ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	488664
ToLID	**xgNatSter1**
Species	Naticarius stercusmuscarum
Class	Gastropoda
Order	Littorinimorpha

Genome Traits	Expected	Observed
Haploid size (bp)	2,228,076,832	2,137,685,731
Haploid Number	17 (source: ancestor)	17
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q57

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

. BUSCO single copy value is less than 90% for collapsed

Curator notes

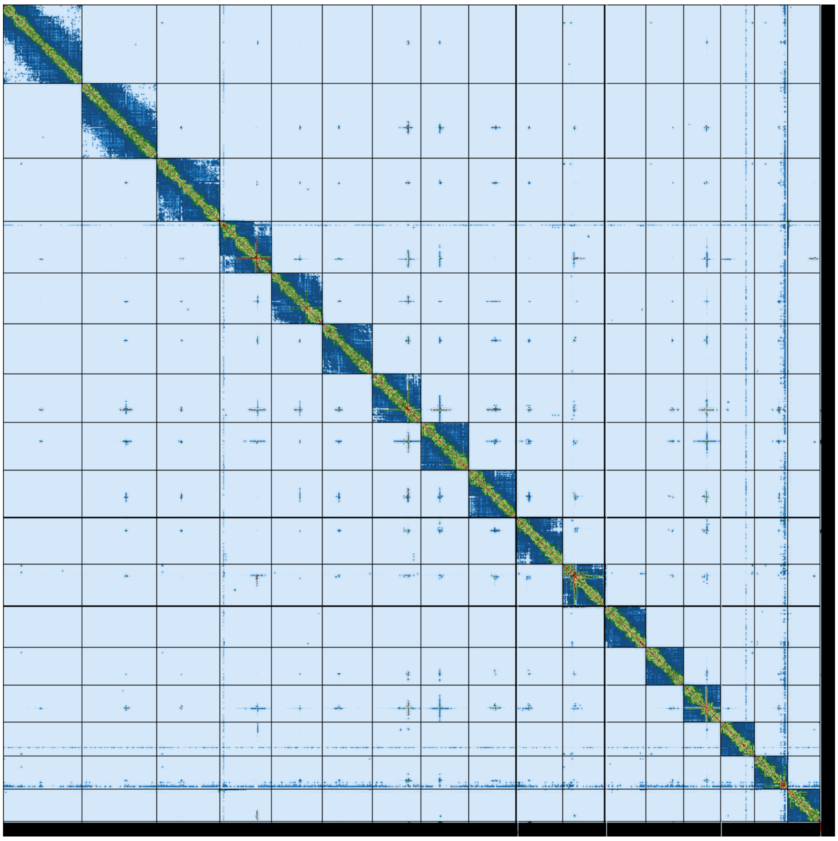
. Interventions/Gb: 36
. Contamination notes: ""
. Other observations: "The assembly of Naticarius stercusmuscarum (xgNatSter1) is based on 32X PacBio and ONT data and Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio and ONT 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, 32 contigs were identified as contaminants (bacterial), totaling 10 Mb (with the largest being 2.8 Mb). Additionally, 361 regions totaling 116 Mb (with the largest being 7.7 Mb) were identified as haplotypic duplications and removed by purge_dups. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, no regions were tagged as allelic duplications or contaminants. 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	2,137,680,670	2,137,685,731
GC %	41.62	41.62
Gaps/Gbp	283.95	299.39
Total gap bp	60,700	69,400
Scaffolds	223	177
Scaffold N50	120,737,815	125,243,207
Scaffold L50	8	7
Scaffold L90	16	15
Contigs	830	817
Contig N50	6,089,937	6,089,937
Contig L50	99	99
Contig L90	364	364
QV	43.3929	57.8324
Kmer compl.	97.8412	97.8538
BUSCO sing.	78.5%	78.6%
BUSCO dupl.	0.6%	0.6%
BUSCO frag.	5.1%	5.1%
BUSCO miss.	15.8%	15.7%

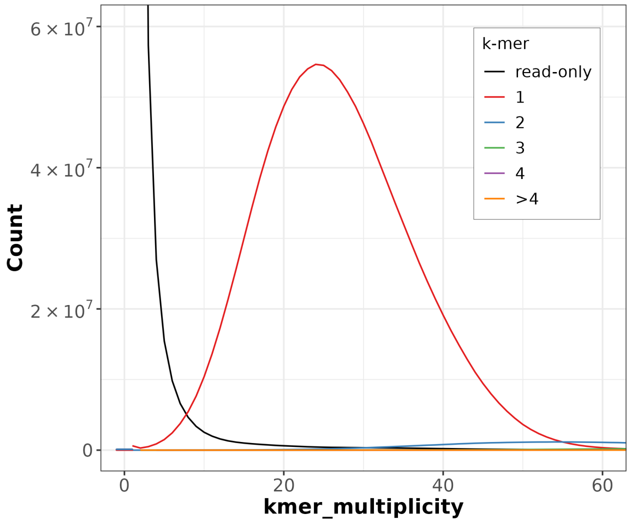
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: mollusca_odb10 (genomes:7, BUSCOs:5295)

HiC contact map of curated assembly

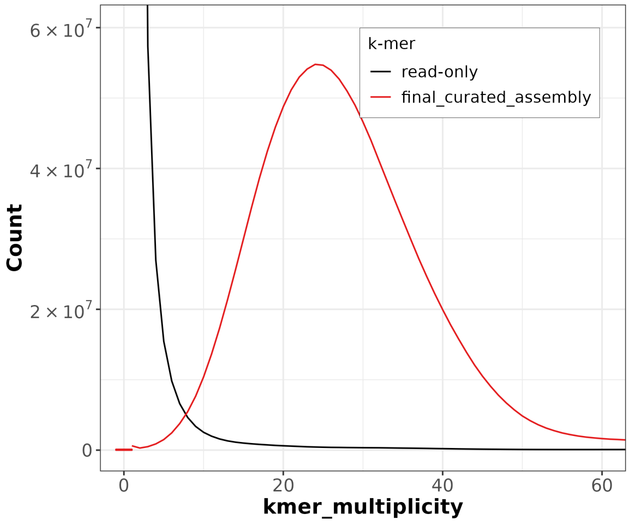


collapsed [LINK]

K-mer spectra of curated assembly

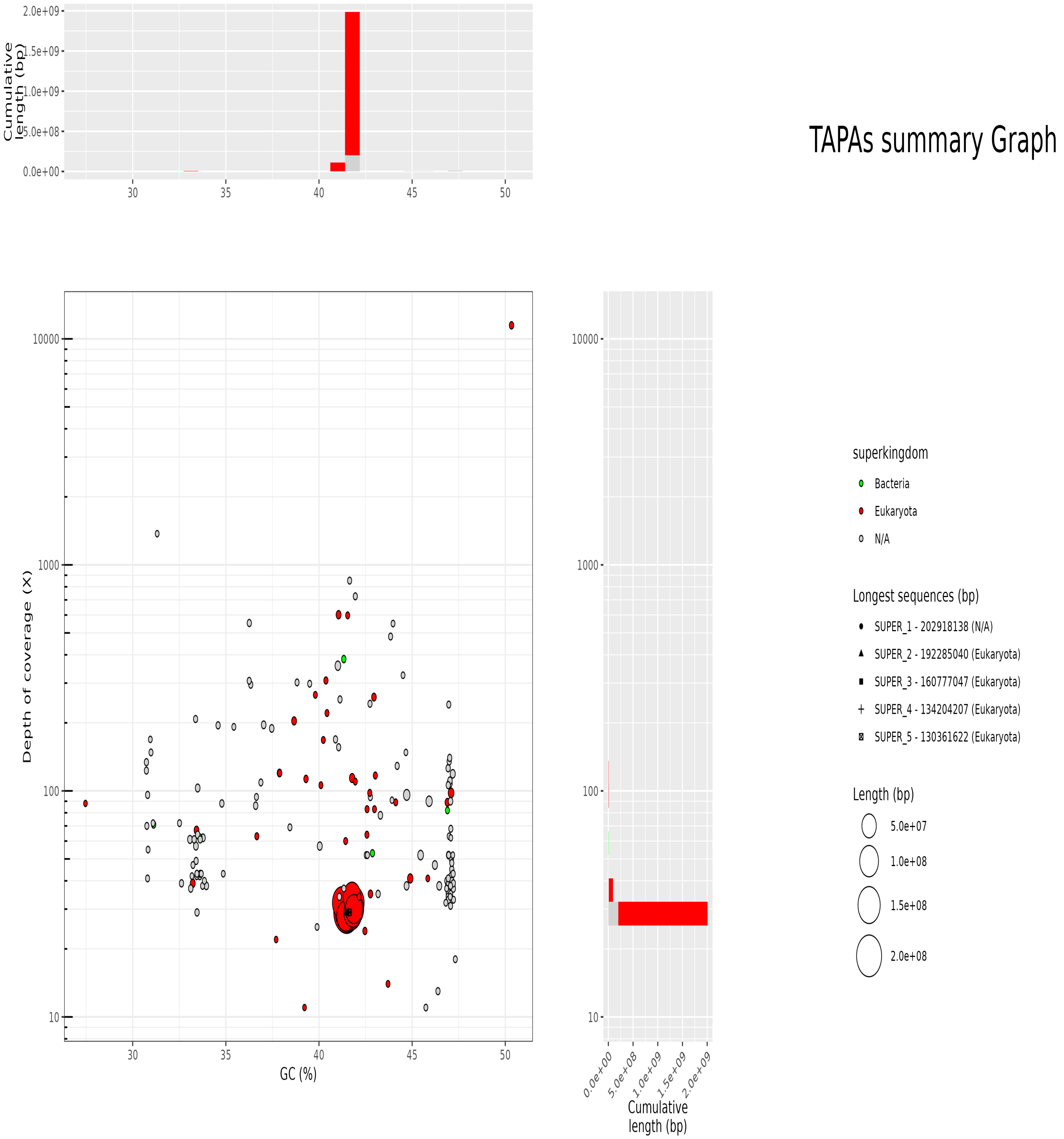


Distribution of k-mer counts per copy numbers found in asm



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	PACBIO/ONT	Arima
Coverage	32	108

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: 2025-02-07 13:11:02 CET