# ERGA Assembly Report

v24.10.15

Tags: Other[INVALID TAG]

| | |
|---|---|
| TxID | 76338 |
| ToLID | **fThaBif1** |
| Species | Thalassoma bifasciatum |
| Class | Actinopteri |
| Order | Labriformes |

| Genome Traits | Expected | Observed |
|---|---|---|
| Haploid size (bp) | 711,967,443 | 774,617,097 |
| Haploid Number | 24 (source: ancestor) | 24 |
| Ploidy | 2 (source: ancestor) | 2 |
| Sample Sex | Unknown | Unknown |

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q65

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

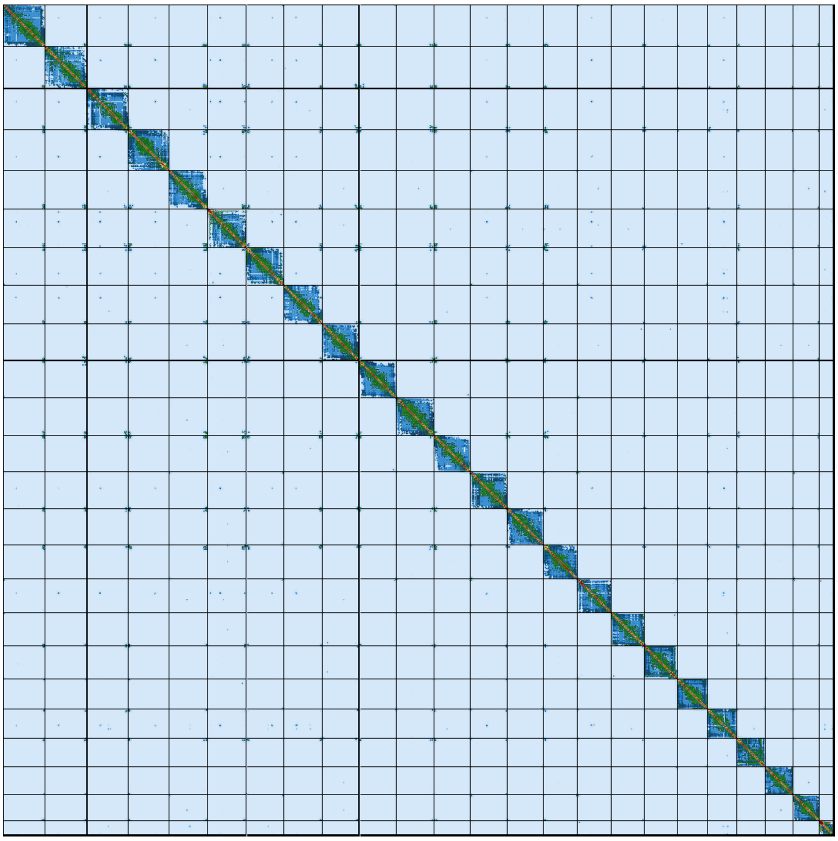
. Interventions/Gb: 60
. Contamination notes: ""
. Other observations: "The assembly of Thalassoma bifasciatum (fThaBif1) was generated as part of the Vertebrate Genomes Project (https://vertebrategenomesproject.org/). Both haplotypes were analyzed and manually improved using Pretext. "

# Quality metrics table

| Metrics | Pre-curation collapsed | Curated collapsed |
|---|---|---|
| Total bp | 781,939,372 | 774,617,097 |
| GC % | 39.79 | 39.8 |
| Gaps/Gbp | 94.64 | 111.02 |
| Total gap bp | 14,800 | 17,200 |
| Scaffolds | 81 | 59 |
| Scaffold N50 | 34,399,871 | 34,308,015 |
| Scaffold L50 | 11 | 11 |
| Scaffold L90 | 21 | 21 |
| Contigs | 155 | 145 |
| Contig N50 | 23,038,093 | 23,178,480 |
| Contig L50 | 14 | 13 |
| Contig L90 | 37 | 34 |
| QV | 65.5036 | 65.0462 |
| Kmer compl. | 75.7393 | 75.6886 |
| BUSCO sing. | 97.3% | 99.4% |
| BUSCO dupl. | 0.2% | 0.2% |
| BUSCO frag. | 0.7% | 0.0% |
| BUSCO miss. | 1.8% | 0.4% |

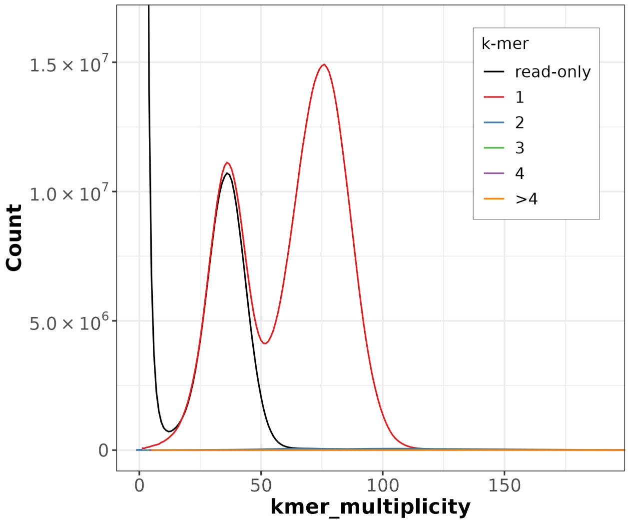
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

# HiC contact map of curated assembly

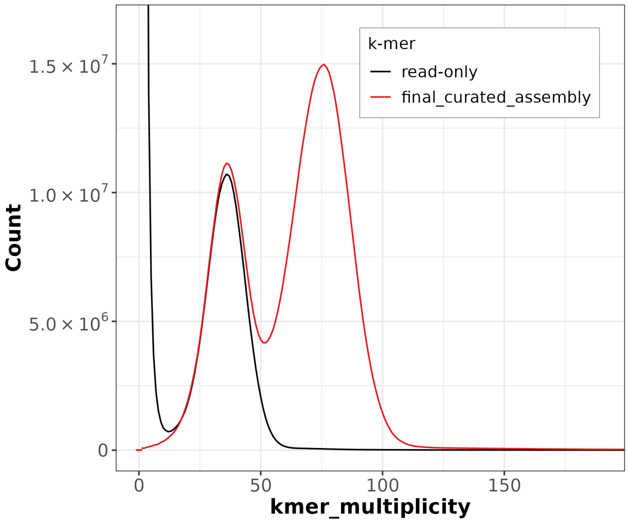


**collapsed** [LINK]

# K-mer spectra of curated assembly

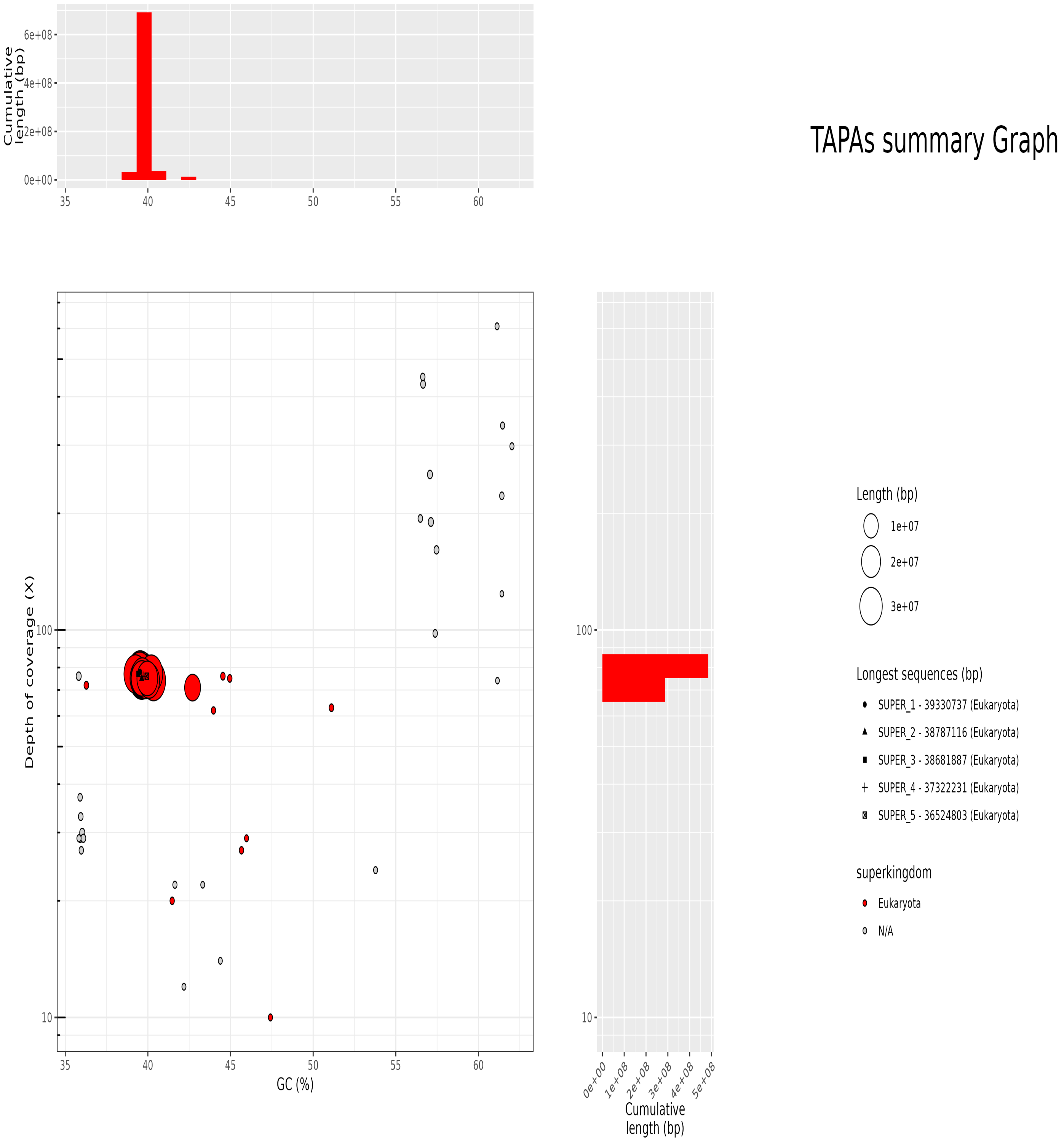


Distribution of k-mer counts per copy numbers found in asm



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

# Post-curation contamination screening



**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 | NA | NA |

# 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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2025-11-14 20:19:45 CET