

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	3244224
ToLID	xgAgaFasc1
Species	Agathistoma fasciatum
Class	Gastropoda
Order	Trochida

Genome Traits	Expected	Observed
Haploid size (bp)	1,564,806,700	1,562,765,212
Haploid Number	9 (source: ancestor)	18
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q50

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: 22
. Contamination notes: ""
. Other observations: "The assembly of *Agathistoma fasciatum* (xgAgaFasc1) is based on 44X PacBio data and 152X 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, 36 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 2.107 Mb (with the largest being 0.35 Mb). Additionally, 409 regions totaling 56.106 Mb (with the largest being 4.067 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 8 haplotypic regions were removed, totaling 9.799Mb (with the largest being 3.62Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

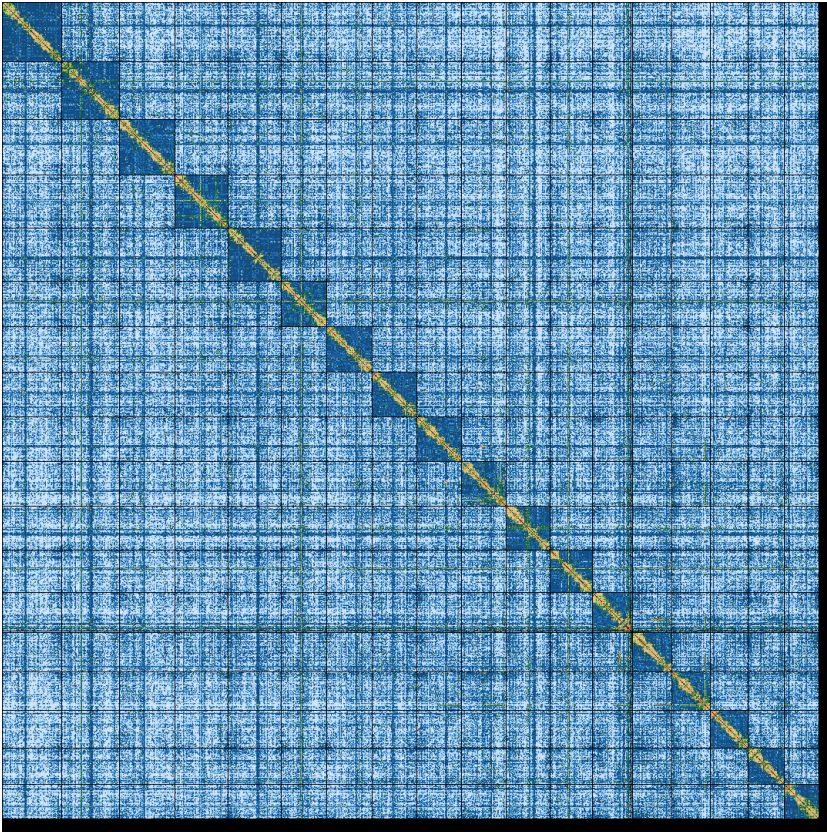
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,572,598,816	1,562,765,212
GC %	37.97	37.97
Gaps/Gbp	94.11	97.9
Total gap bp	14,800	17,000
Scaffolds	250	243
Scaffold N50	84,059,597	84,059,597
Scaffold L50	9	9
Scaffold L90	17	17
Contigs	398	396
Contig N50	17,353,252	17,640,249
Contig L50	30	29
Contig L90	96	95
QV	50.806	50.8016
Kmer compl.	69.3586	69.1247
BUSCO sing.	97.8%	95.2%
BUSCO dupl.	1.6%	1.0%
BUSCO frag.	0.2%	3.1%
BUSCO miss.	0.5%	0.7%

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)

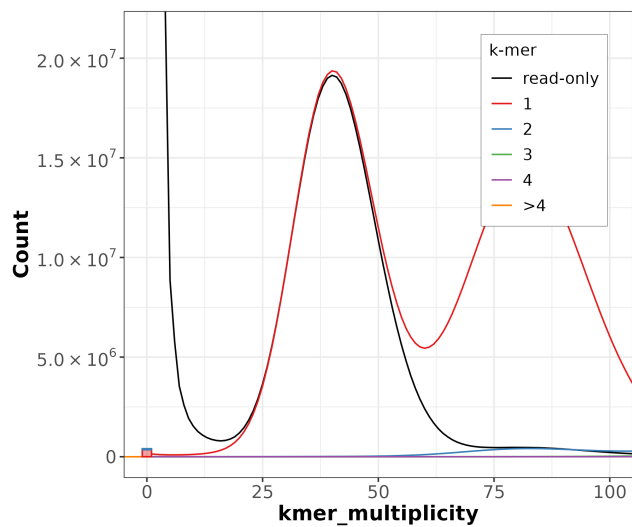
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

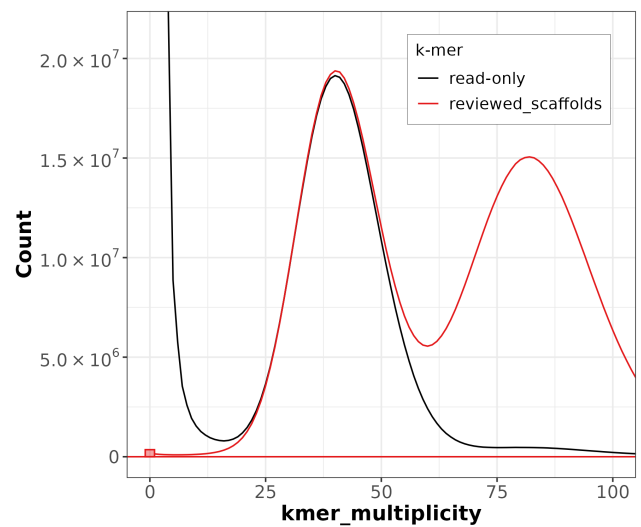


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

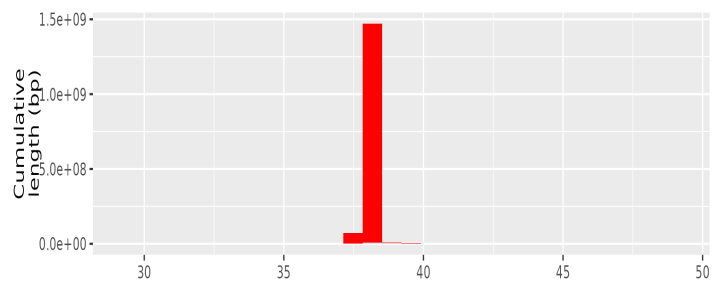


Distribution of k-mer counts per copy numbers found in asm

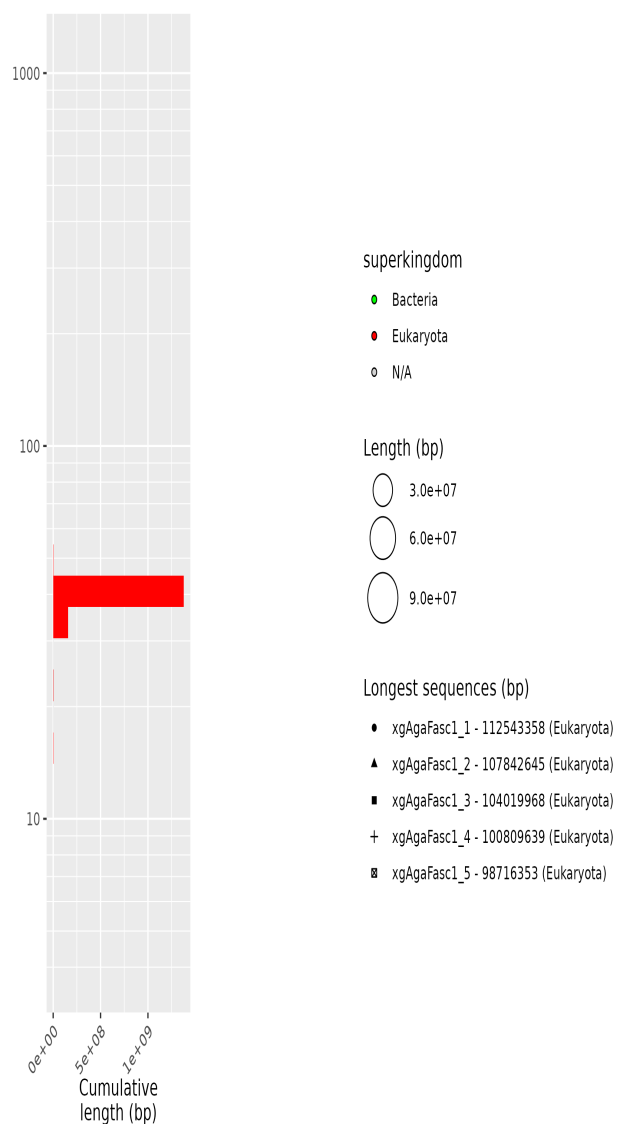
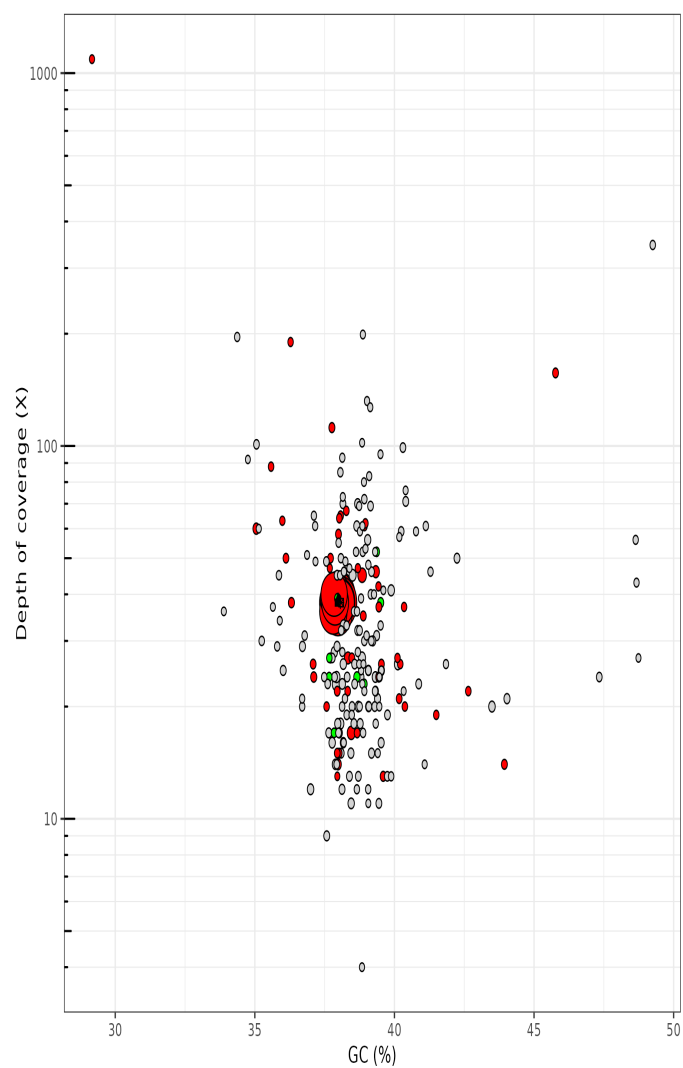


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	44	152

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: Mohamed Meftah Saoues

Affiliation: Genoscope

Date and time: 2026-07-14 06:19:30 CEST