ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	581068
ToLID	**kaDisVari1**
Species	Distomus variolosus
Class	Ascidiacea
Order	Stolidobranchia

Genome Traits	Expected	Observed
Haploid size (bp)	818,517,592	839,323,762
Haploid Number	16 (source: ancestor)	22
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q47

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 collapsed

Curator notes

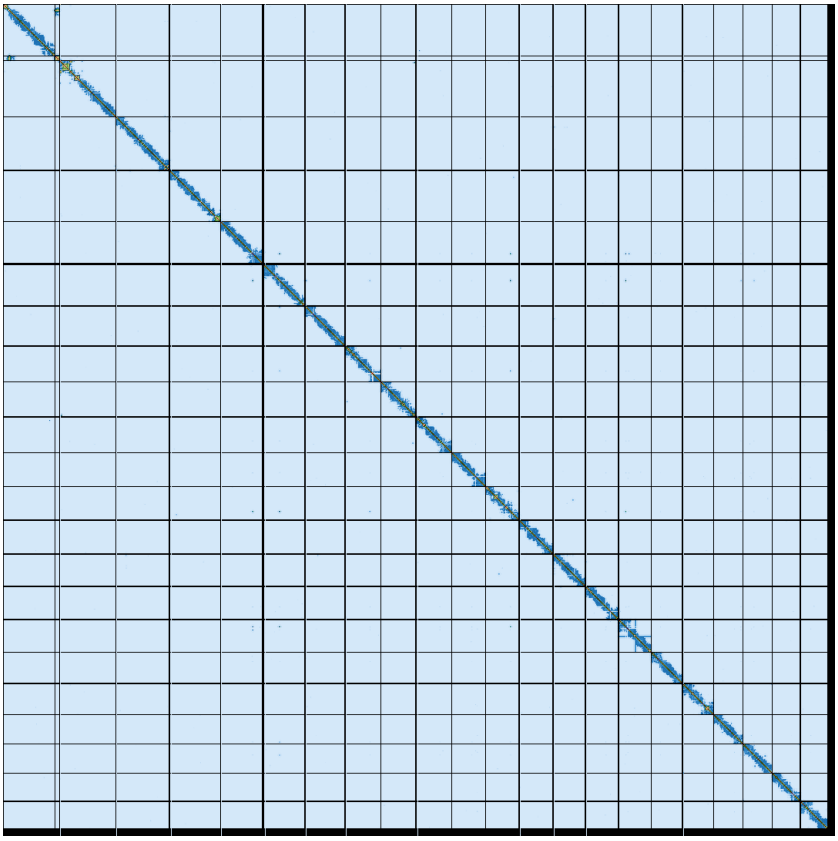
. Interventions/Gb: 124
. Contamination notes: ""
. Other observations: "The assembly of Distomus variolosus (kaDisVari1) is based on 40X PacBio data and 119X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 171 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 11.37 Mb (with the largest being 5.915 Mb). Additionally, 559 regions totaling 72.685 Mb (with the largest being 0.874 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 13 haplotypic regions were removed, totaling 3.725Mb (with the largest being 0.668Mb). 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	843,025,012	839,323,762
GC %	40.9	40.9
Gaps/Gbp	301.3	330.03
Total gap bp	25,400	33,500
Scaffolds	156	134
Scaffold N50	34,362,501	34,943,046
Scaffold L50	10	10
Scaffold L90	22	20
Contigs	410	411
Contig N50	4,660,217	4,660,217
Contig L50	59	59
Contig L90	184	184
QV	47.2303	47.2341
Kmer compl.	84.1679	83.9488
BUSCO sing.	92.7%	92.9%
BUSCO dupl.	4.2%	4.0%
BUSCO frag.	1.2%	1.3%
BUSCO miss.	1.9%	1.8%

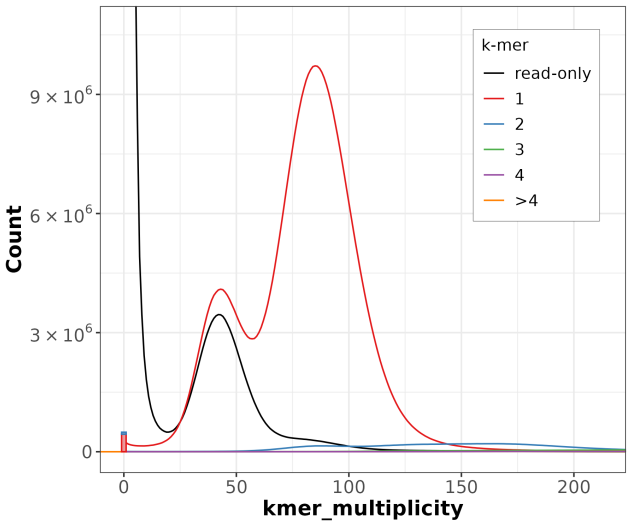
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

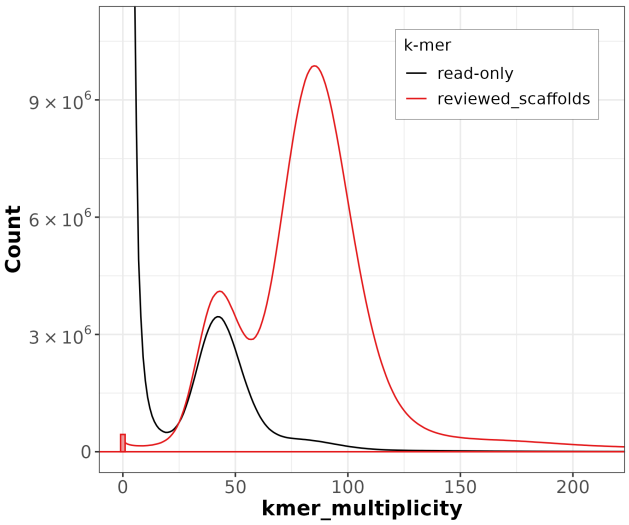


collapsed [LINK]

K-mer spectra of curated assembly

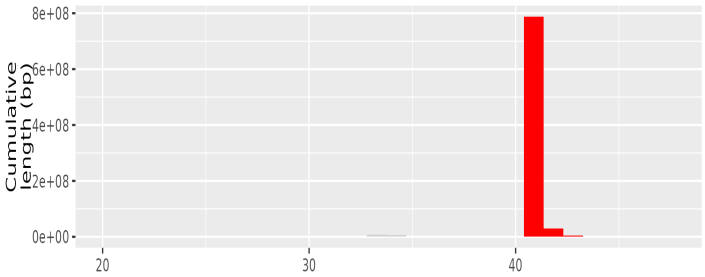


Distribution of k-mer counts per copy numbers found in asm

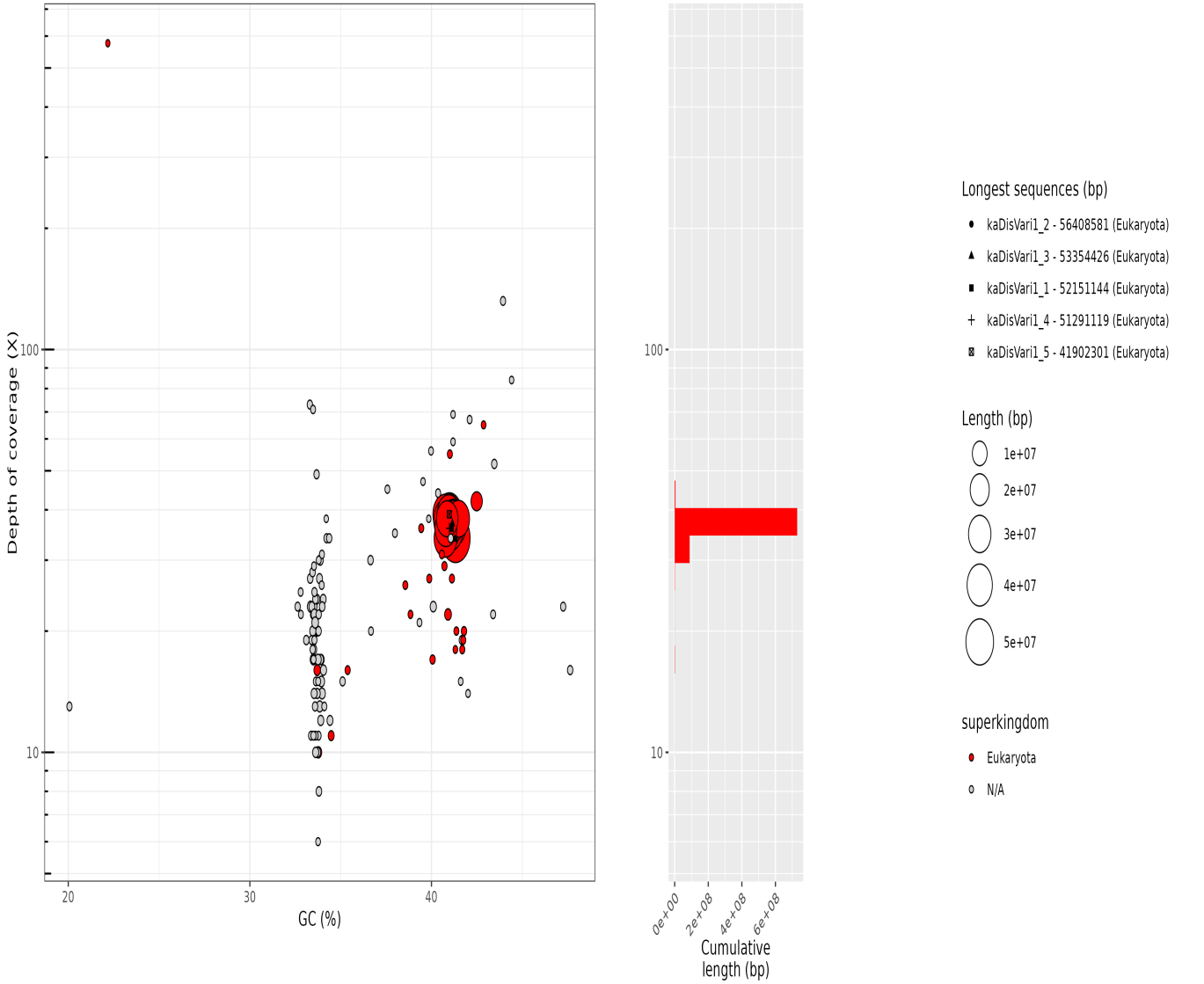


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	Long reads	Arima
Coverage	40	120

Assembly pipeline

- **Hifiasm**
  |_ *ver:* 0.19.5-r593
  |_ *key param:* NA
- **purge_dups**
  |_ *ver:* 1.2.5
  |_ *key param:* NA
- **YaHS**
  |_ *ver:* 1.2
  |_ *key param:* NA

Curation pipeline

- **PretextMap**
  |_ *ver:* 0.1.9
  |_ *key param:* NA
- **PretextView**
  |_ *ver:* 0.2.5
  |_ *key param:* NA

Submitter: Emilie Teodori
Affiliation: Genoscope

Date and time: 2026-05-03 12:56:38 CEST