ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1564611
ToLID	**xgFelPict1**
Species	Felimare picta
Class	Gastropoda
Order	Nudibranchia

Genome Traits	Expected	Observed
Haploid size (bp)	1,663,143,705	1,434,865,120
Haploid Number	12 (source: ancestor)	13
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q34

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

. Observed Haploid Number is different from Expected

. QV value is less than 40 for collapsed
. Kmer completeness value is less than 90 for collapsed
. BUSCO single copy value is less than 90% for collapsed

Curator notes

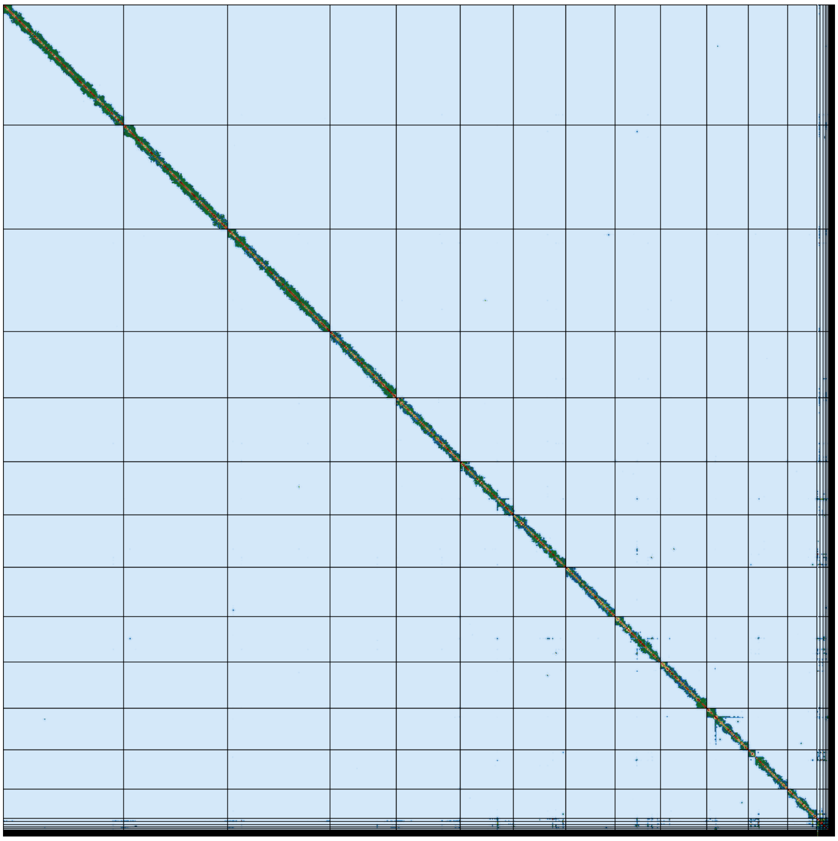
. Interventions/Gb: 48
. Contamination notes: ""
. Other observations: "The assembly of Felimare picta (xgFelPict1) is based on 28X ONT data and 29X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio assembly generation with Nextdenovo, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 5 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 450 Mb (with the largest being 143 kb). Additionally, 605 regions totaling 92 Mb (with the largest being 2 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 890 kb (with the largest being 385 kb). 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	1,437,924,397	1,434,865,120
GC %	38.03	38.03
Gaps/Gbp	860.27	866.28
Total gap bp	123,700	128,000
Scaffolds	156	76
Scaffold N50	109,353,802	110,165,539
Scaffold L50	5	5
Scaffold L90	12	12
Contigs	1,393	1,319
Contig N50	3,724,955	3,760,560
Contig L50	112	111
Contig L90	584	580
QV	34.0101	34.8013
Kmer compl.	81.3738	80.5855
BUSCO sing.	86.1%	86.1%
BUSCO dupl.	0.8%	0.8%
BUSCO frag.	2.7%	2.7%
BUSCO miss.	10.4%	10.4%

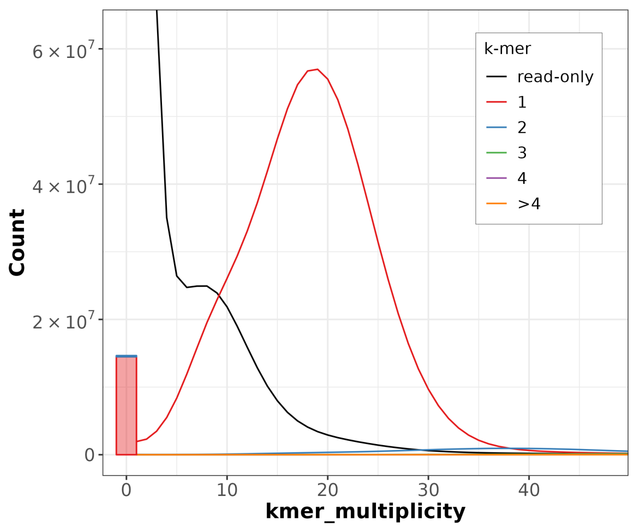
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: mollusca_odb10 (genomes:7, BUSCOs:5295)

HiC contact map of curated assembly

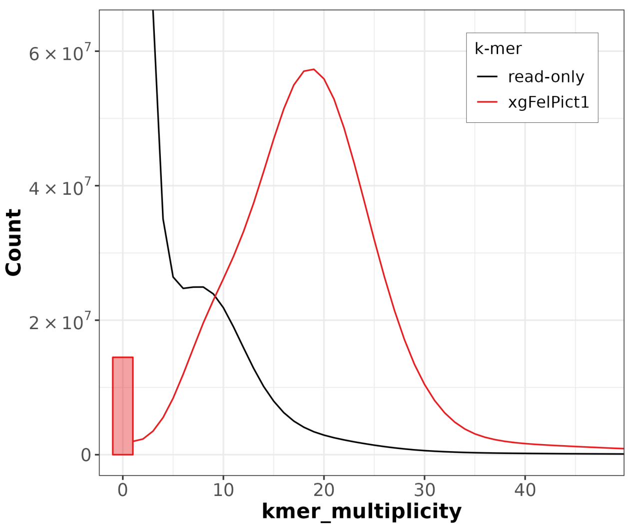


collapsed [LINK]

K-mer spectra of curated assembly

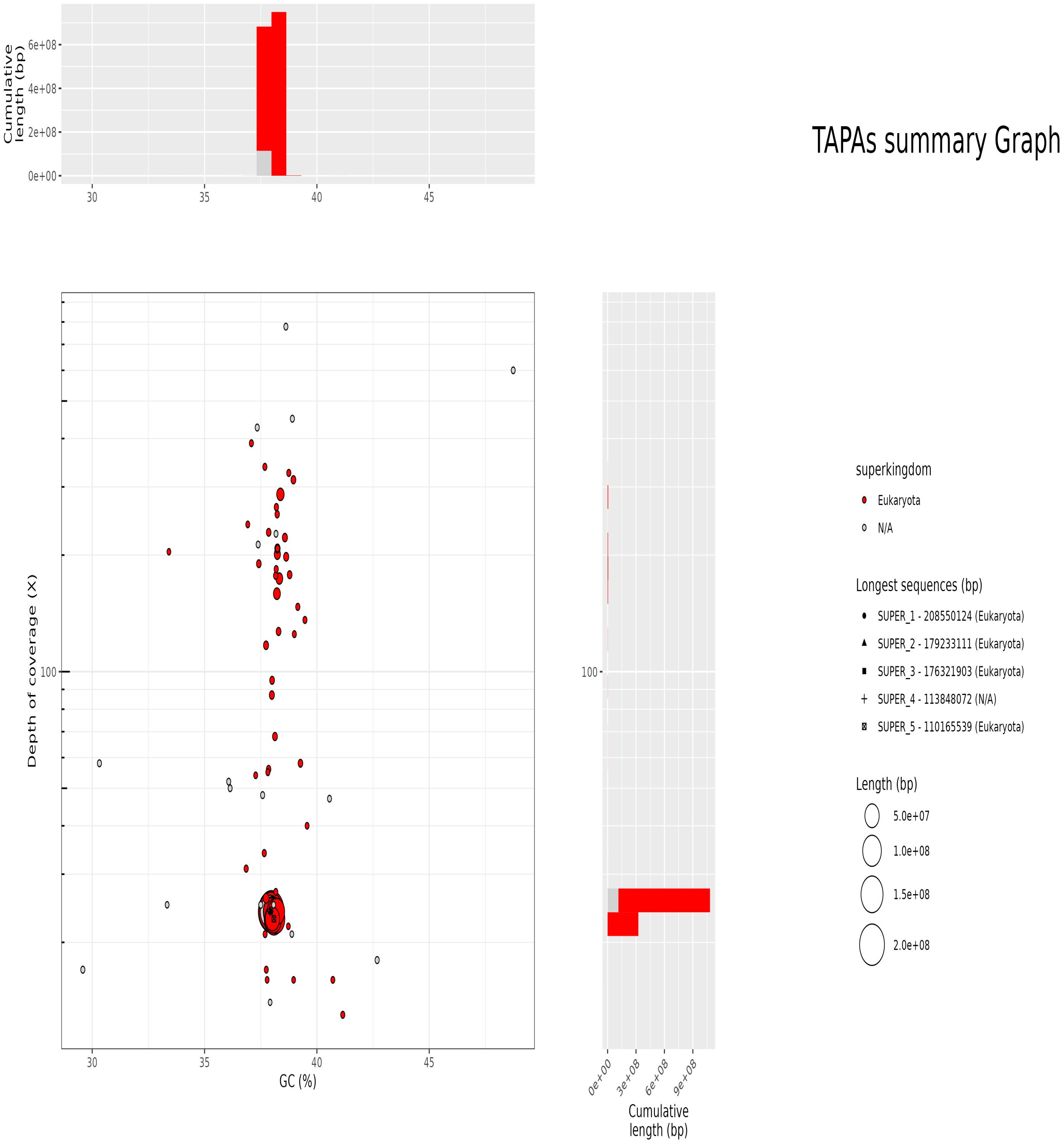


Distribution of k-mer counts per copy numbers found in asm



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	28	29

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: Sophie Mangenot
Affiliation: Genoscope

Date and time: 2024-12-18 00:17:51 CET