ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	2759737
ToLID	**jaPteGris1.1**
Species	Pteroeides griseum
Class	Anthozoa
Order	Scleralcyonacea

Genome Traits	Expected	Observed
Haploid size (bp)	495,096,660	477,838,095
Haploid Number	6 (source: ancestor)	20
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q53

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

Curator notes

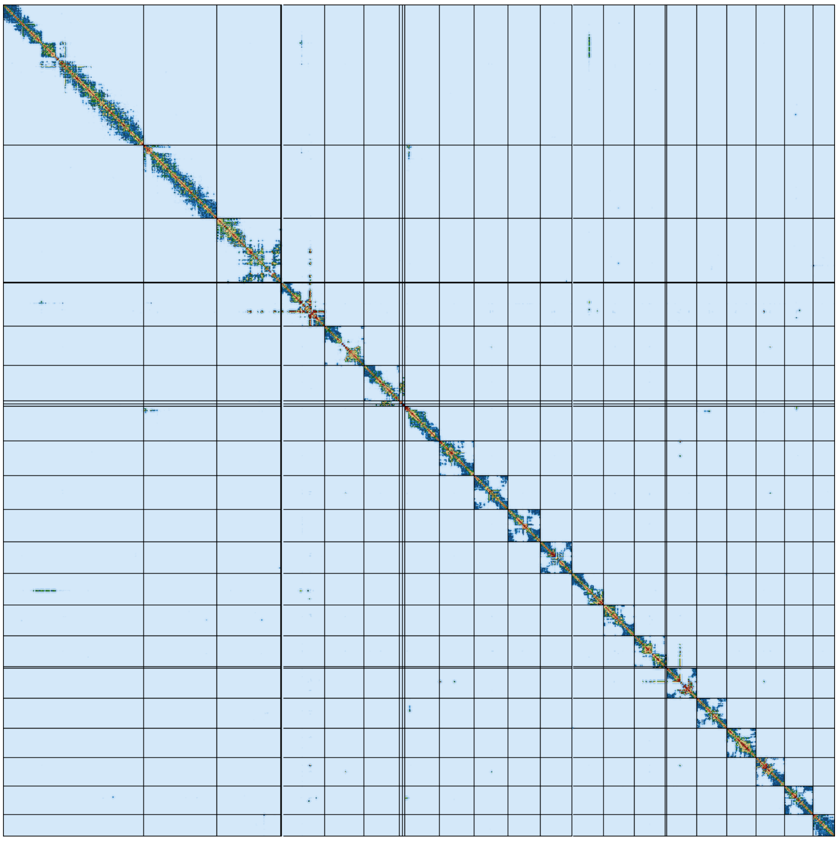
. Interventions/Gb: 314
. Contamination notes: ""
. Other observations: "The assembly of Pteroeides griseum (jaPteGris1.1) is based on 61X PacBio data and Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. No contigs were found to be contaminants (bacterial, archaeal, or viral). Additionally, 578 regions totaling 217 Mb (with the largest being 4.46 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, 35 haplotypic regions were removed, totaling 7.65 Mb (with the largest being 1.88 Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. Chromosome 3 contains a large centromeric region where the organization from 30 Mb to the end remains uncertain. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	485,525,557	477,838,095
GC %	35.5	35.49
Gaps/Gbp	436.64	571.32
Total gap bp	21,200	36,300
Scaffolds	125	37
Scaffold N50	18,772,541	20,131,959
Scaffold L50	8	7
Scaffold L90	20	18
Contigs	337	310
Contig N50	3,220,859	3,343,201
Contig L50	45	41
Contig L90	143	136
QV	45.7426	53.2862
Kmer compl.	82.3605	82.3368
BUSCO sing.	86.0%	86.1%
BUSCO dupl.	0.9%	0.8%
BUSCO frag.	4.9%	4.9%
BUSCO miss.	8.2%	8.2%

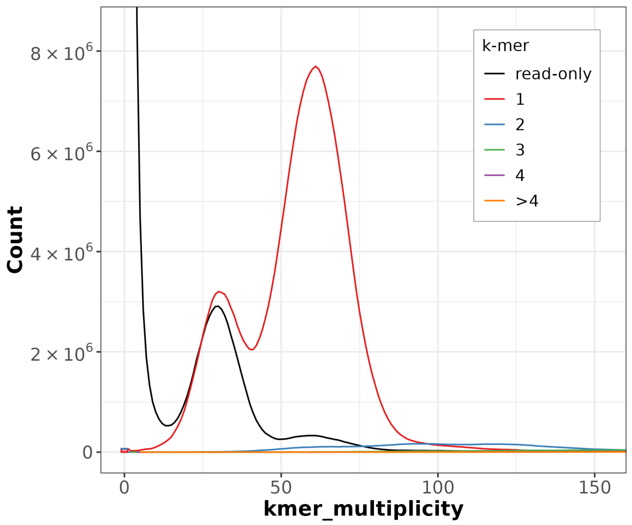
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

HiC contact map of curated assembly

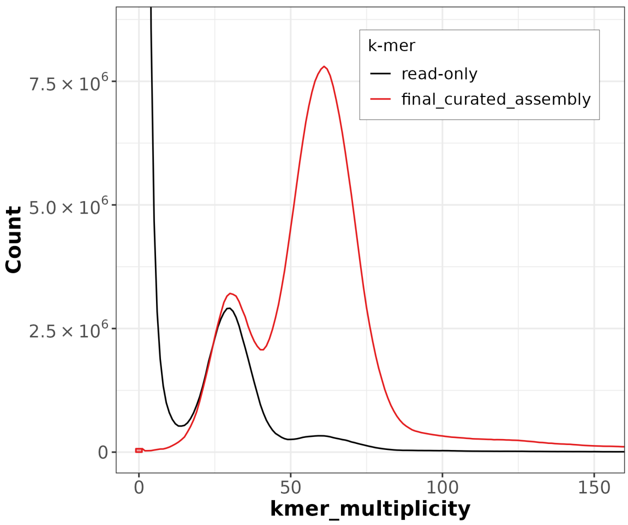


collapsed [LINK]

K-mer spectra of curated assembly

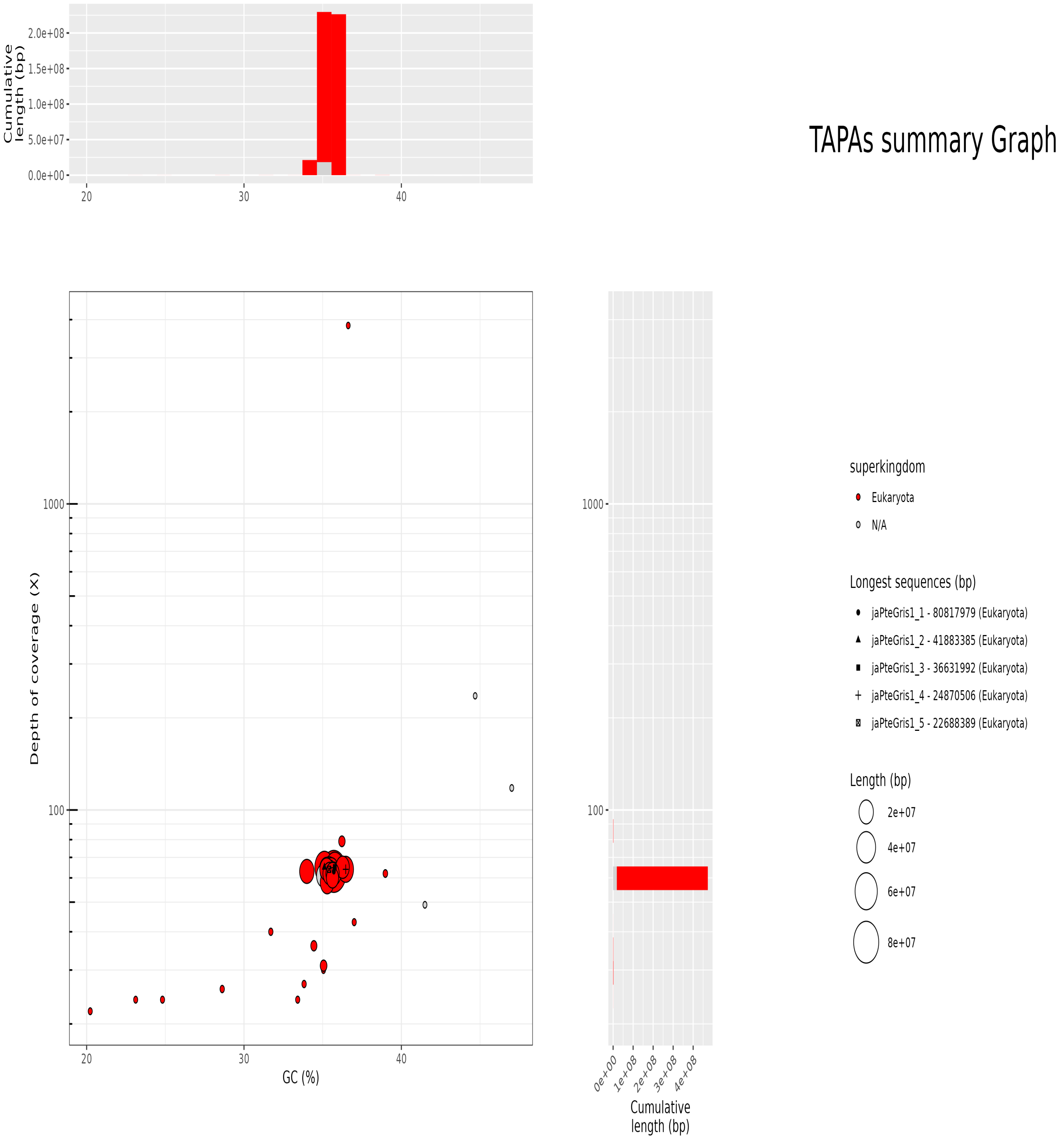


Distribution of k-mer counts per copy numbers found in asm



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	61	99

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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2024-12-17 21:12:35 CET