

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	3158904
ToLID	iqOecDulc4
Species	Oecanthus dulcisonans
Class	Insecta
Order	Orthoptera

Genome Traits	Expected	Observed
Haploid size (bp)	960,397,348	1,072,582,059
Haploid Number	10 (source: ancestor)	9
Ploidy	2 (source: ancestor)	2
Sample Sex	X0	X0

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 8.8.Q68

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 pri
- . Assembly length loss > 3% for pri

Curator notes

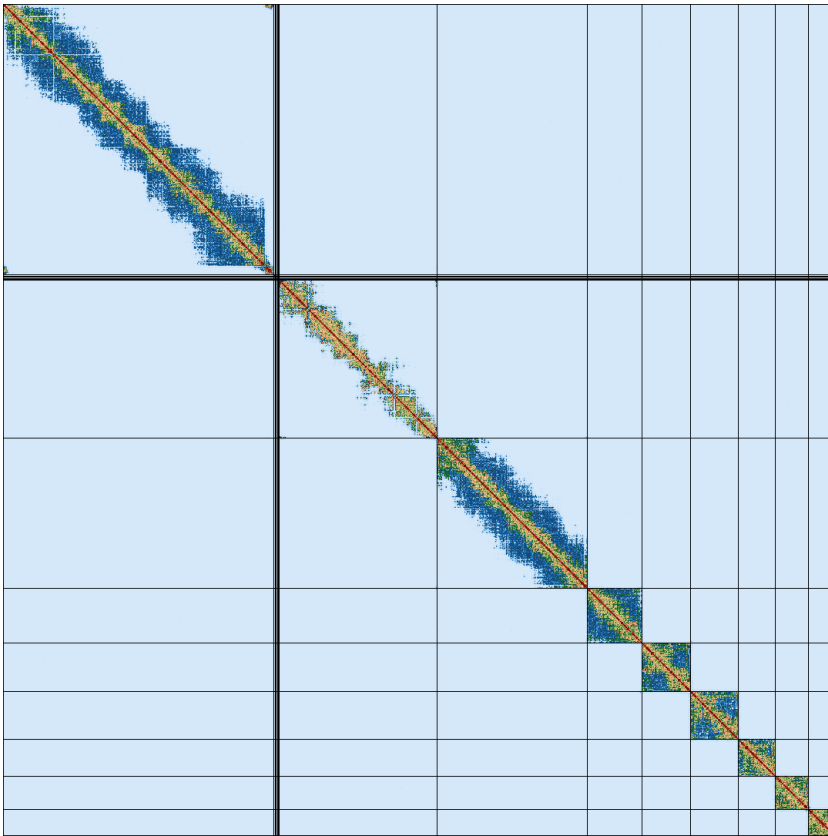
- . Interventions/Gb: 0
- . Contamination notes: "The BTK pipeline detected 530 contaminants, 7 no-hit (coverage and GC) and 523 belonging to the following Phylum: Bacillota (336), Bacterioidota (177), Actinomycetota (3), Chordata (2), Pseudomonadota (1), Streptophyta (1), Uroviricota (1). These contaminants have been removed from the assembly. In addition, 5 scaffolds shorter than 68K contain partial or complete mitogenome insertions. They will be removed from the reviewed assembly before ENA upload (I already checked that BUSCO and QV will stay the same)."
- . Other observations: "Hifiasm produced a primary assembly with 9 superscaffolds holding telomeres at both sides. HiC belongs to a different individual (iqOecDulc5) but also a male. One of the 9 super-scaffolds was identified as the X chromosome based on the coverage. Both, the contact map and the telomere locations, strongly support a haploid number of 9. The curation was minimal, we just placed scaffolds larger than 119K into SUPER_1 using the multimaps pretext (mq=0). They seem to be centromeric repeats that can fit nearby a gap but they are impossible to order and orient with confidence. Regarding to the mitochondrial genome, this was assembled with FOAM into a single circular contig of 17,493 bp."

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	1,129,341,404	1,072,582,059
GC %	41.47	41.31
Gaps/Gbp	1.77	1.86
Total gap bp	400	400
Scaffolds	569	39
Scaffold N50	193,308,966	204,502,310
Scaffold L50	3	2
Scaffold L90	8	7
Contigs	571	41
Contig N50	193,308,966	193,308,966
Contig L50	3	3
Contig L90	9	8
QV	56.9282	68.2709
Kmer compl.	84.8353	83.6885
BUSCO sing.	96.0%	96.0%
BUSCO dupl.	2.6%	2.6%
BUSCO frag.	0.7%	0.7%
BUSCO miss.	0.7%	0.7%

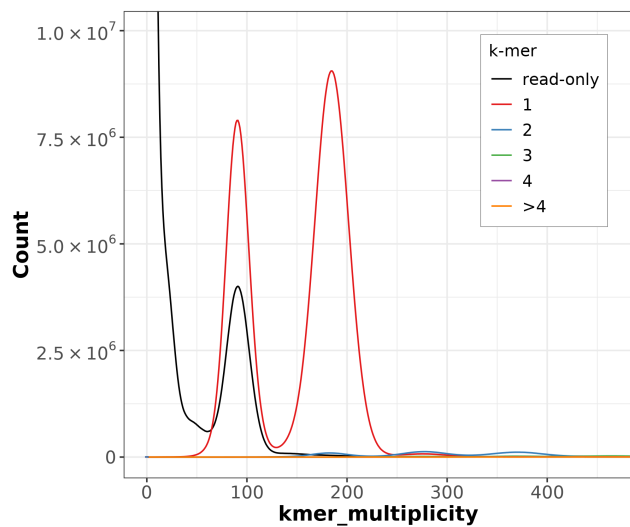
BUSCO: 5.5.0 (euk_genome_met, metaeuk) / Lineage: insecta_odb10 (genomes:75, BUSCOs:1367)

HiC contact map of curated assembly

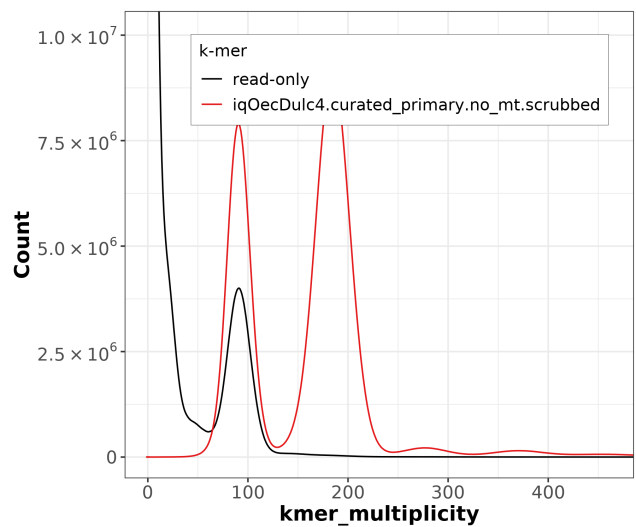


pri [\[LINK\]](#)

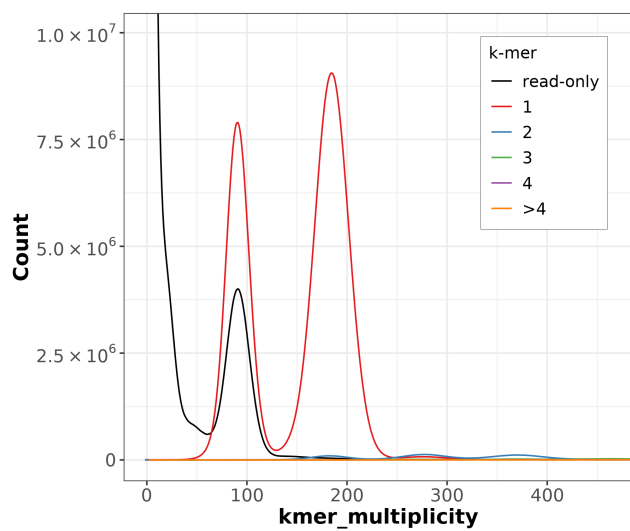
K-mer spectra of curated assembly



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

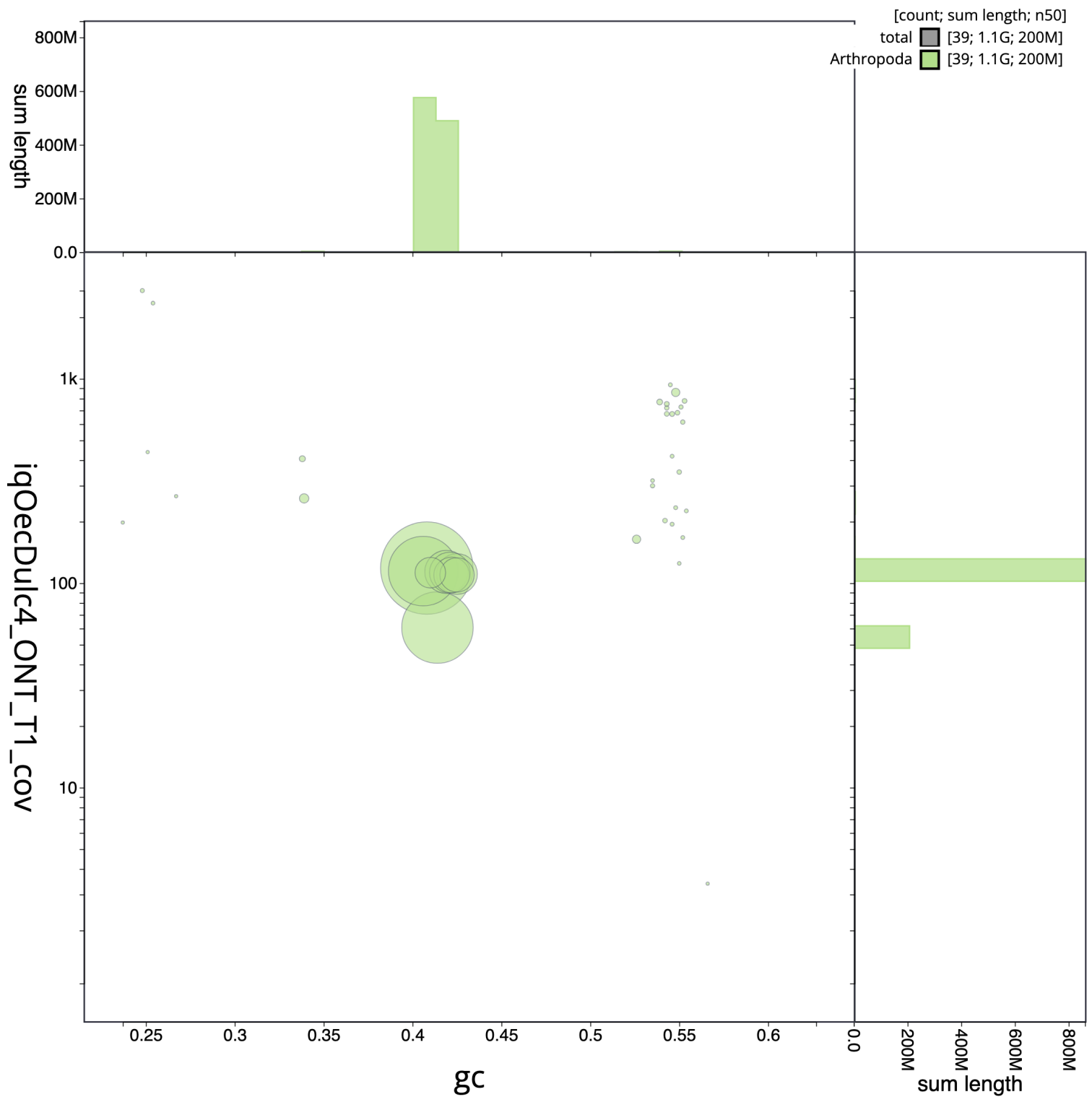


Distribution of k-mer counts coloured by their presence in reads/assemblies



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

Post-curation contamination screening



pri. 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	107	95	78

Assembly pipeline

- **CLAWS pipeline**
 - |_ *ver*: 2.3.0
 - |_ *key param*: NA
- **Trim_galore**
 - |_ *ver*: 0.6.7
 - |_ *key param*: NA
- **Filtlong**
 - |_ *ver*: 0.2.1
 - |_ *key param*: NA
- **hifiasm**
 - |_ *ver*: 0.24.0
 - |_ *key param*: NA
- **YaHS**
 - |_ *ver*: 1.2a
 - |_ *key param*: NA
- **FOAM pipeline**
 - |_ *ver*: 0.5
 - |_ *key param*: NA
- **mitos**
 - |_ *ver*: 2.1.3
 - |_ *key param*: NA

Curation pipeline

- **Blobtoolkit Nextflow pipeline (latest)**
 - |_ *ver*: 0.6
 - |_ *key param*: NA
- **PretextViewAI**
 - |_ *ver*: 1.0.3
 - |_ *key param*: NA

Submitter: Fernando Cruz

Affiliation: CNAG

Date and time: 2025-09-23 18:45:27 CEST