

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	154550
ToLID	fLepWhi1
Species	Lepidorhombus whiffiagonis
Class	Actinopteri
Order	Pleuronectiformes

Genome Traits	Expected	Observed
Haploid size (bp)	561,789,746	540,570,312
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.Q39

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

- . QV value is less than 40 for collapsed
- . Kmer completeness value is less than 90 for collapsed

Curator notes

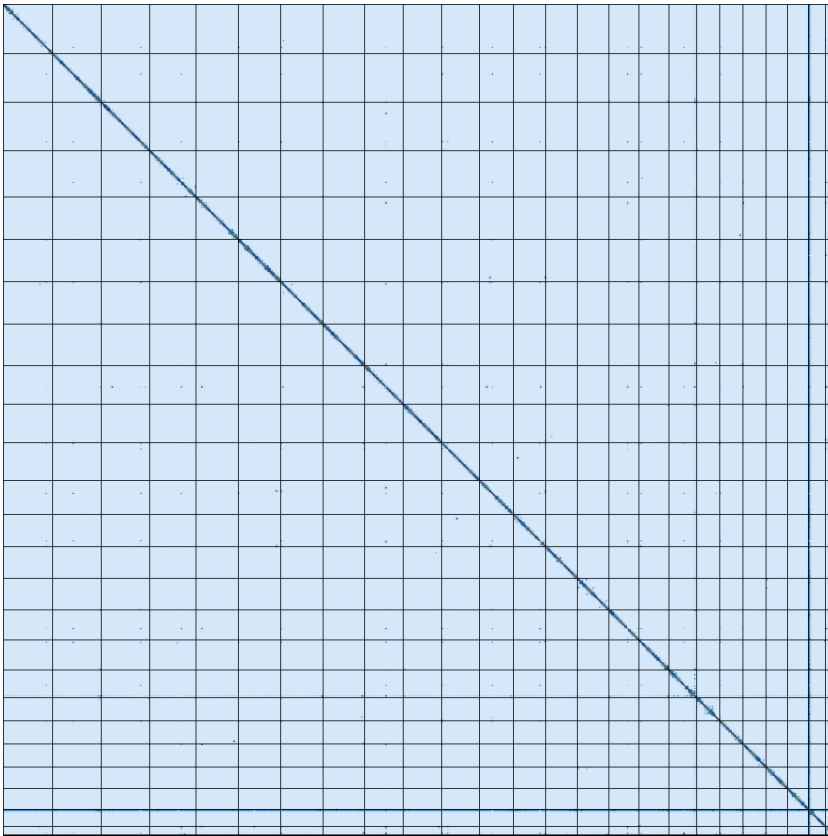
- . Interventions/Gb: 9
- . Contamination notes: ""
- . Other observations: "The assembly of Lepidorhombus whiffiagonis (fLepWhi1) is based on 41X ONT data and Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>).The assembly process included the following steps: initial ONT assembly generation with Nextdenovo, 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, 28 regions totaling 5 Mb (with the largest being 0.799 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. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	540,591,285	540,570,312
GC %	44.45	44.45
Gaps/Gbp	68.44	74
Total gap bp	3,700	4,400
Scaffolds	36	33
Scaffold N50	25,103,352	25,032,925
Scaffold L50	10	10
Scaffold L90	20	21
Contigs	73	73
Contig N50	17,714,167	17,714,167
Contig L50	12	12
Contig L90	29	29
QV	38.7252	39.3955
Kmer compl.	87.4864	86.9349
BUSCO sing.	96.9%	96.9%
BUSCO dupl.	0.3%	0.3%
BUSCO frag.	0.8%	0.8%
BUSCO miss.	2.0%	2.0%

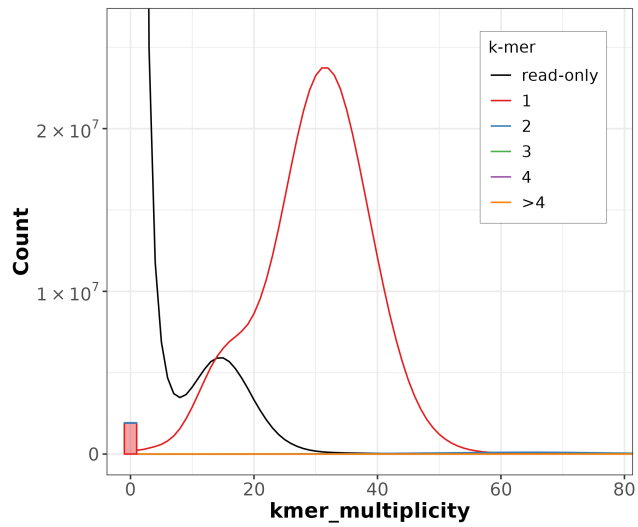
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

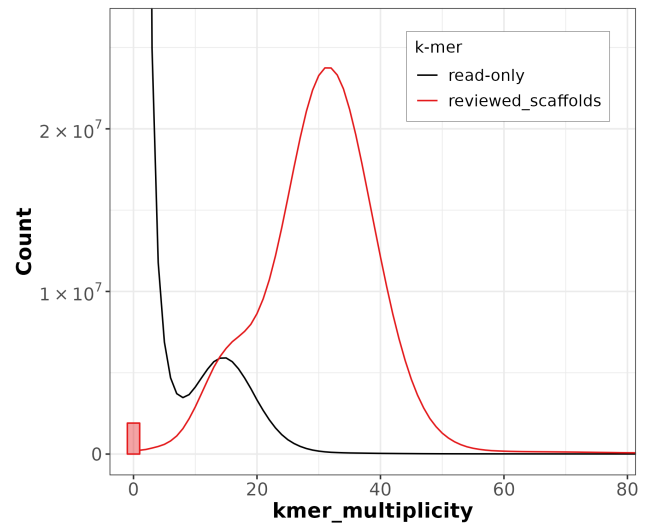


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

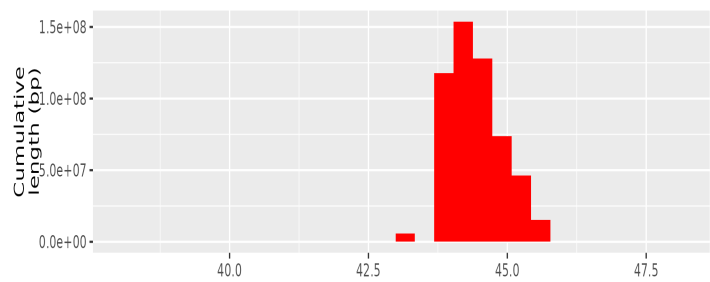


Distribution of k-mer counts per copy numbers found in asm

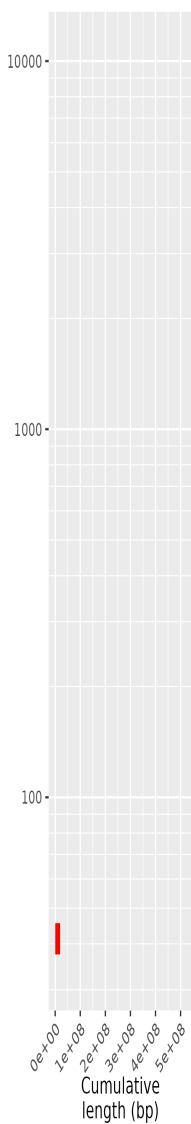
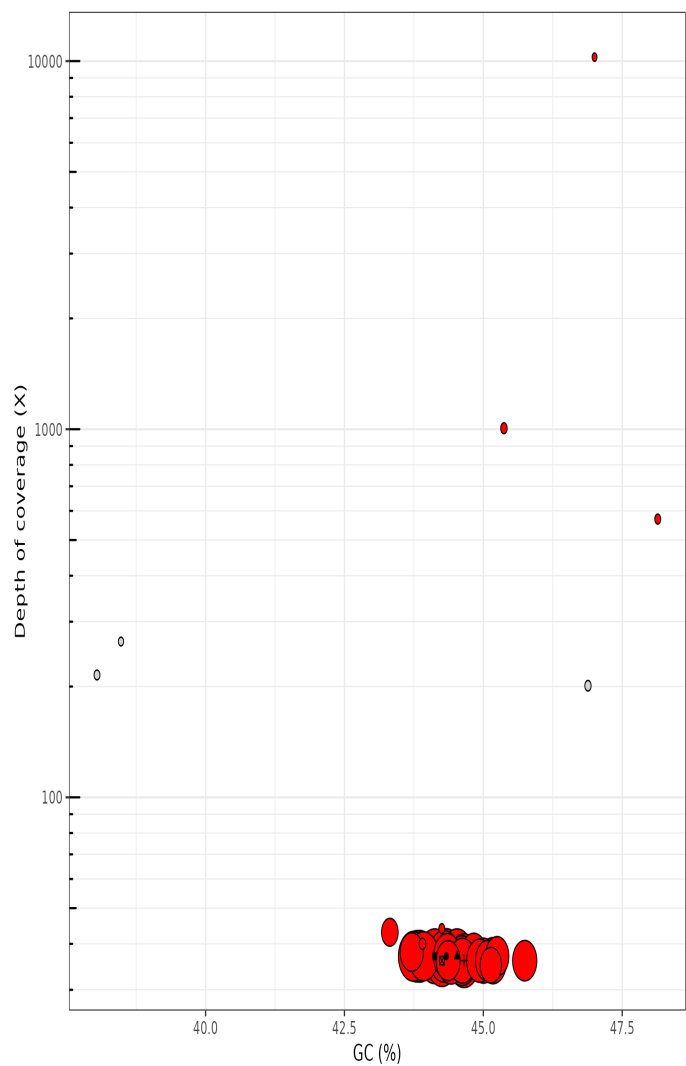


Distribution of k-mer counts coloured by their presence in reads/assemblies

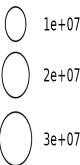
Post-curation contamination screening



TAPAs summary Graph



Length (bp)



Longest sequences (bp)

- fLepWhi1_1 - 32162095 (Eukaryota)
- ▲ fLepWhi1_2 - 31551915 (Eukaryota)
- fLepWhi1_3 - 31549387 (Eukaryota)
- + fLepWhi1_4 - 30187417 (Eukaryota)
- ⊠ fLepWhi1_5 - 27632939 (Eukaryota)

superkingdom

- Eukaryota
- N/A

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	41	195

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 00:32:44 CEST