ERGA Assembly Report
v24.10.15

Tags: ERGA-BGE

TxID	544679
ToLID	**tzSchErrt1**
Species	Schizoporella errata
Class	Gymnolaemata
Order	Cheilostomatida

Genome Traits	Expected	Observed
Haploid size (bp)	343,009,412	396,390,615
Haploid Number	11 (source: ancestor)	11
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q59

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

. Kmer completeness value is less than 90 for collapsed
. More than 1000 gaps/Gbp for collapsed
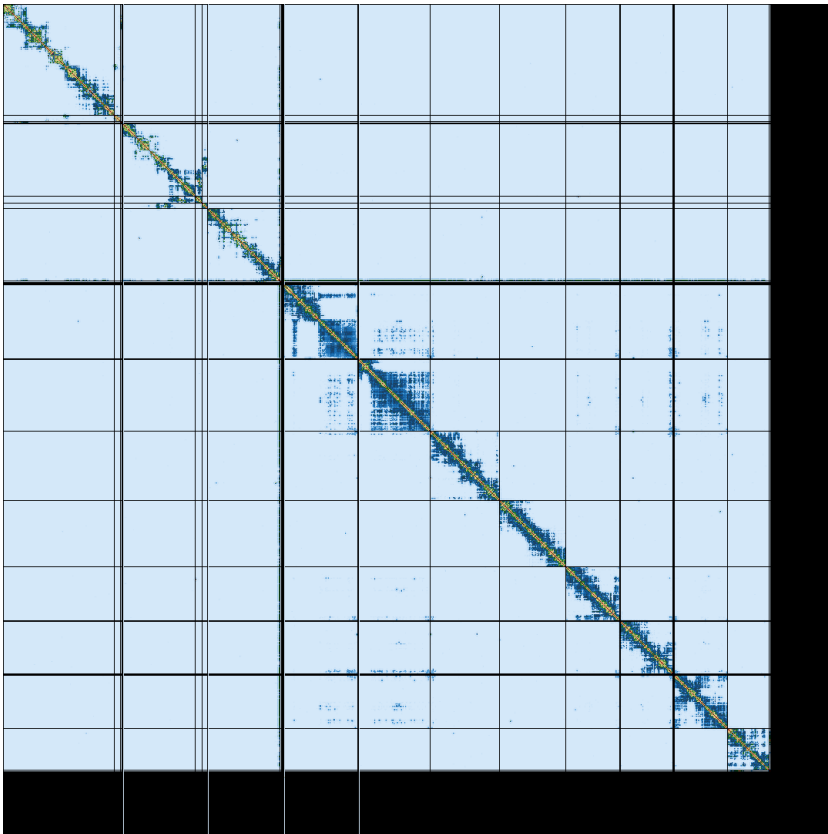. Not 90% of assembly in chromosomes for collapsed

Curator notes

. Interventions/Gb: 181
. Contamination notes: ""
. Other observations: "The assembly of Schizoporella errata (tzSchErrt1) is based on
63,48X PacBio data and 282,50X 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, 2501 contigs were
identified as contaminants (bacterial, archaeal, or viral), totaling 203.001 Mb (with
the largest being 11.897 Mb). Additionally, 633 regions totaling 18.648 Mb (with the
largest being 0.641 Mb) were identified as haplotypic duplications and removed.
During manual curation, 12 haplotypic regions were removed, totaling 2.634306Mb (with
the largest being 1.003723Mb). The telomeric pattern TTTGGGG was identified with
TelFinder and used to generate the Pretext Telomeres tracks. Chromosome-scale
scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

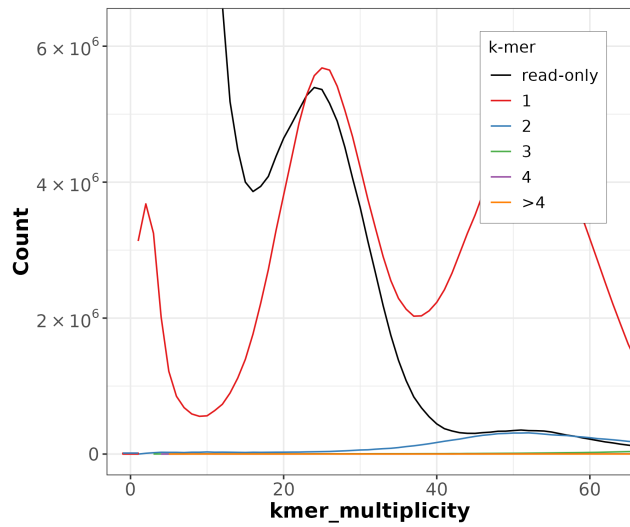| Metrics	Pre-curation collapsed	Curated collapsed
Total bp	399,442,775	396,390,615
GC %	41.78	41.77
Gaps/Gbp	1,011.41	1,104.97
Total gap bp	40,400	48,300
Scaffolds	925	754
Scaffold N50	25,342,757	32,950,285
Scaffold L50	6	6
Scaffold L90	22	13
Contigs	1,329	1,192
Contig N50	3,538,013	3,358,205
Contig L50	31	34
Contig L90	151	158
QV	59.2567	59.3069
Kmer compl.	70.2415	70.0386
BUSCO sing.	91.4%	91.9%
BUSCO dupl.	3.4%	2.8%
BUSCO frag.	2.2%	2.2%
BUSCO miss.	3.0%	3.0%

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

HiC contact map of curated assembly

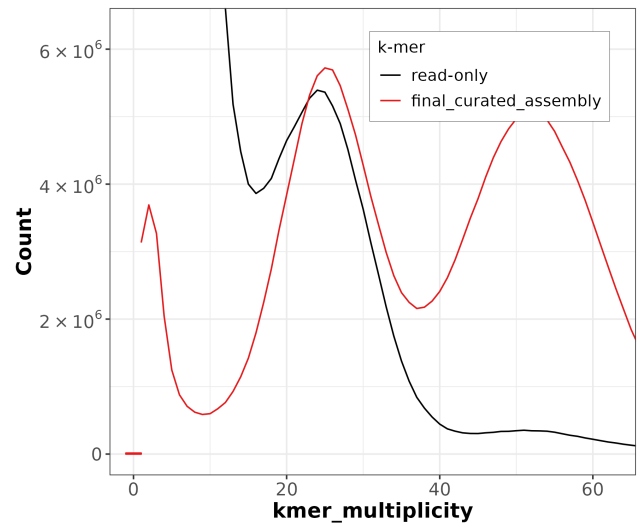**collapsed** [LINK]

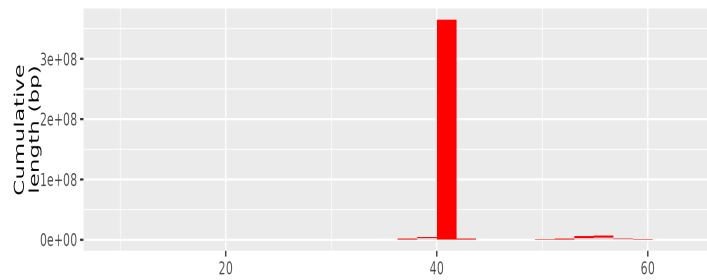# K-mer spectra of curated assembly

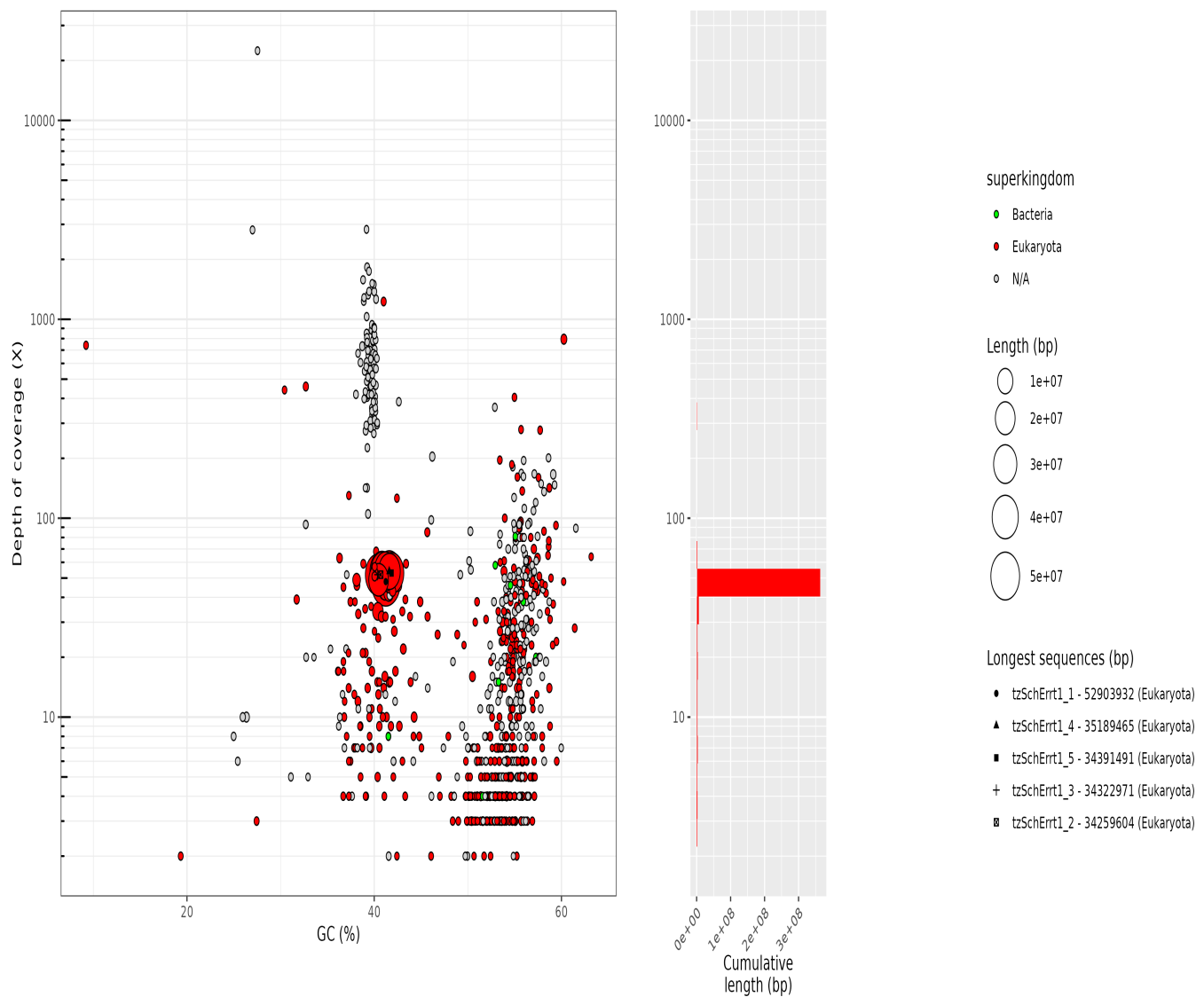Distribution of k-mer counts per copy
numbers found in asm

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	Arima
Coverage	63	282

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: Phuong Doan
Affiliation: Genoscope

Date and time: 2025-11-18 05:33:30 CET