

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	237119
ToLID	odAxiVerr1
Species	<i>Axinella verrucosa</i>
Class	Demospongiae
Order	Axinellida

Genome Traits	Expected	Observed
Haploid size (bp)	219,258,830	247,466,917
Haploid Number	5 (source: ancestor)	13
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q60

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: 56
- . Contamination notes: ""
- . Other observations: "The assembly of *AXINELLA VERRUCOSA* (odAxiVerr1) is based on 97X PacBio data and 161X Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA, <https://www.erga-biodiversity.eu/>) via the Biodiversity Genomics Europe project (BGE, <https://biodiversitygenomics.eu/>). 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, 703 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 71,633,625 pb (with the largest being 3,087,110 pb). Additionally, 186 regions totaling 15,617,024 pb (with the largest being 834,842 pb) 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, 42 contigs were identified as contaminants, totaling 917,636 pb (with the largest being 58,401 pb) and 3 haplotypic region was removed, totaling 2,339,198 pb (with the largest being 1,680,985 pb). Chromosome-scale scaffolds confirmed by Hi-C data were

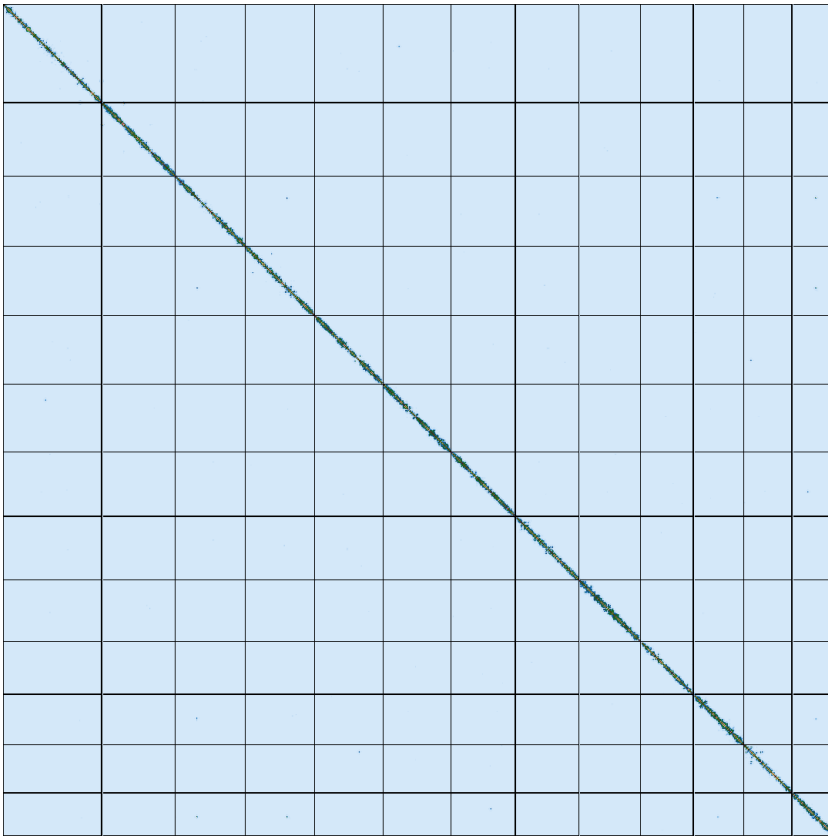
named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	251,161,513	247,466,917
GC %	45.4	45.43
Gaps/Gbp	740.56	767.78
Total gap bp	18,600	20,200
Scaffolds	69	23
Scaffold N50	20,322,608	20,338,932
Scaffold L50	6	6
Scaffold L90	12	12
Contigs	255	213
Contig N50	2,127,225	2,127,225
Contig L50	31	31
Contig L90	120	117
QV	60.3866	60.7305
Kmer compl.	72.9915	72.4708
BUSCO sing.	56.2%	56.2%
BUSCO dupl.	0.7%	0.7%
BUSCO frag.	27.2%	27.2%
BUSCO miss.	15.8%	15.8%

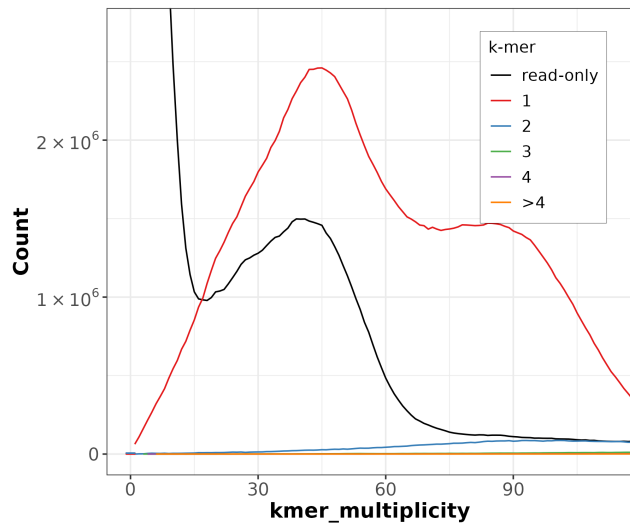
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

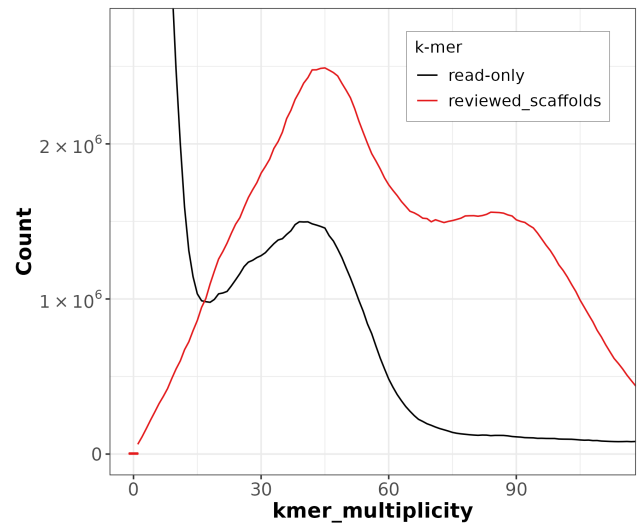


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

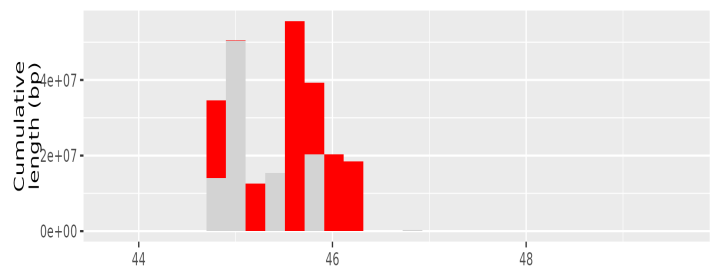


Distribution of k-mer counts per copy numbers found in asm

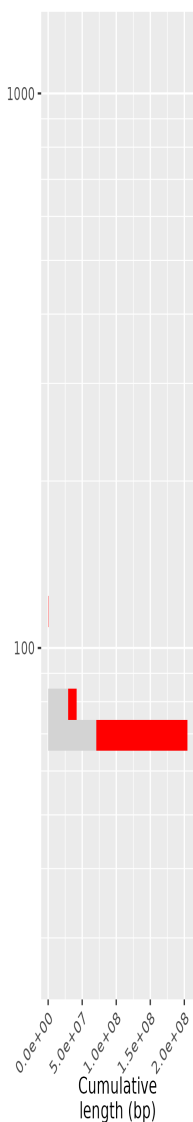
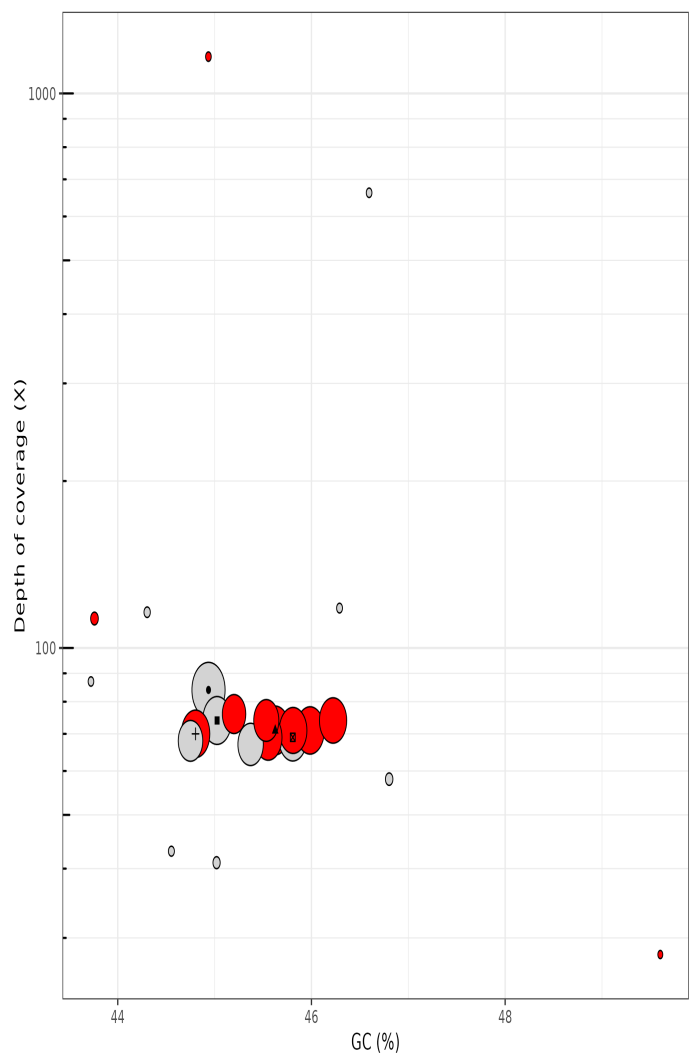


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



Longest sequences (bp)

- odAxiverr1_1 - 29300985 (N/A)
- ▲ odAxiverr1_2 - 21861230 (Eukaryota)
- odAxiverr1_3 - 20888526 (N/A)
- + odAxiverr1_4 - 20467386 (Eukaryota)
- ▣ odAxiverr1_5 - 20345779 (N/A)

Length (bp)

- 1e+07
- 2e+07

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	Long reads	Arima
Coverage	97	161

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: Lola Demirdjian

Affiliation: Genoscope

Date and time: 2025-09-27 20:55:47 CEST