ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	256063
ToLID	**xpIscRiss1**
Species	Ischnochiton rissoi
Class	Polyplacophora
Order	Chitonida

Genome Traits	Expected	Observed
Haploid size (bp)	1,494,612,057	1,496,897,274
Haploid Number	9 (source: ancestor)	12
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q66

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
. BUSCO single copy value is less than 90% for collapsed
. Assembly length loss > 3% for collapsed
. Not 90% of assembly in chromosomes for collapsed

Curator notes

. Interventions/Gb: 36
. Contamination notes: ""
. Other observations: "The assembly of \'Ischnochiton rissoi\' (xpIscRiss1.1) is based on 36X PacBio data and 127X 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, 12 contigs were identified as contaminants (bacterial), totaling 326.7 Kb (with the largest being 119 Kb). Additionally, 428 regions totaling 759.4 Mb (with the largest being 45.08 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, 24 haplotypic regions were removed, totaling 81.4 Mb (with the largest being 13.5 Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. Chromosome-scale scaffolds

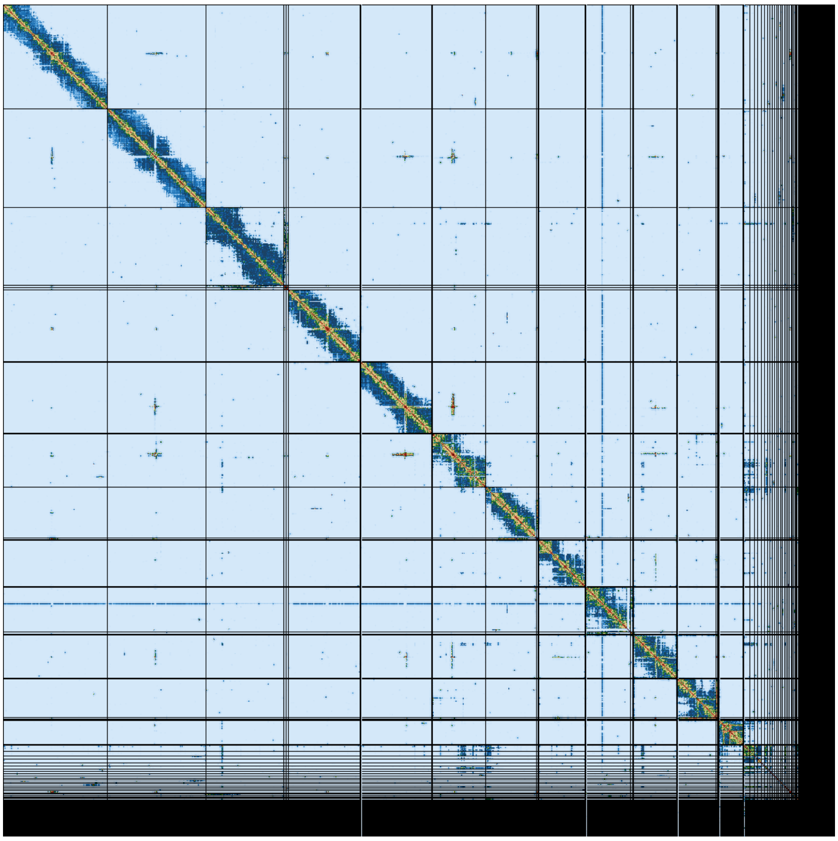
confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,578,474,016	1,496,897,274
GC %	39.52	39.53
Gaps/Gbp	94.39	78.16
Total gap bp	14,900	12,400
Scaffolds	330	317
Scaffold N50	131,887,338	127,715,546
Scaffold L50	5	5
Scaffold L90	22	20
Contigs	479	434
Contig N50	15,451,000	21,887,000
Contig L50	29	22
Contig L90	118	91
QV	49.0409	66.4181
Kmer compl.	64.3149	62.2459
BUSCO sing.	72.0%	74.4%
BUSCO dupl.	6.3%	3.6%
BUSCO frag.	4.6%	4.7%
BUSCO miss.	17.1%	17.3%

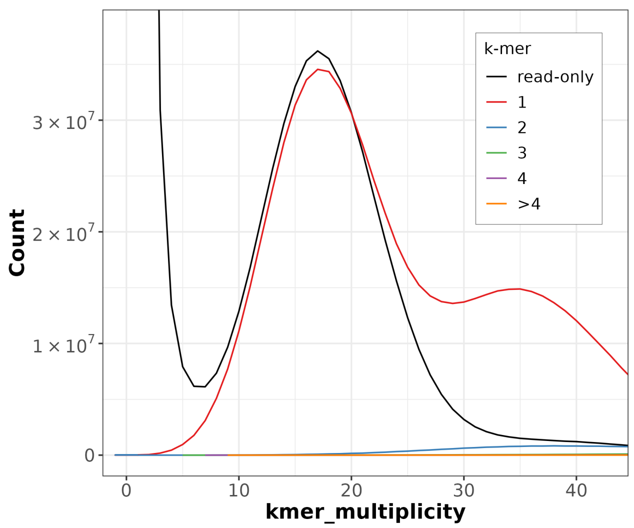
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: mollusca_odb10 (genomes:7, BUSCOs:5295)

HiC contact map of curated assembly

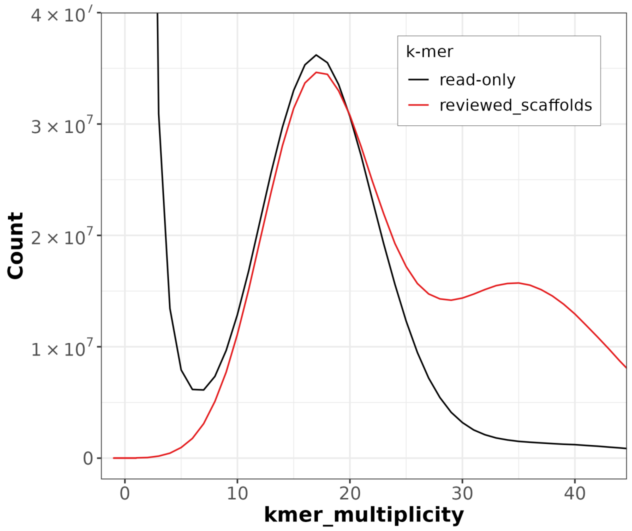


collapsed [LINK]

K-mer spectra of curated assembly

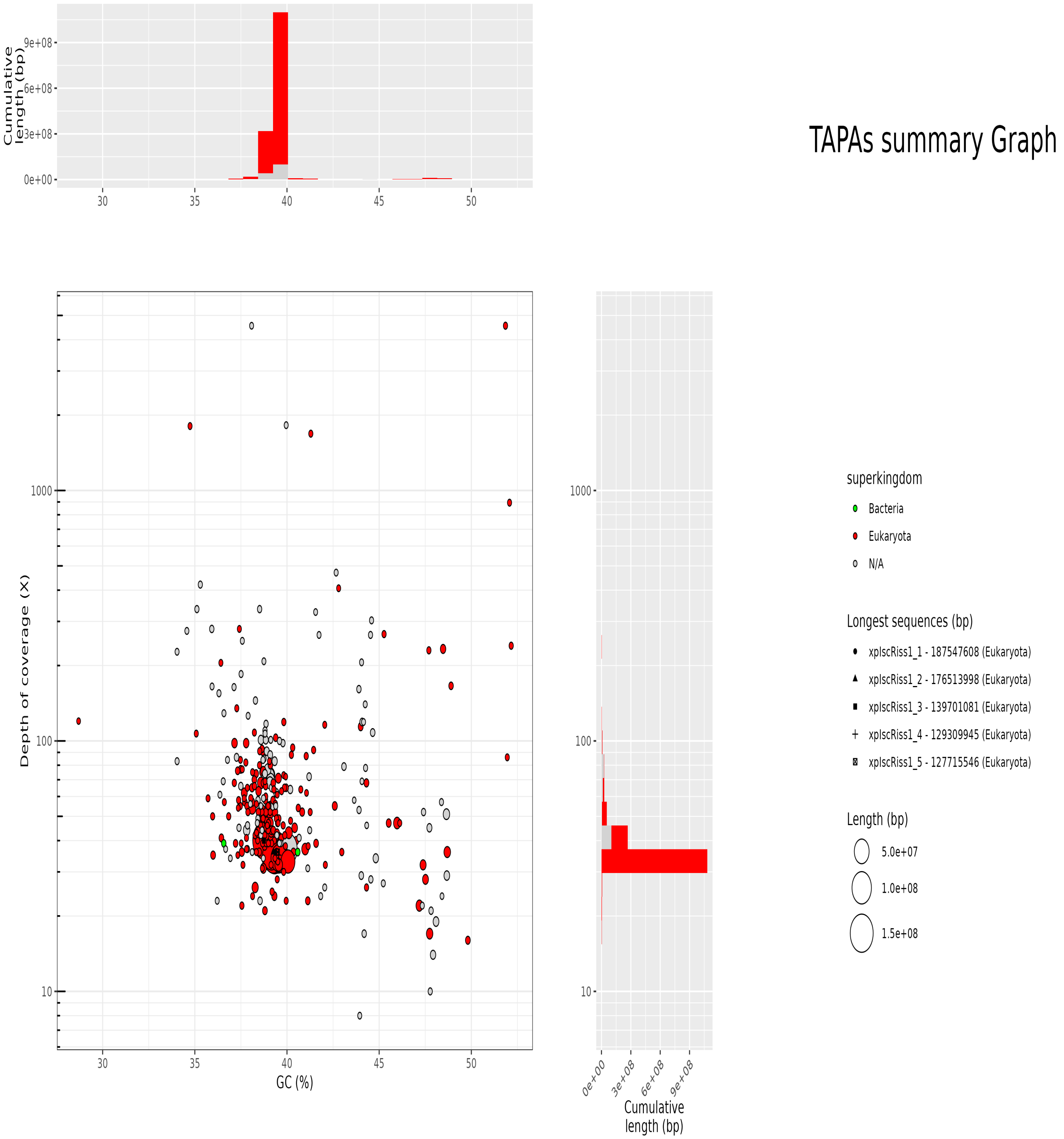


Distribution of k-mer counts per copy numbers found in asm



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	PACBIO Hifi	Arima
Coverage	36	127

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: Simone Duprat
Affiliation: Genoscope

Date and time: 2025-04-05 17:55:25 CEST