

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	316389
ToLID	fHalPoe1
Species	Halichoeres poeyi
Class	Actinopteri
Order	Labriiformes

Genome Traits	Expected	Observed
Haploid size (bp)	777,318,343	785,400,893
Haploid Number	24 (source: direct)	24
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q48

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

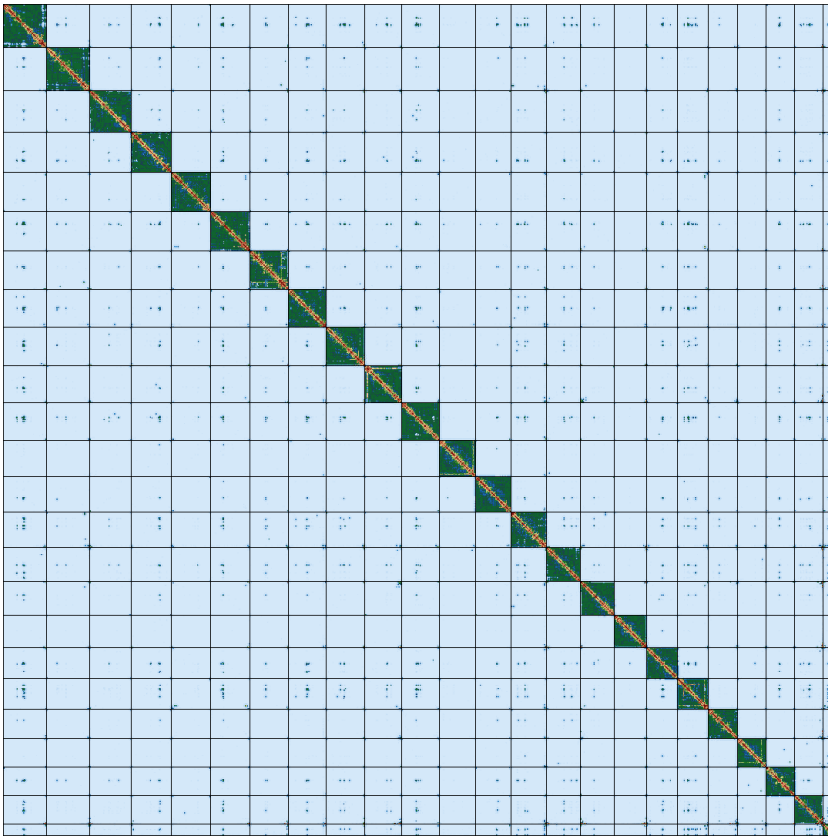
- . Interventions/Gb: 1
- . Contamination notes: ""
- . Other observations: "The assembly of Halichoeres poeyi (fHalPoe1) is based on 86X PacBio data and 175X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial 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, 2 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.129 Mb (with the largest being 0.07 Mb). Additionally, 39 regions totaling 8.331 Mb (with the largest being 3.302 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. 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	785,417,130	785,400,893
GC %	40.88	40.88
Gaps/Gbp	14.01	15.28
Total gap bp	1,100	1,300
Scaffolds	31	30
Scaffold N50	35,198,123	35,198,123
Scaffold L50	11	11
Scaffold L90	21	21
Contigs	42	42
Contig N50	33,572,800	33,572,800
Contig L50	11	11
Contig L90	22	22
QV	48.7213	48.7212
Kmer compl.	73.0675	73.0675
BUSCO sing.	96.4%	96.4%
BUSCO dupl.	0.2%	0.2%
BUSCO frag.	0.6%	0.6%
BUSCO miss.	2.9%	2.9%

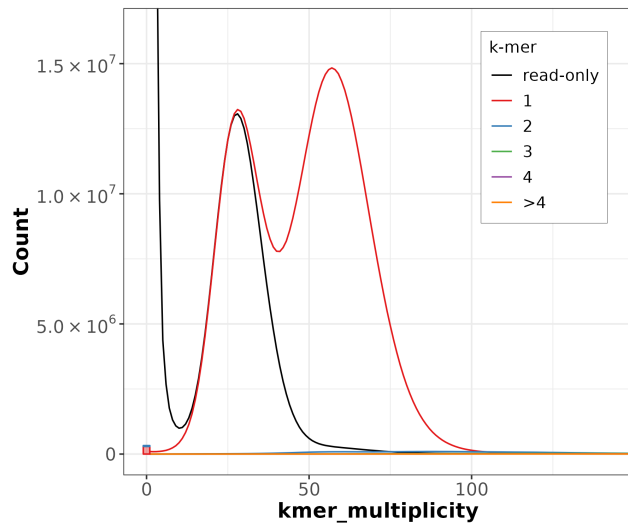
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

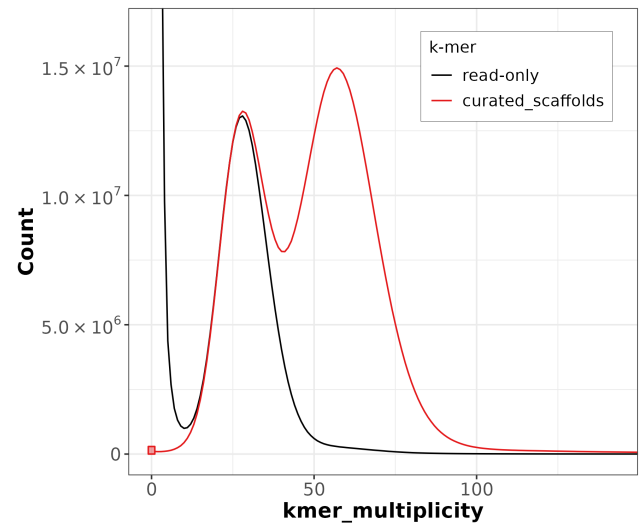


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

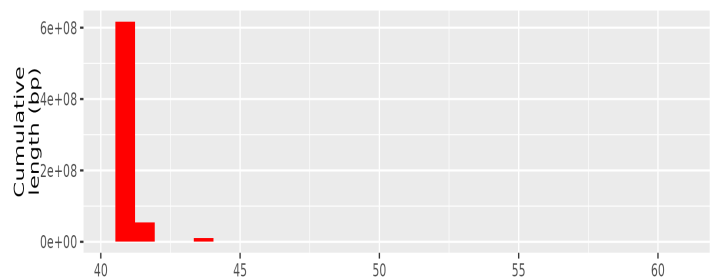


Distribution of k-mer counts per copy numbers found in asm

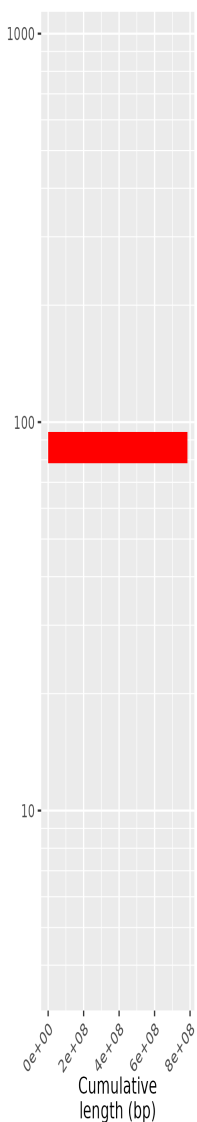
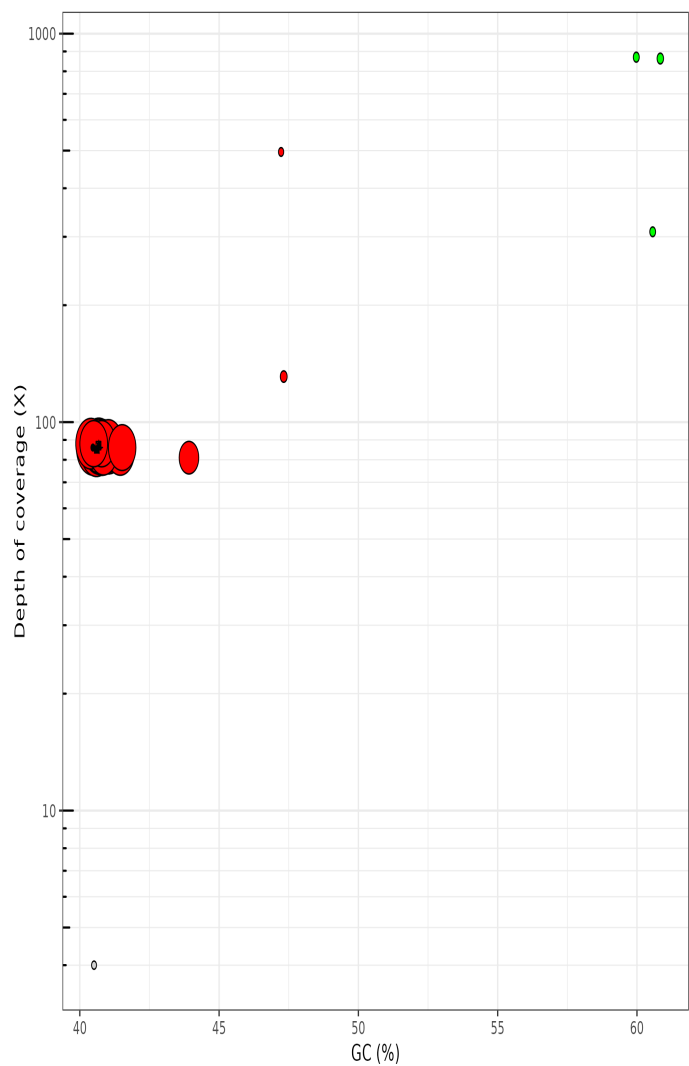


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



Longest sequences (bp)

- fHalPoe1_1 - 41145859 (Eukaryota)
- ▲ fHalPoe1_2 - 40734800 (Eukaryota)
- fHalPoe1_3 - 39019540 (Eukaryota)
- + fHalPoe1_4 - 37941311 (Eukaryota)
- ▣ fHalPoe1_5 - 37168138 (Eukaryota)

superkingdom

- Bacteria
- Eukaryota
- N/A

Length (bp)

- 1e+07
- 2e+07
- 3e+07
- 4e+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	PACBIO Hifi	Arima
Coverage	86	175

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: 2025-07-24 14:34:05 CEST