ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	2821464
ToLID	**fZeuReg1**
Species	Zeugopterus regius
Class	Actinopteri
Order	Pleuronectiformes

Genome Traits	Expected	Observed
Haploid size (bp)	454,675,421	476,591,588
Haploid Number	21 (source: ancestor)	14
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q43

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

. Observed Haploid Number is different from Expected

Curator notes

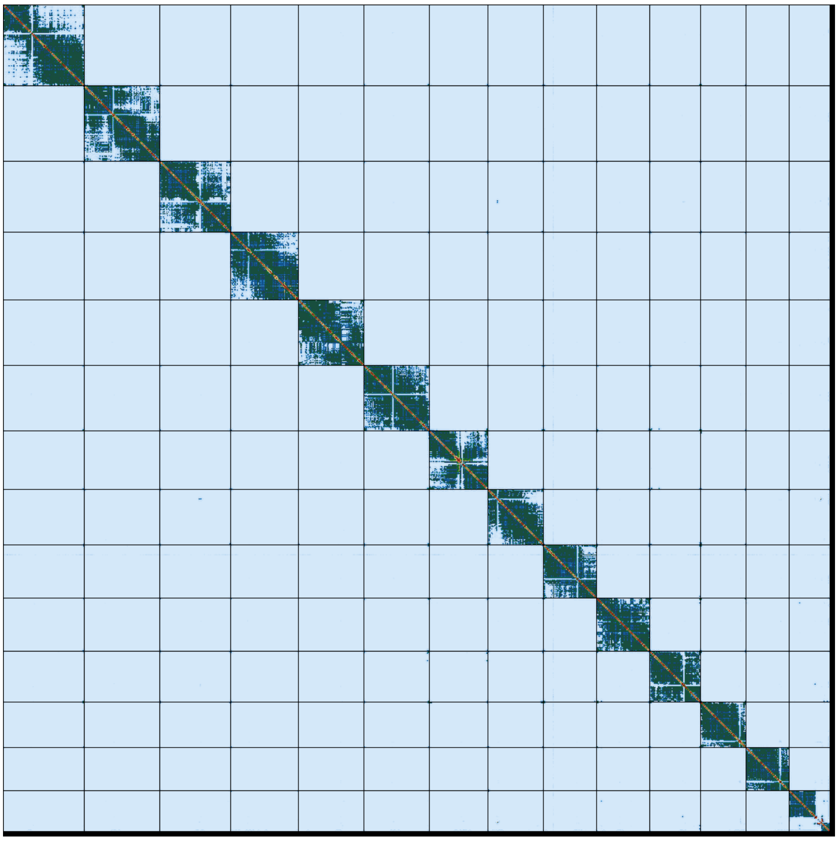
. Interventions/Gb: 3
. Contamination notes: ""
. Other observations: "The assembly of Zeugopterus regius (fZeuReg1) is based on 146X ONT data and 191X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial ONT assembly 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 0.012 Mb (with the largest being 0.003 Mb). Additionally, 613 regions totaling 9.027 Mb (with the largest being 0.453 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using ptGAUL. Finally, the primary assembly was analyzed and manually improved using Pretext. Several chromosomal regions, likely corresponding to centromeres, exhibit unusual patterns on the contact map, suggesting the presence of heterozygous variation. 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	478,355,387	476,591,588
GC %	44.98	44.98
Gaps/Gbp	39.72	46.16
Total gap bp	1,900	2,600
Scaffolds	1,260	76
Scaffold N50	37,332,450	37,332,450
Scaffold L50	6	6
Scaffold L90	13	13
Contigs	1,279	98
Contig N50	17,049,368	17,349,238
Contig L50	12	11
Contig L90	25	25
QV	42.9354	43.2744
Kmer compl.	94.5909	94.5899
BUSCO sing.	96.8%	96.8%
BUSCO dupl.	0.2%	0.2%
BUSCO frag.	0.6%	0.6%
BUSCO miss.	2.4%	2.4%

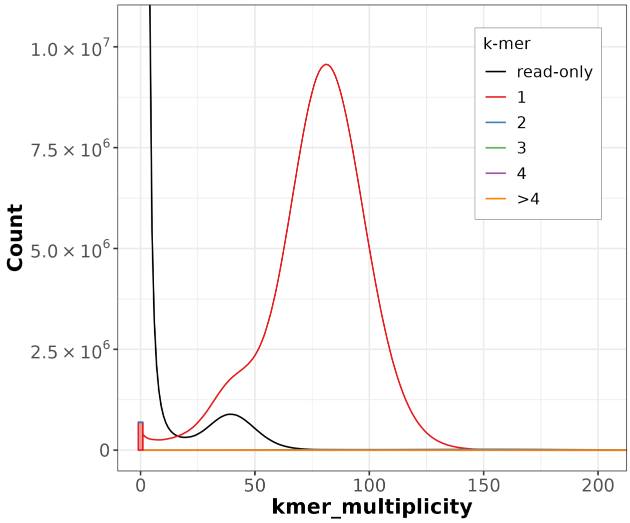
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

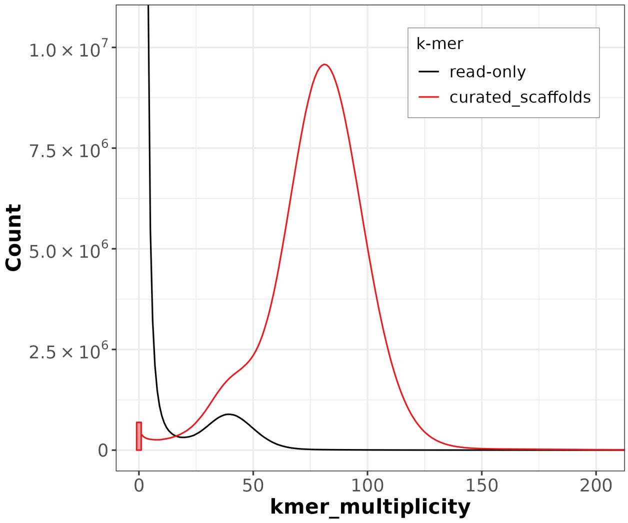


collapsed [LINK]

K-mer spectra of curated assembly

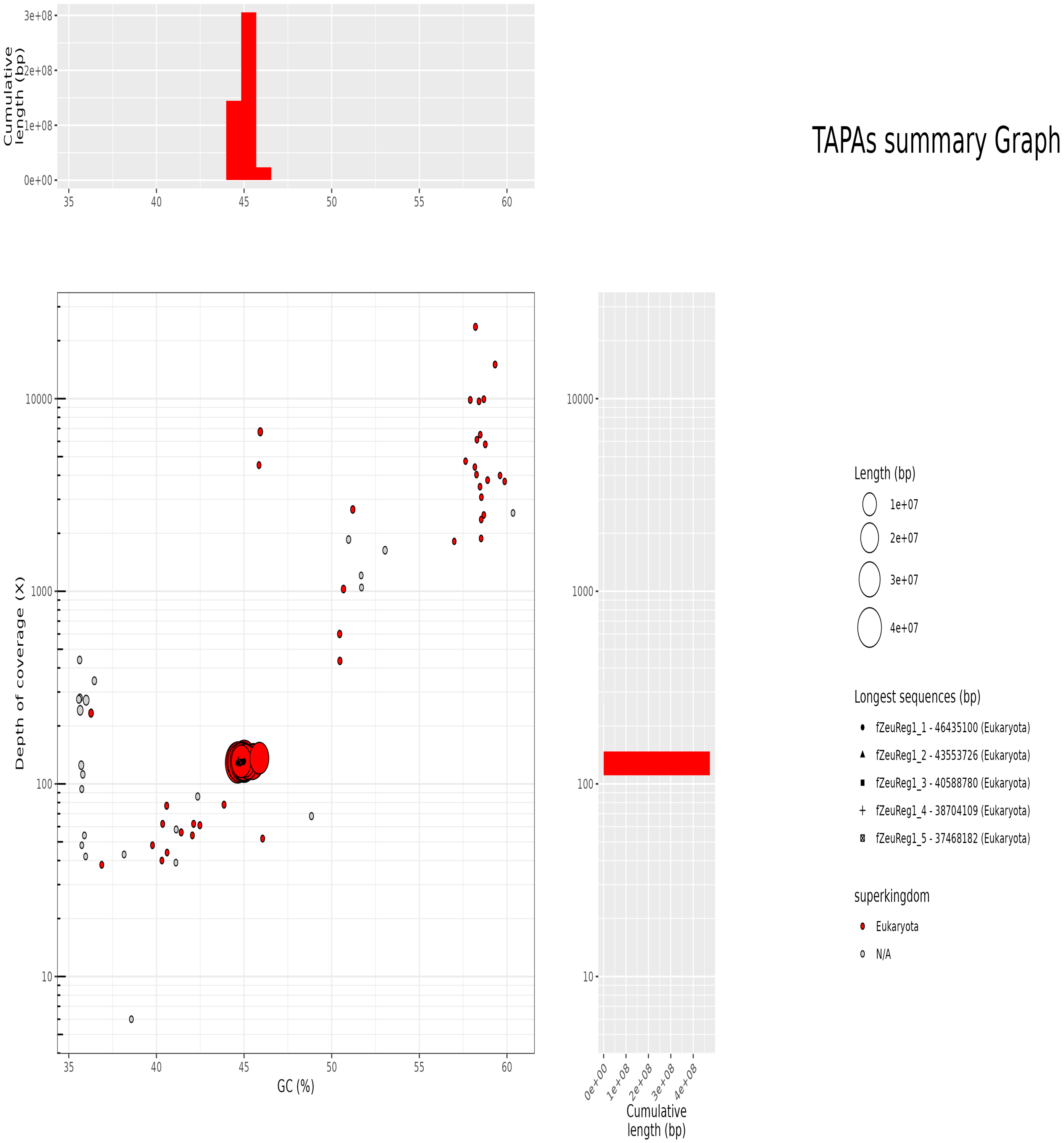


Distribution of k-mer counts per copy numbers found in asm



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	145	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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2025-07-21 08:21:12 CEST