ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	171871
ToLID	**fParGat1**
Species	Parablennius gattorugine
Class	Actinopteri
Order	Blenniiformes

Genome Traits	Expected	Observed
Haploid size (bp)	650,747,671	768,167,284
Haploid Number	19 (source: ancestor)	24
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q58

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

. Observed Haploid Number is different from Expected

Curator notes

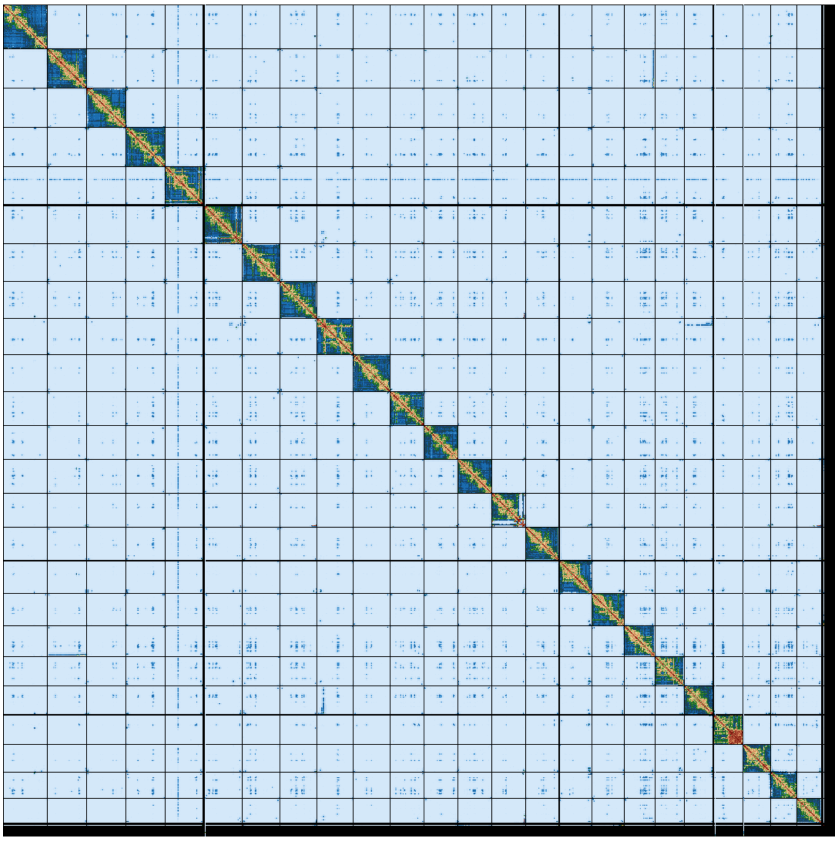
. Interventions/Gb: 66
. Contamination notes: ""
. Other observations: "The assembly of Parablennius gattorugine (fParGat1) is based on 56X PacBio data and Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included the following steps: initial PacBio 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, no contigs were identified as contaminants (bacterial, archaeal, or viral). Additionally, 382 regions totaling 19 Mb (with the largest being 1.5 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 4 haplotypic regions was removed,totaling 3.5 Mb (with the largest being 2.7 Mb "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	771,931,693	768,167,284
GC %	42.45	42.46
Gaps/Gbp	887.38	942.5
Total gap bp	68,500	76,600
Scaffolds	325	253
Scaffold N50	31,208,515	31,625,660
Scaffold L50	12	11
Scaffold L90	23	22
Contigs	1,010	977
Contig N50	6,015,568	6,015,568
Contig L50	31	31
Contig L90	251	253
QV	45.7805	58.3323
Kmer compl.	93.1668	93.4652
BUSCO sing.	95.2%	95.8%
BUSCO dupl.	1.0%	0.3%
BUSCO frag.	1.3%	1.2%
BUSCO miss.	2.6%	2.6%

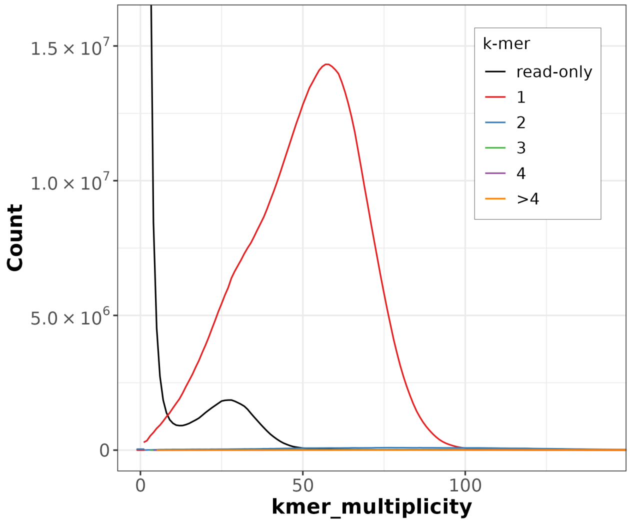
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

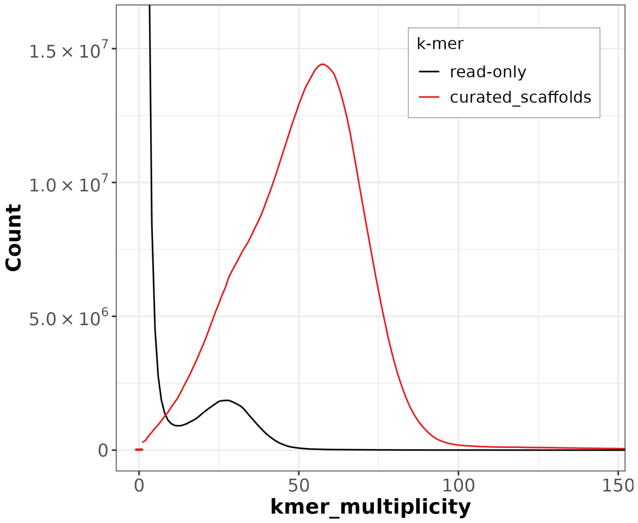


collapsed [LINK]

K-mer spectra of curated assembly

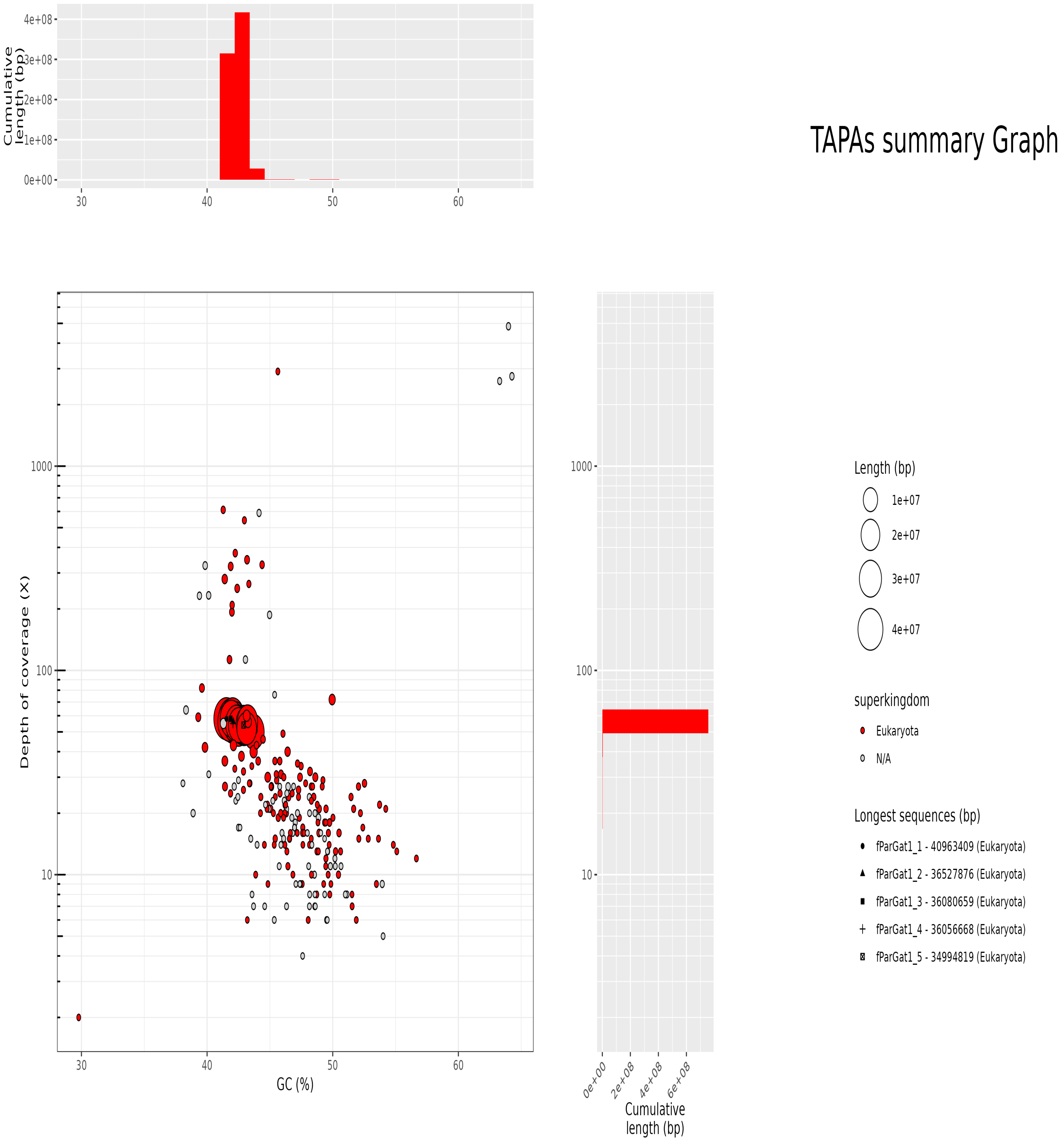


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	PACBIO Hifi	Arima
Coverage	56	181

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: Caroline Menguy
Affiliation: Genoscope

Date and time: 2025-04-18 00:48:53 CEST