

# ERGA Assembly Report

v24.10.15

Tags: Other[INVALID TAG]

TxID	30501
ToLID	<b>sCarTaul_hap1</b>
Species	Carcharias taurus
Class	Chondrichthyes
Order	Lamniformes

Genome Traits	Expected	Observed
Haploid size (bp)	4,296,075,824	5,133,528,597
Haploid Number	36 (source: ancestor)	38
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q59

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

- . Observed Haploid Number is different from Expected
- . Not 90% of assembly in chromosomes for collapsed

### Curator notes

- . Interventions/Gb: 25
- . Contamination notes: ""
- . Other observations: "The assembly of Carcharias taurus (sCarTaul) was generated as part of the Vertebrate Genomes Project (<https://vertebrategenomesproject.org/>). Both haplotypes were analyzed and manually improved using Pretext "

# Quality metrics table

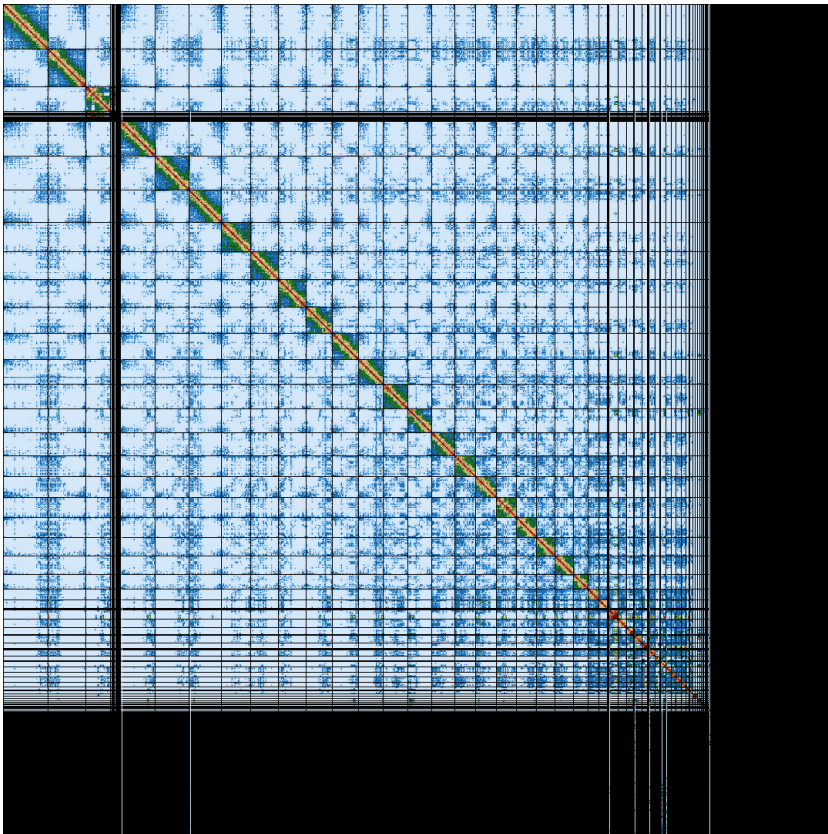
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	4,818,022,793	5,133,528,597
GC %	45.24	45.4
Gaps/Gbp	328.97	353.95
Total gap bp	9,456,865	11,113,672
Scaffolds	1,766	2,803
Scaffold N50	149,215,395	148,143,658
Scaffold L50	12	14
Scaffold L90	93	177
Contigs	3,351	4,620
Contig N50	6,910,477	6,312,879
Contig L50	180	198
Contig L90	915	1,199
QV	60.6841	59.9726
Kmer compl.	97.4221	97.1237
BUSCO sing.	93.2%	95.1%
BUSCO dupl.	2.0%	2.4%
BUSCO frag.	2.4%	1.3%
BUSCO miss.	2.4%	1.2%

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)

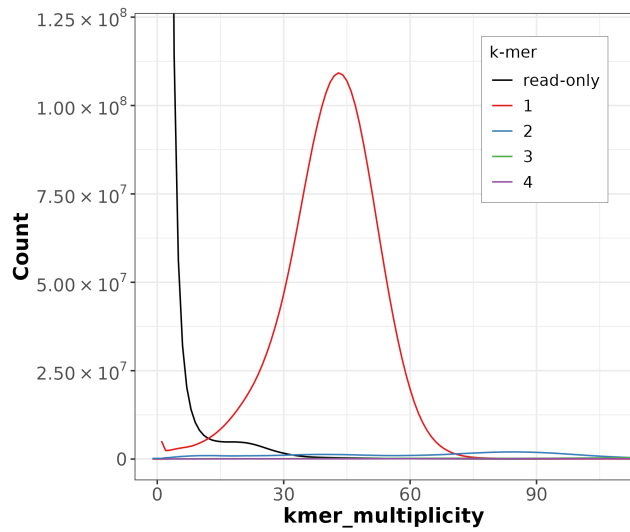
BUSCO: 5.8.2 (euk\_genome\_min, miniprot) / Lineage: vertebrata\_odb12 (genomes:119, BUSCOs:3390)

# HiC contact map of curated assembly

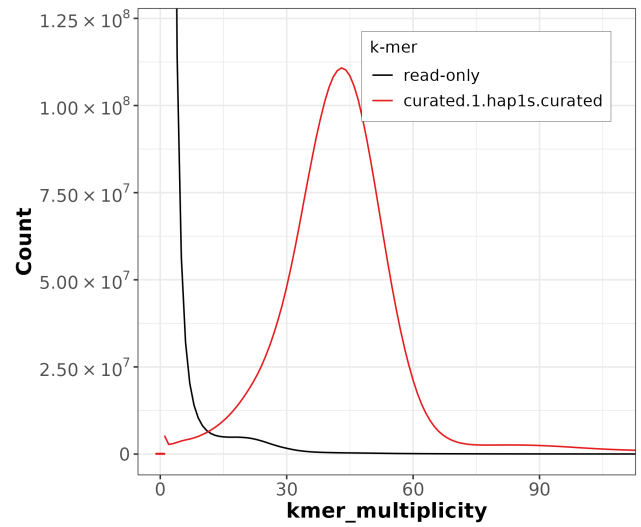


collapsed [\[LINK\]](#)

# K-mer spectra of curated assembly

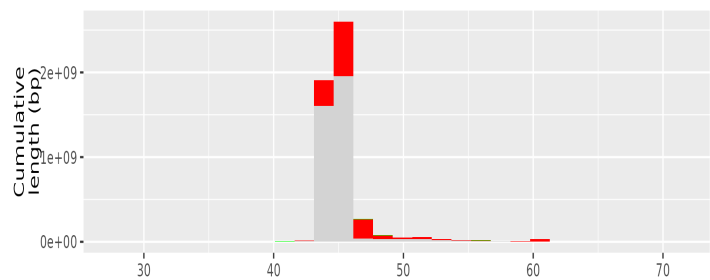


Distribution of k-mer counts per copy numbers found in asm

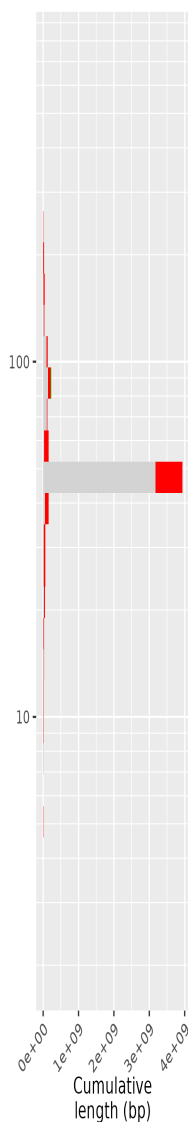
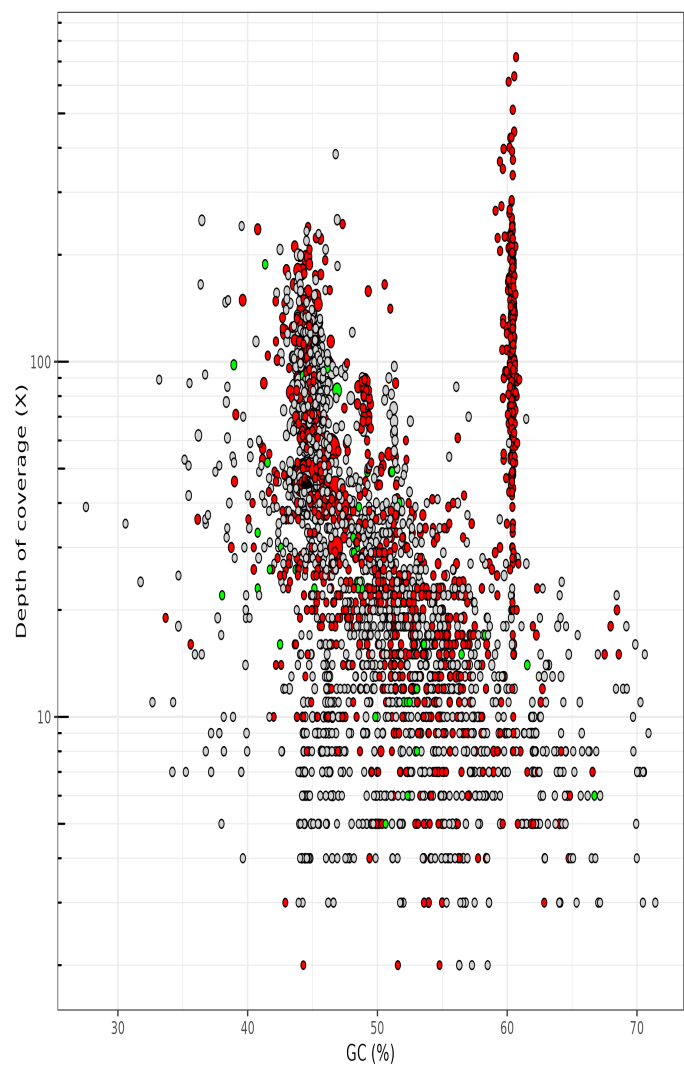


Distribution of k-mer counts coloured by their presence in reads/assemblies

# Post-curation contamination screening



TAPAs summary Graph



- Longest sequences (bp)
- sCarTau1\_1 - 277141663 (N/A)
  - ▲ sCarTau1\_2 - 232641887 (N/A)
  - sCarTau1\_4 - 213981726 (N/A)
  - + sCarTau1\_5 - 206397951 (N/A)
  - ⊠ sCarTau1\_6 - 198813373 (N/A)

- superkingdom
- Bacteria
  - Eukaryota
  - N/A

- Length (bp)
- 1e+08
  - 2e+08

**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	Arima
Coverage	NA	NA

# 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: Emilie Teodori

Affiliation: Genoscope

Date and time: 2025-08-08 07:45:20 CEST