ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1365678
ToLID	**fCorLie1**
Species	Corcyrogobius liechtensteini
Class	Actinopteri
Order	Gobiiformes

Genome Traits	Expected	Observed
Haploid size (bp)	955,782,275	999,123,557
Haploid Number	24 (source: ancestor)	23
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

Curator notes

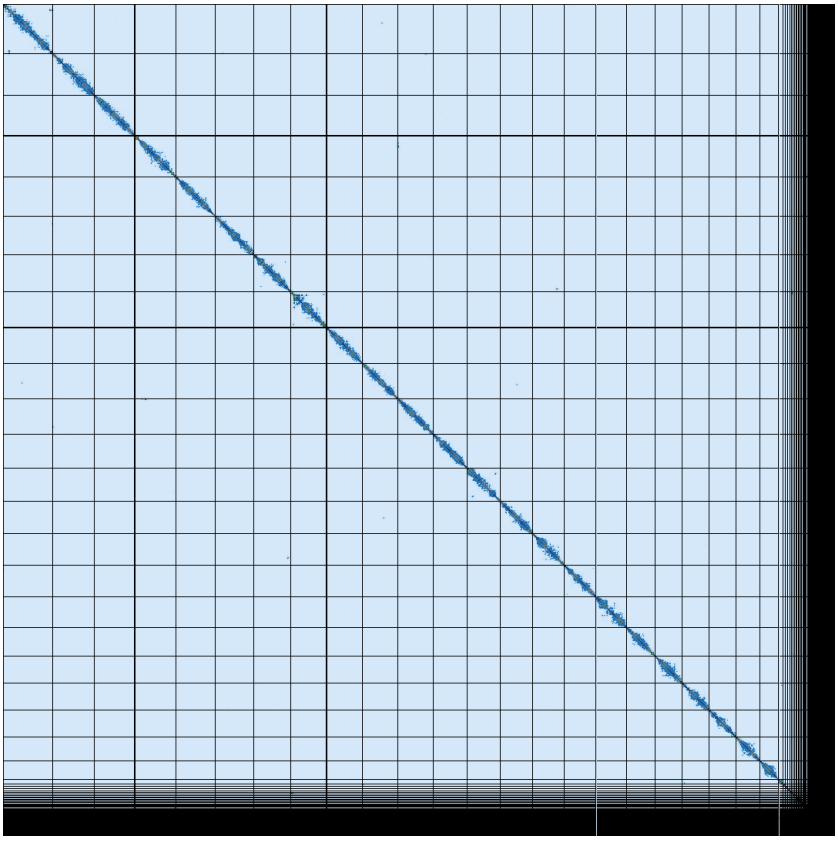
. Interventions/Gb: 24
. Contamination notes: ""
. Other observations: "The assembly of Corcyrogobius liechtensteini (fCorLie1) is based on 62X ONT data and 101X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial ONT 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, 85 regions totaling 32.474 Mb (with the largest being 7.916 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, 15 haplotypic regions were removed, totaling 7.281Mb (with the largest being 2.612Mb). 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	1,007,169,522	999,123,557
GC %	40.8	40.79
Gaps/Gbp	42.69	55.05
Total gap bp	4,300	7,000
Scaffolds	257	213
Scaffold N50	38,673,000	42,357,318
Scaffold L50	12	11
Scaffold L90	24	22
Contigs	300	268
Contig N50	32,099,000	31,911,770
Contig L50	14	14
Contig L90	42	40
QV	49.8385	49.8885
Kmer compl.	84.5828	84.4923
BUSCO sing.	97.1%	97.8%
BUSCO dupl.	1.3%	0.6%
BUSCO frag.	0.3%	0.3%
BUSCO miss.	1.2%	1.2%

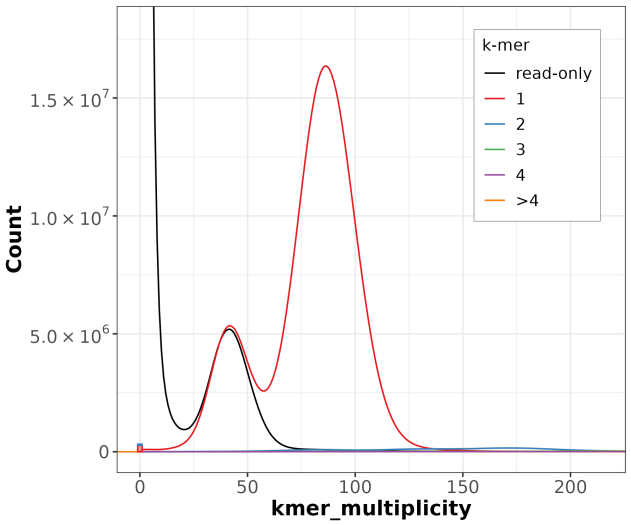
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

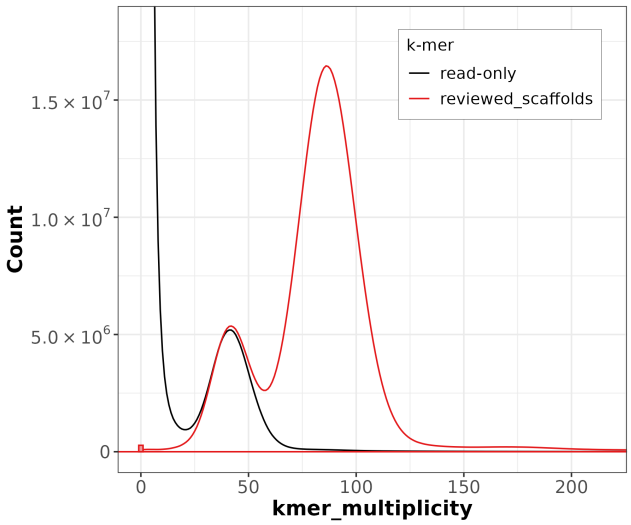


collapsed [LINK]

K-mer spectra of curated assembly

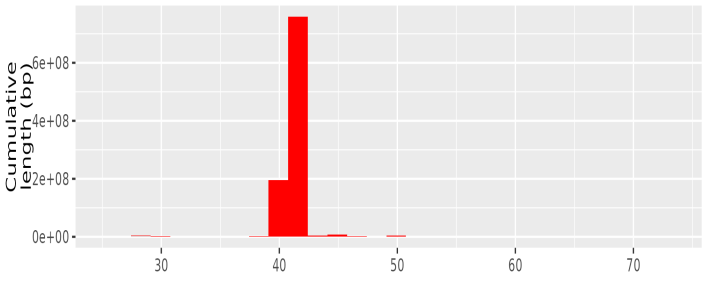


Distribution of k-mer counts per copy numbers found in asm



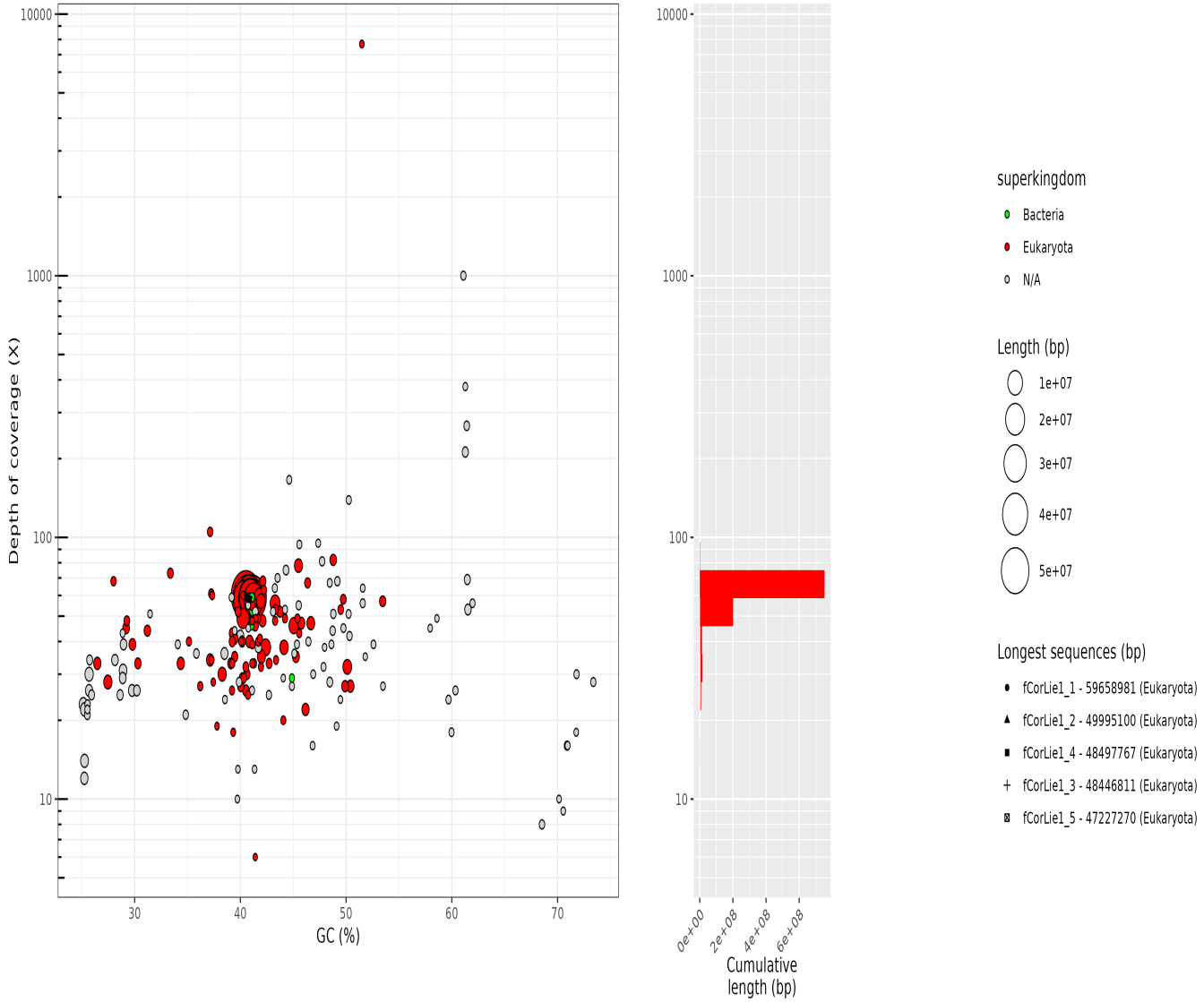
Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph

(5 0X contigs have been hidden)



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	62	102

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: 2026-03-28 12:50:04 CET