

ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	483417
ToLID	qmEriVerr1
Species	<i>Eriphia verrucosa</i>
Class	Malacostraca
Order	Decapoda

Genome Traits	Expected	Observed
Haploid size (bp)	2,117,157,944	1,277,833,293
Haploid Number	32 (source: ancestor)	46
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 5.7.Q37

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
- . QV value is less than 40 for collapsed
- . Kmer completeness value is less than 90 for collapsed
- . More than 1000 gaps/Gbp for collapsed

Curator notes

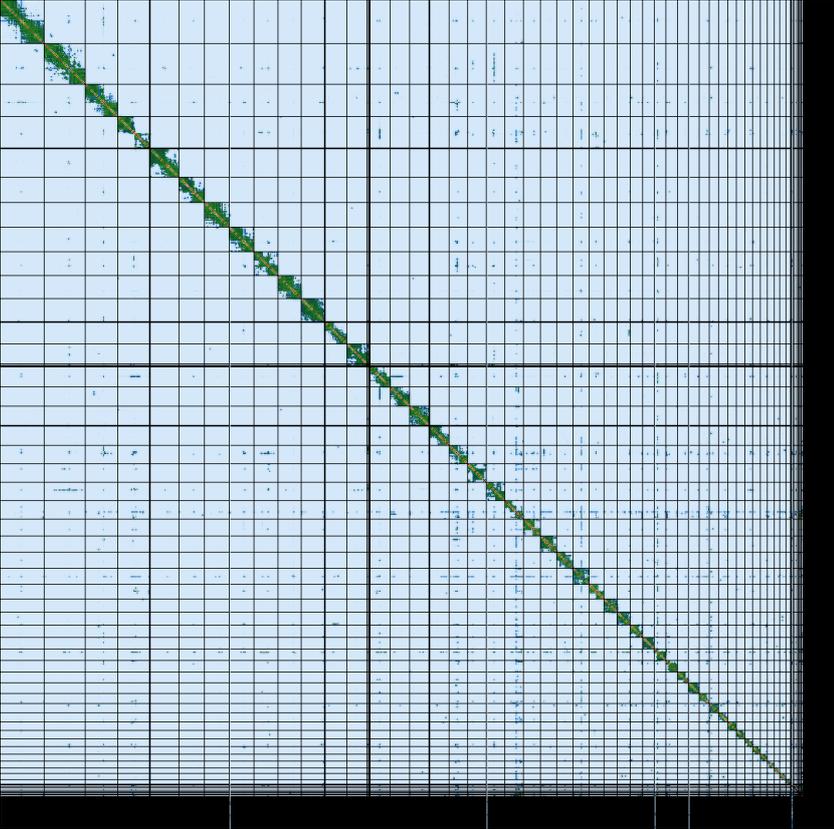
- . Interventions/Gb: 83
- . Contamination notes: ""
- . Other observations: "The assembly of *Eriphia verrucosa* (qmEriVerr1) is based on 30X ONT data and 76X Arima Hi-C data generated as part of the ATLASea programme (<https://www.atlasea.fr/>). The assembly process included the following steps: initial ONT assembly generation with Nextdenovo, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 11 contigs were identified as contaminants (bacterial or viral), totaling 1 Mb (with the largest being 202 kb). Additionally, 1120 regions totaling 136 Mb (with the largest being 909 kb) 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, no regions were tagged as allelic duplications or contaminants. 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	1,277,935,017	1,277,833,293
GC %	41.28	41.28
Gaps/Gbp	1,802.13	1,836.7
Total gap bp	230,300	242,300
Scaffolds	407	318
Scaffold N50	26,275,345	29,056,483
Scaffold L50	17	16
Scaffold L90	57	41
Contigs	2,710	2,665
Contig N50	969,605	960,000
Contig L50	292	296
Contig L90	1,624	1,632
QV	37.1706	37.658
Kmer compl.	76.9326	76.8975
BUSCO sing.	92.9%	92.9%
BUSCO dupl.	1.2%	1.2%
BUSCO frag.	2.4%	2.4%
BUSCO miss.	3.5%	3.5%

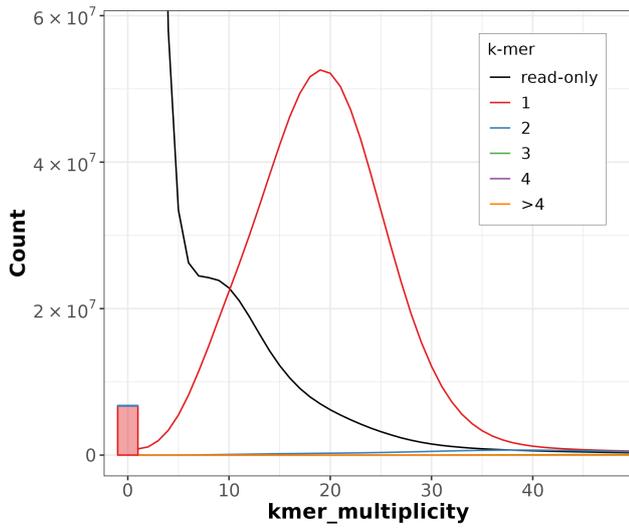
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: arthropoda_odb10 (genomes:90, BUSCOs:1013)

HiC contact map of curated assembly

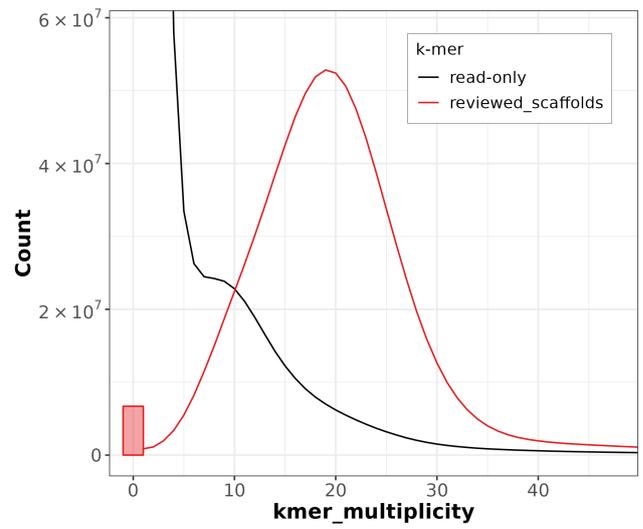


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

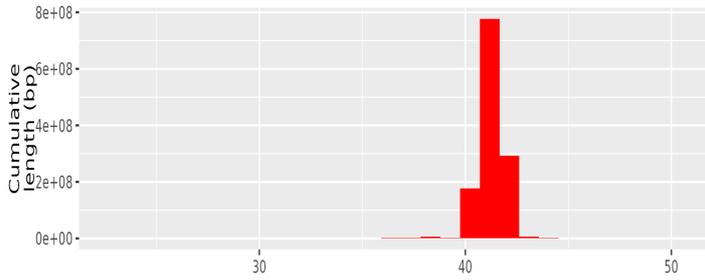


Distribution of k-mer counts per copy numbers found in asm

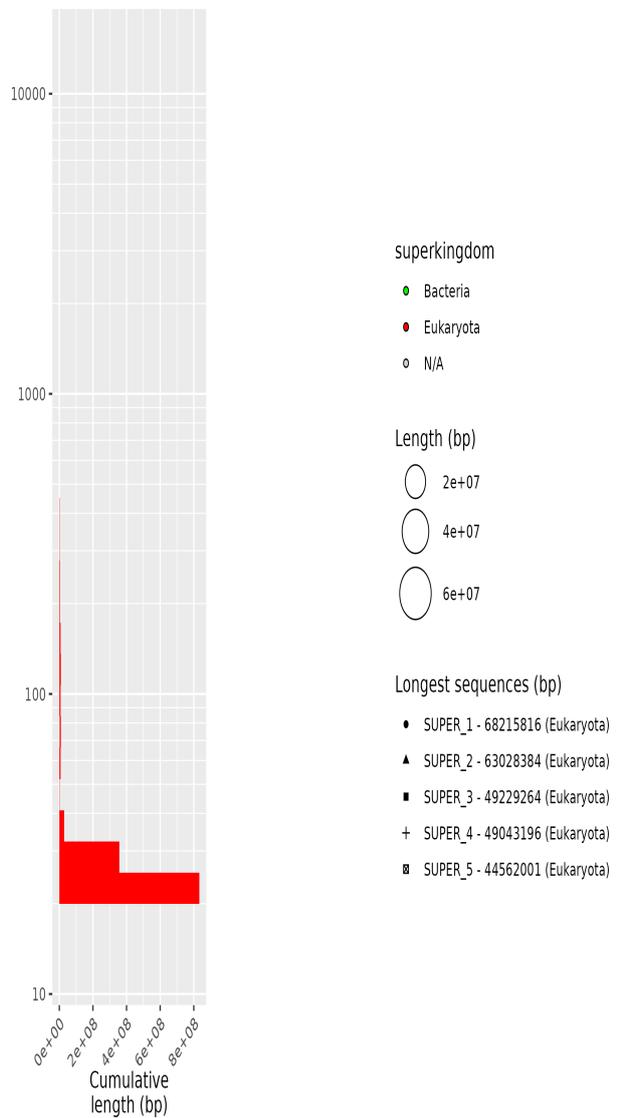
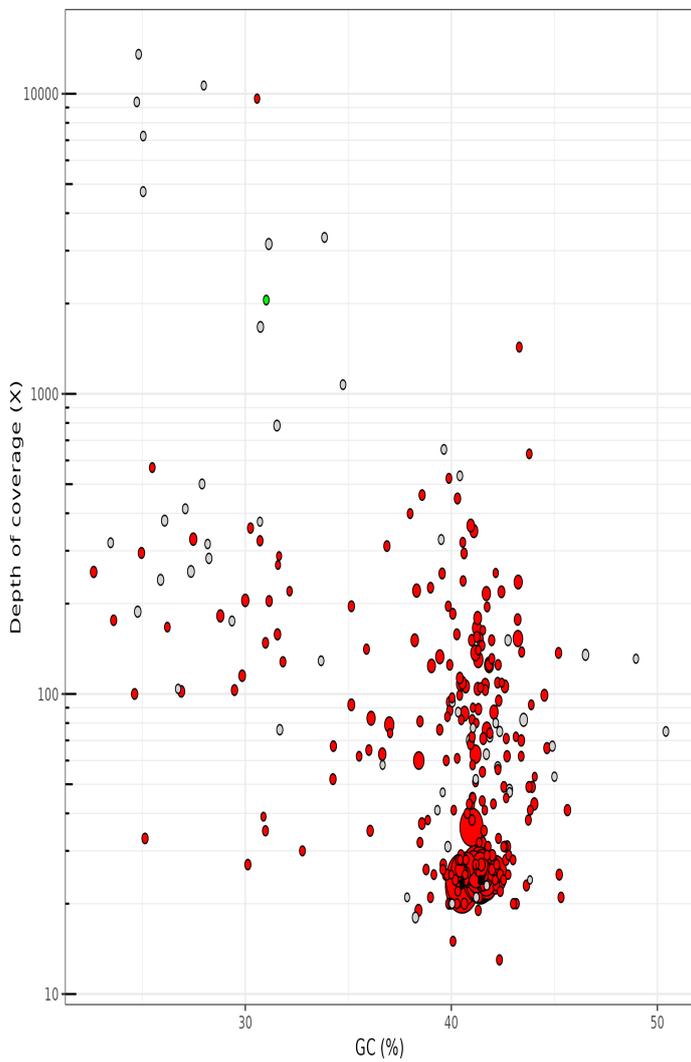


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	PACBIO Hifi	Arima
Coverage	30	51

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: 2025-02-15 08:30:24 CET