

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	2566713
ToLID	fGobHai1
Species	Gobioclinus haitiensis
Class	Actinopteri
Order	Blenniiformes

Genome Traits	Expected	Observed
Haploid size (bp)	592,142,888	718,314,559
Haploid Number	19 (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.Q50

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
- . Not 90% of assembly in chromosomes for collapsed

Curator notes

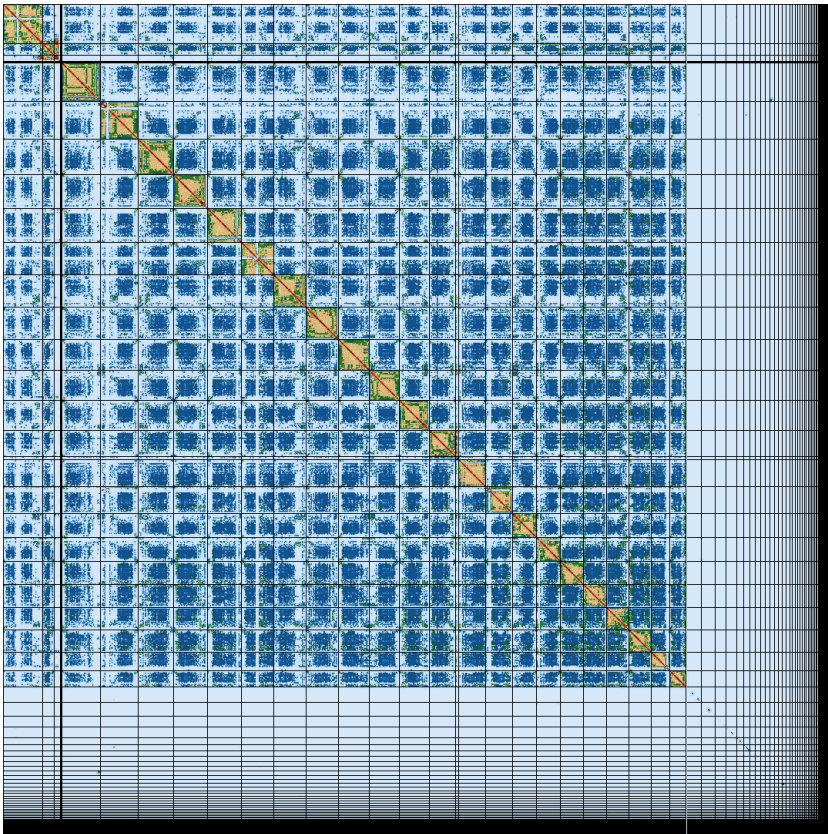
. Interventions/Gb: 39
. Contamination notes: ""
. Other observations: "The assembly of Gobioclinus haitiensis (fGobHai1) is based on 76X PacBio data and 221X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial PacBio assembly generation with Hifiiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. No contigs were found to be contaminants (bacterial, archaeal, or viral). Additionally, 160 regions totaling 43.676 Mb (with the largest being 3.026 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, 5 haplotypic regions were removed, totaling 10.1Mb, (with the largest being 4.6Mb). 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	728,510,911	718,314,559
GC %	46.18	46.19
Gaps/Gbp	130.4	129.47
Total gap bp	10,500	11,600
Scaffolds	131	127
Scaffold N50	23,372,818	23,372,818
Scaffold L50	13	13
Scaffold L90	34	33
Contigs	220	220
Contig N50	19,327,000	19,327,000
Contig L50	16	16
Contig L90	54	56
QV	50.0468	50.0805
Kmer compl.	80.637	80.3208
BUSCO sing.	94.2%	96.0%
BUSCO dupl.	2.3%	0.4%
BUSCO frag.	1.1%	1.1%
BUSCO miss.	2.4%	2.4%

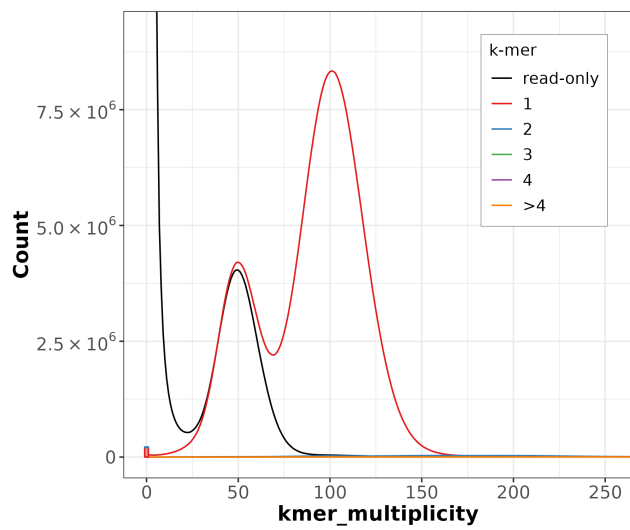
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

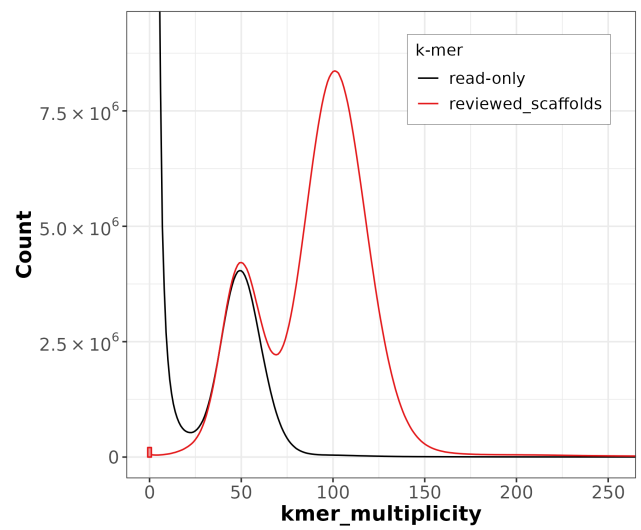


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

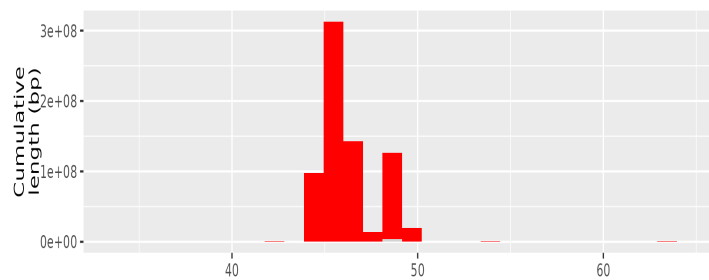


Distribution of k-mer counts per copy numbers found in asm

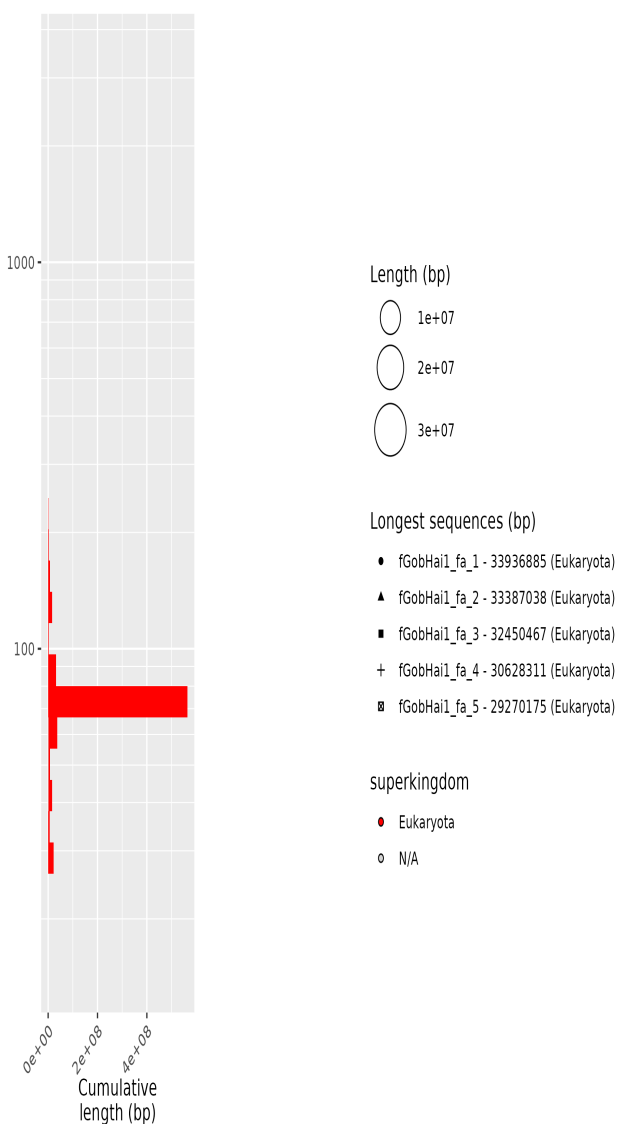
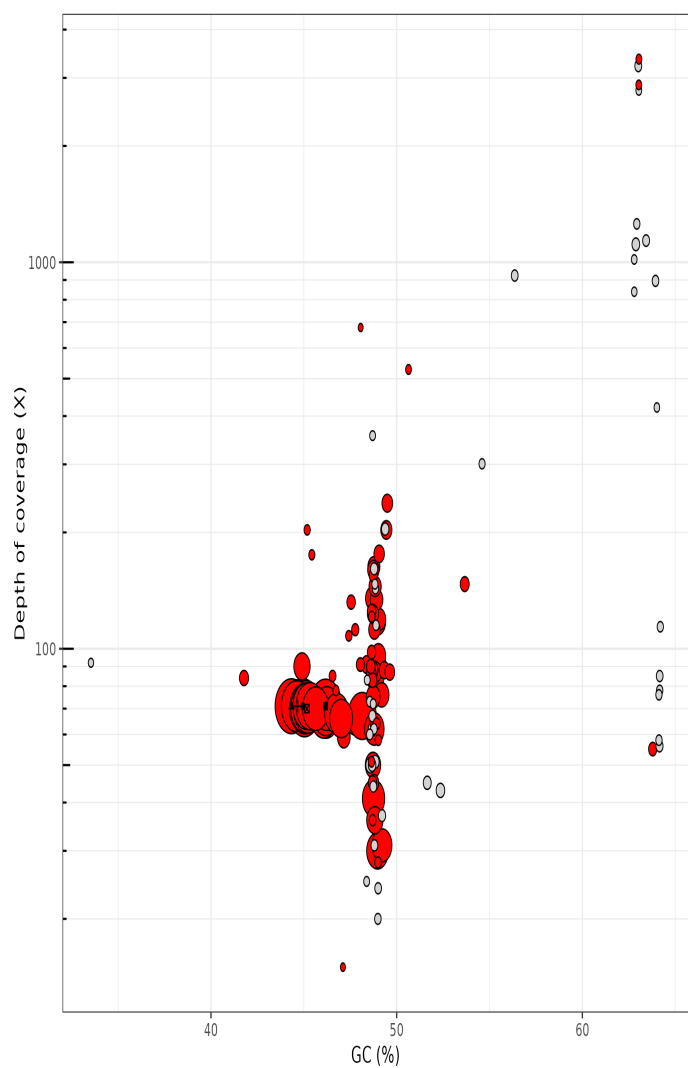


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	76	221

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: Benjamin Istace

Affiliation: Genoscope

Date and time: 2025-09-30 23:08:52 CEST