

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	146147
ToLID	fUmbCan1
Species	Umbrina canariensis
Class	Actinopteri
Order	NA

Genome Traits	Expected	Observed
Haploid size (bp)	673,519,270	672,282,499
Haploid Number	24 (source: ancestor)	24
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q49

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

- . Kmer completeness value is less than 90 for collapsed

Curator notes

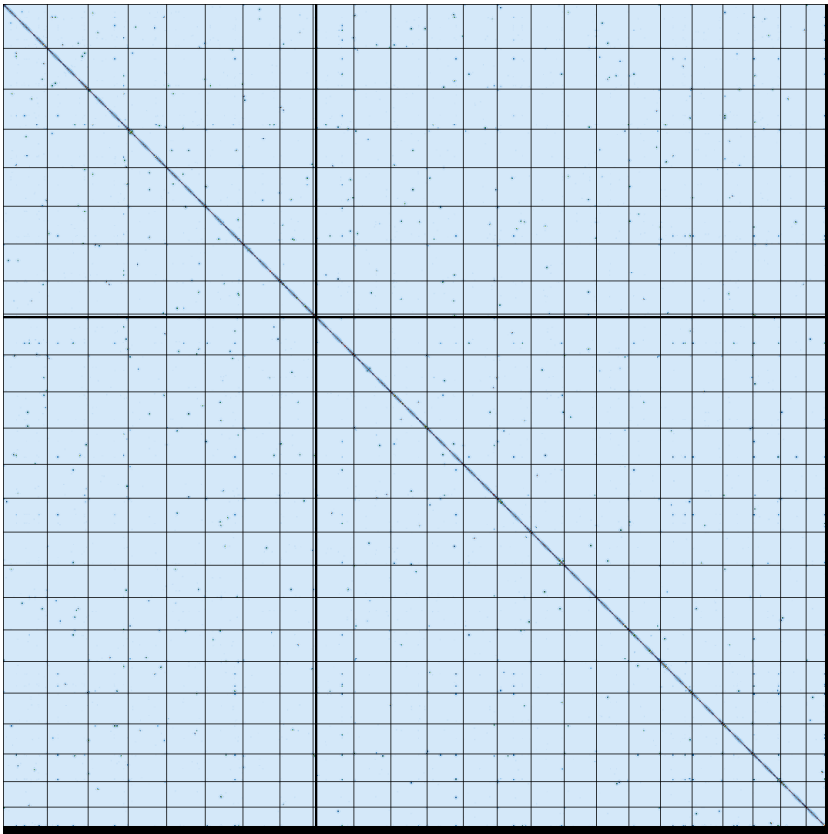
- . Interventions/Gb: 11
- . Contamination notes: ""
- . Other observations: "The assembly of *Umbrina canariensis* (fUmbCan1) is based on 31X PacBio data and 145X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: 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, 1 contigs was identified as contaminants (bacterial, archaeal, or viral), totaling 0.032 Mb (with the largest being 0.032 Mb). Additionally, 86 regions totaling 11.971 Mb (with the largest being 0.509 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using ptGAUL. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 1 haplotypic region was removed, totaling 0.205Mb. The telomeric pattern TTAGGG was identified with TelFinder and used to generate the Pretext Telomeres tracks. 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	672,844,752	672,282,499
GC %	41.06	41.05
Gaps/Gbp	47.56	49.09
Total gap bp	3,200	3,500
Scaffolds	138	136
Scaffold N50	29,789,594	29,067,398
Scaffold L50	10	11
Scaffold L90	21	22
Contigs	170	169
Contig N50	24,884,000	24,884,000
Contig L50	13	13
Contig L90	27	27
QV	49.2964	49.3343
Kmer compl.	89.6611	89.6552
BUSCO sing.	98.4%	98.4%
BUSCO dupl.	0.3%	0.3%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	1.2%	1.2%

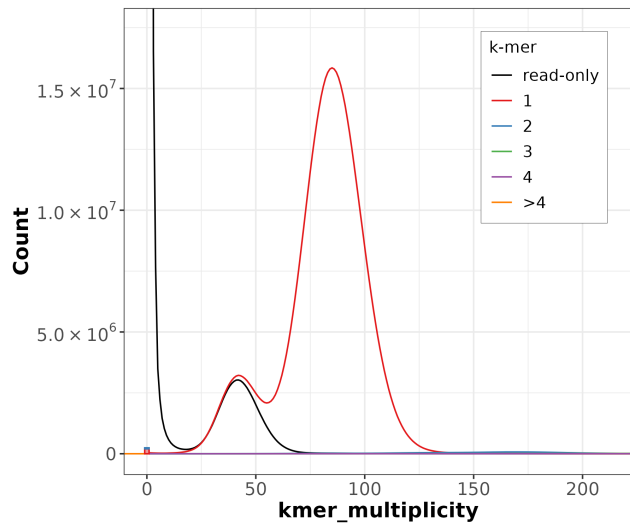
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12.2 (genomes:79, BUSCOs:9048)

HiC contact map of curated assembly

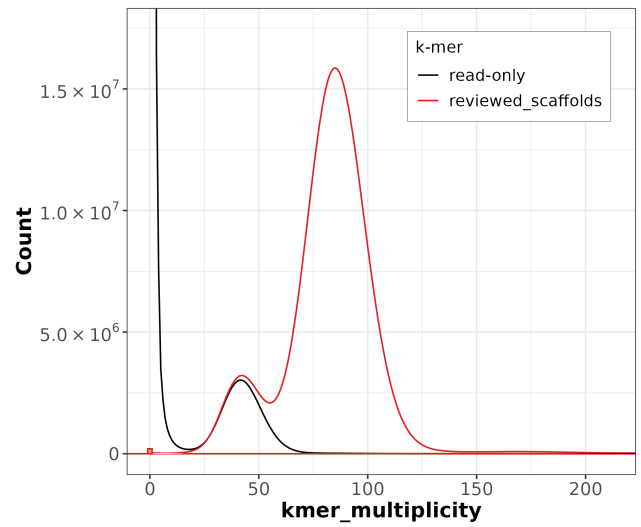


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

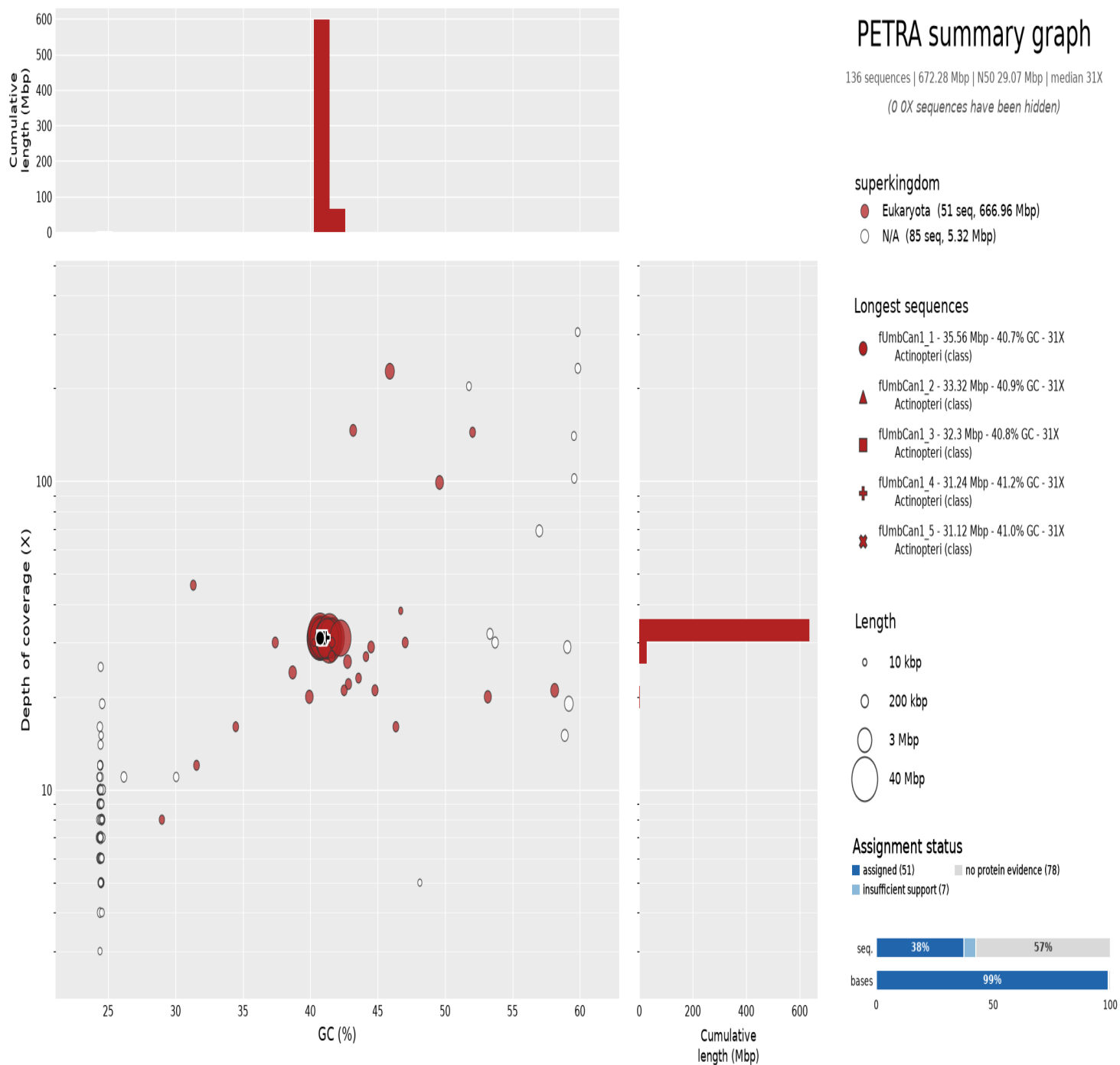


Distribution of k-mer counts per copy numbers found in asm



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	31	145

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: Eleanore Lacoste

Affiliation: Genoscope

Date and time: 2026-09-09 20:14:54 CEST