ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	596122
ToLID	**fChrZer1**
Species	Chromogobius zebratus
Class	Actinopteri
Order	Gobiiformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,125,006,134	1,161,606,913
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.Q49

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
. Not 90% of assembly in chromosomes for collapsed

Curator notes

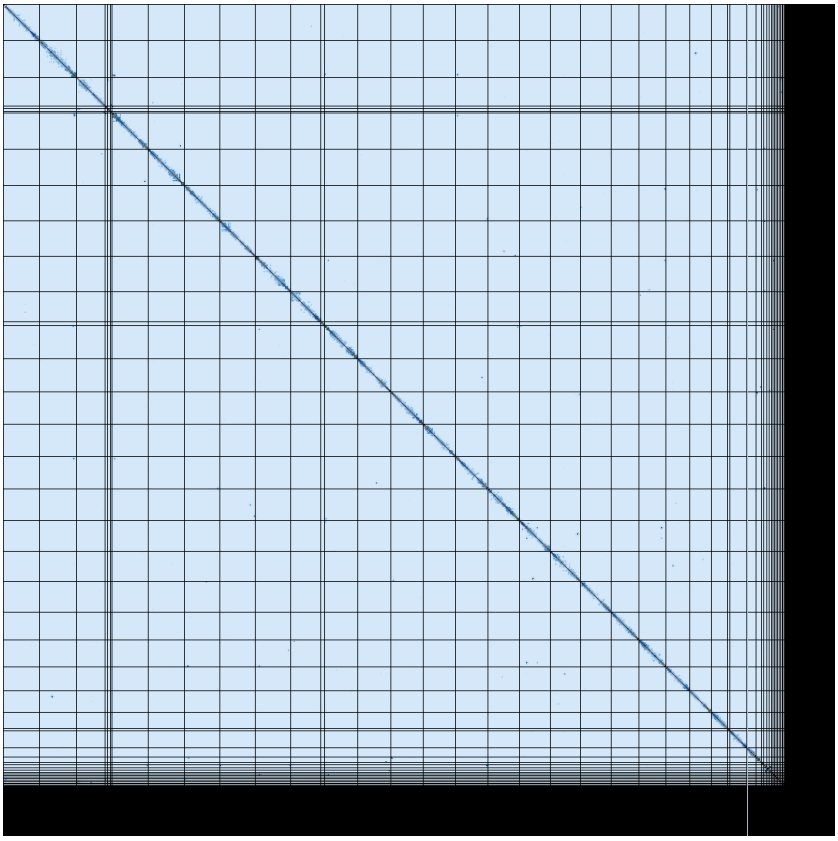
. Interventions/Gb: 9
. Contamination notes: ""
. Other observations: "The assembly of Chromogobius zebratus (fChrZer1) is based on 58X PacBio data and 124X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 41 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 2.114 Mb (with the largest being 0.438 Mb). Additionally, 185 regions totaling 70.145 Mb (with the largest being 4.45 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 2 haplotypic regions were removed, totaling 2.597Mb (with the largest being 1.307Mb). 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	1,175,444,532	1,161,606,913
GC %	41.13	41.13
Gaps/Gbp	116.55	115.36
Total gap bp	13,700	13,800
Scaffolds	409	399
Scaffold N50	44,593,100	44,593,100
Scaffold L50	13	13
Scaffold L90	29	28
Contigs	546	533
Contig N50	39,076,000	39,076,000
Contig L50	14	14
Contig L90	55	53
QV	49.7761	49.7806
Kmer compl.	80.4652	80.0506
BUSCO sing.	96.8%	97.3%
BUSCO dupl.	1.5%	1.0%
BUSCO frag.	0.3%	0.3%
BUSCO miss.	1.4%	1.4%

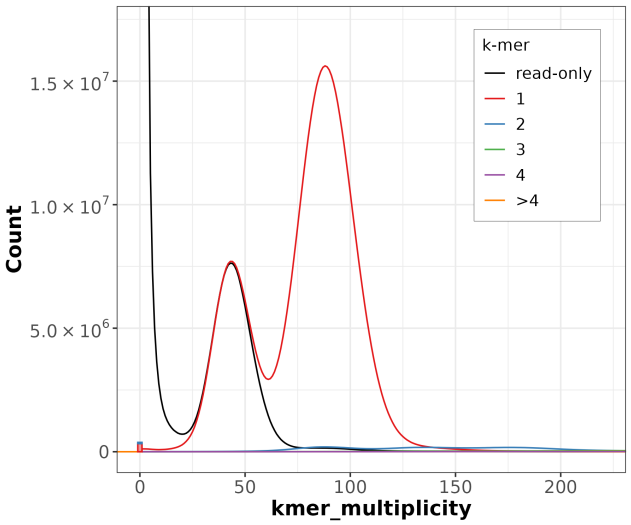
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

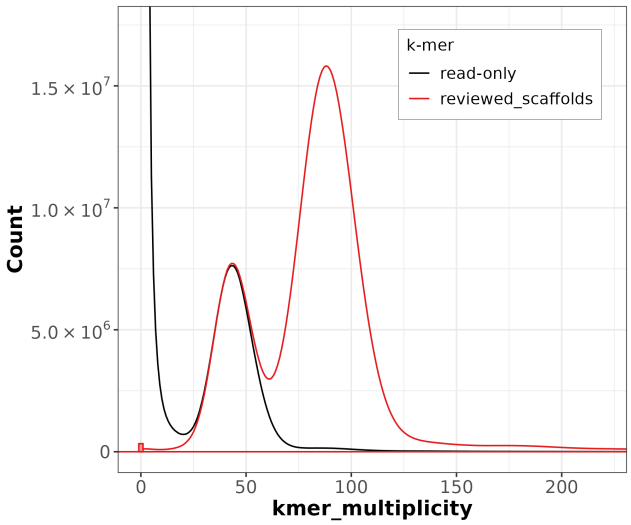


collapsed [LINK]

K-mer spectra of curated assembly

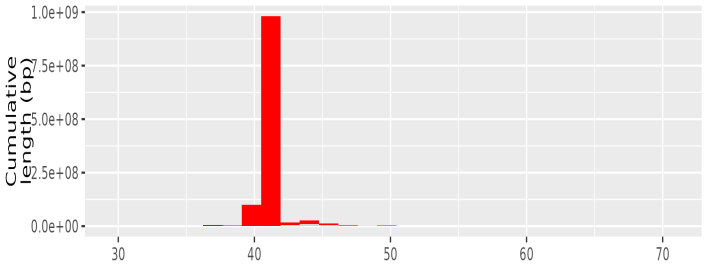


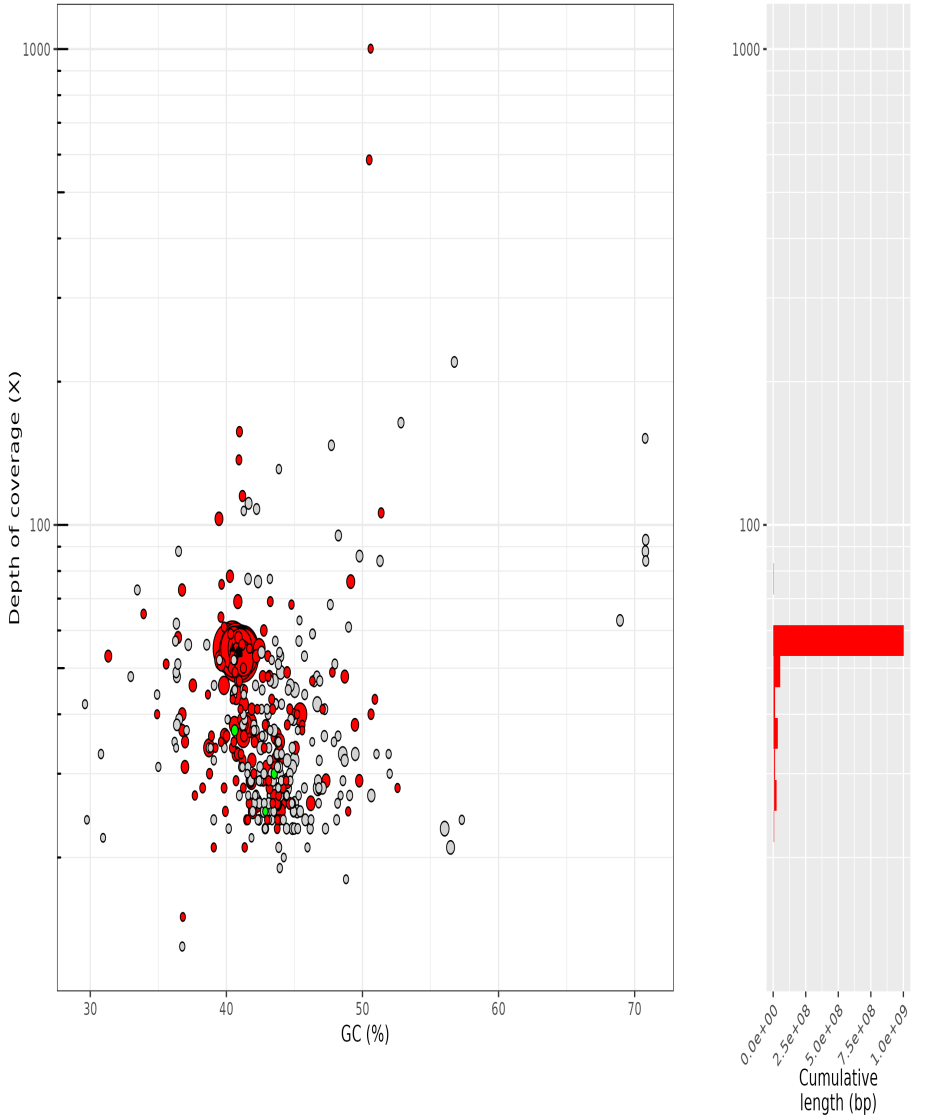
Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening





TAPAs summary Graph

(3 0X contigs have been hidden)

Longest sequences (bp)

- fChrZer1_1 - 51523564 (Eukaryota)
- fChrZer1_2 - 50692000 (Eukaryota)
- fChrZer1_4 - 50426800 (Eukaryota)
- fChrZer1_5 - 49954703 (Eukaryota)
- fChrZer1_6 - 49679489 (Eukaryota)

superkingdom

- Bacteria
- Eukaryota
- N/A

Length (bp)

- 1e+07
- 2e+07
- 3e+07
- 4e+07
- 5e+07

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	58	125

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: 2026-04-27 15:57:31 CEST