

ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1400855
ToLID	jaVerCynol
Species	Veretillum cynomorium
Class	Anthozoa
Order	Scleralcyonacea

Genome Traits	Expected	Observed
Haploid size (bp)	749,740,417	669,772,940
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.Q57

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

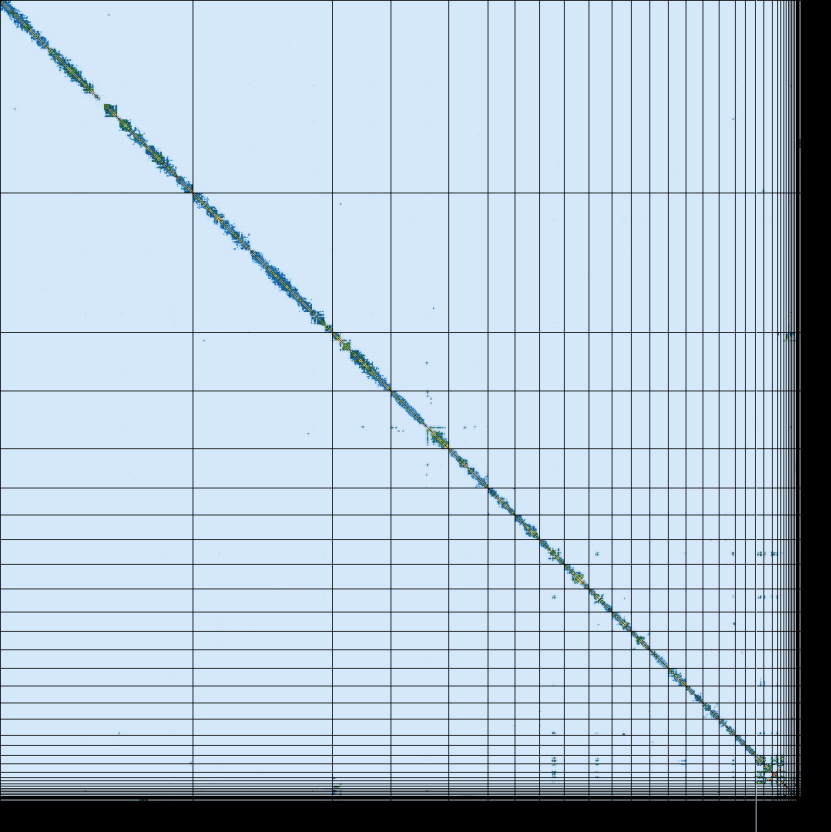
- . Interventions/Gb: 64
- . Contamination notes: ""
- . Other observations: "The assembly of Veretillum cynomorium (jaVerCynol.1) is based on 60X of PacBio data and Arima Hi-C data generated as part of the ATLASea programme (<https://www.atlasea.fr>). The assembly process included the following steps: 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, 27 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 2.8 Mb (with the largest being 0.57 Mb). Additionally, 475 regions totaling 250 Mb (with the largest being 18.9 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, 3 haplotypic regions and 4 contaminant sequences were removed, totaling 1.4 Mb and 2.2 Mb, respectively (with the largest being 0.96 Mb and 1 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	673,430,521	669,772,940
GC %	37.47	37.56
Gaps/Gbp	294.02	309.06
Total gap bp	19,800	22,800
Scaffolds	274	263
Scaffold N50	46,610,113	46,610,213
Scaffold L50	4	4
Scaffold L90	22	19
Contigs	472	470
Contig N50	5,241,857	5,241,857
Contig L50	43	43
Contig L90	136	137
QV	44.6854	57.6485
Kmer compl.	72.0682	73.1805
BUSCO sing.	84.8%	84.7%
BUSCO dupl.	0.7%	0.8%
BUSCO frag.	5.8%	5.8%
BUSCO miss.	8.7%	8.7%

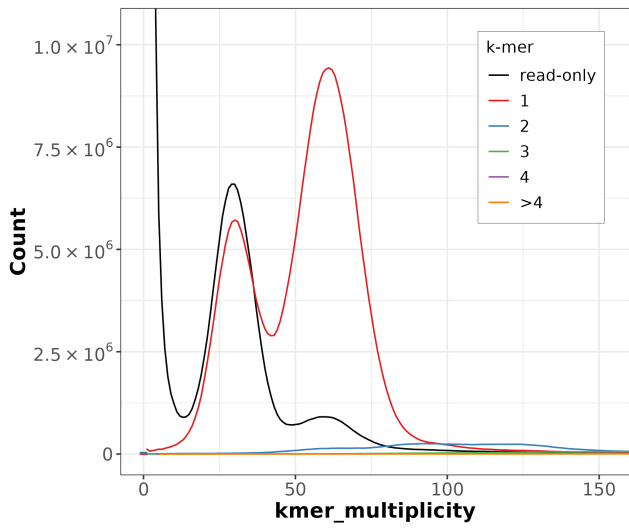
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

HiC contact map of curated assembly

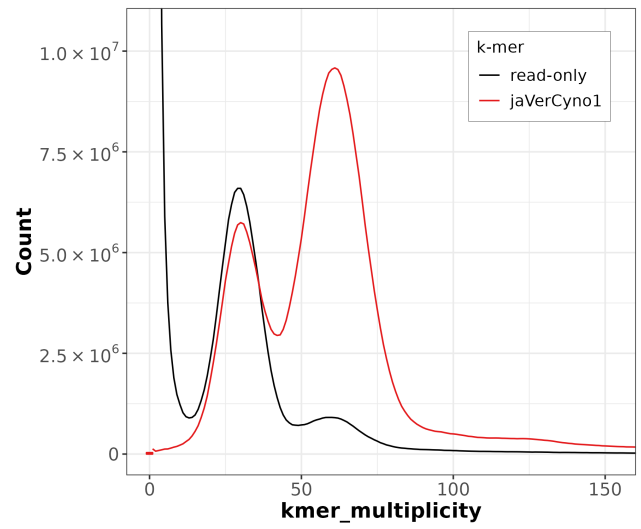


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

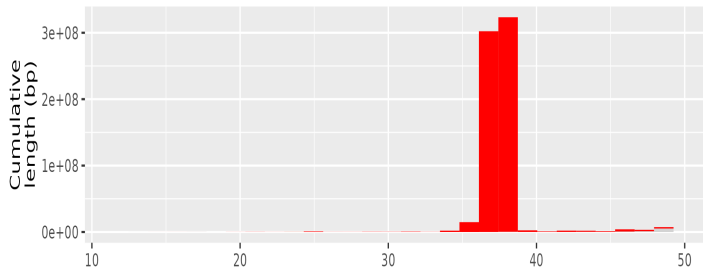


Distribution of k-mer counts per copy numbers found in asm



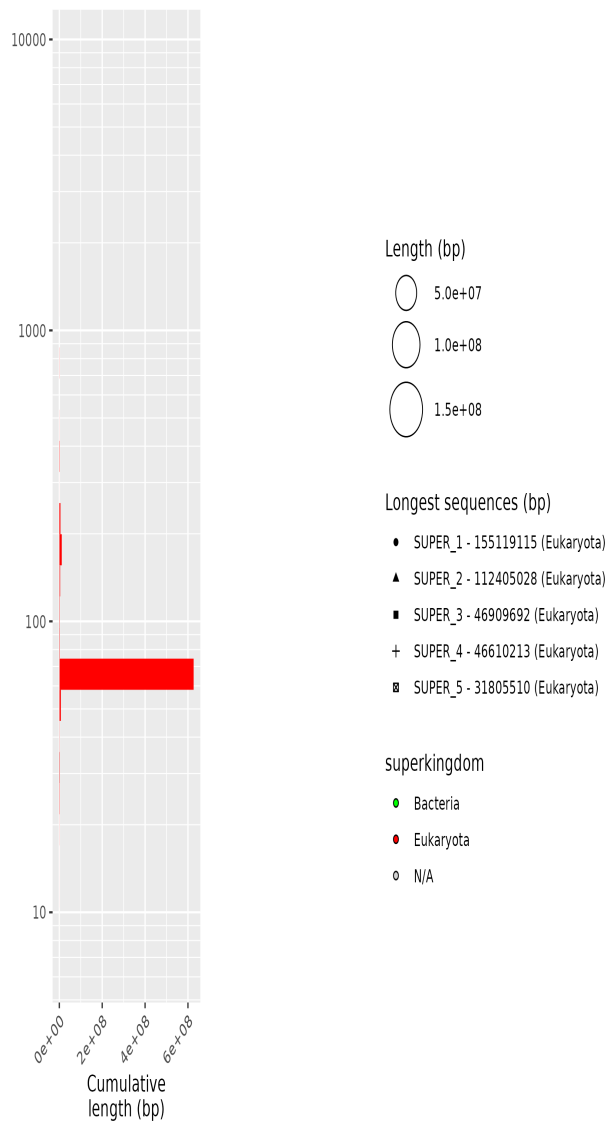
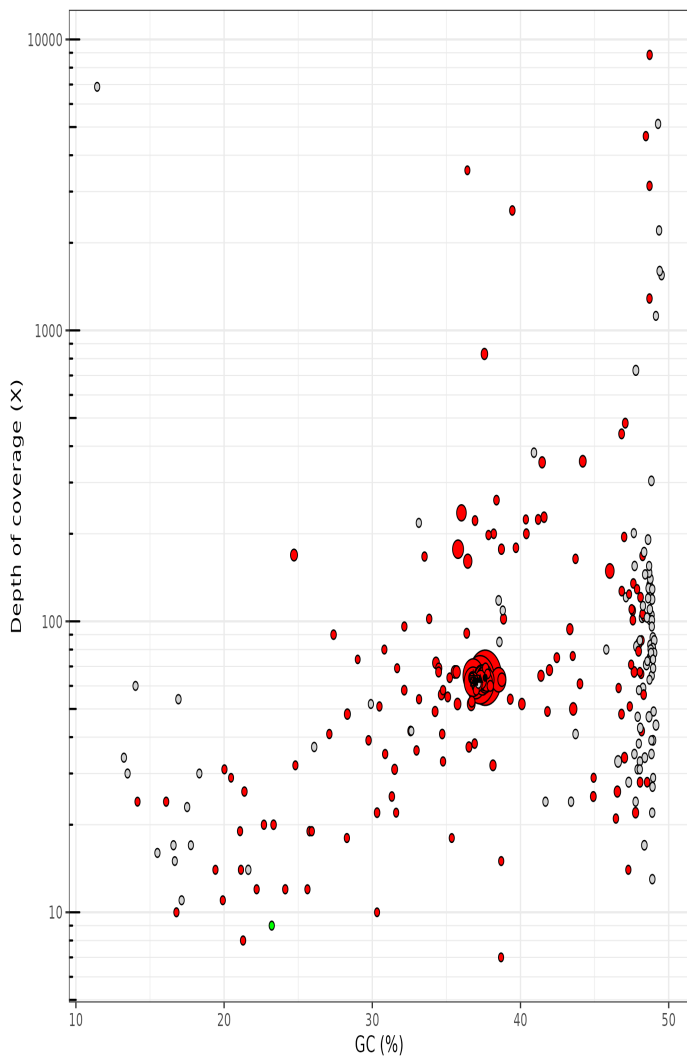
Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph

(1 0X contig has been hidden)



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	60	49

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: 2024-12-17 19:26:14 CET