ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	658148
ToLID	**fCorVen1**
Species	Coryphopterus venezuelae
Class	Actinopteri
Order	Gobiiformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,098,288,990	1,143,957,084
Haploid Number	21 (source: ancestor)	21
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q50

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

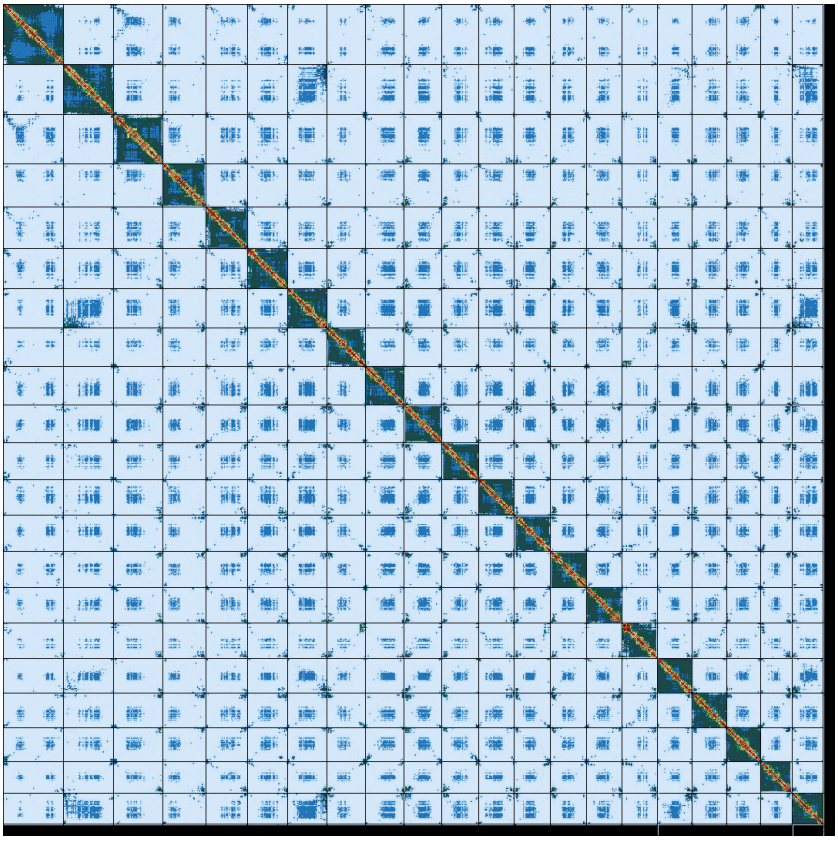
. Interventions/Gb: 17
. Contamination notes: ""
. Other observations: "The assembly of Coryphopterus venezuelae (fCorVen1) is based on 56X PacBio data and 176X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 3 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.059 Mb (with the largest being 0.022 Mb). Additionally, 204 regions totaling 50.783 Mb (with the largest being 2.449 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 9 haplotypic regions and 229 contaminant sequences were removed, totaling 13.31Mb and 4.69Mb, respectively (with the largest being 4.37Mb and 0.15Mb). 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	1,161,956,739	1,143,957,084
GC %	38.98	38.96
Gaps/Gbp	87.78	81.3
Total gap bp	10,200	10,000
Scaffolds	339	95
Scaffold N50	50,791,912	52,110,546
Scaffold L50	10	10
Scaffold L90	20	19
Contigs	441	188
Contig N50	44,879,000	44,879,000
Contig L50	11	11
Contig L90	28	25
QV	44.0844	50.4595
Kmer compl.	75.4278	74.89
BUSCO sing.	97.6%	97.9%
BUSCO dupl.	1.0%	0.6%
BUSCO frag.	0.3%	0.3%
BUSCO miss.	1.2%	1.2%

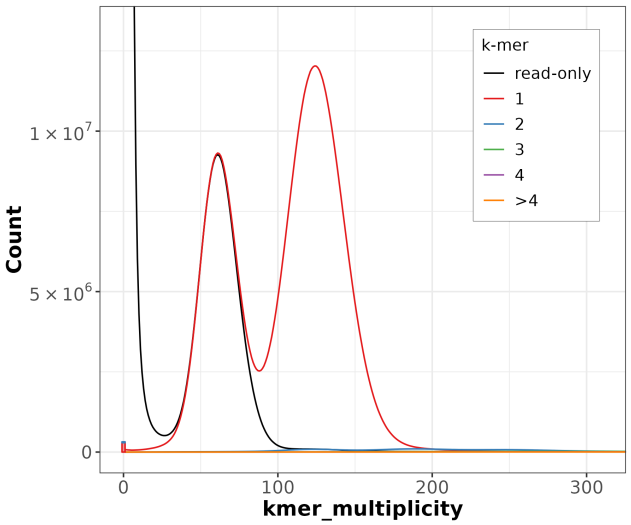
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

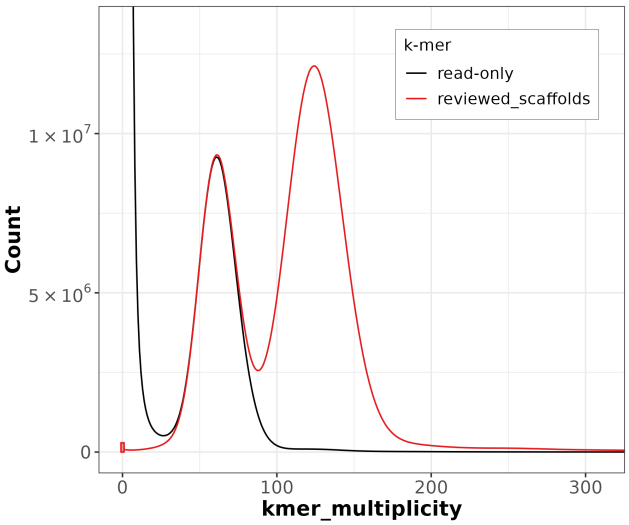


collapsed [LINK]

K-mer spectra of curated assembly

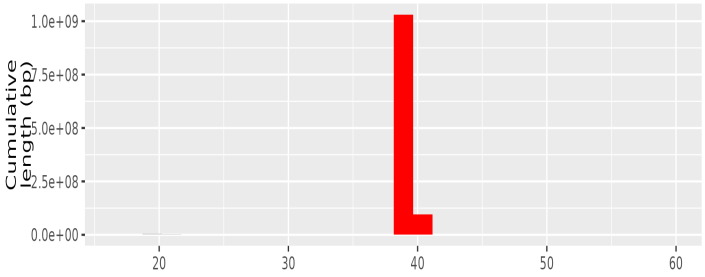


Distribution of k-mer counts per copy numbers found in asm

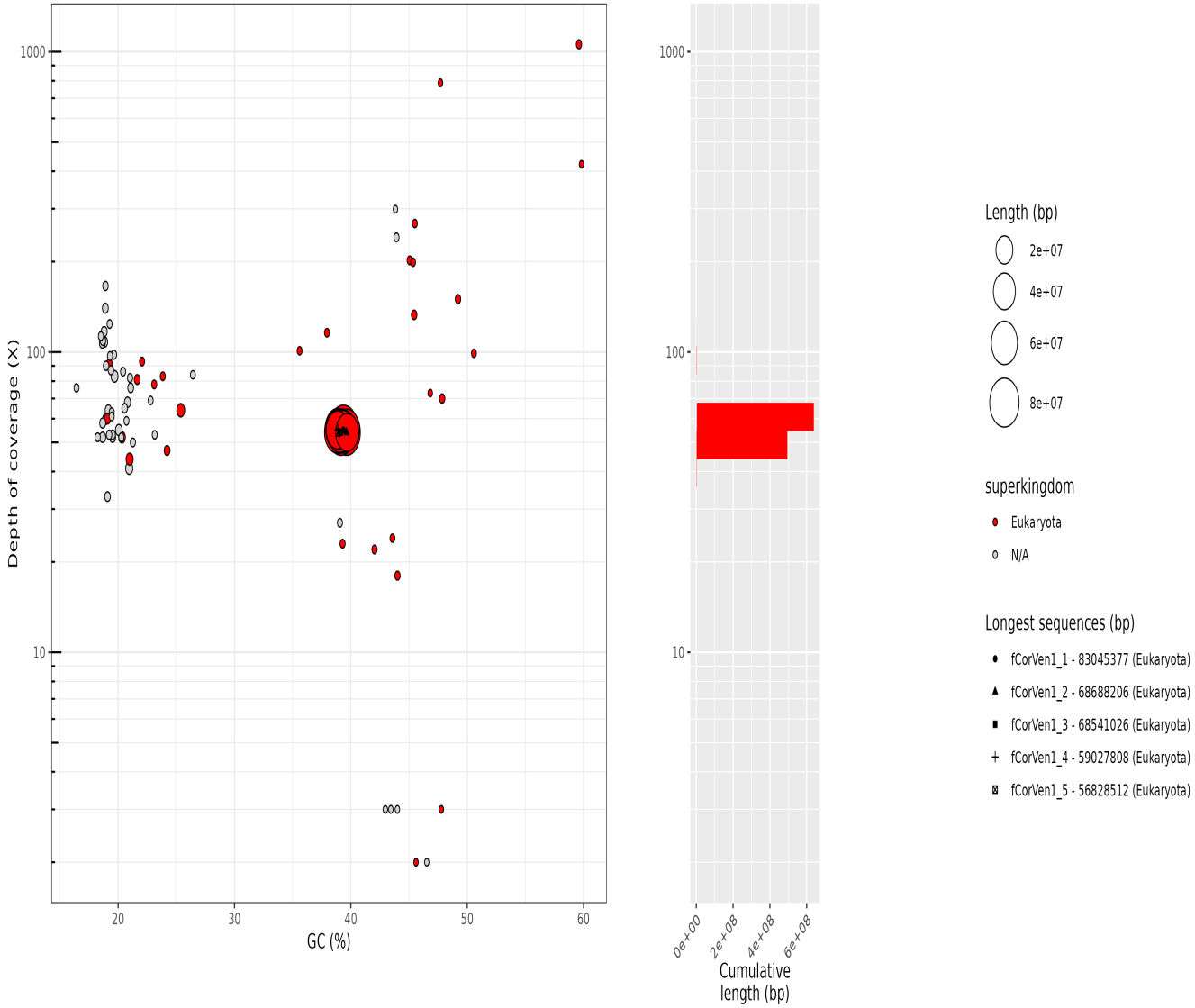


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	56	176

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: 2025-11-09 14:35:56 CET