ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	460703
ToLID	**xgGloHiku1**
Species	Glossodoris hikuerensis
Class	Gastropoda
Order	Nudibranchia

Genome Traits	Expected	Observed
Haploid size (bp)	2,377,389,626	2,506,359,050
Haploid Number	12 (source: ancestor)	12
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q41

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

. Kmer completeness value is less than 90 for collapsed

Curator notes

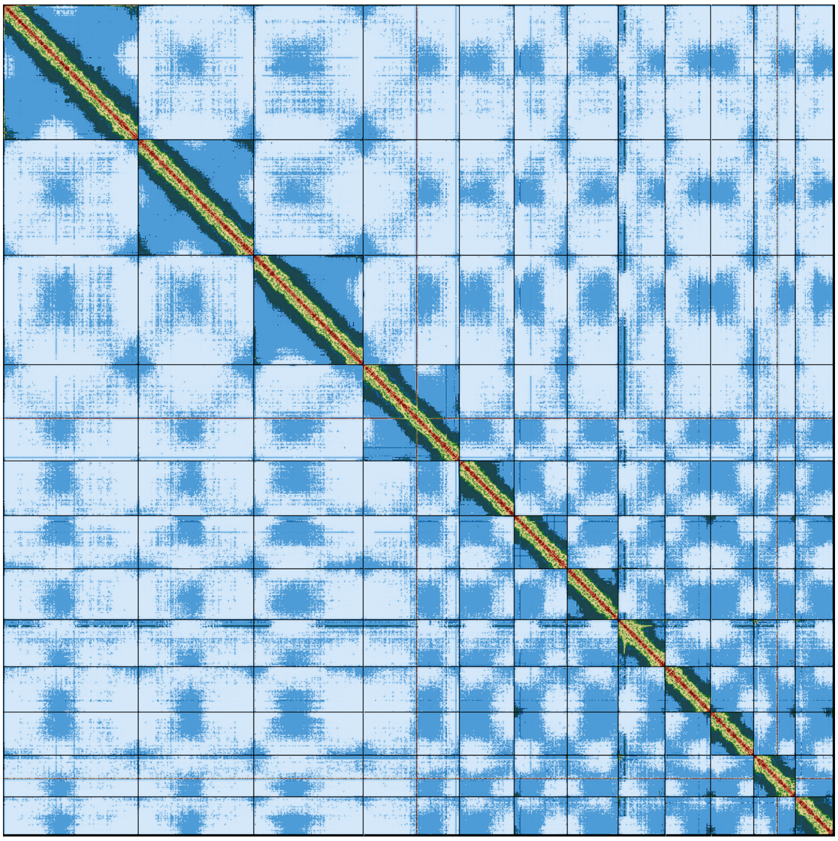
. Interventions/Gb: 43
. Contamination notes: ""
. Other observations: "The assembly of Glossodoris hikuerensis (xgGloHiku1) is based on 35X ONT data and 157X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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 9.212 Mb (with the largest being 8.728 Mb). Additionally, 509 regions totaling 65.704 Mb (with the largest being 1.969 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, 35 haplotypic regions and 2 contaminant sequences were removed, totaling 10.6Mb and 0.60Mb, respectively (with the largest being 0.96Mb and 0.35Mb). 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	2,517,602,200	2,506,359,050
GC %	37.53	37.52
Gaps/Gbp	202.18	205.88
Total gap bp	50,900	55,500
Scaffolds	78	66
Scaffold N50	290,496,467	288,763,574
Scaffold L50	4	4
Scaffold L90	11	10
Contigs	587	582
Contig N50	8,068,234	8,165,290
Contig L50	97	96
Contig L90	312	312
QV	41.2494	41.2511
Kmer compl.	86.9497	86.8184
BUSCO sing.	97.1%	97.2%
BUSCO dupl.	1.2%	1.0%
BUSCO frag.	0.4%	0.4%
BUSCO miss.	1.2%	1.3%

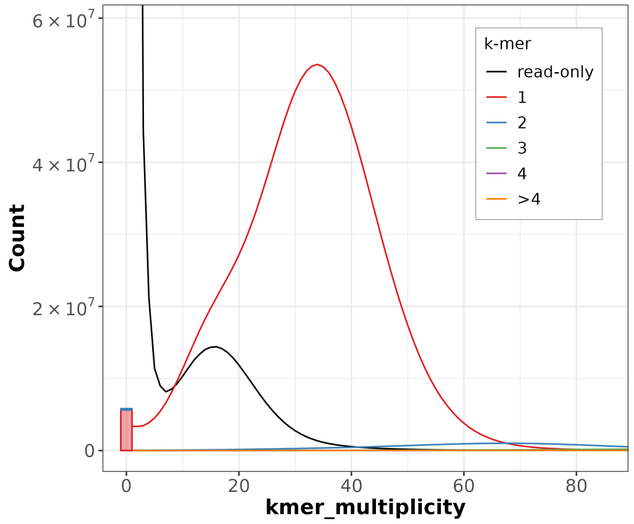
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

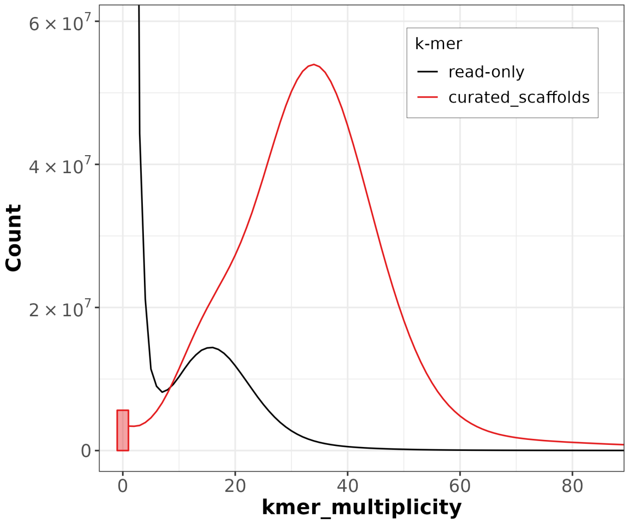


collapsed [LINK]

K-mer spectra of curated assembly

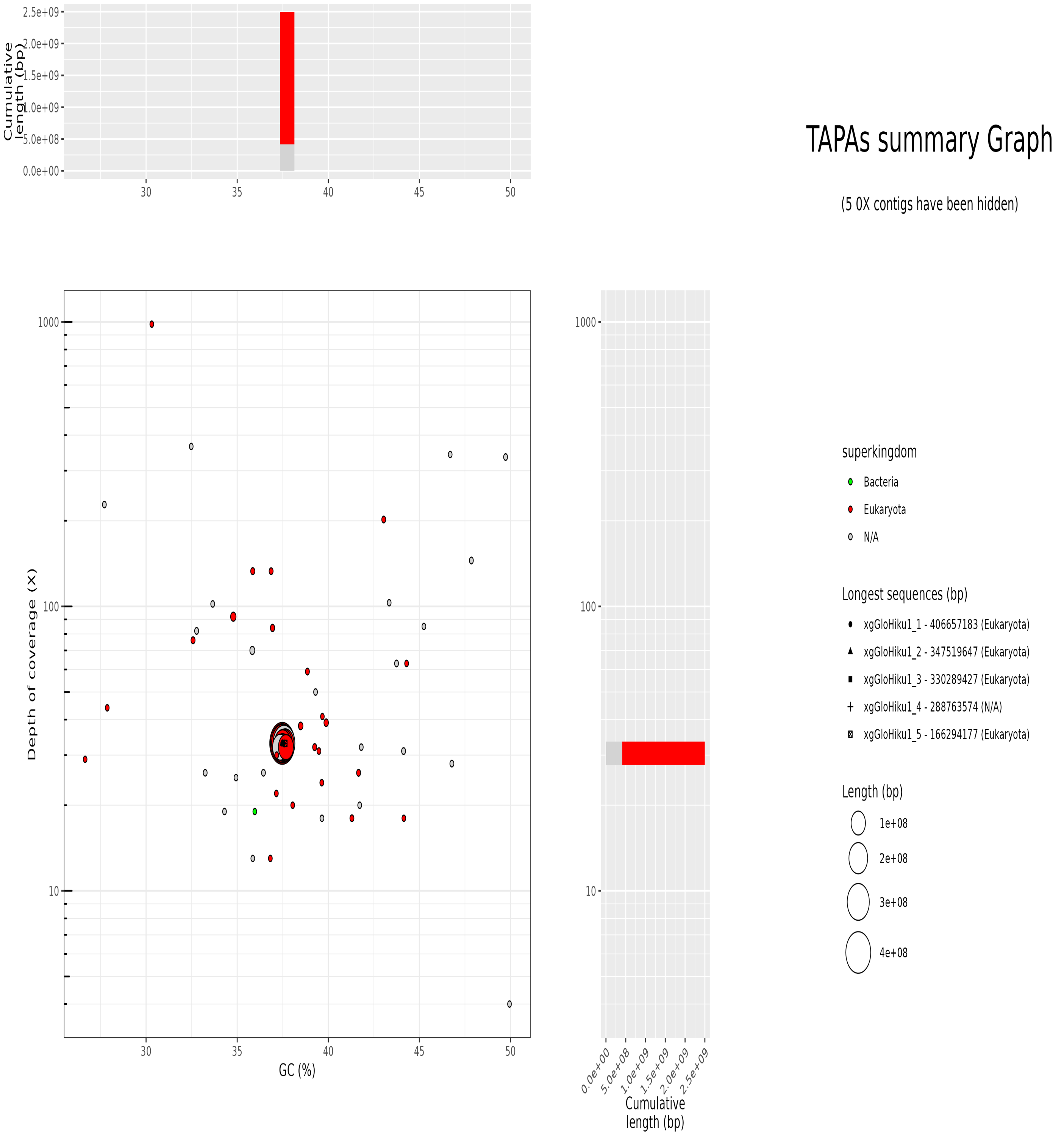


Distribution of k-mer counts per copy numbers found in asm



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	35	156

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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2025-12-09 17:11:44 CET