ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	104710
ToLID	**wpGlyTrid2**
Species	Glycera tridactyla
Class	Polychaeta
Order	Phyllodocida

Genome Traits	Expected	Observed
Haploid size (bp)	1,886,247,747	2,038,292,761
Haploid Number	10 (source: ancestor)	19
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q45

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
. BUSCO single copy value is less than 90% for collapsed
. Assembly length loss > 3% for collapsed

Curator notes

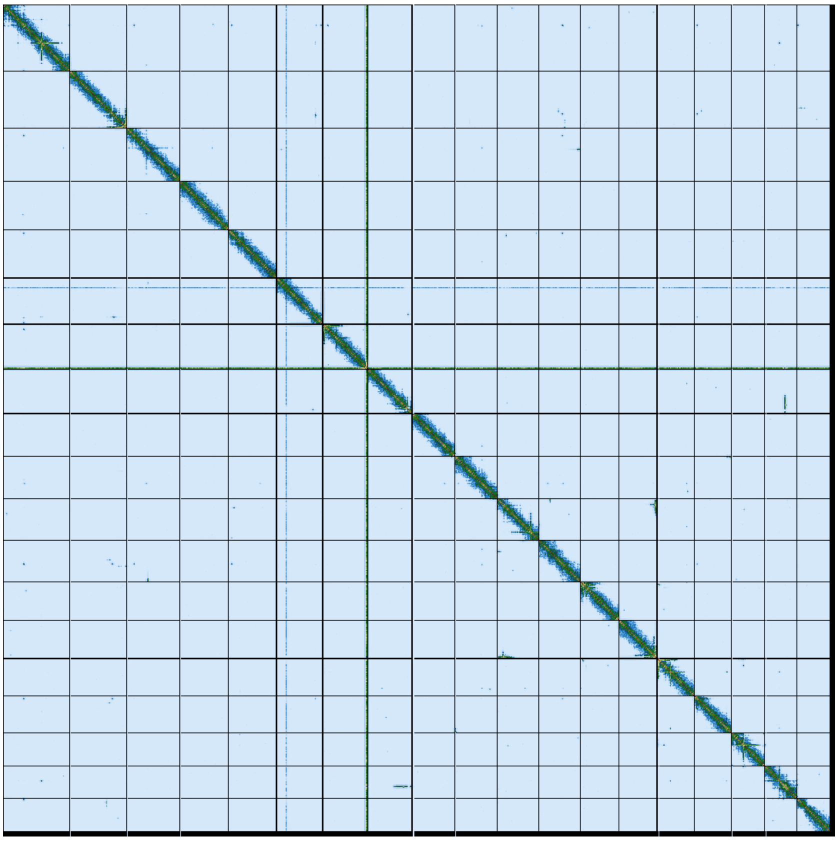
. Interventions/Gb: 175
. Contamination notes: ""
. Other observations: "The assembly of Glycera tridactyla (wpGlyTrid2.1) is based on 50X of PacBio data and 205X of Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: 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, 20 contigs were identified as contaminants (bacterial and archea), totaling 15.11 Mb (with the largest being 13.08 Mb). Additionally, 1378 regions totaling 377.85 Mb (with the largest being 4.42 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 204 haplotypic regions totaling 354.15 Mb were removed (with the largest being 10.27 Mb). 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	2,392,577,811	2,038,292,761
GC %	42.47	42.49
Gaps/Gbp	402.49	387.58
Total gap bp	96,300	87,300
Scaffolds	331	207
Scaffold N50	119,128,413	103,788,147
Scaffold L50	9	9
Scaffold L90	19	17
Contigs	1,294	997
Contig N50	4,086,370	5,291,057
Contig L50	172	120
Contig L90	612	415
QV	46.0317	45.9926
Kmer compl.	71.3344	63.5972
BUSCO sing.	77.1%	88.1%
BUSCO dupl.	14.0%	2.1%
BUSCO frag.	6.9%	7.3%
BUSCO miss.	2.0%	2.6%

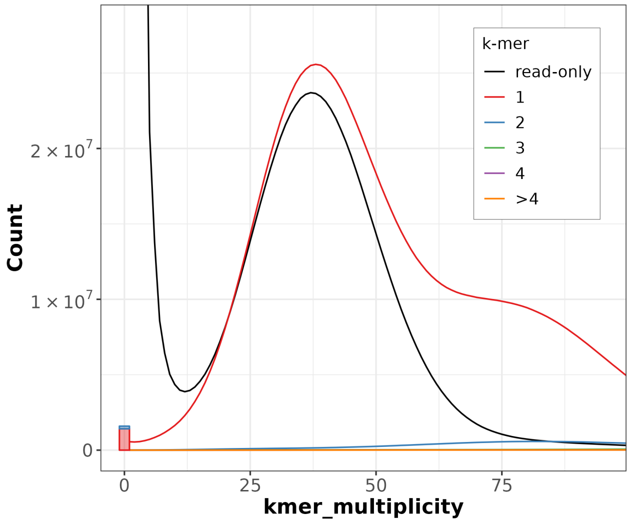
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)

HiC contact map of curated assembly

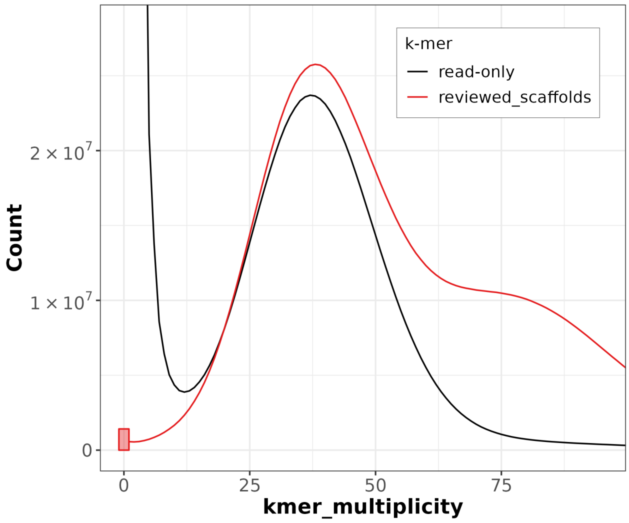


collapsed [LINK]

K-mer spectra of curated assembly

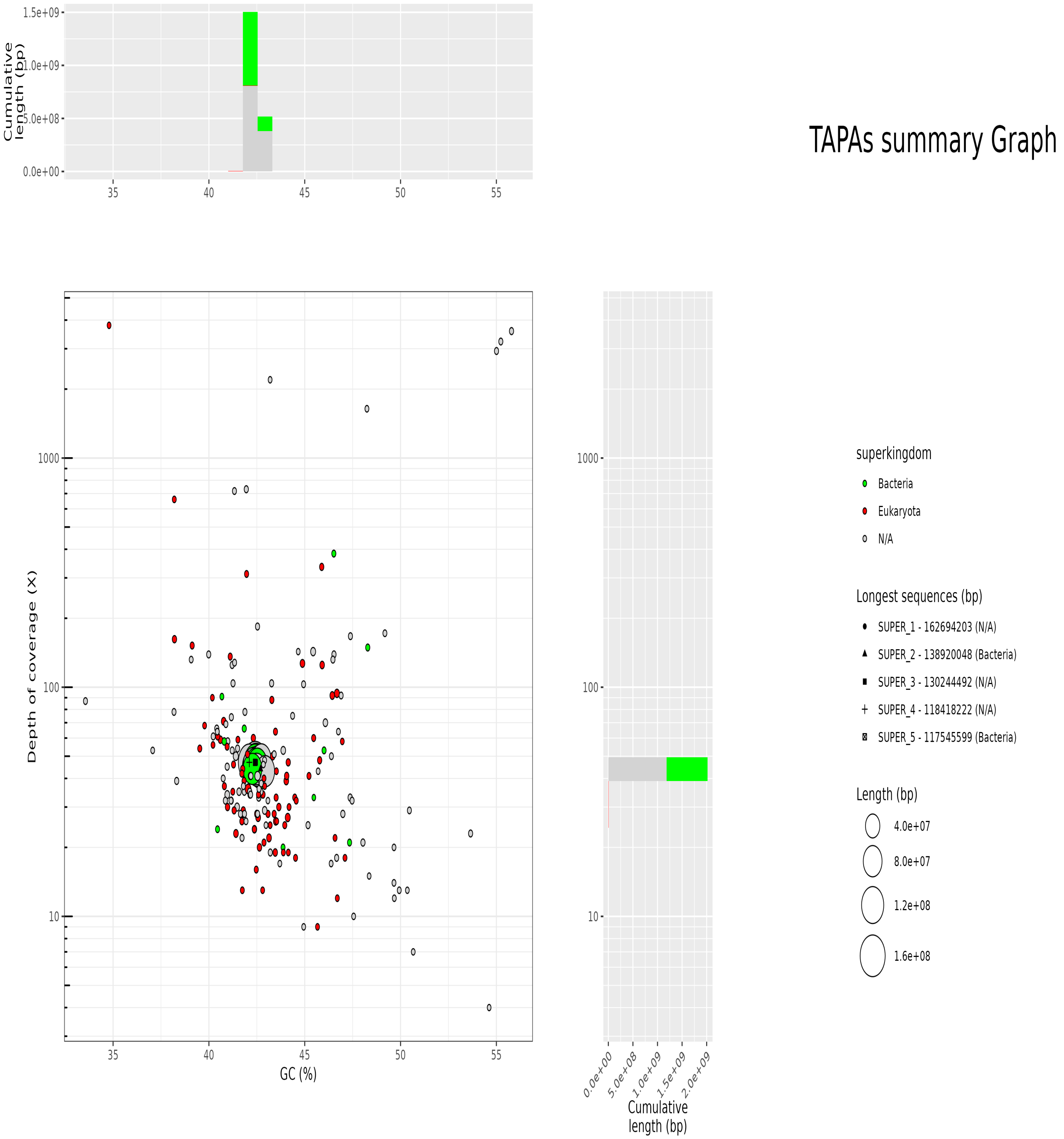


Distribution of k-mer counts per copy numbers found in asm



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 Hifi	Omnic
Coverage	50	205

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: Adama Ndar
Affiliation: Genoscope

Date and time: 2025-06-14 17:05:49 CEST