ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	109184
ToLID	**eaLinGuil1**
Species	Linckia guildingi
Class	Asteroidea
Order	Valvatida

Genome Traits	Expected	Observed
Haploid size (bp)	572,963,098	578,569,425
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: 6.7.Q47

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

. Observed Haploid Number is different from Expected

Curator notes

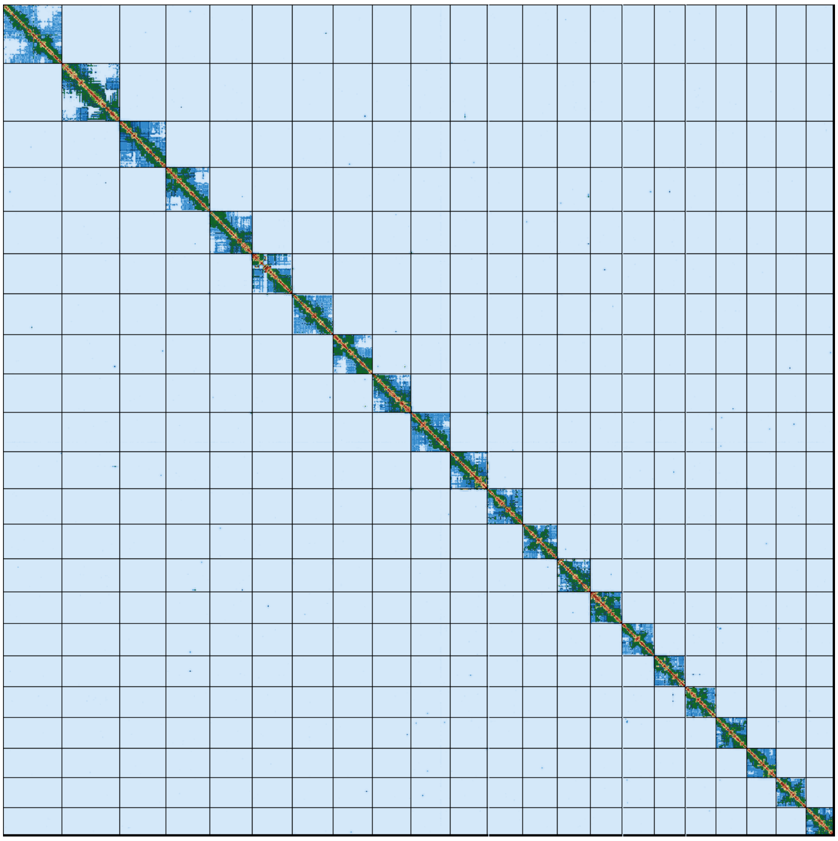
. Interventions/Gb: 67
. Contamination notes: ""
. Other observations: "The assembly of Linckia guildingi (eaLinGuil1) is based on 48X PacBio data and 137X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 6 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.148 Mb (with the largest being 0.046 Mb). Additionally, 117 regions totaling 10.792 Mb (with the largest being 0.905 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. During manual curation, 16 haplotypic regions were removed, totaling 4.51 Mb (with the largest being 0.83 Mb). Chromosome 2 contains a large region from 18 Mb to 30 Mb that appears to be a heterozygous deletion; indeed, both the depth of coverage and the density of Hi-C contacts are low in this interval. 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	583,067,024	578,569,425
GC %	41.34	41.34
Gaps/Gbp	250.4	267.9
Total gap bp	14,600	17,400
Scaffolds	58	44
Scaffold N50	26,581,213	26,789,821
Scaffold L50	10	10
Scaffold L90	20	20
Contigs	204	199
Contig N50	5,673,591	5,673,591
Contig L50	33	33
Contig L90	103	102
QV	47.5392	47.5232
Kmer compl.	90.5935	90.478
BUSCO sing.	98.2%	98.4%
BUSCO dupl.	0.6%	0.4%
BUSCO frag.	1.0%	1.0%
BUSCO miss.	0.1%	0.1%

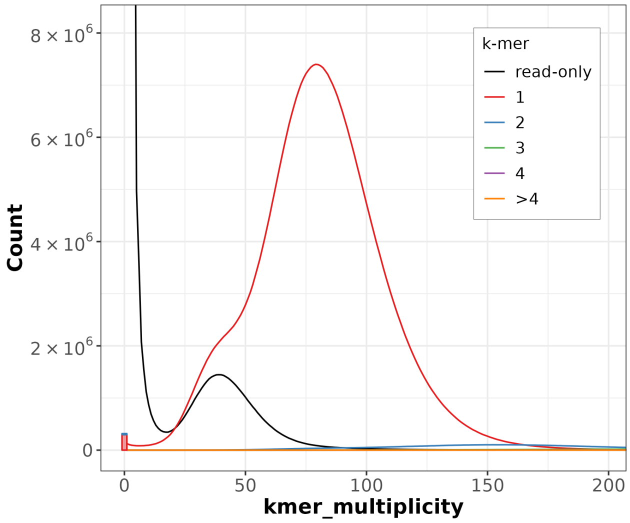
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

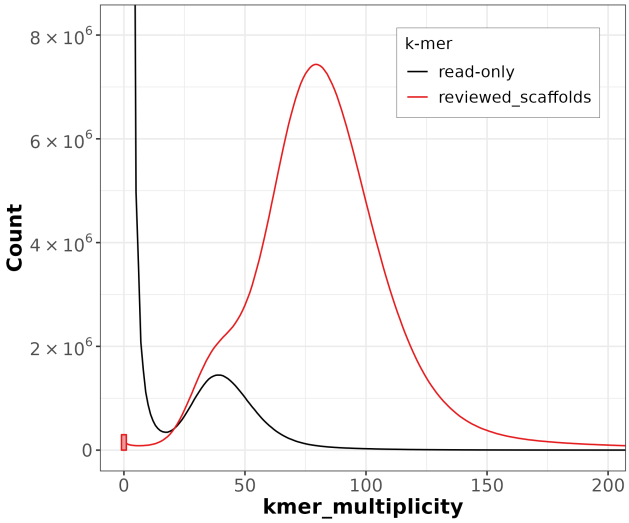


collapsed [LINK]

K-mer spectra of curated assembly

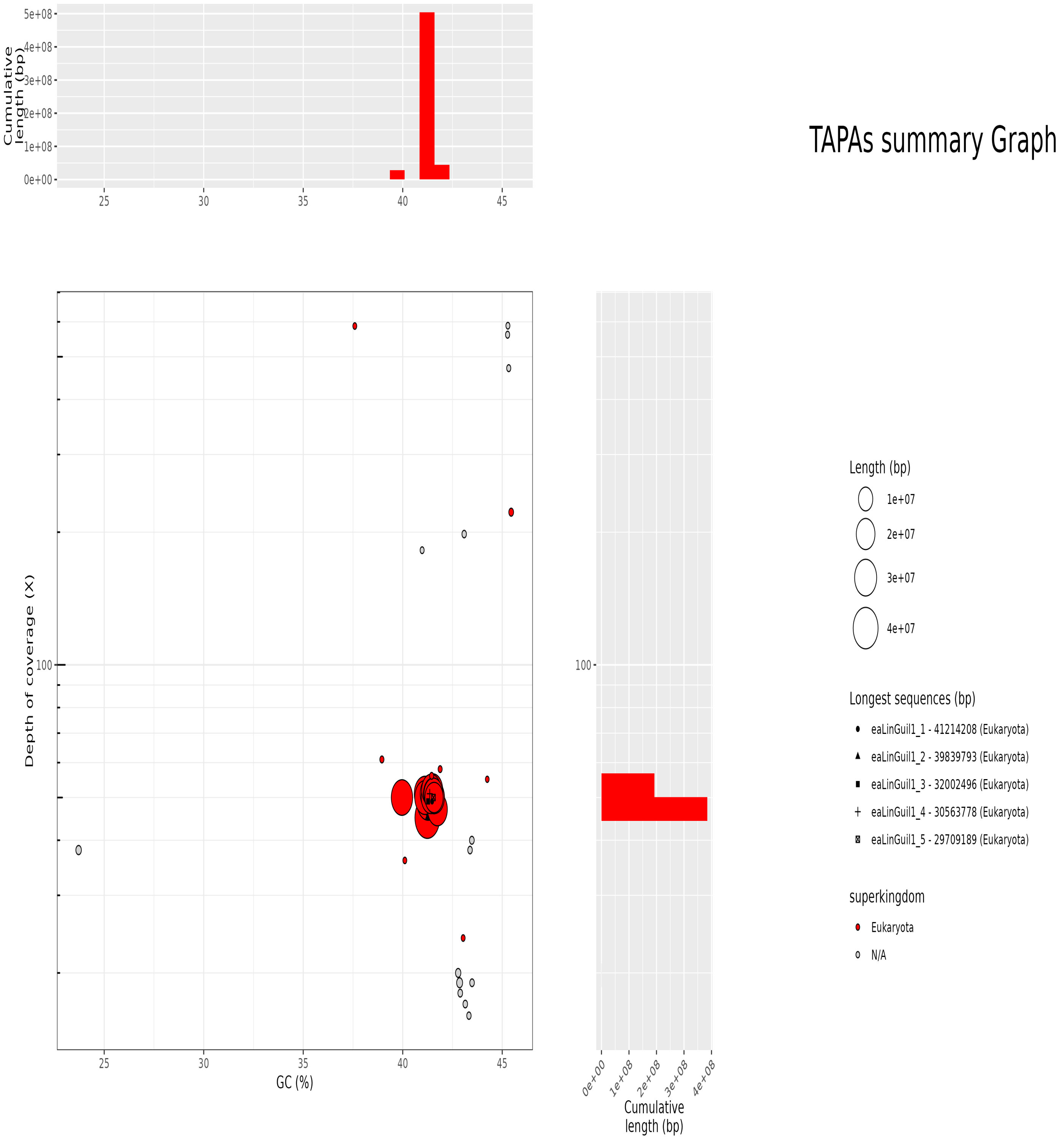


Distribution of k-mer counts per copy numbers found in asm



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	48	137

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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2025-11-23 17:09:21 CET