ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	1235830
ToLID	**ioSomMeri1**
Species	Somatochlora meridionalis
Class	Insecta
Order	Odonata

Genome Traits	Expected	Observed
Haploid size (bp)	2,238,591,471	2,424,650,310
Haploid Number	13 (source: ancestor)	12
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q62

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
. Not 90% of assembly in chromosomes for collapsed

Curator notes

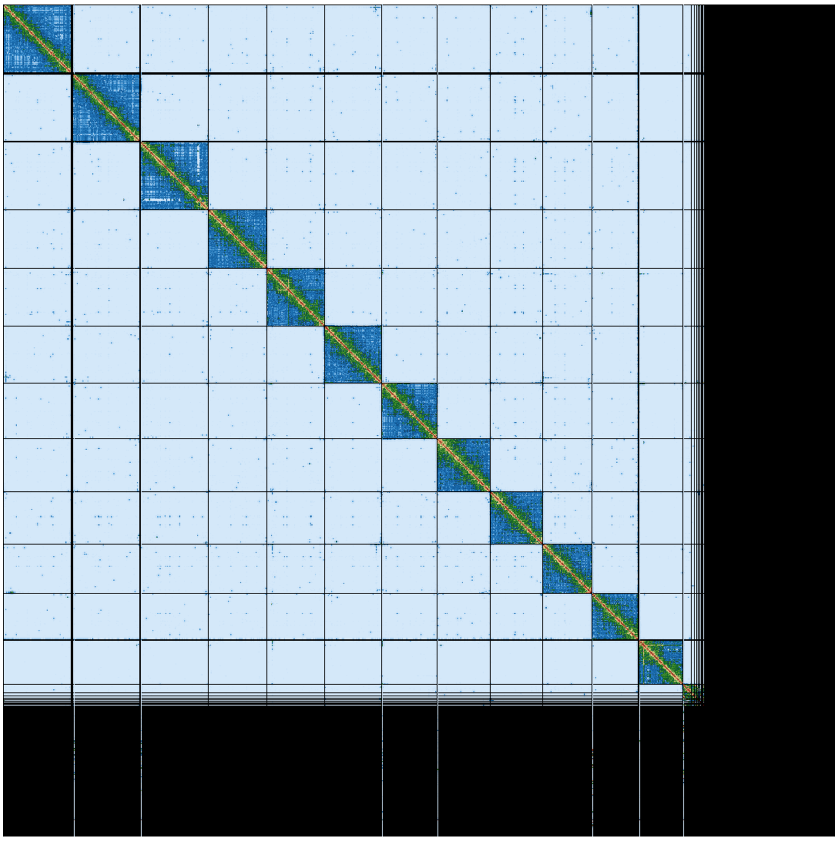
. Interventions/Gb: 38
. Contamination notes: ""
. Other observations: "The assembly of Somatochlora meridionalis (ioSomMeri1.1) is based on 29X PacBio data and 169X of Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA, https://www.erga-biodiversity.eu/) via the Biodiversity Genomics Europe project (BGE, https://biodiversitygenomics.eu/). 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, 14 contigs were identified as contaminants (bacterial), totaling 3.926 Mb (with the largest being 1.620 Mb). Additionally, 1832 regions totaling 129.562 Mb (with the largest being 1.031 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 contaminant sequences and 17 haplotypic regions were removed, totaling 0.292 Mb and 12.431 Mb (with the largest being 0.128 Mb and 1.277 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	2,437,390,213	2,424,650,310
GC %	34.8	34.8
Gaps/Gbp	256.42	266.84
Total gap bp	65,000	74,500
Scaffolds	2,653	2,577
Scaffold N50	151,448,127	161,429,810
Scaffold L50	7	7
Scaffold L90	248	197
Contigs	3,252	3,224
Contig N50	7,673,000	7,849,553
Contig L50	69	68
Contig L90	701	691
QV	62.8632	62.8642
Kmer compl.	83.7788	83.7149
BUSCO sing.	91.7%	92.1%
BUSCO dupl.	1.4%	1.0%
BUSCO frag.	3.7%	3.7%
BUSCO miss.	3.2%	3.2%

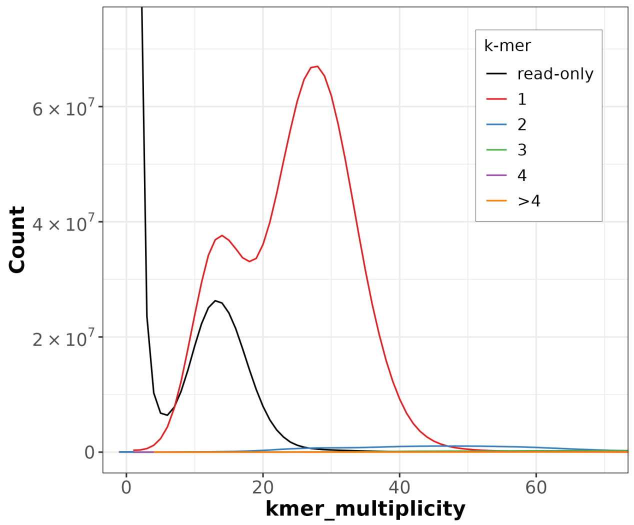
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: insecta_odb12 (genomes:79, BUSCOs:3114)

HiC contact map of curated assembly

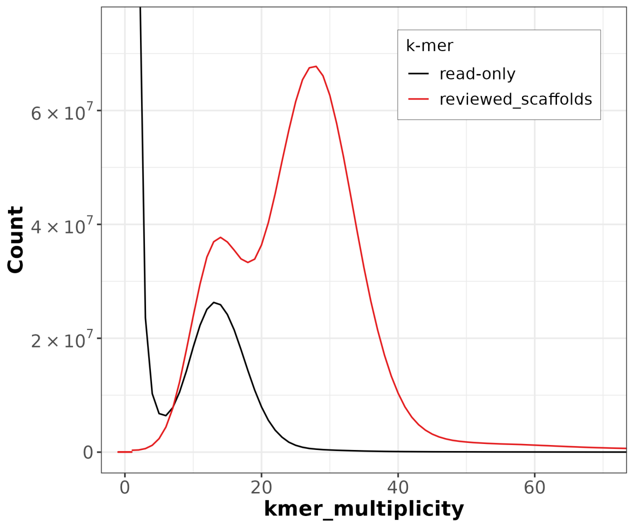


collapsed [LINK]

K-mer spectra of curated assembly

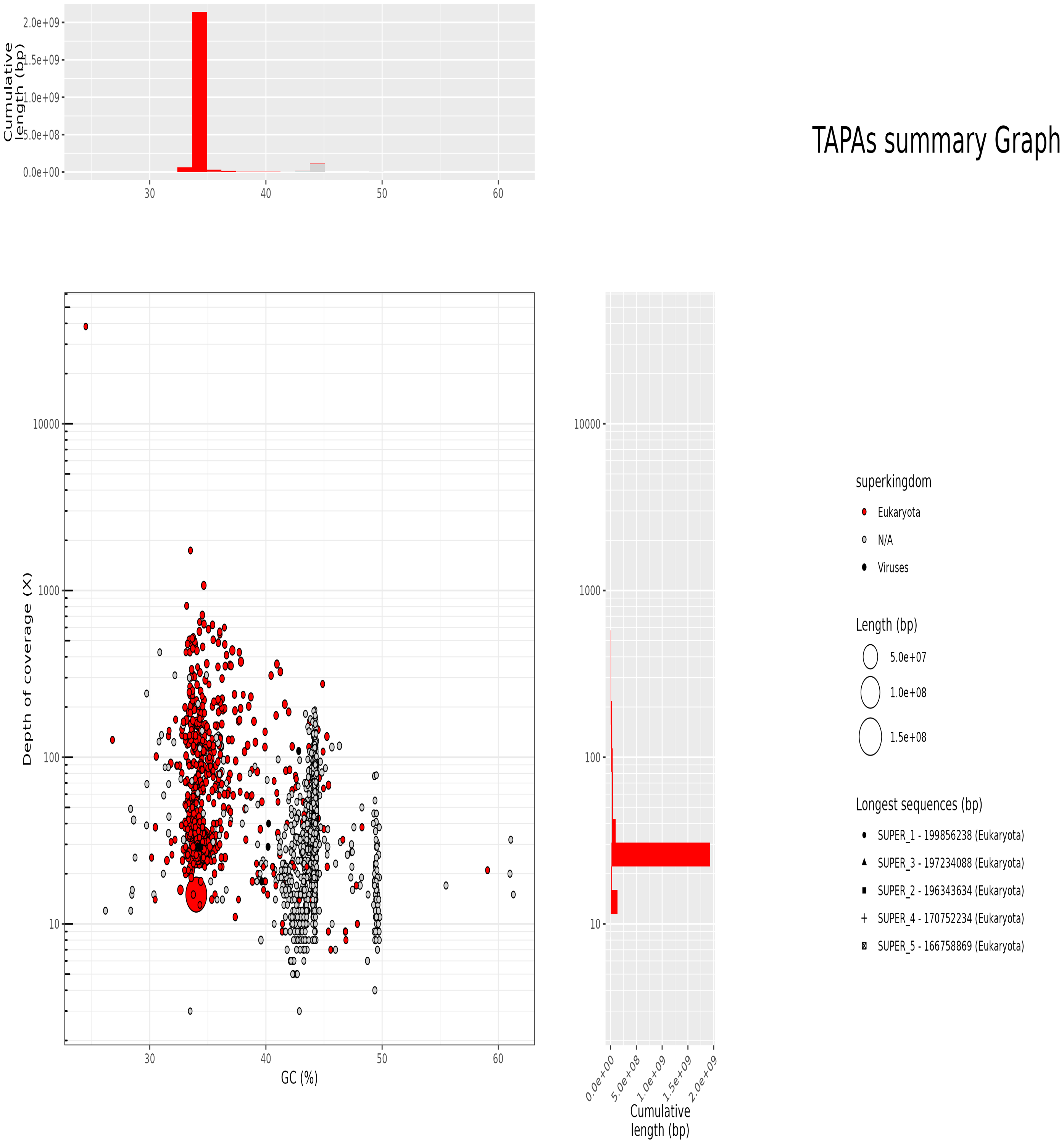


Distribution of k-mer counts per copy numbers found in asm



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	PACBIO Hifi	Arima
Coverage	29	169

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: 2025-09-16 12:41:48 CEST