ERGA Assembly Report
v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	584774
ToLID	**xbDosExol1**
Species	Dosinia exoleta
Class	Bivalvia
Order	Venerida

Genome Traits	Expected	Observed
Haploid size (bp)	1,533,956,437	1,644,291,676
Haploid Number	19 (source: ancestor)	19
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q48

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

. Kmer completeness value is less than 90 for collapsed

Curator notes

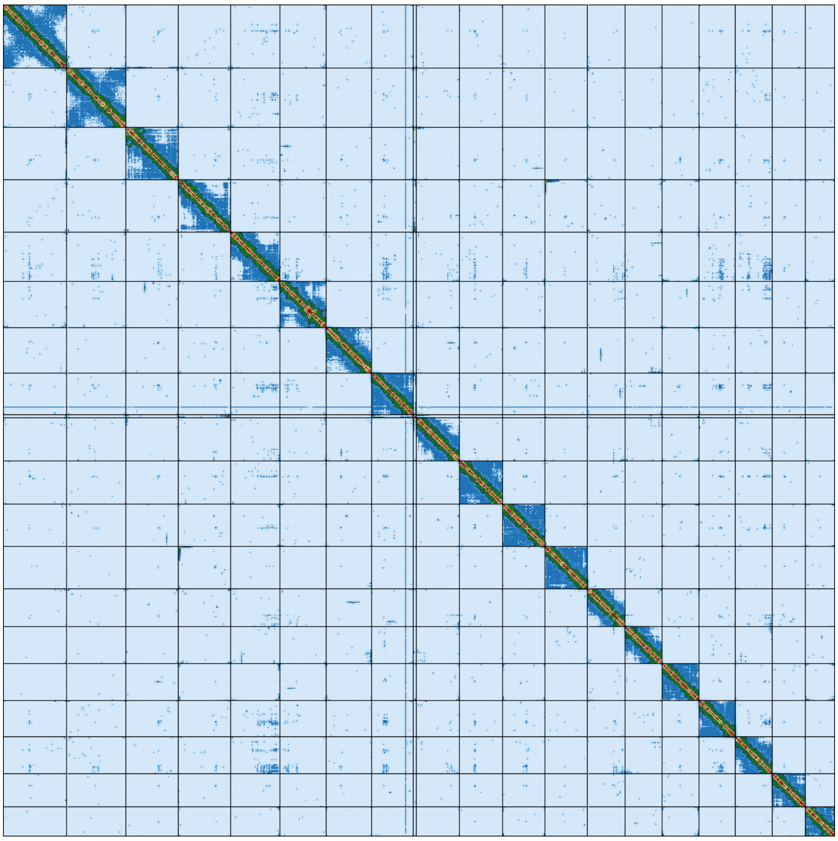
. Interventions/Gb: 63
. Contamination notes: ""
. Other observations: "The assembly of Dosinia exoleta (xbDosExol1) is based on 38X PacBio data and 149X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 3 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.269 Mb (with the largest being 0.154 Mb). Additionally, 148 regions totaling 46.43 Mb (with the largest being 9.695 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. During manual curation, 11 haplotypic regions were removed, totaling 4.636Mb (with the largest being 6.196Mb). 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	1,683,932,459	1,644,291,676
GC %	35.36	35.36
Gaps/Gbp	34.44	38.31
Total gap bp	5,800	8,000
Scaffolds	42	31
Scaffold N50	85,470,933	85,252,745
Scaffold L50	9	9
Scaffold L90	17	17
Contigs	100	94
Contig N50	51,077,196	51,077,196
Contig L50	14	14
Contig L90	39	37
QV	48.6257	48.6397
Kmer compl.	57.8541	56.9594
BUSCO sing.	96.2%	98.0%
BUSCO dupl.	3.3%	1.5%
BUSCO frag.	0.3%	0.3%
BUSCO miss.	0.2%	0.2%

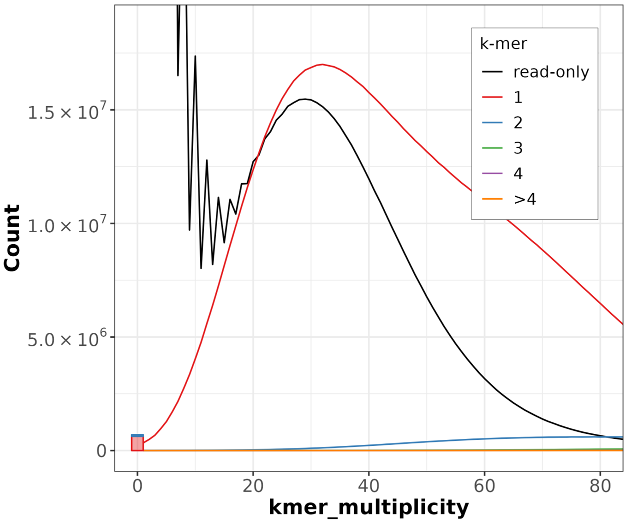
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

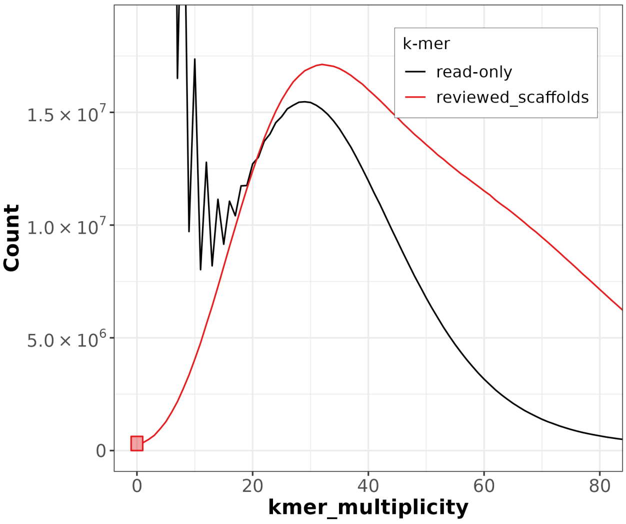


collapsed [LINK]

K-mer spectra of curated assembly

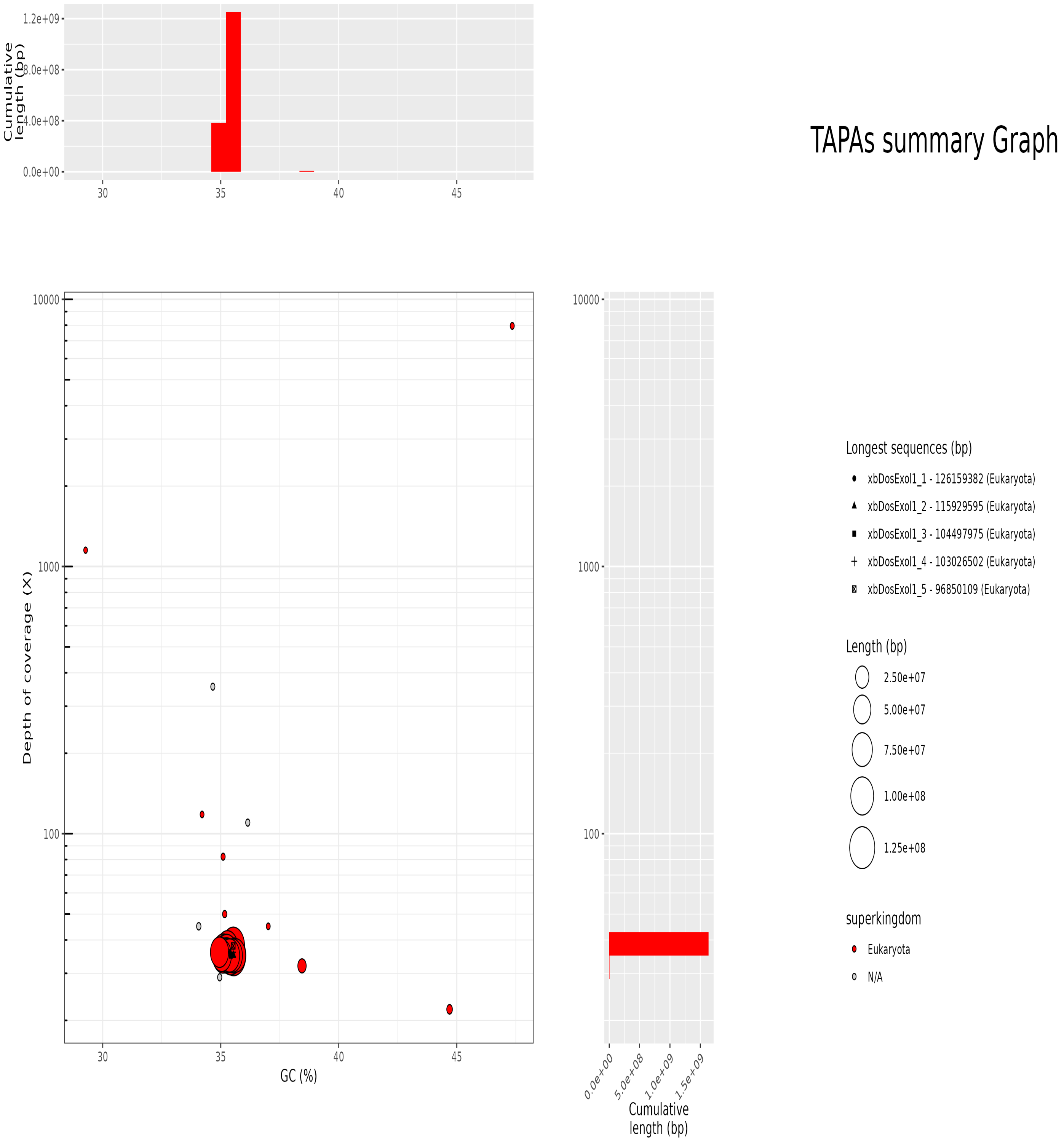


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	38	149

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: Benjamin Istace
Affiliation: Genoscope

Date and time: 2025-11-10 07:34:23 CET