ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	1483045
ToLID	**fKniMra**
Species	Knipowitschia mrakovcici
Class	Actinopteri
Order	Gobiiformes

Genome Traits	Expected	Observed
Haploid size (bp)	390,697,359	918,020,362
Haploid Number	21 (source: ancestor)	23
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q62

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

. Observed Haploid size (bp) has >20% difference with Expected
. Observed Haploid Number is different from Expected

Curator notes

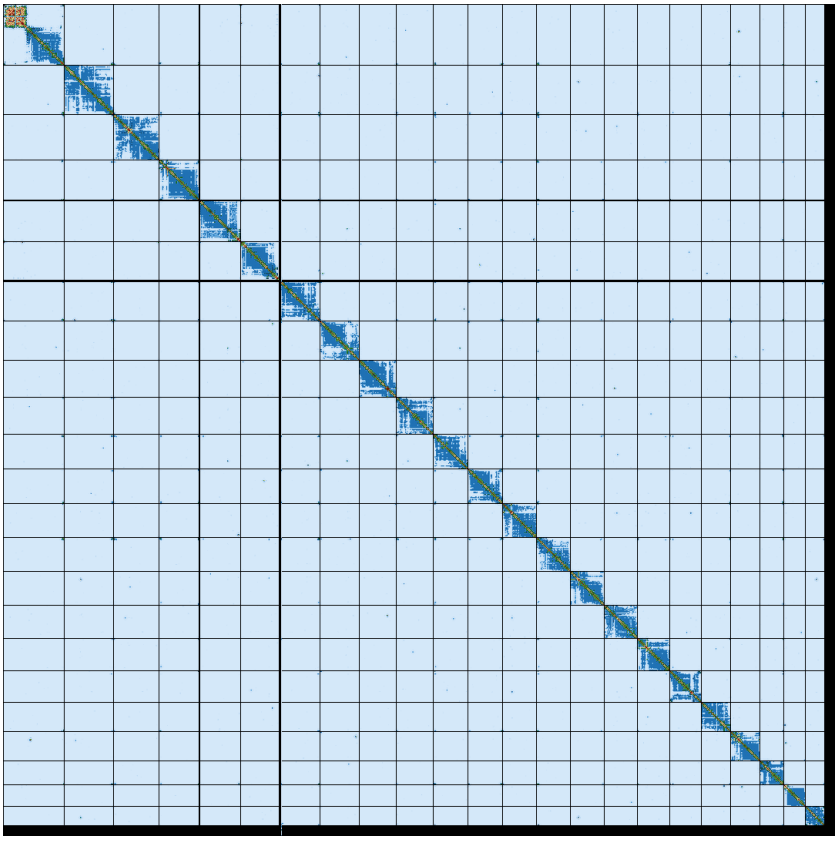
. Interventions/Gb: 94
. Contamination notes: ""
. Other observations: "The assembly of Knipowitschia mrakovcici (fKniMra) is based on 184,48X PacBio data and 459,73X Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA, https://www.erga-biodiversity.eu/) via the Biodiversity 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, 1 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.027 Mb (with the largest being 0.027 Mb). Additionally, 247 regions totaling 25.344 Mb (with the largest being 0.986 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. During manual curation, 2 haplotypic regions were removed, totaling 0.574978Mb (with the largest being 0.524169Mb). 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	918,792,218	918,020,362
GC %	42.39	42.39
Gaps/Gbp	341.75	385.61
Total gap bp	31,400	41,300
Scaffolds	240	146
Scaffold N50	38,227,930	40,594,045
Scaffold L50	11	10
Scaffold L90	21	20
Contigs	554	500
Contig N50	13,107,143	13,107,143
Contig L50	24	24
Contig L90	128	128
QV	62.1186	62.1554
Kmer compl.	97.044	97.0365
BUSCO sing.	90.9%	96.1%
BUSCO dupl.	0.4%	0.8%
BUSCO frag.	2.7%	0.4%
BUSCO miss.	6.0%	2.7%

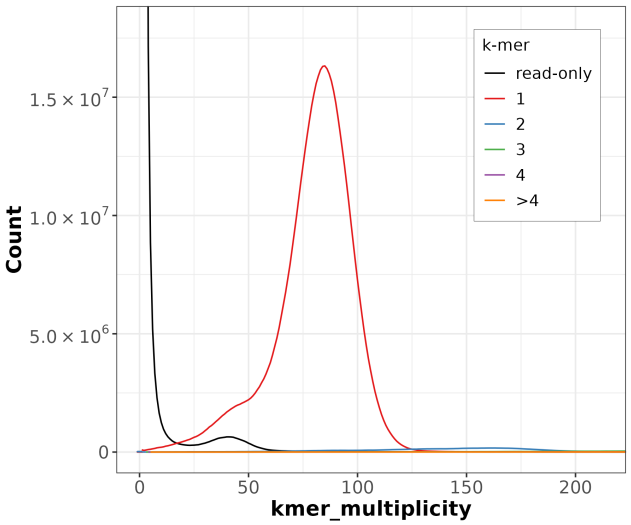
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

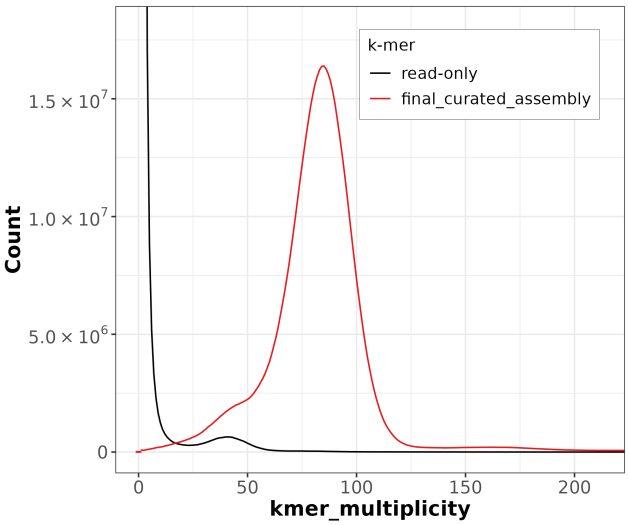


collapsed [LINK]

K-mer spectra of curated assembly

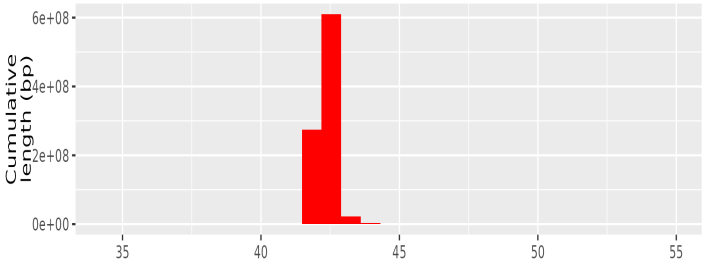


Distribution of k-mer counts per copy numbers found in asm

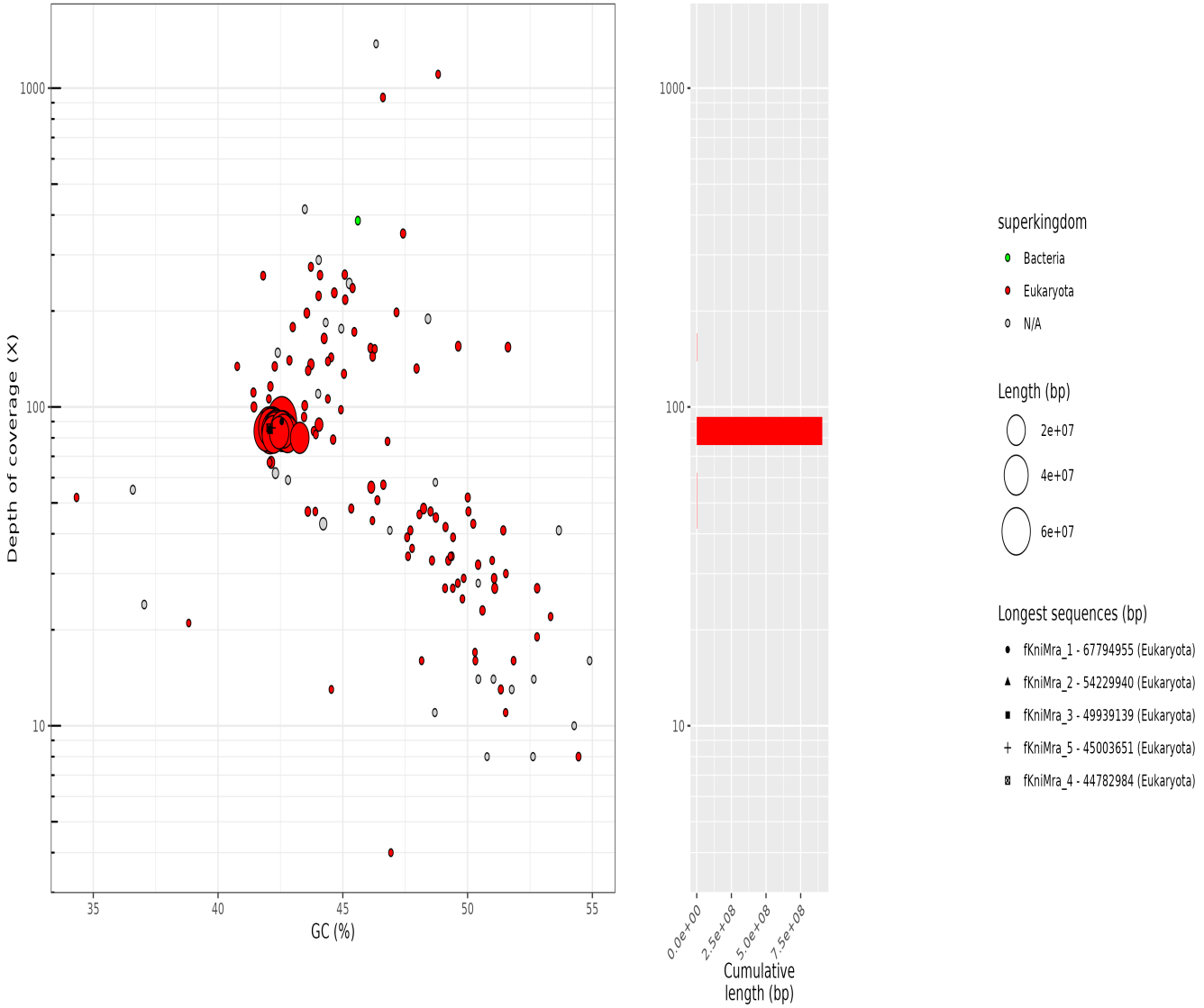


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	184	459

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: 2025-09-26 04:53:27 CEST