ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	188456
ToLID	**wjPauBell1**
Species	Paucibranchia bellii
Class	Polychaeta
Order	Eunicida

Genome Traits	Expected	Observed
Haploid size (bp)	3,305,276,407	3,834,031,046
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: 6.8.Q42

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

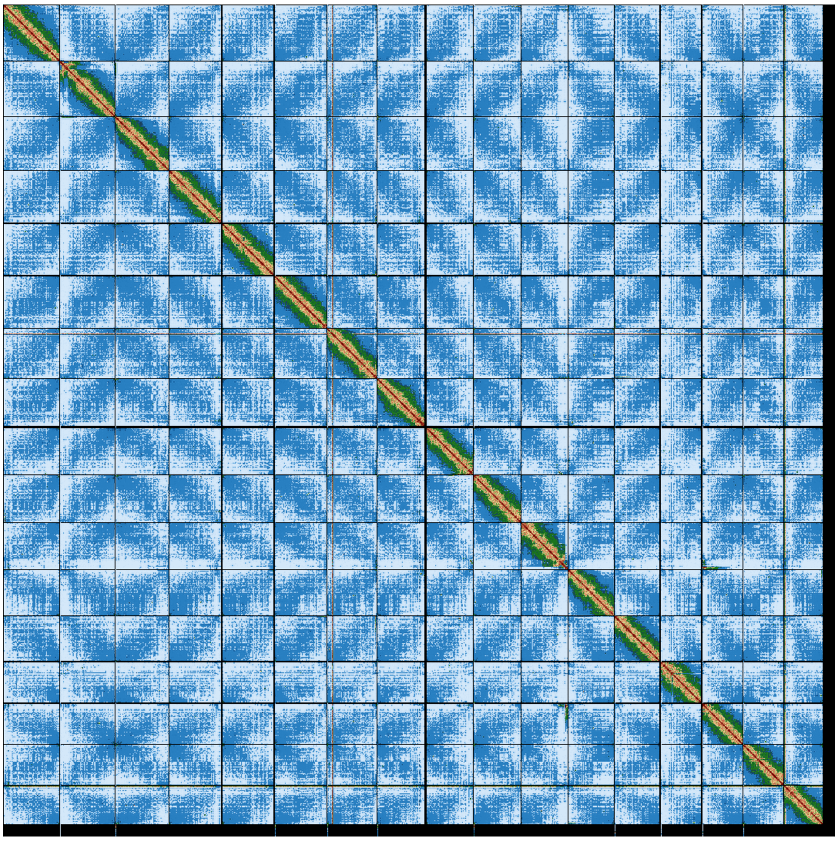
. Interventions/Gb: 191
. Contamination notes: ""
. Other observations: "The assembly of Paucibranchia belii (wjPauBell1) is based on 43X PacBio data and Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included the following steps: initial PacBio 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, 84 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 6.4 Mb (with the largest being 0.697Mb). Additionally, 3,371 regions totaling 800 Mb (with the largest being 3.5 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, 285 haplotypic regions were removed,totaling 1.431 Gb (with the largest being 78 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	5,267,016,779	3,834,031,046
GC %	37.4	37.51
Gaps/Gbp	368.9	422.27
Total gap bp	194,300	186,400
Scaffolds	588	395
Scaffold N50	257,295,204	219,199,435
Scaffold L50	9	8
Scaffold L90	27	16
Contigs	2,531	2,014
Contig N50	4,279,788	4,182,163
Contig L50	369	275
Contig L90	1,251	939
QV	43.0378	42.8502
Kmer compl.	78.2827	61.4393
BUSCO sing.	74.5%	84.8%
BUSCO dupl.	17.8%	2.6%
BUSCO frag.	6.5%	6.6%
BUSCO miss.	1.1%	5.9%

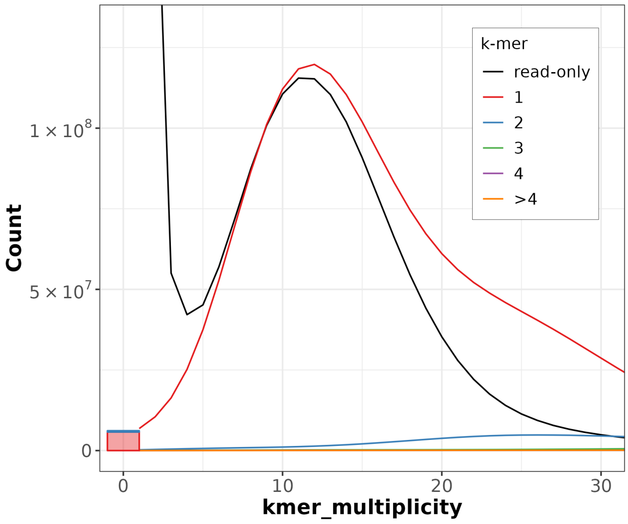
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)

HiC contact map of curated assembly

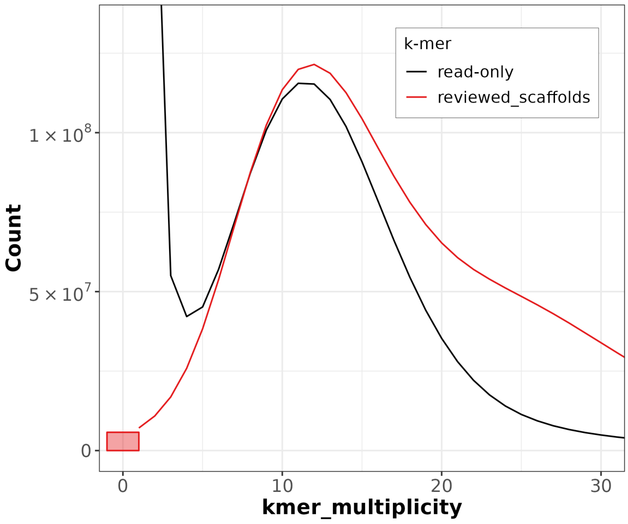


collapsed [LINK]

K-mer spectra of curated assembly

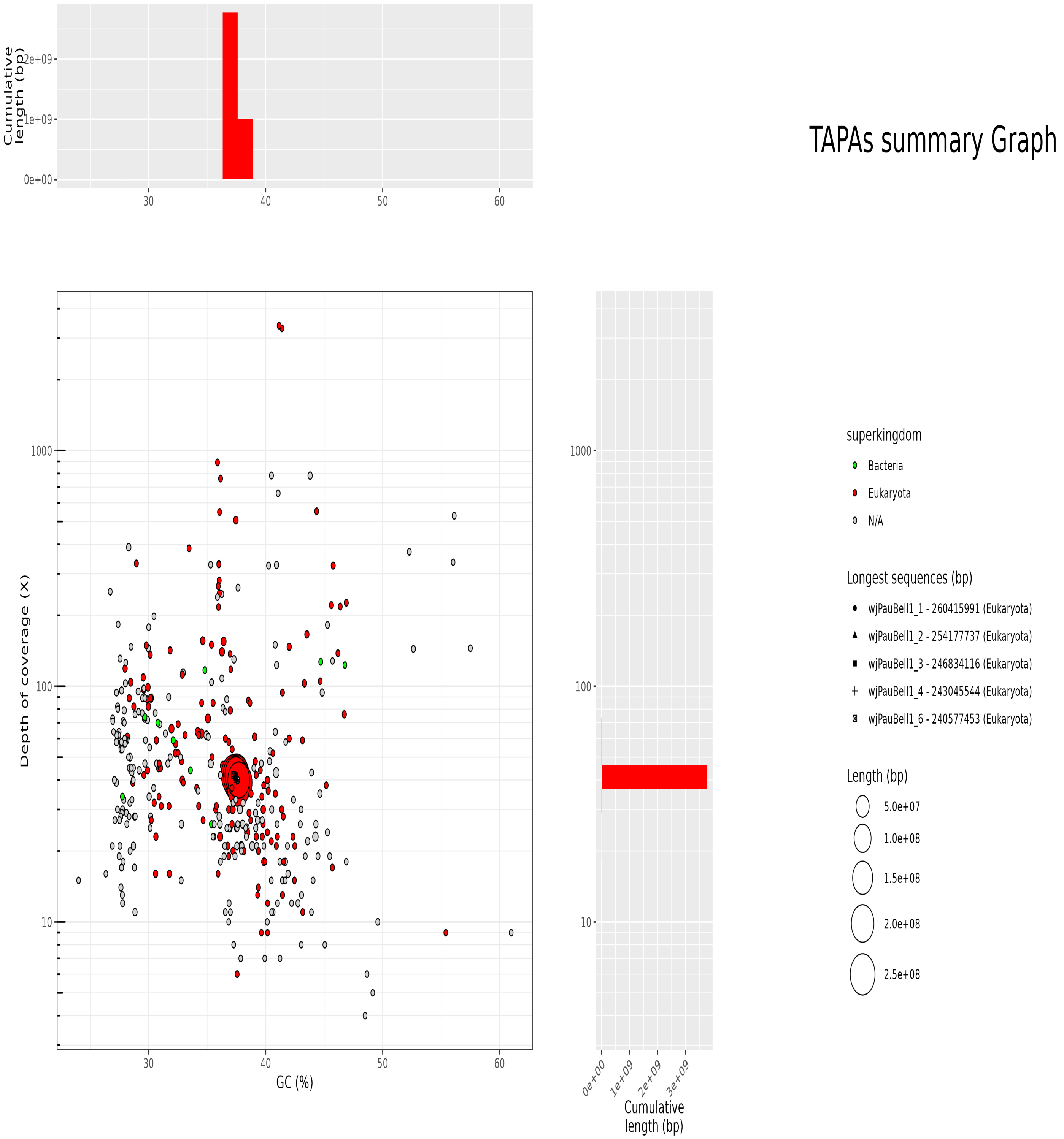


Distribution of k-mer counts per copy numbers found in asm



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	Long reads	Arima
Coverage	43	190

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: Caroline Menguy
Affiliation: Genoscope

Date and time: 2025-09-29 00:27:01 CEST