ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	173309
ToLID	**fCarCry1**
Species	Caranx crysos
Class	Actinopteri
Order	Carangiformes

Genome Traits	Expected	Observed
Haploid size (bp)	565,802,719	578,101,140
Haploid Number	21 (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.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

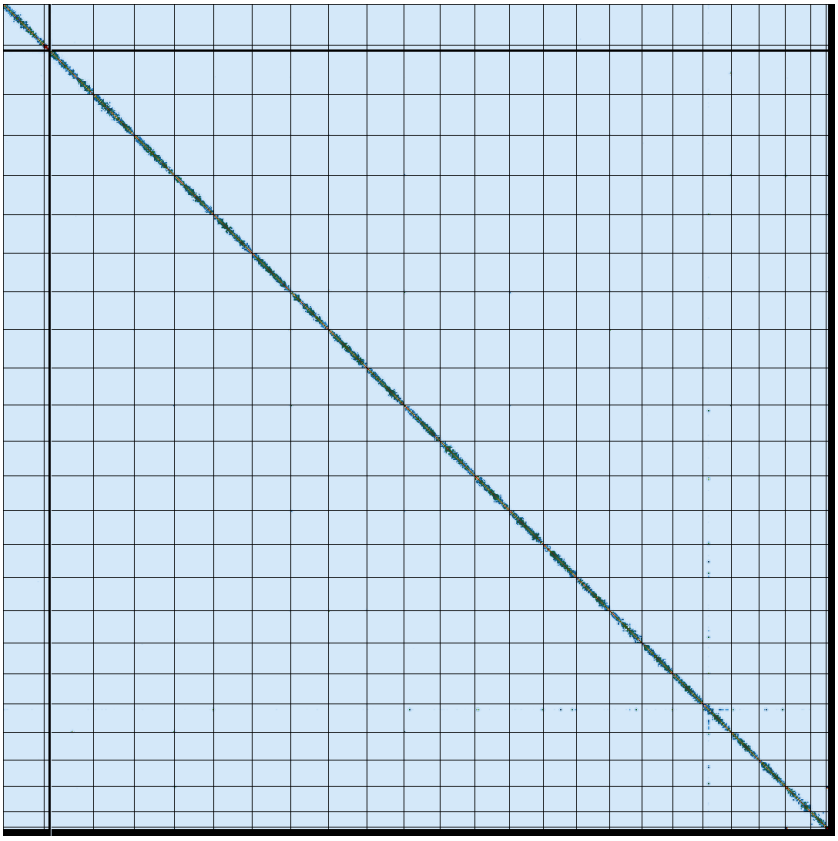
. Interventions/Gb: 12
. Contamination notes: ""
. Other observations: "The assembly of Caranx crysos (fCarCry1) is based on 46X PacBio data and 178X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 21 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.055 Mb (with the largest being 0.129 Mb). Additionally, 158 regions totaling 7.013 Mb (with the largest being 0.332 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. 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	578,083,760	578,101,140
GC %	42.87	42.87
Gaps/Gbp	345.97	351.15
Total gap bp	20,000	20,800
Scaffolds	80	78
Scaffold N50	24,970,639	25,082,788
Scaffold L50	11	11
Scaffold L90	21	21
Contigs	280	281
Contig N50	20,001,237	20,001,237
Contig L50	12	12
Contig L90	32	32
QV	46.0579	46.058
Kmer compl.	86.9777	86.9804
BUSCO sing.	99.2%	99.2%
BUSCO dupl.	0.2%	0.2%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.5%	0.5%

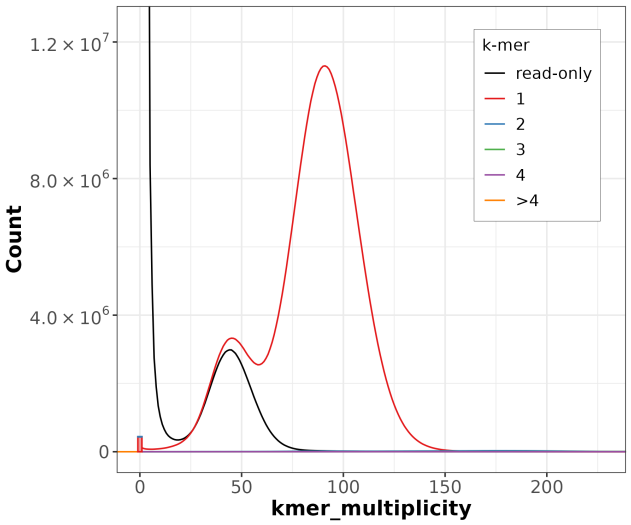
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

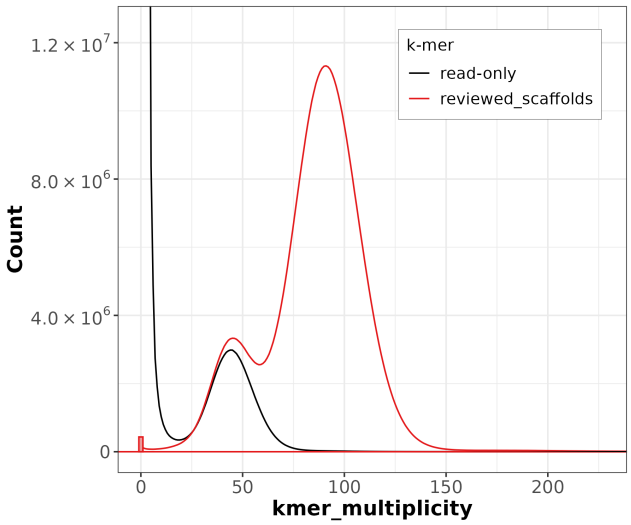


collapsed [LINK]

K-mer spectra of curated assembly

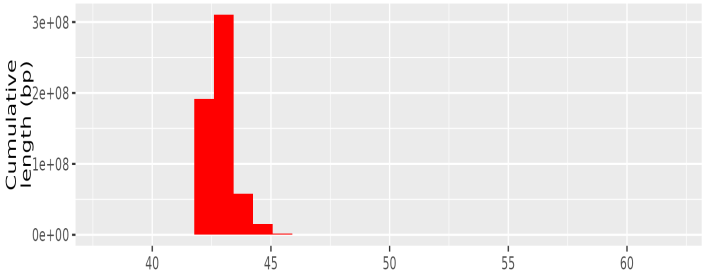


Distribution of k-mer counts per copy numbers found in asm

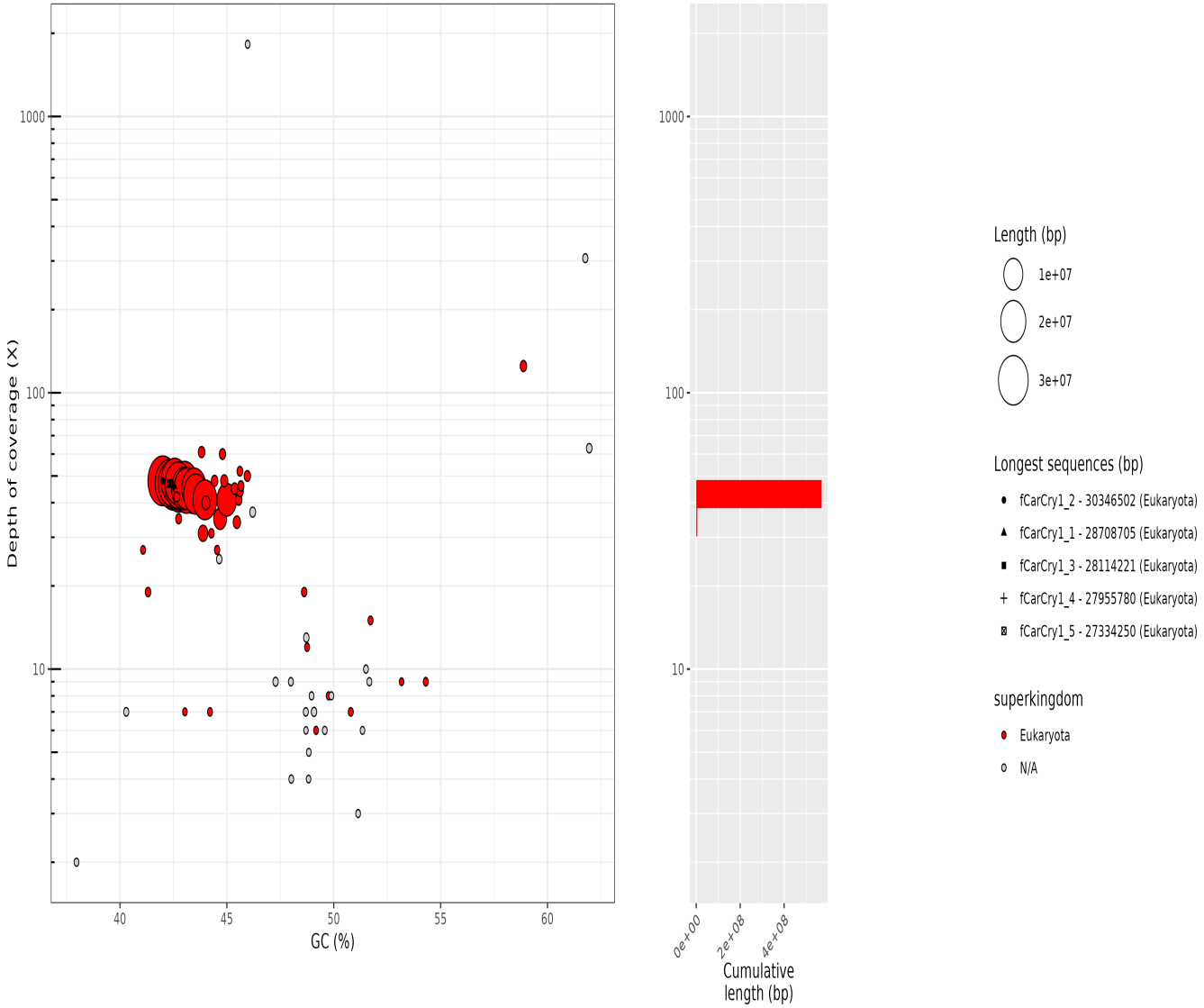


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	45	178

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: Mohamed Meftah Saoues
Affiliation: Genoscope

Date and time: 2026-05-17 05:44:54 CEST