

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

| | |
|---------|----------------------|
| TxID | 243007 |
| ToLID | iqAioThal3 |
| Species | Aiolopus thalassinus |
| Class | Insecta |
| Order | Orthoptera |

| Genome Traits | Expected | Observed |
|-------------------|----------------------|---------------|
| Haploid size (bp) | 6,979,785,760 | 7,281,159,242 |
| Haploid Number | 11 (source: direct) | 12 |
| Ploidy | 2 (source: ancestor) | 2 |
| Sample Sex | X0 | X0 |

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 7.8.Q72

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

Curator notes

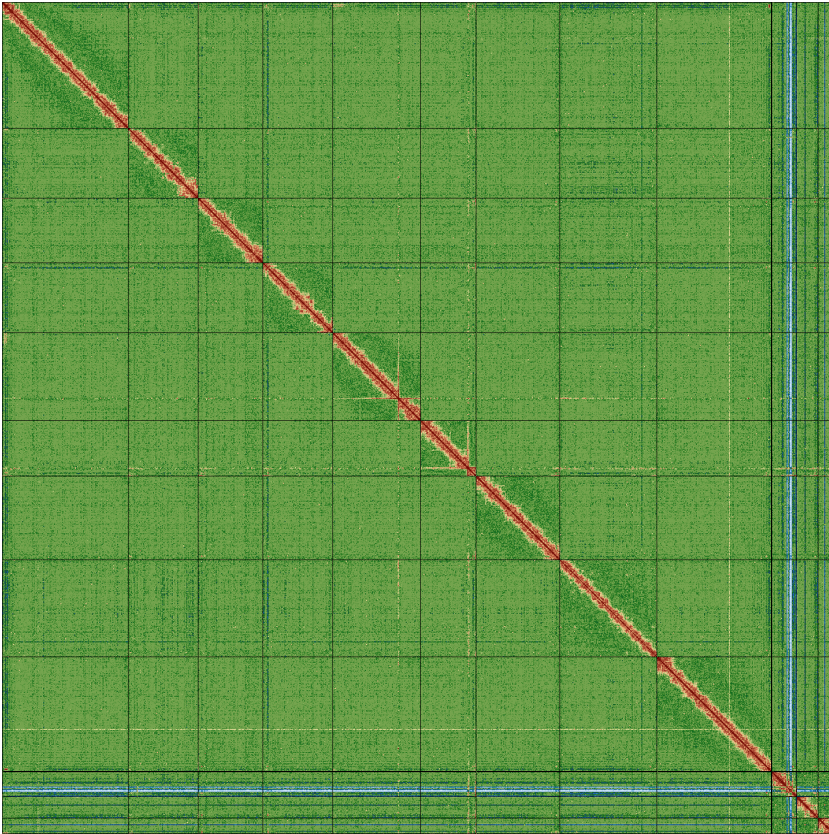
. Interventions/Gb: 39
. Contamination notes: "Contaminations were identified and removed with FCS-GX (removed: 4 contigs, 150Kb). Mitochondrial genome was removed from the assembly."
. Other observations: "PacBio reads were assembled with hifiiasm. FCS-GX detected a few contaminations which were removed. Purge_dups was run to remove haplotypic duplications. But it removed too much sequence and even the Busco score dropped significantly. Therefore the Busco gene set was used to guide the purging. Afterwards Tiara was used to remove archaea|bacteria|organelle hits. Due to time issues and compute cluster down-time the scaffolding was done manually within the curation step. This worked well as the input assembly consisted of just 230 contigs. The curation revealed more contig misjoins than usual (76 contigs cuts were necessary). This is probably caused by the large heterozygosity (1.6%). Several large-scale structural haplotypic variations can be seen in the HiC-map. The X-chromosome was identified by synteny analysis with other Acrididae (e.g. GCF_021461395.2_iqSchAmer2) assemblies, but also the PacBio read coverage showed a 1n coverage pattern."

Quality metrics table

| Metrics | Pre-curation pri | Curated pri |
|--------------|---------------------|----------------|
| Total bp | 7,497,130,708 | 7,281,159,242 |
| GC % | 41.47 | 41.47 |
| Gaps/Gbp | 0 | 28.02 |
| Total gap bp | 0 | 40,800 |
| Scaffolds | 227 | 48 |
| Scaffold N50 | 63,728,264 | 765,756,381 |
| Scaffold L50 | 33 | 4 |
| Scaffold L90 | 115 | 9 |
| Contigs | 227 | 252 |
| Contig N50 | 63,728,264 | 52,464,910 |
| Contig L50 | 33 | 36 |
| Contig L90 | 115 | 127 |
| QV | 72.0441 | 72.0023 |
| Kmer compl. | 77.349 | 76.2696 |
| BUSCO sing. | 93.4% | 96.1% |
| BUSCO dupl. | 4.9% | 2.5% |
| BUSCO frag. | 0.6% | 0.4% |
| BUSCO miss. | 1.1% | 1.1% |

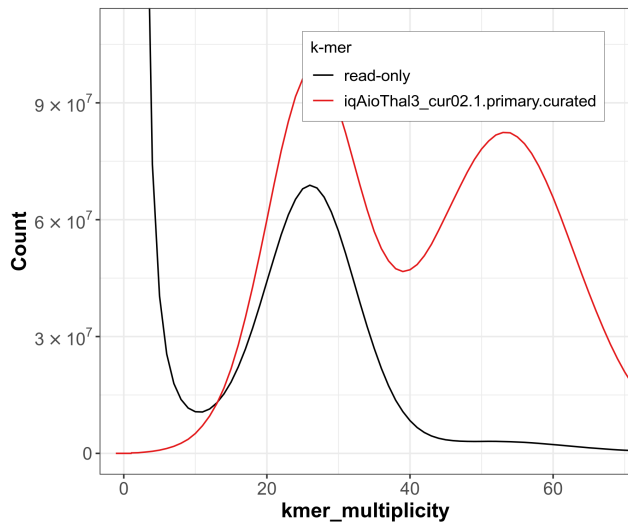
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: insecta_odb12 (genomes:79, BUSCOs:3114)

HiC contact map of curated assembly

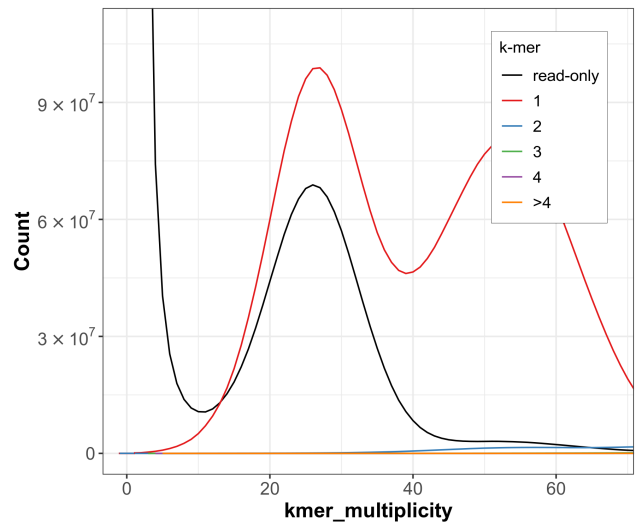


pri [\[LINK\]](#)

K-mer spectra of curated assembly

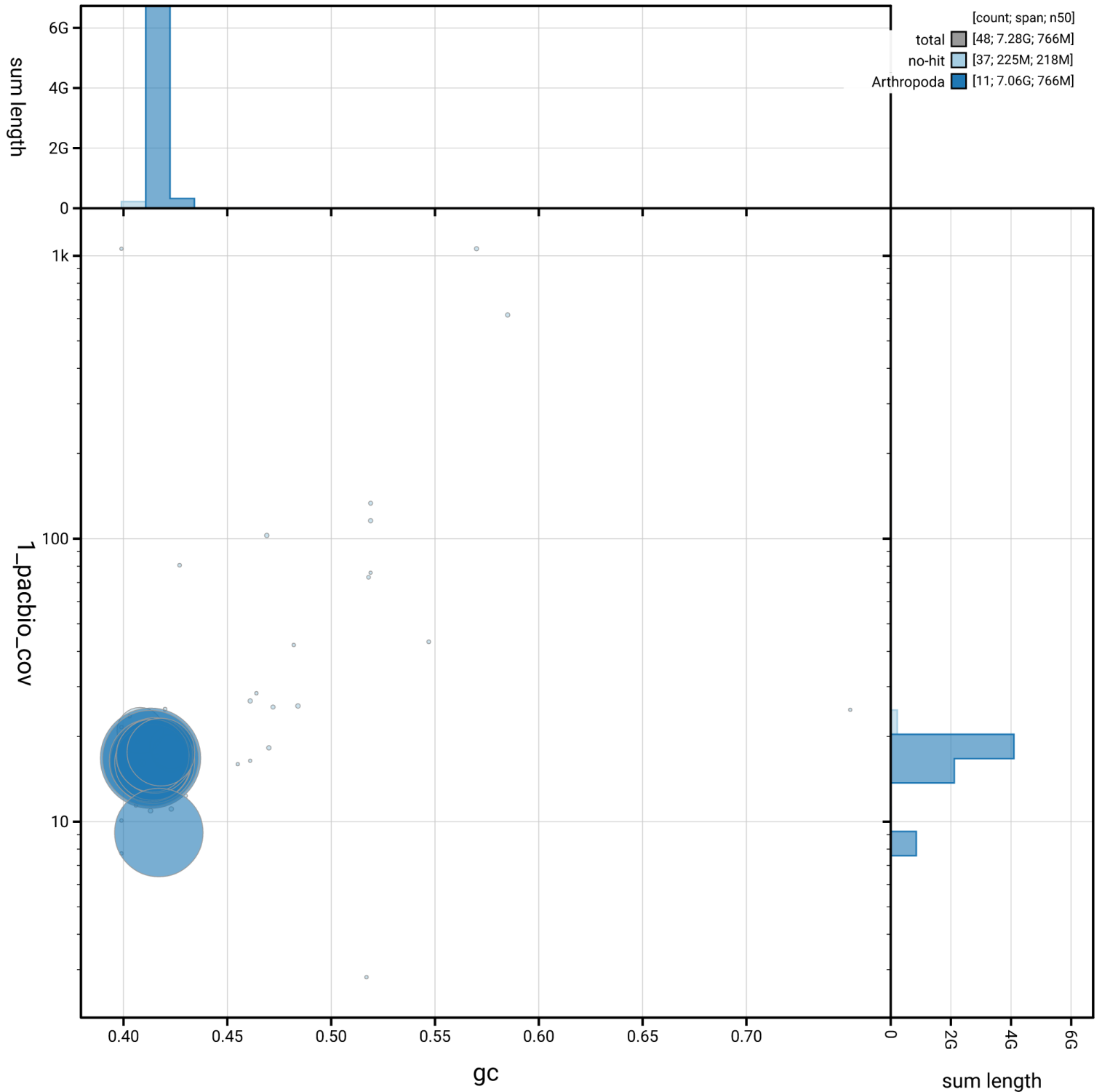


Distribution of k-mer counts coloured by their presence in reads/assemblies



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 | 53x | 210x |

Assembly pipeline

- **Hifiasm**
 - |_ *ver*: 0.25.0-r726
 - |_ *key param*: 13
- **purge_dups**
 - |_ *ver*: 1.2.6
 - |_ *key param*: NA

Curation pipeline

- **GRIT_Rapid**
 - |_ *ver*: 1a3d79a8
 - |_ *key param*: NA

Submitter: Guilherme Dias Martin Pippel

Affiliation: SciLifeLab

Date and time: 2026-02-16 17:02:44 CET