ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	109186
ToLID	**eaLinMult1**
Species	Linckia multifora
Class	Asteroidea
Order	Valvatida

Genome Traits	Expected	Observed
Haploid size (bp)	572,878,748	594,564,371
Haploid Number	20 (source: ancestor)	22
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q47

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

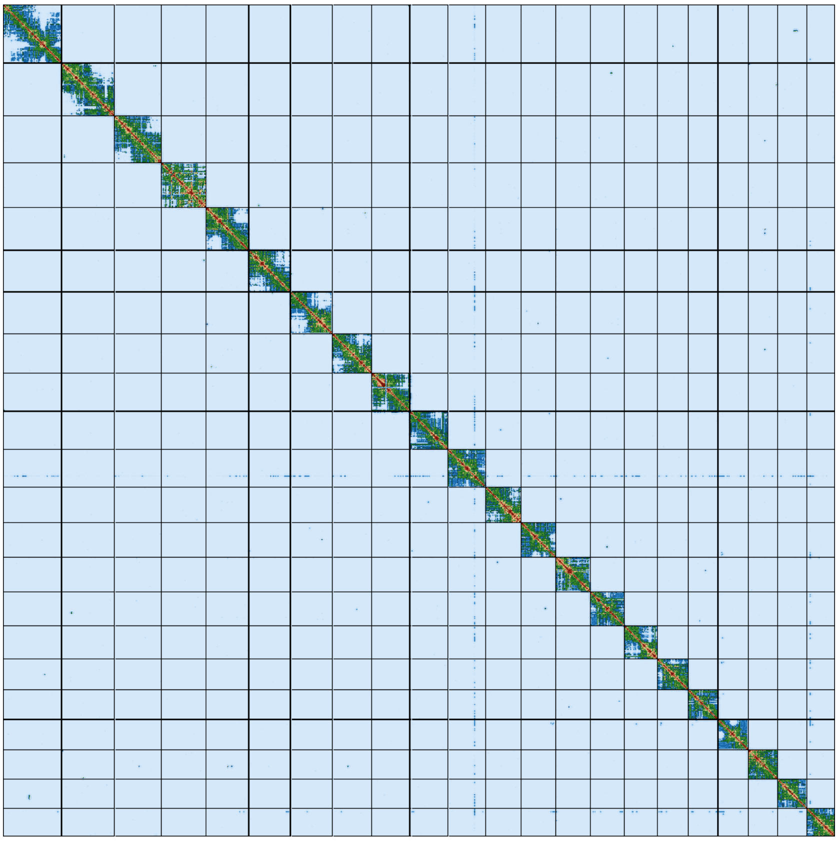
. Interventions/Gb: 69
. Contamination notes: ""
. Other observations: "The assembly of Linckia multifora (eaLinMult1) is based on 78X PacBio data and 95X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 10 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.719 Mb (with the largest being 0.208 Mb). Additionally, 172 regions totaling 17.404 Mb (with the largest being 1.657 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, 1 haplotypic regions and 7 contaminant sequences were removed, totaling 0.25Mb and 0.36Mb, respectively (with the largest being 0.25Mb and 0.08Mb). 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	595,169,698	594,564,371
GC %	41.32	41.32
Gaps/Gbp	141.14	141.28
Total gap bp	9,600	11,600
Scaffolds	57	44
Scaffold N50	26,577,889	26,860,312
Scaffold L50	10	10
Scaffold L90	20	20
Contigs	135	128
Contig N50	12,517,539	12,517,539
Contig L50	17	17
Contig L90	54	54
QV	47.2878	47.4637
Kmer compl.	70.8103	70.7708
BUSCO sing.	87.5%	87.5%
BUSCO dupl.	0.0%	0.0%
BUSCO frag.	11.5%	11.5%
BUSCO miss.	1.0%	1.0%

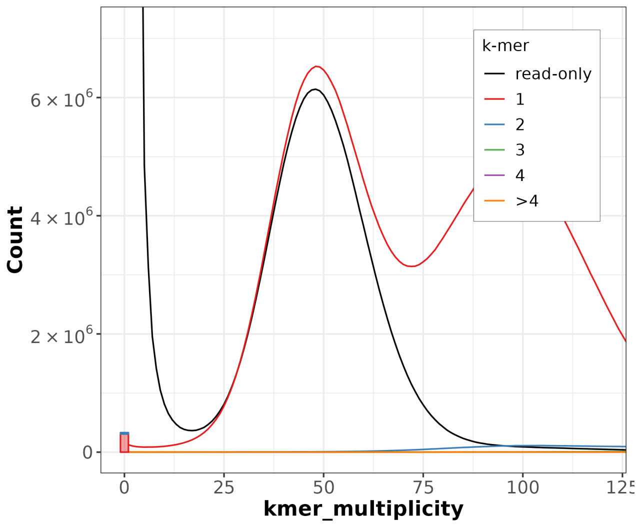
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

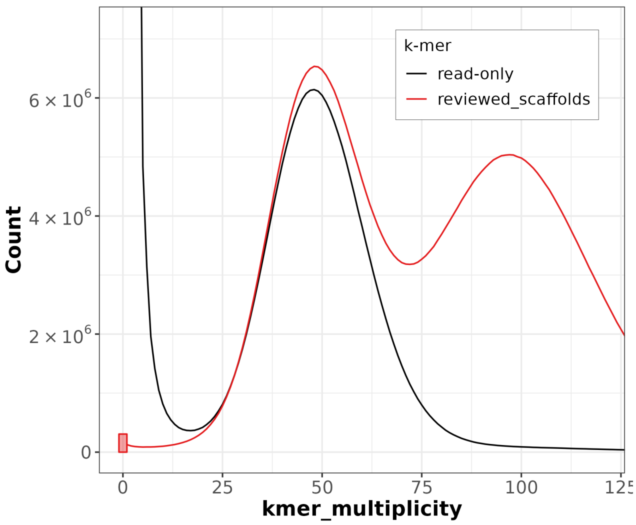


collapsed [LINK]

K-mer spectra of curated assembly

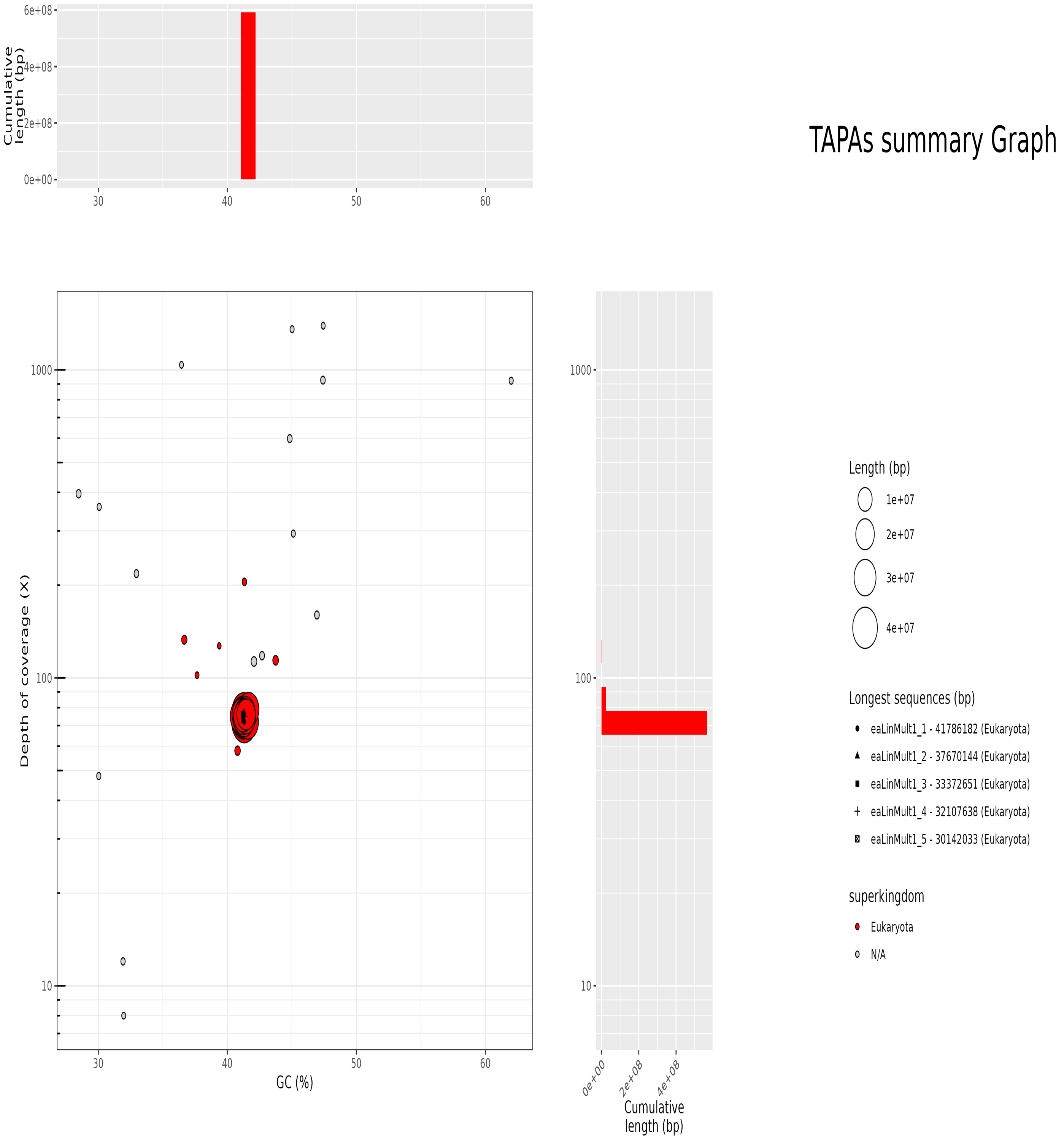


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	77	95

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: 2025-09-25 08:44:33 CEST