ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	85420
ToLID	**fPomQug**
Species	Pomatoschistus quagga
Class	Actinopteri
Order	Gobiiformes

Genome Traits	Expected	Observed
Haploid size (bp)	606,248,907	732,558,495
Haploid Number	21 (source: ancestor)	10
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q64

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

. Kmer completeness value is less than 90 for collapsed
. Assembly length loss > 3% for collapsed

Curator notes

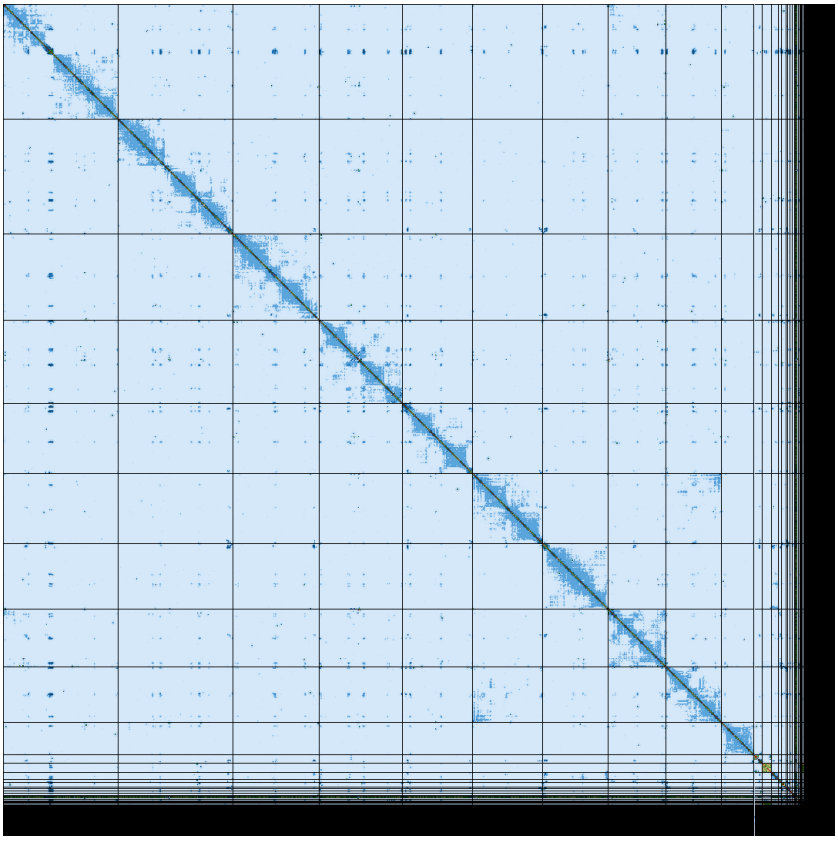
. Interventions/Gb: 141
. Contamination notes: ""
. Other observations: "The assembly of Pomatoschistus quagga (fPomQug) is based on 95,45X PacBio data and 516,48X Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA, https://www.erga-biodiversity.eu/) via the Biodiversity Genomics 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, 52 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.707 Mb (with the largest being 0.076 Mb). Additionally, 465 regions totaling 49.724 Mb (with the largest being 1.689 Mb) were identified as haplotypic duplications and removed. During manual curation, 49 haplotypic regions were removed, totaling 33.375178Mb (with the largest being 0.884338Mb). 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	766,245,325	732,558,495
GC %	43.31	43.19
Gaps/Gbp	438.5	428.63
Total gap bp	33,600	35,000
Scaffolds	360	256
Scaffold N50	64,181,764	61,878,084
Scaffold L50	5	5
Scaffold L90	14	10
Contigs	696	570
Contig N50	17,752,867	16,601,000
Contig L50	17	17
Contig L90	118	97
QV	63.8376	64.106
Kmer compl.	78.9366	77.3255
BUSCO sing.	86.9%	95.1%
BUSCO dupl.	3.6%	0.9%
BUSCO frag.	2.7%	0.4%
BUSCO miss.	6.7%	3.5%

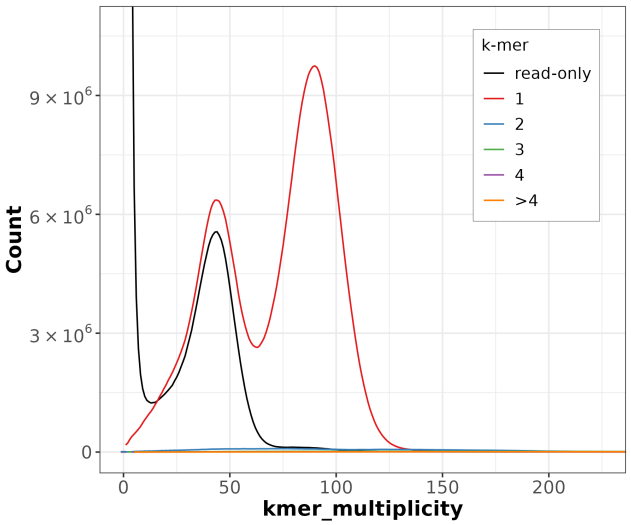
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

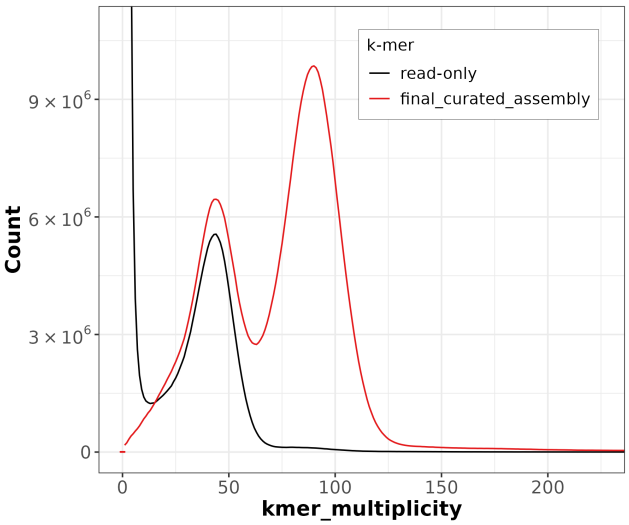


collapsed [LINK]

K-mer spectra of curated assembly

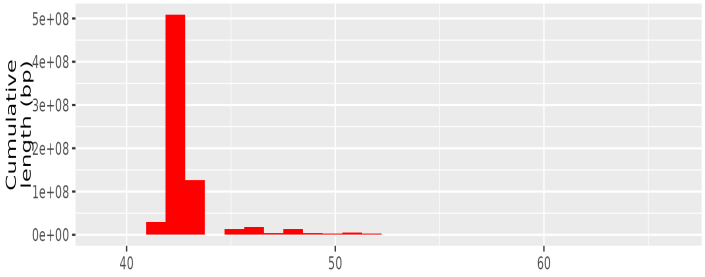


Distribution of k-mer counts per copy numbers found in asm

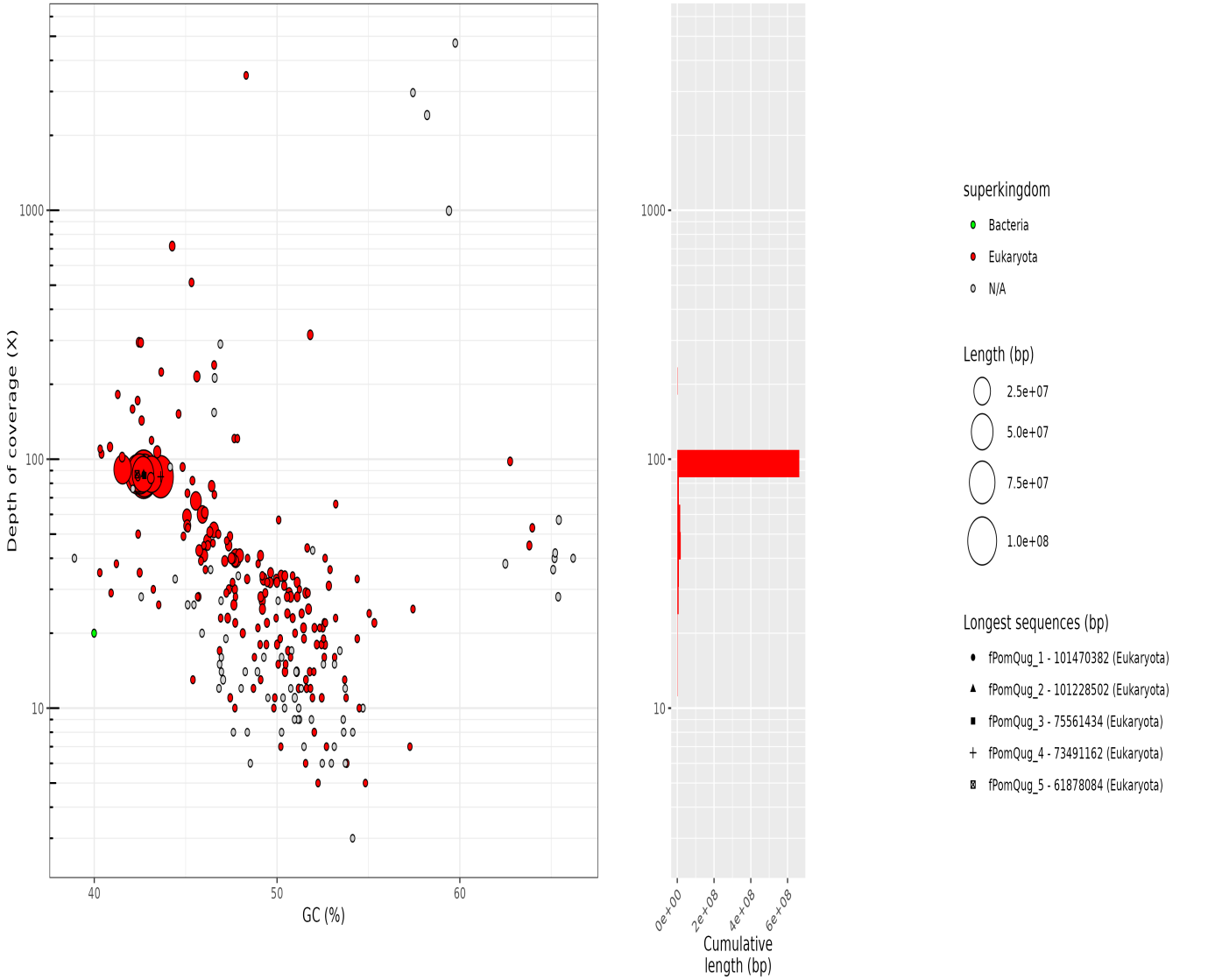


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	95	516

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: 2025-11-26 08:37:02 CET