

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	526623
ToLID	fZeuPun1.fa
Species	Zeugopterus punctatus
Class	Actinopteri
Order	Pleuronectiformes

Genome Traits	Expected	Observed
Haploid size (bp)	505,844,142	535,409,667
Haploid Number	21 (source: ancestor)	23
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q41

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
- . More than 1000 gaps/Gbp for collapsed

Curator notes

- . Interventions/Gb: 284
- . Contamination notes: ""
- . Other observations: "The assembly of Zeugopterus punctatus (fZeuPun1) is based on 43X PacBio data and 130X 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 Hifiiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 18 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 3.024 Mb (with the largest being 2.09 Mb). Additionally, 364 regions totaling 38.401 Mb (with the largest being 1.473 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. During manual curation, 73 haplotypic regions were removed, totaling 49Mb (with the largest being 1.8Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

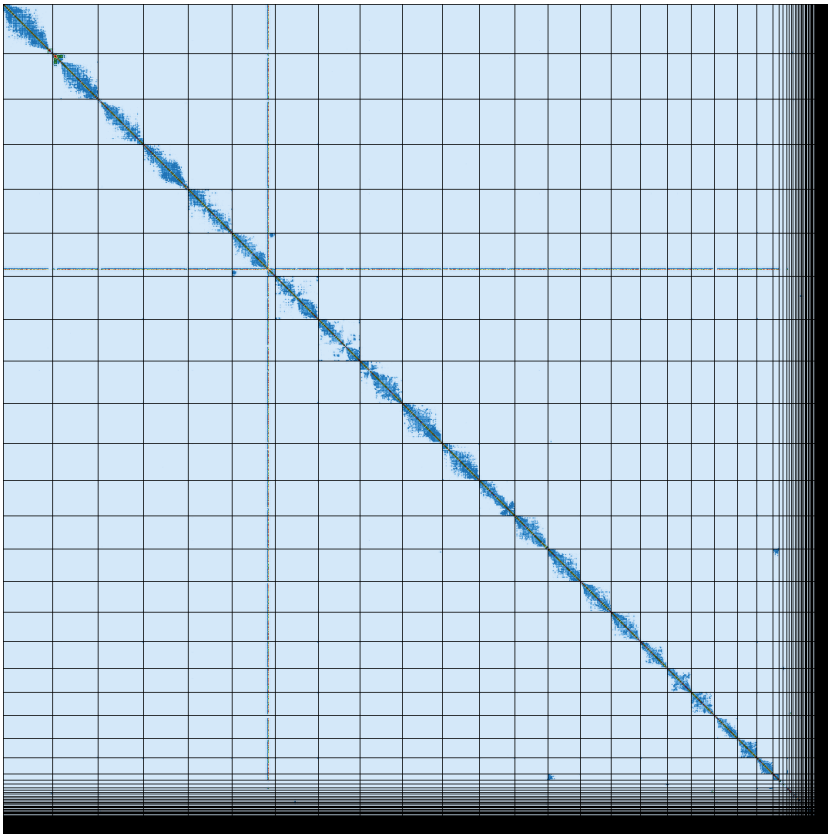
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	605,197,644	535,409,667
GC %	44.2	44.34
Gaps/Gbp	1,265.7	1,486.71
Total gap bp	76,600	88,200
Scaffolds	266	121
Scaffold N50	23,261,663	26,165,760
Scaffold L50	11	10
Scaffold L90	49	21
Contigs	1,032	917
Contig N50	1,376,608	1,418,811
Contig L50	123	104
Contig L90	487	427
QV	41.5517	41.7203
Kmer compl.	87.6806	85.9983
BUSCO sing.	87.0%	98.6%
BUSCO dupl.	9.9%	0.7%
BUSCO frag.	0.7%	0.1%
BUSCO miss.	2.4%	0.7%

Warning! BUSCO versions or lineage datasets are not the same across results:

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

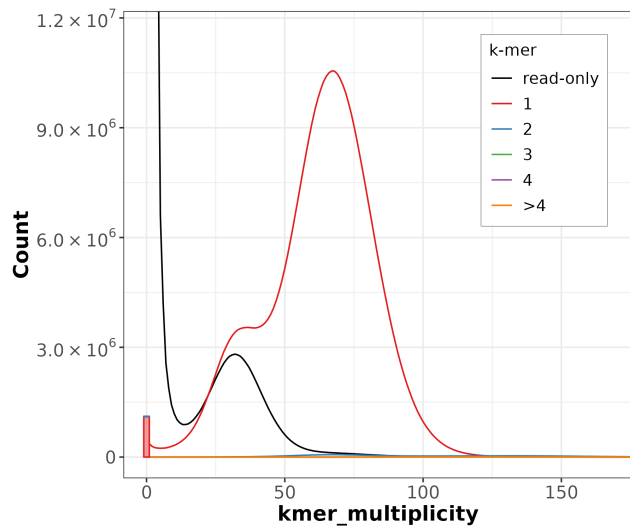
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

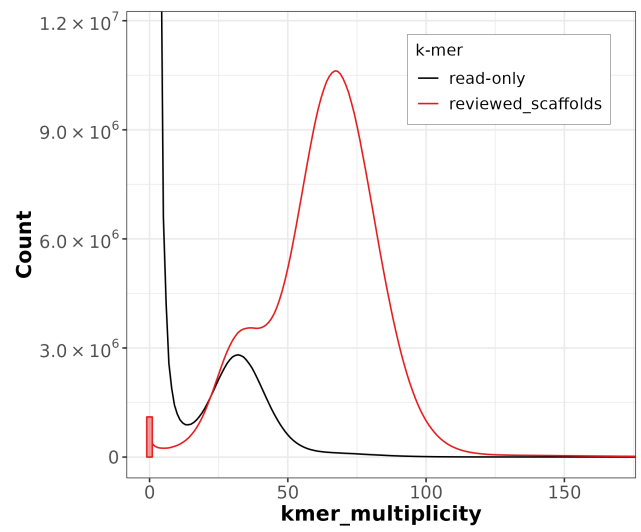


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

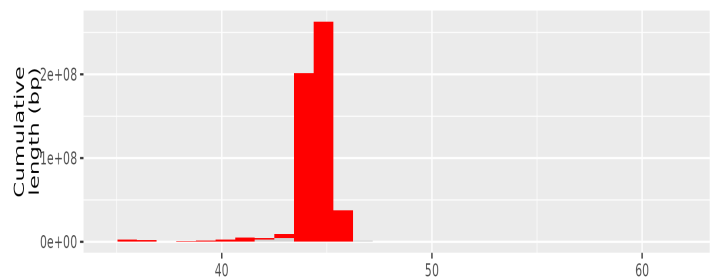


Distribution of k-mer counts per copy numbers found in asm

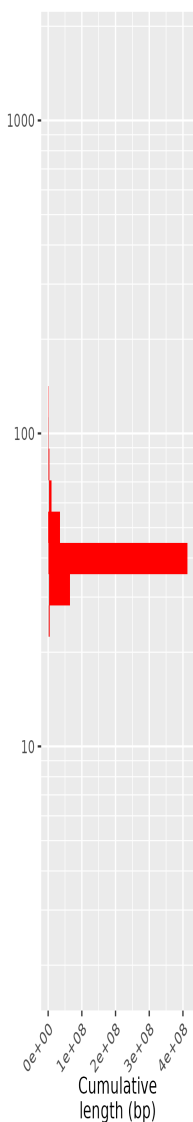
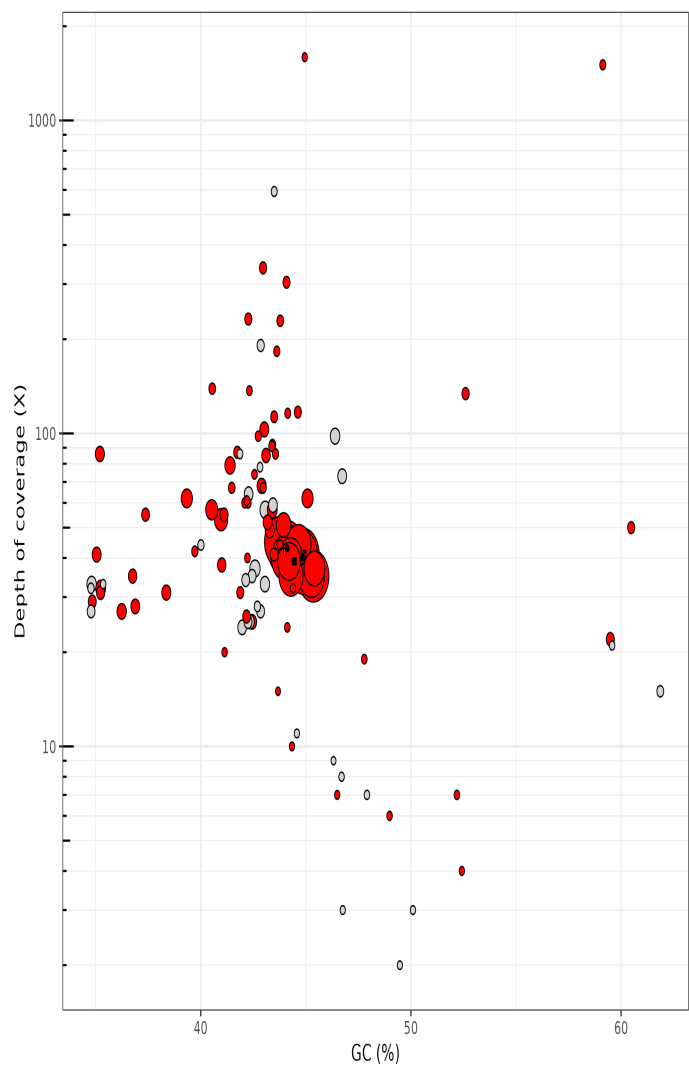


Distribution of k-mer counts coloured by their presence in reads/assemblies

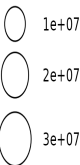
Post-curation contamination screening



TAPAs summary Graph



Length (bp)



superkingdom



Longest sequences (bp)

- fZeuPun1_1 - 31997471 (Eukaryota)
- ▲ fZeuPun1_2 - 29337204 (Eukaryota)
- fZeuPun1_3 - 29044322 (Eukaryota)
- + fZeuPun1_4 - 28794678 (Eukaryota)
- ▣ fZeuPun1_5 - 28328803 (Eukaryota)

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	130

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: Benjamin Istace

Affiliation: Genoscope

Date and time: 2025-10-13 00:20:31 CEST