

1 **Chromosome-scale genome assembly and annotation of the two-spotted** 2 **cricket *Gryllus bimaculatus* (Orthoptera: Gryllidae)**

3 Kosuke Kataoka^{1, 2*}, Ryuto Sanno⁴, Tomasz Gaczorek⁴, Upendra Raj Bhattarai^{5, 6}, Yuki Ito³,
4 Shintaro Inoue⁷, Kei Yura^{2, 3, 8}, Toru Asahi^{2, 3}, Guillem Ylla^{4*}, Taro Mito^{7*}, Cassandra G.
5 Extavour^{5, 9, 10*}

- 6 1. Institute of Engineering, Tokyo University of Agriculture and Technology, 2-24-16 Naka-
7 cho, Koganei-shi, Tokyo 184-8588, Japan.
- 8 2. Comprehensive Research Organization, Waseda University, 2-2 Wakamatsu-cho,
9 Shinjuku-ku, Tokyo 162-8480, Japan.
- 10 3. Graduate School of Advanced Science and Engineering, Waseda University, 2-2
11 Wakamatsu-cho, Shinjuku-ku, Tokyo 162-8480, Japan.
- 12 4. Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, 7
13 Gronostajowa, Kraków 30-387, Poland.
- 14 5. Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford
15 Street, Cambridge, MA 02138, USA.
- 16 6. Harvard T.H. Chan School of Public Health, Harvard University, 677 Huntington Avenue,
17 Boston, MA 02115, USA.
- 18 7. Bio-Innovation Research Center, Tokushima University, 2272-2 Ishii, Ishii, Ishii-cho,
19 Myozai-gun, Tokushima 779-3233, Japan.
- 20 8. Graduate School of Humanities and Sciences, Ochanomizu University, 2-1-1 Otsuka,
21 Bunkyo-ku, Tokyo 112-8610, Japan.
- 22 9. Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue,
23 Cambridge, MA 02138, USA.
- 24 10. Howard Hughes Medical Institute, 4000 Jones Bridge Road, Chevy Chase, MD 20815,
25 USA.

26

1 * Correspondence:

2 Kosuke Kataoka, Institute of Engineering, Tokyo University of Agriculture and Technology,
3 Tokyo, Japan. Email: kataokak@go.tuat.ac.jp

4 Guillem Ylla, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University,
5 Kraków, Poland. Email: guillem.ylla@uj.edu.pl

6 Taro Mito, Bio-Innovation Research Center, Tokushima University, Tokushima, Japan. Email:
7 mito.taro@tokushima-u.ac.jp

8 Cassandra G. Extavour, Department of Organismic and Evolutionary Biology, Harvard
9 University, Cambridge, MA, USA. Email: extavour@oeb.harvard.edu

10

11 **Abstract**

12

13 The two-spotted cricket, *Gryllus bimaculatus*, is a key hemimetabolous model organism for
14 developmental biology, neuroscience, and regeneration. The existing reference genome is,
15 however, highly fragmented into 47,877 scaffolds, hampering chromosome-scale analyses
16 for these fields. Here, we report a high-quality, chromosome-scale genome assembly for the
17 white-eyed mutant strain of this cricket, generated using a combination of Nanopore and
18 PacBio HiFi long reads, integrated with Hi-C data. The final 1.62 Gbp assembly achieves a
19 scaffold N50 of 107.4 Mbp, a significant improvement in contiguity over the previous 6.3 Mbp
20 N50. We anchored 94.45% of the assembly into 15 pseudomolecules, consistent with the
21 known karyotype ($n = 15$). The genome completeness (BUSCO v6.0.0 insecta_odb12)
22 reached 98.1%. We also updated the annotation, identifying 14,964 protein-coding genes.
23 This gene set shows markedly improved completeness (BUSCO v6.0.0 insecta_odb12:
24 95.7%) compared with the previous annotation (81.2%) and successfully recovers all nine
25 essential neuropeptide genes previously reported as missing from the draft assembly. This
26 chromosome-scale genomic resource provides an essential foundation for comparative and
27 functional genomics in *G. bimaculatus*.

28

29 **Keywords:** Orthoptera; Gryllidae; Two-spotted cricket; White-eyed mutant strain; Genome
30 assembly

1 Introduction

2

3 The two-spotted cricket, *Gryllus bimaculatus*, is a model organism for hemimetabolous
4 insects (Horch et al., 2017) (Fig. 1). Unlike holometabolous insects such as the fruit fly
5 (*Drosophila melanogaster*) or the red flour beetle (*Tribolium castaneum*), hemimetabolous
6 insects undergo direct development, where nymphs hatch and grow through successive
7 molts to become adults without larval or pupal stages. This ancestral developmental mode
8 is crucially important for understanding insect evolution. Due to its ease of rearing, short
9 generation time, and the applicability of efficient RNA interference (RNAi) (Miyawaki et al.,
10 2004) and genome-editing techniques such as CRISPR/Cas9 (Matsuoka et al., 2025), *G.*
11 *bimaculatus* has been established as a powerful experimental model in a wide range of
12 fields, including evolutionary developmental biology (Donoughe & Extavour, 2016),
13 regeneration biology (Nakamura et al., 2008), neuroscience (Matsumoto et al., 2018), and
14 ethology (Abe et al., 2021; Kuriwada, 2022). It is also gaining significant global attention as a
15 novel, sustainable food source due to its high protein content and efficient rearing (Kataoka
16 et al., 2020, 2022; Mito et al., 2022).

17

18 In the first report of a genomic resource for *G. bimaculatus*, the white-eyed mutant strain,
19 which is standardly used in many functional studies, was sequenced (Ylla et al., 2021). The
20 assembly (GenBank accession: GCA_017312745.1) was a useful resource with a genome
21 size of approximately 1.66 Gb and a scaffold N50 of 6.3 Mb, contributing to analyses of gene
22 family evolution and DNA methylation. However, this previous assembly was fragmented
23 into 47,877 scaffolds and was not assembled to chromosome scale. While chromosome-
24 scale genomes have recently been reported for other related cricket species, such as *Gryllus*
25 *assimilis* (Ito et al., 2025) and *Acheta domesticus* (Dossey et al., 2023), such a resource has
26 remained unavailable for *G. bimaculatus*, which serves as the primary model for functional,
27 developmental, and neurobiological studies. The lack of a chromosome-scale assembly for
28 this specific species has been a significant limitation for large-scale comparative genomic
29 analyses based on synteny (conservation of gene order), understanding the structural
30 arrangement of transposons and repetitive sequences on chromosomes, and for
31 quantitative trait locus mapping and the accurate identification of genome-editing off-target
32 sites.

33

34 In this study, to fill the gap, we constructed a high-quality, chromosome-scale genome
35 assembly using the same, white-eyed mutant strain, used in the previous study. By

1 combining Nanopore and PacBio HiFi long-read sequencing methods, and Hi-C chromatin
2 conformation capture technology, we anchored 94.45% of the entire genome into 15
3 pseudomolecules, corresponding to the *G. bimaculatus* karyotype ($n = 14$ autosomes + X)
4 (Yoshimura et al., 2006). This assembly achieves a contig N50 of 4.58 Mb and a scaffold N50
5 of 107.39 Mb, representing a significant improvement in contiguity. Furthermore, we report
6 an updated gene annotation comprising 14,964 protein-coding genes. This chromosome-
7 scale genome resource will strengthen the foundation for all genomic research using *G.*
8 *bimaculatus* and accelerate new insights into the biology of hemimetabolous insects.

10 **Materials and Methods**

12 **Animals**

14 A white-eyed mutant strain of *G. bimaculatus* (Mito & Noji, 2008) was housed in plastic cases
15 at $30\text{ }^{\circ}\text{C} \pm 1\text{ }^{\circ}\text{C}$ and 30%–40% relative humidity under a 10 h light and 14 h dark photoperiod.
16 This strain has been widely used in genetic and developmental studies and was also used in
17 the previous genome assembly, enabling direct comparison with earlier genomic resources
18 (Miyawaki et al., 2004; Ylla et al., 2021). They were nourished with an artificial fish food
19 (4971618–011312, Kyorin, Japan).

21 **Library preparation and sequencing**

23 A single alive *G. bimaculatus* individual (Fig. 1) was used for genomic DNA extraction. Total
24 genomic DNA was extracted from the head and hind legs of a male *G. bimaculatus* using
25 NucleoBond® HMW DNA (Macherey-Nagel, Germany) according to the manufacturer's
26 instructions. This species exhibits an XX/X0 sex determination system, in which males are
27 hemizygous for the X chromosome; therefore, a male individual was selected to enable clear
28 identification and chromosome-scale assembly of the X chromosome. The resulting
29 genomic DNA was size-selected using a Short Read Eliminator Kit (PacBio, CA, USA). DNA
30 purity and concentrations were measured by spectrometry using NanoPhotometer NP80-
31 TOUCH (Implen, Germany) and fluorometry using Qubit 4 (Thermo Fisher Scientific, MA,
32 USA).

1
2 For long-read sequencing, Oxford Nanopore Technologies (ONT) libraries were constructed
3 using the Ligation Sequencing Kit V14 and sequenced on the PromethION 2 Solo platform
4 (Oxford Nanopore Technologies, UK) with a Flow Cell R10.4.1. Base-calling was performed
5 using Dorado v0.3.0 (model: dna_r10.4.1_e8.2_400bps_sup@v4.2.0). The resulting raw
6 reads were adapter-trimmed using Porechop_ABI v0.5.0 (Bonenfant et al., 2023), and reads
7 with a mean quality score below Q10 were removed using NanoFilt v2.8.0 (De Coster et al.,
8 2018). Additionally, a SMRTbell library was prepared and sequenced on a PacBio Sequel IIe
9 system.

10
11 For chromosome-scale scaffolding, two separate Hi-C libraries were prepared. The first
12 library was prepared from the hind legs of a single male *G. bimaculatus* using the Dovetail™
13 Omni-C™ Kit (Dovetail Genomics, CA, USA) following the manufacturer's instructions. The
14 second Hi-C library was generated from the thorax and legs of an adult male using the
15 Proximo Hi-C (Animal) kit (KT2045) (Phase Genomics, Seattle, US), following the
16 manufacturer's instructions. Both libraries were sequenced on the Illumina NovaSeq 6000
17 platform, and the sequencing data were combined for downstream scaffolding analysis.

18
19 Genome size estimation

20
21 Genome size estimation was performed using ONT reads. A total of 45.0 Gbp of ONT
22 sequencing data was used for k-mer-based analysis. Reads were counted using Jellyfish
23 v2.2.10 (Marçais & Kingsford, 2011) with a k-mer size of 21. A k-mer frequency histogram was
24 generated and used as input for GenomeScope v2.0 (Ranallo-Benavidez et al., 2020) to
25 estimate genome size and sequence characteristics. GenomeScope was run with default
26 parameters, and the resulting model was used to infer genome size based on the major k-
27 mer peak.

28
29 Genome *de novo* assembly

30
31 The initial draft genome was assembled by combining the filtered ONT long reads and the
32 PacBio HiFi reads using Flye v2.9.5 (Kolmogorov et al., 2019). To correct errors in the resulting

1 contigs, we retrieved publicly available Illumina short-read data (DDBJ Sequence Read
2 Archive [DRA] accessions DRR272308–DRR272313), which were originally sequenced on an
3 Illumina HiSeq 2000. These reads were downsampled to an approximate 93.9x coverage and
4 used for polishing. The assembly underwent two rounds of error correction using POLCA
5 v4.1.0 (Zimin & Salzberg, 2020) with default settings.

6
7 Potential contamination in the assembly was removed using BlobToolKit v1.1.1 (Challis et
8 al., 2020), which analyzes unexpected coverage, GC content, or similarity to bacterial and
9 other contaminant sequences. Sequence coverage was determined by mapping Illumina
10 reads with bwa v0.7.17-r1188. Similarity analysis was performed using BLASTn v2.13.0+
11 against NCBI NT database v5 (options: -task megablast culling_limit 10 -evalue 1e-25 -
12 outfmt '6 qseqid staxids bitscore std sscinames sskindoms stitle') to assign putative
13 taxonomic origins of contigs as part of contamination screening and assembly quality
14 control. Based on these analyses, no contigs could be confidently identified as apparent
15 contaminants.

16
17 The resulting contigs were then corrected for misjoins, ordered, oriented, and anchored into
18 a chromosome-scale assembly using Omni-C™ data with Juicer v1.9.9 (Durand et al., 2016)
19 and 3D-DNA v180419 (Dudchenko et al., 2017). Candidate assembly was reviewed with
20 Juicebox Assembly Tools v1.9.9 (Durand et al., 2016) for quality control and interactive
21 corrections. The contact map was visualized using Juicebox Assembly Tools v1.9.9. The
22 completeness of the final genome assembly was assessed using BUSCO v6.0.0 (Tegenfeldt
23 et al., 2025) against the insecta_odb12 lineage datasets. In addition, k-mer-based assembly
24 completeness and consensus accuracy were evaluated using Merqury v1.3 (Rhie et al.,
25 2020).

26
27 Mitochondrial genome assembly

28
29 The mitochondrial genome was assembled from PacBio HiFi reads using MitoHiFi v2 (Uliano-
30 Silva et al., 2023), following the default workflow. Mitochondrial sequences were excluded
31 from the nuclear genome assembly statistics but are provided as part of the released
32 genome resources for transparency.

33

1 Prediction of repeat regions

2

3 In *de novo* repeat prediction, RepeatModeler v2.0.6 (Flynn et al., 2020) was first used for *de*
4 *novo* repeat identification. Because standard libraries are insufficient for effective masking
5 in *Gryllus* genomes (Szrajer et al., 2024), this library was supplemented with the custom
6 repeat library previously generated (Ylla et al., 2021). This custom library was constructed by
7 integrating multiple complementary homology-based and *de novo* repeat identification
8 approaches, enabling comprehensive annotation of diverse repeat classes
9 (https://github.com/guillemylla/Crickets_Genome_Annotation). The combined repeat
10 library was then used to identify and softmask repetitive elements in the *G. bimaculatus*
11 genome using RepeatMasker v4.2.1 (Smit et al., 2015).

12

13 Structural gene annotation

14

15 Structural annotation for protein-coding genes was performed on the softmasked genome
16 using *ab initio* prediction and RNA-seq-based prediction. Both methods used publicly
17 available RNA-seq data (Sequence Read Archive [SRA] accessions: SRR10619411,
18 SRR10619415, SRR10619417, SRR10619418, SRR10619421, SRR10619423, SRR10619425,
19 SRR10619429, SRR10619431, SRR10619432, SRR10619434, SRR10619437, SRR10619439,
20 SRR10619440, SRR14026720–SRR14026726) as input.

21

22 To remove noisy RNA-seq reads potentially arising from erroneous transcription and splicing,
23 *de novo* transcriptome assembly was first performed using Trinity v2.15.1 (Haas et al., 2013)
24 to generate contigs. The original RNA-seq reads were then mapped back to these contigs
25 using HISAT2 v2.2.1 (Kim et al., 2019) with default parameters, allowing filtration of reads
26 that did not map correctly in the proper orientation as paired-end reads. After removing
27 these noisy reads, the remaining reads were subsequently used for gene predictions.

28

29 The *ab initio* prediction was carried out using BRAKER v3.0.8 (Brůna et al., 2021; Gabriel et
30 al., 2024; Hoff et al., 2016, 2019; Stanke et al., 2008, 2006), incorporating protein data from
31 the OrthoDB 12 arthropods dataset (retrieved from [https://bioinf.uni-](https://bioinf.uni-greifswald.de/bioinf/partitioned_odb12/)
32 [greifswald.de/bioinf/partitioned_odb12/](https://bioinf.uni-greifswald.de/bioinf/partitioned_odb12/)) and the mapping data of the filtered RNA-seq
33 reads. This BRAKER prediction served as the foundation for our gene set. To complement this

1 base set, StringTie2 v2.2.1 (Kovaka et al., 2019) was used for RNA-seq-based prediction.
2 During this process, StringTie2-derived transcripts were compared against the BRAKER gene
3 models, and only non-redundant transcript classes classified by GffCompare v0.12.6
4 (Pertea & Pertea, 2020) as “u” (intergenic), “i” (intronic), or “y” (contains a reference
5 transcript within its intron) were retained and added to the BRAKER-based gene set.
6 Redundant or overlapping predictions were discarded. This procedure resulted in a final,
7 non-redundant consensus gene set.

8

9 Functional gene annotation

10

11 Gene functional annotation was conducted using eggNOG-mapper online ([http://eggnog-](http://eggnog-mapper.embl.de/)
12 [mapper.embl.de/](http://eggnog-mapper.embl.de/)) (Cantalapiedra et al., 2021) and BLASTp-based methods. For the BLASTp-
13 based annotation, we used databases including *Homo sapiens*, *Mus musculus*,
14 *Caenorhabditis elegans*, *D. melanogaster*, *Tribolium castaneum*, and UniProt Swiss-Prot to
15 identify the best hits for annotation (E-value < 1.0×10^{-10}).

16

17 The completeness of the final predicted gene models was assessed using BUSCO v6.0.0
18 (Tegenfeldt et al., 2025) against the insecta_odb12 lineage dataset. In addition, annotation
19 quality was further evaluated using OMArk v0.4.1 (Nevers et al., 2025). For these analyses,
20 the longest isoform for each gene was first extracted from the annotation file using AGAT
21 v0.9.1 (Dainat et al., 2022).

22

23 Phylogenomic analysis

24

25 We inferred a phylogenetic tree using single-copy orthologs identified by BUSCO
26 (insecta_odb12) from predicted protein-coding genes of *G. bimaculatus* and related insect
27 species. BUSCO analyses were performed using BUSCO v6.0.0, and only single-copy
28 BUSCO proteins shared across all taxa were retained. For each BUSCO ortholog, amino-acid
29 sequences were aligned using MAFFT v7.525 (Kato & Standley, 2013) and subsequently
30 trimmed to remove poorly aligned regions with trimAl v1.5.rev1 (Capella-Gutiérrez et al.,
31 2009). The resulting alignments were concatenated into a supermatrix using AMAS v1.0
32 (Borowiec, 2016). A maximum-likelihood phylogenetic tree was then inferred from the

1 concatenated alignment using IQ-TREE v3.0.1 (Wong et al., 2025), with partitioned model
2 selection and ultrafast bootstrap support.

3

4 Validation of neuropeptide gene loci

5

6 To assess the completeness of our assembly regarding functionally important gene families,
7 we specifically investigated the neuropeptide gene loci previously reported (Mochizuki et al.,
8 2023). The neuropeptide cDNA sequences listed in the study were retrieved. These
9 sequences were mapped against our final chromosome-scale genome assembly using
10 Exonerate v2.4.0 (Slater & Birney, 2005) with the est2genome model to accurately determine
11 exon-intron boundaries. All resulting alignments were manually inspected to validate the
12 gene structures.

13

14 Results and Discussion

15

16 Genome sequencing

17

18 To construct the genome assembly, we generated three types of sequencing data (Table 1).
19 First, we obtained 45.00 Gbp of ONT sequencing data; the average and N50 read lengths
20 were 13.13 Kbp and 24.23 Kbp, respectively. Second, we generated 13.44 Gbp of PacBio HiFi
21 data, with an average read length of 13.61 Kbp and an N50 of 14.17 Kbp. Finally, for
22 chromosome-scale scaffolding, 243.22 Gbp of Hi-C raw data was generated from the
23 Illumina platform.

24

25 Genome *de novo* assembly statistics

26

27 Genome size estimation based on Nanopore long-read data using a k-mer-based approach
28 yielded an estimated genome size of approximately 1.64 Gbp (Fig. S1). This estimate is highly
29 consistent with the final assembled genome size (1.62 Gbp) and with previous estimates
30 reported for *G. bimaculatus* (Ylla et al., 2021).

1 The hybrid assembly strategy combining ONT and HiFi reads, followed by two rounds of
2 Illumina-based polishing, yielded a 1.63 Gbp draft genome. This polished assembly
3 comprised 3,789 contigs and achieved a contig N50 of 4.58 Mbp (Table S1).

4
5 Following Hi-C-based scaffolding, the final assembly resulted in a genome of 1.62 Gbp in
6 size, visualized by a snail plot (Fig. 2A). This assembly consists of 196 scaffolds with a
7 scaffold N50 length of 107 Mbp (Table 2). This represents a substantial improvement in
8 contiguity compared to the previous *G. bimaculatus* assembly (Ylla et al., 2021), which had
9 a scaffold N50 of 6.3 Mbp and was fragmented into 47,877 scaffolds (Fig. 2B). To assess
10 genomic completeness, we performed a BUSCO v6.0.0 analysis using the insecta_odb12
11 dataset. Our assembly achieved a completeness score of 98.1%
12 (C:98.1%[S:96.0%,D:2.2%],F:0.4%,M:1.5%), demonstrating a higher level of completeness
13 compared to the 96.0% (C:96.0%[S:94.3%,D:1.7%],F:1.4%,M:2.6%) of the previous
14 genome (Table 3). A k-mer-based evaluation using Merqury showed that 85.5% of the
15 reliable k-mers derived from the sequencing reads were represented in the genome
16 assembly, with an estimated consensus quality value (QV) of 33.6, corresponding to a
17 base-level accuracy of approximately 99.96%.

18
19 A total of 15 pseudochromosomes were identified based on Hi-C-guided scaffolding using
20 3D-DNA, and accounted for 94.45% of the total genome length (Table 4). This number (n =
21 15) is in agreement with the established karyotype for *G. bimaculatus* (n = 14 autosomes +
22 X), which was previously determined by cytogenetic analysis (Yoshimura et al., 2006) and is
23 also consistent with that of the recently reported species *G. assimilis* (Ito et al., 2025), which
24 is closely related to *G. bimaculatus*. The X chromosome was identified as the longest,
25 accounting for 15.68% of the genome, which is in accordance with the karyotype of this
26 species (Yoshimura et al., 2006). This was further validated by the observation that the X
27 chromosome displayed half of the read coverage compared to the autosomal
28 chromosomes, calculated using a genomic short-read library (DRR272308) from a single
29 hemizygous male (Fig. S2).

30
31 Repetitive sequence annotation revealed that 823.1 Mbp of the genome, corresponding to
32 50.72% of the assembled sequence, is composed of repetitive elements (Table 5). This
33 includes a substantial contribution from both Class I retroelements (13.95%) and Class II
34 DNA transposons (10.53%), as well as a notable fraction of unclassified repeats (18.15%).
35 The proportion of annotated repetitive sequences is markedly higher than that reported in

1 the previous *G. bimaculatus* assembly, in which only 33.69% of the genome was annotated
2 as repetitive (Ylla et al., 2021), likely reflecting improved assembly contiguity in the present
3 study.

4

5 Gene annotation

6

7 The final consensus gene set, derived from merging BRAKER *ab initio* predictions and
8 StringTie2 RNA-seq-based predictions, comprised 14,964 protein-coding genes (Table 6).
9 This total is less than the 17,871 genes reported previously (Ylla et al. 2021), likely because
10 the improved scaffold contiguity of our assembly allows for the correct assembly of genes
11 previously split across multiple scaffolds. Of the 14,964 predicted genes, functional
12 annotation was assigned using eggNOG-mapper and BLASTp (E-value < 1.0×10^{-10}) against
13 several model organisms' annotation datasets. eggNOG-mapper annotated 71.96% of the
14 genes. The BLASTp searches (E-value < 1.0×10^{-10}) against databases including *H. sapiens*,
15 *M. musculus*, *C. elegans*, *D. melanogaster*, *T. castaneum*, and UniProt Swiss-Prot yielded hit
16 rates ranging from 44.84% to 71.89% (Table 7).

17

18 The completeness of this annotation set was validated using BUSCO v6.0.0. The analysis
19 identified 95.7% of the expected complete insecta BUSCOs
20 (C:95.7%[S:93.6%,D:2.0%],F:1.5%,M:2.9%) (Table 8). OMArk analysis indicated a
21 completeness of 92.67% and a consistency score of 71.54% (Table S2). Together, these
22 independent metrics support the overall quality of the predicted gene models.

23

24 We next inferred a maximum-likelihood phylogeny using single-copy BUSCO orthologs
25 shared across *G. bimaculatus* and closely related orthopteran species (Fig. S3). The
26 resulting topology places *G. bimaculatus* within Gryllidae and in the expected relationship
27 to other *Gryllus* and cricket lineages, providing an additional consistency check and a
28 comparative framework for future evolutionary analyses using this updated genome
29 resource.

30

31 Recovery of missing neuropeptide genes

32

1 A recent comprehensive study (Mochizuki et al., 2023) highlighted significant gaps in the
2 previous draft genome assembly (Ylla et al., 2021). They reported that several crucial
3 neuropeptide genes (e.g., ACP (Adipokinetic hormone/corazonin-related peptide),
4 Allatotropin, Kinin, etc.) were missing from the draft genome and could only be identified
5 within *de novo* transcriptome assemblies, suggesting these loci were absent from the
6 previous reference.

7
8 We asked if our new chromosome-scale assembly was more complete in this regard, by
9 mapping the cDNA sequences of these previously missing neuropeptides. We successfully
10 located all nine of these genes encoding neuropeptides (i.e., ACP, Allatostatin CC (Ast CC),
11 Allatotropin, CCHamide-1, CCHamide-2, CRF/DH (Corticotropin releasing factor-like
12 diuretic hormone), Kinin (Leucokinin), Neuropeptide F1a (NPF1a), Neuropeptide F1b
13 (NPF1b)), which are now correctly anchored onto our pseudomolecules. For example, the
14 ACP gene, previously missing, was successfully mapped to Chromosome X, where it spans
15 11,668 bp and is composed of three exons, revealing its complete exon-intron structure (Fig.
16 S4). This demonstrates that our assembly not only improves contiguity to the chromosome
17 scale but also recovers functionally critical genes that were absent in the previous reference,
18 providing a more complete and reliable resource for functional genomics in *G. bimaculatus*.

19

20 **Conclusions**

21

22 We have generated a high-quality, chromosome-scale genome assembly and updated gene
23 annotation for the key hemimetabolous model organism *G. bimaculatus*. This assembly
24 represents a substantial upgrade to the previous draft sequence (Ylla et al., 2021), increasing
25 the scaffold N50 from 6.3 Mbp to 107.4 Mbp and anchoring 94.45% of the sequence into 15
26 pseudomolecules, consistent with the known karyotype (Yoshimura et al., 2006).

27

28 Crucially, our assembly resolves significant gaps present in the previous version, evidenced
29 by the recovery of nine essential neuropeptide genes previously reported as missing
30 (Mochizuki et al., 2023). This improved completeness is further supported by superior
31 BUSCO scores for both the genome (98.1% vs. 96.0%) and the gene set (95.7% vs. 81.2%).
32 This highly contiguous and complete genome sequence provides an essential new
33 foundation for the *G. bimaculatus* research community, facilitating advanced genetic and

1 genomic analyses, such as synteny comparisons, QTL mapping, and the precise design of
2 genome-editing experiments.

3

4 **Data Availability Statement**

5

6 The scripts used for the analyses in this study are available in GitHub
7 (https://github.com/Kataoka-K-Lab/Gryllus_bimaculatus_genome_gbim_v2.2). All
8 bioinformatics tools used in this study followed their respective manuals and protocols. The
9 software versions, codes, and parameters are provided in the Materials and Methods
10 section. Unless otherwise specified, default parameters were used.

11

12 The assembled genome and annotation datasets are available in FigShare
13 (<https://doi.org/10.6084/m9.figshare.30472754.v1>) (Kataoka, 2025).

14

15 The genomic WGS sequencing data were deposited in the NCBI Sequence Read Archive
16 (SRA) database under the BioProject PRJNA1347939. All data and assembly statistics
17 reported in this study correspond to the version before contamination screening and
18 removal performed by NCBI.

19

20 **Funding**

21

22 This study was supported by the Cabinet Office, Government of Japan, Cross-ministerial
23 Moonshot Agriculture, Forestry and Fisheries Research and Development Program,
24 “Technologies for Smart Bio-industry and Agriculture” (BRAIN) [JPJ009237] (K.K., S.I., T.A.,
25 K.Y., and T.M.), the Strategic Programme Excellence Initiative at the Jagiellonian University –
26 BioS PRA (T.G. and G.Y.), Foundation for Polish Science (FNP) START 2025 (T.G.), and the
27 National Science Foundation Award [IOS-2220747] (C.G.E.). C.G.E. is an investigator of the
28 Howard Hughes Medical Institute.

29

30

1 **Competing Interests**

2

3 The authors declare no conflict of interest.

4

5 **References**

6 Abe, T., Tada, C., & Nagayama, T. (2021). Winner and loser effects of juvenile cricket *Gryllus*
7 *bimaculatus*. *Journal of Ethology*, 39(1), 47–54.

8 Bonenfant, Q., Noé, L., & Touzet, H. (2023). Porechop_ABI: discovering unknown adapters in
9 Oxford Nanopore Technology sequencing reads for downstream trimming.
10 *Bioinformatics Advances*, 3(1), vbac085.

11 Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation and computing of
12 summary statistics. *PeerJ*, 4(e1660), e1660.

13 Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M., & Borodovsky, M. (2021). BRAKER2: automatic
14 eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a
15 protein database. *NAR Genomics and Bioinformatics*, 3(1), lqaa108.

16 Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021).
17 eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain
18 Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, 38(12), 5825–
19 5829.

20 Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for
21 automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*
22 *(Oxford, England)*, 25(15), 1972–1973.

23 Challis, R., Richards, E., Rajan, J., Cochrane, G., & Blaxter, M. (2020). BlobToolKit - interactive
24 quality assessment of genome assemblies. *G3*, 10(4), 1361–1374.

25 Dainat, J., Hereñú, D., Davis, E., Crouch, K., LucileSol, Pascal-Git, & Tayyrov. (2022).
26 *NBISweden/AGAT: AGAT-v0.9.1*. Zenodo. <https://doi.org/10.5281/ZENODO.6488306>

- 1 De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack:
2 visualizing and processing long-read sequencing data. *Bioinformatics (Oxford,*
3 *England)*, 34(15), 2666–2669.
- 4 Donoughe, S., & Extavour, C. G. (2016). Embryonic development of the cricket *Gryllus*
5 *bimaculatus*. *Developmental Biology*, 411(1), 140–156.
- 6 Dossey, A. T., Oppert, B., Chu, F.-C., Lorenzen, M. D., Scheffler, B., Simpson, S., Koren, S.,
7 Johnston, J. S., Kataoka, K., & Ide, K. (2023). Genome and genetic engineering of the
8 house cricket (*Acheta domestica*): A resource for sustainable agriculture.
9 *Biomolecules*, 13(4), 589.
- 10 Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim,
11 M. S., Machol, I., Lander, E. S., Aiden, A. P., & Aiden, E. L. (2017). *De novo* assembly of
12 the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*,
13 356(6333), 92–95.
- 14 Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., & Aiden,
15 E. L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C
16 experiments. *Cell Systems*, 3(1), 95–98.
- 17 Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020).
18 RepeatModeler2 for automated genomic discovery of transposable element families.
19 *Proceedings of the National Academy of Sciences of the United States of America*,
20 117(17), 9451–9457.
- 21 Gabriel, L., Bruna, T., Hoff, K. J., Ebel, M., Lomsadze, A., Borodovsky, M., & Stanke, M. (2024).
22 BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence
23 with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Research*, 34(5), 769–777.
- 24 Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M.
25 B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi,
26 F., Weeks, N., Westerman, R., William, T., Dewey, C. N., ... Regev, A. (2013). *De novo*

- 1 transcript sequence reconstruction from RNA-seq using the Trinity platform for
2 reference generation and analysis. *Nature Protocols*, 8(8), 1494–1512.
- 3 Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., & Stanke, M. (2016). BRAKER1:
4 Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and
5 AUGUSTUS. *Bioinformatics*, 32(5), 767–769.
- 6 Hoff, K. J., Lomsadze, A., Borodovsky, M., & Stanke, M. (2019). Whole-genome annotation
7 with BRAKER. *Methods in Molecular Biology (Clifton, N.J.)*, 1962, 65–95.
- 8 Horch, H. W., Mito, T., Popadic, A., Ohuchi, H., & Noji, S. (Eds.). (2017). *The cricket as a model*
9 *organism: Development, regeneration, and behavior*. Springer.
- 10 Ito, Y., Sanno, R., Ashikari, S., Yura, K., Asahi, T., Ylla, G., & Kataoka, K. (2025). Chromosome-
11 scale whole genome assembly and annotation of the Jamaican field cricket *Gryllus*
12 *assimilis*. *Scientific Data*, 12(1), 826.
- 13 Kataoka, K. (2025). *Chromosome-scale genome assembly and annotation of the two-*
14 *spotted cricket Gryllus bimaculatus (Orthoptera: Gryllidae)* [Data set]. figshare.
15 <https://doi.org/10.6084/M9.FIGSHARE.30472754.V1>
- 16 Kataoka, K., Minei, R., Ide, K., Ogura, A., Takeyama, H., Takeda, M., Suzuki, T., Yura, K., &
17 Asahi, T. (2020). The draft genome dataset of the Asian cricket *Teleogryllus occipitalis*
18 for molecular research toward entomophagy. *Frontiers in Genetics*, 11, 470.
- 19 Kataoka, K., Togawa, Y., Sanno, R., Asahi, T., & Yura, K. (2022). Dissecting cricket genomes
20 for the advancement of entomology and entomophagy. *Biophysical Reviews*, 14(1),
21 75–97.
- 22 Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7:
23 improvements in performance and usability. *Molecular Biology and Evolution*, 30(4),
24 772–780.

- 1 Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome
2 alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*,
3 37(8), 907–915.
- 4 Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads
5 using repeat graphs. *Nature Biotechnology*, 37(5), 540–546.
- 6 Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L., & Pertea, M. (2019).
7 Transcriptome assembly from long-read RNA-seq alignments with StringTie2.
8 *Genome Biology*, 20(1), 278.
- 9 Kuriwada, T. (2022). Encounter with heavier females changes courtship and fighting efforts
10 of male field crickets *Gryllus bimaculatus* (Orthoptera: Gryllidae). *Journal of*
11 *Ethology*, 40(2), 145–151.
- 12 Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting
13 of occurrences of k-mers. *Bioinformatics (Oxford, England)*, 27(6), 764–770.
- 14 Matsumoto, Y., Matsumoto, C. S., & Mizunami, M. (2018). Signaling Pathways for Long-Term
15 Memory Formation in the Cricket. *Frontiers in Psychology*, 9, 1014.
- 16 Matsuoka, Y., Nakamura, T., Watanabe, T., Barnett, A. A., Tomonari, S., Ylla, G., Whittle, C.
17 A., Noji, S., Mito, T., & Extavour, C. G. (2025). Establishment of CRISPR/Cas9-based
18 knock-in in a hemimetabolous insect: targeted gene tagging in the cricket *Gryllus*
19 *bimaculatus*. *Development*, 152(1), dev199746.
- 20 Mito, T., Ishimaru, Y., Watanabe, T., Nakamura, T., Ylla, G., Noji, S., & Extavour, C. G. (2022).
21 Cricket: The third domesticated insect. *Current Topics in Developmental Biology*,
22 147, 291–306.
- 23 Mito, T., & Noji, S. (2008). The two-spotted cricket *Gryllus bimaculatus*: An emerging model
24 for developmental and regeneration studies. *Cold Spring Harbor Protocols*, 2008(12),
25 db.emo110.

- 1 Miyawaki, K., Mito, T., Sarashina, I., Zhang, H., Shinmyo, Y., Ohuchi, H., & Noji, S. (2004).
2 Involvement of Wingless/Armadillo signaling in the posterior sequential
3 segmentation in the cricket, *Gryllus bimaculatus* (Orthoptera), as revealed by RNAi
4 analysis. *Mechanisms of Development*, 121(2), 119–130.
- 5 Mochizuki, T., Sakamoto, M., Tanizawa, Y., Seike, H., Zhu, Z., Zhou, Y. J., Fukumura, K.,
6 Nagata, S., & Nakamura, Y. (2023). Best Practices for Comprehensive Annotation of
7 Neuropeptides of *Gryllus bimaculatus*. *Insects*, 14(2), 121.
- 8 Nakamura, T., Mito, T., Bando, T., Ohuchi, H., & Noji, S. (2008). Dissecting insect leg
9 regeneration through RNA interference: Dissecting insect leg regeneration through
10 RNA interference. *Cellular and Molecular Life Sciences*, 65(1), 64–72.
- 11 Nevers, Y., Warwick Vesztröcy, A., Rossier, V., Train, C.-M., Altenhoff, A., Dessimoz, C., &
12 Glover, N. M. (2025). Quality assessment of gene repertoire annotations with OMArk.
13 *Nature Biotechnology*, 43(1), 124–133.
- 14 Pertea, G., & Pertea, M. (2020). GFF utilities: GffRead and GffCompare. *F1000Research*,
15 9(304), 304.
- 16 Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. (2020). GenomeScope 2.0 and
17 Smudgeplot for reference-free profiling of polyploid genomes. *Nature*
18 *Communications*, 11(1), 1432.
- 19 Rhie, A., Walenz, B. P., Koren, S., & Phillippy, A. M. (2020). Merqury: reference-free quality,
20 completeness, and phasing assessment for genome assemblies. *Genome Biology*,
21 21(1), 245.
- 22 Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological
23 sequence comparison. *BMC Bioinformatics*, 6(1), 31.
- 24 Smit, A. F. A., Hubley, R., & Green, P. (2015). *RepeatMasker* (Open-4.0) [Computer software].

- 1 Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically
2 mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics*, *24*(5),
3 637–644.
- 4 Stanke, M., Schöffmann, O., Morgenstern, B., & Waack, S. (2006). Gene prediction in
5 eukaryotes with a generalized hidden Markov model that uses hints from external
6 sources. *BMC Bioinformatics*, *7*(1), 62.
- 7 Szrajter, S., Gray, D., & Ylla, G. (2024). The genome assembly and annotation of the cricket
8 *Gryllus longicercus*. *Scientific Data*, *11*(1), 708.
- 9 Tegenfeldt, F., Kuznetsov, D., Manni, M., Berkeley, M., Zdobnov, E. M., & Kriventseva, E. V.
10 (2025). OrthoDB and BUSCO update: annotation of orthologs with wider sampling of
11 genomes. *Nucleic Acids Research*, *53*(D1), D516–D522.
- 12 Uliano-Silva, M., Ferreira, J. G. R. N., Krasheninnikova, K., Darwin Tree of Life Consortium,
13 Formenti, G., Abueg, L., Torrance, J., Myers, E. W., Durbin, R., Blaxter, M., & McCarthy,
14 S. A. (2023). MitoHiFi: a python pipeline for mitochondrial genome assembly from
15 PacBio high fidelity reads. *BMC Bioinformatics*, *24*(1), 288.
- 16 Wong, T. K. F., Ly-Trong, N., Ren, H., Baños, H., Roger, A. J., Susko, E., Bielow, C., De Maio, N.,
17 Goldman, N., Hahn, M. W., Huttley, G., Lanfear, R., & Minh, B. Q. (2025, April 7). *IQ-*
18 *TREE 3: Phylogenomic Inference Software using Complex Evolutionary Models*.
19 EcoEvoRxiv. <https://ecoevorxiv.org/repository/view/8916/>
- 20 Ylla, G., Nakamura, T., Itoh, T., Kajitani, R., Toyoda, A., Tomonari, S., Bando, T., Ishimaru, Y.,
21 Watanabe, T., Fuketa, M., Matsuoka, Y., Barnett, A. A., Noji, S., Mito, T., & Extavour, C.
22 G. (2021). Insights into the genomic evolution of insects from cricket genomes.
23 *Communications Biology*, *4*(1), 733.
- 24 Yoshimura, A., Nakata, A., Mito, T., & Noji, S. (2006). The characteristics of karyotype and
25 telomeric satellite DNA sequences in the cricket, *Gryllus bimaculatus* (Orthoptera,
26 Gryllidae). *Cytogenetic and Genome Research*, *112*(3–4), 329–336.

1 Zimin, A. V., & Salzberg, S. L. (2020). The genome polishing tool POLCA makes fast and
 2 accurate corrections in genome assemblies. *PLoS Computational Biology*, 16(6),
 3 e1007981.

4

5 **Tables**

6

7

Table 1. Statistics for the DNA-seq data of the *G. bimaculatus* genome.

Platform	Raw Data (bp)	Average Read Length (bp)	N50 Read Length (bp)	Coverage (X)
ONT	45,004,338,081	13,129.8	24,234	27.73
PacBio HiFi	13,437,781,216	13,611.4	14,170	8.28
Hi-C libraries	243,215,968,945	150	150	149.87

8

9

Table 2. Summary statistics for the chromosome-scale assembly of *G. bimaculatus*.

Features	Statistic
Number of chromosomes	15
Number of scaffolds	196
Scaffold N50 (bp)	107,386,002
GC content (%)	40.36
Max scaffold size (bp)	254,522,992
Total N (bp)	1,440,159

Gaps (%)	0.089
Total size (bp)	1,622,898,960

1

2 **Table 3.** BUSCO completeness assessment of the *G. bimaculatus* genome assembly.

BUSCO v6.0.0 (Insecta)	This Study	Ylla et al. 2021
Complete BUSCOs	98.1%	96.0%
Single-copy complete BUSCOs	96.0%	94.3%
Duplicated complete BUSCOs	2.2%	1.7%
Fragmented BUSCOs	0.4%	1.4%
Missing BUSCOs	1.5%	2.6%

3

4 **Table 4.** Statistics of the assembled pseudochromosomes in *G. bimaculatus*.

Chromosome	Length (bp)	Proportion in Genome
chrX	254,523,492	15.68%
chr1	148,286,872	9.14%
chr2	109,203,633	6.73%
chr3	109,044,855	6.72%
chr4	108,018,790	6.66%
chr5	107,387,002	6.62%

chr6	100,738,338	6.21%
chr7	98,895,268	6.09%
chr8	83,461,245	5.14%
chr9	81,413,916	5.02%
chr10	78,899,412	4.86%
chr11	69,666,237	4.29%
chr12	66,393,650	4.09%
chr13	63,483,624	3.91%
chr14	53,360,289	3.29%
Total	1,532,776,623	94.45%

1
2 **Table 5.** Classification of repetitive sequences of the *G. bimaculatus* genome assembly.

Type	Length (bp)	Proportion in genome (%)
Retroelements	226,335,103	13.95
SINEs	20,383,703	1.26
Penelope	1,718,510	0.11
LINES	156,279,698	9.63
LTR elements	47,953,192	2.95
DNA transposons	170,814,501	10.53

Unclassified	294,555,500	18.15
Total interspersed repeats	691,705,104	42.62
Rolling-circles	20,709,593	1.28
Other repeats	113,971,824	7.02
Total	823,137,644	50.72

1

2

Table 6. Summary statistics for gene prediction in *G. bimaculatus*.

Features	
Number of protein-coding genes	14,964
Mean exons per mRNA	7.0
Mean gene length (bp)	29,659
Mean exon length (bp)	2,32
Mean intron length (bp)	4,696

3

4

Table 7. Structural statistics of the *G. bimaculatus* gene set.

Annotation Sources	
<i>Homo sapiens</i> (GRCh38)	44.84%
<i>Mus musculus</i> (GRCm39)	71.89%
<i>Caenorhabditis elegans</i> (WBcel235)	55.18%

<i>Drosophila melanogaster</i> (BDGP6.32)	67.16%
<i>Tribolium castaneum</i> (icTriCast1.1)	75.76%
Uniprot/Swissprot (release: 2020_06)	66.74%
eggNOG-mapper	71.96%

1

2

Table 8. Functional annotation summary for the *G. bimaculatus* gene set.

BUSCO v6.0.0 (Insecta)	This study	Ylla et al. 2021
Complete BUSCOs	95.7%	81.2%
Single-copy complete BUSCOs	93.6%	79.9%
Duplicated complete BUSCOs	2.0%	1.3%
Fragmented BUSCOs	1.5%	8.9%
Missing BUSCOs	2.9%	10.0%

3

4 **Figure Legends**

5

6 **Figure 1.** Adult male (left) and female (right) of the white-eyed mutant strain *G.*
7 *bimaculatus*.

8

9 **Figure 2.** Assembly statistics and contiguity of the *Gryllus bimaculatus* (white-eyed
10 strain) chromosome-scale genome.

11 **(A)** Snail plot visualizing the key statistics of the final assembly. The cumulative assembly
12 length (1.62 Gbp) is plotted in light orange, with the longest scaffold (254.5 Mbp) shown in

1 red. The N50 length (107.4 Mbp) is indicated by the orange arc. The inner purple spiral plots
2 the cumulative number of scaffolds (196 total) on a log scale, with white scale lines drawn
3 at successive orders of magnitude from 10 scaffolds onwards. The circumferential axis
4 indicates the base composition of the assembly (GC: 40.4%, AT: 59.6%, N: 0.1%). The donut
5 chart (top right) displays the BUSCO completeness (insecta_odb12), showing 98.1% total
6 complete ("Comp.") genes (light green), which includes 96.0% single-copy and 2.2%
7 duplicated ("Dup.") genes (dark green). An additional 0.4% were fragmented ("Frag.") genes
8 (pale green). An interactive version is available at [https://kataokaklab.github.io/snailplot-](https://kataokaklab.github.io/snailplot-assembly-stats/)
9 [assembly-stats/](https://kataokaklab.github.io/snailplot-assembly-stats/).

10 **(B)** Cumulative length plot comparing the contiguity of this assembly (blue line) with the
11 previous draft assembly (Ylla et al., 2021) (orange line). The y-axis shows the percentage of
12 the total genome length covered, and the x-axis (log scale) shows the number of scaffolds.

ACCEPTED MANUSCRIPT

Figure 1

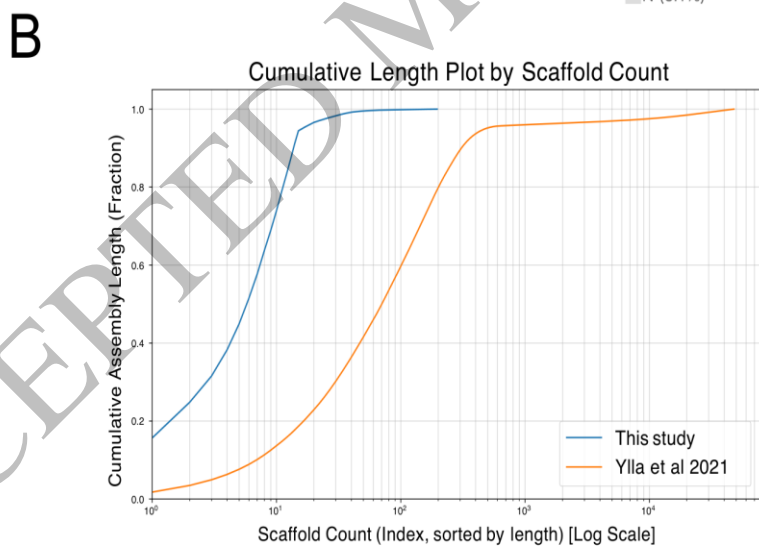
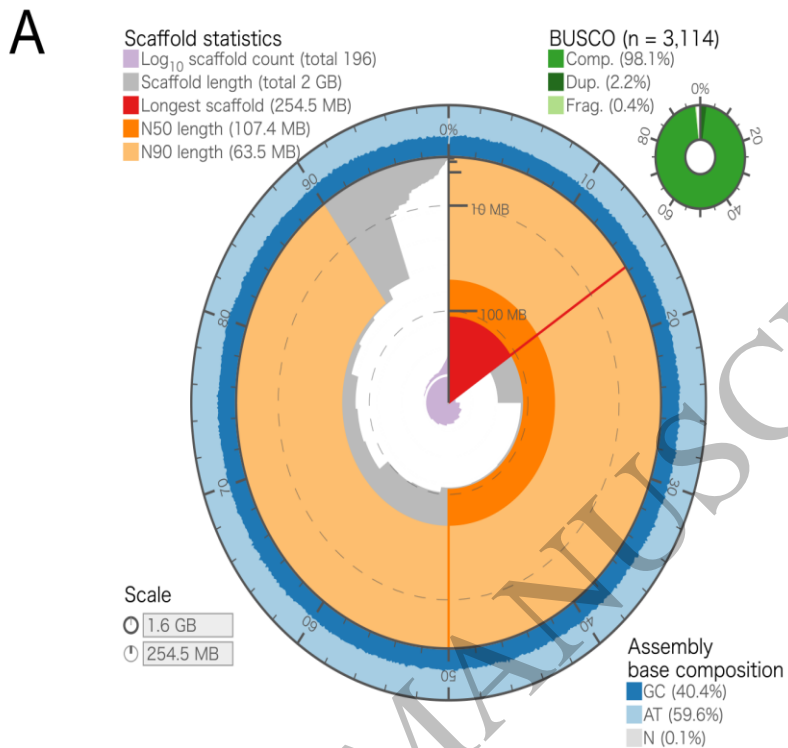


ACCEPTED MANUSCRIPT

1
2
3

Figure 1
210x297 mm (x DPI)

Figure 2



1
2
3

Figure 2
210x297 mm (x DPI)

Supplementary Figures

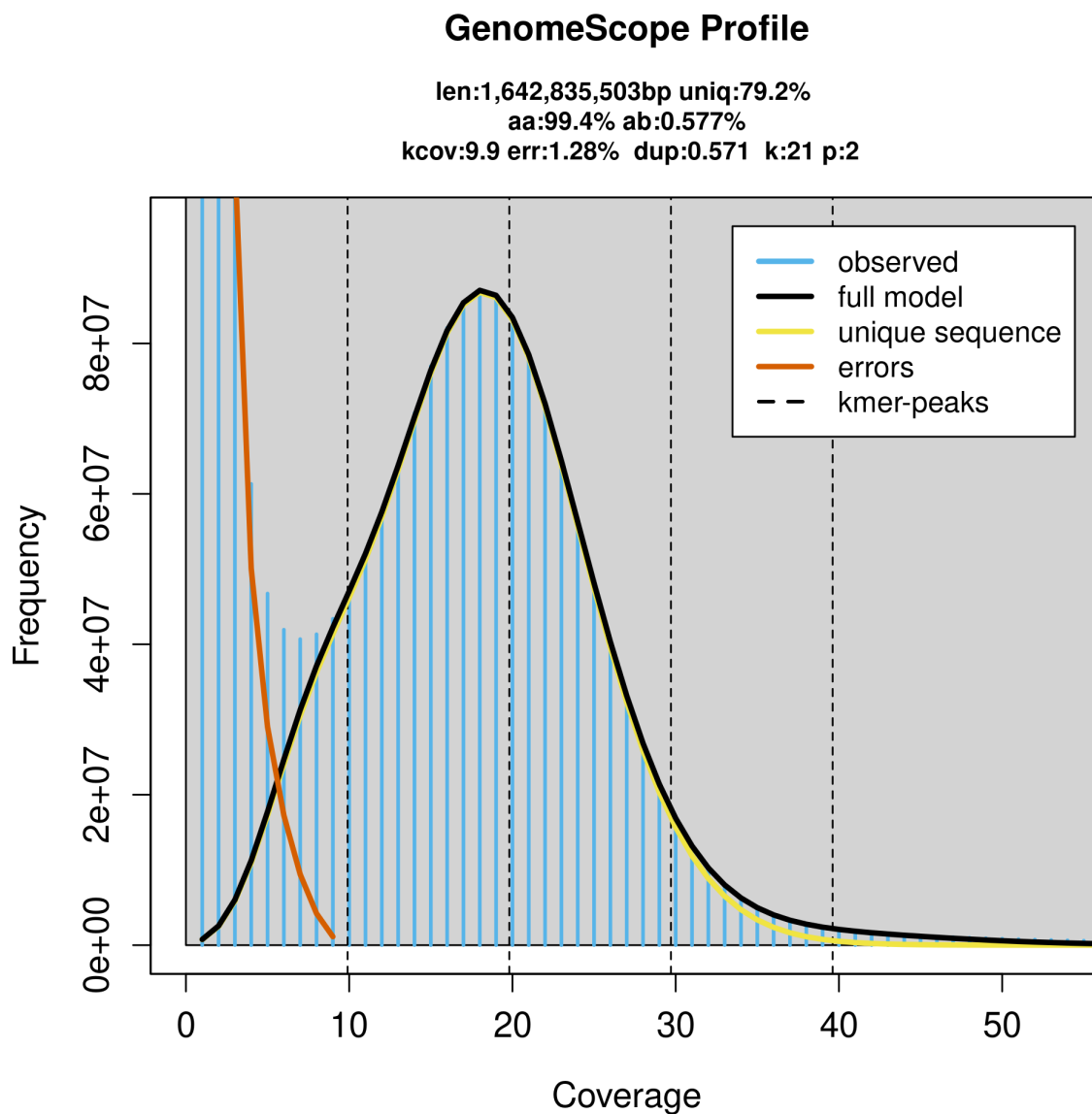


Figure S1. Genome size estimation of *Gryllus bimaculatus* based on ONT long-read data.

A k-mer frequency histogram (k = 21) generated from ONT long-reads was analyzed using GenomeScope. The model estimates a genome size of approximately 1.64 Gbp.

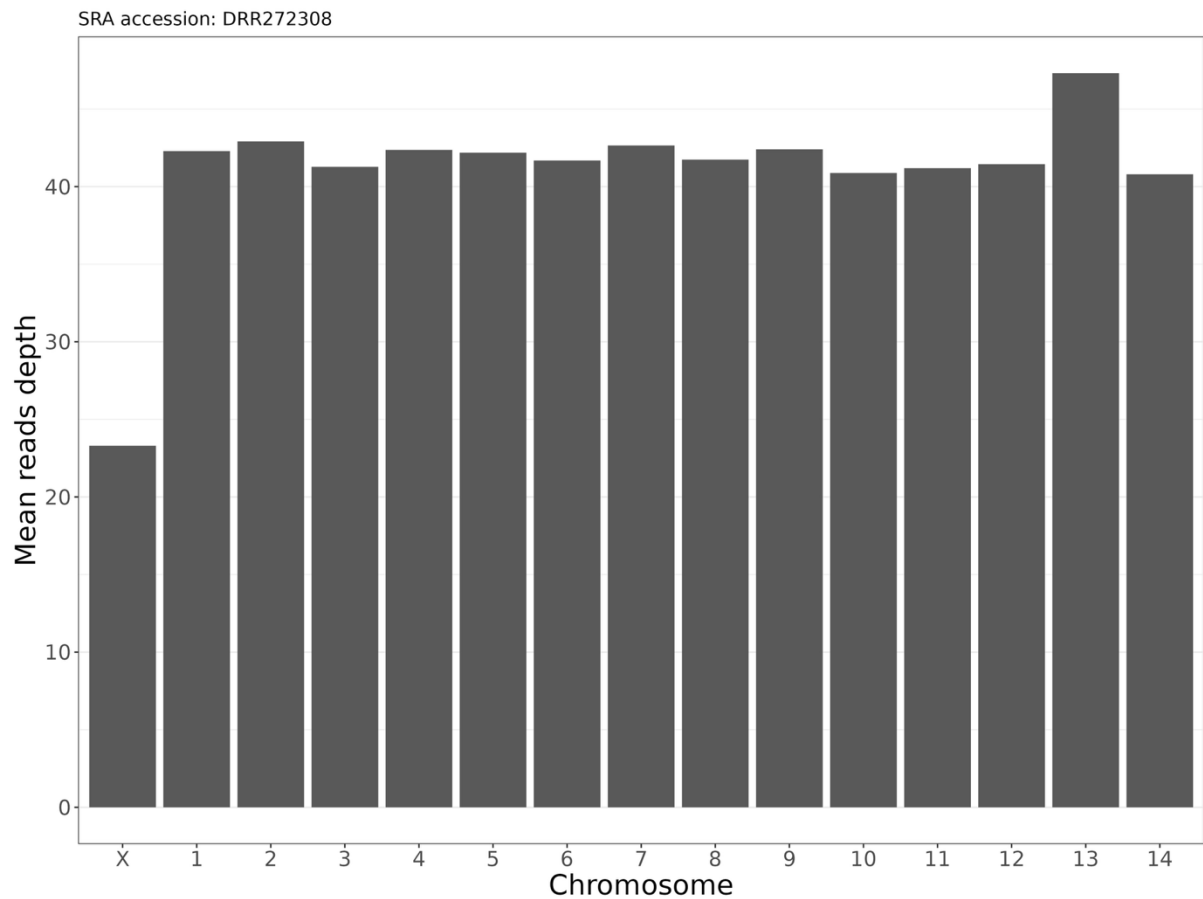


Figure S2. Genomic read coverage confirms X chromosome hemizyosity.

Mean read coverage depth across all assembled chromosomes derived from the mapping of genomic short reads of a single male *Gryllus bimaculatus* individual (SRA: DRR272308). The X chromosome displays approximately half the average autosomal coverage depth, consistent with the expected pattern of male hemizyosity.

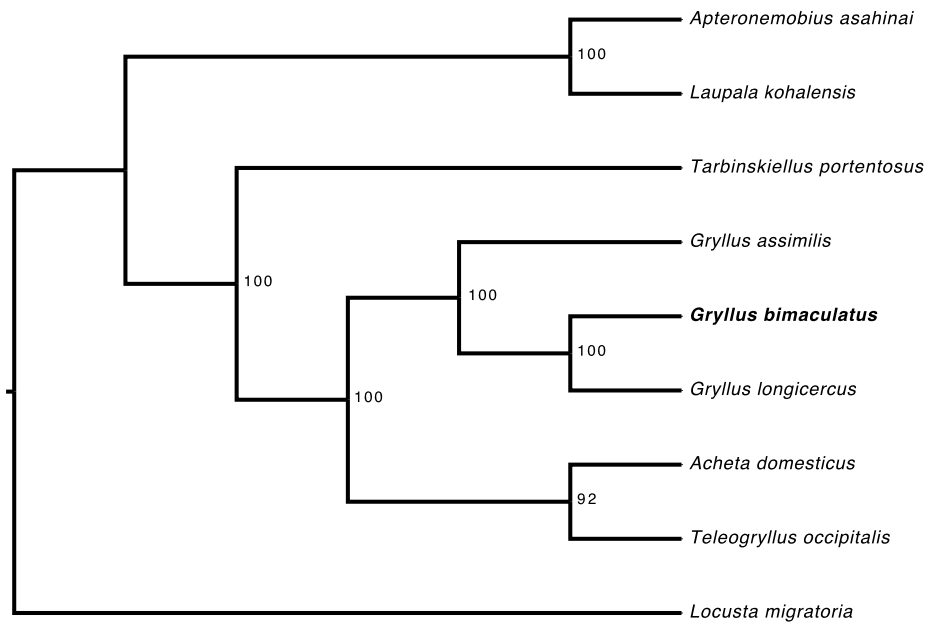


Figure S3. BUSCO-based phylogeny of *Gryllus bimaculatus* and related insects
 Single-copy BUSCO orthologs shared across all included taxa were extracted (insecta_odb12), aligned at the amino-acid level, trimmed, concatenated into a supermatrix, and used to infer a maximum-likelihood phylogeny with partitioned model selection. Node support values indicate bootstrap support.

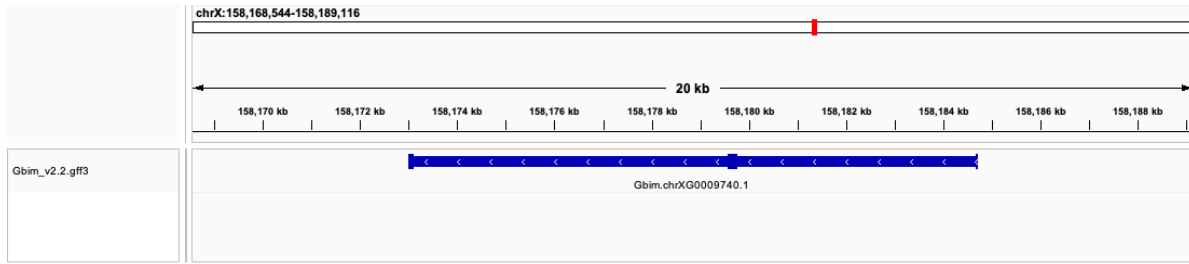


Figure S4. Recovery of the Adipokinetic hormone/corazonin-related peptide (ACP) gene, previously missing from the draft genome.

The image displays an Integrative Genomics Viewer (IGV) screenshot of the *G. bimaculatus* chromosome-scale assembly. The complete gene model for ACP (Gbim.chrXG0009740.1), one of the nine neuropeptide genes reported missing from the first assembly report (Mochizuki et al., 2023), is shown. The gene is now successfully anchored and annotated on Chromosome X (chrX), spanning a region of approximately 20 kb. The blue track (Gbim_v2.2.gff3) shows the full exon-intron structure (exons as thick blocks, introns as thin lines).

Table S1. Assembly statistics at major stages of the *Gryllus bimaculatus* genome assembly pipeline

Assembly Method	Flye v2.9.5	Flye v2.9.5 + POLCA v4.0.5 × 2	Flye v2.9.5 + POLCA v4.0.5 × 2 + 3D-DNA v180419
Number of sequences		3,725	3,789
Total length (nt)		1,637,450,406	1,634,638,790
N50 sequence length (nt)		4,224,697	4,578,840
L50 sequence count		112	99
GC-content (%)		40.39	40.38
BUSCO v5.1.2 (Arthropoda)	C:99.5%[S:96.8%,D:2.7%],F:0.1%,M:0.4%	C:99.4%[S:96.8%,D:2.6%],F:0.2%,M:0.4%	C:99.5%[S:96.7%,D:2.8%],F:0.1%,M:0.4%

Table S2. Evaluation of gene annotation completeness using OMArk (Neoptera)

Completeness assessment	Number of genes	Genes (%)
Total Completeness	4,284	92.67
Single	4,108	88.86
Duplicated	176	3.81
Duplicated, unexpected	174	3.76
Duplicated, expected	2	0.04
Missing	339	7.33
Consistency assessment		
Total Consistent	10,703	71.54
Consistent, partial hits	485	3.24
Consistent, fragmented	486	3.25
Total Inconsistent	1,809	12.09
Inconsistent, partial hits	485	3.24
Inconsistent, fragmented	486	3.25
Total Contamination	0	0.00
Contamination, partial hits	0	0.00
Contamination, fragmented	0	0.00
Total Unknown	2,449	16.37