

ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	77183
ToLID	odPacJohn2
Species	Pachymatisma johnstonia
Class	Demospongiae
Order	Tetractinellida

Genome Traits	Expected	Observed
Haploid size (bp)	495,469,631	159,956,449
Haploid Number	5 (source: ancestor)	21
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.6.Q45

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
- . BUSCO single copy value is less than 90% for collapsed
- . Assembly length loss > 3% for collapsed
- . More than 1000 gaps/Gbp for collapsed

Curator notes

- . Interventions/Gb: 919
- . Contamination notes: ""
- . Other observations: "The assembly of Pachymatisma johnstonia (odPacJohn2) is based on 67X PacBio data and 379X 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, 19716 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1173.685 Mb (with the largest being 4.808 Mb). Additionally, 164 regions totaling 8.063 Mb (with the largest being 0.163 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, 15 haplotypic regions and 79 contaminant sequences were removed, totaling 1.73 Mb and 4.03 Mb, respectively (with

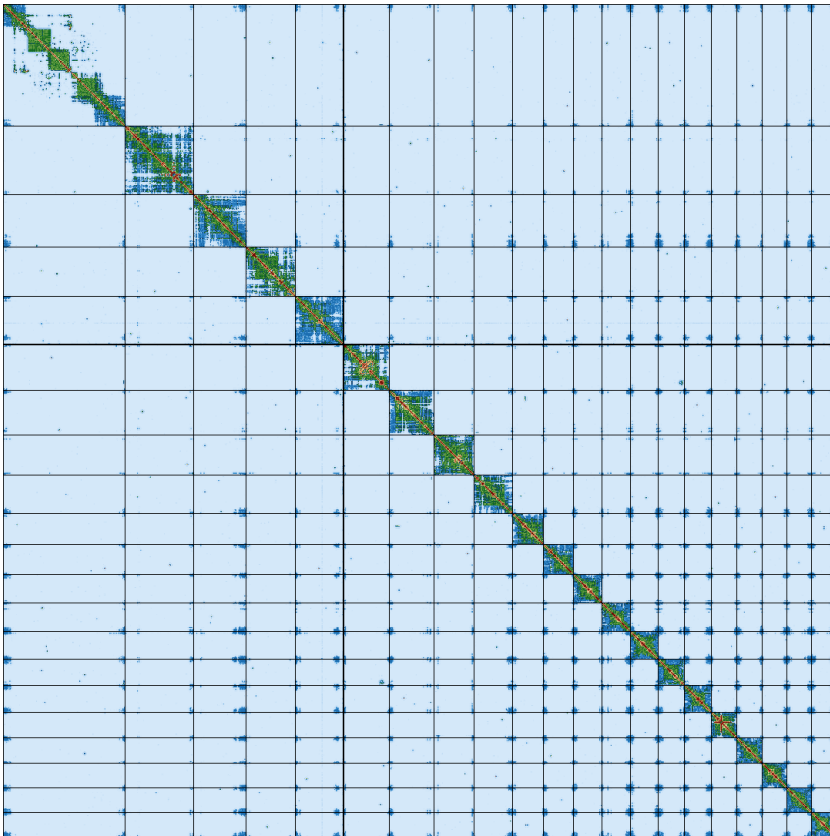
the largest being 0.562 Mb and 0.689 Mb). During manual curation, 15 haplotypic regions were removed, totaling 1.73Mb, (with the largest being 0.562Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	165,712,830	159,956,449
GC %	44.05	43.78
Gaps/Gbp	1,816.4	2,163.09
Total gap bp	30,100	43,800
Scaffolds	165	22
Scaffold N50	7,693,635	8,629,306
Scaffold L50	9	7
Scaffold L90	21	18
Contigs	466	368
Contig N50	1,435,604	1,952,367
Contig L50	27	25
Contig L90	174	155
QV	43.6864	45.2888
Kmer compl.	20.3785	12.3915
BUSCO sing.	49.9%	50.7%
BUSCO dupl.	1.2%	0.1%
BUSCO frag.	29.0%	28.6%
BUSCO miss.	19.9%	20.5%

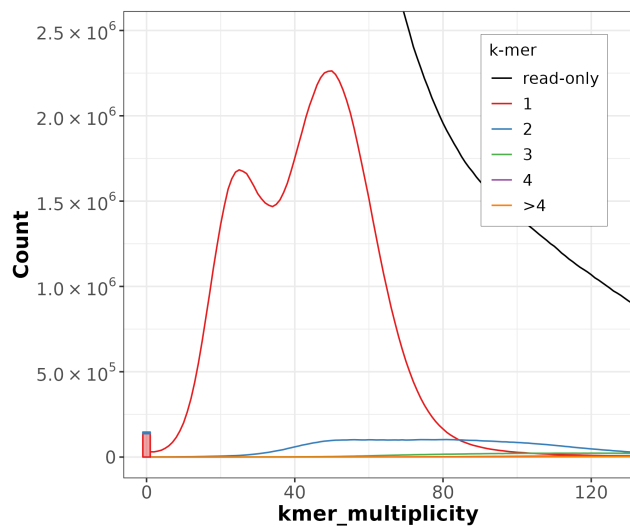
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

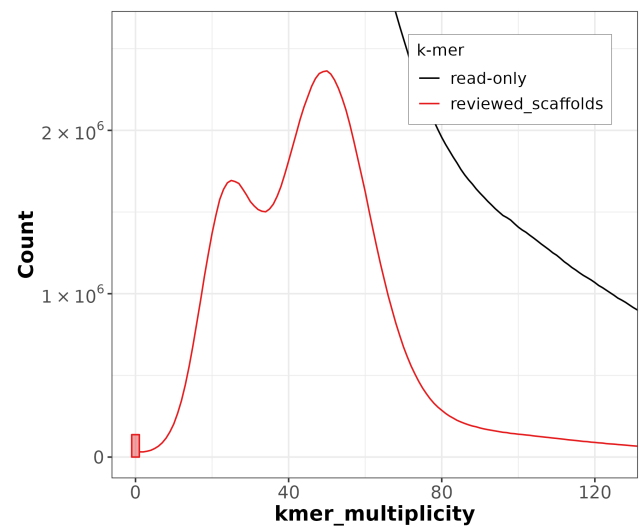


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

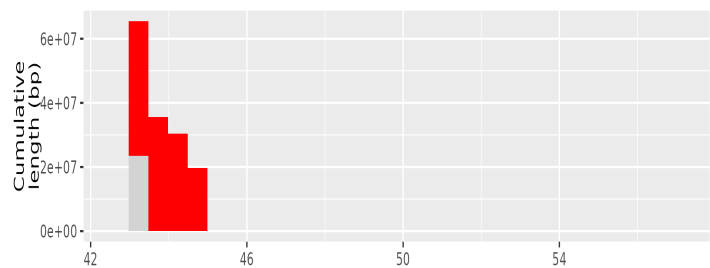


Distribution of k-mer counts per copy numbers found in asm

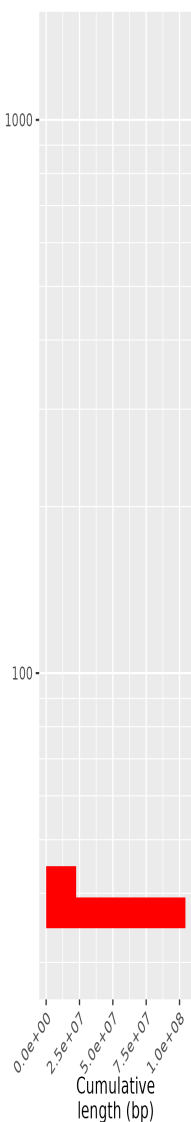
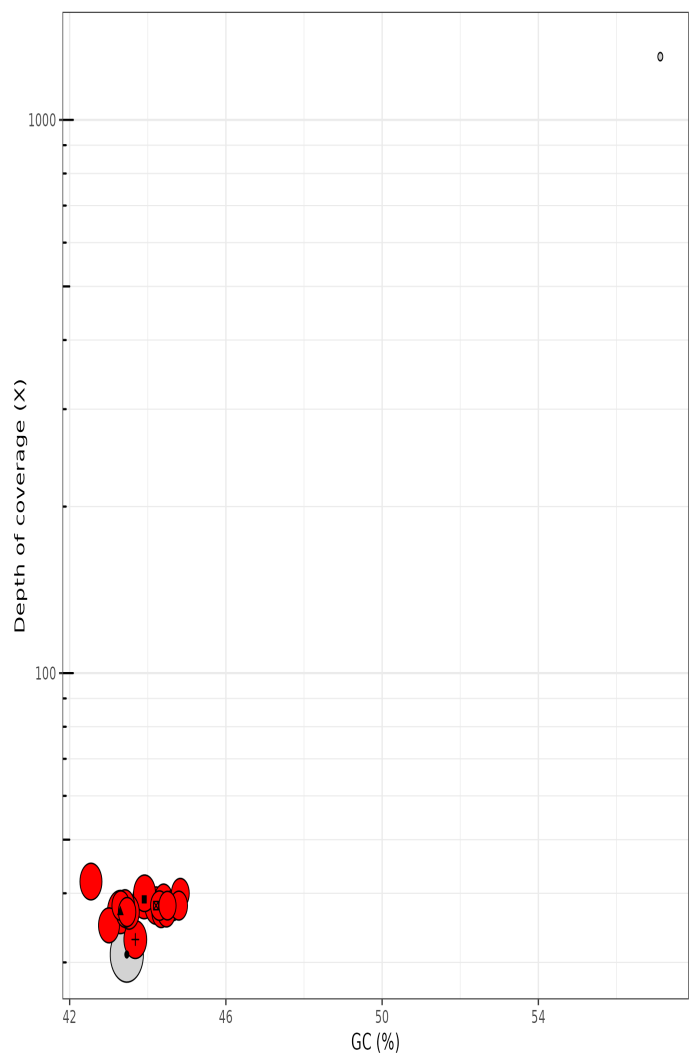


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



- Longest sequences (bp)
- odPacJohn2_1 - 23514268 (N/A)
 - odPacJohn2_2 - 13117769 (Eukaryota)
 - odPacJohn2_3 - 10055960 (Eukaryota)
 - odPacJohn2_4 - 9534154 (Eukaryota)
 - odPacJohn2_5 - 9131168 (Eukaryota)

- superkingdom
- Eukaryota
 - N/A

- Length (bp)
- 5.0e+06
 - 1.0e+07
 - 1.5e+07
 - 2.0e+07

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	67	379

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: 2025-07-30 11:21:20 CEST