

ERGA Assembly Report

v24.09.10

Tags: ERGA-BGE

TxID	2053936
ToLID	xcStiArct1
Species	Stigmatoteuthis arcturi
Class	Cephalopoda
Order	Oegopsida

Genome Traits	Expected	Observed
Haploid size (bp)	3,669,337,499	3,249,387,216
Haploid Number	6 (source: ancestor)	46
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q43

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

- . Observed Haploid Number is different from Expected
- . Kmer completeness value is less than 90 for collapsed

Curator notes

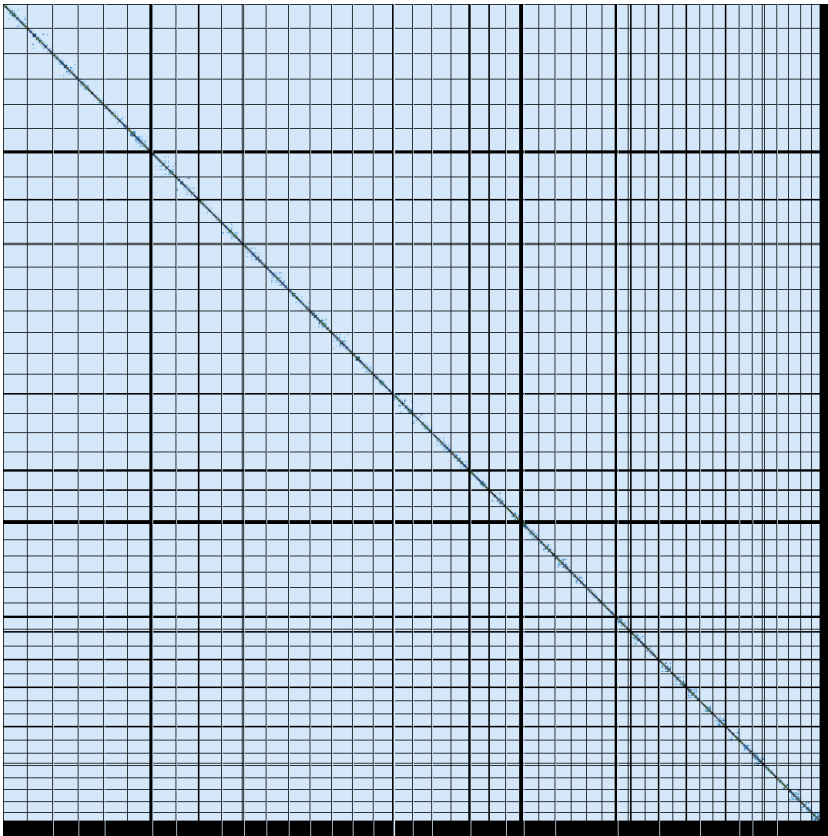
- . Interventions/Gb: 9
- . Contamination notes: "No clear contaminants were found. However, we observed that BlobTools produces best hits to species other than molluscs in a small percentage of the assembly. Upon closer scrutiny of some unlocs or unplaced scaffolds matching other taxa, we found that many of these sequences actually either do match mollusca sequence but with lower scores or match several different unrelated taxa. We suspect that the main issue is that the mollusca databases are quite incomplete, and most data represents a limited number of organisms that could be evolutionarily distant from this species"
- . Other observations: "The curation process was complicated by the fairly poor quality of the Hi-C data we received, resulting in a sparser pretext map (approximately 3.5 million valid pairs, including cis >= 1kb and trans. per Gbase of sequence) compared to most projects with better Hi-C libraries. However, the pre-scaffolded NextDenovo assembly was already quite good, which helped us generate a draft assembly that meets most of the stringent EBP requirements. In this final version of the EAR report we addressed all of the reviewer concerns and produced a new pretext map and assembly based on that. The blobtools plot was not modified as that was deemed not to be necessary."

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	3,249,385,616	3,249,387,216
GC %	34.92	34.92
Gaps/Gbp	234.81	237.28
Total gap bp	152,600	154,200
Scaffolds	503	497
Scaffold N50	73,833,701	74,550,748
Scaffold L50	19	19
Scaffold L90	42	42
Contigs	1,266	1,268
Contig N50	10,980,000	10,980,000
Contig L50	88	88
Contig L90	343	344
QV	43.3597	43.3596
Kmer compl.	79.2506	79.2506
BUSCO sing.	93.8%	93.8%
BUSCO dupl.	0.9%	0.9%
BUSCO frag.	1.9%	1.9%
BUSCO miss.	3.4%	3.4%

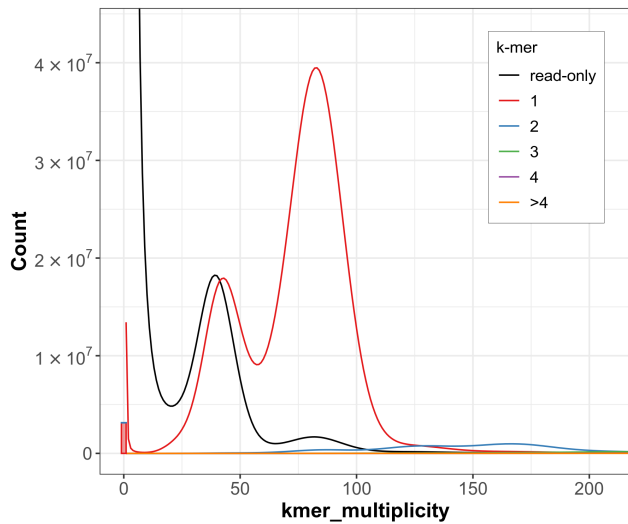
BUSCO 5.4.0 Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

HiC contact map of curated assembly

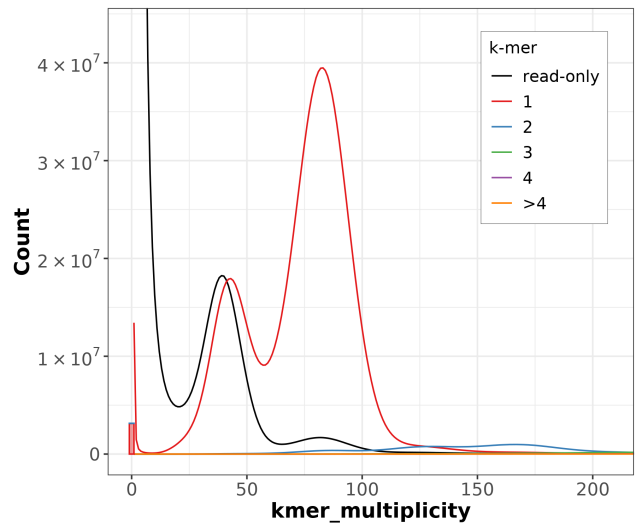


collapsed [\[LINK\]](#)

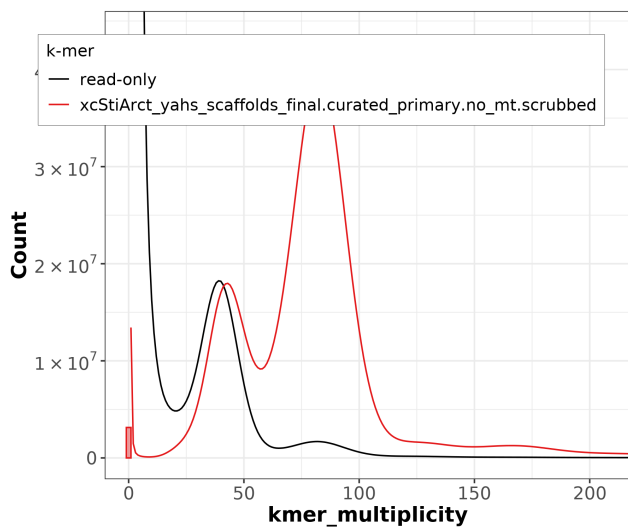
K-mer spectra of curated assembly



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

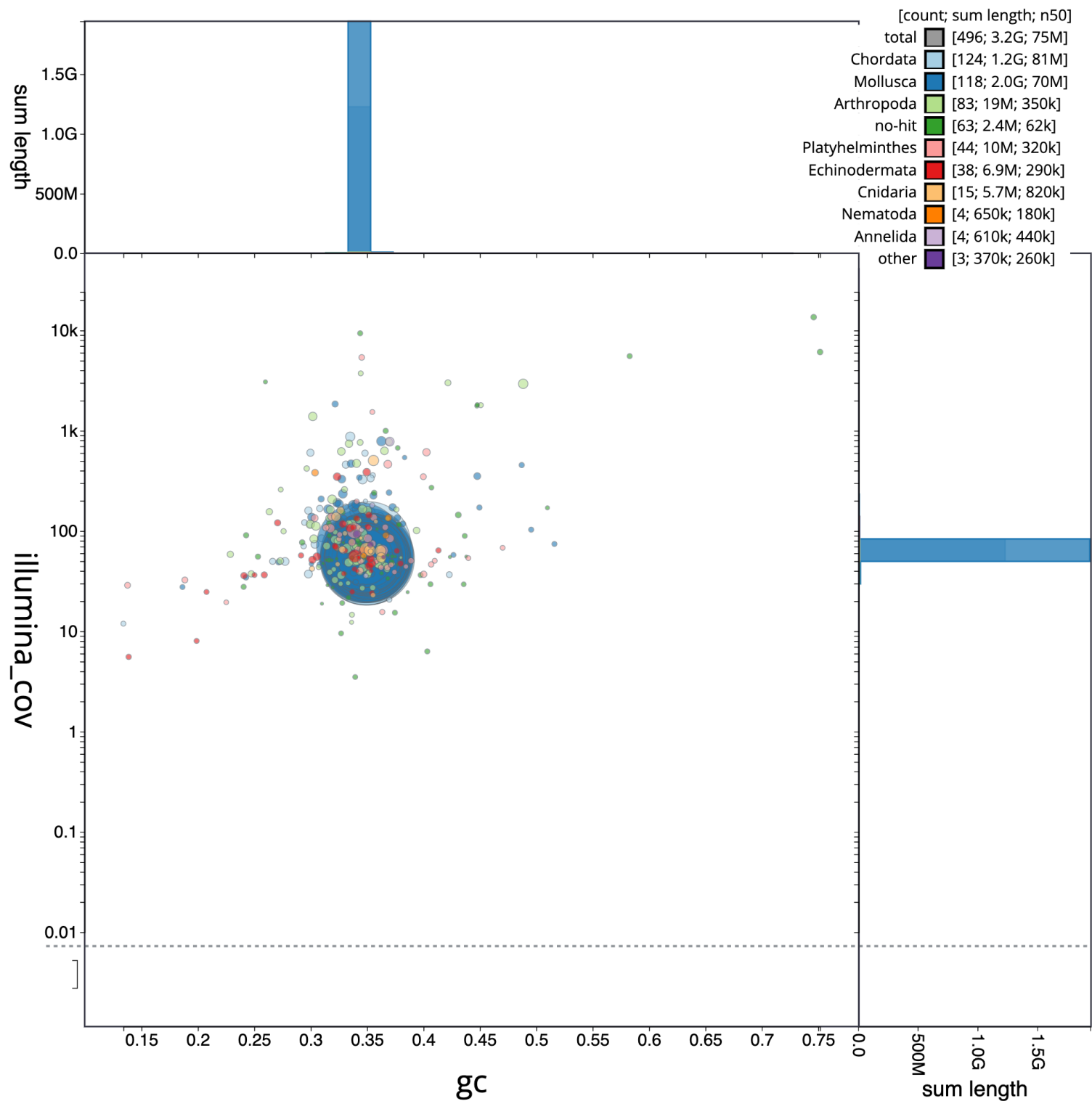


Distribution of k-mer counts per copy numbers found in asm



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	ONT	Illumina	OmniC
Coverage	56x	63x	60x

Assembly pipeline

- **Trim_galore**
 - |_ *ver*: 0.6.7
 - |_ *key param*: --gzip
 - |_ *key param*: =q 20
 - |_ *key param*: --paired
 - |_ *key param*: retain_unpaired
- **Filtlong**
 - |_ *ver*: 0.2.1
 - |_ *key param*: NA
- **nextdenovo**
 - |_ *ver*: 2.5.0
 - |_ *key param*: NA
- **hypo**
 - |_ *ver*: 1.0.3
 - |_ *key param*: NA
- **purge_dups**
 - |_ *ver*: 1.2.6
 - |_ *key param*: NA
- **YaHS**
 - |_ *ver*: 1.2a
 - |_ *key param*: NA

Curation pipeline

- **PretextView**
 - |_ *ver*: 0.2.5
 - |_ *key param*: NA
- **GRIT_Rapid**
 - |_ *ver*: 2.0
 - |_ *key param*: NA

Submitter: Francisco Camara

Affiliation: CNAG Barcelona

Date and time: 2024-10-16 10:43:32 CEST