ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	983666
ToLID	**hrPseSple1**
Species	Pseudobiceros splendidus
Class	Rhabditophora
Order	Polycladida

Genome Traits	Expected	Observed
Haploid size (bp)	3,033,532,512	3,206,858,410
Haploid Number	9 (source: ancestor)	9
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q63

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

Curator notes

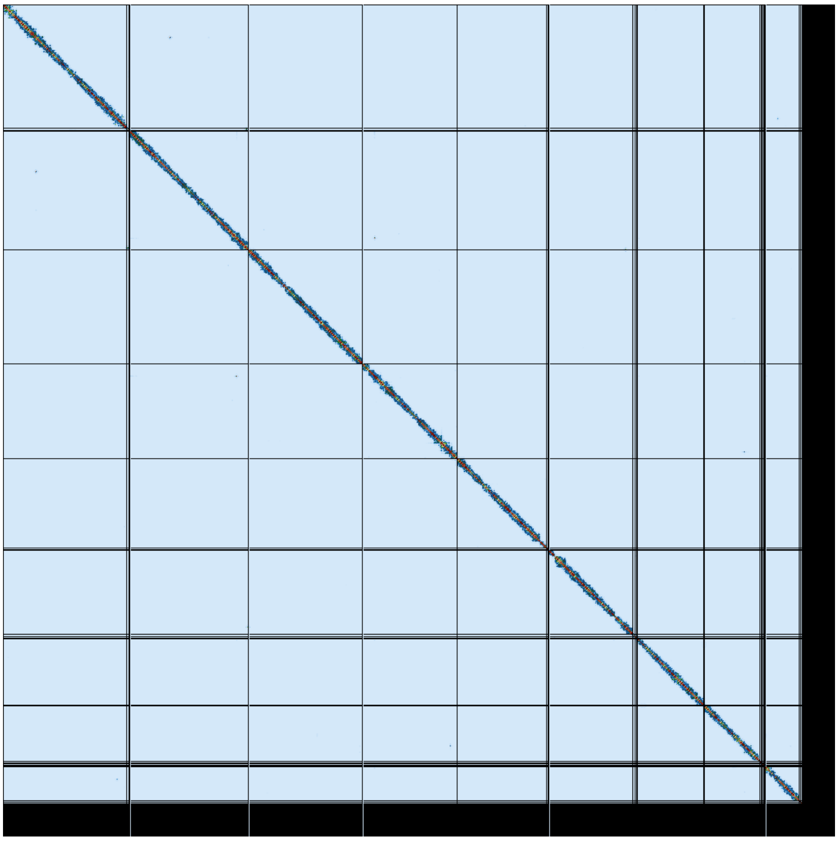
. Interventions/Gb: 74
. Contamination notes: ""
. Other observations: "The assembly of Pseudobiceros splendidus (hrPseSple1.1) is based on 40X of PacBio data and Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 6 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.6 Mb (with the largest being 0.54 Mb). Additionally, 1002 regions totaling 246 Mb were identified as haplotypic duplications and removed (with the largest being 11 Mb). The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 47 haplotypic regions and 0 contaminant sequences were removed, totaling 26 Mb (with the largest being 2.4 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	3,234,438,130	3,206,858,410
GC %	40.14	40.13
Gaps/Gbp	297.73	306.22
Total gap bp	96,300	109,300
Scaffolds	609	495
Scaffold N50	361,629,564	365,166,217
Scaffold L50	4	4
Scaffold L90	9	9
Contigs	1,572	1,477
Contig N50	6,394,019	6,449,511
Contig L50	152	150
Contig L90	533	522
QV	49.8337	63.3433
Kmer compl.	87.0633	87.4328
BUSCO sing.	81.2%	
BUSCO dupl.	1.3%	
BUSCO frag.	6.1%	
BUSCO miss.	11.4%	

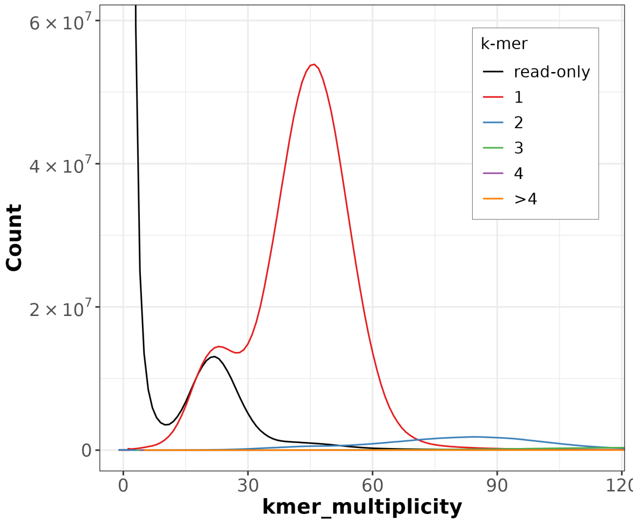
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

HiC contact map of curated assembly

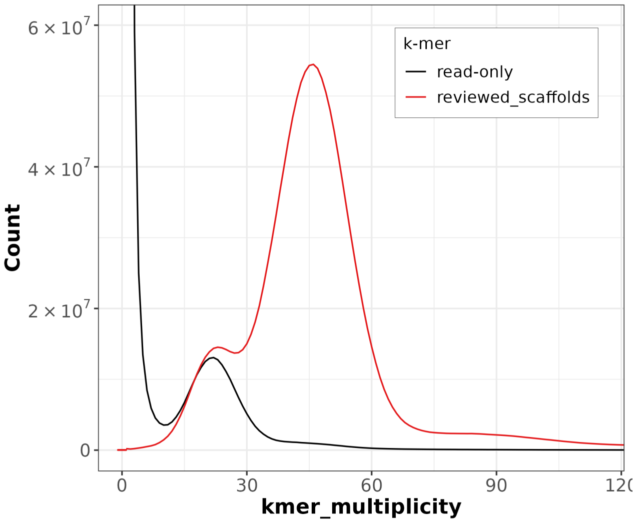


collapsed [LINK]

K-mer spectra of curated assembly

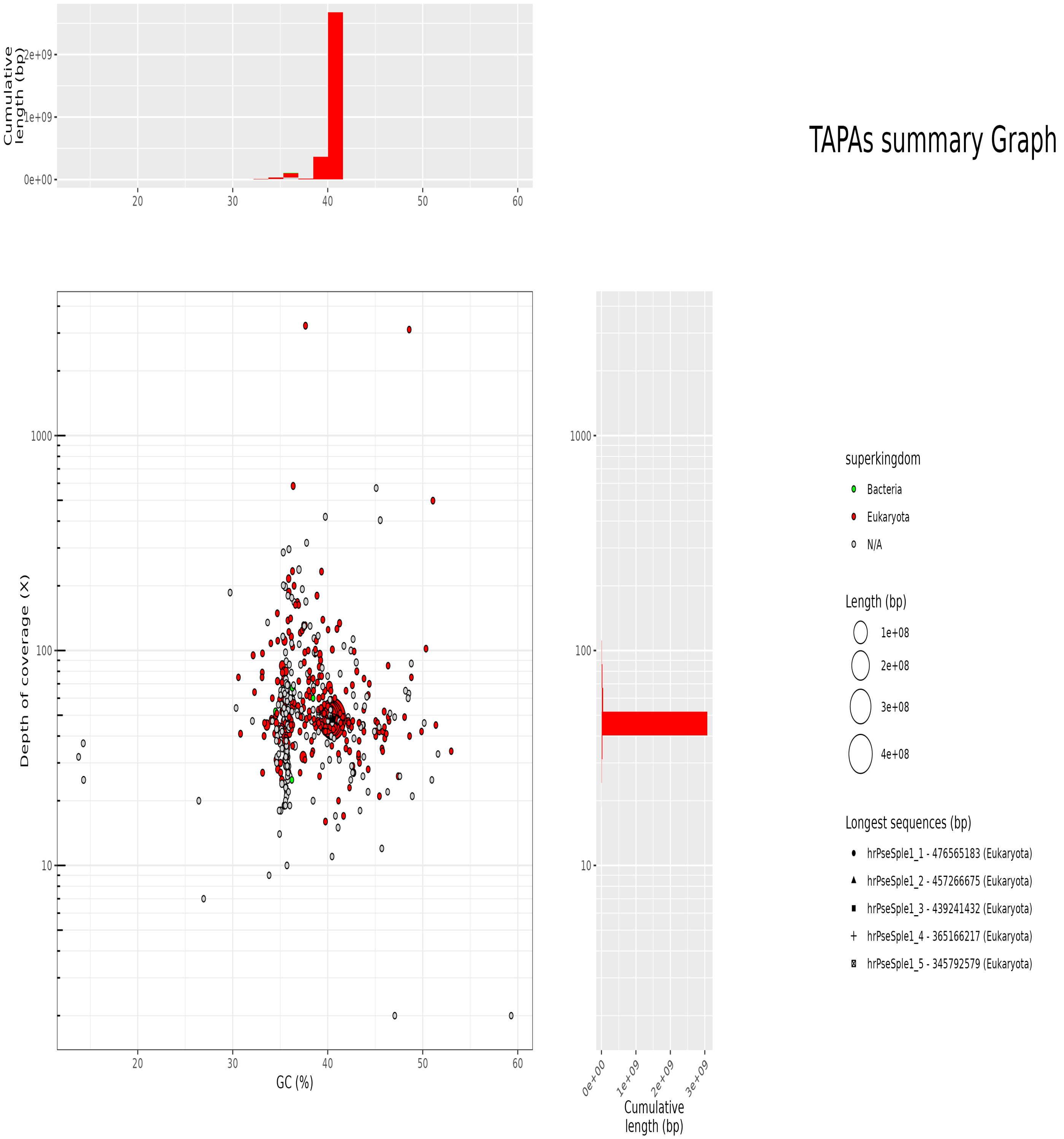


Distribution of k-mer counts per copy numbers found in asm



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	Arima
Coverage	44	24

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: Arnaud Couloux
Affiliation: Genoscope

Date and time: 2025-03-25 10:01:24 CET