ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	152538
ToLID	**fCorGal2**
Species	Coryphoblennius galerita
Class	Actinopteri
Order	Blenniiformes

Genome Traits	Expected	Observed
Haploid size (bp)	646,306,855	775,345,180
Haploid Number	19 (source: ancestor)	24
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q59

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

. Observed Haploid Number is different from Expected

Curator notes

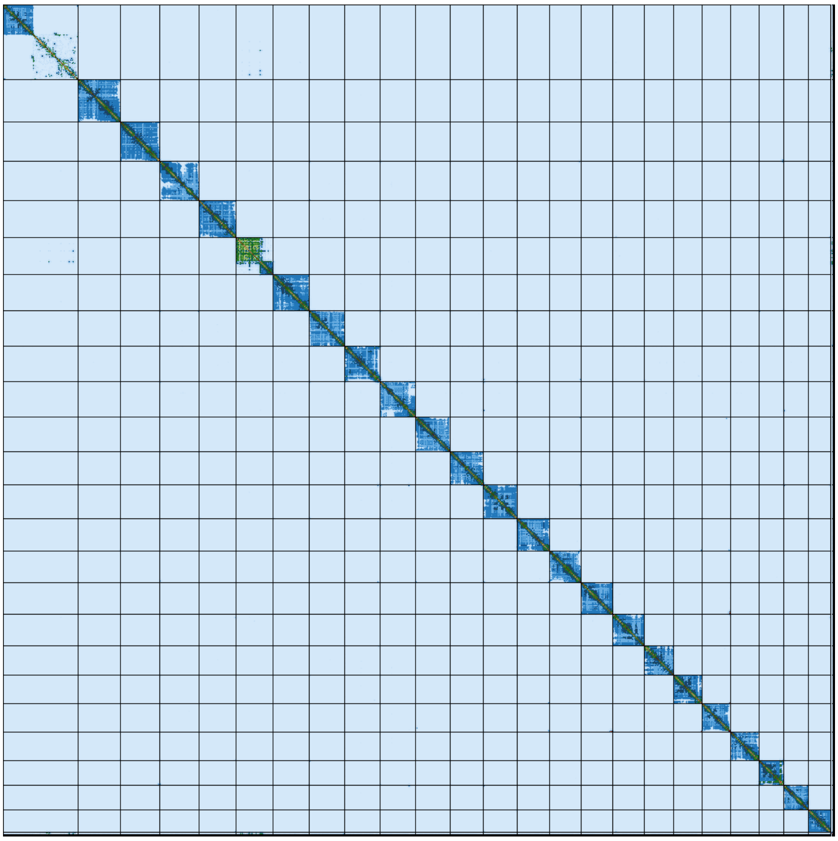
. Interventions/Gb: 46
. Contamination notes: ""
. Other observations: "The assembly of Coryphoblennius galerita (fCorGal2) is based on 42X PacBio data and Dovetail Omni-C Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included the following steps: 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, no contigs were identified as contaminants (bacterial, archaeal, or viral). Additionally, 289 regions totaling 28,8 Mb (with the largest being 0.993 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, 1 haplotypic regions was removed,totaling 0.165 Mb (with the largest being 0.165 Mb). The organization of the fCorGal2_1 and fCorGal2_6 chromosomes are uncertain. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	775,548,726	775,345,180
GC %	42.7	42.7
Gaps/Gbp	569.92	622.95
Total gap bp	44,200	52,600
Scaffolds	118	77
Scaffold N50	30,396,407	32,081,767
Scaffold L50	12	11
Scaffold L90	24	21
Contigs	560	560
Contig N50	3,971,000	3,971,000
Contig L50	55	55
Contig L90	219	220
QV	44.6523	59.3347
Kmer compl.	88.4709	88.6369
BUSCO sing.	97.1%	
BUSCO dupl.	0.8%	
BUSCO frag.	0.7%	
BUSCO miss.	1.4%	

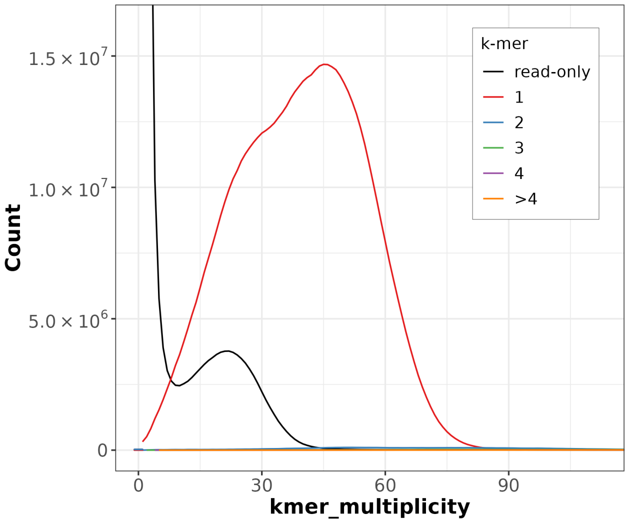
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb10 (genomes:26, BUSCOs:3640)

HiC contact map of curated assembly

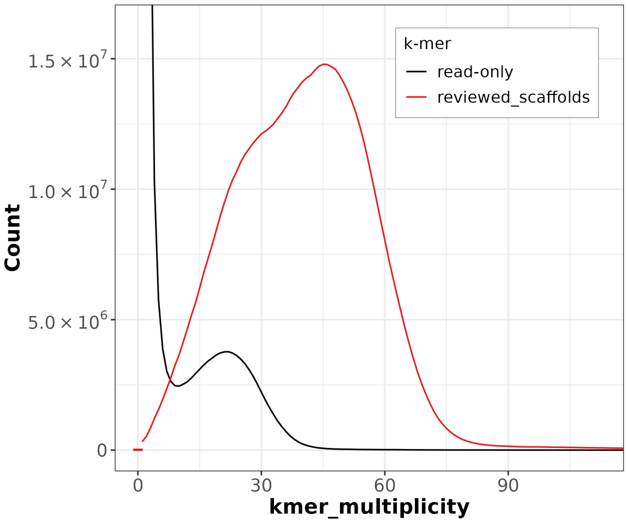


collapsed [LINK]

K-mer spectra of curated assembly

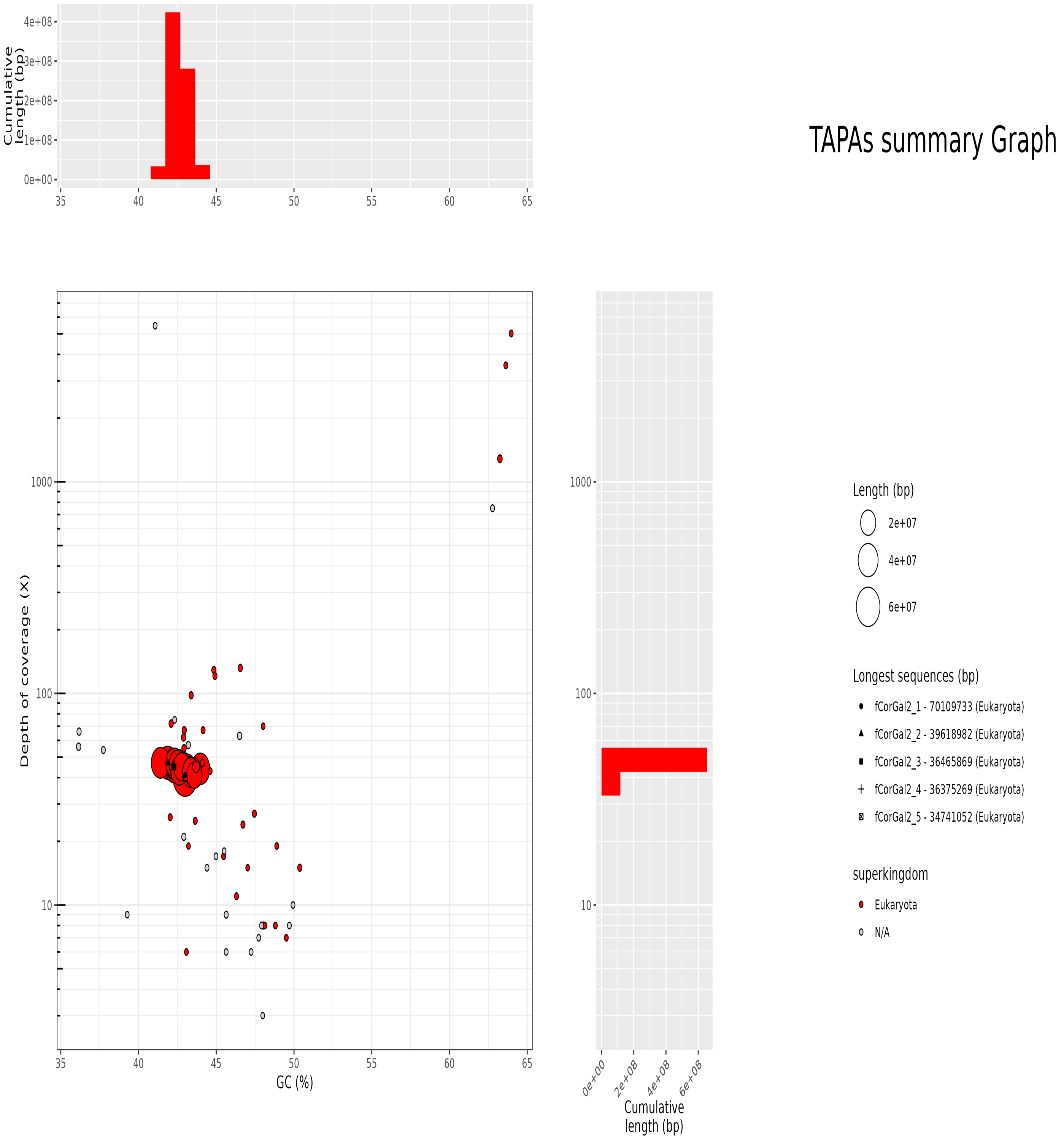


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	PACBIO Hifi	Omnic
Coverage	46	181

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: Caroline Menguy
Affiliation: Genoscope

Date and time: 2025-03-25 09:59:03 CET