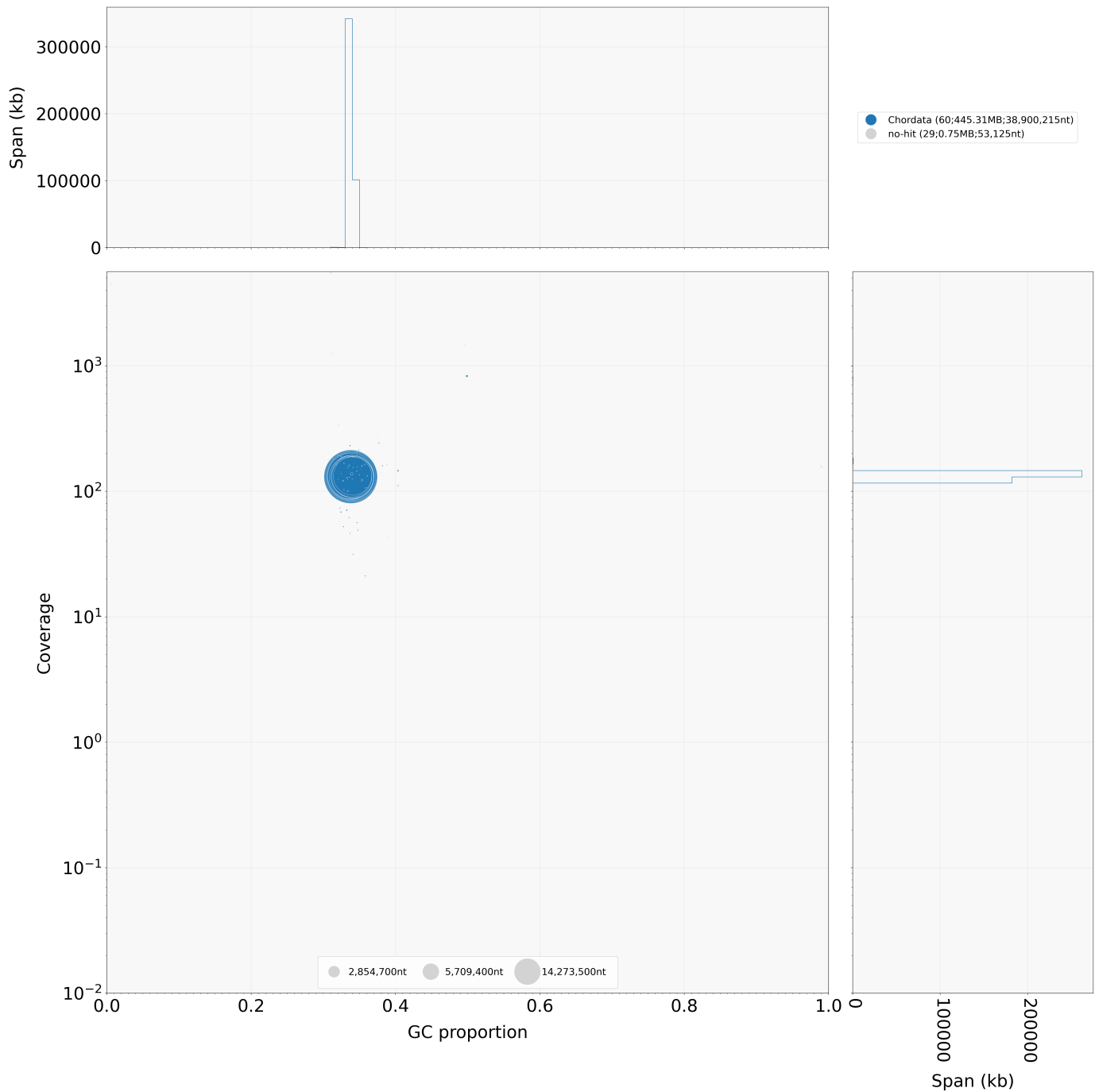
Distribution of k-mer counts coloured by their presence in reads/assemblies



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	247x	105x

## 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 16:37:20 CET