

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	588211
ToLID	fAstPun1
Species	Astrapogon puncticulatus
Class	Actinopteri
Order	Kurtiiformes

Genome Traits	Expected	Observed
Haploid size (bp)	752,857,960	770,648,391
Haploid Number	21 (source: ancestor)	20
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q46

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 *Astrapogon puncticulatus* (fAstPun1) is based on 50X PacBio data and 198X 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, 2 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.086 Mb (with the largest being 0.047 Mb). Additionally, 134 regions totaling 9.298 Mb (with the largest being 0.582 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. During manual curation, 2 haplotypic regions were removed, totaling 0.17Mb (with the largest being 0.091Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

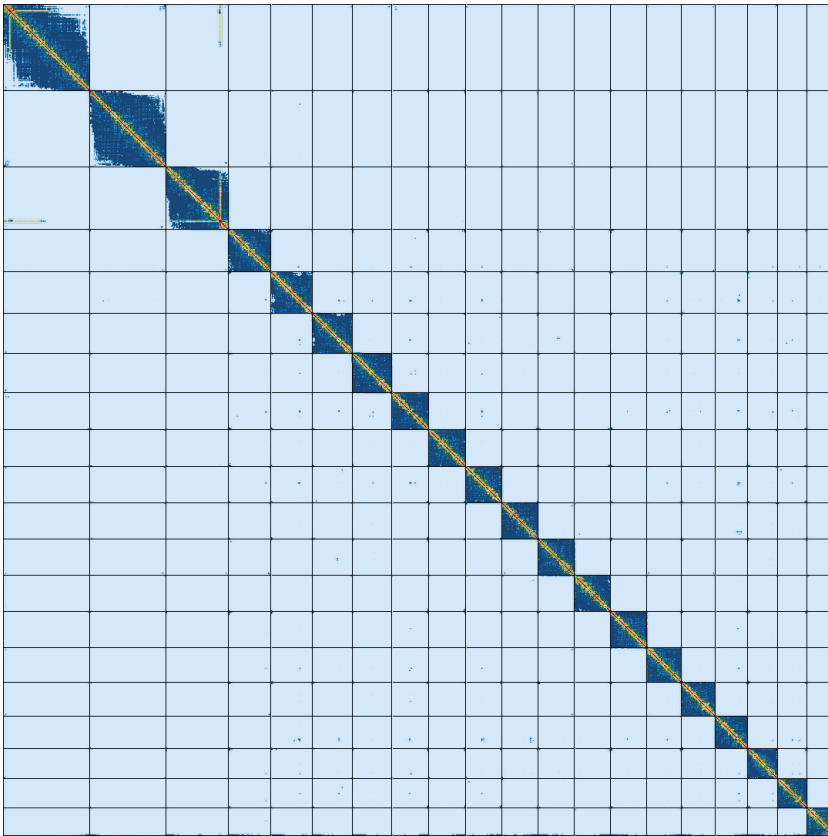
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	770,861,254	770,648,391
GC %	40.13	40.13
Gaps/Gbp	243.88	266.01
Total gap bp	20,400	25,700
Scaffolds	65	36
Scaffold N50	34,106,620	34,106,620
Scaffold L50	8	8
Scaffold L90	18	18
Contigs	250	241
Contig N50	21,736,283	21,736,283
Contig L50	15	15
Contig L90	48	48
QV	46.2349	46.2427
Kmer compl.	87.1921	87.182
BUSCO sing.	96.6%	99.2%
BUSCO dupl.	0.2%	0.2%
BUSCO frag.	0.8%	0.1%
BUSCO miss.	2.3%	0.5%

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)

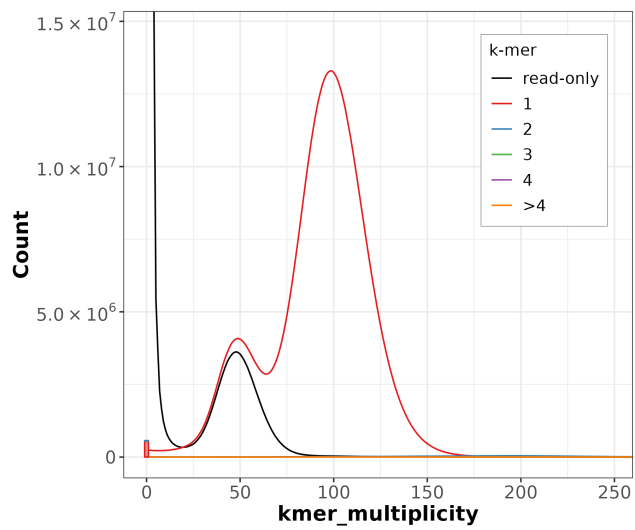
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

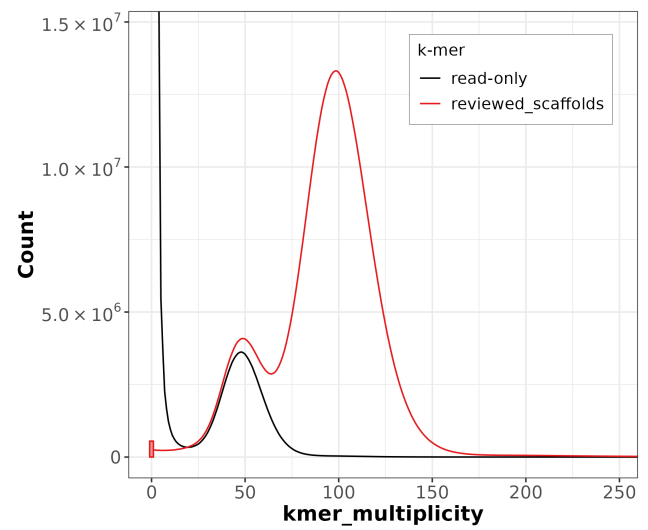


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

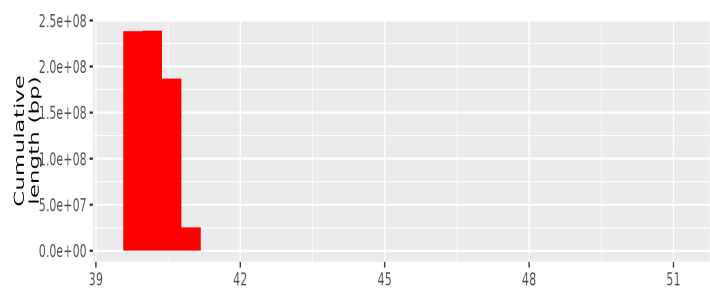


Distribution of k-mer counts per copy numbers found in asm

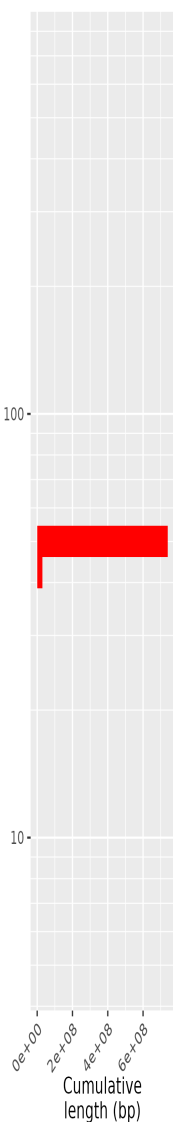
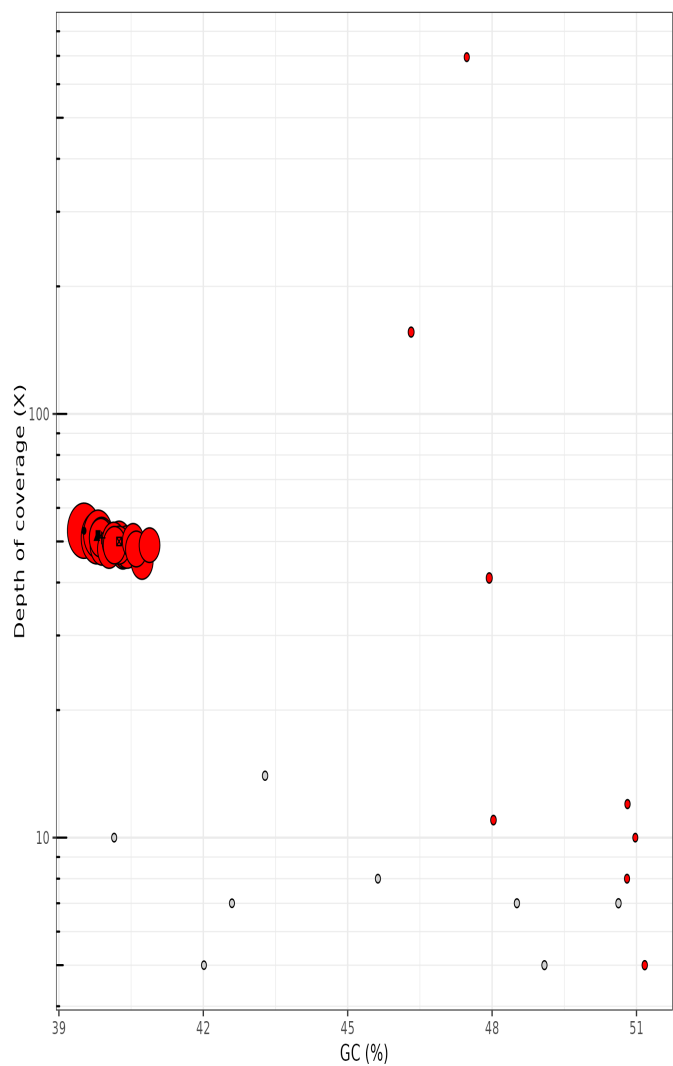


Distribution of k-mer counts coloured by their presence in reads/assemblies

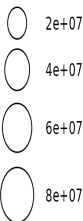
Post-curation contamination screening



TAPAs summary Graph



Length (bp)



superkingdom



Longest sequences (bp)

- fAstPun1_1 - 80242105 (Eukaryota)
- ▲ fAstPun1_2 - 70494256 (Eukaryota)
- fAstPun1_3 - 58310084 (Eukaryota)
- + fAstPun1_4 - 39134168 (Eukaryota)
- ▣ fAstPun1_5 - 38111908 (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	49	198

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: 2025-10-28 15:32:06 CET