

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	87759
ToLID	fSarSall
Species	Sarpa salpa
Class	Actinopteri
Order	Spariformes

Genome Traits	Expected	Observed
Haploid size (bp)	774,680,592	782,260,259
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: 12
- . Contamination notes: ""
- . Other observations: "The assembly of Sarpa salpa (fSarSall) is based on 45X PacBio data and 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 Hifiiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 2 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 55 Kb (with the largest being 31Kb). Additionally, 68 regions totaling 7 Mb (with the largest being 1.7 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. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size "

Quality metrics table

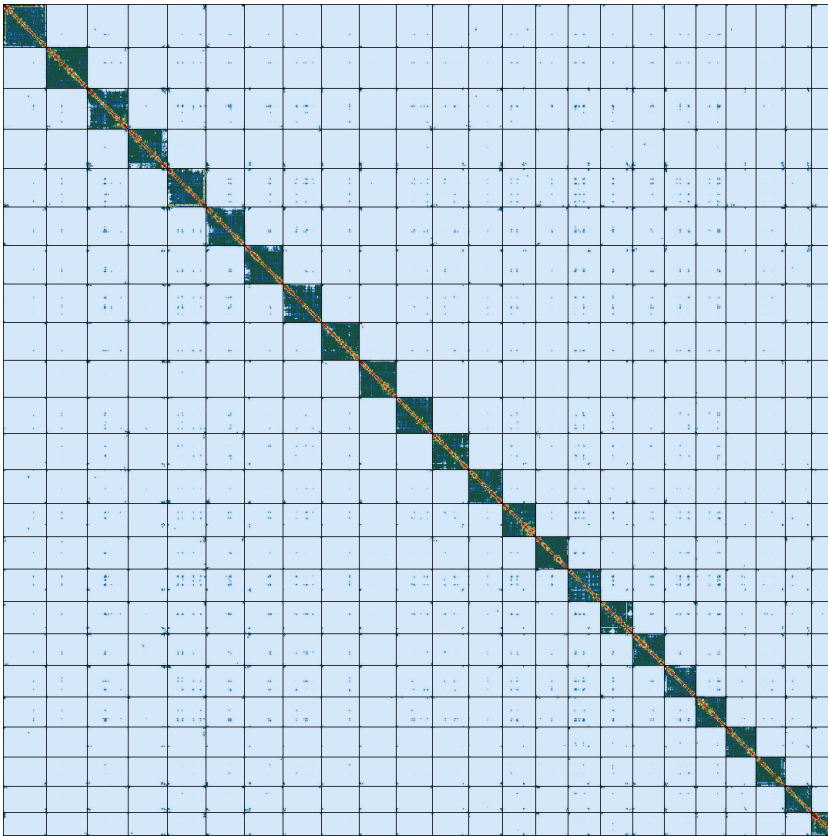
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	782,262,659	782,260,259
GC %	41.91	41.91
Gaps/Gbp	80.54	81.81
Total gap bp	6,300	6,900
Scaffolds	37	35
Scaffold N50	34,095,251	34,095,251
Scaffold L50	11	11
Scaffold L90	21	21
Contigs	100	99
Contig N50	29,016,000	29,016,000
Contig L50	13	13
Contig L90	28	28
QV	48.5154	48.5159
Kmer compl.	84.3844	84.3843
BUSCO sing.	97.4%	99.5%
BUSCO dupl.	0.3%	0.2%
BUSCO frag.	0.5%	0.1%
BUSCO miss.	1.8%	0.3%

Warning! BUSCO versions or lineage datasets are not the same across results:

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

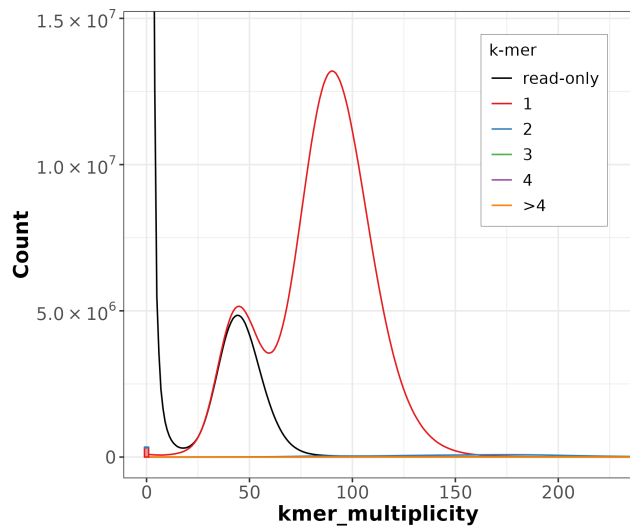
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

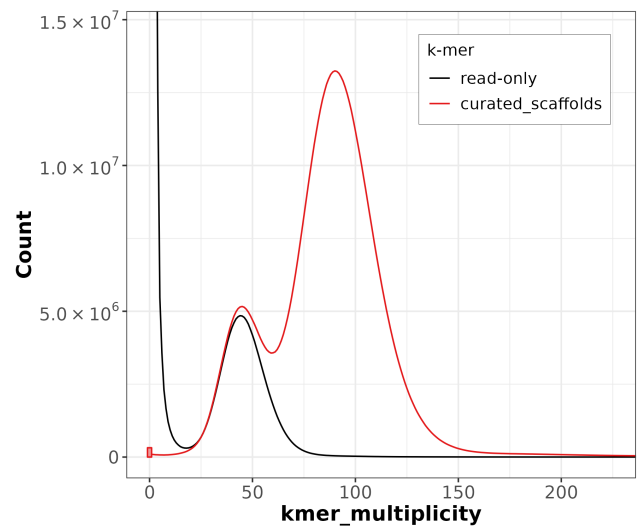


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

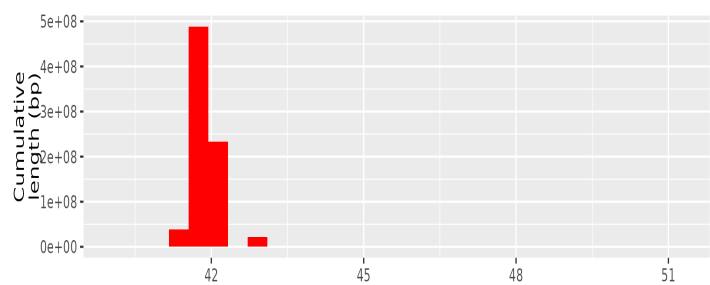


Distribution of k-mer counts per copy numbers found in asm

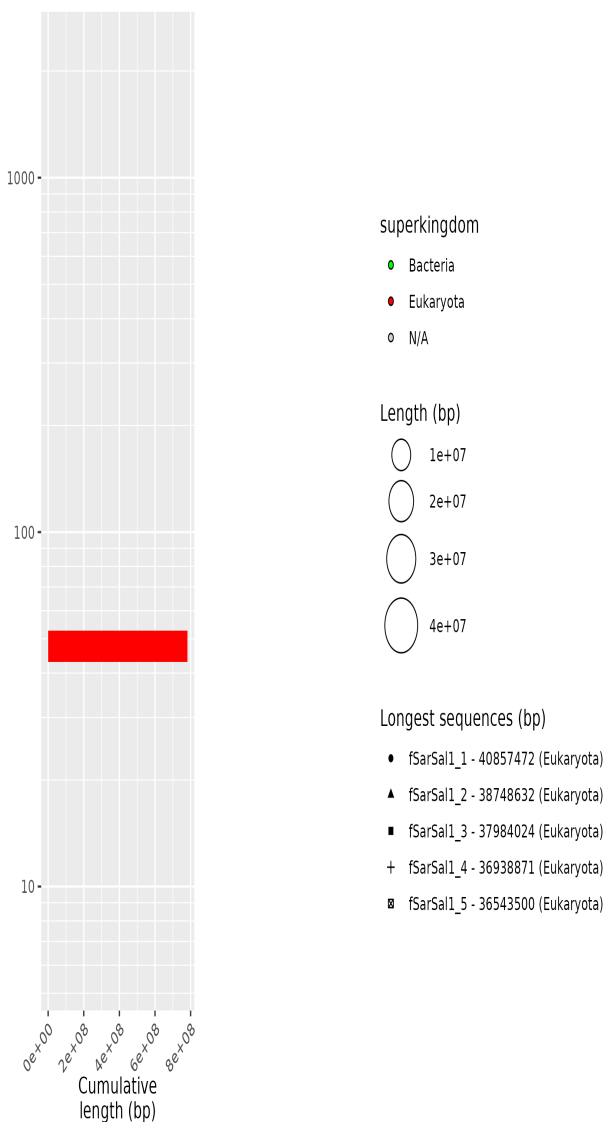
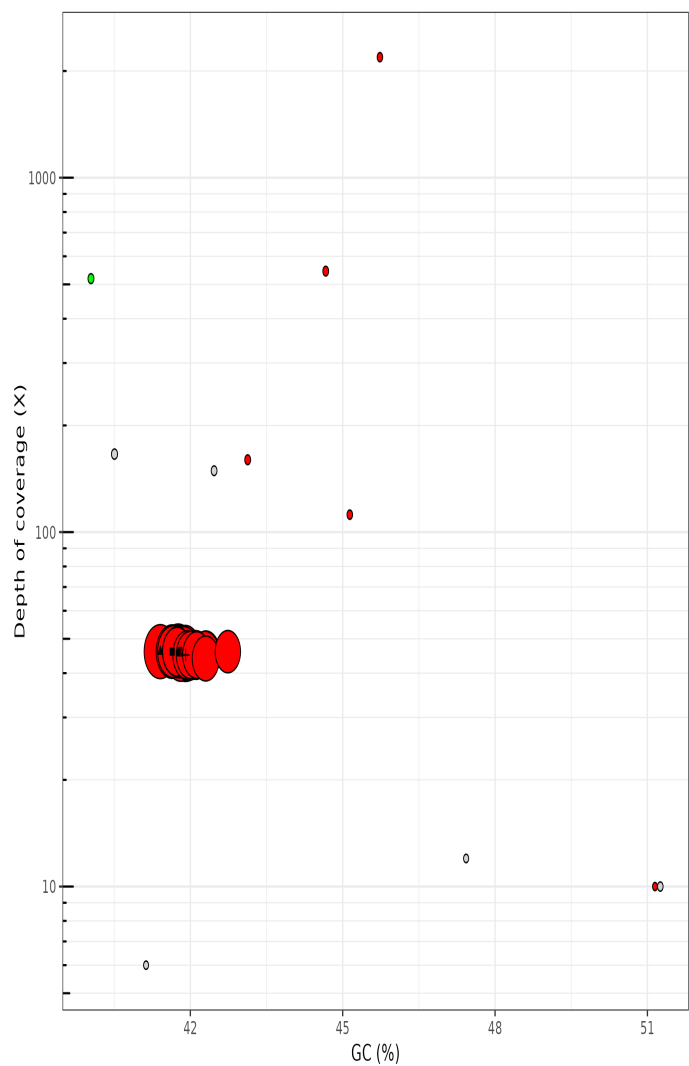


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	45	192

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: 2025-10-08 15:00:45 CEST