ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1676568
ToLID	**xbParScab1**
Species	Parvicardium scabrum
Class	Bivalvia
Order	Cardiida

Genome Traits	Expected	Observed
Haploid size (bp)	948,753,578	972,009,762
Haploid Number	26 (source: ancestor)	18
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q50

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
. Assembly length loss > 3% for collapsed

Curator notes

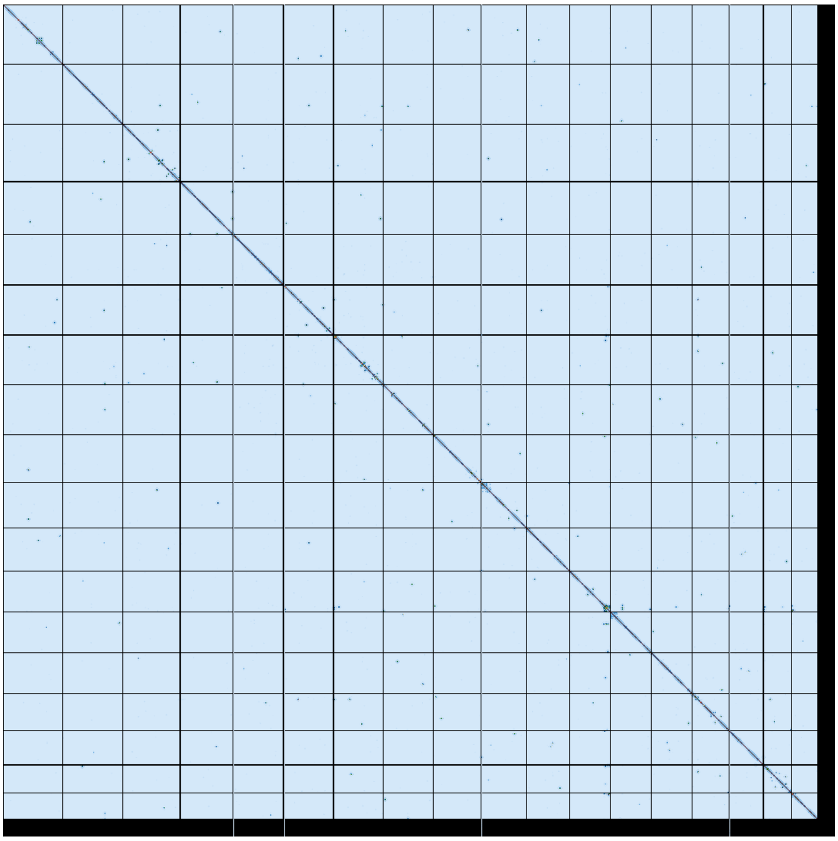
. Interventions/Gb: 639
. Contamination notes: ""
. Other observations: "The assembly of Parvicardium scabrum (xbParScab1) is based on 51X PacBio data and 134X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included the following steps: initial 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, 9 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.1 Mb (with the largest being 0.492 Mb). Additionally, 521 regions totaling 169 Mb (with the largest being 2.6 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 278 supplementary haplotypic regions were removed, totaling 380 Mb (with the largest being 7.4 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	1,354,208,090	972,009,762
GC %	36.53	36.55
Gaps/Gbp	629.89	731.47
Total gap bp	85,300	98,300
Scaffolds	687	388
Scaffold N50	22,633,511	57,944,666
Scaffold L50	16	8
Scaffold L90	70	16
Contigs	1,540	1,099
Contig N50	2,747,083	2,870,191
Contig L50	161	105
Contig L90	503	354
QV	50.4382	50.3892
Kmer compl.	78.2847	60.3771
BUSCO sing.	58.8%	95.7%
BUSCO dupl.	40.3%	2.1%
BUSCO frag.	0.5%	0.7%
BUSCO miss.	0.4%	1.6%

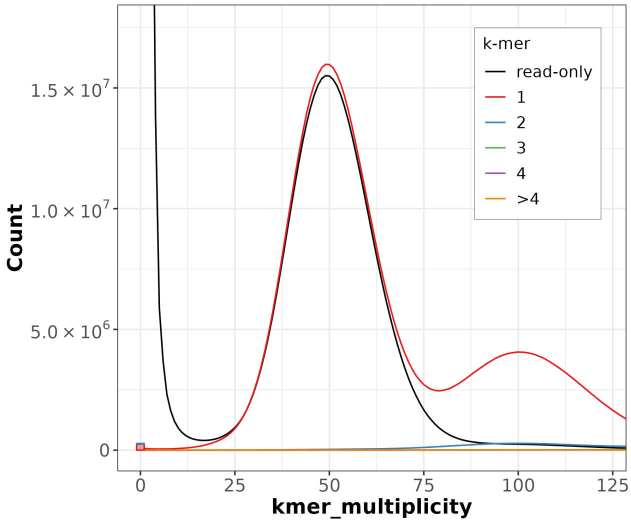
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

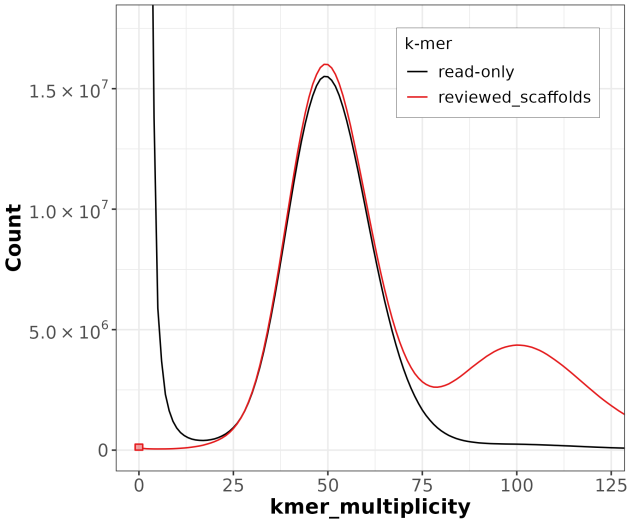


collapsed [LINK]

K-mer spectra of curated assembly

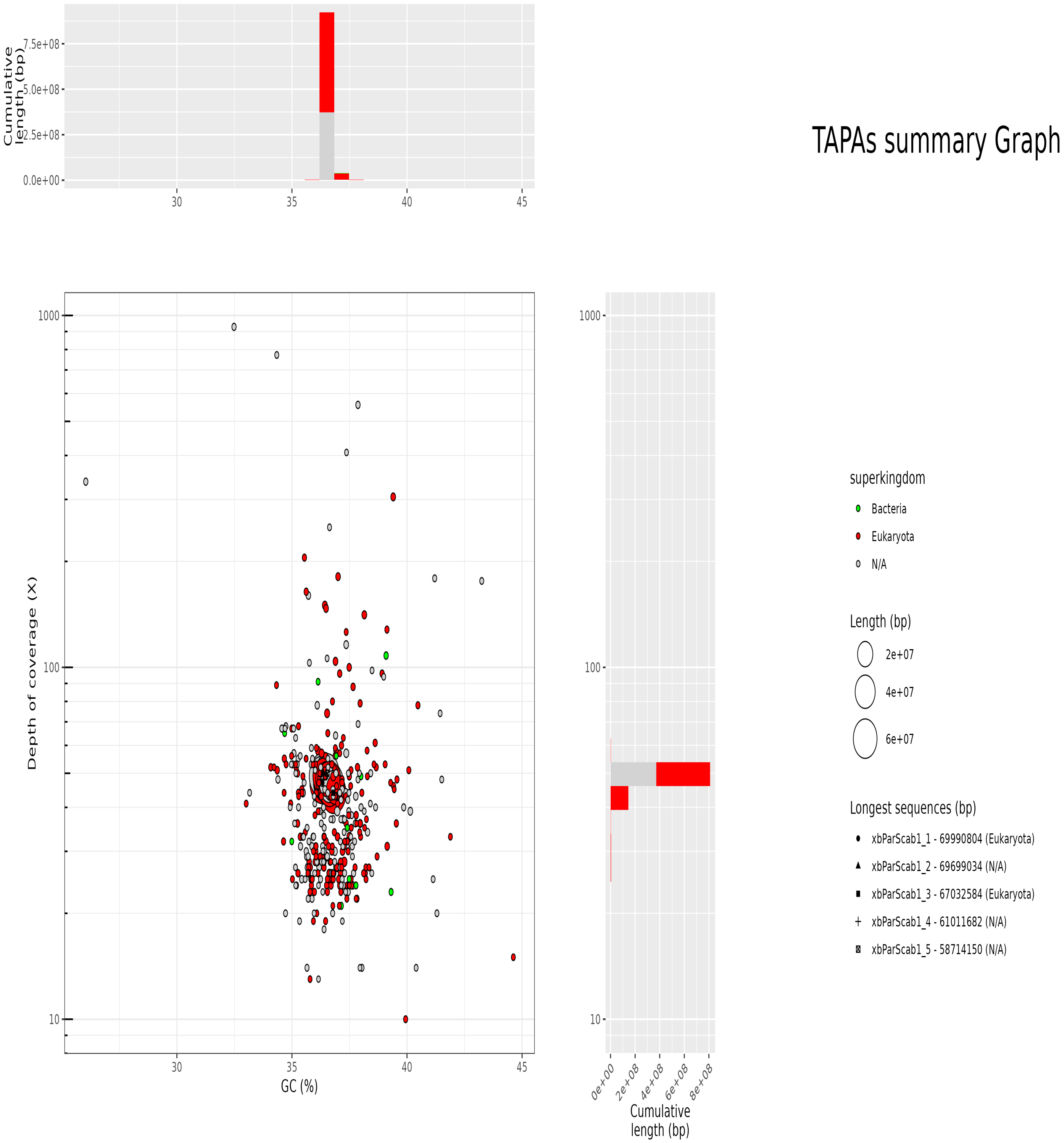


Distribution of k-mer counts per copy numbers found in asm



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	51	134

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: Caroline Menguy
Affiliation: Genoscope

Date and time: 2025-11-18 16:00:06 CET