

ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	6108
ToLID	jaAneSulc1
Species	Anemonia sulcata
Class	Anthozoa
Order	Actiniaria

Genome Traits	Expected	Observed
Haploid size (bp)	313,567,667	308,320,409
Haploid Number	6 (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.Q59

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

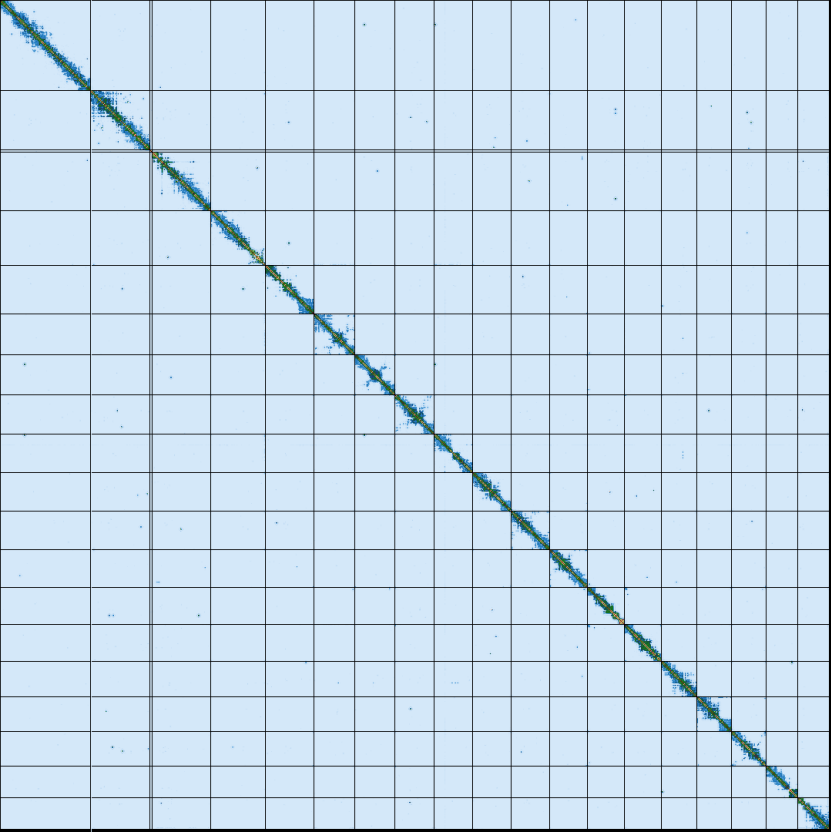
. Interventions/Gb: 29
. Contamination notes: ""
. Other observations: "The assembly of *Anemonia sulcata* (jaAneSulc1.1) is based on 55X of PacBio data and Arima Hi-C data generated as part of the ATLASea programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial 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. In total, 4 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.97 Mb (with the largest being 0.43 Mb). Additionally, 239 regions totaling 47,39 Mb (with the largest being 7,94 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, 4 haplotypic regions and 7 contaminant sequences were removed, totaling 2.5 Mb and 0.23 Mb, respectively (with the largest being 1.5 Mb and 0.07 Mb). 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	311,111,649	308,320,409
GC %	37.66	37.64
Gaps/Gbp	99.64	110.27
Total gap bp	3,100	3,900
Scaffolds	55	41
Scaffold N50	14,454,390	14,454,390
Scaffold L50	8	8
Scaffold L90	18	17
Contigs	86	75
Contig N50	9,596,318	9,596,318
Contig L50	12	12
Contig L90	35	33
QV	46.9376	59.8753
Kmer compl.	69.8906	72.6281
BUSCO sing.	90.4%	90.7%
BUSCO dupl.	0.8%	0.5%
BUSCO frag.	1.8%	1.8%
BUSCO miss.	7.0%	7.0%

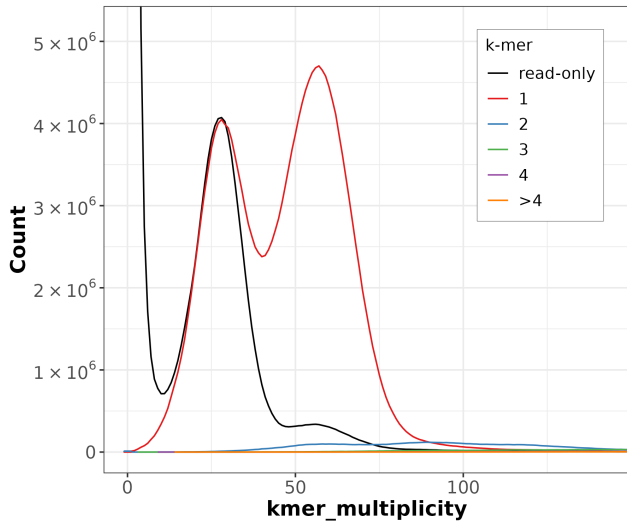
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

HiC contact map of curated assembly

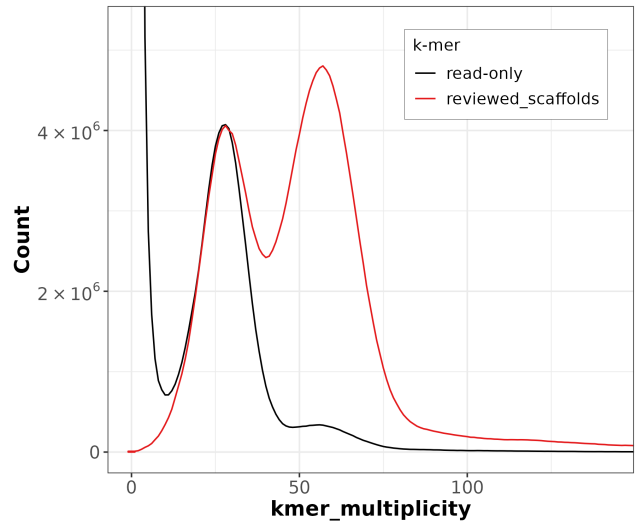


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

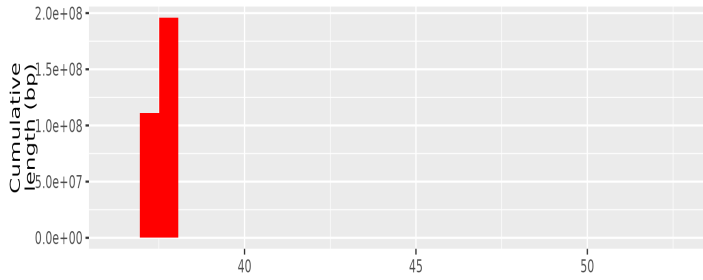


Distribution of k-mer counts per copy numbers found in asm

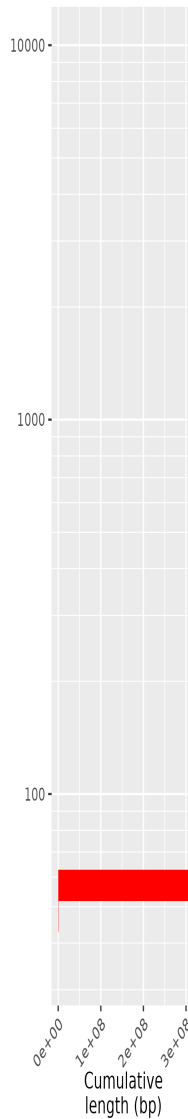
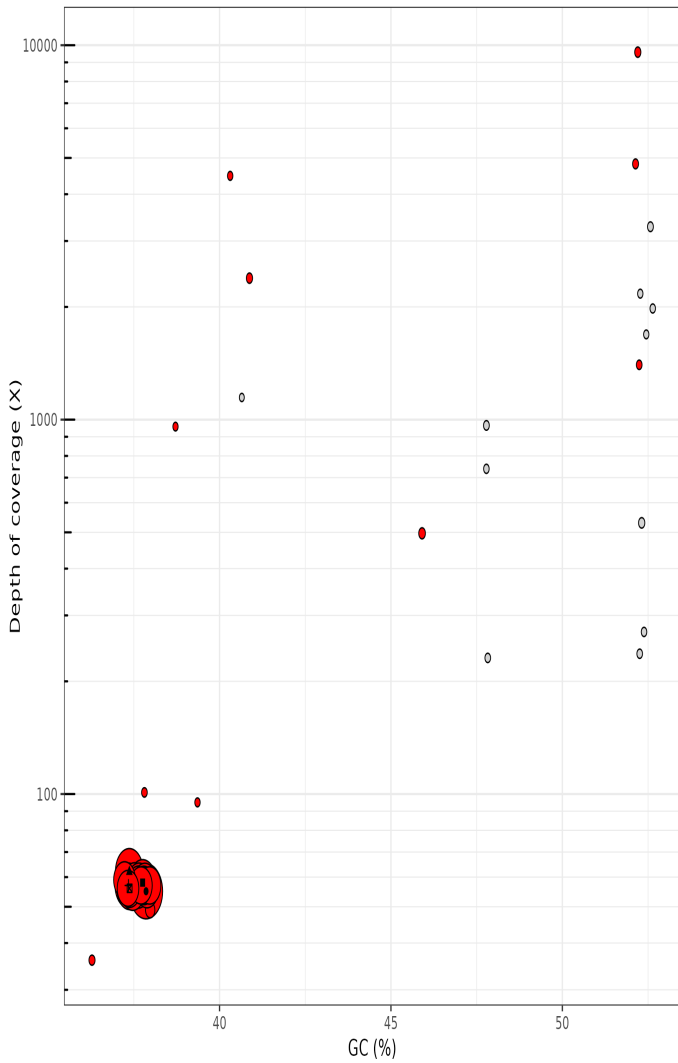


Distribution of k-mer counts coloured by their presence in reads/assemblies

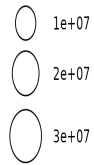
Post-curation contamination screening



TAPAs summary Graph



Length (bp)



Longest sequences (bp)

- SUPER_1 - 33432664 (Eukaryota)
- ▲ SUPER_2 - 22144057 (Eukaryota)
- SUPER_3 - 21576832 (Eukaryota)
- + SUPER_4 - 20302814 (Eukaryota)
- ▣ SUPER_5 - 17876028 (Eukaryota)

superkingdom

- Eukaryota
- N/A

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	55	246

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: Adama Ndar
Affiliation: Genoscope

Date and time: 2025-02-25 13:41:13 CET