ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	3258647
ToLID	**wjParGratt1**
Species	Parasalenia gratiosa
Class	Echinoidea
Order	Camarodonta

Genome Traits	Expected	Observed
Haploid size (bp)	930,920,920	996,298,710
Haploid Number	16 (source: ancestor)	21
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q44

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

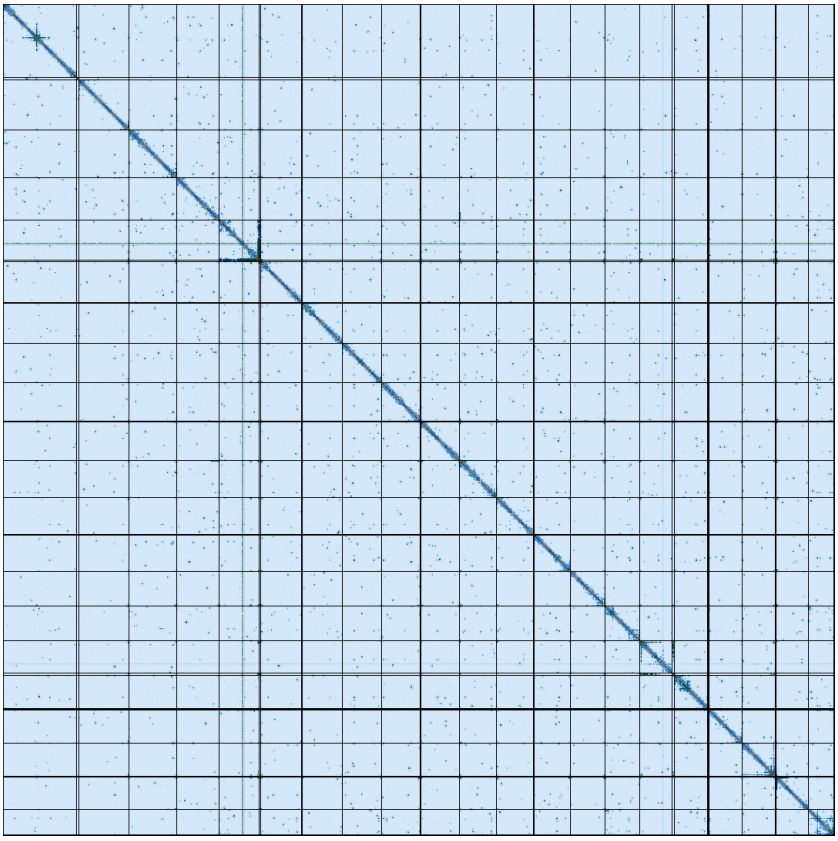
. Interventions/Gb: 56
. Contamination notes: ""
. Other observations: "The assembly of Parasalenia gratiosa (wjParGratt1) is based on 56.02X PacBio data and 122.52X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: 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, 41 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.987 Mb (with the largest being 0.264 Mb). Additionally, 289 regions totaling 118.077 Mb (with the largest being 1.525 Mb) were identified as haplotypic duplications and removed. 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 7.981Mb (with the largest being 1.647Mb). The telomeric pattern AACCCT was identified with TelFinder and used to generate the Pretext Telomeres tracks. 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	1,155,160,492	996,298,710
GC %	37.98	37.96
Gaps/Gbp	193.05	207.77
Total gap bp	22,300	27,200
Scaffolds	86	61
Scaffold N50	49,160,839	46,216,764
Scaffold L50	10	10
Scaffold L90	22	19
Contigs	309	268
Contig N50	7,054,441	7,288,556
Contig L50	52	44
Contig L90	167	138
QV	44.0509	44.114
Kmer compl.	68.761	61.819
BUSCO sing.	82.6%	98.1%
BUSCO dupl.	17.3%	0.9%
BUSCO frag.	0.1%	0.4%
BUSCO miss.	0.0%	0.6%

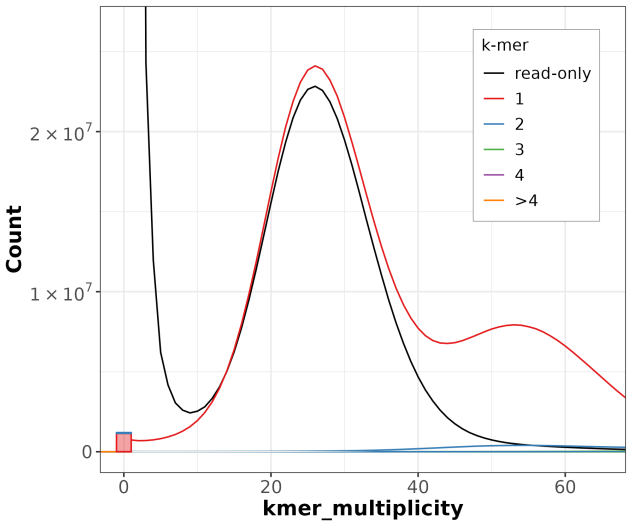
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

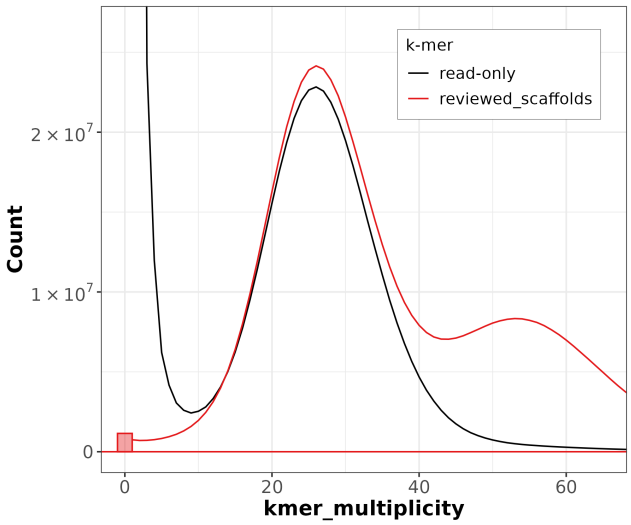


collapsed [LINK]

K-mer spectra of curated assembly

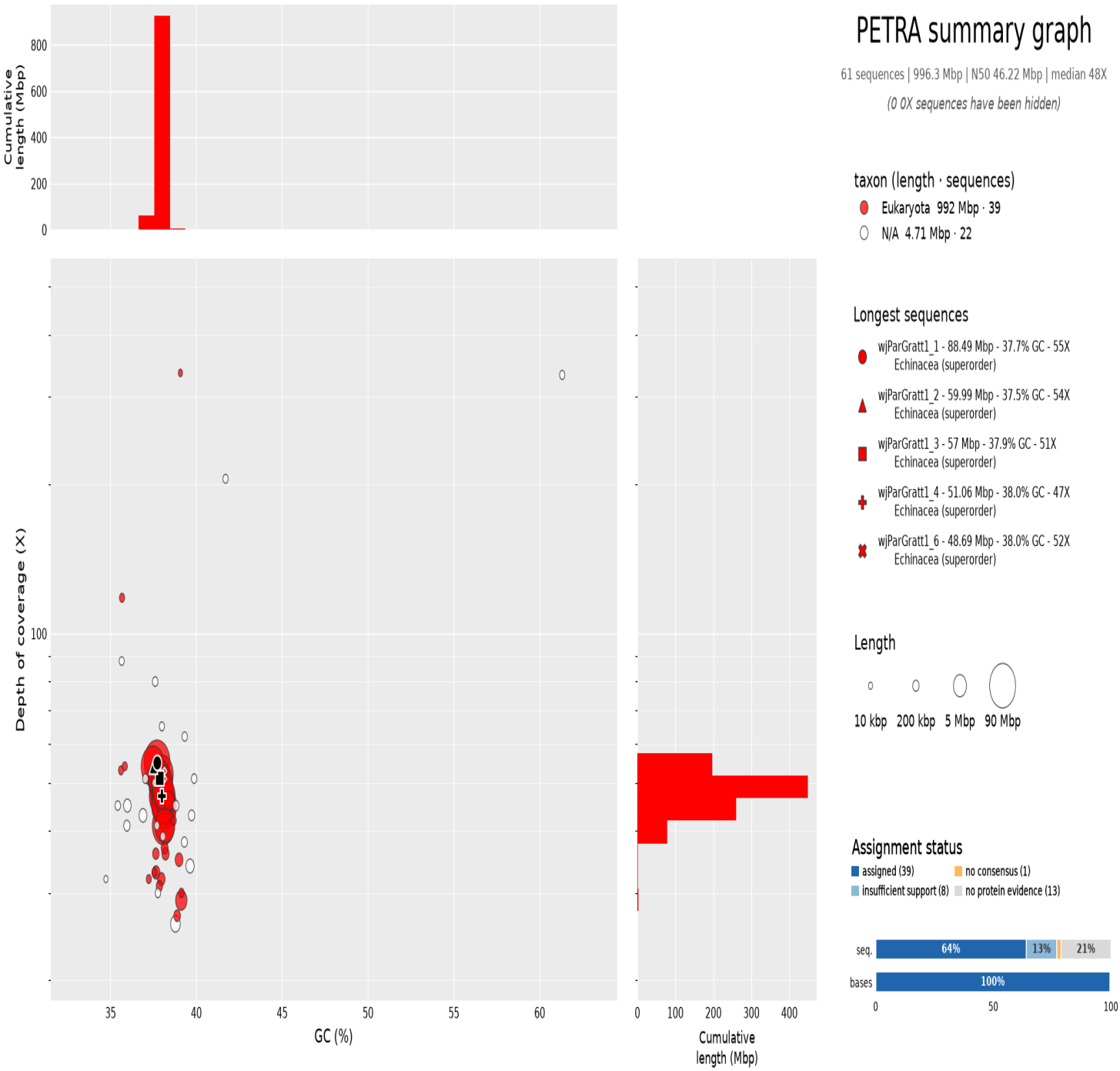


Distribution of k-mer counts per copy numbers found in asm



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	64	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:* 0.2.4
  |_ *key param:* NA
- **PretextView**
  |_ *ver:* 1.0.7
  |_ *key param:* NA

Submitter: Eleanore Lacoste
Affiliation: Genoscope

Date and time: 2026-10-08 08:10:01 CEST