ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	349661
ToLID	**fSerScr1**
Species	Serranus scriba
Class	Actinopteri
Order	Perciformes

Genome Traits	Expected	Observed
Haploid size (bp)	837,754,929	924,817,704
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:

. Not 90% of assembly in chromosomes for collapsed

Curator notes

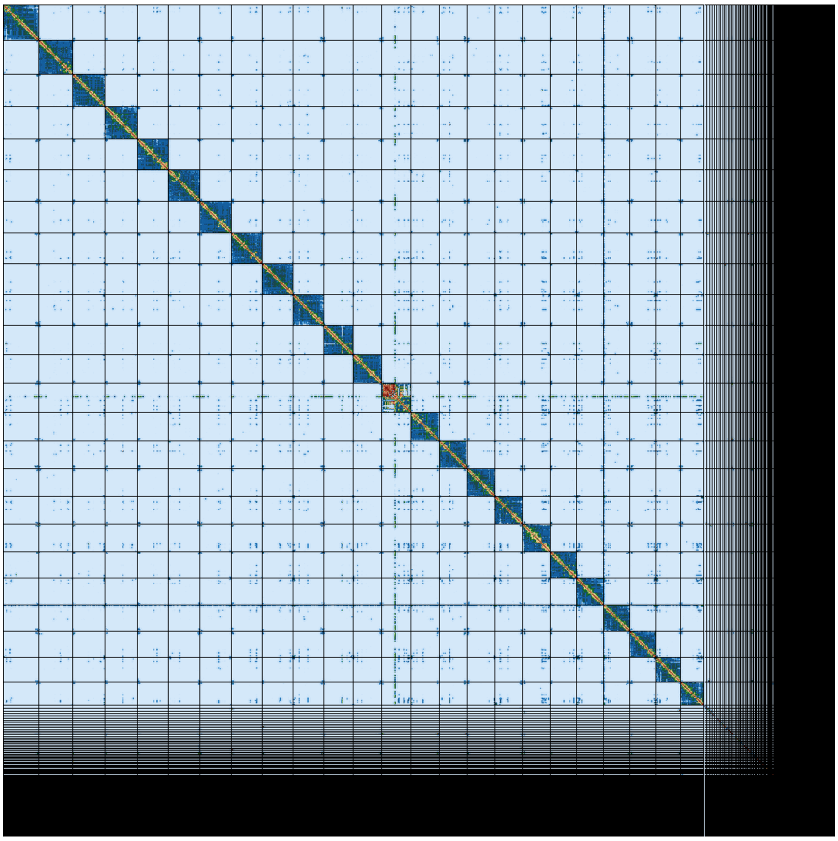
. Interventions/Gb: 19
. Contamination notes: ""
. Other observations: "The assembly of Serranus scriba (fSerScr1) is based on 55X ONT data and 224X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial ONT assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 1 contig of 0.069 Mbwas identified as contaminant (bacterial, archaeal, or viral). Additionally, 128 regions totaling 27.172 Mb (with the largest being 0.848 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using ptGAUL. During manual curation, 1 haplotypic region of 0.66Mb was removed. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. Note that the organization of chromosome 13 from 6Mb to 15 Mb is a bit uncertain. The position and orientation of the small contigs relative to each other are not fully reliable. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	925,478,003	924,817,704
GC %	40.42	40.42
Gaps/Gbp	19.45	35.68
Total gap bp	1,800	4,900
Scaffolds	269	254
Scaffold N50	31,017,000	31,366,288
Scaffold L50	14	14
Scaffold L90	57	52
Contigs	287	287
Contig N50	27,104,000	27,104,000
Contig L50	15	15
Contig L90	71	71
QV	48.3796	48.378
Kmer compl.	91.2278	91.2169
BUSCO sing.	99.4%	99.5%
BUSCO dupl.	0.4%	0.3%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.3%	0.3%

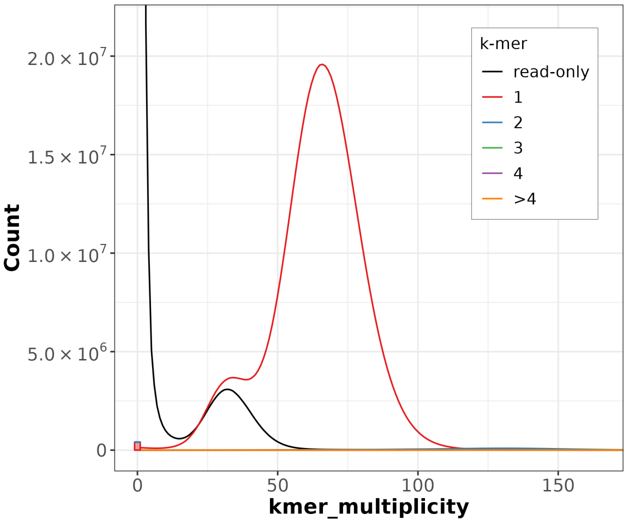
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

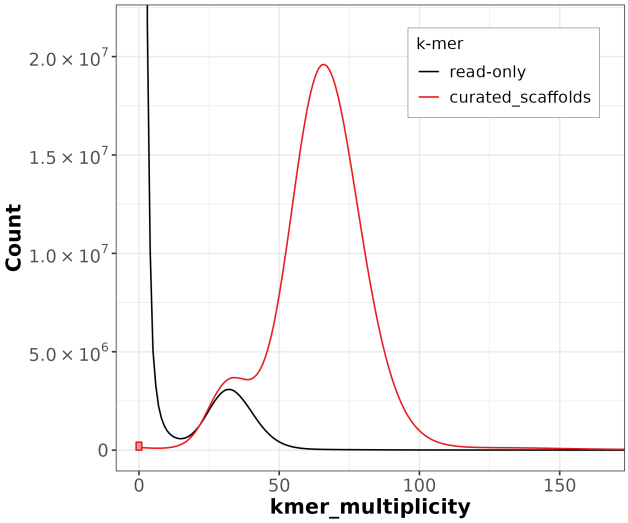


collapsed [LINK]

K-mer spectra of curated assembly

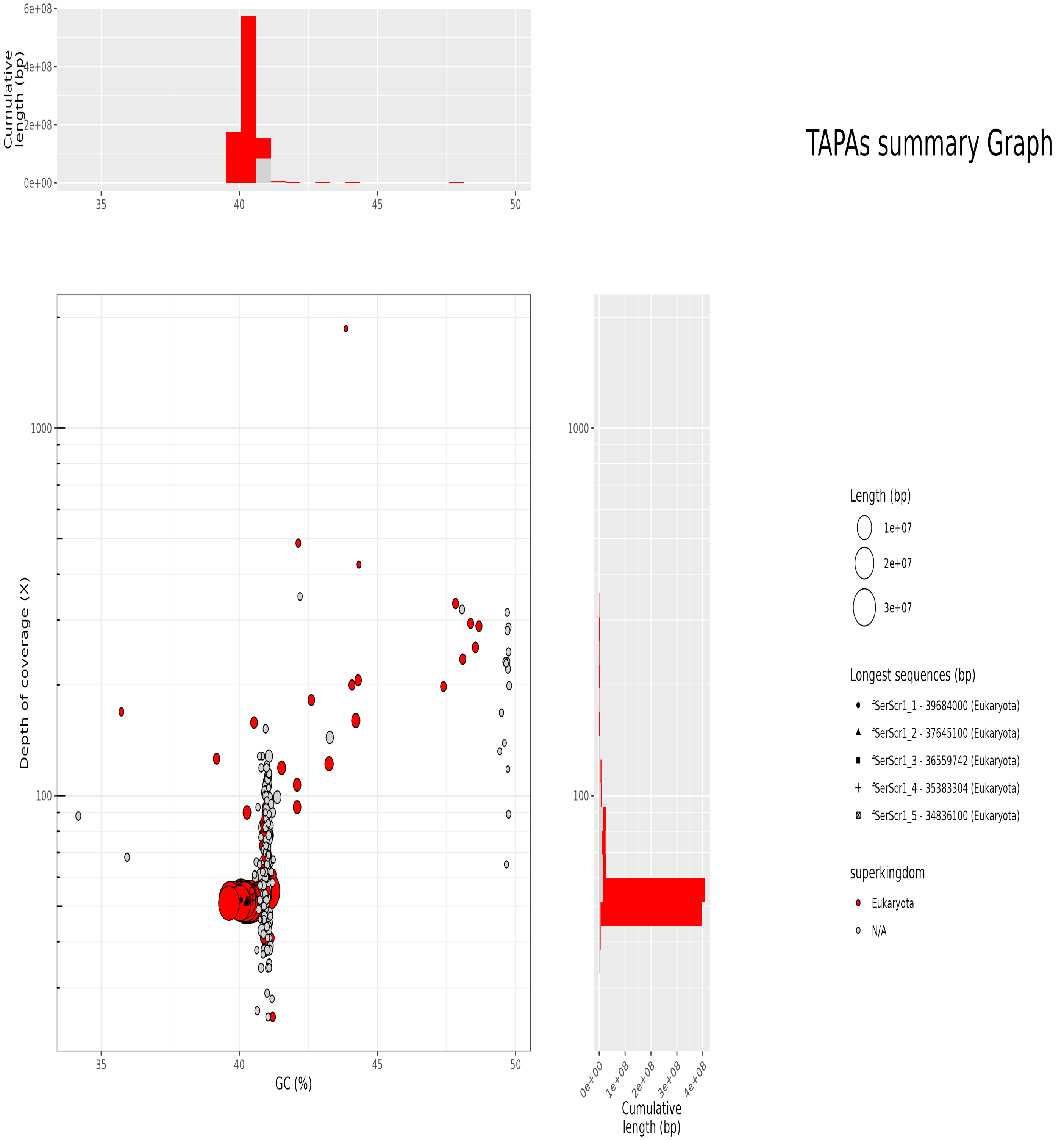


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	55	224

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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2025-12-02 08:38:51 CET