ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	94953
ToLID	**fSerCab1**
Species	Serranus cabrilla
Class	Actinopteri
Order	Perciformes

Genome Traits	Expected	Observed
Haploid size (bp)	672,904,498	735,069,690
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.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

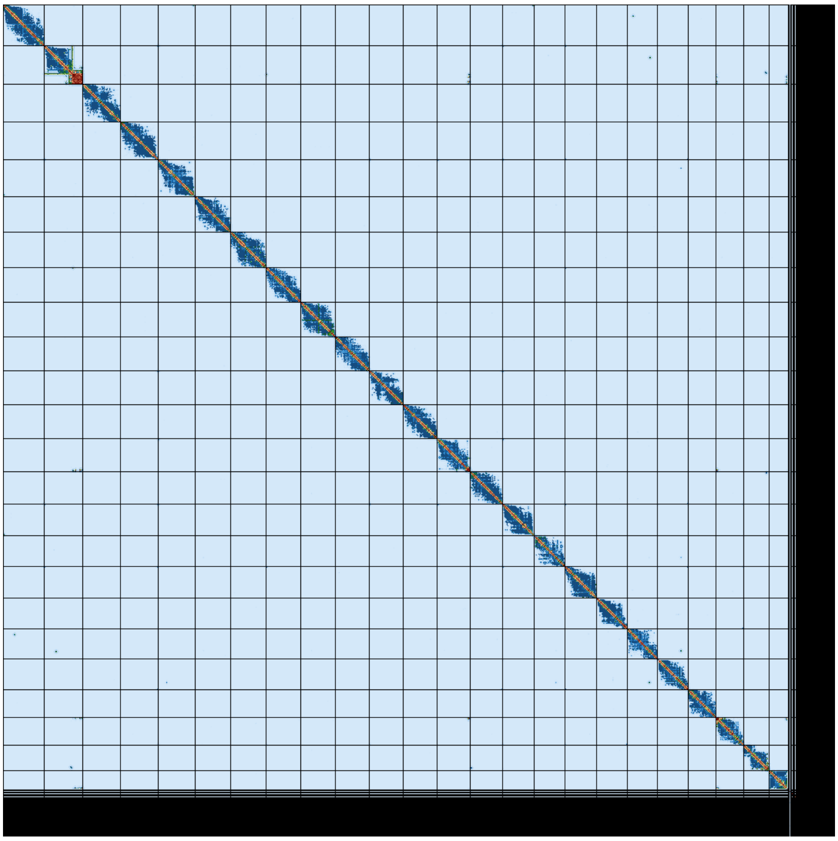
. Interventions/Gb: 4
. Contamination notes: ""
. Other observations: "The assembly of Serranus cabrilla (fSerCab1) is based on 87X ONT data and 304X Omnic Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 3 contigs were identified as contaminants (bacterial), totaling 3.2 kb (with the largest being 1.4 kb). Additionally, 194 regions totaling 10 Mb (with the largest being 5.4 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using ptGAUL. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 1 haplotypic regions and 2 contaminant sequences were removed, totaling 1.2 Mb and 0.39 Mb, respectively (with the largest being 1.2 Mb and 0.2 Mb). 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	737,618,191	735,069,690
GC %	39.33	39.33
Gaps/Gbp	0	4.08
Total gap bp	0	600
Scaffolds	637	220
Scaffold N50	30,008,223	30,008,223
Scaffold L50	12	12
Scaffold L90	23	23
Contigs	637	223
Contig N50	30,008,223	30,008,223
Contig L50	12	12
Contig L90	23	23
QV	48.1293	48.3892
Kmer compl.	78.3289	78.3051
BUSCO sing.	97.3%	97.5%
BUSCO dupl.	0.3%	0.2%
BUSCO frag.	0.5%	0.5%
BUSCO miss.	1.9%	1.9%

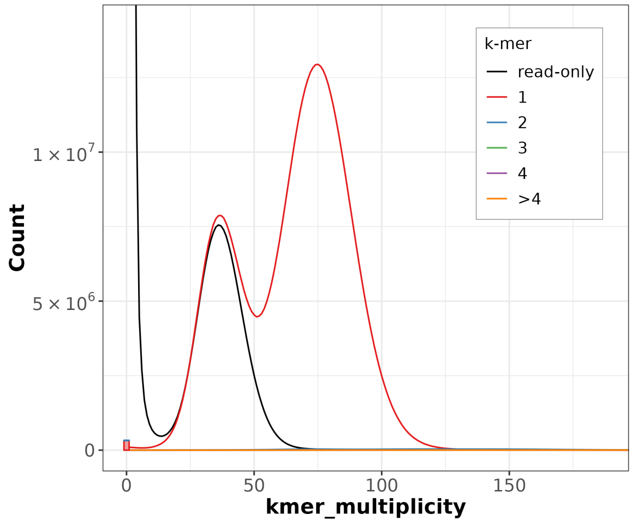
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

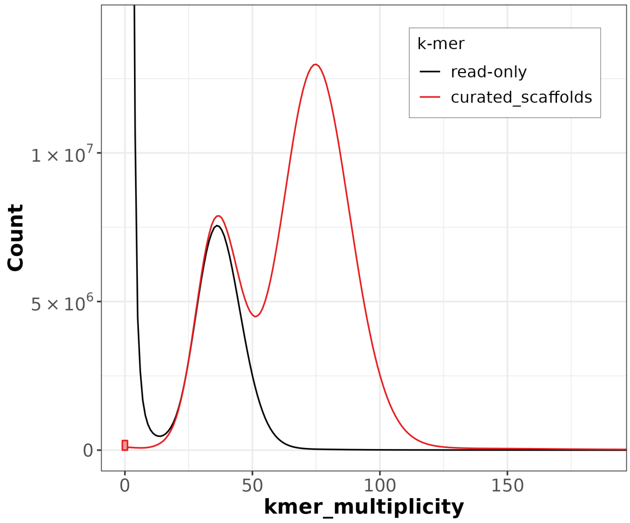


collapsed [LINK]

K-mer spectra of curated assembly

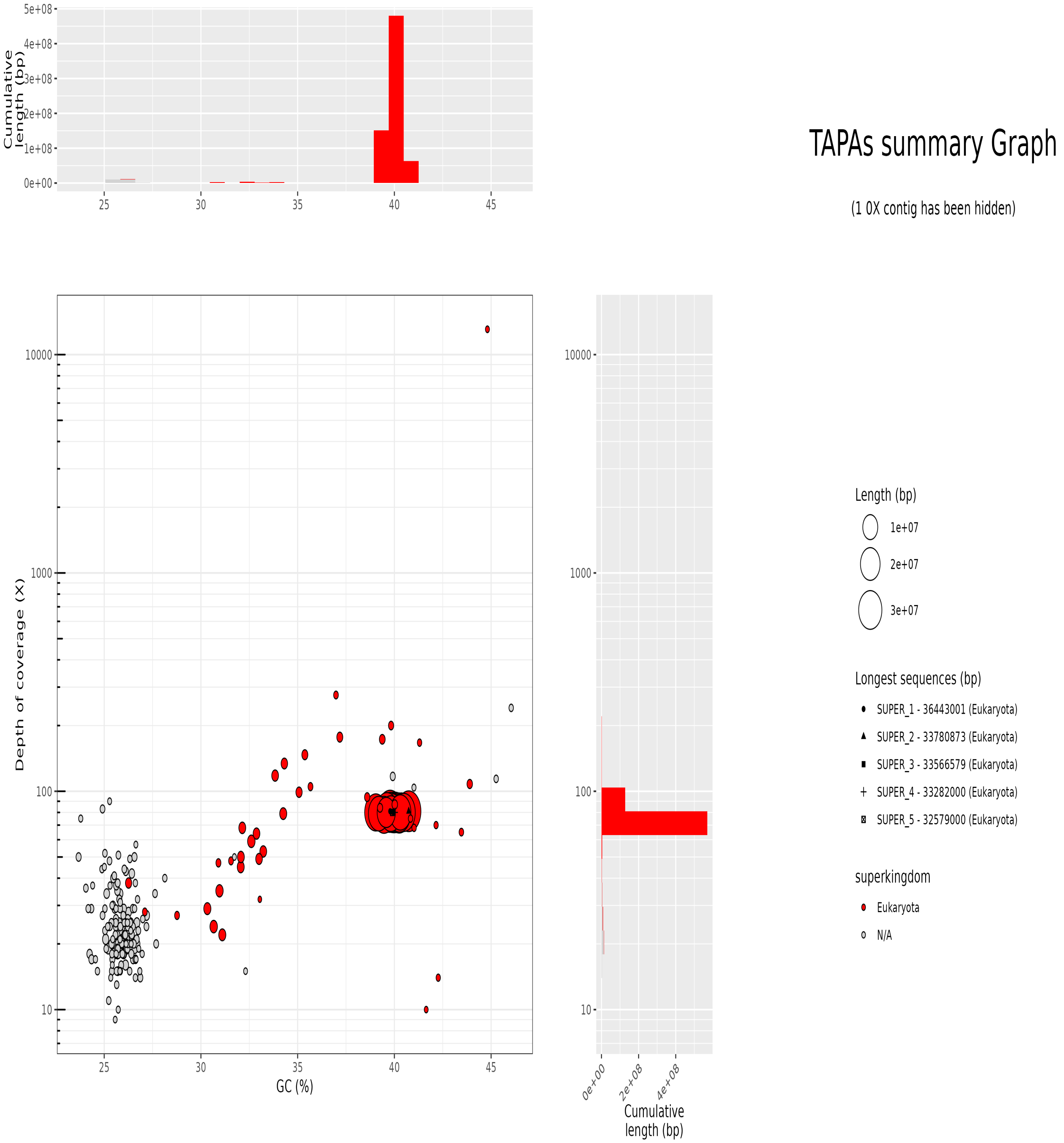


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	87	304

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: 2025-07-09 17:25:11 CEST