ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	427284
ToLID	**waEupFoli1**
Species	Euphrosine foliosa
Class	Polychaeta
Order	Eunicida

Genome Traits	Expected	Observed
Haploid size (bp)	1,666,250,507	1,682,655,180
Haploid Number	3 (source: ancestor)	18
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q48

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

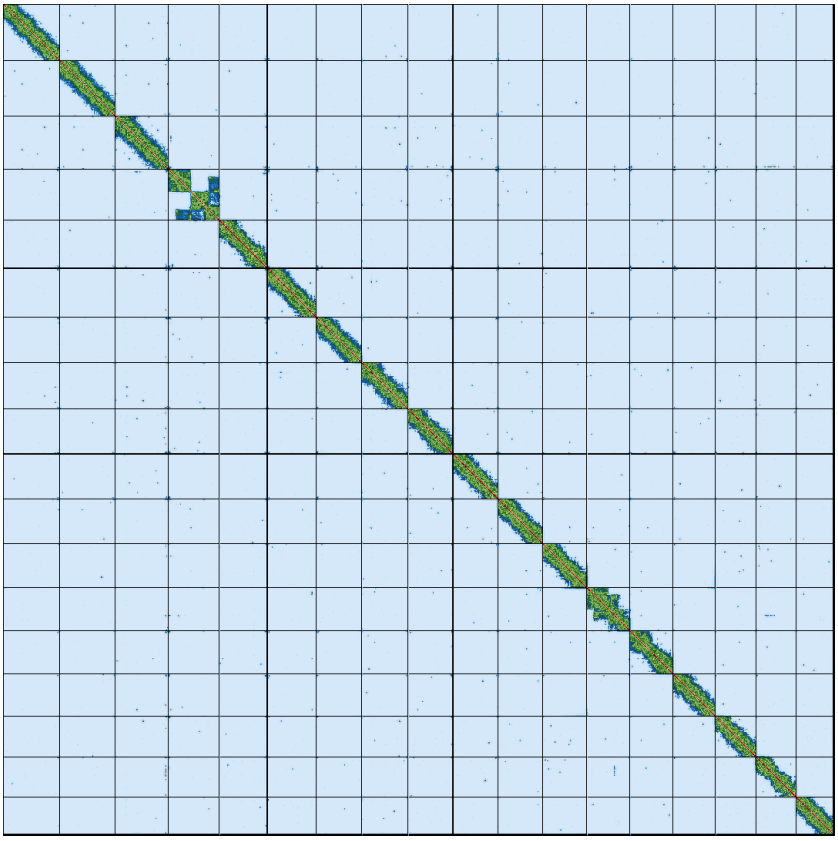
. Interventions/Gb: 25
. Contamination notes: ""
. Other observations: "The assembly of Euphrosine foliosa (waEupFoli1) is based on 51X PacBio data and 413X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 79 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 8.716 Mb (with the largest being 3.99 Mb). Additionally, 440 regions totaling 129.167 Mb (with the largest being 9.701 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, 18 haplotypic regions and 1 contaminant sequence were removed, totaling 42.036Mb and 0.025Mb, respectively (with the largest being 14.333Mb and 0.025Mb). 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	1,724,725,198	1,682,655,180
GC %	38.81	38.81
Gaps/Gbp	49.86	51.7
Total gap bp	8,600	10,500
Scaffolds	88	73
Scaffold N50	92,531,715	91,571,014
Scaffold L50	9	9
Scaffold L90	17	16
Contigs	174	160
Contig N50	27,865,000	27,865,000
Contig L50	21	21
Contig L90	62	61
QV	48.5983	48.6603
Kmer compl.	67.1792	66.315
BUSCO sing.	96.1%	97.4%
BUSCO dupl.	2.5%	1.1%
BUSCO frag.	0.4%	0.5%
BUSCO miss.	1.0%	1.0%

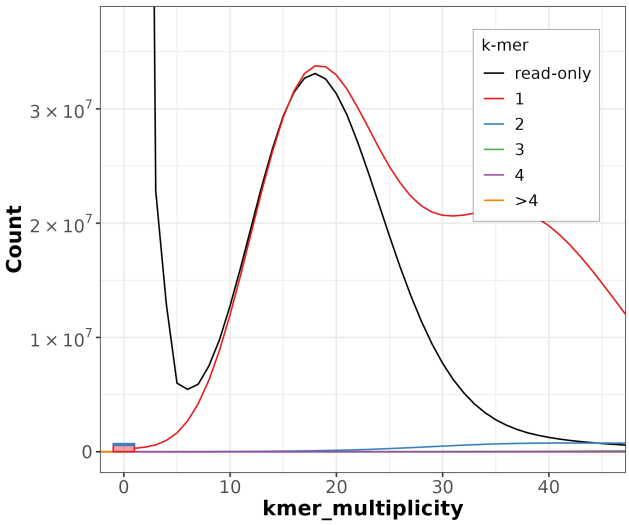
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)

HiC contact map of curated assembly

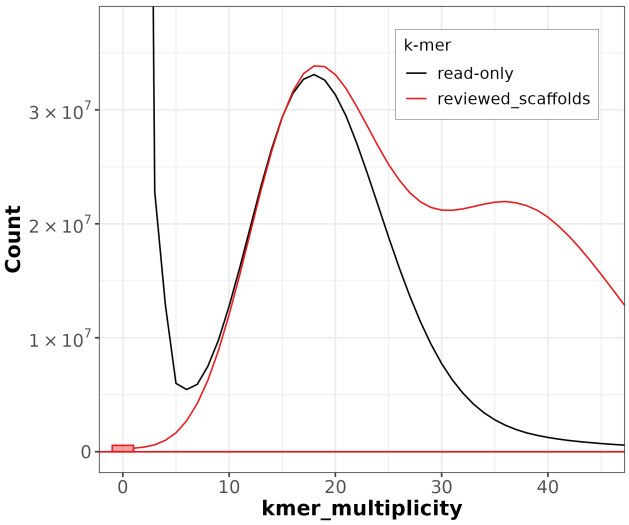


collapsed [LINK]

K-mer spectra of curated assembly

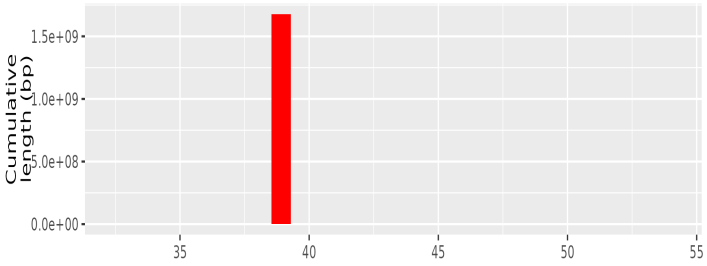


Distribution of k-mer counts per copy numbers found in asm

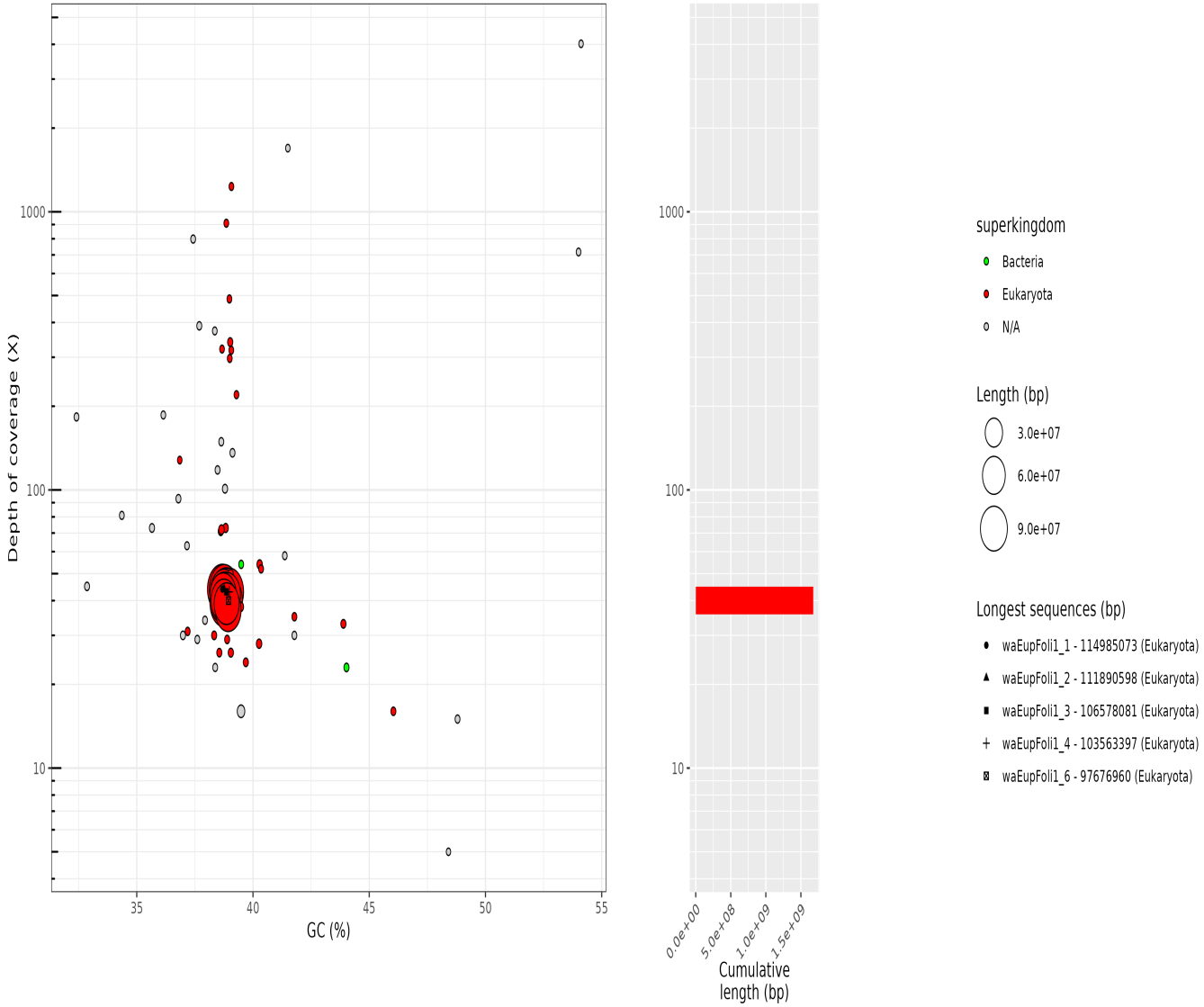


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	Omnic,Arima
Coverage	51	412

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: Arnaud Couloux
Affiliation: Genoscope

Date and time: 2026-05-13 01:15:03 CEST