

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	166372
ToLID	icSisScha5
Species	<i>Sisyphus schaefferi</i>
Class	Insecta
Order	Coleoptera

Genome Traits	Expected	Observed
Haploid size (bp)	383,238,411	398,499,069
Haploid Number	10 (source: direct)	10
Ploidy	2 (source: direct)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q66

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

- . Kmer completeness value is less than 90 for collapsed

Curator notes

- . Interventions/Gb: 20
- . Contamination notes: ""
- . Other observations: "The assembly of *Sisyphus schaefferi* (icSisScha5) is based on 115X PacBio data and 580X 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 Hifiiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. The mitochondrial genome has not been assembled, and mitochondrial contigs may be present in the assembly. No contigs were found to be contaminants (bacterial, archaeal, or viral). Additionally, 301 regions totaling 15.045 Mb (with the largest being 2.354 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 1 haplotypic regions and 18 contaminant sequences were removed, totaling 0.181Mb and 0.412 Mb respectively, (with the largest being 0.181Mb and 0.060 Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

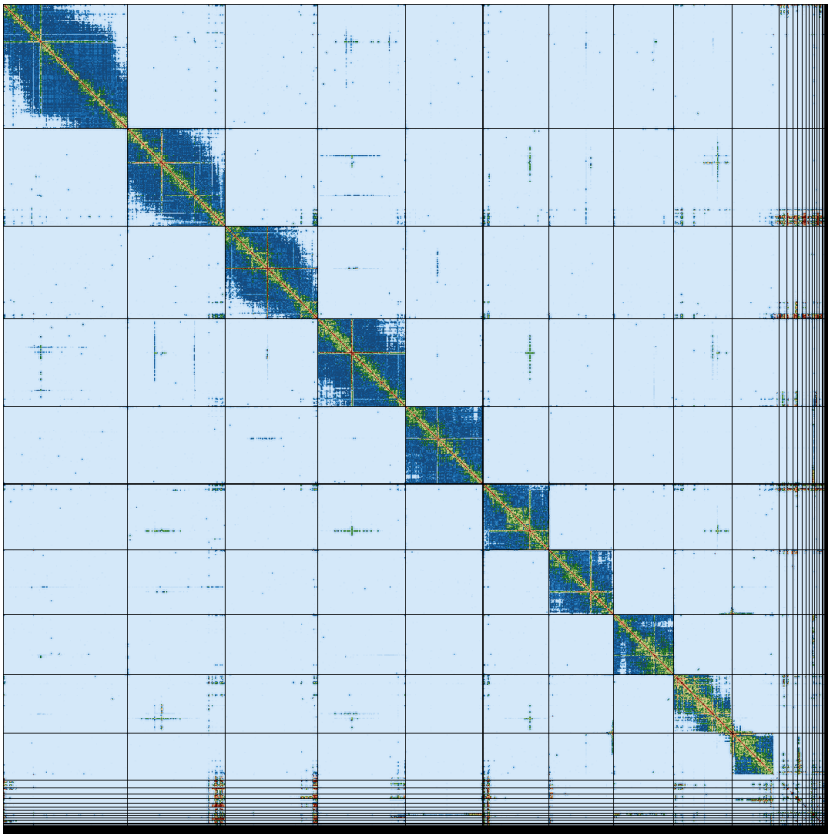
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	399,091,625	398,499,069
GC %	34.72	34.72
Gaps/Gbp	40.09	42.66
Total gap bp	1,600	2,100
Scaffolds	120	101
Scaffold N50	37,205,113	37,205,113
Scaffold L50	5	5
Scaffold L90	10	10
Contigs	136	118
Contig N50	19,673,571	19,673,571
Contig L50	7	7
Contig L90	20	20
QV	66.3288	66.4649
Kmer compl.	70.1479	70.1356
BUSCO sing.	95.3%	97.2%
BUSCO dupl.	0.6%	0.9%
BUSCO frag.	1.5%	0.3%
BUSCO miss.	2.5%	1.6%

Warning! BUSCO versions or lineage datasets are not the same across results:

BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: polyphaga_odb12 (genomes:60, BUSCOs:4010)

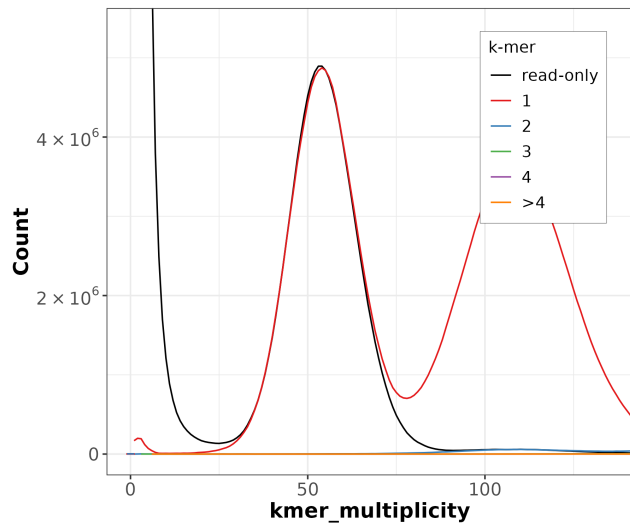
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: polyphaga_odb12 (genomes:60, BUSCOs:4010)

HiC contact map of curated assembly

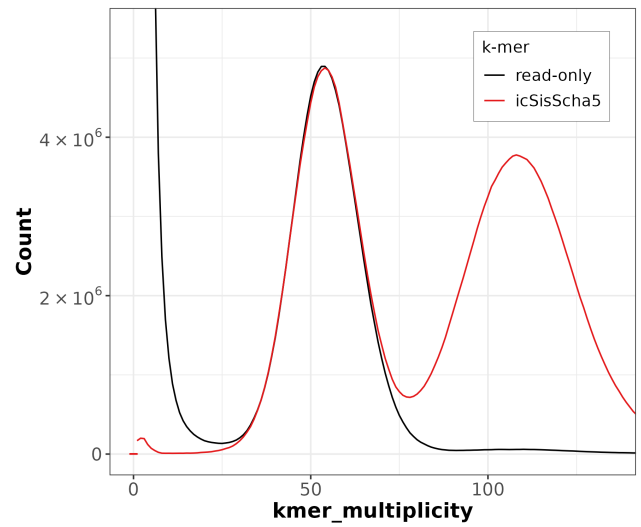


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

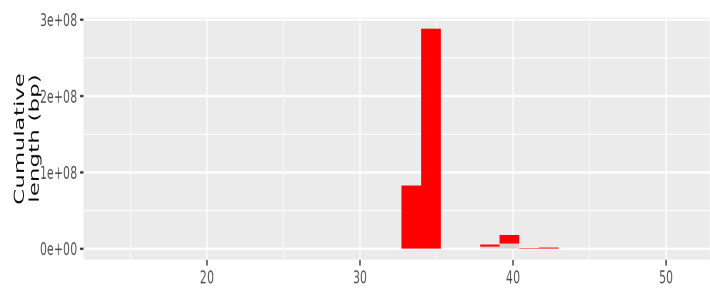


Distribution of k-mer counts per copy numbers found in asm

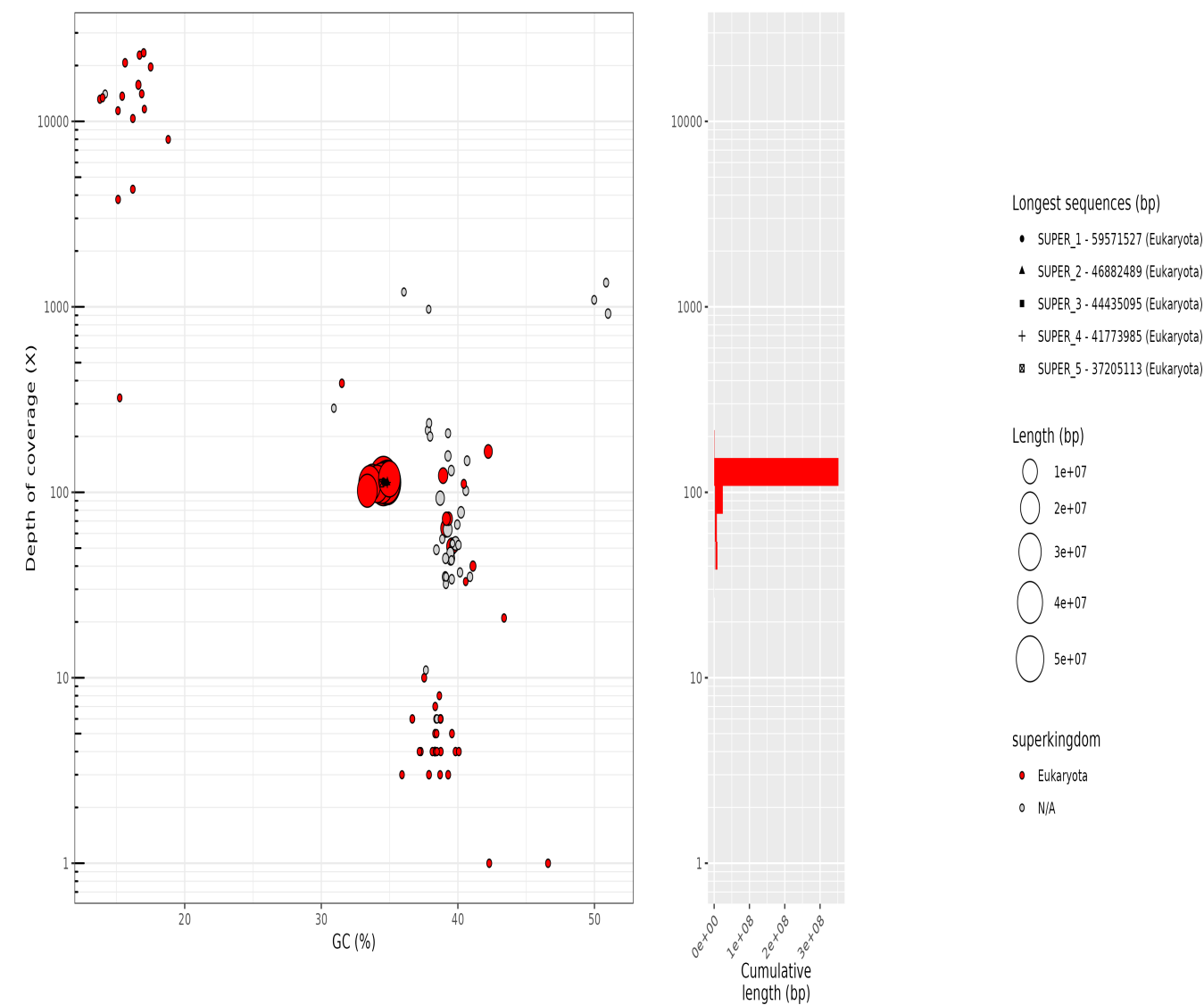


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	Arima
Coverage	115	580

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: Sophie Mangenot

Affiliation: Genoscope

Date and time: 2025-09-20 19:05:38 CEST