

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	105367
ToLID	eeTriVent1
Species	<i>Tripneustes ventricosus</i>
Class	Echinoidea
Order	Temnopleuroida

Genome Traits	Expected	Observed
Haploid size (bp)	939,295,031	980,979,081
Haploid Number	16 (source: direct)	18
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q47

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

- . Observed Haploid Number is different from Expected
- . Kmer completeness value is less than 90 for collapsed

Curator notes

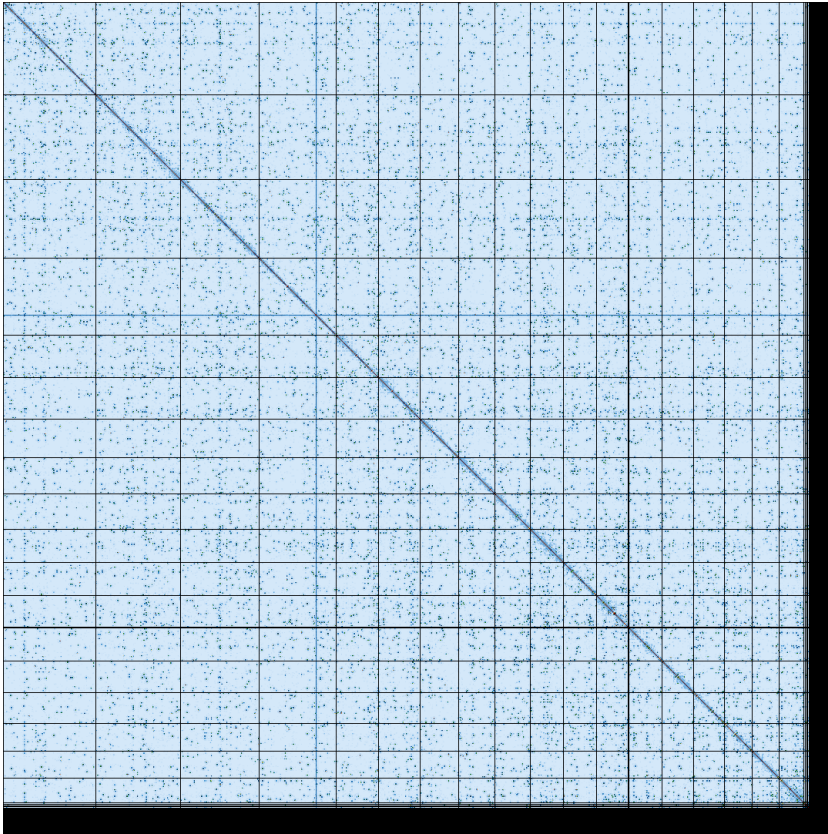
. Interventions/Gb: 31
. Contamination notes: ""
. Other observations: "The assembly of *Tripneustes ventricosus* (eeTriVent1) is based on 43X PacBio data and 182X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: 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, 3 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.043 Mb (with the largest being 0.016 Mb). Additionally, 181 regions totaling 41.803 Mb (with the largest being 2.146 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 7 haplotypic regions were removed, totaling 9.591Mb (with the largest being 4.51Mb). 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	990,563,152	980,979,081
GC %	35.77	35.77
Gaps/Gbp	114.08	118.25
Total gap bp	11,300	13,200
Scaffolds	322	316
Scaffold N50	45,380,828	49,258,133
Scaffold L50	7	6
Scaffold L90	17	17
Contigs	435	432
Contig N50	22,740,766	24,758,000
Contig L50	12	11
Contig L90	43	42
QV	46.979	47.0126
Kmer compl.	81.531	81.3032
BUSCO sing.	98.4%	99.4%
BUSCO dupl.	1.3%	0.3%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	0.1%	0.1%

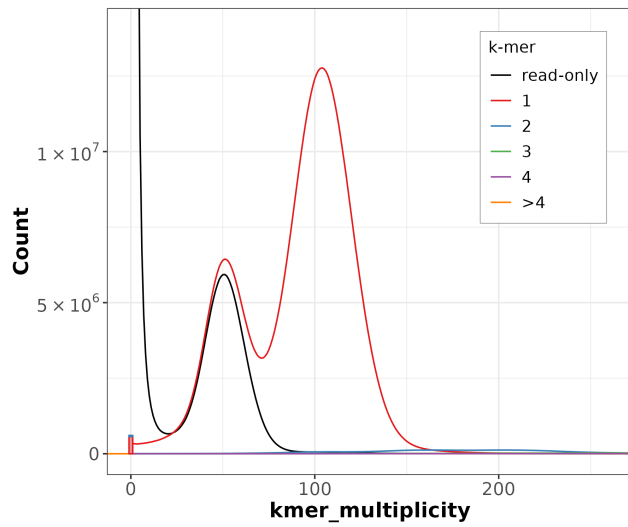
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

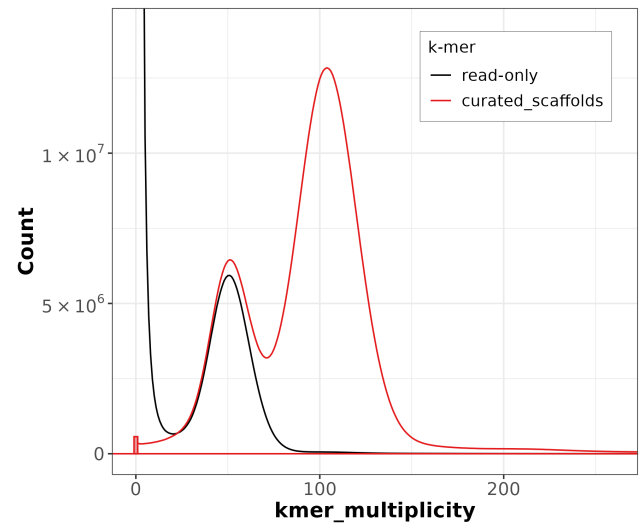


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

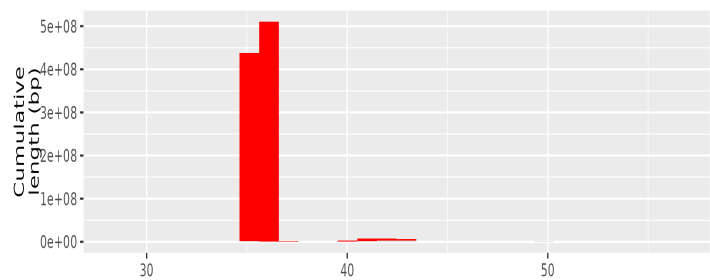


Distribution of k-mer counts per copy numbers found in asm

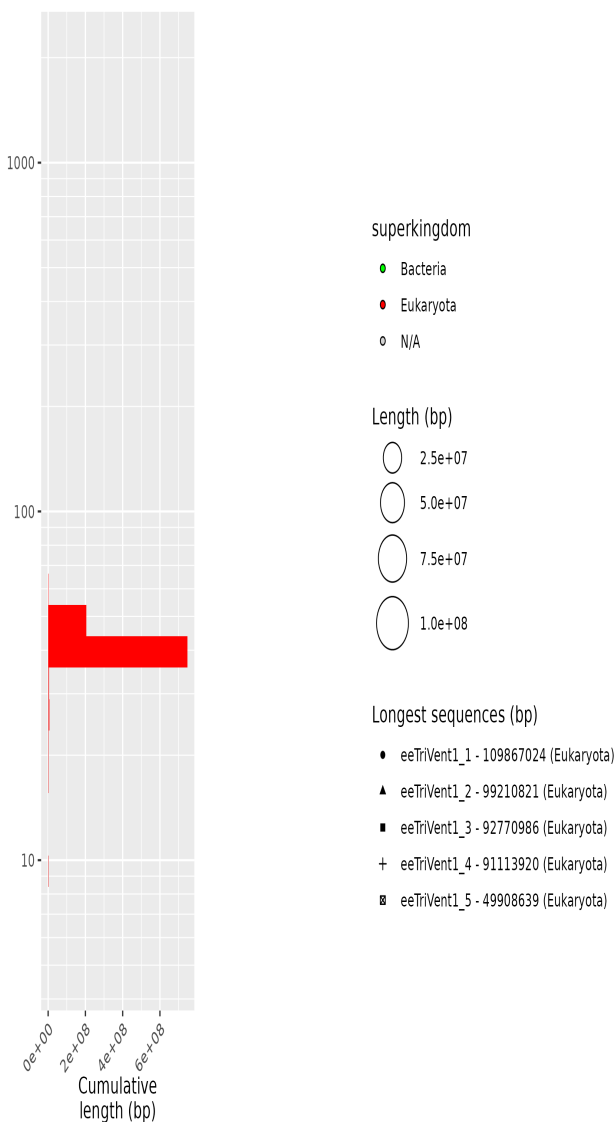
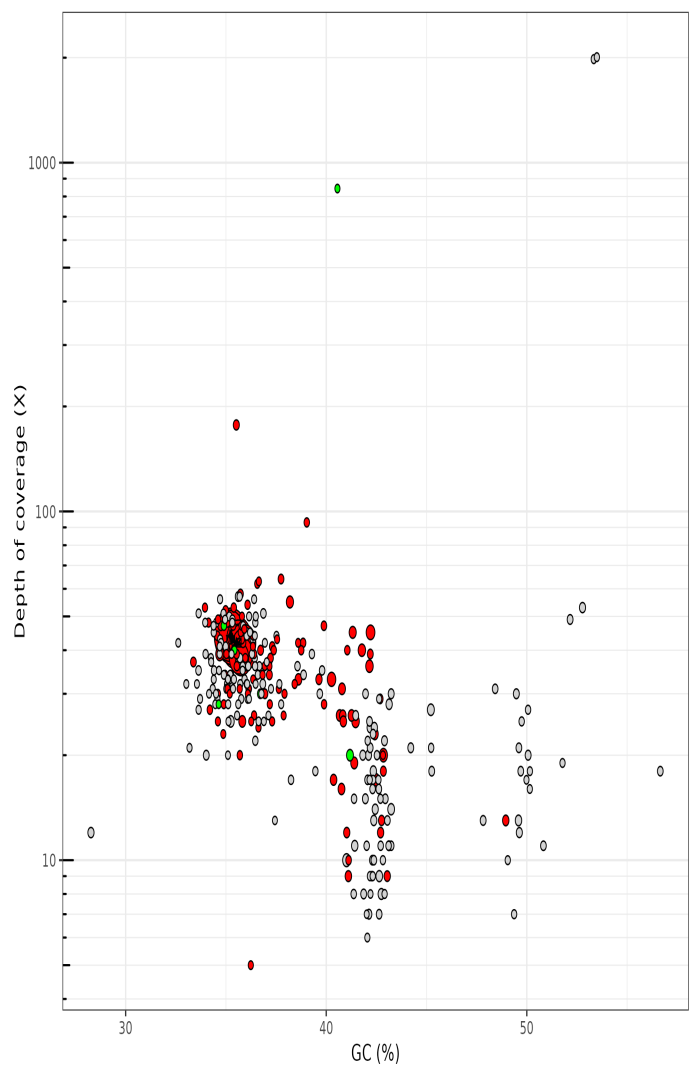


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	43	184

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: Jean-Marc Aury

Affiliation: Genoscope

Date and time: 2026-08-03 18:36:16 CEST