ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1182262
ToLID	**fRhiTri1**
Species	Rhinesomus triqueter
Class	Actinopteri
Order	Tetraodontiformes

Genome Traits	Expected	Observed
Haploid size (bp)	823,211,852	846,532,836
Haploid Number	17 (source: ancestor)	24
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q46

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

. Observed Haploid Number is different from Expected

Curator notes

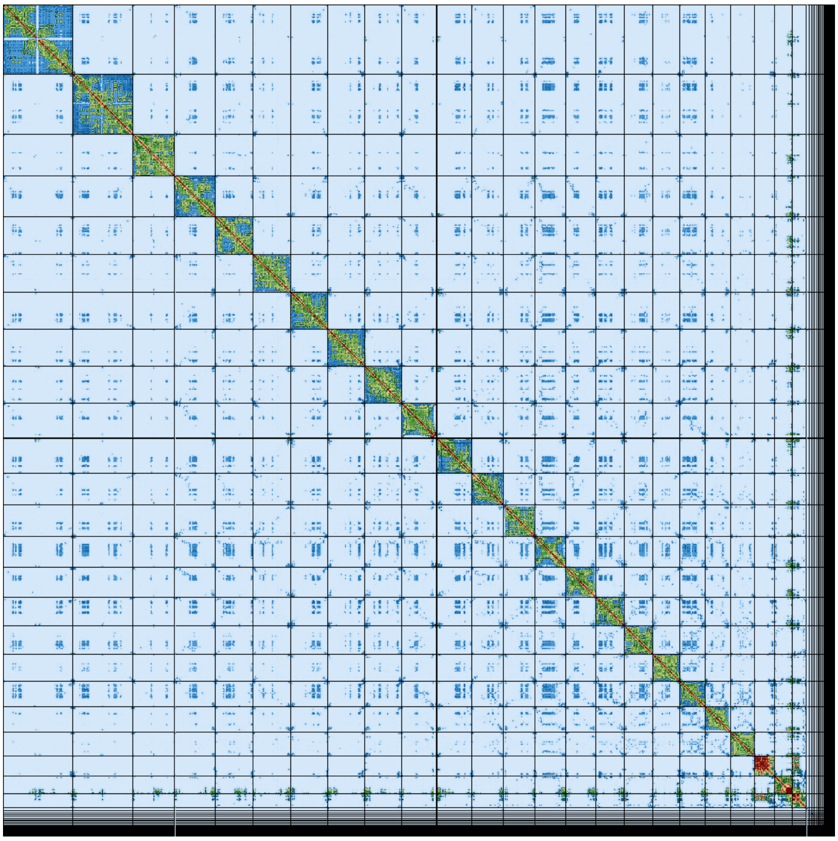
. Interventions/Gb: 18
. Contamination notes: ""
. Other observations: "The assembly of Rhinesomus triqueter (fRhiTri1) is based on 58X PacBio data and 168X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 1 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.02 Mb (with the largest being 0.02 Mb). Additionally, 262 regions totaling 22.607 Mb (with the largest being 5.035 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, 1 haplotypic regions and 254 contaminant sequences were removed, totaling 0.663Mb and 7.5Mb, respectively (with the largest being 0.663Mb and 0.093Mb). 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	854,703,673	846,532,836
GC %	41.22	41.39
Gaps/Gbp	40.95	53.16
Total gap bp	3,500	5,700
Scaffolds	369	105
Scaffold N50	35,473,571	35,473,571
Scaffold L50	10	10
Scaffold L90	22	21
Contigs	404	150
Contig N50	25,761,061	25,761,061
Contig L50	14	14
Contig L90	39	37
QV	38.2434	46.5755
Kmer compl.	93.1125	93.1047
BUSCO sing.	99.4%	99.4%
BUSCO dupl.	0.2%	0.2%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.3%	0.3%

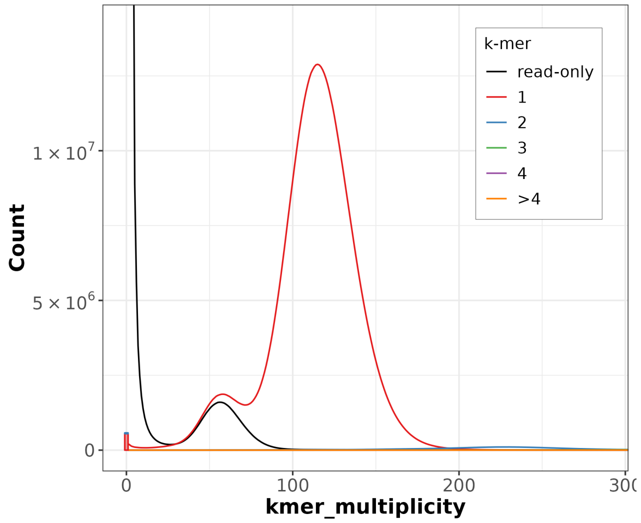
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

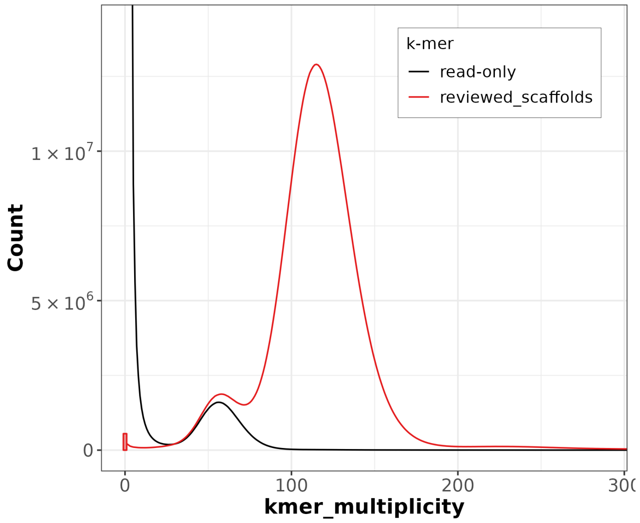


collapsed [LINK]

K-mer spectra of curated assembly

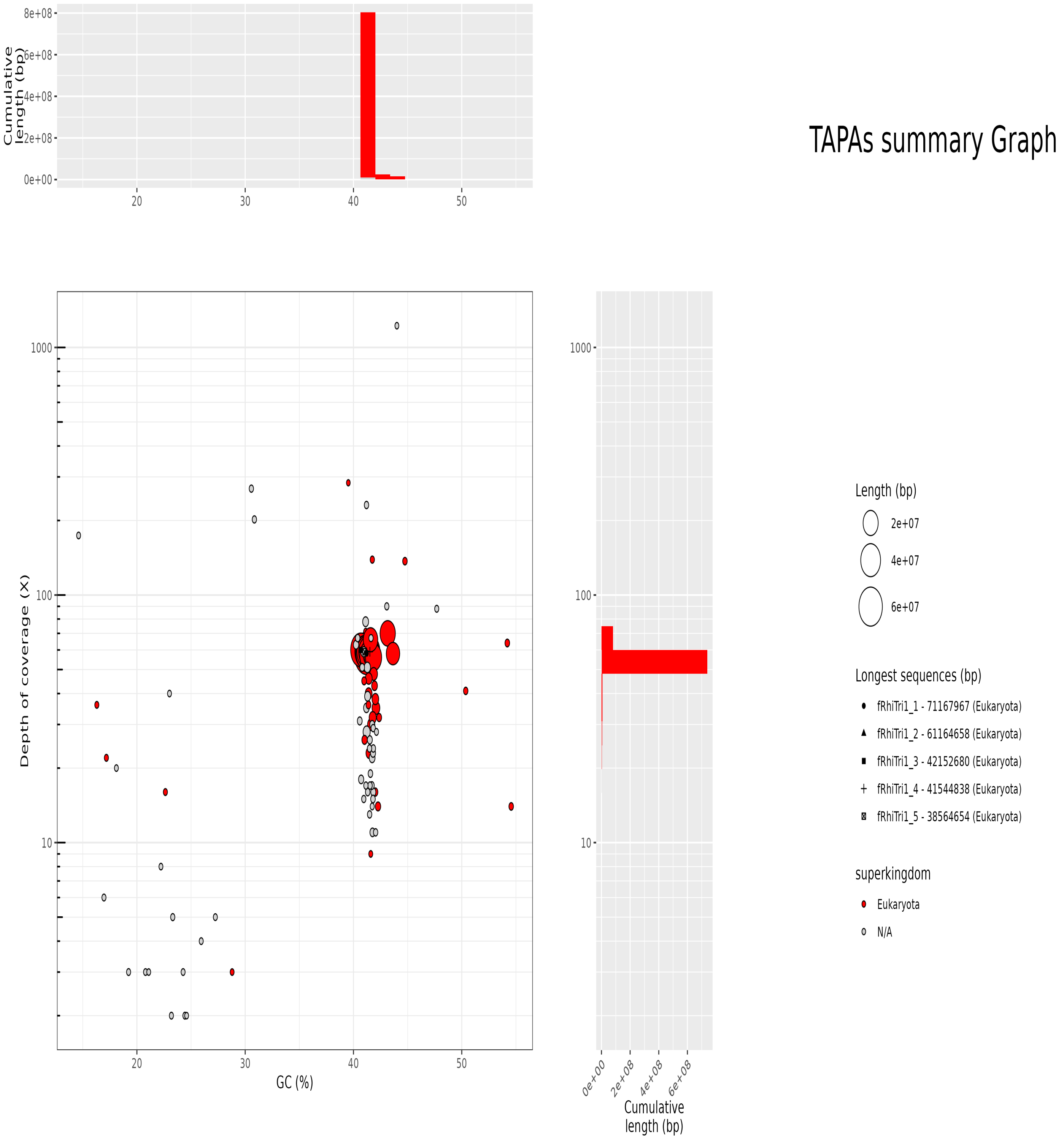


Distribution of k-mer counts per copy numbers found in asm



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	58	168

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: 2025-11-27 02:05:14 CET