

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	204497
ToLID	<b>ilChaSubt4</b>
Species	Charissa subtaurica
Class	Insecta
Order	Lepidoptera

Genome Traits	Expected	Observed
Haploid size (bp)	651,036,455	675,591,976
Haploid Number	30 (source: ancestor)	15
Ploidy	2 (source: ancestor)	2
Sample Sex	ZW	ZW

## EBP metrics summary and curation notes

Obtained EBP quality metric for hap1: 7.7.Q66

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 hap1

### Curator notes

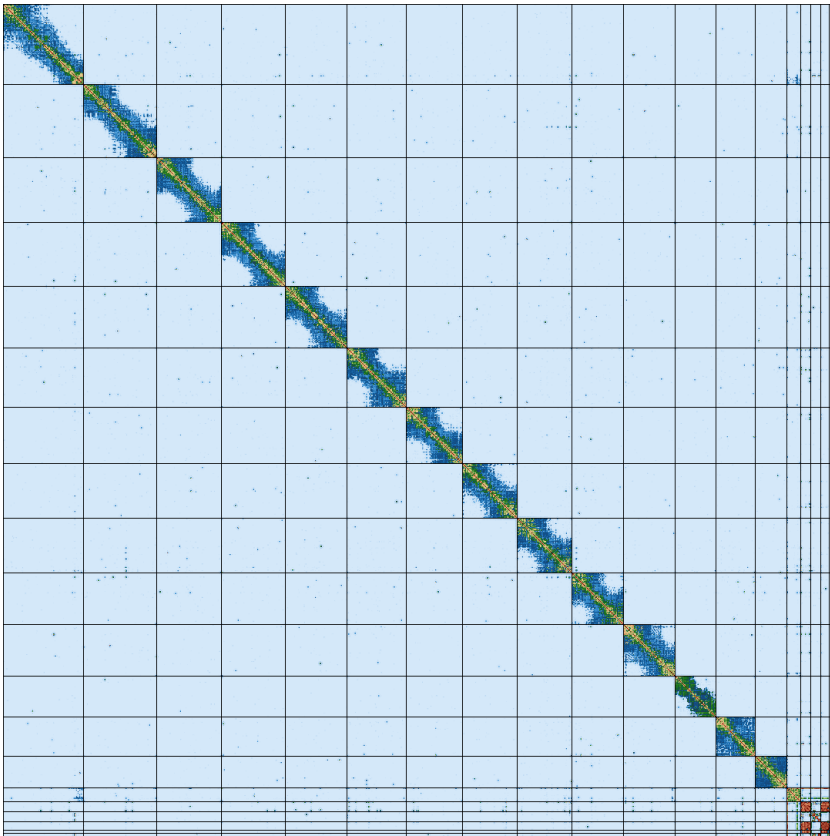
. Interventions/Gb: 7  
. Contamination notes: "We could not find any contaminants on this this assembly"  
. Other observations: "The assembly of each haplotype for this species was nearly telomere-to-telomere (T-to-T). Hap2 was slightly more contiguous than hap1-hap1 contained two gaps, whereas hap2 was gap-free. However, hap1 included the SUPER scaffold identified as Z, as well as most of the W chromosome. To leverage this, we created a pseudo-hap1 assembly by combining the entire hap2 assembly with the Z and W scaffolds from hap1. We then remapped the Hi-C data against this new combined assembly, which we designated as hap1. This resulted in a T-to-T assembly, with the exception of SUPER\_W, which retained two gaps. We attempted to reorganize the SUPER\_W sequences, but due to the high repetitiveness of the region, our confidence in the scaffold's structural accuracy remains low. We assigned the longest contig corresponding to W as the SUPER, and the remaining fragments were labeled as Unlocs. A save\_state of the Pretext map has been placed in the shared folder for your review"

# Quality metrics table

Metrics	Pre-curation hap1	Curated hap1
Total bp	633,077,418	675,591,976
GC %	35.82	35.81
Gaps/Gbp	0	4.44
Total gap bp	0	600
Scaffolds	19	20
Scaffold N50	47,879,875	46,187,926
Scaffold L50	6	7
Scaffold L90	12	13
Contigs	19	23
Contig N50	47,879,875	46,187,926
Contig L50	6	7
Contig L90	12	13
QV	67.6823	66.3434
Kmer compl.	77.7251	81.7073
BUSCO sing.	95.8%	98.5%
BUSCO dupl.	0.4%	0.4%
BUSCO frag.	0.8%	0.8%
BUSCO miss.	3.0%	0.3%

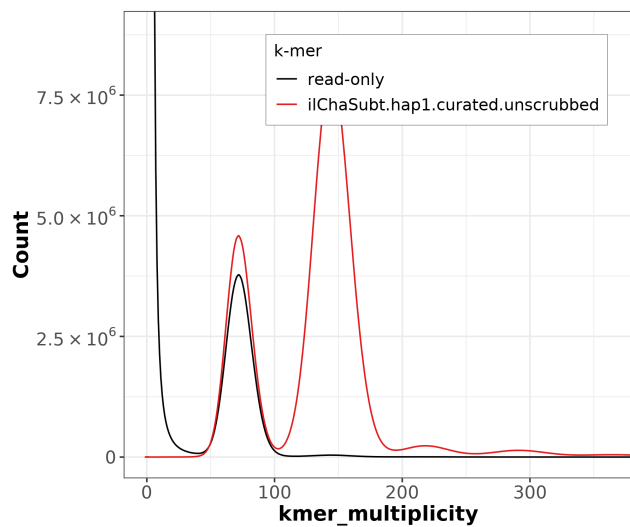
BUSCO: 5.5.0 (euk\_genome\_met, metaeuk) / Lineage: arthropoda\_odb10 (genomes:90, BUSCOs:1013)

# HiC contact map of curated assembly

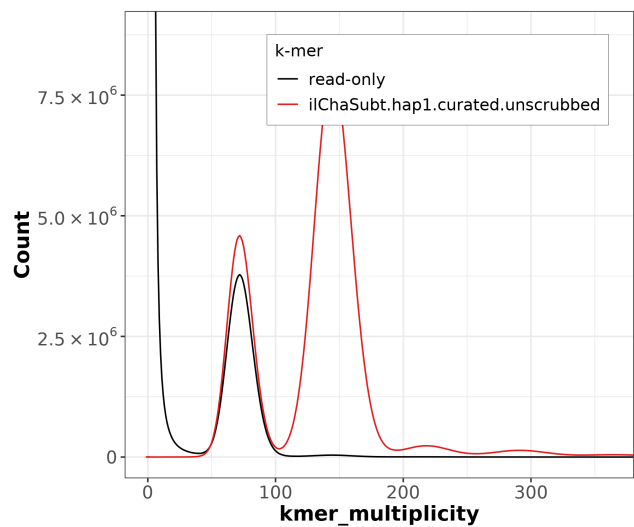


hap1 [\[LINK\]](#)

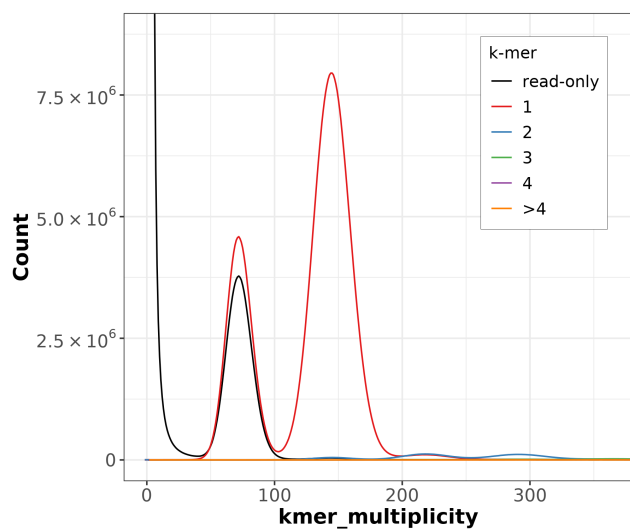
# K-mer spectra of curated assembly



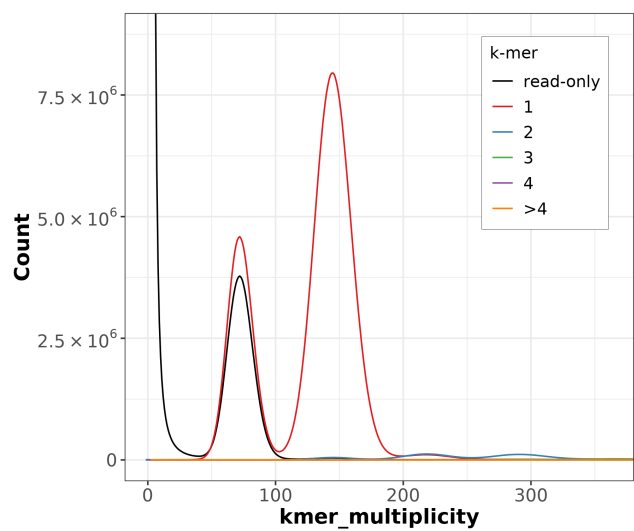
Distribution of k-mer counts coloured by their presence in reads/assemblies



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

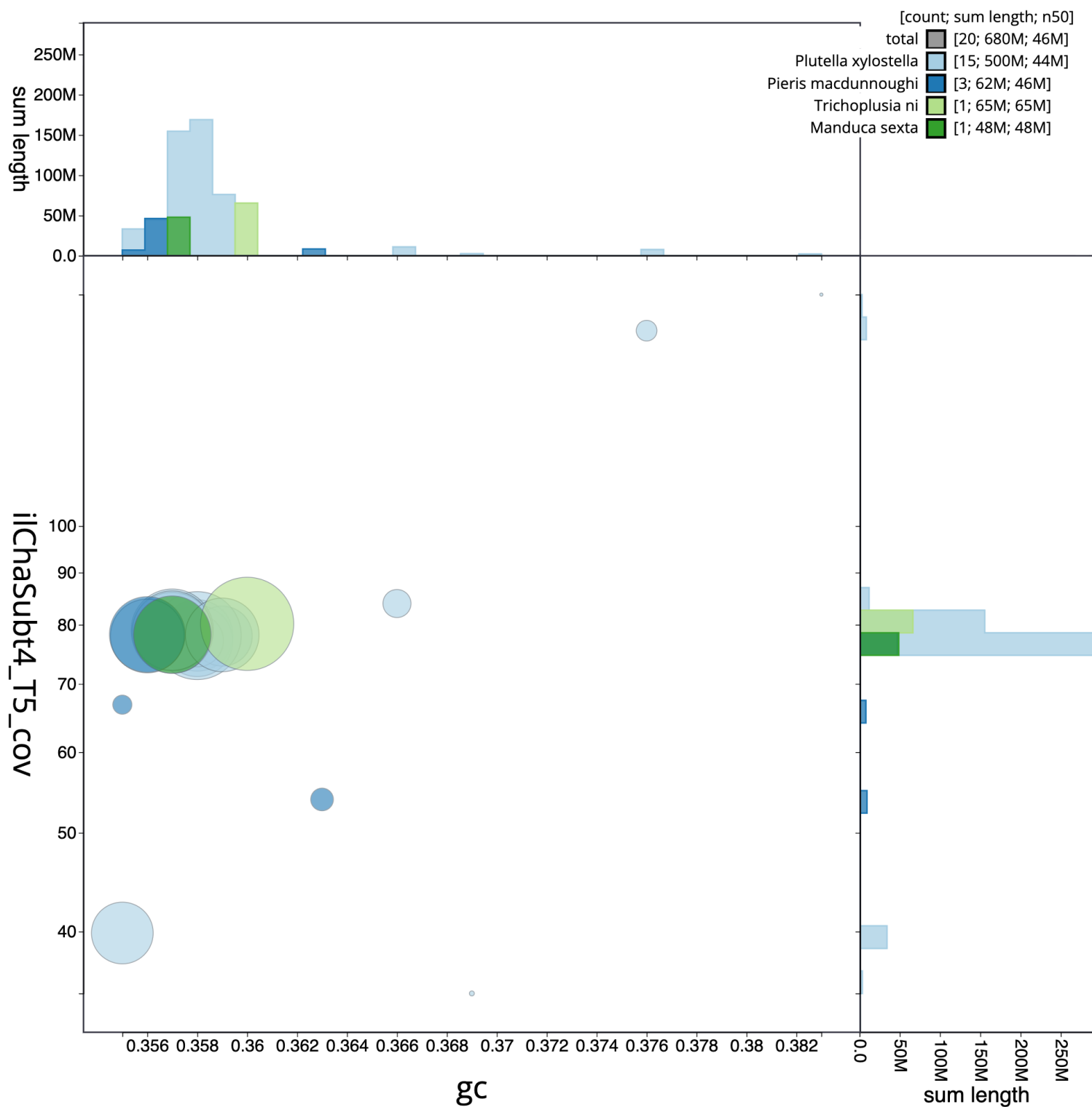


Distribution of k-mer counts per copy numbers found in asm



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

# Post-curation contamination screening



**hap1.** 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 WGS	Omnich
Coverage	74x	91x	98x

# Assembly pipeline

- **hifiasm**
  - |\_ *ver*: 0.24.0-r702
  - |\_ *key param*: NA
- **purge\_dups**
  - |\_ *ver*: 1.2.6
  - |\_ *key param*: NA
- **YaHS**
  - |\_ *ver*: 1.2a
  - |\_ *key param*: NA
- **CLAWS**
  - |\_ *ver*: 2.3
  - |\_ *key param*: NA

# Curation pipeline

- **PretextViewAI**
  - |\_ *ver*: 1.0.4
  - |\_ *key param*: NA
- **GRIT\_Rapid**
  - |\_ *ver*: 2.0
  - |\_ *key param*: NA

Submitter: Francisco Camara

Affiliation: CNAG

Date and time: 2025-10-02 16:00:59 CEST