ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	3475048
ToLID	**xgGouSpea1**
Species	Gourmya sp. A-9AQBERFAU
Class	Gastropoda
Order	NA

Genome Traits	Expected	Observed
Haploid size (bp)	788,313,183	808,719,870
Haploid Number	15 (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.Q44

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

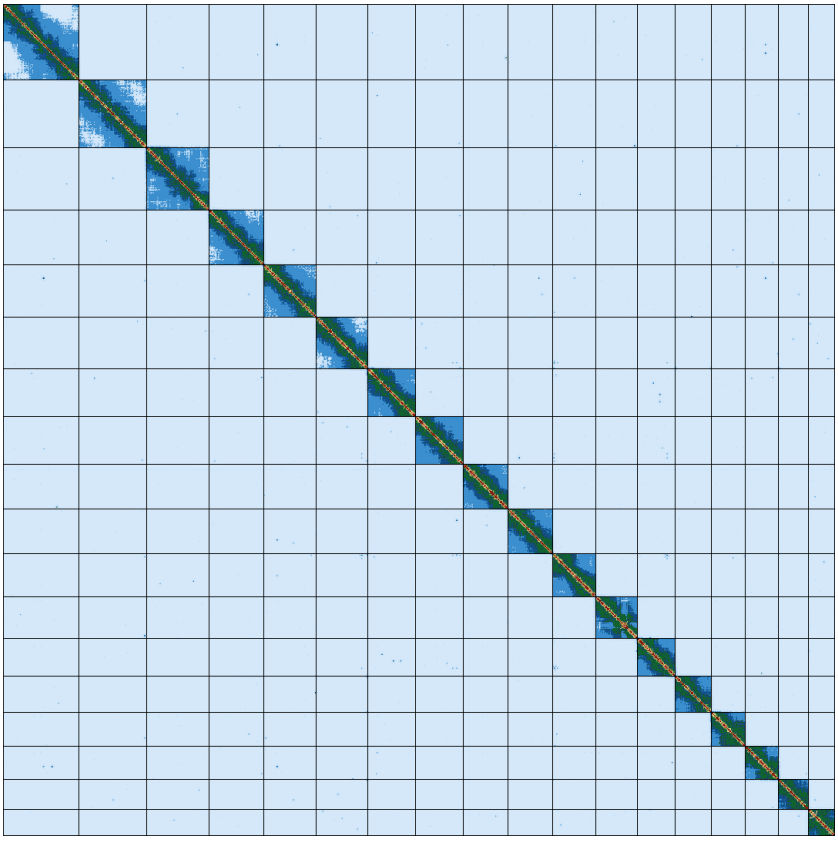
. Interventions/Gb: 29
. Contamination notes: ""
. Other observations: "The assembly of Gourmya sp. A-9AQBERFAU (xgGouSpea1) is based on 94X ONT data and 335X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial ONT assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 10 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 2.424 Mb (with the largest being 2.144 Mb). Additionally, 40 regions totaling 10.994 Mb (with the largest being 3.004 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. 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	818,212,699	808,719,870
GC %	46.47	46.47
Gaps/Gbp	19.55	23.49
Total gap bp	1,600	2,700
Scaffolds	27	21
Scaffold N50	44,029,030	45,966,108
Scaffold L50	8	8
Scaffold L90	16	16
Contigs	43	40
Contig N50	34,926,373	36,834,883
Contig L50	10	9
Contig L90	23	22
QV	44.8864	44.8861
Kmer compl.	87.7911	87.6637
BUSCO sing.	96.4%	97.4%
BUSCO dupl.	1.8%	0.8%
BUSCO frag.	1.1%	1.0%
BUSCO miss.	0.7%	0.8%

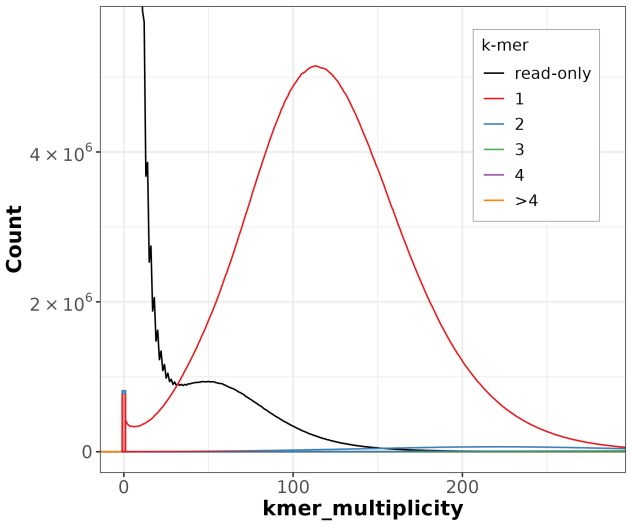
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

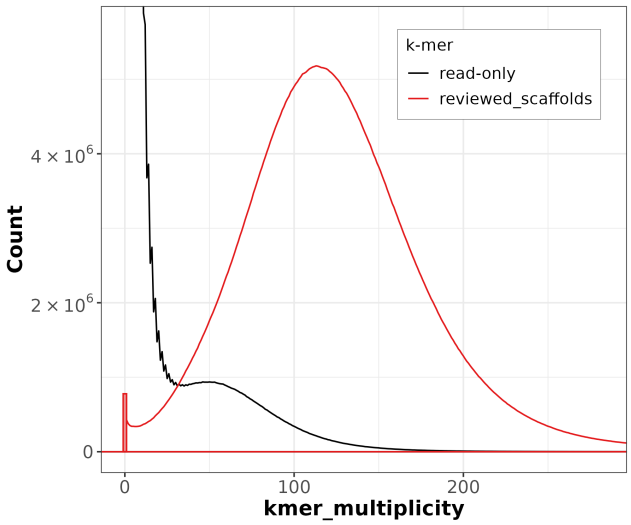


collapsed [LINK]

K-mer spectra of curated assembly

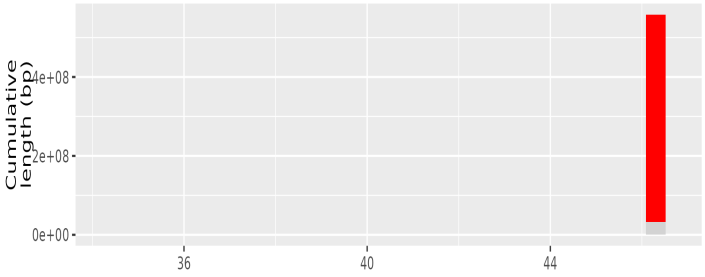


Distribution of k-mer counts per copy numbers found in asm

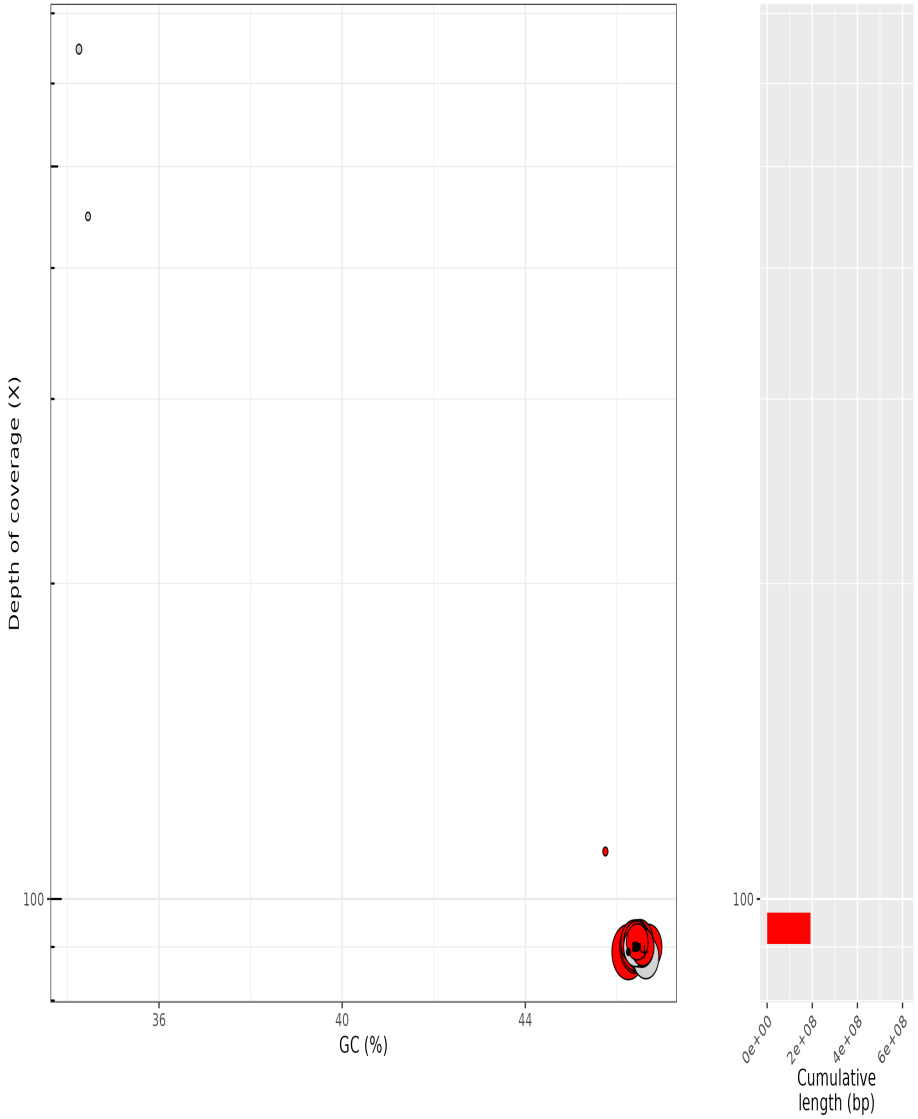


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



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	95	339

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: 2026-04-19 03:21:27 CEST