ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	52918
ToLID	**xsAntVulg1**
Species	Antalis vulgaris
Class	Scaphopoda
Order	Dentaliida

Genome Traits	Expected	Observed
Haploid size (bp)	2,742,322,835	4,073,233,002
Haploid Number	9 (source: ancestor)	10
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 5.8.Q41

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

. Observed Haploid size (bp) has >20% difference with Expected
. Observed Haploid Number is different from Expected

. Kmer completeness value is less than 90 for collapsed
. BUSCO single copy value is less than 90% for collapsed
. BUSCO duplicated value is more than 5% for collapsed
. Assembly length loss > 3% for collapsed
. More than 1000 gaps/Gbp for collapsed
. Not 90% of assembly in chromosomes for collapsed

Curator notes

. Interventions/Gb: 167
. Contamination notes: ""
. Other observations: "The assembly of Antalis vulgaris (xsAntVulg1) is based on 23X PacBio data and 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, 497 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 59.9 Mb (with the largest being 2.8Mb). Additionally, 8,741 regions totaling 1 Gb (with the largest being 0.9 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using oatk. Finally, the primary assembly was analyzed and manually

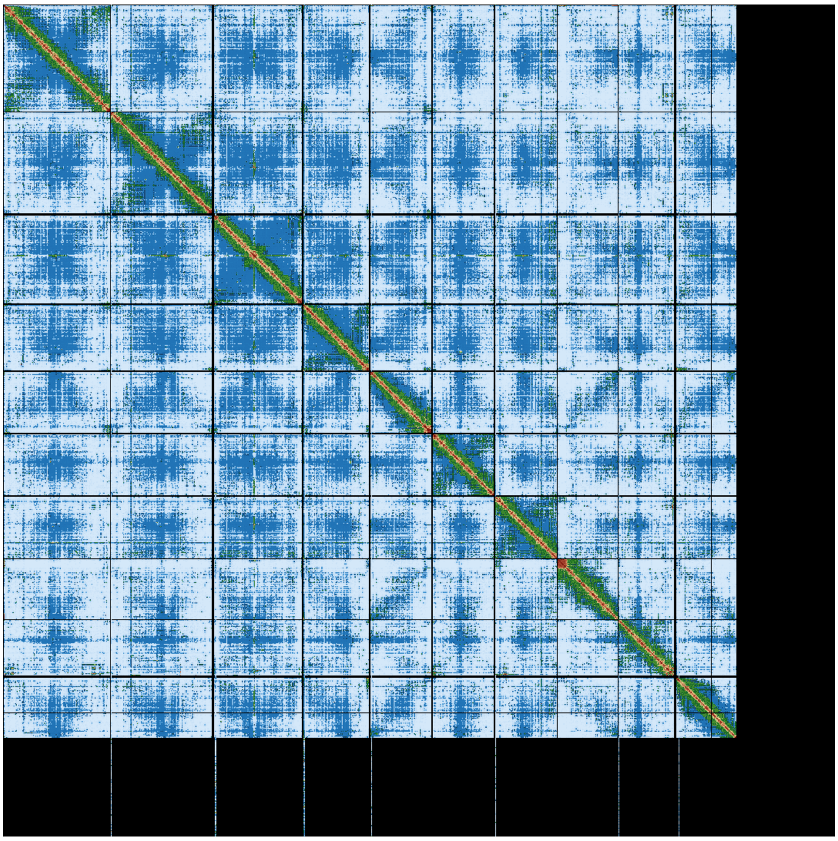
improved using Pretext. During manual curation, 287 haplotypic regions were removed, totaling 1.427 Gb (with the largest being 78 Mb). Based on Hi-C contacts, the contigs corresponding to 2 parasites (Minchinia sp. and Aponurus sp.) were removed, totaling 1.620 Gb (with the largest being 101 Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	5,910,075,484	4,073,233,002
GC %	39.21	39.59
Gaps/Gbp	1,565.29	1,816.49
Total gap bp	1,038,100	939,800
Scaffolds	7,323	4,065
Scaffold N50	248,538,169	302,300,967
Scaffold L50	7	5
Scaffold L90	665	255
Contigs	14,692	11,464
Contig N50	976,147	673,000
Contig L50	1,093	1,752
Contig L90	6,215	6,325
QV	42.9734	41.9373
Kmer compl.	76.6086	56.7338
BUSCO sing.	64.5%	78.4%
BUSCO dupl.	17.9%	12.7%
BUSCO frag.	12.7%	4.0%
BUSCO miss.	4.9%	4.9%

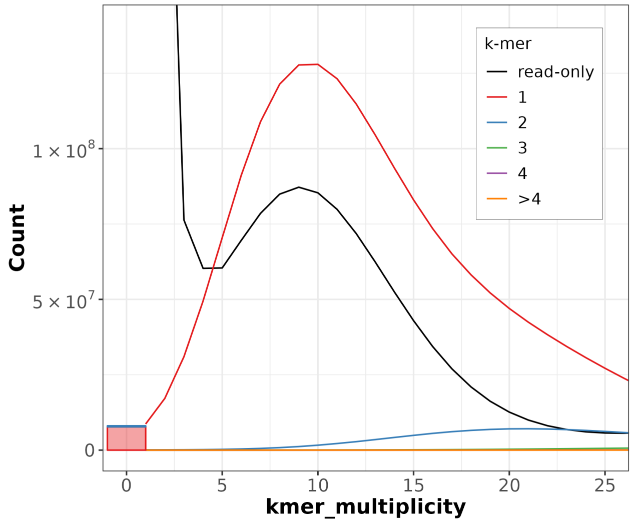
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

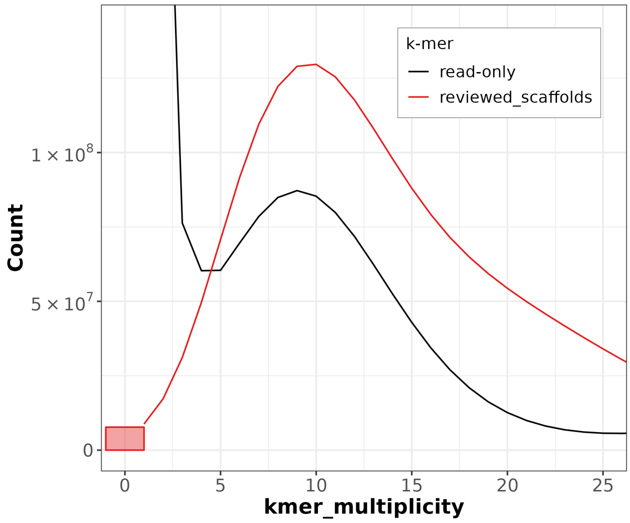


collapsed [LINK]

K-mer spectra of curated assembly

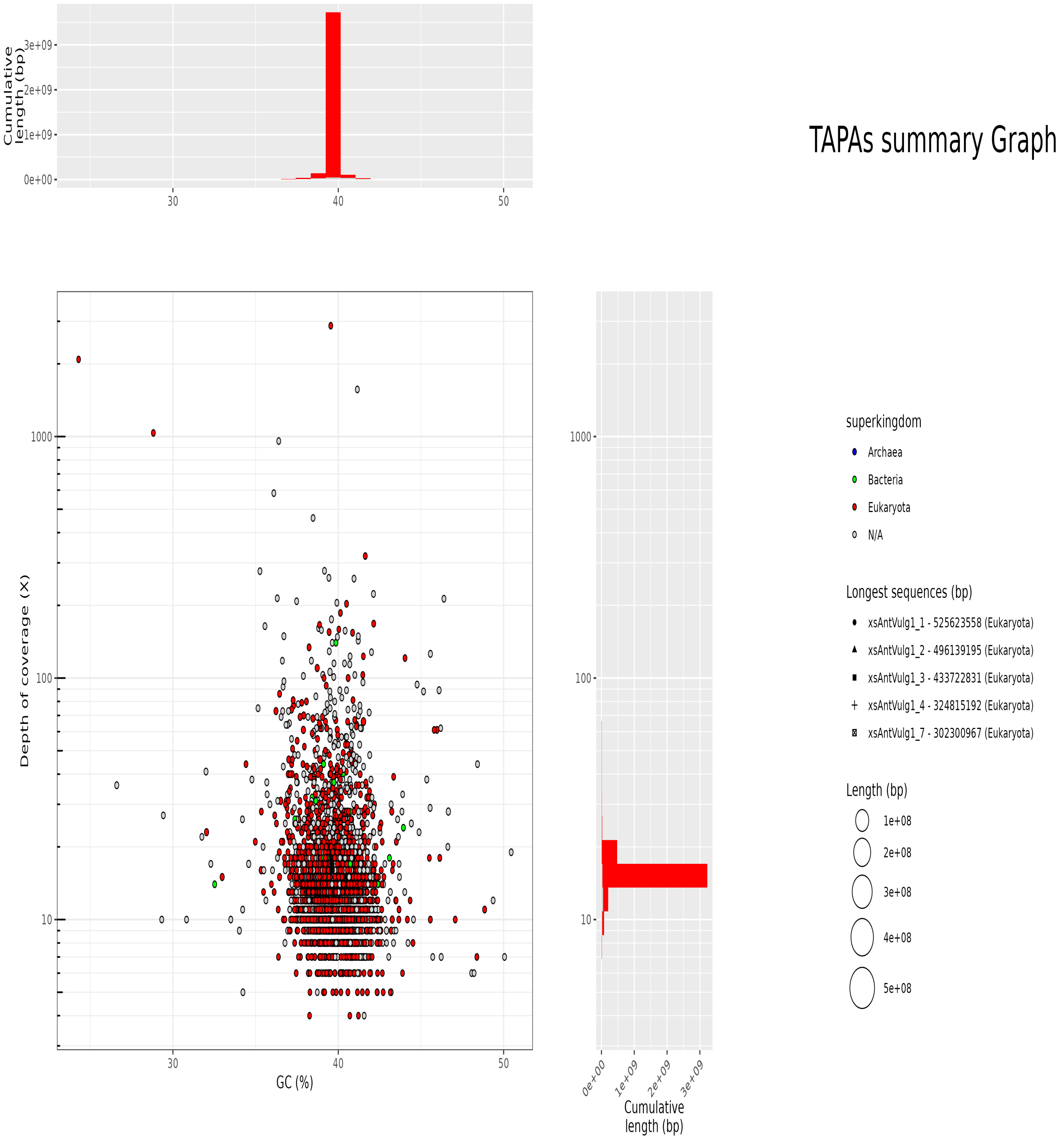


Distribution of k-mer counts per copy numbers found in asm



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	49	182

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: Caroline Menguy
Affiliation: Genoscope

Date and time: 2025-10-10 15:36:00 CEST