ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1367176
ToLID	**wsBraBomb5**
Species	Branchiomma bombyx
Class	Polychaeta
Order	Sabellida

Genome Traits	Expected	Observed
Haploid size (bp)	1,217,366,717	1,268,251,189
Haploid Number	4 (source: ancestor)	14
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.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
. Assembly length loss > 3% for collapsed

Curator notes

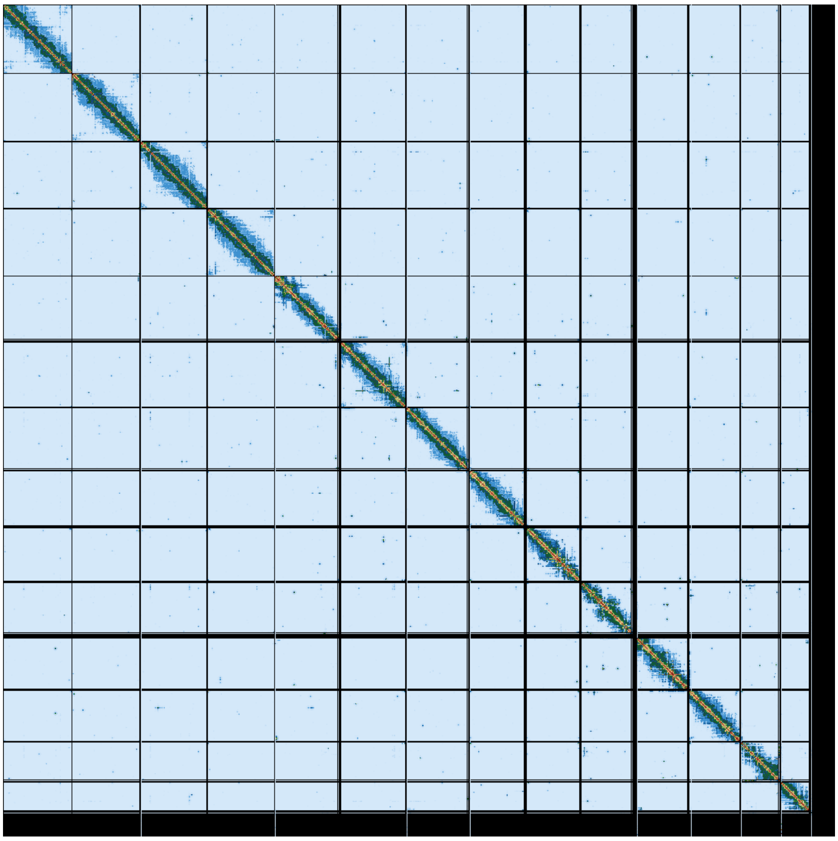
. Interventions/Gb: 73
. Contamination notes: ""
. Other observations: "The assembly of Branchiomma bombyx (wsBraBomb5) is based on 48X PacBio data and 114X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, archaeal, or viral), totaling 2.749 Mb (with the largest being 0.708 Mb). Additionally, 445 regions totaling 193.207 Mb (with the largest being 6.952 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. During manual curation, 38 haplotypic regions were removed, totaling 490Mb (with the largest being 44.7Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. Large haplotypic duplications removed during manual curation. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,758,320,352	1,268,251,189
GC %	35.64	35.67
Gaps/Gbp	219.53	234.18
Total gap bp	39,800	36,000
Scaffolds	295	267
Scaffold N50	88,153,495	92,569,470
Scaffold L50	8	7
Scaffold L90	23	13
Contigs	668	564
Contig N50	6,834,668	6,641,100
Contig L50	65	46
Contig L90	251	193
QV	50.3031	50.4215
Kmer compl.	77.842	60.2219
BUSCO sing.	56.1%	97.0%
BUSCO dupl.	32.4%	1.0%
BUSCO frag.	9.2%	0.8%
BUSCO miss.	2.3%	1.2%

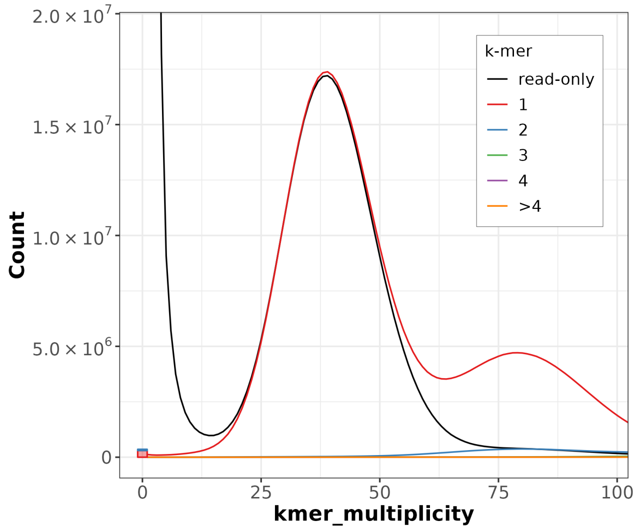
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)

HiC contact map of curated assembly

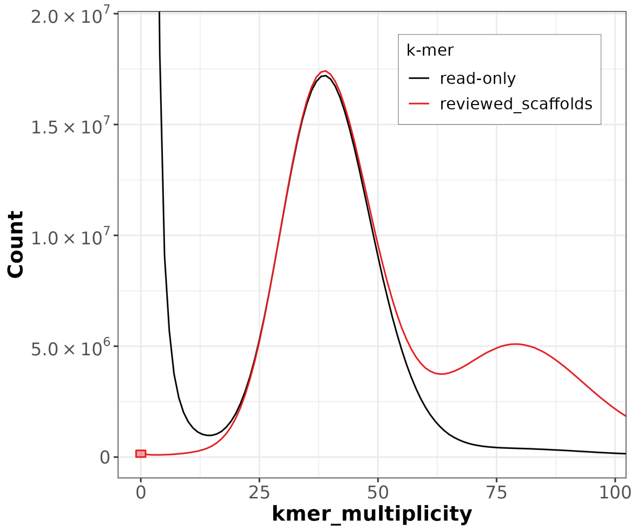


collapsed [LINK]

K-mer spectra of curated assembly

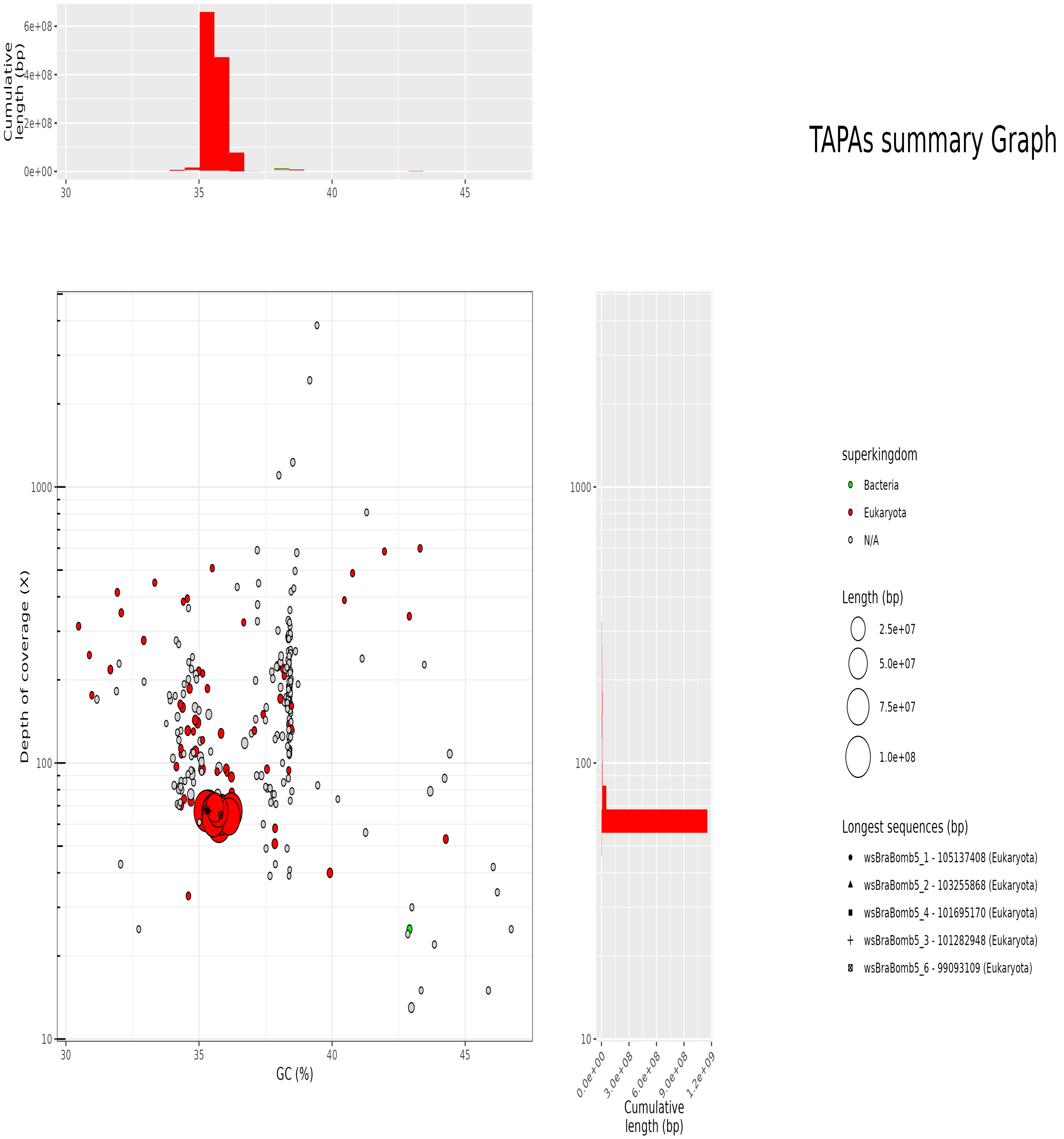


Distribution of k-mer counts per copy numbers found in asm



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	Long reads	Arima
Coverage	69	164

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: Sophie Layac
Affiliation: Genoscope

Date and time: 2025-11-27 04:44:17 CET