

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	1934731
ToLID	xgTriMuta1
Species	<i>Tritia mutabilis</i>
Class	Gastropoda
Order	Neogastropoda

Genome Traits	Expected	Observed
Haploid size (bp)	4,788,613	2,677,822,450
Haploid Number	14 (source: ancestor)	35
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q16

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
- . QV value is less than 40 for collapsed
- . BUSCO single copy value is less than 90% for collapsed
- . BUSCO duplicated value is more than 5% for collapsed

Curator notes

- . Interventions/Gb: 155
- . Contamination notes: ""
- . Other observations: "The assembly of *Tritia mutabilis* (xgTriMuta1) is based on 32X ONT data and 115X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial ONT assembly generation with Hifiiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 111 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 10.591 Mb (with the largest being 0.489 Mb). Additionally, 1675 regions totaling 164.602 Mb (with the largest being 0.728 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using ptGAUL. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 80 haplotypic regions and 4 contaminant sequences were removed, totaling 18.492Mb and 0.458Mb, respectively (with the largest being 0.778Mb and 0.14Mb). The telomeric pattern TTAGGG was identified with TelFinder and used to

generate the Pretext Telomeres tracks. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

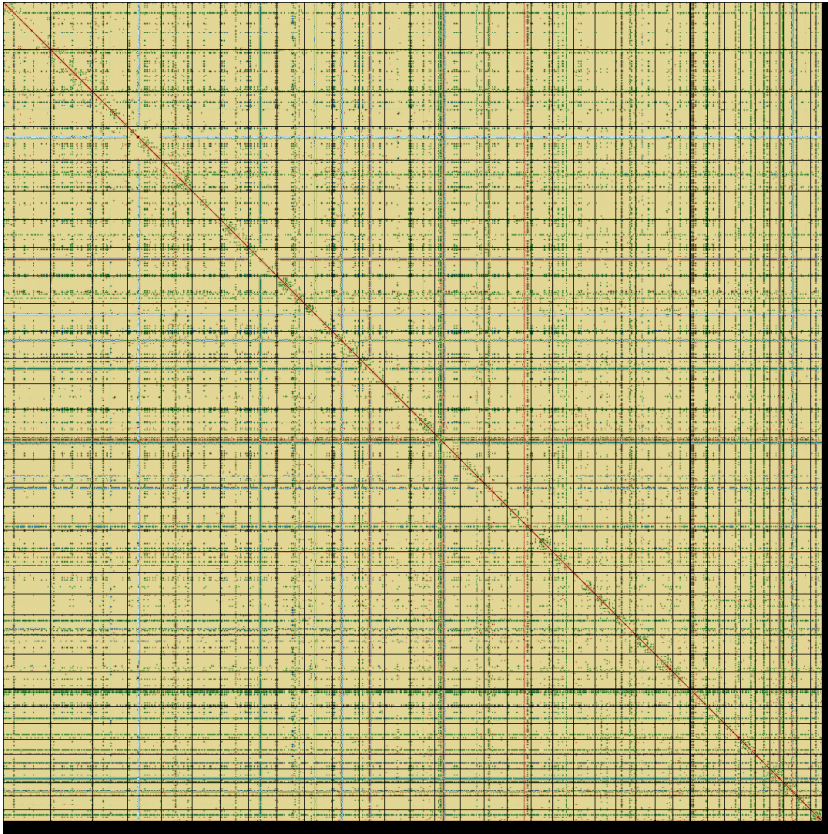
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	2,718,644,366	2,677,822,450
GC %	41.89	41.87
Gaps/Gbp	553.22	552.69
Total gap bp	150,400	166,000
Scaffolds	937	374
Scaffold N50	83,462,173	80,327,854
Scaffold L50	13	14
Scaffold L90	31	30
Contigs	2,441	1,854
Contig N50	3,087,573	3,092,269
Contig L50	269	264
Contig L90	908	878
QV	16.3278	16.3371
Kmer compl.	99.8518	99.6649
BUSCO sing.	80.4%	78.8%
BUSCO dupl.	16.6%	15.2%
BUSCO frag.	1.1%	4.0%
BUSCO miss.	1.9%	2.0%

Warning! BUSCO versions or lineage datasets are not the same across results:

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

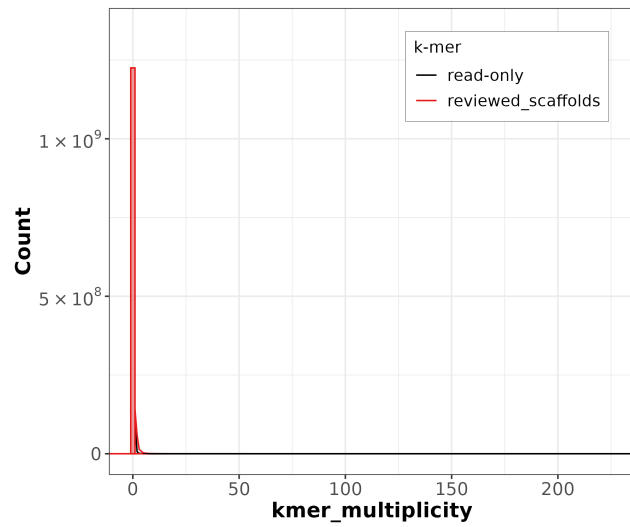
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly



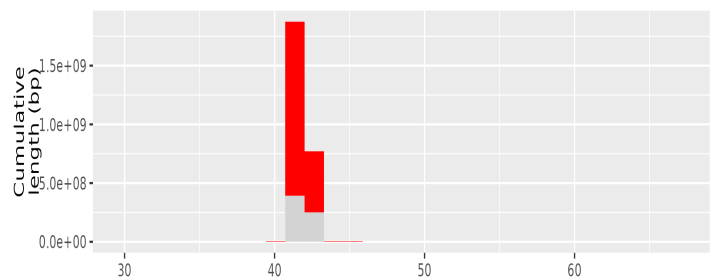
collapsed [\[LINK\]](#)

K-mer spectra of curated assembly



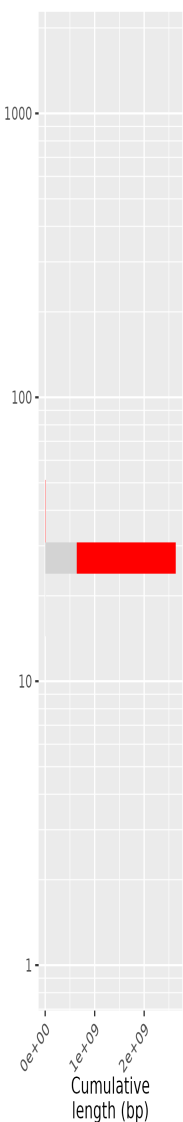
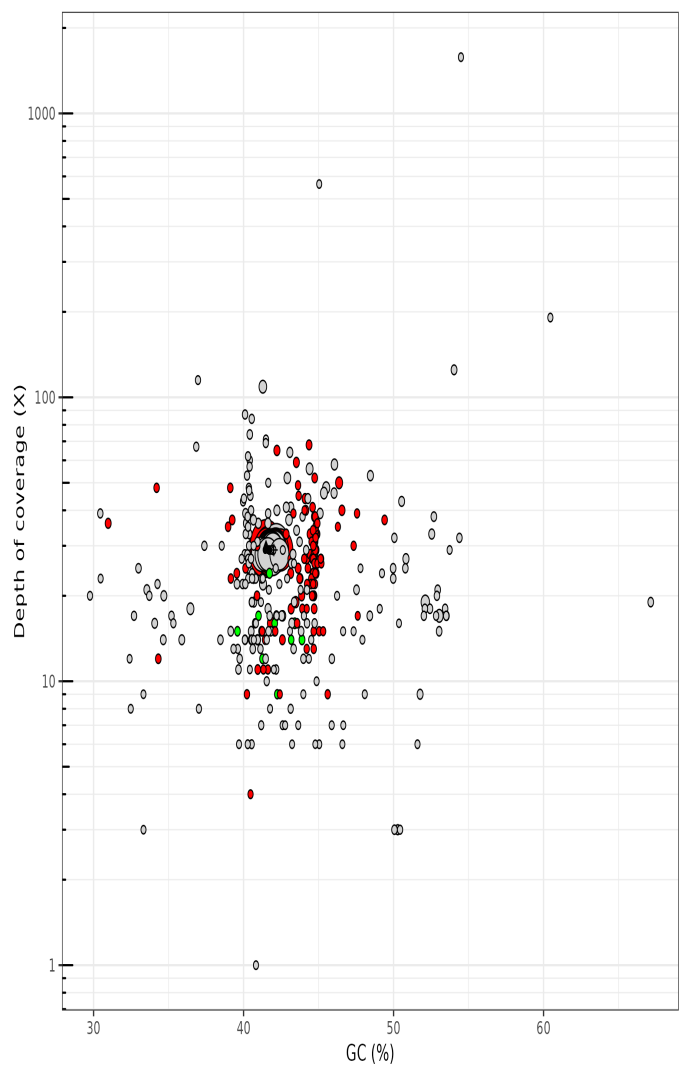
Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph

(10 0X contigs have been hidden)



superkingdom

- Bacteria
- Eukaryota
- N/A

Longest sequences (bp)

- SUPER_1 - 153206238 (Eukaryota)
- ▲ SUPER_2 - 134497015 (Eukaryota)
- SUPER_3 - 110373627 (Eukaryota)
- + SUPER_4 - 109002527 (Eukaryota)
- SUPER_5 - 99635582 (Eukaryota)

Length (bp)

- 5.0e+07
- 1.0e+08
- 1.5e+08

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	33	117

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: Sophie Layac

Affiliation: Genoscope

Date and time: 2026-07-18 16:46:57 CEST