ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	3093271
ToLID	**wjEunSchi1**
Species	Eunice schizobranchia
Class	Polychaeta
Order	Eunicida

Genome Traits	Expected	Observed
Haploid size (bp)	1,836,355,602	1,837,776,588
Haploid Number	3 (source: ancestor)	13
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q47

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

Curator notes

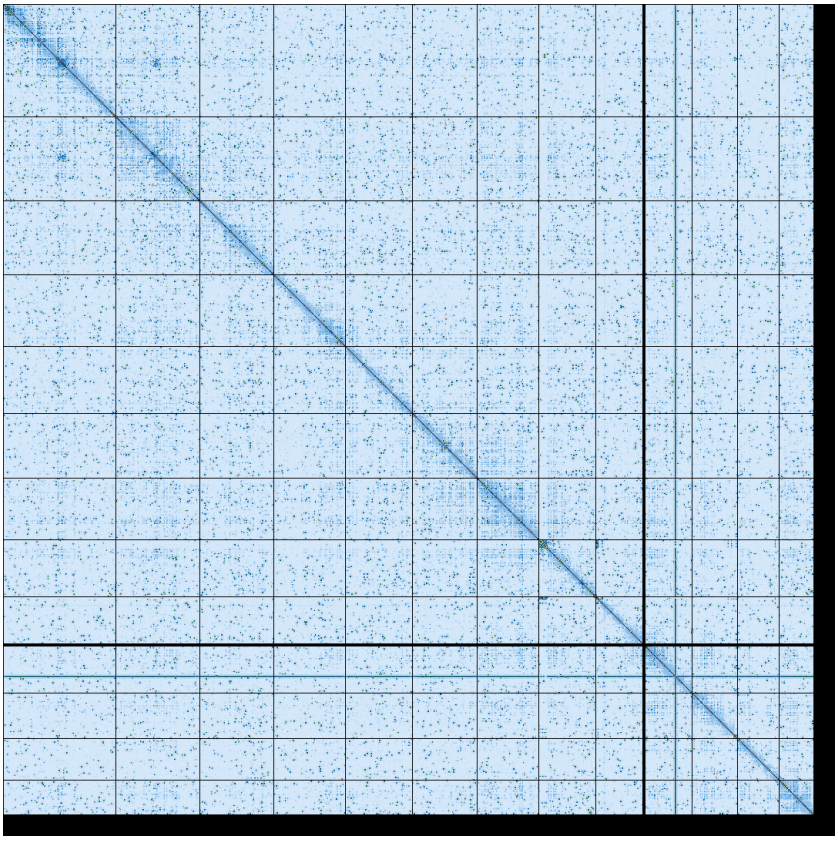
. Interventions/Gb: 69
. Contamination notes: ""
. Other observations: "The assembly of \'Eunice schizobranchia\' (wjEunSchi1) is based on 46X PacBio data and 195X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 17 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 3.278 Mb (with the largest being 1.744 Mb). Additionally, 881 regions totaling 78.19 Mb (with the largest being 2.774 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 31 haplotypic regions were removed, totaling 12.432Mb (with the largest being 2.539Mb). 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	1,851,112,789	1,837,776,588
GC %	38.89	38.88
Gaps/Gbp	283.61	315.05
Total gap bp	52,500	67,100
Scaffolds	851	735
Scaffold N50	143,538,930	143,875,112
Scaffold L50	6	6
Scaffold L90	12	12
Contigs	1,376	1,314
Contig N50	13,898,000	13,898,000
Contig L50	38	38
Contig L90	148	146
QV	47.5683	47.5748
Kmer compl.	70.4902	70.2497
BUSCO sing.	96.6%	95.1%
BUSCO dupl.	2.5%	1.4%
BUSCO frag.	0.6%	2.8%
BUSCO miss.	0.4%	0.6%

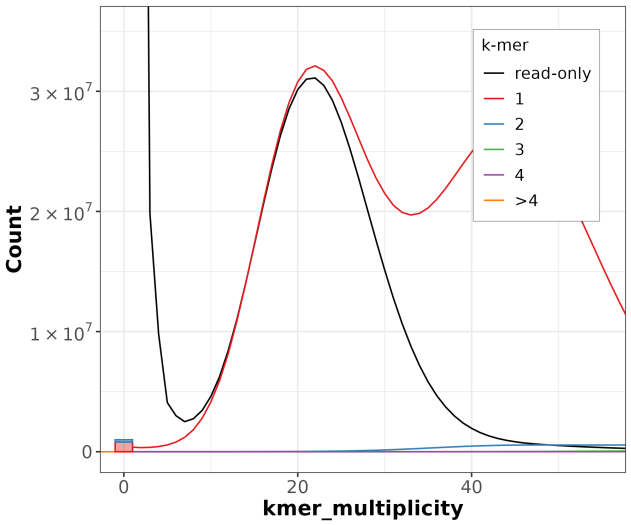
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)

HiC contact map of curated assembly

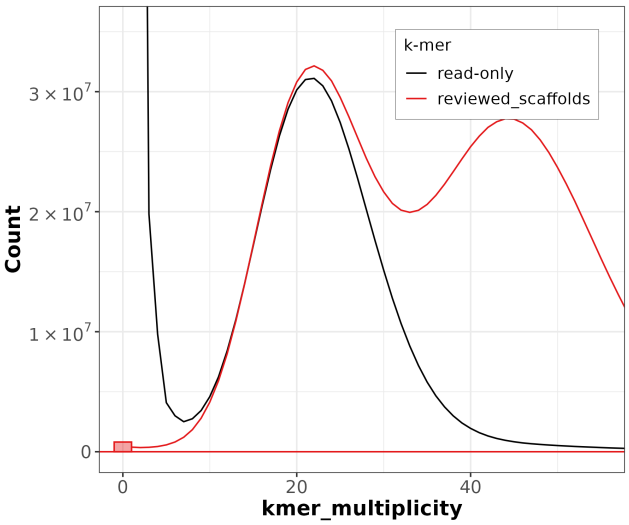


collapsed [LINK]

K-mer spectra of curated assembly

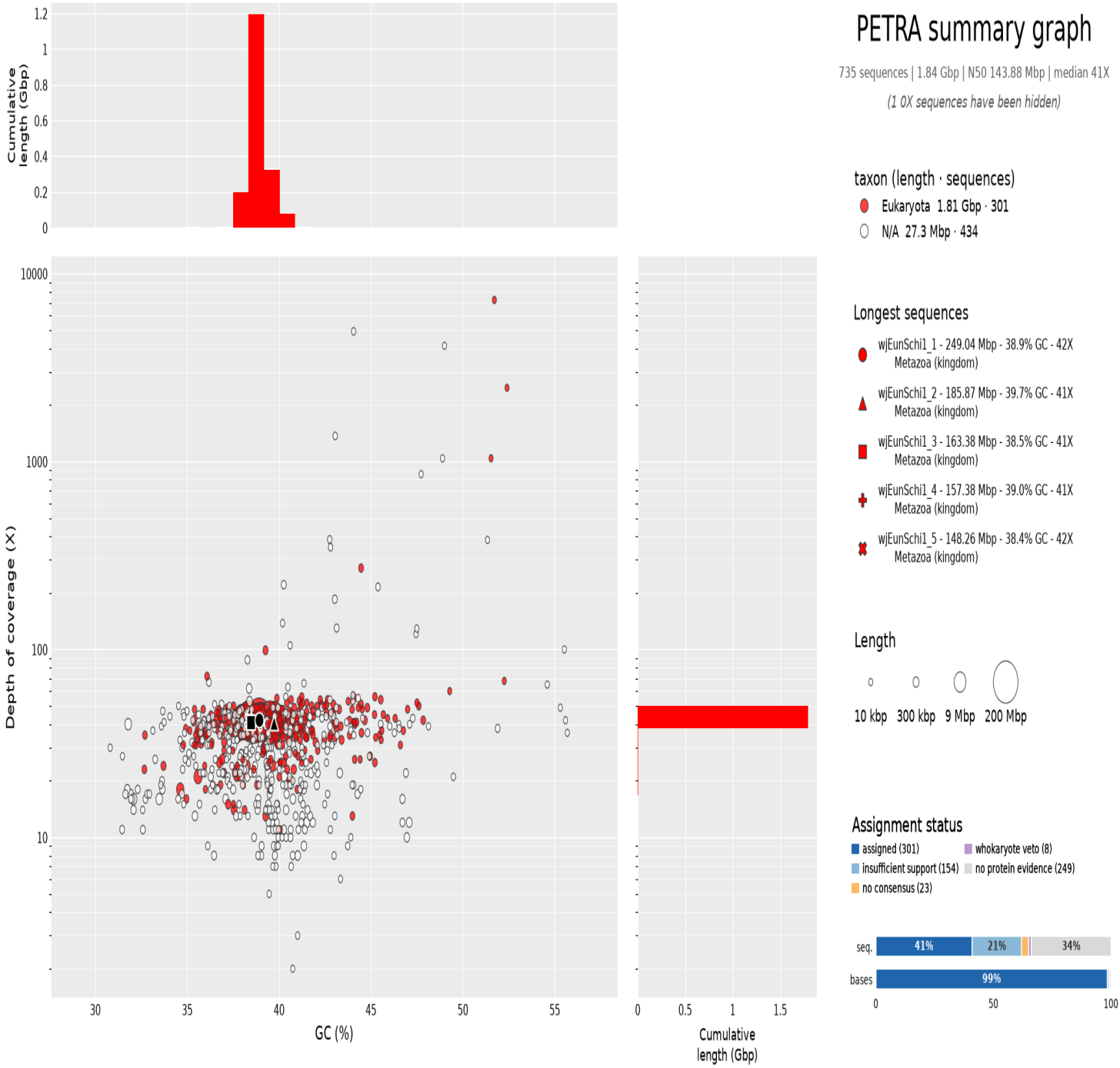


Distribution of k-mer counts per copy numbers found in asm



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	46	196

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:* 0.2.4
  |_ *key param:* NA
- **PretextView**
  |_ *ver:* 1.0.7
  |_ *key param:* NA

Submitter: Simone Duprat
Affiliation: Genoscope

Date and time: 2026-10-17 11:00:41 CEST