

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	2862683
ToLID	wtEupGili2
Species	Eupolymnia gili
Class	Polychaeta
Order	Terebellida

Genome Traits	Expected	Observed
Haploid size (bp)	1,266,851,558	1,248,093,907
Haploid Number	3 (source: ancestor)	17
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.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

Curator notes

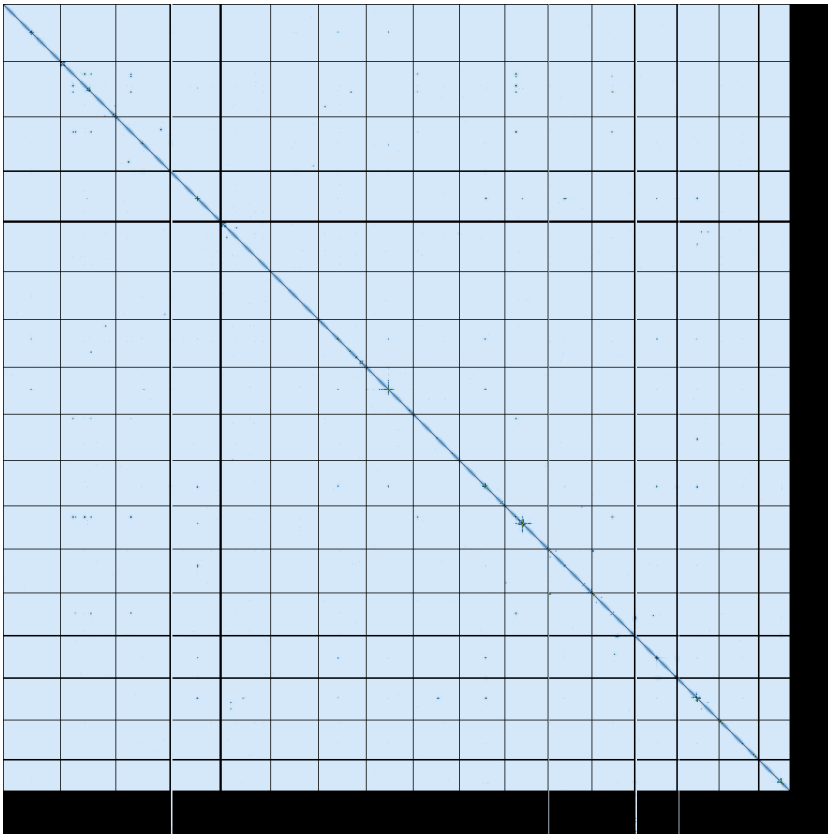
. Interventions/Gb: 71
. Contamination notes: ""
. Other observations: "The assembly of Eupolymnia gili (wtEupGili2) is based on 79X PACBIO data and 269X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial PACBIO assembly generation with Hifiiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 15 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.369 Mb (with the largest being 0.389 Mb). Additionally, 949 regions totaling 81.938 Mb (with the largest being 4.242 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, 9 haplotypic regions were removed, totaling 22.413 Mb (with the largest being 5.133 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	1,270,964,297	1,248,093,907
GC %	41.53	41.51
Gaps/Gbp	95.99	116.98
Total gap bp	12,200	19,900
Scaffolds	840	753
Scaffold N50	68,719,366	69,543,071
Scaffold L50	9	9
Scaffold L90	17	16
Contigs	962	899
Contig N50	15,181,000	15,308,000
Contig L50	28	27
Contig L90	109	106
QV	45.1382	45.328
Kmer compl.	69.0118	68.3915
BUSCO sing.	88.1%	89.1%
BUSCO dupl.	2.9%	1.7%
BUSCO frag.	7.2%	7.3%
BUSCO miss.	1.8%	1.8%

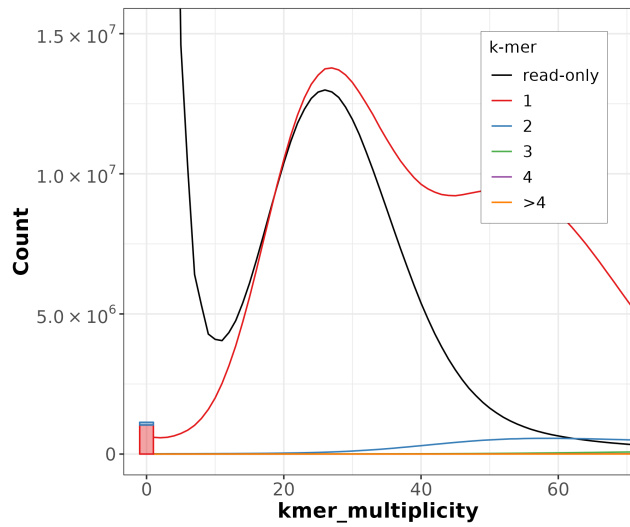
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)

HiC contact map of curated assembly

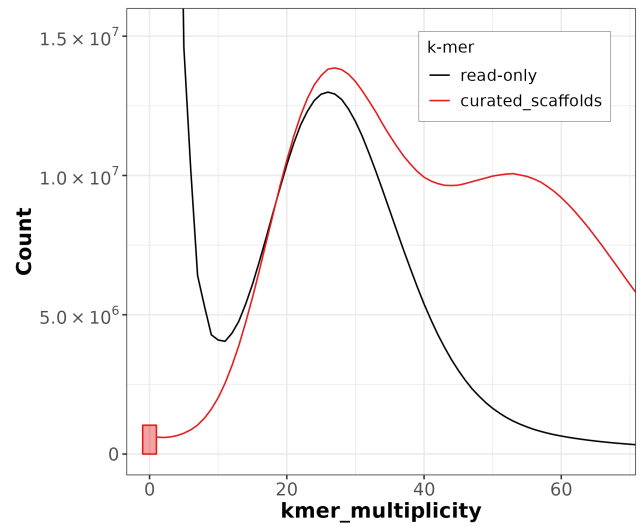


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

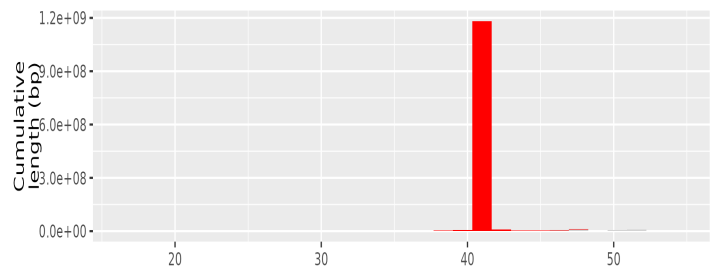


Distribution of k-mer counts per copy numbers found in asm



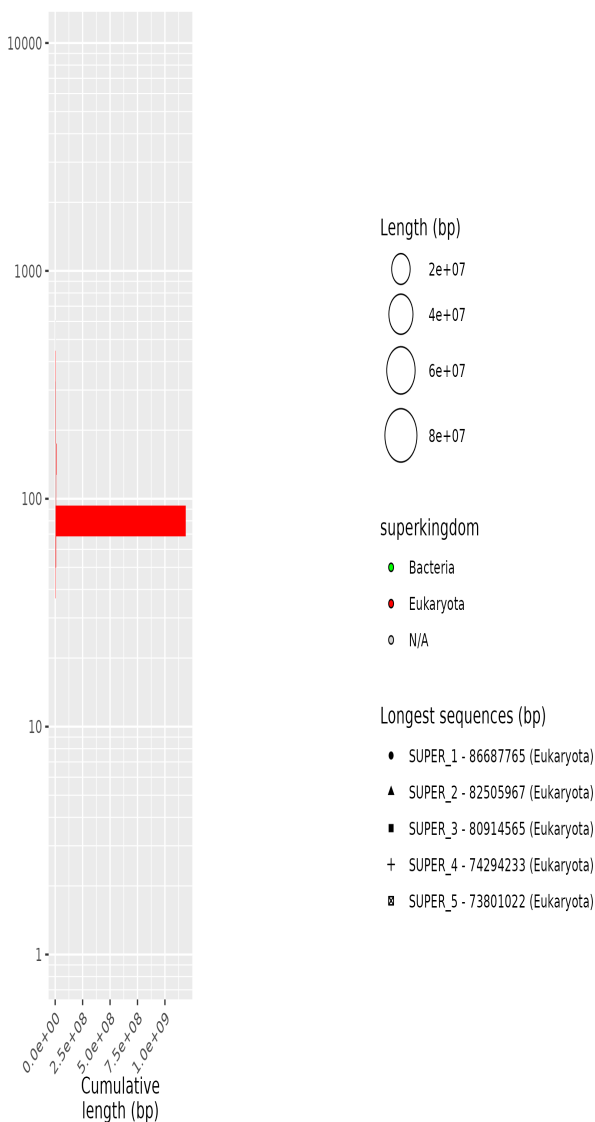
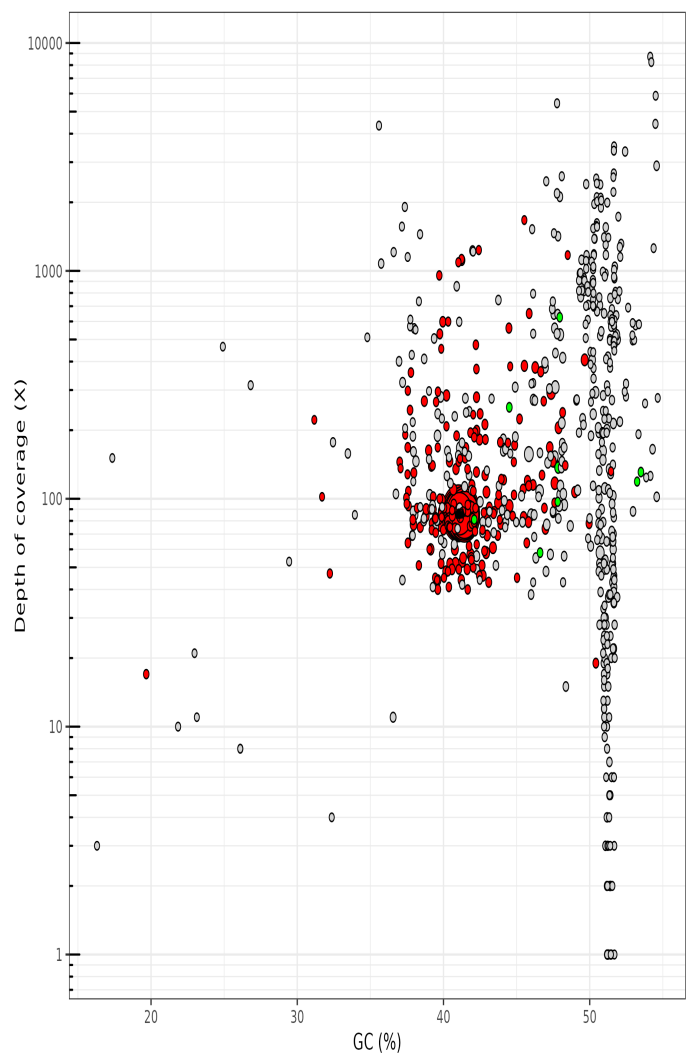
Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph

(40 0X contigs have been hidden)



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	79	269

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: 2025-09-16 15:57:36 CEST