

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	673854
ToLID	sDasPas1
Species	Dasyatis pastinaca
Class	Chondrichthyes
Order	Myliobatiformes

Genome Traits	Expected	Observed
Haploid size (bp)	3,861,507,254	4,119,270,395
Haploid Number	34 (source: ancestor)	32
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q48

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

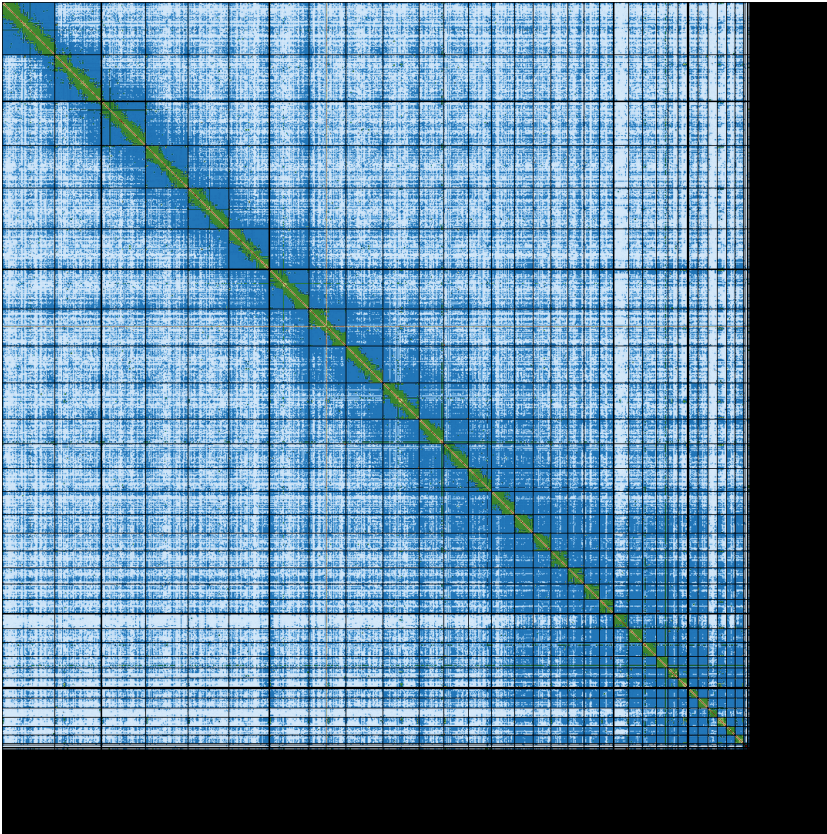
. Interventions/Gb: 32
. Contamination notes: ""
. Other observations: "The assembly of *Dasyatis pastinaca* (sDasPas1) is based on 35X PacBio data and 146X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial 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, 5 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.181 Mb (with the largest being 0.052 Mb). Additionally, 857 regions totaling 155.42 Mb (with the largest being 1.493 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 7 haplotypic regions and 21 contaminant sequences were removed, totaling 8.126Mb and 3.404Mb respectively (with the largest being 3.286Mb and 0.787Mb). The telomeric pattern TTAGGG was identified with TelFinder and used to generate the Pretext Telomeres tracks. 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	4,130,771,558	4,119,270,395
GC %	43.45	43.44
Gaps/Gbp	394.12	403.47
Total gap bp	162,800	173,600
Scaffolds	986	929
Scaffold N50	120,578,597	122,042,294
Scaffold L50	11	11
Scaffold L90	62	43
Contigs	2,614	2,591
Contig N50	4,237,000	4,237,000
Contig L50	260	259
Contig L90	1,085	1,080
QV	48.1635	48.1746
Kmer compl.	96.8163	96.7849
BUSCO sing.	96.0%	96.1%
BUSCO dupl.	2.3%	2.3%
BUSCO frag.	0.3%	0.3%
BUSCO miss.	1.4%	1.4%

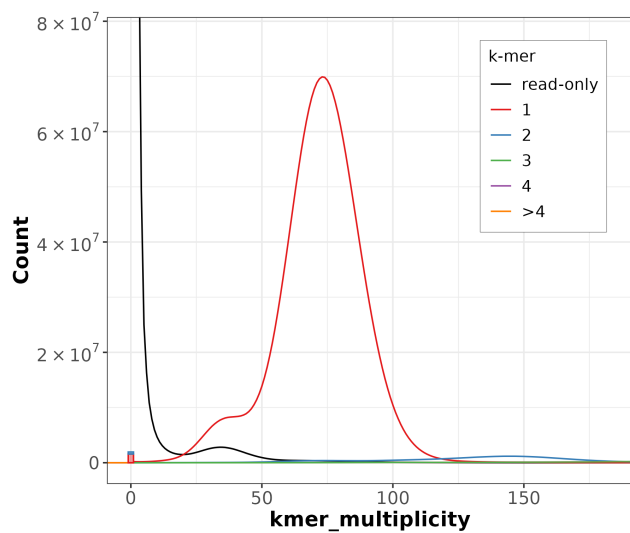
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: vertebrata_odb12 (genomes:119, BUSCOs:3390)

HiC contact map of curated assembly

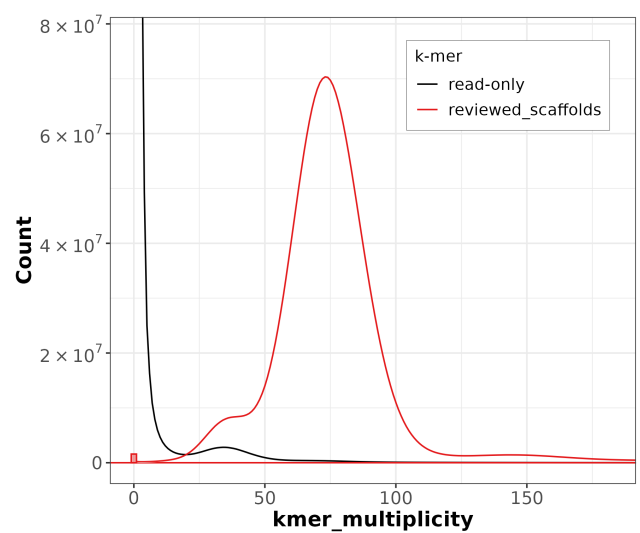


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

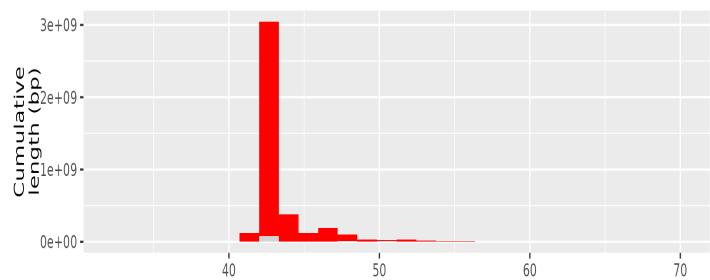


Distribution of k-mer counts per copy numbers found in asm

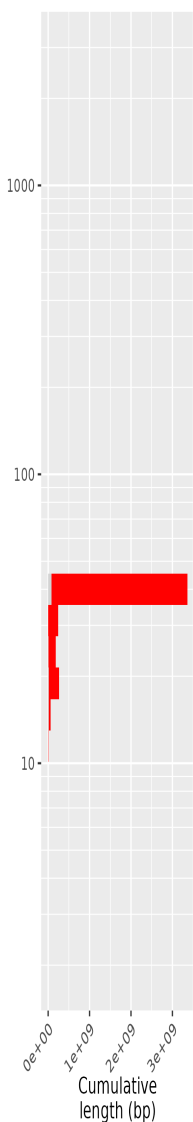
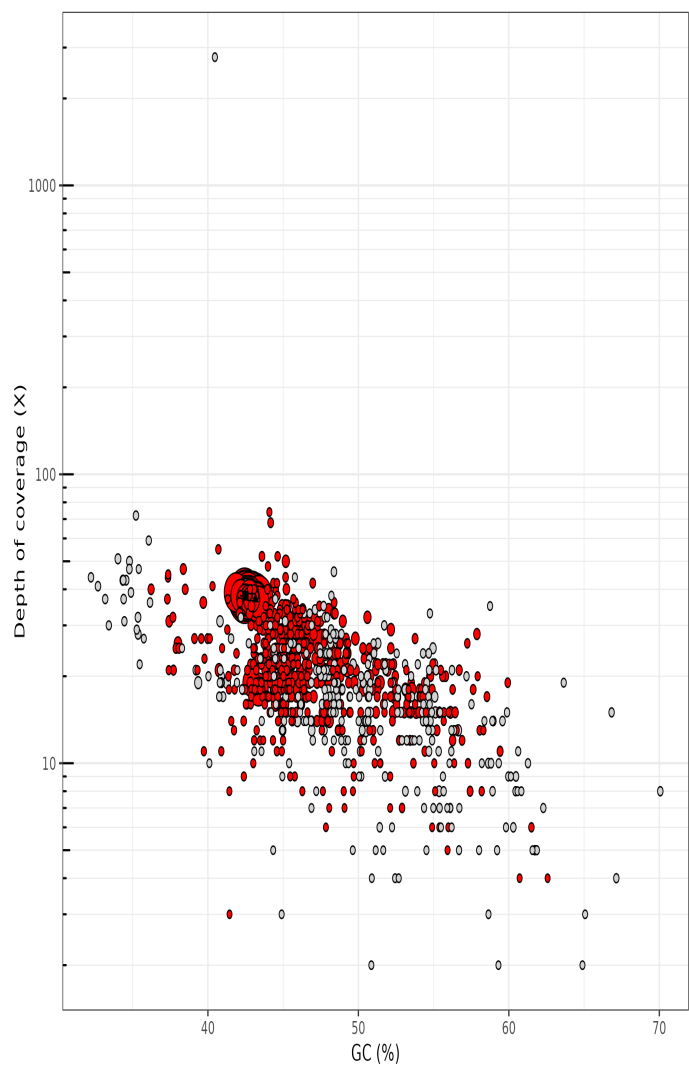


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



Longest sequences (bp)

- SUPER_1 - 262559035 (Eukaryota)
- ▲ SUPER_2 - 228824716 (Eukaryota)
- SUPER_3 - 219045104 (Eukaryota)
- + SUPER_4 - 208559491 (Eukaryota)
- ▣ SUPER_5 - 201111394 (Eukaryota)

Length (bp)

- 5.0e+07
- 1.0e+08
- 1.5e+08
- 2.0e+08
- 2.5e+08

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	35	147

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: 2026-07-10 00:12:59 CEST