

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	1483046
ToLID	<b>fKniRad1</b>
Species	Knipowitschia radovici
Class	Actinopteri
Order	Gobiiformes

Genome Traits	Expected	Observed
Haploid size (bp)	744,287,080	1,076,409,648
Haploid Number	21 (source: ancestor)	0
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.6.Q55

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
- . BUSCO single copy value is less than 90% for collapsed
- . BUSCO duplicated value is more than 5% for collapsed
- . Not 90% of assembly in chromosomes for collapsed

### Curator notes

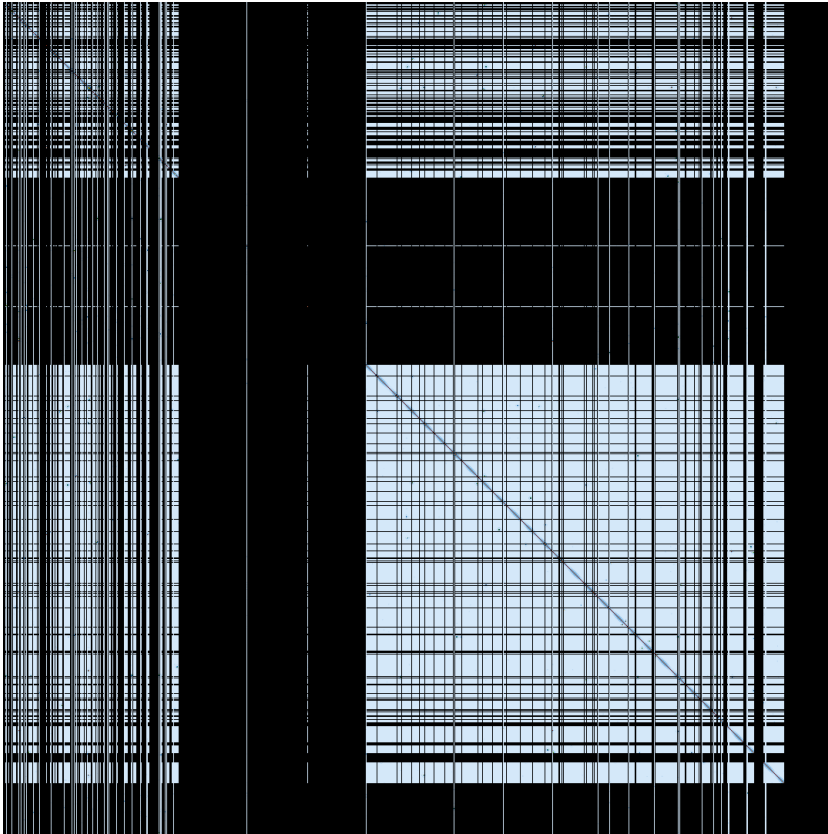
. Interventions/Gb: 0  
. Contamination notes: ""  
. Other observations: "The assembly of *Knipowitschia radovici* (fKniRad1) is based on 40X ONT data and 489X Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA, <https://www.erga-biodiversity.eu/>) via the Biodiversity Genomics Europe project (BGE, <https://biodiversitygenomics.eu/>). The assembly process included the following steps: initial ONT 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, 58 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.501 Mb (with the largest being 0.088 Mb). Additionally, 3735 regions totaling 66.756 Mb (with the largest being 0.121 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. Genome submitted to the contig scale "

## Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,076,582,148	1,076,409,648
GC %	42.25	42.25
Gaps/Gbp	1,602.29	0
Total gap bp	172,500	0
Scaffolds	8,851	7,108
Scaffold N50	3,695,328	5,342,815
Scaffold L50	51	43
Scaffold L90	2,612	2,780
Contigs	10,576	7,108
Contig N50	1,985,558	5,342,815
Contig L50	77	43
Contig L90	4,158	2,780
QV	55.089	55.0886
Kmer compl.	90.7045	90.7078
BUSCO sing.	74.2%	74.2%
BUSCO dupl.	16.4%	17.4%
BUSCO frag.	3.3%	2.8%
BUSCO miss.	6.1%	5.7%

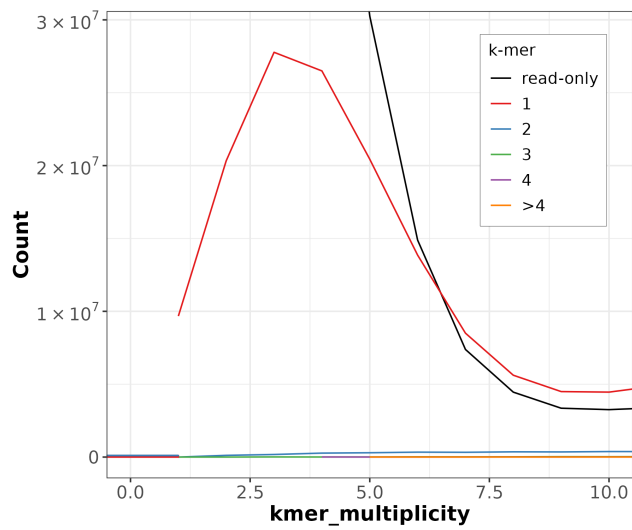
BUSCO: 5.8.2 (euk\_genome\_met, metaeuk) / Lineage: actinopterygii\_odb12 (genomes:75, BUSCOs:7207)

# HiC contact map of curated assembly

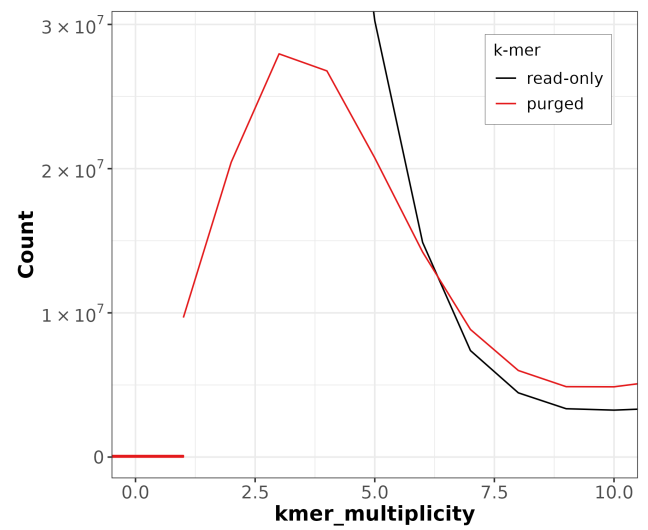


collapsed [\[LINK\]](#)

# K-mer spectra of curated assembly

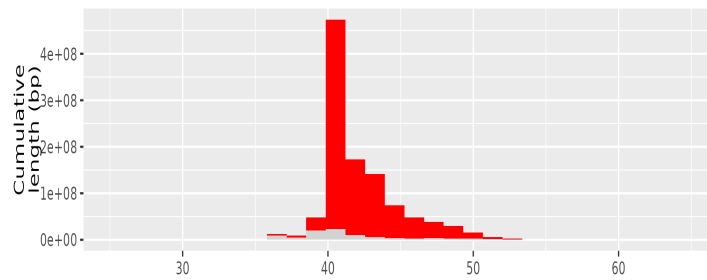


Distribution of k-mer counts per copy numbers found in asm

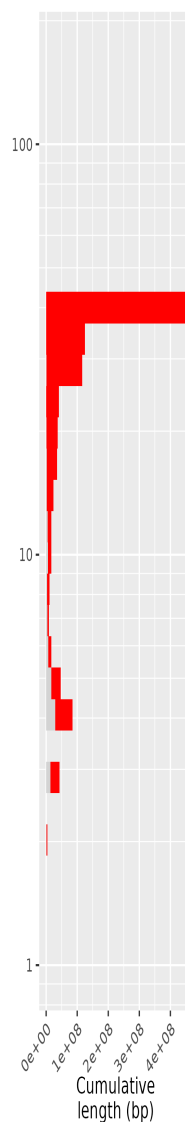
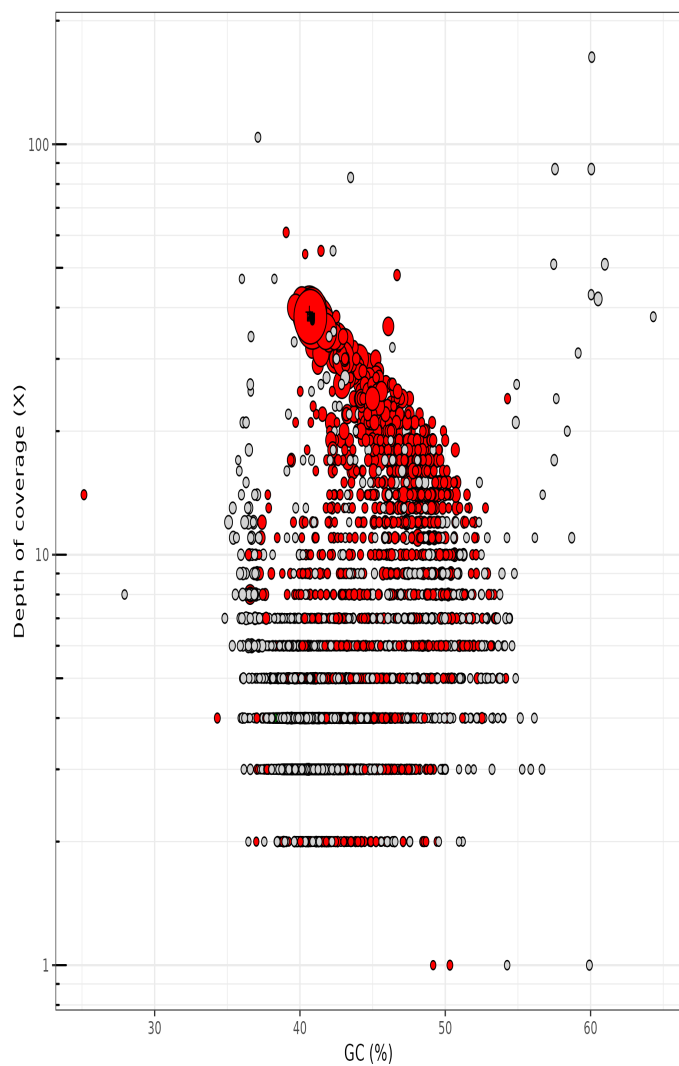


Distribution of k-mer counts coloured by their presence in reads/assemblies

# Post-curation contamination screening



## TAPAs summary Graph



### superkingdom

- Bacteria
- Eukaryota
- N/A

### Longest sequences (bp)

- ptg0001361 - 28868688 (Eukaryota)
- ▲ ptg0020411 - 28027703 (Eukaryota)
- ptg0000671 - 27310492 (Eukaryota)
- + ptg0000031 - 25192182 (Eukaryota)
- ⊠ ptg0000921 - 24813696 (Eukaryota)

### Length (bp)

- 1e+07
- 2e+07

**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	40	489

# 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: Lola Demirdjian

Affiliation: Genoscope

Date and time: 2026-02-27 11:02:55 CET