ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	7788
ToLID	**sTorMar1**
Species	Torpedo marmorata
Class	Chondrichthyes
Order	Torpediniformes

Genome Traits	Expected	Observed
Haploid size (bp)	6,010,064,597	6,221,104,460
Haploid Number	43 (source: direct)	40
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q47

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

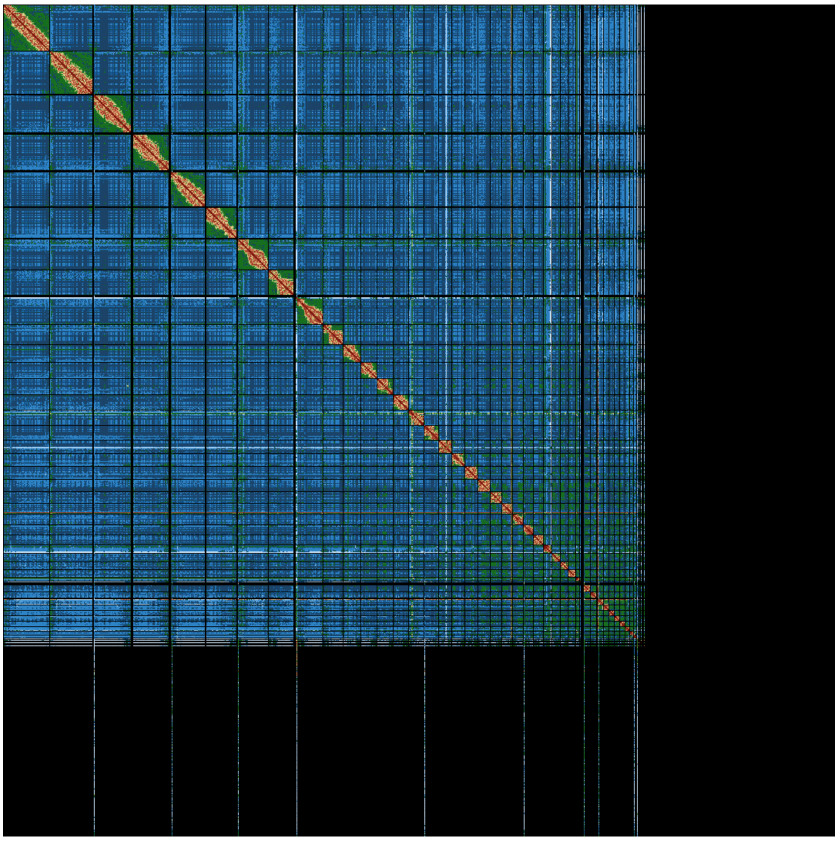
. Interventions/Gb: 17
. Contamination notes: ""
. Other observations: "The assembly of Torpedo marmorata (sTorMar1) is based on 30X PacBio data and Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 99 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 17.1 Mb (with the largest being 1.9Mb). Additionally, 6,609 regions totaling 413 Mb (with the largest being 0.68 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using oatk. A linear sequence of 14Kb was obtained. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, no haplotypic regions were removed. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. The orientation of some chromosome arms could not be confirmed. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	6,222,118,416	6,221,104,460
GC %	43.83	43.83
Gaps/Gbp	540.81	546.85
Total gap bp	336,500	349,900
Scaffolds	7,899	4,566
Scaffold N50	91,663,292	111,696,861
Scaffold L50	17	16
Scaffold L90	1,552	1,025
Contigs	11,264	7,968
Contig N50	2,919,481	3,050,324
Contig L50	458	441
Contig L90	3,930	3,322
QV		47.5122
Kmer compl.	93.5415	93.5378
BUSCO sing.	91.7%	91.7%
BUSCO dupl.	1.2%	1.2%
BUSCO frag.	2.8%	2.8%
BUSCO miss.	4.3%	4.2%

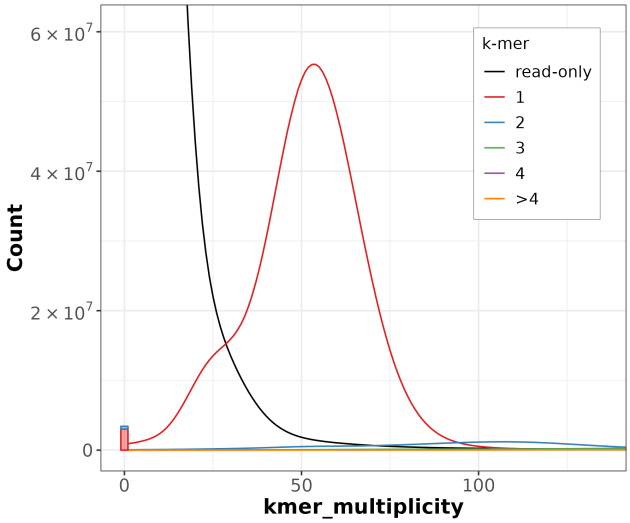
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: vertebrata_odb12 (genomes:119, BUSCOs:3390)

HiC contact map of curated assembly

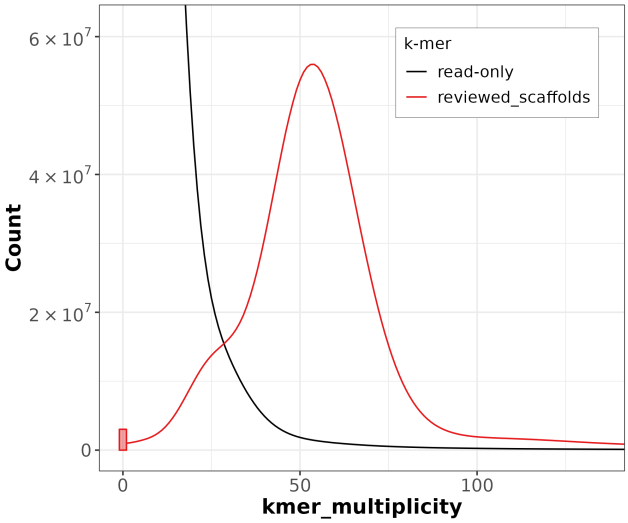


collapsed [LINK]

K-mer spectra of curated assembly

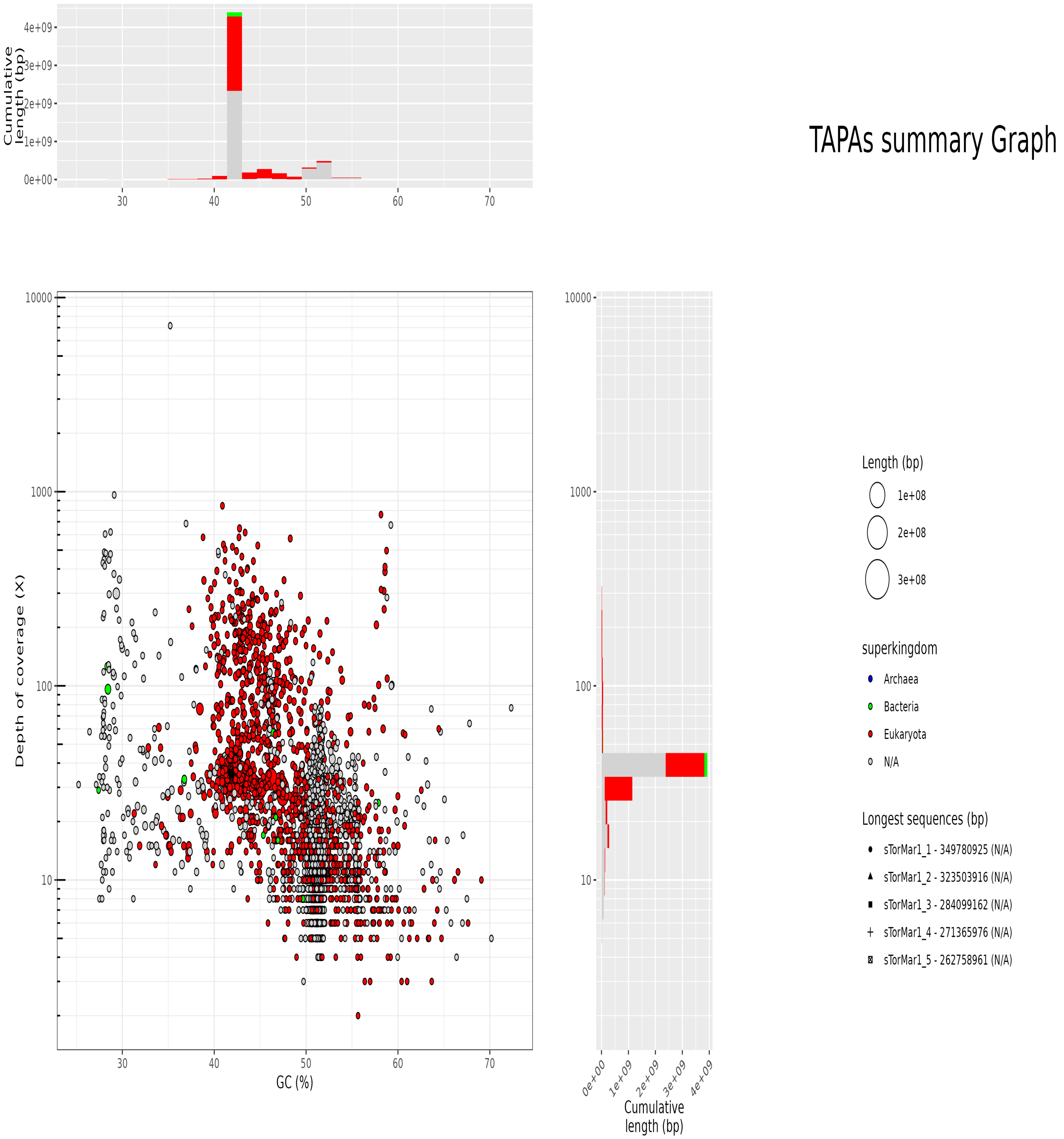


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	30	167

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: 2025-10-28 10:43:28 CET