ERGA Assembly Report
v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1933096
ToLID	**eeGraAcut1**
Species	Gracilechinus acutus
Class	Echinoidea
Order	Camarodonta

Genome Traits	Expected	Observed
Haploid size (bp)	820,704,990	1,125,641,135
Haploid Number	16 (source: ancestor)	19
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q55

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

. Observed Haploid size (bp) has >20% difference with Expected
. Observed Haploid Number is different from Expected

. Kmer completeness value is less than 90 for collapsed
. More than 1000 gaps/Gbp for collapsed

Curator notes

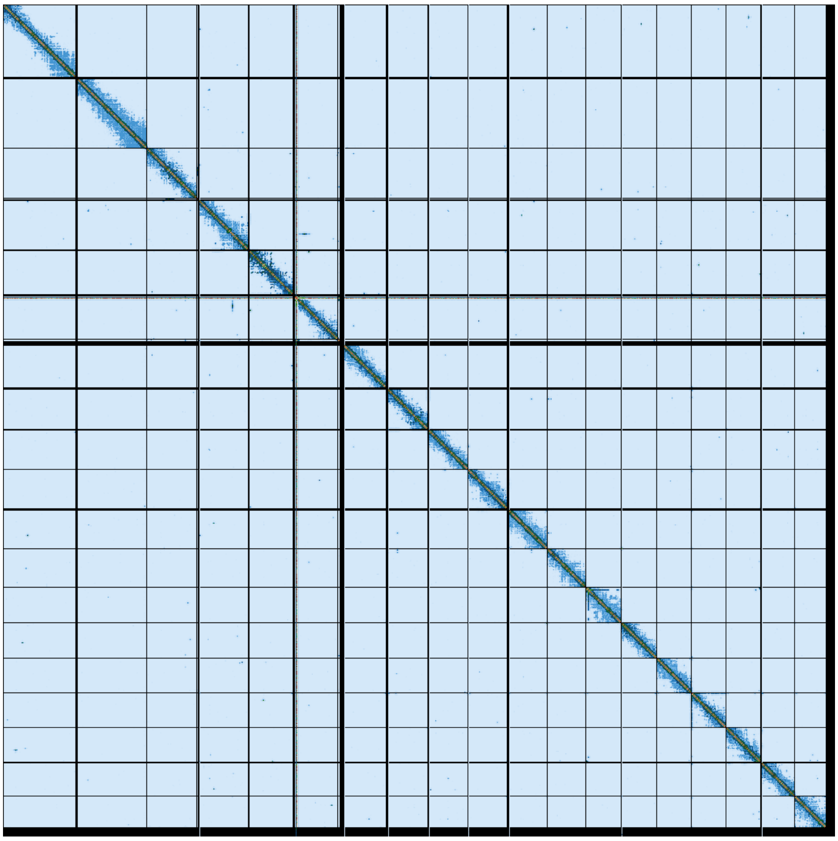
. Interventions/Gb: 209
. Contamination notes: ""
. Other observations: "The assembly of Gracilechinus acutus (eeGraAcut1) is based on 40X PacBio data and Arima HighCoverage Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included the following steps: 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, 11 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 501 Kb (with the largest being 124 Kb). Additionally, 1,621 regions totaling 1.5 Gb (with the largest being 6.4 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, 65 haplotypic regions were removed,totaling 24 Mb (with the largest being 876 Kb) "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,150,244,363	1,125,641,135
GC %	37.99	37.99
Gaps/Gbp	1,277.12	1,293.49
Total gap bp	146,900	154,700
Scaffolds	368	239
Scaffold N50	53,196,899	53,305,607
Scaffold L50	7	9
Scaffold L90	16	18
Contigs	1,837	1,695
Contig N50	1,168,000	1,210,738
Contig L50	306	292
Contig L90	1,011	959
QV	41.0749	55.7334
Kmer compl.	74.4505	74.377
BUSCO sing.	94.4%	95.4%
BUSCO dupl.	1.9%	0.8%
BUSCO frag.	1.7%	1.7%
BUSCO miss.	2.0%	2.1%

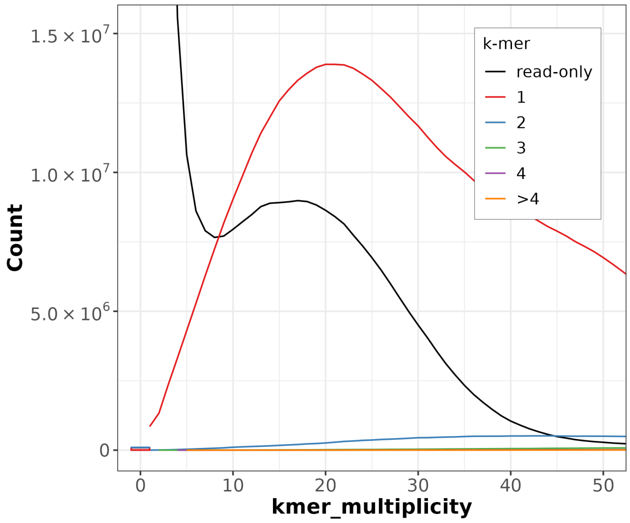
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

HiC contact map of curated assembly

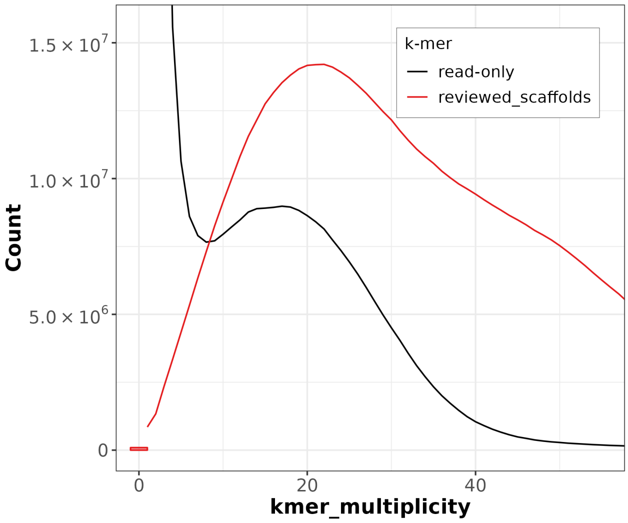


collapsed [LINK]

K-mer spectra of curated assembly

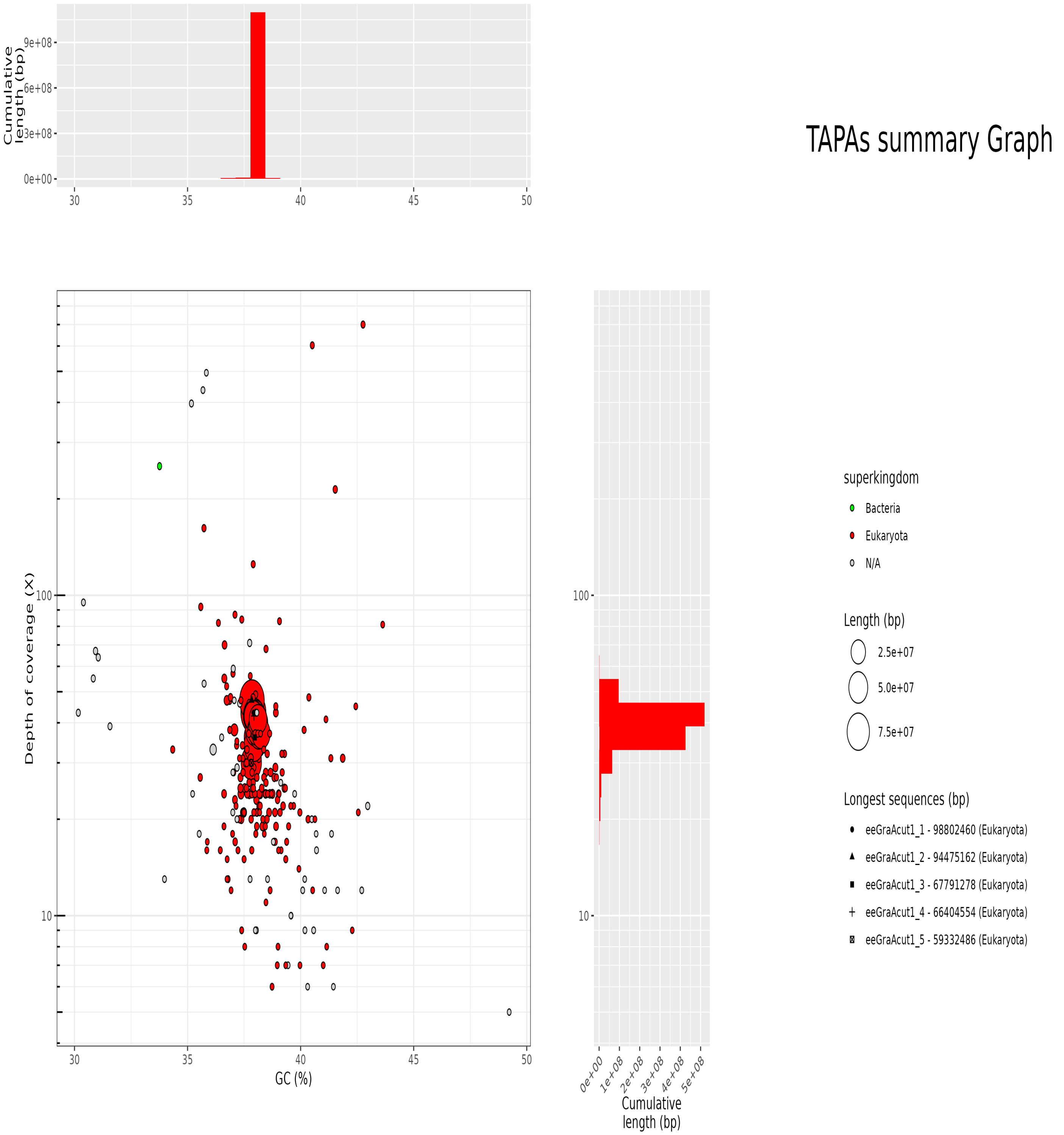


Distribution of k-mer counts per copy numbers found in asm



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	42	211

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: 2025-02-16 05:24:42 CET