

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	104758
ToLID	jaPorAstr1
Species	Porites astreoides
Class	Anthozoa
Order	Scleractinia

Genome Traits	Expected	Observed
Haploid size (bp)	536,006,072	559,172,316
Haploid Number	6 (source: ancestor)	14
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q40

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

Curator notes

. Interventions/Gb: 16
. Contamination notes: ""
. Other observations: "The assembly of *Porites astreoides* (jaPorAstr1) is based on 57X PacBio data and 430X 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 Hifiiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 200 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 4.044 Mb (with the largest being 0.068 Mb). Additionally, 644 regions totaling 21.83 Mb (with the largest being 1.227 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, 3 haplotypic regions and 28 contaminant sequences were removed, totaling 2.173Mb and 0.944Mb, respectively (with the largest being 1.831Mb and 0.18Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

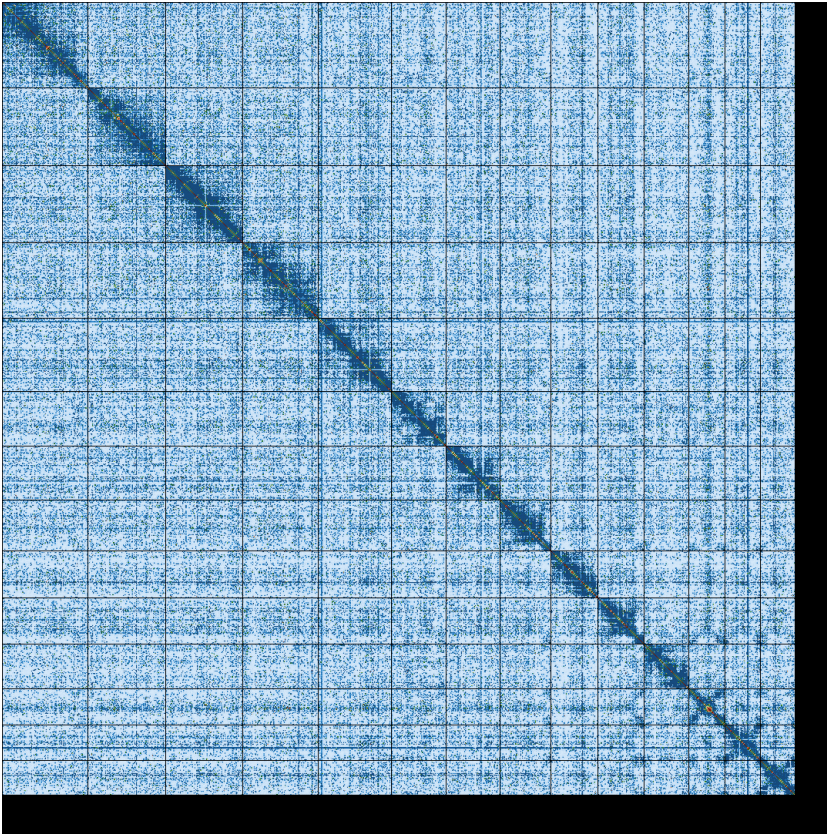
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	562,348,711	559,172,316
GC %	39.55	39.55
Gaps/Gbp	113.81	125.19
Total gap bp	6,400	7,700
Scaffolds	1,214	1,171
Scaffold N50	36,311,000	36,734,000
Scaffold L50	6	6
Scaffold L90	14	13
Contigs	1,278	1,241
Contig N50	29,988,000	29,988,000
Contig L50	8	8
Contig L90	22	22
QV	40.9664	40.9916
Kmer compl.	79.5004	79.3266
BUSCO sing.	98.8%	97.5%
BUSCO dupl.	1.1%	0.9%
BUSCO frag.	0.0%	1.2%
BUSCO miss.	0.1%	0.4%

Warning! BUSCO versions or lineage datasets are not the same across results:

BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: anthozoa_odb12 (genomes:12, BUSCOs:3649)

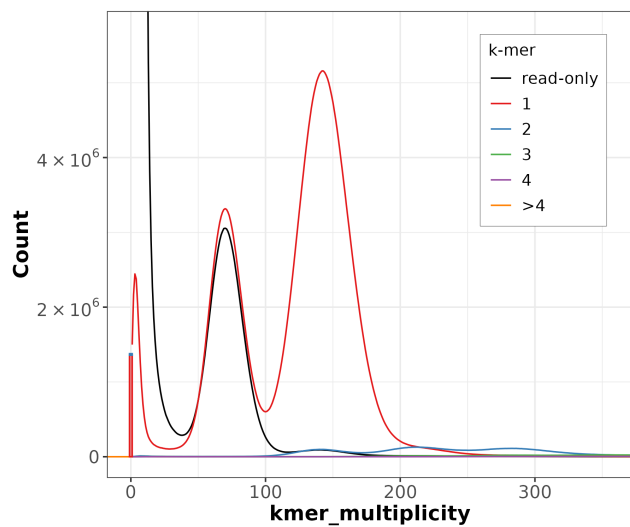
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: anthozoa_odb12 (genomes:12, BUSCOs:3649)

HiC contact map of curated assembly

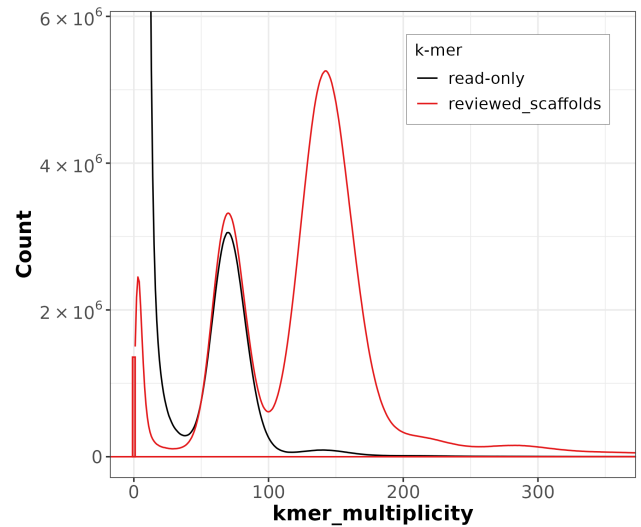


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

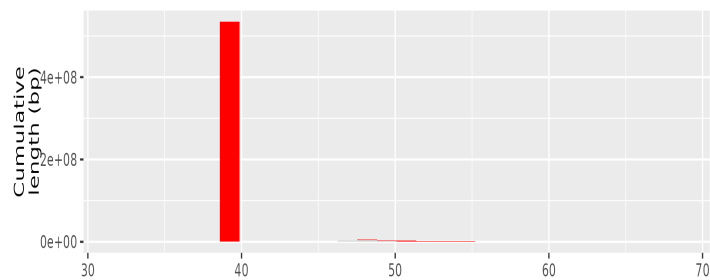


Distribution of k-mer counts per copy numbers found in asm

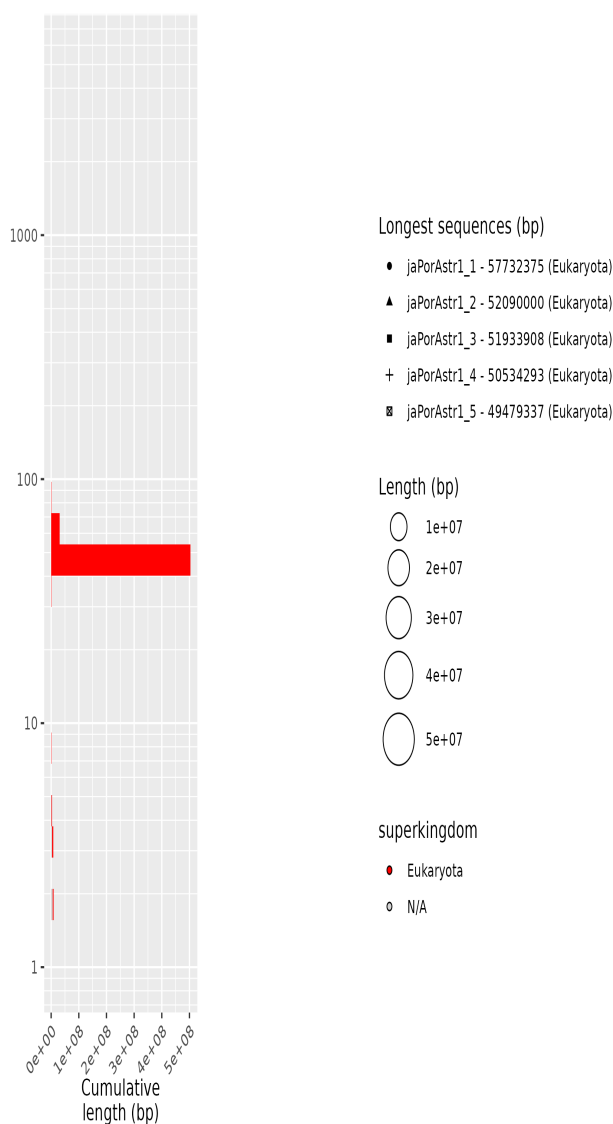
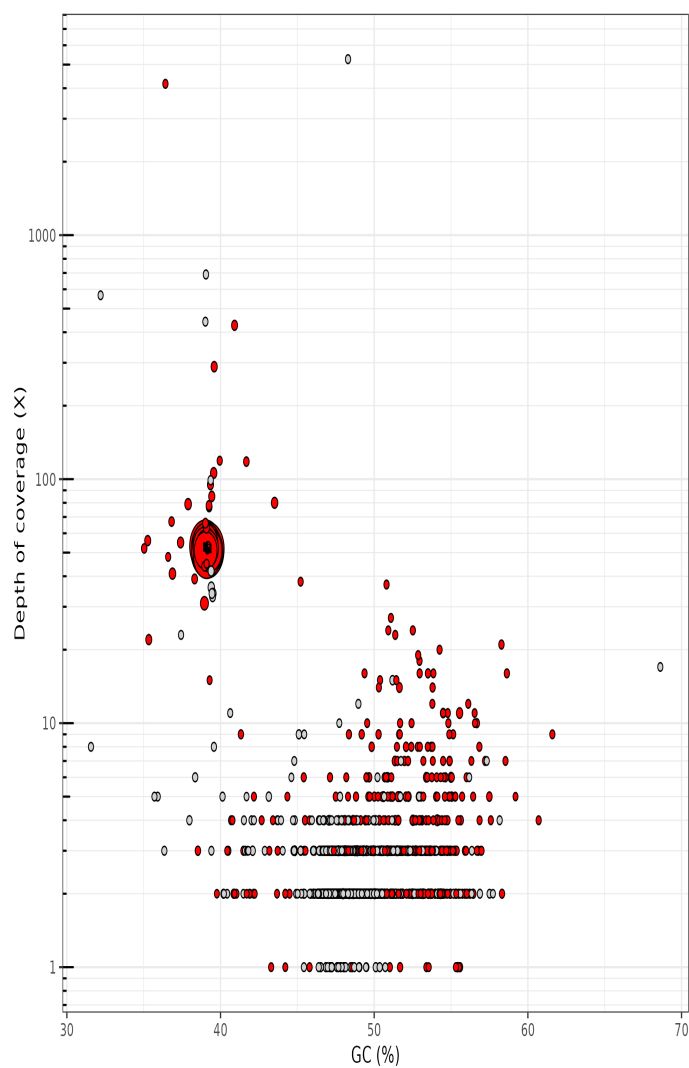


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	57	432

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: Emilie Teodori

Affiliation: Genoscope

Date and time: 2026-07-24 20:22:15 CEST