

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	3158463
ToLID	iyCerCypr5
Species	Ceratina cypriaca
Class	Insecta
Order	Hymenoptera

Genome Traits	Expected	Observed
Haploid size (bp)	233,648,048	372,795,564
Haploid Number	17 (source: ancestor)	20
Ploidy	1 (source: ancestor)	2
Sample Sex	Female	NA

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 7.7.Q68

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

- . Observed Haploid size (bp) has >20% difference with Expected
- . Observed Haploid Number is different from Expected
- . Observed Ploidy is different from Expected
- . Observed sex is different from Sample sex
- . Assembly length loss > 3% for pri

Curator notes

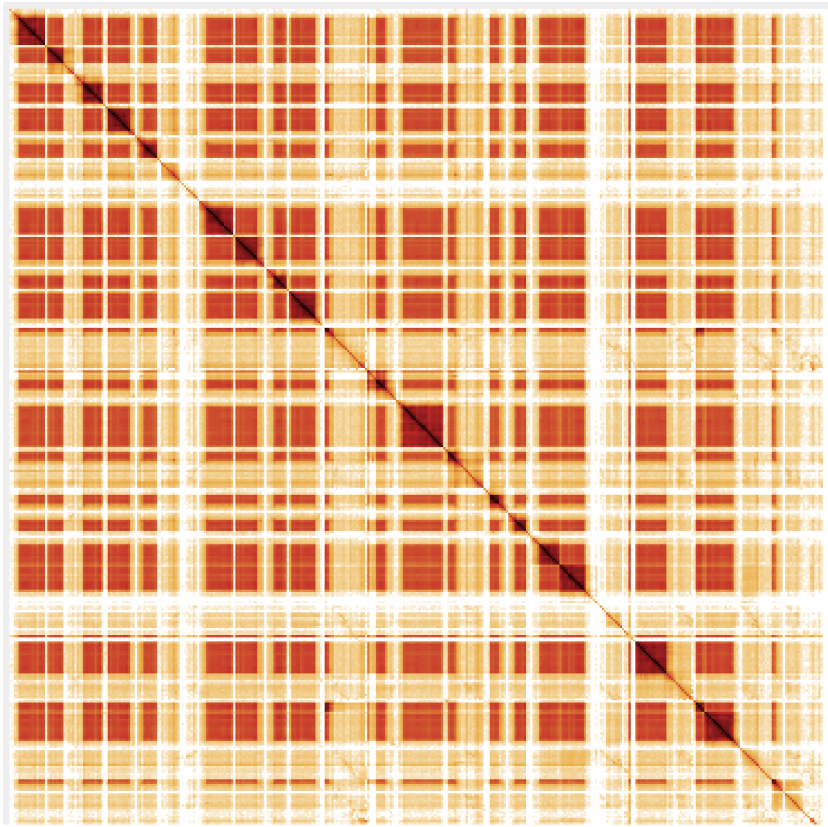
- . Interventions/Gb: 1
- . Contamination notes: "No contaminant sequences detected"
- . Other observations: "YaHS joined a lot of the sequences together, so I worked with the raw contigs from purge-dups. I was only able to join two contigs together and couldn't make out much from the rest. The expected chromosome number based on other species in this genus is 14 or 17."

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	652,012,055	372,795,564
GC %	43.46	42.21
Gaps/Gbp	0	0
Total gap bp	0	0
Scaffolds	301	25
Scaffold N50	14,152,198	19,086,850
Scaffold L50	16	7
Scaffold L90	40	18
Contigs	301	25
Contig N50	14,152,198	19,086,850
Contig L50	16	7
Contig L90	40	18
QV	64.6511	68.5826
Kmer compl.	94.8796	94.2628
BUSCO sing.	98.0%	97.7%
BUSCO dupl.	1.3%	1.1%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	0.7%	1.1%

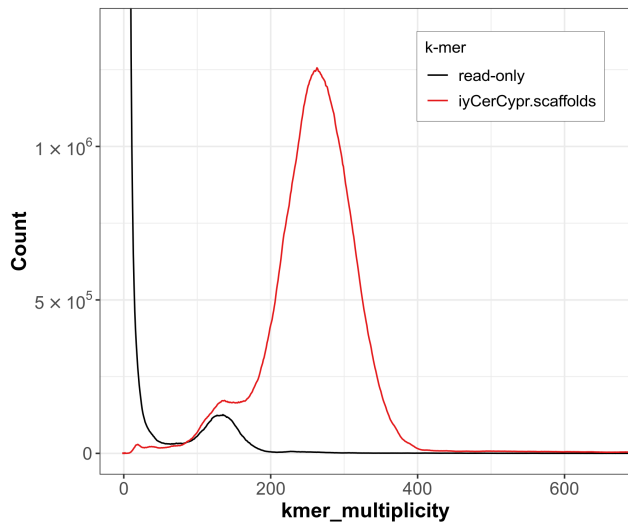
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: hymenoptera_odb12 (genomes:78, BUSCOs:5920)

HiC contact map of curated assembly

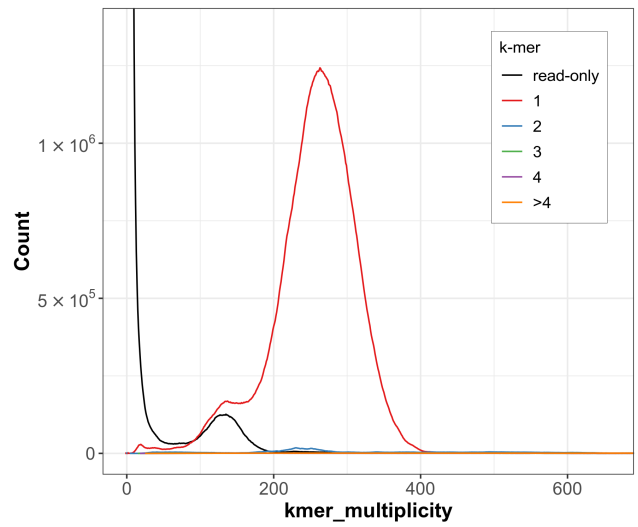


pri [\[LINK\]](#)

K-mer spectra of curated assembly

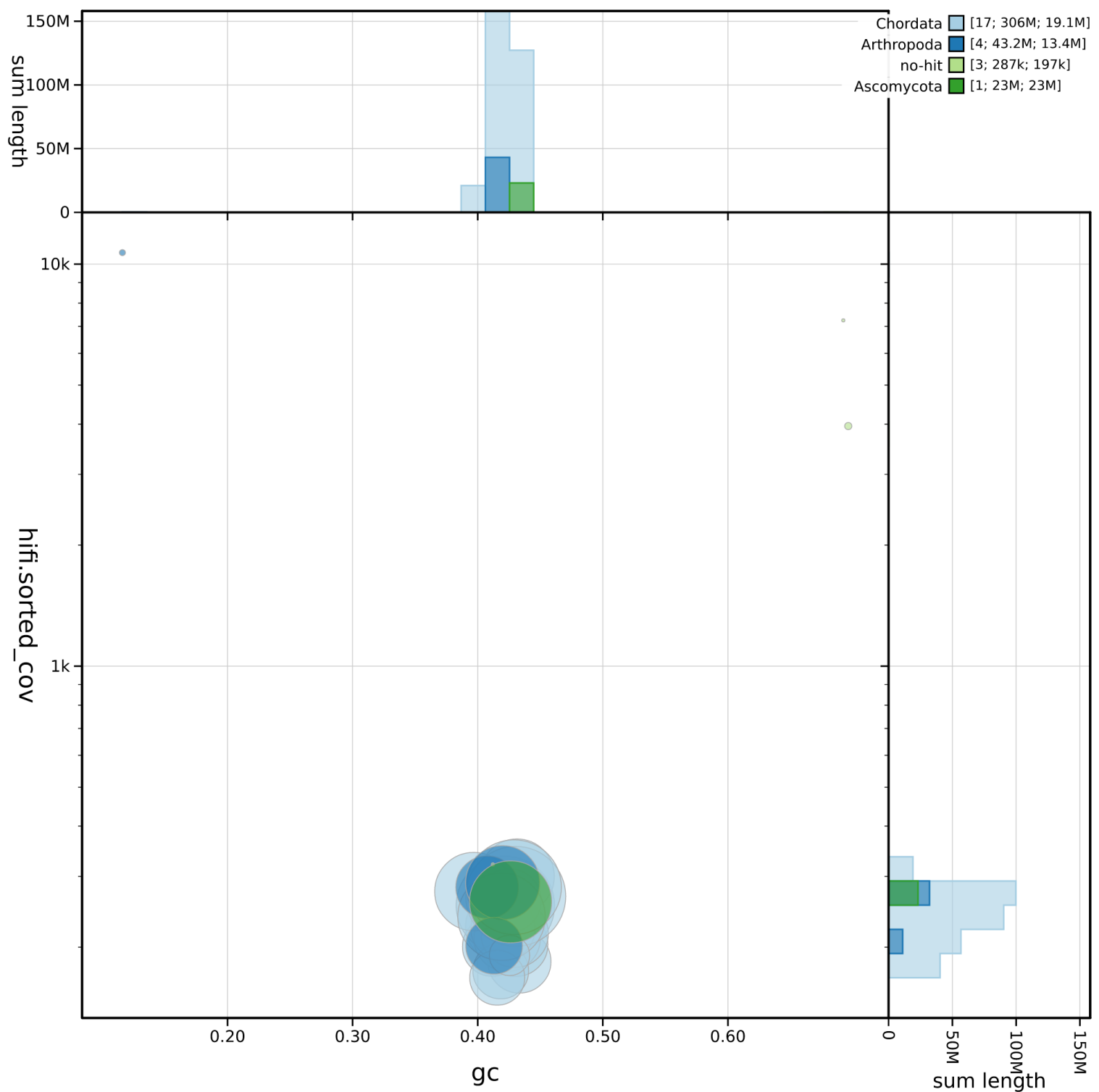


Distribution of k-mer counts coloured by their presence in reads/assemblies



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	HiFi	HiC
Coverage	260	478

Assembly pipeline

- **HiFiasm**
 - |_ *ver*: 0.25.0
 - |_ *key param*: NA
- **purge-dups**
 - |_ *ver*: 1.2.6
 - |_ *key param*: --hom-cov 260
- **fcs-gx**
 - |_ *ver*:
 - |_ *key param*: scratch
 - |_ *key param*: brown
 - |_ *key param*: asm_iyCerCypr
 - |_ *key param*: assembly
 - |_ *key param*: scaffolding
 - |_ *key param*: genome.fasta.stats

Curation pipeline

- **Higlass**
 - |_ *ver*: Higlass
 - |_ *key param*: NA
- **PretextAI**
 - |_ *ver*: 1.0.0
 - |_ *key param*: NA

Submitter: Tom Brown

Affiliation: IZW

Date and time: 2025-10-08 14:18:32 CEST