ERGA Assembly Report
v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	182855
ToLID	**sRajUnd1**
Species	Raja undulata
Class	Chondrichthyes
Order	Rajiformes

Genome Traits	Expected	Observed
Haploid size (bp)	2,855,542,641	2,933,668,520
Haploid Number	49 (source: ancestor)	49
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q45

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: 89
. Contamination notes: ""
. Other observations: "The assembly of Raja undulata (sRajUnd1) is based on 33X PacBio data and 97X 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. In total, 24 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1 Mb (with the largest being 375 kb).Additionally, 723 regions totaling 85 Mb (with the largest being 1.6Mb) 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, 34 haplotypic regions and 5 contaminant sequences were removed, totaling 26.942 Mb and 1.109Mb, respectively (with the largest being 4.442Mb and 0.784). 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	2,962,121,652	2,933,668,520
GC %	45.06	45.05
Gaps/Gbp	822.72	866.83
Total gap bp	243,700	269,800
Scaffolds	1,560	1,253
Scaffold N50	68,556,055	72,868,749
Scaffold L50	14	14
Scaffold L90	77	42
Contigs	3,997	3,796
Contig N50	2,804,833	2,806,603
Contig L50	293	289
Contig L90	1,344	1,323
QV	45.169	45.1627
Kmer compl.	88.4296	88.3027
BUSCO sing.	92.2%	96.0%
BUSCO dupl.	1.4%	1.5%
BUSCO frag.	2.7%	0.7%
BUSCO miss.	3.7%	1.8%

Warning! BUSCO versions or lineage datasets are not the same across results:
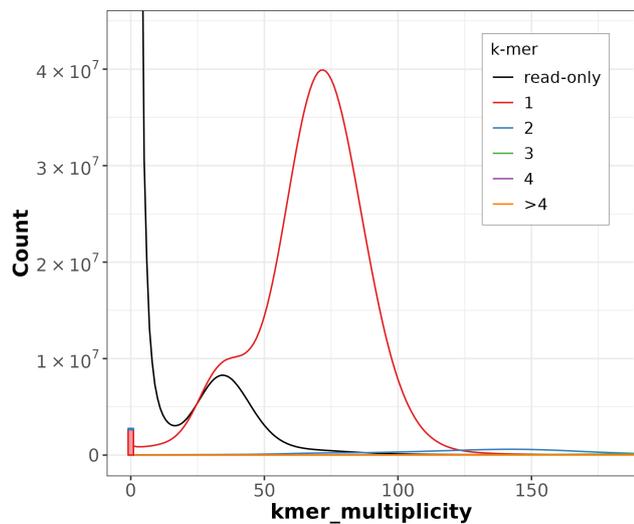BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: vertebrata_odb12 (genomes:119, BUSCOs:3390)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: vertebrata_odb12 (genomes:119, BUSCOs:3390)
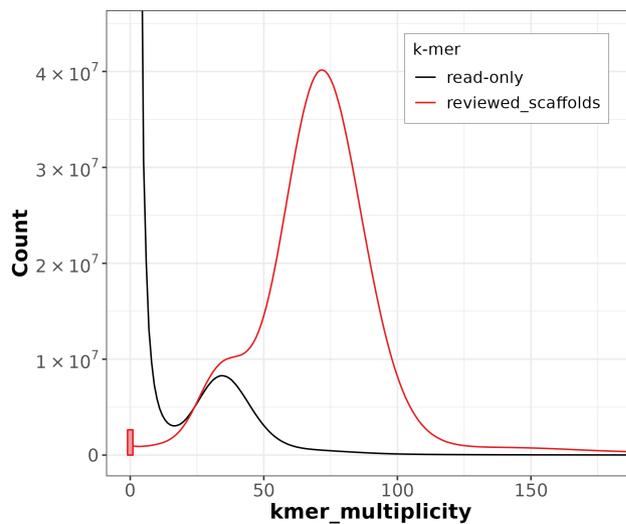
HiC contact map of curated assembly
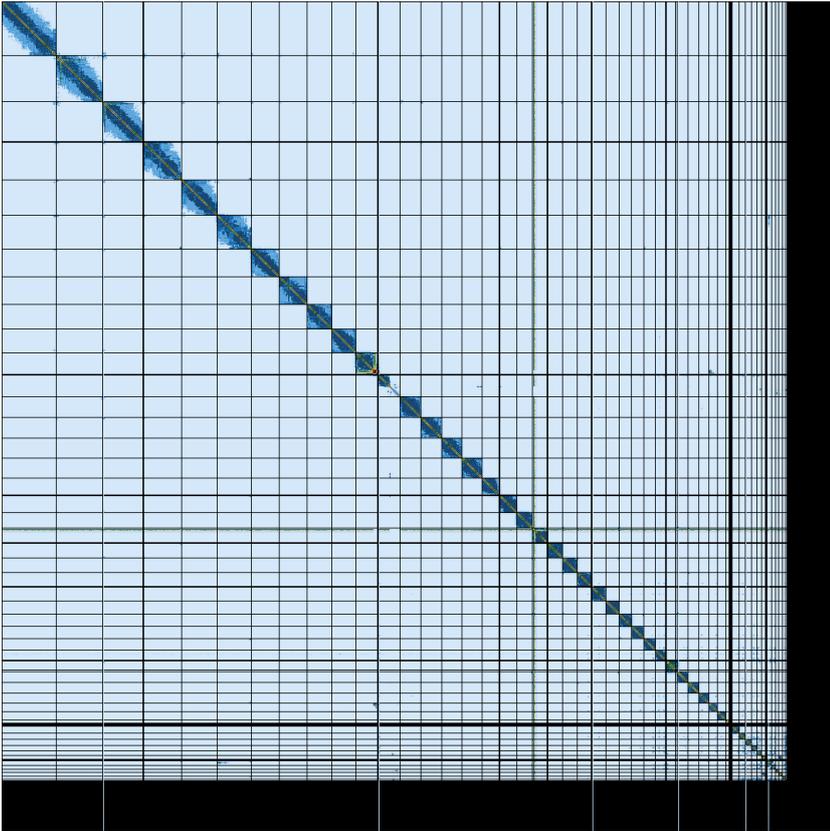
collapsed [LINK]

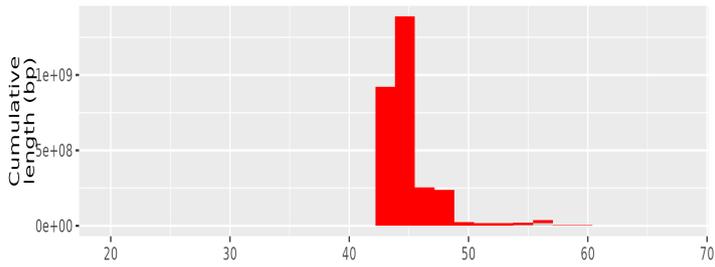# K-mer spectra of curated assembly

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

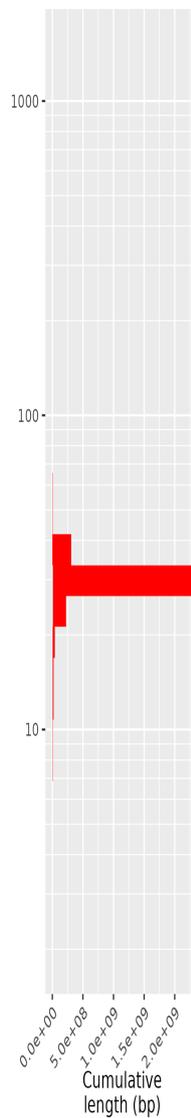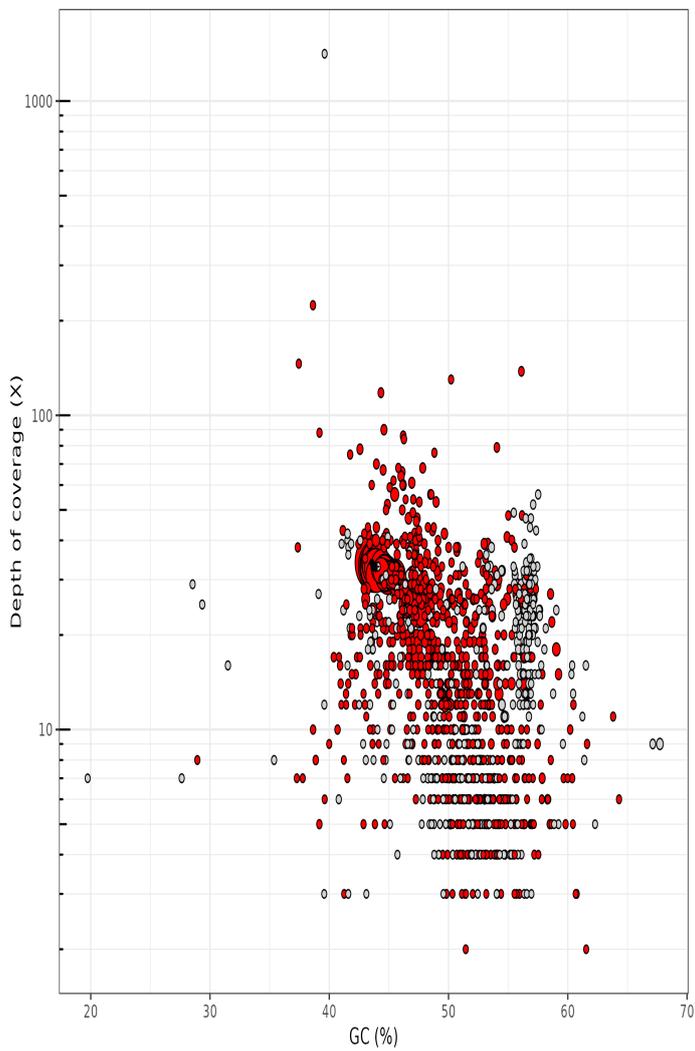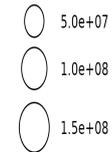Distribution of k-mer counts coloured by
their presence in reads/assemblies

Post-curation contamination screening

TAPAs summary Graph

Length (bp)
- 5.0e+07
- 1.0e+08
- 1.5e+08

Longest sequences (bp)
- SUPER_1 - 192021789 (Eukaryota)
- ▲ SUPER_2 - 161455778 (Eukaryota)
- ■ SUPER_3 - 141319640 (Eukaryota)
- + SUPER_4 - 132616081 (Eukaryota)
- ⊠ SUPER_5 - 124582786 (Eukaryota)

superkingdom
- Eukaryota
- N/A

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	32	87

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: 2026-02-03 04:14:03 CET