ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	66828
ToLID	**xbIsoAlat1**
Species	Isognomon alatus
Class	Bivalvia
Order	Pterioida

Genome Traits	Expected	Observed
Haploid size (bp)	656,992,009	669,244,133
Haploid Number	14 (source: ancestor)	14
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:

. Kmer completeness value is less than 90 for collapsed

Curator notes

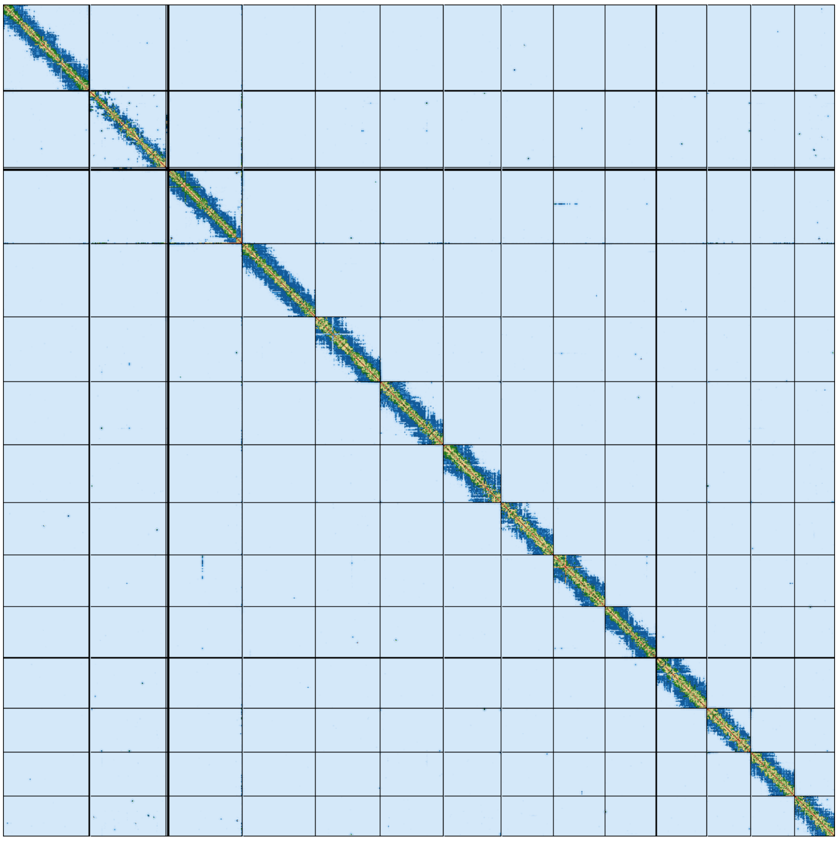
. Interventions/Gb: 87
. Contamination notes: ""
. Other observations: "The assembly of Isognomon alatus (xbIsoAlat1.1) is based on 50X PACBIO data and 142X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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 0.078 Mb (with the largest being 0.059 Mb). Additionally, 173 regions totaling 16.533 Mb (with the largest being 2.6 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 15 haplotypic regions were removed, totaling 6.583 Mb (with the largest being 1.281 Mb). 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	675,797,086	669,244,133
GC %	37	37
Gaps/Gbp	112.46	127.01
Total gap bp	7,600	11,700
Scaffolds	53	36
Scaffold N50	50,599,438	50,599,538
Scaffold L50	6	6
Scaffold L90	13	13
Contigs	129	121
Contig N50	13,403,000	13,403,000
Contig L50	16	16
Contig L90	46	45
QV	47.6466	47.6699
Kmer compl.	75.3107	74.8158
BUSCO sing.	96.9%	97.1%
BUSCO dupl.	1.4%	1.1%
BUSCO frag.	0.5%	0.5%
BUSCO miss.	1.2%	1.2%

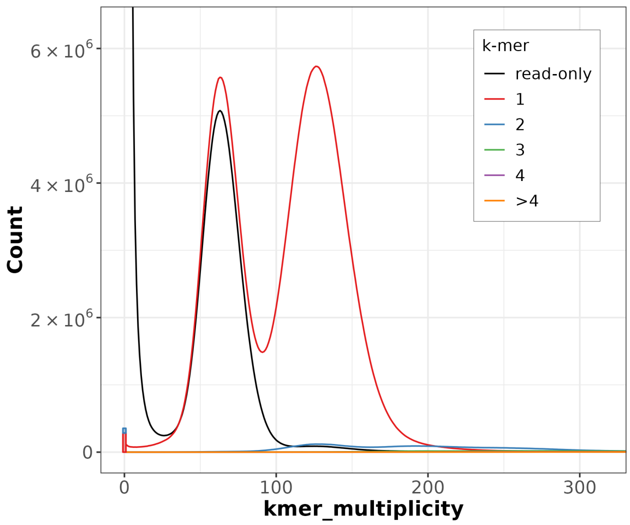
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

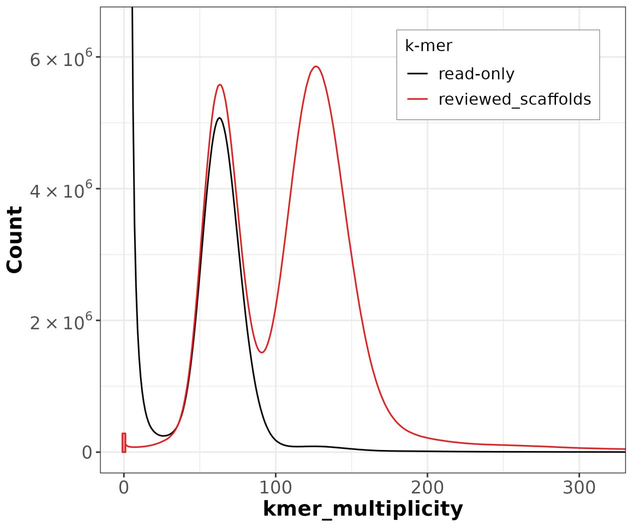


collapsed [LINK]

K-mer spectra of curated assembly

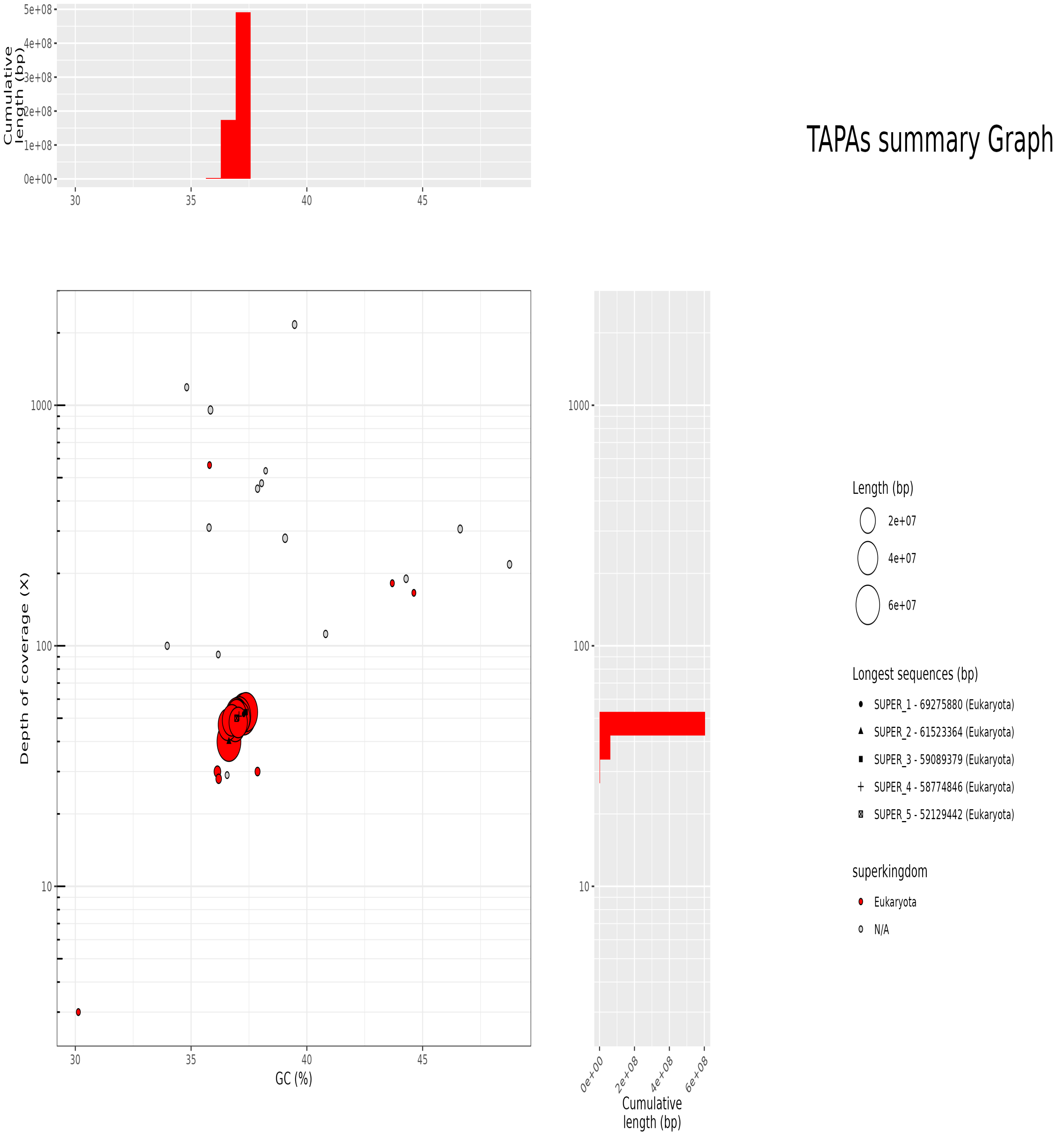


Distribution of k-mer counts per copy numbers found in asm



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	50	142

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: Adama Ndar
Affiliation: Genoscope

Date and time: 2025-11-18 14:16:13 CET