ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1676591
ToLID	**xgTriHomb1**
Species	Tritonia hombergii
Class	Gastropoda
Order	Nudibranchia

Genome Traits	Expected	Observed
Haploid size (bp)	1,479,547,418	1,610,929,061
Haploid Number	11 (source: ancestor)	16
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 5.7.Q43

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

. Observed Haploid Number is different from Expected

. More than 1000 gaps/Gbp for collapsed

Curator notes

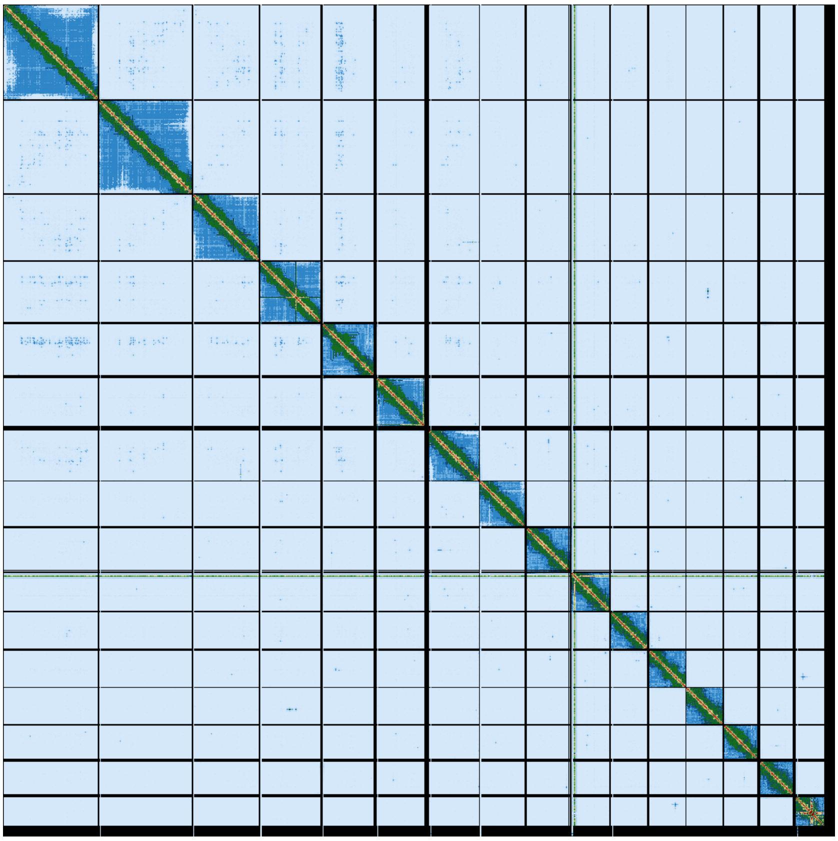
. Interventions/Gb: 600
. Contamination notes: ""
. Other observations: "The assembly of Tritonia hombergii (xgTriHomb1) is based on 27X PacBio data and 158X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 175 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 12.585 Mb (with the largest being 2.052 Mb). Additionally, 1038 regions totaling 57.34 Mb (with the largest being 0.316 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. During manual curation, 6 haplotypic regions were removed, totaling 2,378,219 pb (with the largest being 845,026 pb) and 12 contigs were identified as contaminants. 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	1,614,346,791	1,610,929,061
GC %	39.19	39.19
Gaps/Gbp	2,106.73	2,218.6
Total gap bp	355,000	395,900
Scaffolds	909	557
Scaffold N50	86,028,795	98,039,315
Scaffold L50	7	6
Scaffold L90	16	15
Contigs	4,282	4,131
Contig N50	784,129	784,129
Contig L50	605	604
Contig L90	2,099	2,094
QV	43.1977	43.2245
Kmer compl.	95.1775	95.0858
BUSCO sing.	95.6%	95.6%
BUSCO dupl.	1.2%	1.2%
BUSCO frag.	1.0%	1.1%
BUSCO miss.	2.1%	2.1%

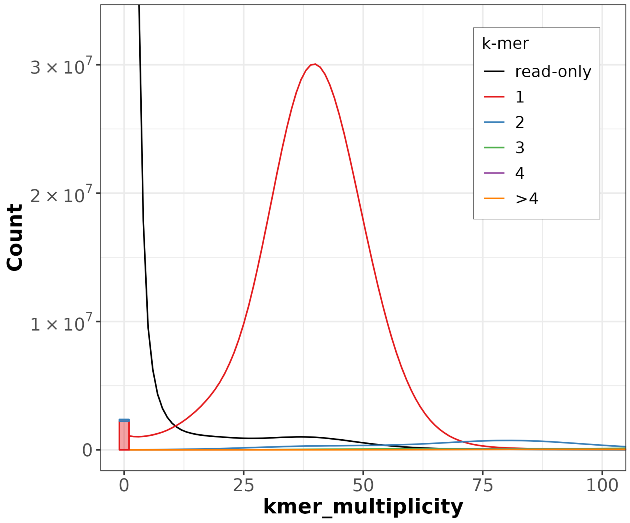
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

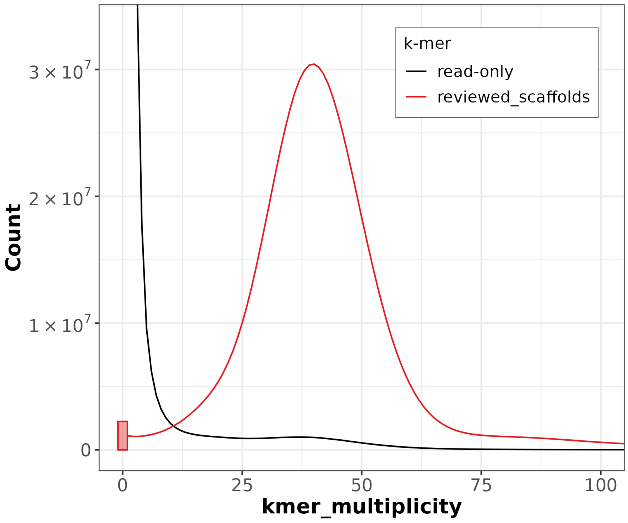


collapsed [LINK]

K-mer spectra of curated assembly

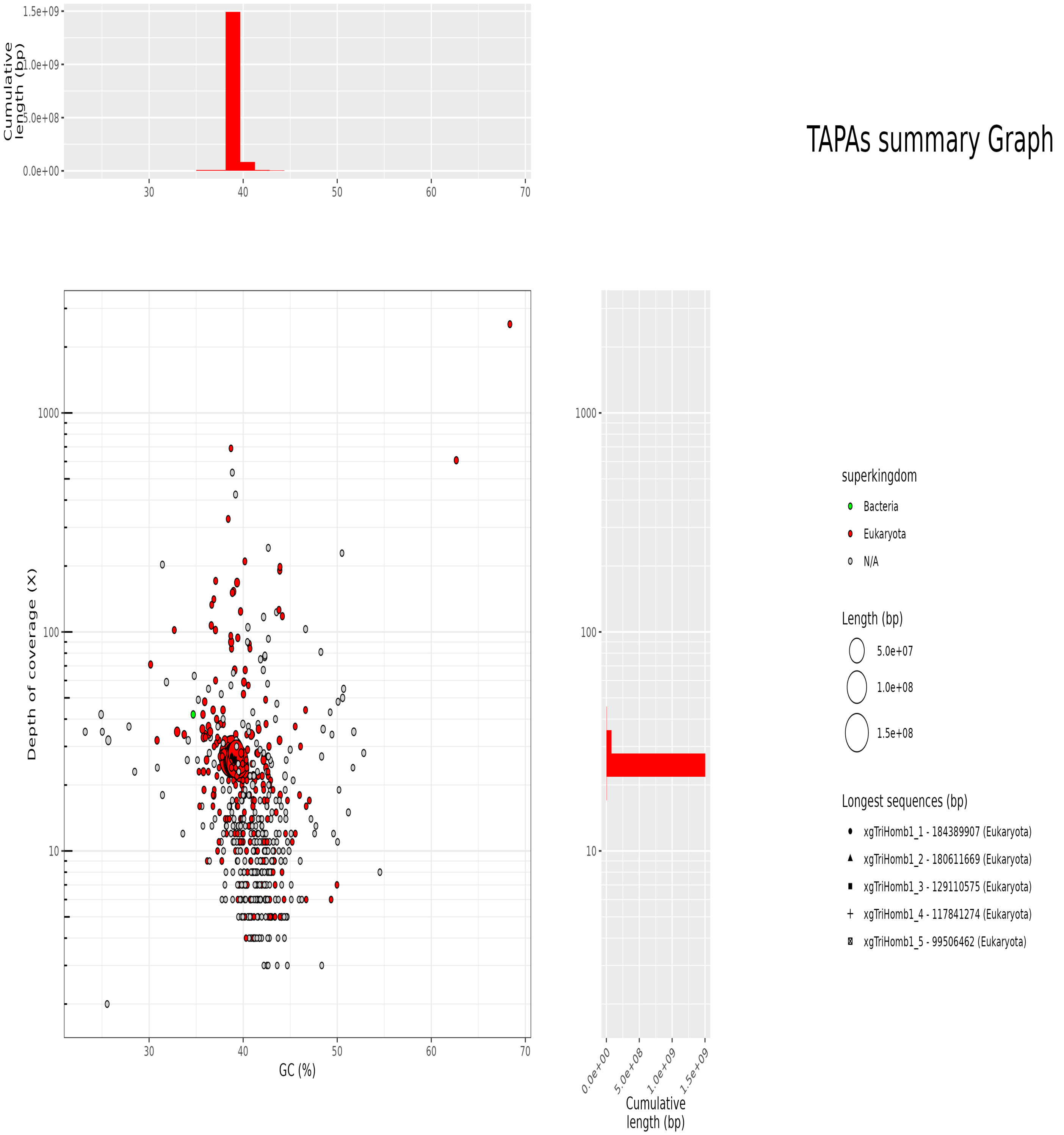


Distribution of k-mer counts per copy numbers found in asm



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	27	158

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: Lola Demirdjian
Affiliation: Genoscope

Date and time: 2025-12-02 13:04:26 CET