

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	349663
ToLID	fCheObs1
Species	Chelidonichthys obscurus
Class	Actinopteri
Order	Perciformes

Genome Traits	Expected	Observed
Haploid size (bp)	643,073,843	669,486,466
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.Q48

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

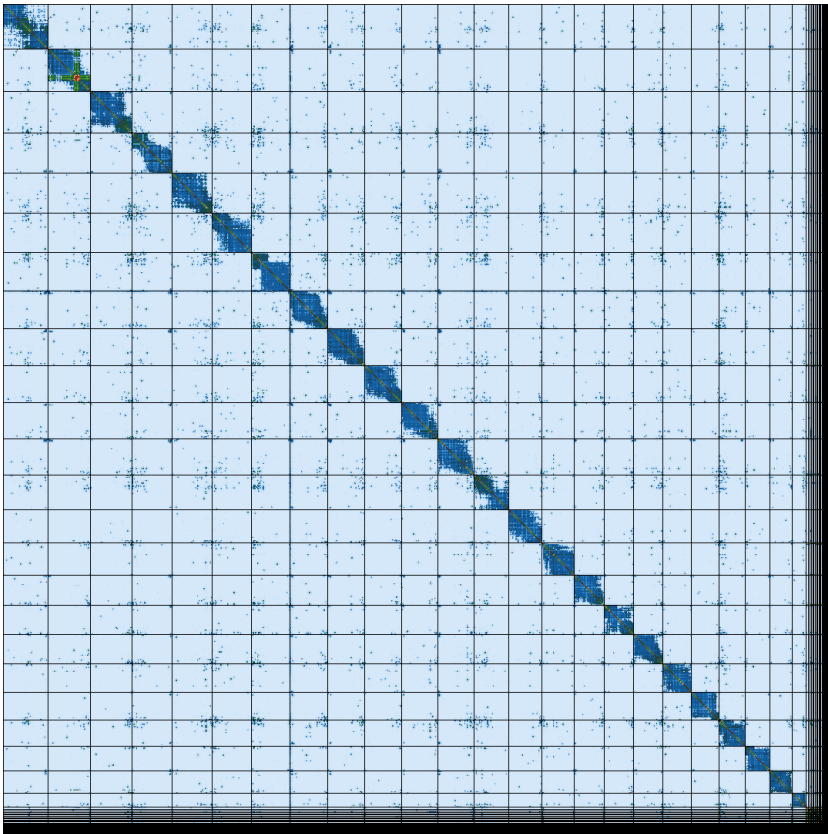
- . Interventions/Gb: 6
- . Contamination notes: ""
- . Other observations: "The assembly of Chelidonichthys obscurus (fCheObs1) is based on 42X PacBio data and 167X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: 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. No contigs were found to be contaminants (bacterial, archaeal, or viral). Additionally, 107 regions totaling 20.67 Mb (with the largest being 2.552 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 1 haplotypic region and 3 contaminant sequences were removed, totaling 1.135Mb and 0.171Mb, respectively (with the largest being 1.135Mb and 0.109Mb). 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	670,775,087	669,486,466
GC %	42.71	42.71
Gaps/Gbp	26.83	29.87
Total gap bp	1,800	2,300
Scaffolds	107	102
Scaffold N50	29,090,559	29,112,875
Scaffold L50	11	11
Scaffold L90	23	22
Contigs	125	122
Contig N50	24,349,045	24,349,045
Contig L50	12	12
Contig L90	26	26
QV	48.1238	48.1209
Kmer compl.	81.4849	81.475
BUSCO sing.	99.3%	99.4%
BUSCO dupl.	0.3%	0.2%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.4%	0.4%

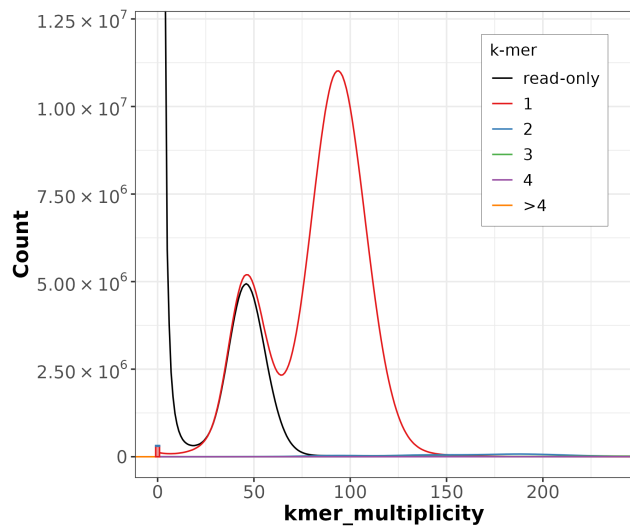
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

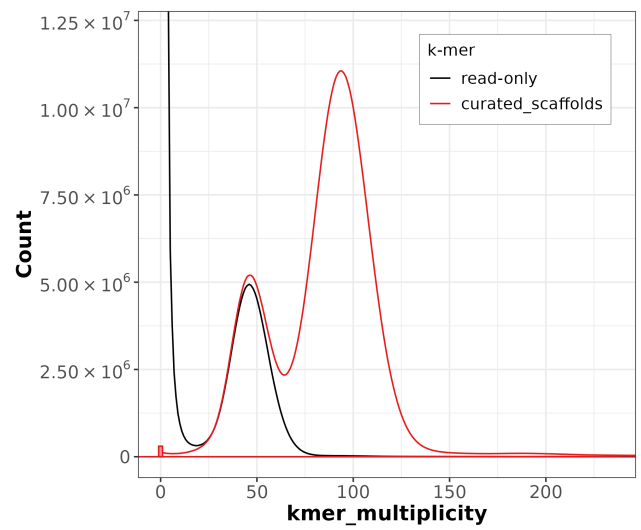


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

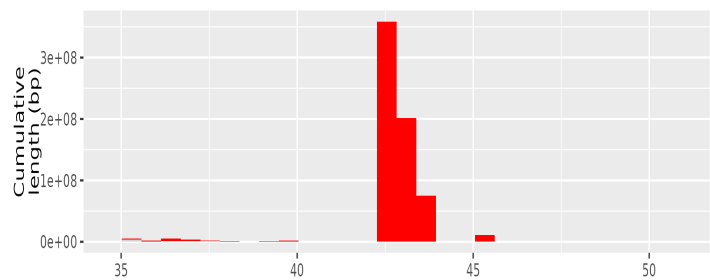


Distribution of k-mer counts per copy numbers found in asm

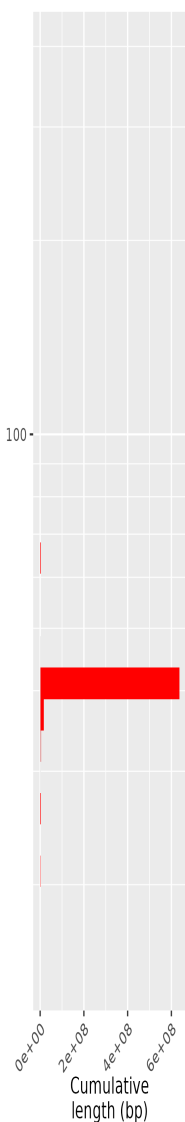
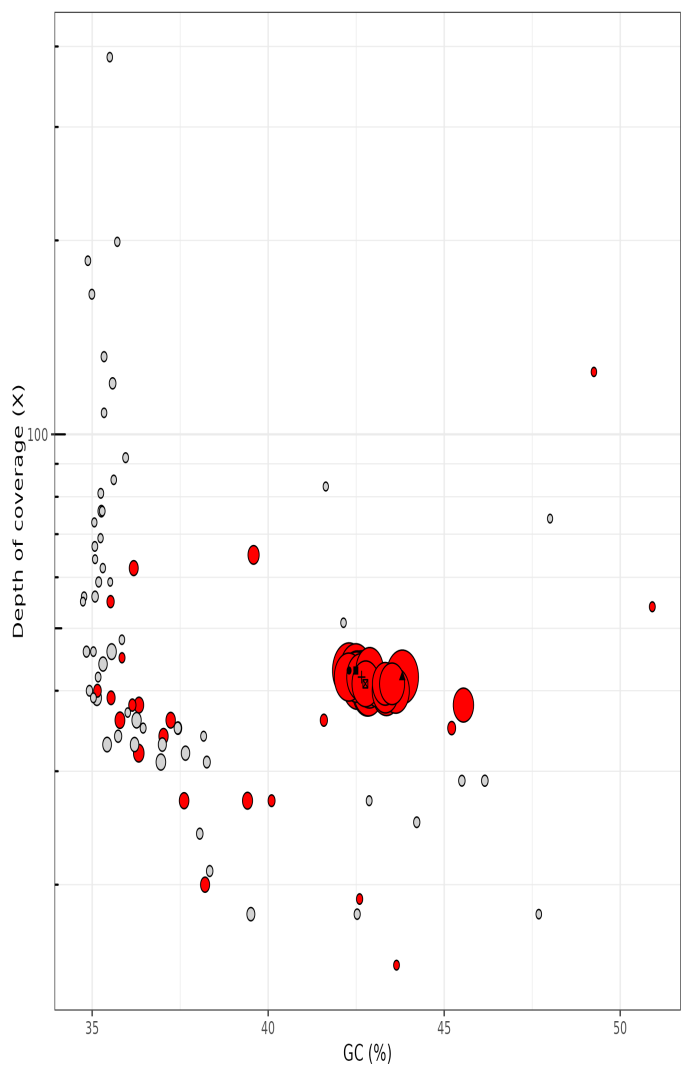


Distribution of k-mer counts coloured by their presence in reads/assemblies

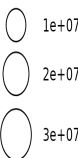
Post-curation contamination screening



TAPAs summary Graph



Length (bp)



Longest sequences (bp)

- fCheObs1_1 - 36100945 (Eukaryota)
- fCheObs1_2 - 34259002 (Eukaryota)
- fCheObs1_3 - 33504915 (Eukaryota)
- fCheObs1_4 - 32231301 (Eukaryota)
- fCheObs1_5 - 32124231 (Eukaryota)

superkingdom

- Eukaryota
- N/A

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	42	167

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: Emilie Teodori

Affiliation: Genoscope

Date and time: 2026-08-06 08:53:44 CEST