

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	869980
ToLID	xgDonBany1
Species	Nemesignis banyulensis
Class	Gastropoda
Order	Nudibranchia

Genome Traits	Expected	Observed
Haploid size (bp)	725,151,085	651,301,231
Haploid Number	11 (source: ancestor)	15
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.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
- . More than 1000 gaps/Gbp for collapsed

Curator notes

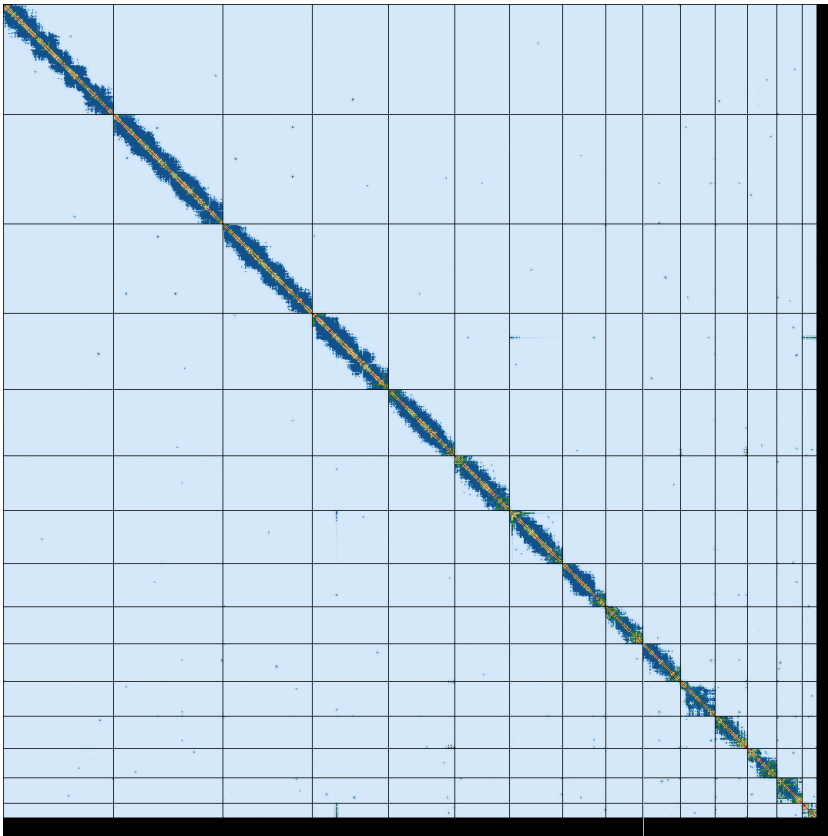
- . Interventions/Gb: 264
- . Contamination notes: ""
- . Other observations: "The assembly of Nemesignis banyulensis (xgDonBany1) is based on 46X ONT data and 451 X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial ONT assembly generation with Nextdenovo , polishing with Medaka, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 11 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.4 Mb (with the largest being 226 kb). Additionally, 449 regions totaling 42 Mb (with the largest being 405 kb) 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	652,747,719	651,301,231
GC %	38.93	38.93
Gaps/Gbp	2,112.61	2,273.91
Total gap bp	137,900	160,600
Scaffolds	721	230
Scaffold N50	51,112,993	51,533,521
Scaffold L50	5	5
Scaffold L90	14	13
Contigs	2,100	1,711
Contig N50	1,361,578	1,370,689
Contig L50	99	98
Contig L90	942	933
QV	34.7975	34.876
Kmer compl.	69.9996	68.1085
BUSCO sing.	79.3%	79.3%
BUSCO dupl.	0.8%	0.7%
BUSCO frag.	3.0%	2.9%
BUSCO miss.	16.9%	17.1%

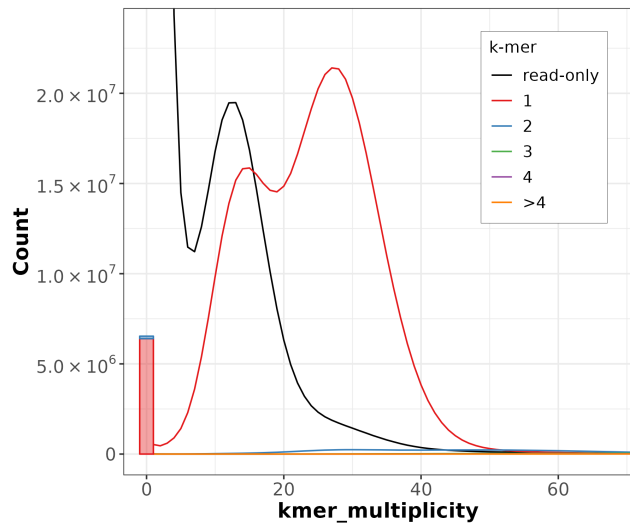
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: mollusca_odb10 (genomes:7, BUSCOs:5295)

HiC contact map of curated assembly

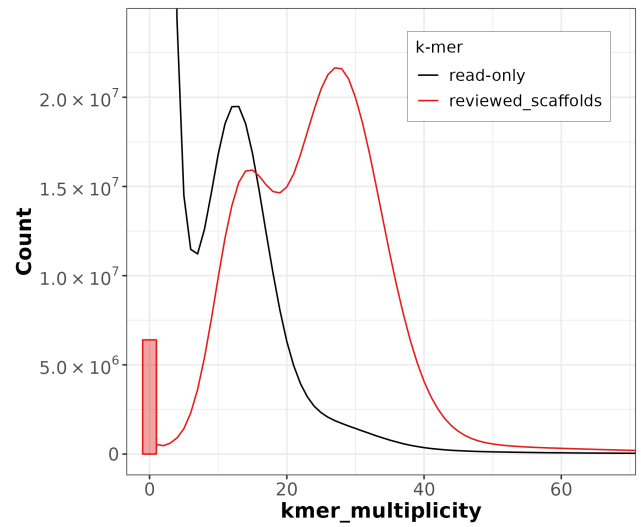


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

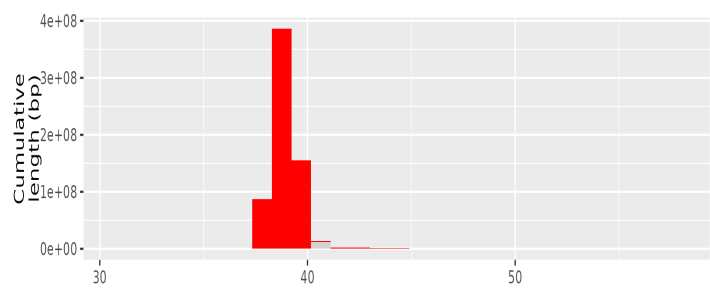


Distribution of k-mer counts per copy numbers found in asm

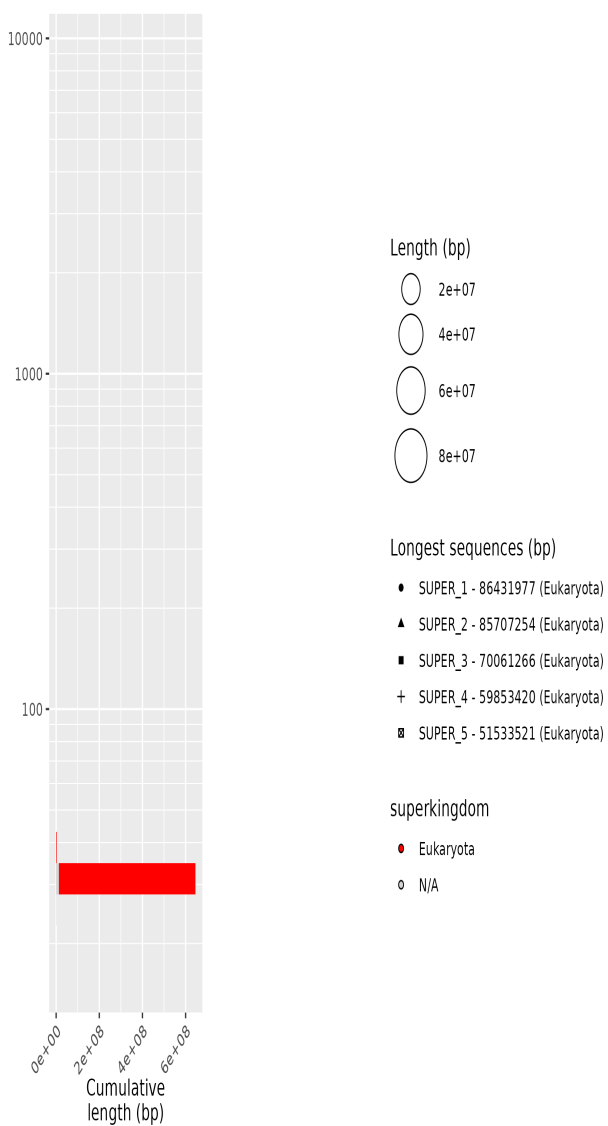
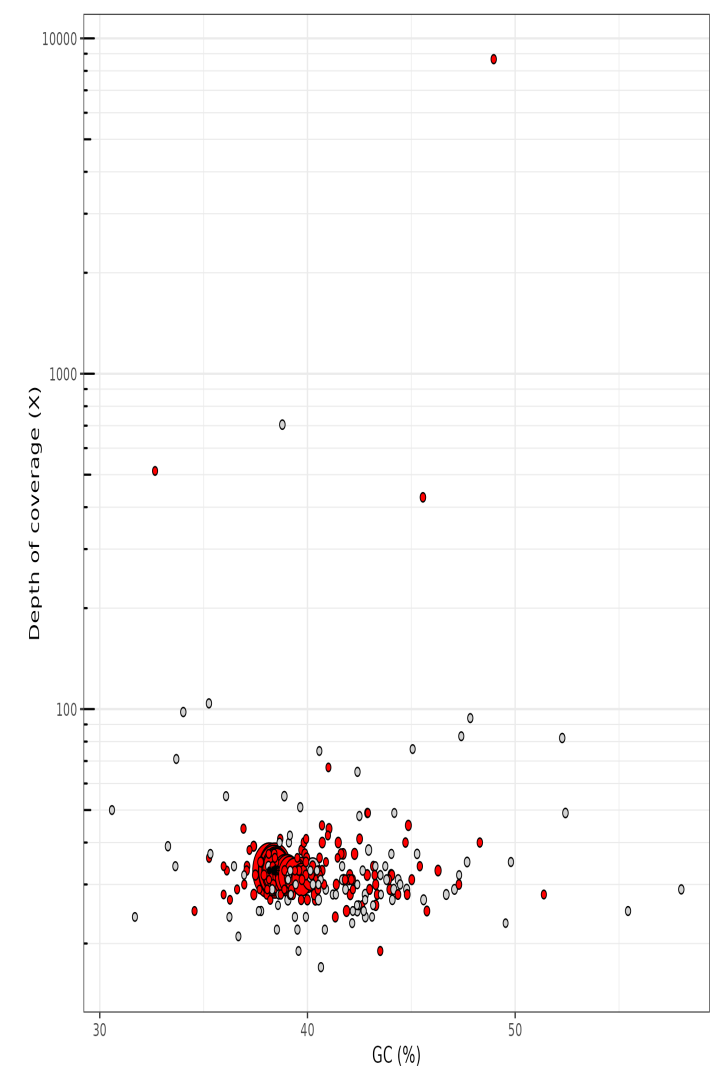


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



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	PACBIO Hifi	Arima
Coverage	46	345

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 Layac

Affiliation: Genoscope

Date and time: 2025-04-18 01:12:35 CEST