ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	223888
ToLID	**sLeuNae2**
Species	Leucoraja naevus
Class	Chondrichthyes
Order	Rajiformes

Genome Traits	Expected	Observed
Haploid size (bp)	2,304,322,291	2,368,913,580
Haploid Number	29 (source: ancestor)	47
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q40

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

. Observed Haploid Number is different from Expected

Curator notes

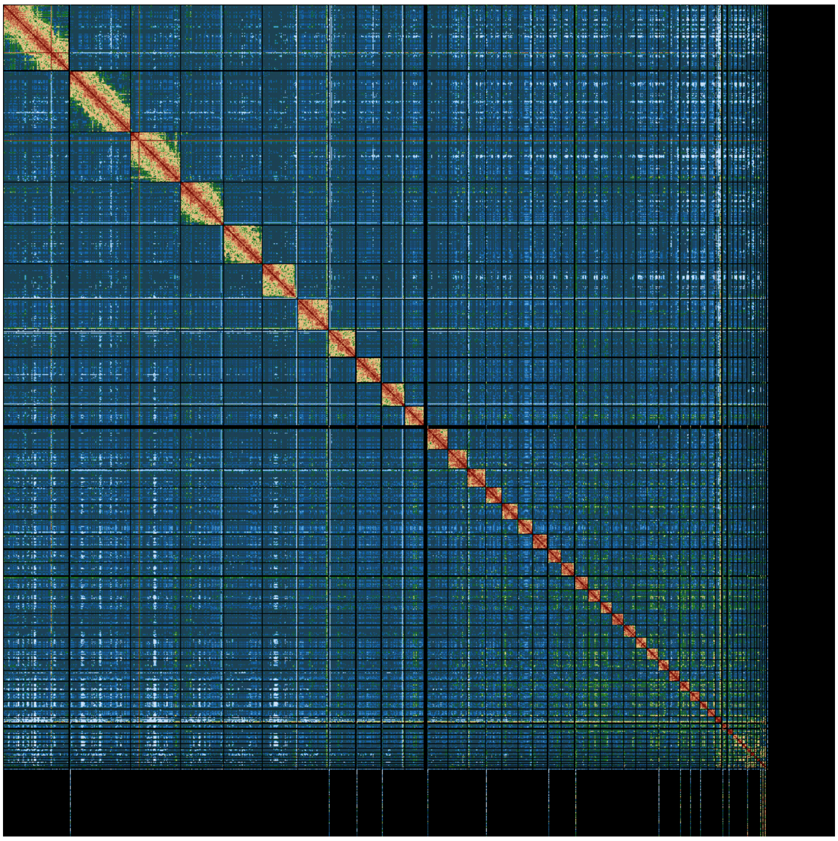
. Interventions/Gb: 346
. Contamination notes: ""
. Other observations: "The assembly of Leucoraja naevus (sLeuNae2) is based on 28X PacBio data and Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included the following steps: initial PacBio assembly 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, 47 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 5.5 Mb (with the largest being 558Kb). Additionally, 2.348 regions totaling 162 Mb (with the largest being 1 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, 44 haplotypic regions were removed,totaling 3.5 Mb (with the largest being 2.5 Mb. 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	2,399,992,255	2,368,913,580
GC %	44.13	44.16
Gaps/Gbp	872.09	917.3
Total gap bp	209,300	228,200
Scaffolds	1,727	1,503
Scaffold N50	55,362,386	60,279,479
Scaffold L50	12	11
Scaffold L90	95	43
Contigs	3,820	3,676
Contig N50	2,355,946	2,392,000
Contig L50	273	267
Contig L90	1,404	1,369
QV	40.8774	40.8751
Kmer compl.	91.684	91.5261
BUSCO sing.	90.0%	91.5%
BUSCO dupl.	3.7%	1.9%
BUSCO frag.	2.6%	2.7%
BUSCO miss.	3.8%	3.9%

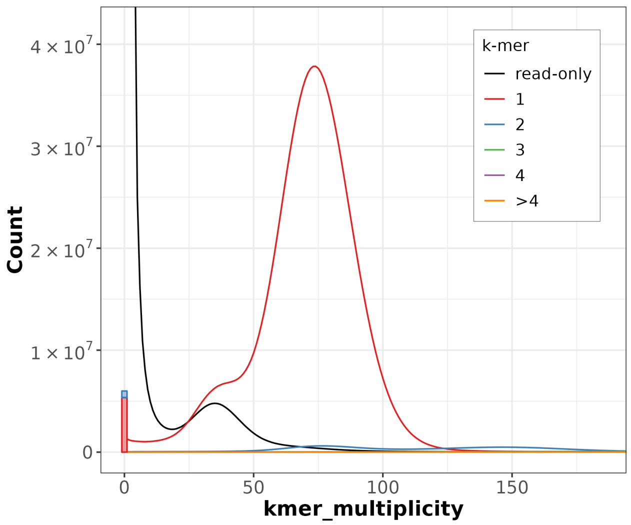
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: vertebrata_odb12 (genomes:119, BUSCOs:3390)

HiC contact map of curated assembly

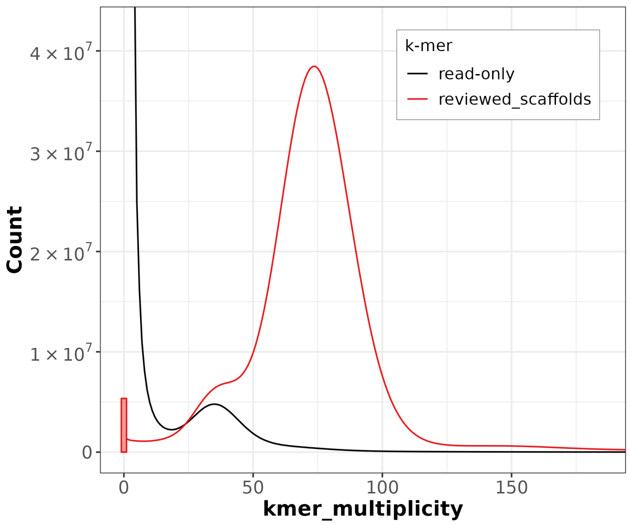


collapsed [LINK]

K-mer spectra of curated assembly

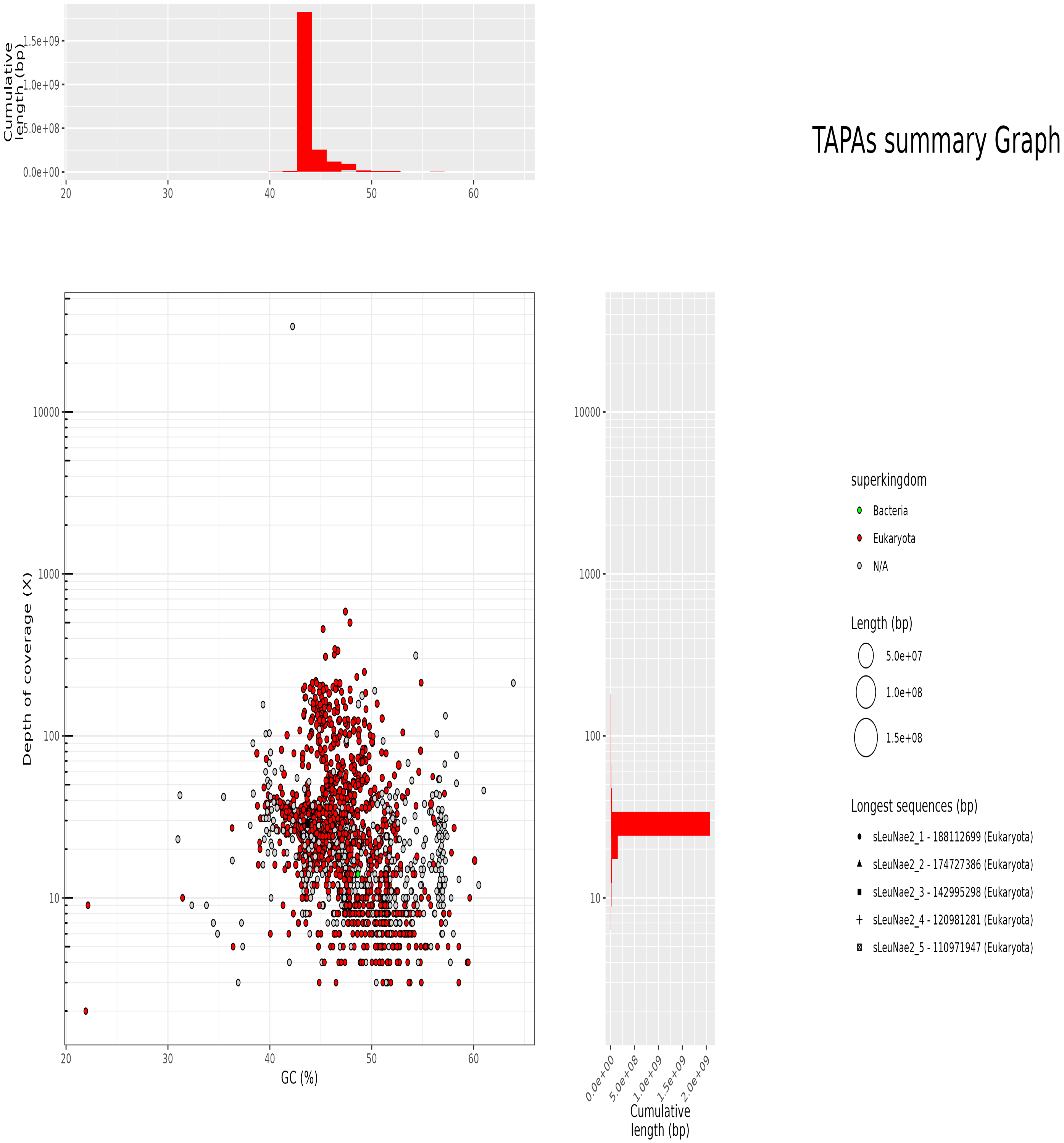


Distribution of k-mer counts per copy numbers found in asm



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	226

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: 2025-07-14 13:00:52 CEST