

ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	126330
ToLID	fCorJull
Species	Coris julis
Class	Actinopteri
Order	Labriformes

Genome Traits	Expected	Observed
Haploid size (bp)	778,444,864	777,341,649
Haploid Number	24 (source: direct)	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:

- . Kmer completeness value is less than 90 for collapsed

Curator notes

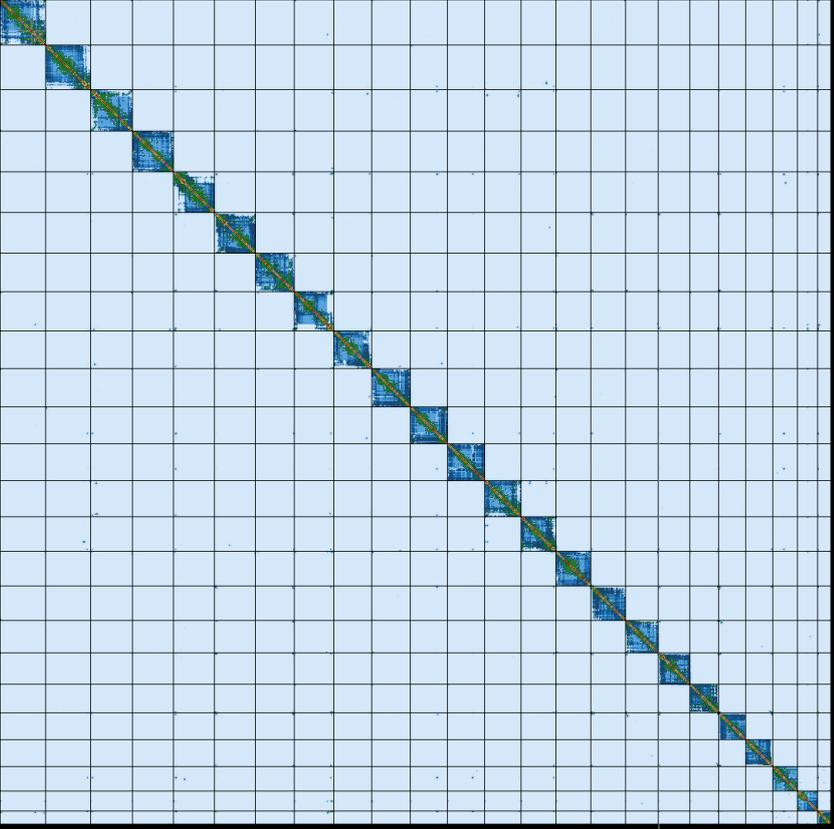
- . Interventions/Gb: 51
- . Contamination notes: ""
- . Other observations: "\The assembly of Coris julis (fCorJull) is based on 68X PacBio data and 139X Arima Hi-C data generated as part of the ATLASea programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial 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, 121 regions totaling 21.956 Mb (with the largest being 1.993 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. Mitochondrial genome was assembled using OATK. During manual curation, 8 haplotypic regions and 1 contaminant sequence were removed, totaling 19.61Mb and 0.028Mb, respectively (with the largest being 7.831Mb and 0.028Mb). 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	796,963,142	777,341,649
GC %	41.99	42.01
Gaps/Gbp	126.73	135.08
Total gap bp	10,100	12,200
Scaffolds	60	51
Scaffold N50	34,288,531	34,860,759
Scaffold L50	11	11
Scaffold L90	21	21
Contigs	161	156
Contig N50	27,439,000	27,439,000
Contig L50	13	13
Contig L90	37	35
QV	49.2093	49.2088
Kmer compl.	76.9629	76.1153
BUSCO sing.	96.2%	98.3%
BUSCO dupl.	2.4%	0.3%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	1.4%	1.4%

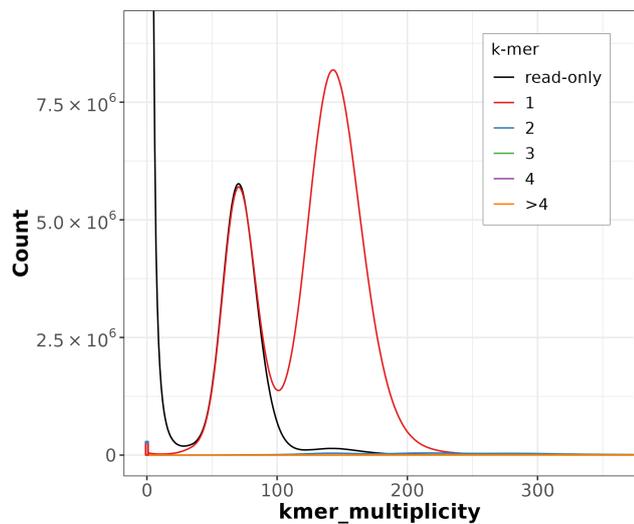
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

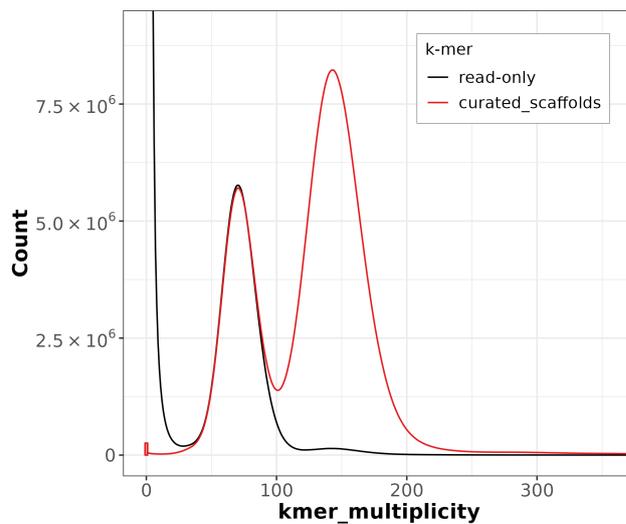


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

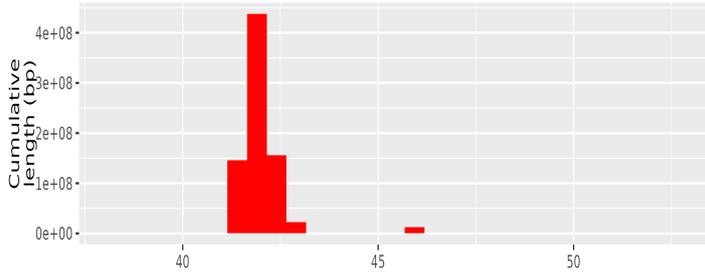


Distribution of k-mer counts per copy numbers found in asm

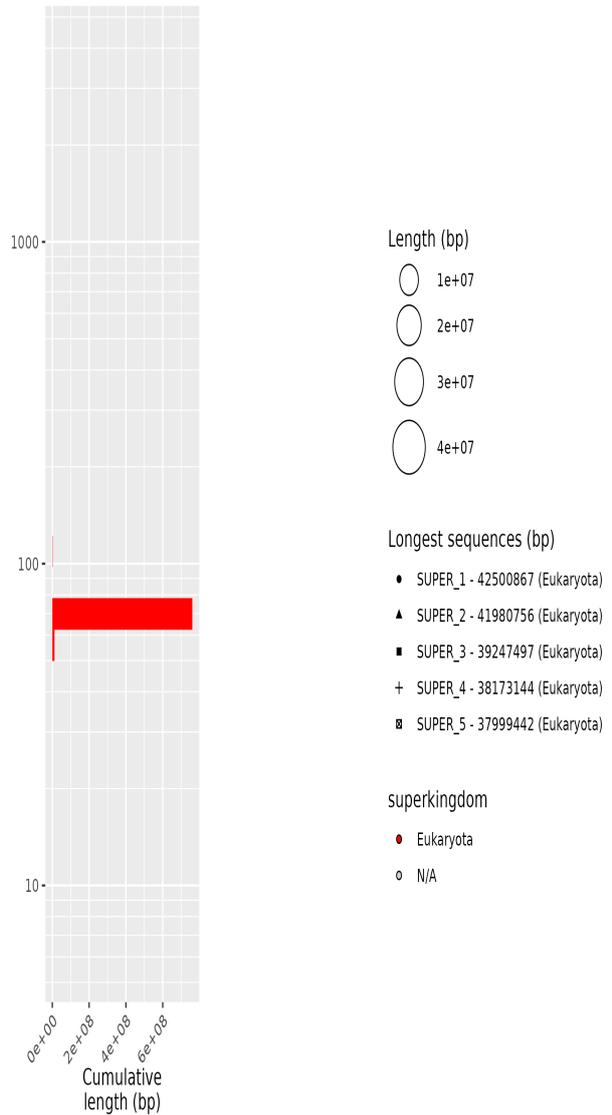
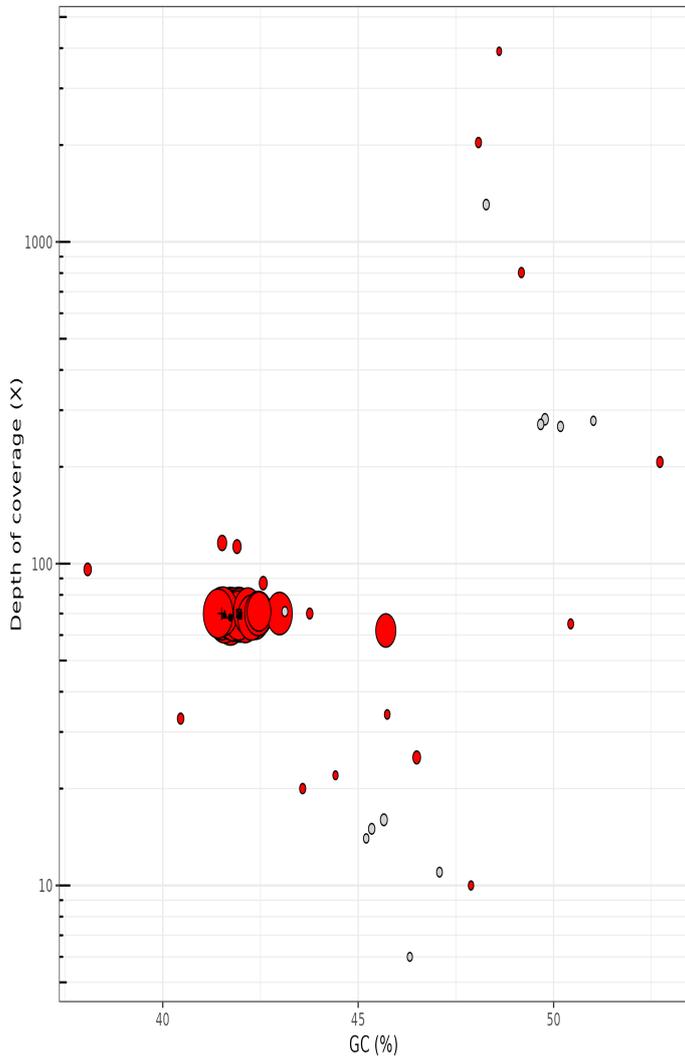


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	68	139

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: Sophie Layac

Affiliation: Genoscope

Date and time: 2026-02-02 13:46:40 CET