Genetic Note

A high-quality reference genome and comparative genomics of the widely-farmed banded cricket (*Gryllodes sigillatus*) identifies selective breeding targets

Shangzhe Zhang[1], Kristin R. Duffield[2,3], Bert Foquet[4,5], Jose L. Ramirez[3], Ben M. Sadd[4], Scott K. Sakaluk[4], John Hunt[6], Nathan W. Bailey[1,*]

[1]School of Biology, University of St Andrews, St Andrews, Fife KY16 9TH, UK
[2]USDA-ARS, Geospatial and Environmental Epidemiology Research Unit, Mississippi State, MS, USA
[3]USDA-ARS, National Center for Agricultural Utilization Research, Crop BioProtection Research Unit, Peoria, IL, USA
[4]School of Biological Sciences, Illinois State University, Normal, IL 61790, USA
[5]McGuire Center for Lepidoptera and Biodiversity, Florida Museum of Natural History, University of Florida, Gainesville, FL, USA
[6]School of Science, Western Sydney University, Hawkesbury Campus, Penrith NSW 2751, Australia

* Correspondence to Nathan W. Bailey: nwb3@st-andrews.ac.uk

Keywords: banded cricket | cricket farming | genome-assisted breeding | *Gryllodes sigillatus* | insect protein

Abstract

Farmed insects have gained attention as an alternative, sustainable source of protein with a lower carbon footprint than traditional livestock. We present a high-quality reference genome for one of the most commonly farmed insects, the banded cricket *Gryllodes sigillatus*. In addition to its agricultural importance, *G. sigillatus* is also a model in behavioural and evolutionary ecology research on reproduction and mating systems. We report comparative genomic analyses that clarify the banded cricket's evolutionary history, identify gene family expansions and contractions unique to this lineage, associate these with agriculturally important traits, and identify targets for genome-assisted breeding efforts. The high-quality *G. sigillatus* genome assembly plus accompanying comparative genomic analyses serve as foundational resources for both applied and basic research on insect farming and behavioural biology, enabling researchers to pinpoint trait-associated genetic variants, unravel functional pathways governing those phenotypes, and accelerate selective breeding efforts to increase the efficacy of large-scale insect farming operations.

Introduction

Global demand for sustainable protein sources has driven intense interest in alternative agriculture. Such efforts include insect farming, which can meet nutritional needs while significantly reducing environmental impacts (Grabowski et al., 2022; van Huis & Oonincx, 2017). Insects are known for their favourable feed conversion efficiency. They require significantly less land, water, and energy resources compared to traditional livestock and produce fewer greenhouse gases (Lange & Nakamura, 2023). These advantages make them an environmentally friendly option for addressing food security challenges (Huis et al., 2021; van Huis & Oonincx, 2017). Crickets (suborder Ensifera) have been farmed and consumed by humans for millennia, and over 60 species are known to be consumed across 49 countries (Magara et al., 2021). Among these, the banded cricket (*Gryllodes sigillatus*, Figure 1a) has gained considerable attention due to its rapid growth rate, high reproductive capacity, and palatability, making it a preferred species for large-scale farming operations (Kong et al., 2024; Magara et al., 2021).

Despite the growing commercial significance of *G. sigillatus*, publicly available, genetic resources for this species are limited. This scarcity of genetic information constrains the potential for selective breeding programs that could enhance desirable traits such as nutritional

value, growth rate, and reproductive efficiency. Recent research on *G. sigillatus* has identified key environmental factors that influence such traits, including temperature, food stress (Rapkin et al., 2018; Sakaluk et al., 2002), and mating strategies (Bateman & MacFadyen, 1999; Champagnon & Cueva del Castillo, 2008, Sakaluk et al., 2019). However, genome-based selection has become central to modern breeding in a wide variety of agricultural species, resulting in rapid and efficient improvement of specific traits (Johnsson, 2023). Beyond their economic importance, field crickets – including *G. sigillatus* - have long been a study organism in evolutionary and behavioural ecology (Sakaluk et al., 2019). Males of this polygynandrous species produce a costly nuptial gift – the spermatophylax – which they transfer to females upon mating, making *G. sigillatus* a prominent research model for examining mating system evolution, mate choice, and sexual selection (Ivy & Sakaluk, 2005; Sakaluk, 1984, 1991). A high-quality genome assembly serves as a foundational resource for such efforts, enabling researchers to pinpoint trait-associated genetic variants, unravel functional pathways governing those phenotypes, and accelerate both selective breeding efforts and fundamental research in behavioural and evolutionary biology (Sinha et al., 2021).

In this study, we produced a high-quality, annotated, chromosome-level reference genome for *G. sigillatus* by integrating long-read PacBio HiFi data with high-throughput chromosome conformation capture (Hi-C) data. Using comparative genomic approaches, we refined the phylogenetic placement and evolutionary history of *G. sigillatus*. We identified candidate gene families linked to economically significant traits that underwent changes during the divergence of *G. sigillatus* from sister taxa and illuminate gene family expansions and contractions relevant to the unique biology of this species. This research provides a foundational resource for the genetic analysis and farming of *G. sigillatus*, offering valuable insights into its genomic features and agriculturally important traits, as well as for basic research on the evolution of mating systems and reproductive behaviour.

Methods

Sampling and DNA sequencing

For whole genomic sequencing, an individual *Gryllodes sigillatus* female was taken from an inbred laboratory stock line descended from approximately 500 adults collected at Las Cruces, New Mexico, in 2001 (Ivy & Sakaluk, 2005). We extracted high molecular weight genomic DNA from liquid nitrogen flash-frozen cricket legs. Following an initial wash with 1% PBS buffer, DNA was extracted using the Macherey-Nagel™ NucleoBond™ High Molecular

91 Weight DNA Kit. DNA quality and concentration were assessed with both Nanodrop and Qubit.
92 Sequencing library preparation and whole genome sequencing were performed using the
93 PacBio SMRT® system by BGI Genomics, resulting in 96.25 Gb of HiFi reads with a coverage
94 depth of approximately 48×. Additionally, flash-frozen muscle tissue was used to prepare and
95 sequence a Hi-C library, also by BGI Genomics, resulting in 151.48 Gb of ca. 72× Hi-C reads
96 (Table S1).
97
98 **Bulk RNA-seq**
99 RNA-seq data used to facilitate genome annotation was generated using two different methods.
100 This was done so as to include reads from as wide a range of tissues as possible to increase the
101 completeness of the genome annotation. We used data from whole bodies and accessory glands
102 of males, plus whole bodies, ovaries, and heads of females (Table S2).
103
104 Whole body RNA-seq data was generated from laboratory stock crickets as follows. These
105 were descended from 500 adult *G. sigillatus* collected in Las Cruces, New Mexico (USA) in
106 2001 (Ivy & Sakaluk, 2005). Rearing followed previously described methods (Duffield et al.,
107 2021, 2022). Approximately 500 crickets were housed in ventilated, 55 L plastic storage bins
108 packed with egg carton. Crickets were given *ad libitum* food (approximately equal parts Purina
109 Cat Chow Complete Cat Food and Teklad Global 18% Protein Extruded Rodent Diet pellets)
110 and water (glass vials plugged with moist cotton). Colonies were housed in an environmental
111 chamber at 32 °C on a 16:8 hour light:dark cycle. Experimental crickets were frozen and stored
112 at -80 °C at 7 days post-eclosion and shipped on dry ice to LC Sciences for (Houston, TX) for
113 further processing. Total RNA was extracted using Trizol (Invitrogen, USA) following the
114 manufacturer's procedure. Total RNA quality and quantity were analysed using a Bioanalyzer
115 2100 and RNA 6000 Nano LabChip Kit (Agilent, USA), with samples obtaining RIN
116 number >7.0. Then, mRNA was purified in two rounds using Dynabeads Oligo (dT) (Thermo
117 Fisher, USA) and subsequently fragmented with a Magnesium RNA Fragmentation Module
118 (New England Biolabs, USA) at 94℃ for 5-7min. The cleaved RNA fragments were reverse-
119 transcribed to cDNA using SuperScript™ II Reverse Transcriptase (Invitrogen, USA) and then
120 used to synthesize U-labelled second-stranded DNAs with *E. coli* DNA polymerase I (New
121 England Biolabs, USA), RNase H (New England Biolabs, USA), and dUTP Solution (Thermo
122 Fisher, USA). Dual-index adapters were ligated to the fragments, and size selection was
123 performed with AMPureXP beads. After the heat-labile UDG enzyme (New England Biolabs,

124 USA) treatment of the U-labelled second-stranded DNAs, the ligated products were amplified
125 with PCR by under the following conditions: initial denaturation at 95°C for 3 min; 8 cycles of
126 denaturation at 98°C for 15 sec, annealing at 60°C for 15 sec, and extension at 72°C for 30 sec;
127 and then final extension at 72°C for 5 min. The average insert size for the final cDNA libraries
128 was 300±50 bp. Sequencing was performed using 2×150bp paired-end sequencing (PE150) on
129 an Illumina Novaseq™ 6000 following the vendor's recommended protocol (Table S2).
130
131 RNA-seq data for male accessory glands, female ovaries, and female heads were generated
132 from a lab-maintained line (Burns-Dunn et al., 2024; McKermitt et al., 2024), that originated
133 from the same crickets collected in Las Cruces, 2001 (Ivy & Sakaluk, 2005). Approximately
134 300 adults were housed in ventilated 62.5 L plastic storage bins packed with egg carton.
135 Nymphs were maintained in 6 L bins. Crickets were given *ad libitum* food (approximately
136 equal parts Purina Cat Chow Complete Cat Food and Teklad Global 18% Protein Extruded
137 Rodent Diet pellets) and water (glass vials plugged with moist cotton). Colonies were
138 maintained at a constant temperature (30 ± 2 °C) and under a reversed photoperiod (14:10 hr
139 light:dark cycle). Seven-day old crickets were mated, and two hours later they were stunned on
140 ice for tissue dissections. Each tissue was immediately snap frozen in liquid nitrogen and stored
141 -80 °C until further processing. RNA extractions were performed as described in Foquet et al.
142 (2023). RNA was extracted using a Trizol (Thermo Fisher Scientific) and 1-bromo-3-
143 chloropropane (BCP, Acros Organics) phase separation using standard procedures, followed
144 by a DNAse treatment using a TURBO DNA-free kit (Thermo Fisher Scientific). RNA
145 qualities and concentration were measured on a MultiSkan GO microplate spectrophotometer
146 with a µDrop adapter plate (ThermoScientific). Only RNA with final 260/230 and 260/280
147 values higher than two were used for RNA sequencing. RNA samples were sent to Novogene
148 for library preparation, sequencing, and read pre-processing. Sample quality was assessed with
149 a Bioanalyzer 2100 (Agilent, USA), with samples obtaining RIN number >5.0 used for
150 sequencing. RNA was processed with an Illumina TruSeq Stranded mRNA Kit with polyA
151 enrichment**,** and sequencing of paired end 150bp reads on a NovaSeq were performed by
152 Novogene.
153
154 **Genome assembly**
155 We used HiFiasm v0.19.5-r587 (Cheng et al., 2021), with default settings to generate a draft
156 assembly of *G. sigillatus*. The primary assembly was retained for subsequent refinement and

157 analysis. To eliminate duplicates within the contig-level assembly, Purge_Dups v1.2.6 (Guan
158 et al., 2020) was employed. This identified redundant contigs by mapping the original HiFi
159 reads back to the assembly, removing repetitive or duplicated sequences that could inflate
160 assembly size and misrepresent genomic content. To remove mitochondrial sequences from
161 the draft assembly, MitoHiFi v3.2.1 (Uliano-Silva et al., 2023) was used. This tool identified
162 and removed contigs associated with the mitochondrial genome by comparing them against the
163 published *G. sigillatus* mitochondrial sequence (MW365703.1, Yang et al., 2021), ensuring
164 that only nuclear genomic sequences remained in the primary assembly.
165
166 Hi-C data were then incorporated to scaffold the contigs and achieve chromosome-level
167 assembly following Arima Hi-C mapping guidelines
168 (https://github.com/ArimaGenomics/mapping_pipeline). Briefly, this involved mapping with
169 BWA-MEM2 v2.0pre2 (Vasimuddin et al., 2019) and removing PCR duplicates using Picard
170 v3.1.0 (https://broadinstitute.github.io/picard). All Hi-C reads were mapped to the contig-level
171 assembly, providing spatial genomic interaction information. YahS v1.1a-r3 (Zhou et al., 2023)
172 was used to scaffold contigs based on these mapping results, integrating long-range contact
173 information to order and orient contigs accurately. To further refine scaffolding, we used Juicer
174 tools (https://github.com/aidenlab/Juicebox) and Juicebox v3.0.0
175 (https://github.com/aidenlab/JuicerTools) to visualise the Hi-C contact heatmap and then
176 manually inspect and correct misassembled regions.
177
178 Once the final scaffolds were generated, BlobTools2 (Challis et al., 2020) was used to screen
179 for potential contaminant sequences originating from other species. This step identified 23
180 scaffolds that likely represented contamination, and these were removed to produce the final,
181 cleaned scaffold-level genome assembly. This comprehensive, iterative workflow maximised
182 the likelihood of obtaining an accurate, high-quality chromosome-level *G. sigillatus* genome
183 assembly suitable for further genomic analyses and annotation.
184
185 **Genome quality assessment**
186 Genome assembly statistics, including total length, N50, and GC content, were calculated using
187 SeqKit v2.4.0 (Shen et al., 2016) to provide a comprehensive overview of the assembly quality.
188 To assess the completeness of the assembled genome, BUSCO v5.4.6 (Manni et al., 2021) was
189 employed, aligning the nucleotide sequences against the Arthropoda dataset (Odb10). This

190 analysis enabled the evaluation of the assembly's coverage and accuracy based on the presence
191 of conserved single-copy orthologs representative of the Arthropoda lineage.
192
193 To evaluate the integrity and accuracy of the genome assembly, we employed the k-mer-based
194 method Merqury v1.3 (Rhie et al., 2020). This approach involves mapping k-mers derived from
195 raw HiFi reads to the contig-level assembly, enabling a detailed assessment of assembly quality.
196 Specifically, Merqury calculates the Quality Value (QV) score, which serves as a metric for
197 the overall accuracy of the assembly and estimates the error rate at the level of individual bases.
198
199 For evaluating the completeness of the protein-coding gene annotation (see below), BUSCO
200 was again utilised with the Insecta dataset in 'transcriptome' mode, which is designed to
201 analyse transcriptome assemblies. This allowed for a thorough assessment of gene prediction
202 quality, ensuring that the annotated protein-coding genes were representative and complete. By
203 using both genome and transcriptome completeness assessments, we ensured a robust
204 evaluation of the assembly's accuracy and the integrity of its functional annotations.
205
206 **Genome annotation**
207 To identify repetitive elements in the genome sequence, three sequential methods were
208 employed to ensure thorough coverage. First, Tandem Repeats Finder v4.09 (TRF; Benson,
209 1999) was used to detect simple sequence repeats. The repetitive regions identified by TRF
210 were masked with 'N' in the genome sequence, effectively obscuring these regions for
211 downstream analyses. Next, transposable elements (TEs) within the genome were identified
212 using RepeatMasker v4.1.2-p1 (Tarailo-Graovac & Chen, 2009), with the previously masked
213 genome sequence serving as input. During this step, the genomic sequences were aligned
214 against the Repbase database v20181026 (Bao et al., 2015) to specifically search for TEs
215 homologous to the Arthropoda lineage. The genome was again subjected to hard-masking
216 based on the identified TEs. Lastly, RepeatModeler2 (Flynn et al., 2020) was applied to conduct
217 a self-alignment of the genomic sequences to uncover any additional repetitive elements. The
218 results from RepeatModeler2 were incorporated into a custom repeat database, and
219 RepeatMasker was run once more to identify repeats based on alignment hits. All repetitive
220 elements identified through this comprehensive protocol were soft-masked in subsequent
221 analyses, preserving their sequence while distinguishing them for further study.
222

For gene prediction, we integrated both homologous protein sequences and RNA-seq data. Protein sequences from *D. melanogaster* (BDGP6.46, Ensembl; Harrison et al., 2024), *G. bimaculatus* (Ylla et al., 2021), *T. occipitalis* (Kataoka et al., 2020) and *T. oceanicus* v2 (Zhang et al., 2024), and reviewed orthologs from Arthropoda were obtained from SwissProt to guide the annotation process. Additionally, RNA-seq data from multiple tissues and both sexes were utilised to facilitate transcriptome-based annotation. Prior to gene prediction, raw RNA-seq reads were quality-trimmed using Fastp v0.23.2 (Chen et al., 2018) and subsequently mapped to the reference genome with HISAT2 v2.2.0 (Kim et al., 2015), ensuring accurate alignment for transcript assembly.

To construct the final gene set, we employed a combined annotation strategy that integrated results from three distinct gene prediction approaches. First, BRAKER3 v3.0 (Gabriel et al., 2024) was used for *ab initio* gene prediction, leveraging both protein and transcript evidence to train and predict gene structures. Second, Miniprot v0.10-r225 (H. Li, 2023) was employed to map homologous protein sequences from closely related species and Arthropoda orthologs to the reference genome, providing potential coordinates and open reading frame (ORF) structures of protein-coding genes. Third, RNA-seq data were assembled using two different tools, StringTie v2.2.1 (Shumate et al., 2022) and Trinity v2.14.0 (Grabherr et al., 2011), to capture transcript diversity. The Trinity-assembled transcriptome was filtered to retain sequences longer than 1 Kb, and redundancy was reduced using CD-HIT v4.8.1 (W. Li & Godzik, 2006). The PASA pipeline v2.5.3 (https://github.com/PASApipeline/PASApipeline) was then utilised to predict gene structures by mapping these transcripts to the reference genome. For the transcriptome assembled by StringTie, potential ORFs were identified using TransDecoder v5.7.1 (https://github.com/TransDecoder/TransDecoder), and coordinates were mapped to the reference genome. Finally, the EvidenceModeler v2.1.0 (Haas et al., 2008) was employed to integrate all prediction results into a unified gene set, while the PASA pipeline refined the final gene models by aligning the transcriptome to predicted gene structures. This comprehensive approach resulted in a high-confidence, annotated gene set for the target genome.

Functional annotation

The functions of protein-coding genes were annotated by mapping the isoform sequences against multiple publicly available functional databases. For Non-redundant (Nr; Sayers et al., 2022) and Swiss-Prot/TrEMBL (Boeckmann et al., 2003), protein sequences were mapped to

257 an Insecta-specific subset using DIAMOND blastp v2.0.14.152 (Buchfink et al., 2021),
258 enabling accurate identification of protein functions within this taxonomic group. To identify
259 and annotate functional domains within the protein sequences, InterProScan (Paysan-Lafosse
260 et al., 2023) was employed, which integrates multiple signature databases for a thorough
261 domain analysis. Additionally, EggNOG mapper (Cantalapiedra et al., 2021) was utilised to
262 conduct EggNOG (Huerta-Cepas et al., 2019) analysis, from which Gene Ontology (GO) terms
263 were extracted to provide insights into the biological processes, molecular functions, and
264 cellular components associated with each protein. Finally, KEGG pathway (Kanehisa et al.,
265 2016) annotations were acquired using the KAAS online server (Moriya et al., 2007), allowing
266 for the identification of biochemical pathways and the functional context of the proteins within
267 broader metabolic networks.
268
269 **Ortholog group identification and phylogenetic analysis**
270 Orthologous protein sequences from six other genomes, including *Anabrus simplex* (NCBI
271 RefSeq accession: GCF_040414725.1 by United States Department of Agriculture), *Gryllus*
272 *longicercus* (Szrajer et al., 2024), *Gryllus bimaculatus* (Ylla et al., 2021), *Teleogryllus*
273 *oceanicus* (Zhang et al., 2024), *Teleogryllus occipitalis* (Kataoka et al., 2020) and *Schistocerca*
274 *americana* (iqSchAmer2.1, https://www.dnazoo.org/assemblies/schistocerca_americana),
275 along with our *G. sigillatus* proteins, were collected for ortholog group identification using
276 OrthoFinder v2.2.5 (Emms & Kelly, 2019). To construct a robust phylogenetic framework, we
277 focused on single-copy orthologs, which were used to build a maximum-likelihood
278 phylogenetic tree, providing insights into the evolutionary relationships among these species.
279 To estimate divergence times, we incorporated reference divergence times from fossil evidence
280 obtained from the TimeTree database (Kumar et al., 2022), ensuring that our analysis was
281 anchored to well-established evolutionary benchmarks. The Bayesian inference tool
282 MCMCTree, part of the PAML v4.10.7 package (Reis & Yang, 2011), was then employed to
283 estimate divergence times across all nodes within the phylogeny using only four-fold sites
284 within all single-copy gene sequences.
285
286 **Gene family analysis**
287 Using the ultrametric tree generated by MCMCTree, we analysed the expanded and contracted
288 gene families across all nodes of the phylogeny with CAFE5 v1.1 (Mendes et al., 2021), a
289 software designed for modelling gene family evolution. To optimise the analysis, we tested a
290 range of discrete gamma rate (k) values from 2 to 6, which account for the variation in

evolutionary rates across gene families. The *k*=4 that yielded the highest maximum likelihood was selected for the final model, ensuring an accurate fit to the data. To identify gene families that experienced significant changes specifically in *G. sigillatus*, we conducted a Likelihood Ratio Test (LRT), which enabled us to distinguish gene families that had undergone statistically significant expansions or contractions.

Statistical analysis and visualisation

Statistical analyses were performed using R v4.0.2. Diagrams were generated using ggplot2 in the tidyverse package (Wickham et al., 2019). The genomic circos plot was generated using shinyCircos v2.0 (Wang et al., 2023).

Results and Discussion

Chromosome-level genome assembly of *G. sigillatus*

We generated a high-quality, chromosome level annotated reference genome for *G. sigillatus* (Figure 1b; Table S3). The completed assembly has a total size of 2.17 Gb and comprises 11 pseudochromosomes (Figure 1c; Table S4), consistent with the known karyotype of *G. sigillatus* (You et al., 2007). Based on the latter karyotype study, the longest scaffold in the assembly was marked as chromosome X and autosomes were named by order of length. The assembly has a scaffold N50 of 200.35 Mb and a maximum scaffold length of 420.78 Mb (Figure 1b, Table S3), and k-mer based quality estimation yielded a QV score of 61.21 and per-base error rate of 7.56×10^{-7} (Table S4), indicating a highly contiguous and complete genome assembly. To further assess the completeness and quality of the assembly, we performed Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis using the insecta_odb10, which yielded a BUSCO recovery score of 99.49% (Table S5).

Genome annotation revealed that repetitive sequences accounted for 48.22% of the genome, including DNA transposons (16.22%), long interspersed nuclear elements (LINEs, 24.13%), short interspersed nuclear elements (SINEs, 1.81%), and long terminal repeat (LTR) retrotransposons (0.07%) (Figure 1b; Table S6). This repeat content is comparable to that observed in related species (Szrajer et al., 2024; Ylla et al., 2021; Zhang et al., 2024), suggesting relatively conserved genomic architecture and evolutionary patterns of transposable elements.

324 We predicted a total of 16,179 protein-coding genes using a combination of *ab initio* prediction
325 and evidence-based methods, incorporating RNA-seq data (Figure 1b; Table S2) and protein
326 homology. Functional annotation was successful for 84.56% of the total predicted genes, using
327 alignment against several public databases including the Non-redundant Protein Sequence
328 Database (Nr), TrEMBL, Gene Ontology (GO), EggNOG, and the Kyoto Encyclopedia of
329 Genes and Genomes (KEGG).

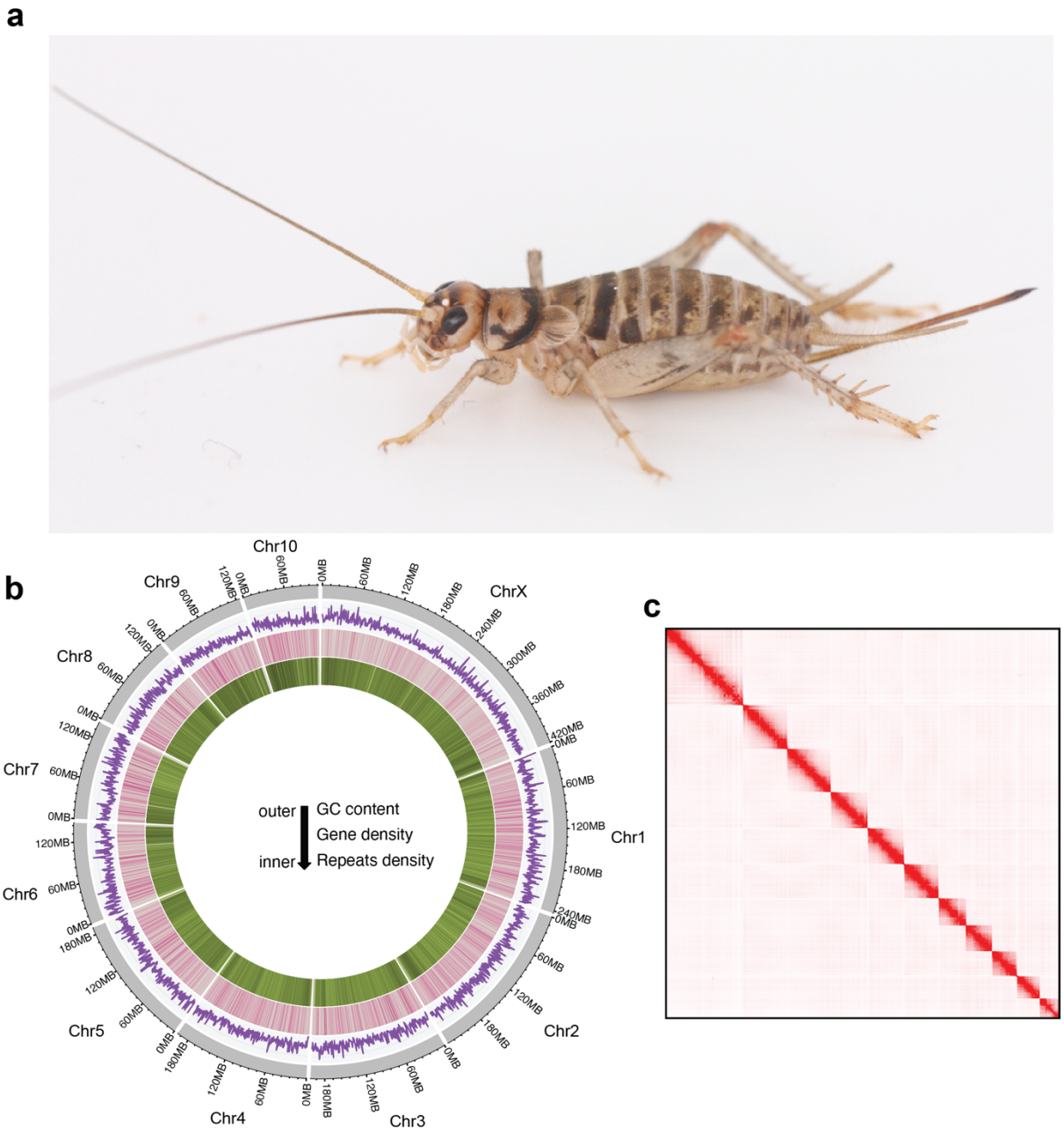


Figure 1. *Gryllodes sigillatus* and its genome characteristics. (a) A female *G. sigillatus*. **(b)** Features of the *G. sigillatus* genome. Outer track (grey) illustrates the 11 chromosomes, with the largest identified as the putative X chromosome on the basis of published karyotypes study (You et al., 2007). Genome position is indicated by the outer numbers. The inner purple track shows GC content, and the inner pink and green tracks show genome-wide gene and repeat density, respectively. **(c)** Hi-C contact map of the genome assembly indicates high contiguity and coverage. **Photo credit**: NW Bailey.

338 **Genome evolution and phylogenetic analysis**

339 Our study provides the first reference genome within the *Gryllodes* genus We also performed
340 a phylogenetic analysis that broadly supports the known evolutionary history of *Gryllodes*
341 (Song et al., 2020). A total of six Ensiferan and one Caeliferan genome (the latter as an
342 outgroup) were selected to identify orthogroups, yielding 15,123 orthogroups of which 4,157
343 comprised single-copy orthologs (Figure S1). A maximum-likelihood species tree was
344 constructed using concatenated alignments of these orthologs, confirming the established
345 evolutionary relationships (Song et al., 2020), validating the inferred speciation timeline and
346 positioning *Gryllodes* within its evolutionary context (Figure 2a). Divergence time estimates
347 indicate that *Gryllodes* diverged from other *Gryllidae* species approximately 141-199 million
348 years ago (Mya).

349

350 Analysis of gene family evolution across orthopterans revealed 377 expanded and 783
351 contracted gene families in *G. sigillatus* (Figure 2a). Focusing on gene families that exhibited
352 significant changes, we identified 28 that were notably expanded in *G. sigillatus* relative to
353 other orthopteran species (Likelihood Ratio Test, $P < 0.01$), consisting of 146 protein-coding
354 genes in *G. sigillatus* (Table S7). Many of these expanded families are linked to histone-related
355 functions. Histones are central to nucleosome formation, which packages DNA within the
356 nucleus and regulates gene expression through chromatin remodelling. In insects, histone
357 modifications influence development (Glastad et al., 2019), stress responses (Gupta & Nair,
358 2024), and genome stability (Glastad et al., 2019), and the expansion of histone-related gene
359 families in *G. sigillatus* may be associated with an enhanced capacity for chromatin
360 remodelling and genome organisation, consistent with adaptations that enable *G. sigillatus* to
361 be resilient in diverse ecological niches.

362

363 Other expanded gene families are associated with metabolism and energy production, including
364 those encoding succinyl-CoA ligase (SCL), NADH dehydrogenase (complex I subunit 1), and
365 trypsin. Succinyl-CoA ligase, a key enzyme in the tricarboxylic acid (TCA) cycle, catalyses
366 the conversion of succinyl-CoA to succinate, producing ATP or GTP through substrate-level
367 phosphorylation. An expansion in SCL genes may enhance TCA cycle efficiency, boosting
368 energy production to meet increased metabolic demands for activities such as locomotion,
369 growth, and reproduction; NADH dehydrogenase, a mitochondrial complex I component,
370 initiates the electron transport chain by transferring electrons from NADH to ubiquinone,
371 driving ATP synthesis via oxidative phosphorylation (Garcia et al., 2017). The expansion of

genes encoding NADH dehydrogenase subunits could increase electron transport efficiency, improving ATP generation. Energy metabolism pathways and their evolution in other systems have been linked to the immunity-metabolism interface (Dolezal et al., 2019), detoxification of and resistance to xenobiotics (Gospodaryov et al., 2020), biological invasion (Wang et al., 2013), dispersal and flight (Yang et al., 2014; Treidel et al., 2023). While flight performance is less relevant in *G. sigillatus*, flight muscles are used for male reproductive signaling which has been shown to be energetically constrained (Houslay et al., 2017). Overall, metabolic enhancement may further provide *G. sigillatus* with greater resilience to environmental fluctuations in energy availability, supporting sustained activity and survival across diverse conditions.

The expansion of the trypsin gene family suggests adaptations in digestive processes. Trypsins are serine proteases involved in protein digestion in the insect midgut, hydrolysing peptide bonds to aid nutrient absorption (Meriño-Cabrera et al., 2022; Muhlia-Almazán et al., 2008). An increased repertoire of trypsin genes may allow *G. sigillatus* to more effectively exploit diverse dietary proteins, supporting rapid growth and high reproductive output. Trypsins are also implicated in immune responses, such as activating the prophenoloxidase system involved in melanisation and pathogen defence (González-Santoyo & Córdoba-Aguilar, 2012; Kanost & Gorman, 2008); thus, their expansion may also enhance immune capabilities in *G. sigillatus*.

Finally, the expansion of the cytochrome P450 (CYP) gene family, known for its roles in detoxification and the immune response, was observed (Feyereisen, 2012). Cytochrome P450 enzymes are involved in the metabolism of endogenous substrates and detoxification of xenobiotics (Nauen et al., 2022), including plant secondary metabolites and synthetic insecticides (Dermauw et al., 2020). Additionally, CYP enzymes contribute to the synthesis and degradation of hormones such as ecdysteroids and juvenile hormones, which regulate development and reproduction. CYP expansion suggests that *G. sigillatus* may possess an enhanced ability to detoxify a broad range of harmful compounds, offering a selective advantage in chemically diverse environments.

These findings highlight potentially distinctive adaptive traits in *G. sigillatus*, shedding light on its unique evolutionary trajectory within Orthoptera. The expansion of gene families related to chromatin remodelling, metabolism, digestion, and detoxification suggests that *G. sigillatus* has evolved genetic adaptations enhancing its physiological functions and ecological resilience,

which is consistent with its global, cosmopolitan distribution in tropical regions and introduction into many non-native habitats. These adaptations likely support rapid responses to environmental stresses, optimise energy use, and strengthen defences against pathogens and toxins, highlighting genetic foundations of traits critical for selective breeding of *G. sigillatus* for optimal reproductive traits in farming applications.

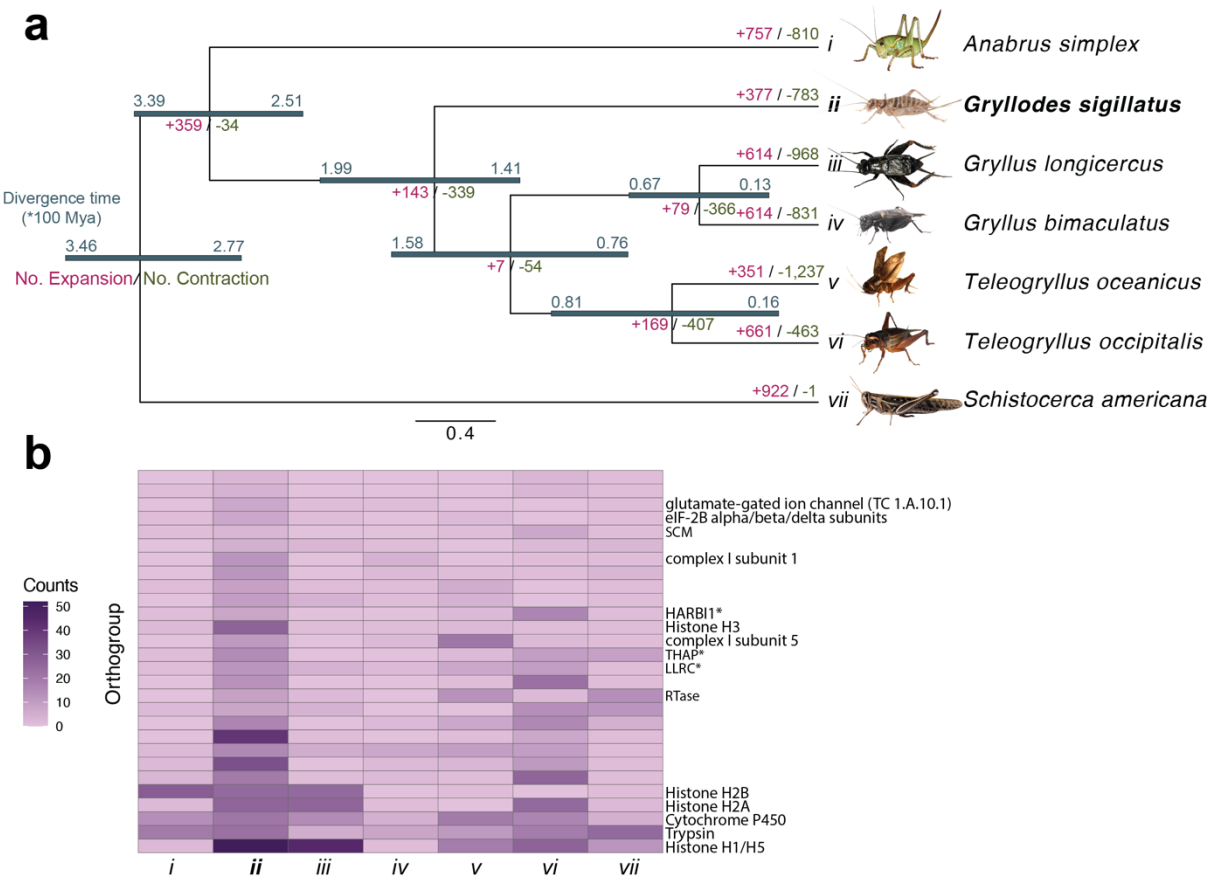


Figure 2. Evolution of gene family expansion and contraction in selected orthopterans. (a) Phylogenetic tree illustrates the divergence times and gene family evolution across seven Orthoptera species. Branch lengths represent evolutionary distances, with upper and lower divergence time estimates (in units of 100 million years ago, Mya) labelled along each branch. Numbers in red and green indicate the number of significantly expanded (+) and contracted (−) gene families along each lineage, respectively. **(b)** Heatmap illustrating counts of gene families that are significantly expanded in *G. sigillatus* (*ii*), with darker shades representing higher counts. Gene families marked with an asterisk (*) represent those that failed to match a verified family name across Insecta but have homologous sequences in human gene families, indicating possible highly conserved functions. **Photo credits:** *Anabrus simplex*, *Gryllodes sigillatus*, *Gryllus bimaculatus* and *Teleogryllus oceanicus* by NW Bailey; *Gryllus longicercus* by Jeff Hollenbeck, modified by Shangzhe Zhang, licensed under CC BY-ND-NC 1.0; *Teleogryllus occipitalis* by Takaaki Hattori, modified by Shangzhe Zhang, licensed under CC-BY 4.0; *Schistocerca americana* by Brandon Woo with permission granted.

Conclusion

There are over 6,000 species of crickets in the infraorder Gryllidea, over 60 of which are known to be consumed and, in some cases, farmed by humans (Cigliano et al., 2018; Magara et al., 2021). Generating a high-quality chromosome-scale assembly of one of the most globally widespread and commonly-farmed species, *Gryllodes sigillatus*, is a key step in promoting the adoption of genome-informed farming improvements for insects. The widespread farming of *G. sigillatus* is likely driven by their worldwide, cosmopolitan distribution and ease of collection (Weissman et al., 2012). However, there are many cricket species with farming potential, and the resources developed here will facilitate their study and use more broadly (Dossey et al., 2023; Magara et al., 2019, 2021). Recently, genomic resources have increasingly deployed to improve strategies for farming crickets (Kataoka et al., 2022; Nakamura et al., 2022). Comparative and phylogenetic approaches continue to advance our understanding of the evolutionary history of crickets and allied taxa (Gray et al., 2020; Song et al., 2020; Dong et al., 2024), and genomic approaches in these species benefit from a longstanding history of major contributions to fundamental research in evolutionary and behavioural biology. Recent work, for example, has shed light on the genomics of speciation (Xu & Shaw, 2021; Yusuf et al., 2024), adaptation in the wild (Pascoal et al., 2020; Zhang et al., 2021; Zhang et al., 2024; Rayner et al. 2024), evolutionary developmental biology (Barnett et al., 2019; Ohde et al., 2022), and sperm competition, mate choice, and sexual selection (Simmons et al., 2013; Xu & Shaw, 2019). Future research efforts will benefit not only from exploring the potential for *G. sigillatus* in greater depth using genome-informed tools for applied and basic research, but also developing genomic resources across a broader taxonomic sampling of commonly studied Gryllids.

Supplementary information

Supplementary Materials include Figs. S1, Table S1 to S7, Data table S1.

Data availability

All genomic data (long-reads sequencing data, and Hi-C sequencing data) have been deposited at European Nucleotide Archive (ENA) under PRJEB84028. Sequences of the genome assembly is publicly available at ENA under accession GCA_965111905. Annotation files of the genome assembly has been deposited at Zenodo as DOI: 10.5281/zenodo.14617131. All transcriptome data have been deposited at xxx.

Acknowledgements

The authors gratefully acknowledge funding support from the Australian Research Council to JH (DP180101708), the UK Natural Environment Research Council to NWB (NE/W001616/1), the China Scholarship Council to SZ (202106180022), and the United States National Science Foundation to SKS, BMS, and JH (IOS 16–54028). This research was also supported in part by the U.S. Department of Agriculture, Agricultural Research Service (Project Number: 6064-32000-001-000D to KRD and 5010-22410-023-00-D to JLR). We thank the UK Biotechnology and Biological Sciences Research Council (BB/S019669/1 and BB/X019683/1) and James Hutton Institute for access to the UK's Crop Diversity Bioinformatics High Performance Computing Cluster, with particularly acknowledgement of Ian Milne's inexhaustible patience as sysadmin. Lianne Baker assisted with sample shipping and Tanya Sneddon with DNA extractions. The authors are grateful for constructive feedback on the manuscript from Xin Du. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

Conflict of interest statement

The authors declare no conflict of interest.

Author contributions

Conceived the study: JH, NWB. Designed experiments: all authors. Secured funding: JH, BMS, JLR, KRD, SKS, NWB. Performed experimental work: KRD, BF, SZ. Analysed data: KRD, BF, SZ. Led manuscript writing: SZ, NWB. Contributed to writing: JH, BF, SKS, BMS, JLR, KRD.

References

Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, *6*(1), 11. https://doi.org/10.1186/s13100-015-0041-9

Barnett, A. A., Nakamura, T., Extavour, C. G. (2019). Hox genes limit germ cell formation in the short germ insect *Gryllus bimaculatus*. *Proceedings of the National Academy of Sciences, U.S.A. 116* (33), 16430-16435. https://doi.org/10.1073/pnas.1816024116

Bateman, P. W., & MacFadyen, D. N. (1999). Mate Guarding in the Cricket Gryllodes sigillatus: Influence of Multiple Potential Partners. *Ethology*, *105*(11), 949–957. https://doi.org/10.1046/j.1439-0310.1999.00484.x

Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research*, *27*(2), 573–580. https://doi.org/10.1093/nar/27.2.573

Blankers, T., Oh, K. P., Bombarely, A., Shaw, K. L. (2018) The genomic architecture of a rapid island radiation: recombination rate variation, chromosome structure, and genome assembly of the Hawaiian cricket *Laupala. Genetics 209*(4), 1329-1344. https://doi.org/10.1534/genetics.118.300894

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., & Schneider, M. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research*, *31*(1), 365–370. https://doi.org/10.1093/nar/gkg095

Buchfink, B., Reuter, K., & Drost, H.-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, *18*(4), 366–368. https://doi.org/10.1038/s41592-021-01101-x

Burns-Dunn, S., Mortys, T., House, C. M., Mitchell, C., Duffield, K. R., Foquet, B., Sadd, B. M., Sakaluk, S. K., & Hunt, J. (2024). Sexually antagonistic coevolution of the male nuptial gift and female feeding behaviour in decorated crickets. *Proceedings of the Royal Society B: Biological Sciences*, *291*(2026), 20240804. https://doi.org/10.1098/rspb.2024.0804

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, *38*(12), 5825–5829. https://doi.org/10.1093/molbev/msab293

Challis, R., Richards, E., Rajan, J., Cochrane, G., & Blaxter, M. (2020). BlobToolKit – Interactive Quality Assessment of Genome Assemblies. *G3 Genes|Genomes|Genetics*, *10*(4), 1361–1374. https://doi.org/10.1534/g3.119.400908

Champagnon, J., & Cueva del Castillo, R. (2008). Female Mate Choice, Calling Song and Genetic Variance in the Cricket, *Gryllodes sigillatus*. *Ethology*, *114*(3), 223–230. https://doi.org/10.1111/j.1439-0310.2007.01470.x

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, *18*(2), Article 2. https://doi.org/10.1038/s41592-020-01056-5

Cigliano, M., Braun, H., Eades, D., & Otte, D. (2018). *"Infraorder Gryllidea". Orthoptera Species File. Version 5.0/5.0*. Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata (UNLP).

Dermauw, W., Van Leeuwen, T., & Feyereisen, R. (2020). Diversity and evolution of the P450 family in arthropods. *Insect Biochemistry and Molecular Biology*, *127*, 103490. https://doi.org/10.1016/j.ibmb.2020.103490

Dolezal, T., Krejcova, G., Bajgar, A., Nedbalova, P., & Strasser, P. (2019). Molecular regulations of metabolism during immune response in insects. *Insect Biochemistry and Molecular Biology*, *109*, 31–42. https://doi.org/10.1016/j.ibmb.2019.04.005

Dong, J., Liu, Y., Tan, M. K., Wahab, R. A., Nattier, R., Chifflet-Belle, P., & Robillard, T. (2024). Museomics allows comparative analyses of mitochondrial genomes in the family Gryllidae (Insecta, Orthoptera) and confirms its phylogenetic relationships. *PeerJ*, *12*, e17734. https://doi.org/10.7717/peerj.17734

Dossey, A. T., Oppert, B., Chu, F.-C., Lorenzen, M. D., Scheffler, B., Simpson, S., Koren, S., Johnston, J. S., Kataoka, K., & Ide, K. (2023). Genome and Genetic Engineering of the House Cricket (Acheta domesticus): A Resource for Sustainable Agriculture. *Biomolecules*, *13*(4), Article 4. https://doi.org/10.3390/biom13040589

Duffield, K. R., Foquet, B., Stasko, J. A., Hunt, J., Sadd, B. M., Sakaluk, S. K., & Ramirez, J. L. (2022). Induction of Multiple Immune Signaling Pathways in *Gryllodes sigillatus* Crickets during Overt Viral Infections. *Viruses*, *14*(12), Article 12. https://doi.org/10.3390/v14122712

Duffield, K. R., Hunt, J., Sadd, B. M., Sakaluk, S. K., Oppert, B., Rosario, K., Behle, R. W., & Ramirez, J. L. (2021). Active and Covert Infections of Cricket Iridovirus and Acheta domesticus Densovirus in Reared *Gryllodes sigillatus* Crickets. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.780796

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238. https://doi.org/10.1186/s13059-019-1832-y

Feyereisen, R. (2012). 8—Insect CYP Genes and P450 Enzymes. In L. I. Gilbert (Ed.), *Insect Molecular Biology and Biochemistry* (pp. 236–316). Academic Press. https://doi.org/10.1016/B978-0-12-384747-8.10008-X

561 Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F.
562 (2020). RepeatModeler2 for automated genomic discovery of transposable element
563 families. *Proceedings of the National Academy of Sciences*, *117*(17), 9451–9457.
564 https://doi.org/10.1073/pnas.1921046117
565 Foquet, B., Rapkin, J., Sharma, M. D., Sadd, B. M., Sakaluk, S. K., & Hunt, J. (2023).
566 Transcriptomic responses of females to consumption of nuptial food gifts as a potential
567 mediator of sexual conflict in decorated crickets. *Journal of Evolutionary Biology*,
568 *36*(1), 183–194. https://doi.org/10.1111/jeb.14114
569 Gabriel, L., Brůna, T., Hoff, K. J., Ebel, M., Lomsadze, A., Borodovsky, M., & Stanke, M.
570 (2024). BRAKER3: Fully automated genome annotation using RNA-seq and protein
571 evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Research*, *34*(5),
572 769–777. https://doi.org/10.1101/gr.278090.123
573 Garcia, C. J., Khajeh, J., Coulanges, E., Chen, E. I., & Owusu-Ansah, E. (2017). Regulation of
574 Mitochondrial Complex I Biogenesis in *Drosophila* Flight Muscles. *Cell Reports*, *20*(1),
575 264–278. https://doi.org/10.1016/j.celrep.2017.06.015
576 Glastad, K. M., Hunt, B. G., & Goodisman, M. A. D. (2019). Epigenetics in Insects: Genome
577 Regulation and the Generation of Phenotypic Diversity. *Annual Review of Entomology*,
578 *64*, 185–203. https://doi.org/10.1146/annurev-ento-011118-111914
579 González-Santoyo, I., & Córdoba-Aguilar, A. (2012). Phenoloxidase: A key component of the
580 insect immune system. *Entomologia Experimentalis et Applicata*, *142*(1), 1–16.
581 https://doi.org/10.1111/j.1570-7458.2011.01187.x
582 Gospodaryov, D. V., Strilbytska, O. M., Semaniuk, U. V., Perkhulyn, N. V., Rovenko, B. M.,
583 Yurkevych, I. S., Barata, A. G., Dick, T. P., Lushchak, O. V., & Jacobs, H. T. (2020).
584 Alternative NADH dehydrogenase extends lifespan and increases resistance to
585 xenobiotics in *Drosophila*. *Biogerontology*, *21*(2), 155–171.
586 https://doi.org/10.1007/s10522-019-09849-8
587 Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis,
588 X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke,
589 A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev,
590 A. (2011). Trinity: Reconstructing a full-length transcriptome without a genome from
591 RNA-Seq data. *Nature Biotechnology*, *29*(7), 644–652.
592 https://doi.org/10.1038/nbt.1883
593 Grabowski, N. T., Abdulmawjood, A., Acheuk, F., Barragán Fonseca, K., Chhay, T., Costa
594 Neto, E. M., Ferri, M., Franco Olivas, J., González Aguilar, D. G., Keo, S.,

Lertpatarakomol, R., Miech, P., Piofczyk, T., Proscia, F., Mitchaothai, J., Guerfali, M. M., Sayed, W., Tchibozo, S., & Plötz, M. (2022). Review: Insects—A Source of Safe and Sustainable Food?— "Jein" (Yes and No). *Frontiers in Sustainable Food Systems*, *5*. https://doi.org/10.3389/fsufs.2021.701797

Gray, D. A., Weissman, D. B., Cole, J. A., Lemmon, E. M., & Lemmon, A. R. (2020). Multilocus phylogeny of Gryllus field crickets (Orthoptera: Gryllidae: Gryllinae) utilizing anchored hybrid enrichment. *Zootaxa*, *4750*(3), Article 3. https://doi.org/10.11646/zootaxa.4750.3.2

Guan, D., McCarthy, S. A., Wood, J., Howe, K., Wang, Y., & Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, *36*(9), 2896–2898. https://doi.org/10.1093/bioinformatics/btaa025

Gupta, A., & Nair, S. (2024). Epigenetic processes in insect adaptation to environmental stress. *Current Opinion in Insect Science*, 101294. https://doi.org/10.1016/j.cois.2024.101294

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C. R., & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology*, *9*(1), R7. https://doi.org/10.1186/gb-2008-9-1-r7

Harrison, P. W., Amode, M. R., Austine-Orimoloye, O., Azov, A. G., Barba, M., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., Bhurji, S. K., Boddu, S., Branco Lins, P. R., Brooks, L., Ramaraju, S. B., Campbell, L. I., Martinez, M. C., Charkhchi, M., Chougule, K., … Yates, A. D. (2024). Ensembl 2024. *Nucleic Acids Research*, *52*(D1), D891–D899. https://doi.org/10.1093/nar/gkad1049

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*(D1), D309–D314. https://doi.org/10.1093/nar/gky1085

Huis, A. van, Rumpold, B., Maya, C., & Roos, N. (2021). Nutritional Qualities and Enhancement of Edible Insects. *Annual Review of Nutrition*, *41*(Volume 41, 2021), 551–576. https://doi.org/10.1146/annurev-nutr-041520-010856

Ivy, T. M., & Sakaluk, S. K. (2005). Polyandry Promotes Enhanced Offspring Survival in Decorated Crickets. *Evolution*, *59*(1), 152–159. https://doi.org/10.1111/j.0014-3820.2005.tb00902.x

Johnsson, M. (2023). Genomics in animal breeding from the perspectives of matrices and molecules. *Hereditas*, *160*(1), 20. https://doi.org/10.1186/s41065-023-00285-w

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, *44*(D1), D457–D462. https://doi.org/10.1093/nar/gkv1070

Kanost, M. R., & Gorman, M. J. (2008). 4—PHENOLOXIDASES IN INSECT IMMUNITY. In N. E. Beckage (Ed.), *Insect Immunology* (pp. 69–96). Academic Press. https://doi.org/10.1016/B978-012373976-6.50006-9

Kataoka, K., Minei, R., Ide, K., Ogura, A., Takeyama, H., Takeda, M., Suzuki, T., Yura, K., & Asahi, T. (2020). The Draft Genome Dataset of the Asian Cricket Teleogryllus occipitalis for Molecular Research Toward Entomophagy. *Frontiers in Genetics*, *11*. https://doi.org/10.3389/fgene.2020.00470

Kataoka, K., Togawa, Y., Sanno, R., Asahi, T., & Yura, K. (2022). Dissecting cricket genomes for the advancement of entomology and entomophagy. *Biophysical Reviews*, *14*(1), 75–97. https://doi.org/10.1007/s12551-021-00924-4

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), Article 4. https://doi.org/10.1038/nmeth.3317

Kong, J. D., Ritchie, M. W., Vadboncoeur, É., MacMillan, H. A., & Bertram, S. M. (2024). Growth, development, and life history of a mass-reared edible insect, *Gryllodes sigillatus* (Orthoptera: Gryllidae) (p. 2024.10.08.617250). *bioRxiv*. https://doi.org/10.1101/2024.10.08.617250

Kumar, S., Suleski, M., Craig, J. M., Kasprowicz, A. E., Sanderford, M., Li, M., Stecher, G., & Hedges, S. B. (2022). TimeTree 5: An Expanded Resource for Species Divergence Times. *Molecular Biology and Evolution*, *39*(8), msac174. https://doi.org/10.1093/molbev/msac174

Lange, K. W., & Nakamura, Y. (2023). Potential contribution of edible insects to sustainable consumption and production. *Frontiers in Sustainability*, *4*. https://doi.org/10.3389/frsus.2023.1112950

Li, H. (2023). Protein-to-genome alignment with miniprot. *Bioinformatics*, *39*(1), btad014. https://doi.org/10.1093/bioinformatics/btad014

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics (Oxford, England)*, *22*(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Magara, H. J. O., Niassy, S., Ayieko, M. A., Mukundamago, M., Egonyu, J. P., Tanga, C. M., Kimathi, E. K., Ongere, J. O., Fiaboe, K. K. M., Hugel, S., Orinda, M. A., Roos, N., & Ekesi, S. (2021). Edible Crickets (Orthoptera) Around the World: Distribution, Nutritional Value, and Other Benefits—A Review. *Frontiers in Nutrition*, *7*. https://doi.org/10.3389/fnut.2020.537915

Magara, H. J. O., Tanga, C. M., Ayieko, M. A., Hugel, S., Mohamed, S. A., Khamis, F. M., Salifu, D., Niassy, S., Sevgan, S., Fiaboe, K. K. M., Roos, N., & Ekesi, S. (2019). Performance of Newly Described Native Edible Cricket *Scapsipedus icipe* (Orthoptera: Gryllidae) on Various Diets of Relevance for Farming. *Journal of Economic Entomology*, *112*(2), 653–664. https://doi.org/10.1093/jee/toy397

Manni, M., Berkeley, M. R., Seppey, M., & Zdobnov, E. M. (2021). BUSCO: Assessing Genomic Data Quality and Beyond. *Current Protocols*, *1*(12), e323. https://doi.org/10.1002/cpz1.323

McKermitt, J. T., Foquet, B., Kuna, W., Hunt, J., Sadd, B. M., & Sakaluk, S. K. (2024). Experimental evolution under varying sex ratio and behavioral plasticity in response to perceived competitive environment independently affect calling effort in male crickets. *Evolution*, *78*(3), 453–462. https://doi.org/10.1093/evolut/qpad224

Mendes, F. K., Vanderpool, D., Fulton, B., & Hahn, M. W. (2021). CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics*, *36*(22–23), 5516–5518. https://doi.org/10.1093/bioinformatics/btaa1022

Meriño-Cabrera, Y. B., Oliveira, M. G. de A., Meriño-Cabrera, Y. B., & Oliveira, M. G. de A. (2022). Trypsins: Structural Characterization and Inhibition Focus in Insects. In *Hydrolases*. IntechOpen. https://doi.org/10.5772/intechopen.102632

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., & Kanehisa, M. (2007). KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*, *35*(suppl_2), W182–W185. https://doi.org/10.1093/nar/gkm321

Muhlia-Almazán, A., Sánchez-Paz, A., & García-Carreño, F. L. (2008). Invertebrate trypsins: A review. *Journal of Comparative Physiology B*, *178*(6), 655–672. https://doi.org/10.1007/s00360-008-0263-y

Nakamura, T., Ylla, G., & Extavour, C. G. (2022). Genomics and genome editing techniques of crickets, an emerging model insect for biology and food science. *Current Opinion in Insect Science*, *50*, 100881. https://doi.org/10.1016/j.cois.2022.100881

Nauen, R., Bass, C., Feyereisen, R., & Vontas, J. (2022). The Role of Cytochrome P450s in Insect Toxicology and Resistance. *Annual Review of Entomology*, *67*(Volume 67, 2022), 105–124. https://doi.org/10.1146/annurev-ento-070621-061328

Ohde, T., Mito, T., & Niimi, T. (2022). A hemimetabolous wing development suggests the wing origin from lateral tergum of a wingless ancestor. *Nature Communications*, *13*(1), 979. https://doi.org/10.1038/s41467-022-28624-x

Pascoal, S., Risse, J. E., Zhang, X., Blaxter, M., Cezard, T., Challis, R. J., Gharbi, K., Hunt, J., Kumar, S., Langan, E., Liu, X., Rayner, J. G., Ritchie, M. G., Snoek, B. L., Trivedi, U., Bailey, N. W. (2020). Field cricket genome reveals the footprint of recent, abrupt adaptation in the wild. *Evolution Letters 4*(1), 19-33. https://doi.org/10.1002/evl3.148

Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunić, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Rivoire, C., … Bateman, A. (2023). InterPro in 2022. *Nucleic Acids Research*, *51*(D1), D418–D427. https://doi.org/10.1093/nar/gkac993

Rapkin, J., Jensen, K., Archer, C. R., House, C. M., Sakaluk, S. K., Castillo, E. del, & Hunt, J. (2018). The Geometry of Nutrient Space–Based Life-History Trade-Offs: Sex-Specific Effects of Macronutrient Intake on the Trade-Off between Encapsulation Ability and Reproductive Effort in Decorated Crickets. *The American Naturalist*, *191*(4), 452–474. https://doi.org/10.1086/696147

Rayner, J. G., Eichenberger, F., Bainbridge, J. V. A., Zhang, S., Zhang, X., Yusuf, L. H. Balenger, S., Gaggiotti, O. E., Bailey, N. W. (2024). Competing adaptations maintain nonadaptive variation in a wild cricket population. *Proceedings of the National Academy of Sciences, U.S.A.* *121*(32), e2317879121. https://doi.org/10.1073/pnas.2317879121

Reis, M. dos, & Yang, Z. (2011). Approximate Likelihood Calculation on a Phylogeny for Bayesian Estimation of Divergence Times. *Molecular Biology and Evolution*, *28*(7), 2161–2172. https://doi.org/10.1093/molbev/msr045

Rhie, A., Walenz, B. P., Koren, S., & Phillippy, A. M. (2020). Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*, *21*(1), 245. https://doi.org/10.1186/s13059-020-02134-9

Sakaluk, S. K. (1984). Male Crickets Feed Females to Ensure Complete Sperm Transfer. *Science*, *223*(4636), 609–610. https://doi.org/10.1126/science.223.4636.609

Sakaluk, S. K. (1991). Post-copulatory mate guarding in decorated crickets. *Animal Behaviour*, *41*(2), 207–216. https://doi.org/10.1016/S0003-3472(05)80472-5

Sakaluk, S. K., Duffield, K. R., Rapkin, J., Sadd, B. M., & Hunt, J. (2019). The troublesome gift: The spermatophylax as a purveyor of sexual conflict and coercion in crickets. In M. Naguib, L. Barrett, S. D. Healy, J. Podos, L. W. Simmons, & M. Zuk (Eds.), *Advances in the Study of Behavior. Volume 51* (pp. 1–30). Academic Press. https://doi.org/10.1016/bs.asb.2018.12.001

Sakaluk, S. K., Schaus, J. M., Eggert, A., Snedden, W. A., & Brady, P. L. (2002). POLYANDRY AND FITNESS OF OFFSPRING REARED UNDER VARYING NUTRITIONAL STRESS IN DECORATED CRICKETS. *Evolution*, *56*(10), 1999–2007. https://doi.org/10.1111/j.0014-3820.2002.tb00126.x

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., … Sherry, S. T. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, *50*(D1), D20–D26. https://doi.org/10.1093/nar/gkab1112

Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLOS ONE*, *11*(10), e0163962. https://doi.org/10.1371/journal.pone.0163962

Shumate, A., Wong, B., Pertea, G., & Pertea, M. (2022). Improved transcriptome assembly using a hybrid of long and short reads with StringTie. *PLOS Computational Biology*, *18*(6), e1009730. https://doi.org/10.1371/journal.pcbi.1009730

Simmons, L. W., Tan, Y.-F., & Millar, A. H. (2013). Sperm and seminal fluid proteomes of the field cricket eleogryllus oceanicus: Identification of novel proteins transferred to females at mating. *Insect Molecular Biology*, *22*(1), 115–130. https://doi.org/10.1111/imb.12007

Sinha, P., Singh, V. K., Bohra, A., Kumar, A., Reif, J. C., & Varshney, R. K. (2021). Genomics and breeding innovations for enhancing genetic gain for climate resilience and nutrition traits. *Theoretical and Applied Genetics*, *134*(6), 1829–1843. https://doi.org/10.1007/s00122-021-03847-6

Song, H., Béthoux, O., Shin, S., Donath, A., Letsch, H., Liu, S., McKenna, D. D., Meng, G., Misof, B., Podsiadlowski, L., Zhou, X., Wipfler, B., & Simon, S. (2020). Phylogenomic analysis sheds light on the evolutionary pathways towards acoustic communication in

Orthoptera. *Nature Communications*, *11*(1), 4939. https://doi.org/10.1038/s41467-020-18739-4

Szrajer, S., Gray, D., & Ylla, G. (2024). The genome assembly and annotation of the cricket *Gryllus longicercus*. *Scientific Data*, *11*(1), 708. https://doi.org/10.1038/s41597-024-03554-z

Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, *Chapter 4*, 4.10.1-4.10.14. https://doi.org/10.1002/0471250953.bi0410s25

Treidel, L. A., Goswami, P., & Williams, C. M. (2023). Changes in mitochondrial function parallel life history transitions between flight and reproduction in wing polymorphic field crickets. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *324*(6), R735–R746. https://doi.org/10.1152/ajpregu.00191.2022

Uliano-Silva, M., Ferreira, J. G. R. N., Krasheninnikova, K., Blaxter, M., Mieszkowska, N., Hall, N., Holland, P., Durbin, R., Richards, T., Kersey, P., Hollingsworth, P., Wilson, W., Twyford, A., Gaya, E., Lawniczak, M., Lewis, O., Broad, G., Martin, F., Hart, M., … Darwin Tree of Life Consortium. (2023). MitoHiFi: A python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. *BMC Bioinformatics*, *24*(1), 288. https://doi.org/10.1186/s12859-023-05385-y

van Huis, A., & Oonincx, D. G. A. B. (2017). The environmental sustainability of insects as food and feed. A review. *Agronomy for Sustainable Development*, *37*(5), 43. https://doi.org/10.1007/s13593-017-0452-8

Vasimuddin, Md., Misra, S., Li, H., & Aluru, S. (2019). Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 314–324. https://doi.org/10.1109/IPDPS.2019.00041

Wang, Y., Jia, L., Tian, G., Dong, Y., Zhang, X., Zhou, Z., Luo, X., Li, Y., & Yao, W. (2023). shinyCircos-V2.0: Leveraging the creation of Circos plot with enhanced usability and advanced features. *iMeta*, *2*(2), e109. https://doi.org/10.1002/imt2.109

Wang, Y.-L., Wang, Y.-J., Luan, J.-B., Yan, G.-H., Liu, S.-S., & Wang, X.-W. (2013). Analysis of the transcriptional differences between indigenous and invasive whiteflies reveals possible mechanisms of whitefly invasion. *PLoS ONE*, *8*(5), e62176. https://doi.org/10.1371/journal.pone.0062176

Weissman, D. B., Gray, D. A., Pham, H. T., & Tijssen, P. (2012). Billions and billions sold: Pet-feeder crickets (Orthoptera: Gryllidae), commercial cricket farms, an epizootic densovirus, and government regulations make for a potential disaster. *Zootaxa*, *3504*(1), Article 1. https://doi.org/10.11646/zootaxa.3504.1.3

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Xu, M. & Shaw, K. L. (2019). Genetic coupling of signal and preference facilitates sexual isolation during rapid speciation. *Proceedings of the Royal Society of London, B 286* (1913), 20191607. https://doi.org/10.1098/rspb.2019.1607

Xu, M., & Shaw, K. L. (2021). Extensive Linkage and Genetic Coupling of Song and Preference Loci Underlying Rapid Speciation in Laupala Crickets. *Journal of Heredity*, *112*(2), 204–213. https://doi.org/10.1093/jhered/esab001

Yang, J., Dong, H., He, M., & Gao, J. (2021). Mitochondrial genome characterization of *Gryllodes sigillatus* (Orthoptera: Gryllidae) and its phylogenetic implications. *Mitochondrial DNA Part B*, *6*(3), 1056–1058. https://doi.org/10.1080/23802359.2021.1899078

Yang, Y., Xu, S., Xu, J., Guo, Y., & Yang, G. (2014). Adaptive Evolution of Mitochondrial Energy Metabolism Genes Associated with Increased Energy Demand in Flying Insects. *PLOS ONE*, *9*(6), e99120. https://doi.org/10.1371/journal.pone.0099120

Ylla, G., Nakamura, T., Itoh, T., Kajitani, R., Toyoda, A., Tomonari, S., Bando, T., Ishimaru, Y., Watanabe, T., Fuketa, M., Matsuoka, Y., Barnett, A. A., Noji, S., Mito, T., & Extavour, C. G. (2021). Insights into the genomic evolution of insects from cricket genomes. *Communications Biology*, *4*(1), Article 1. https://doi.org/10.1038/s42003-021-02197-9

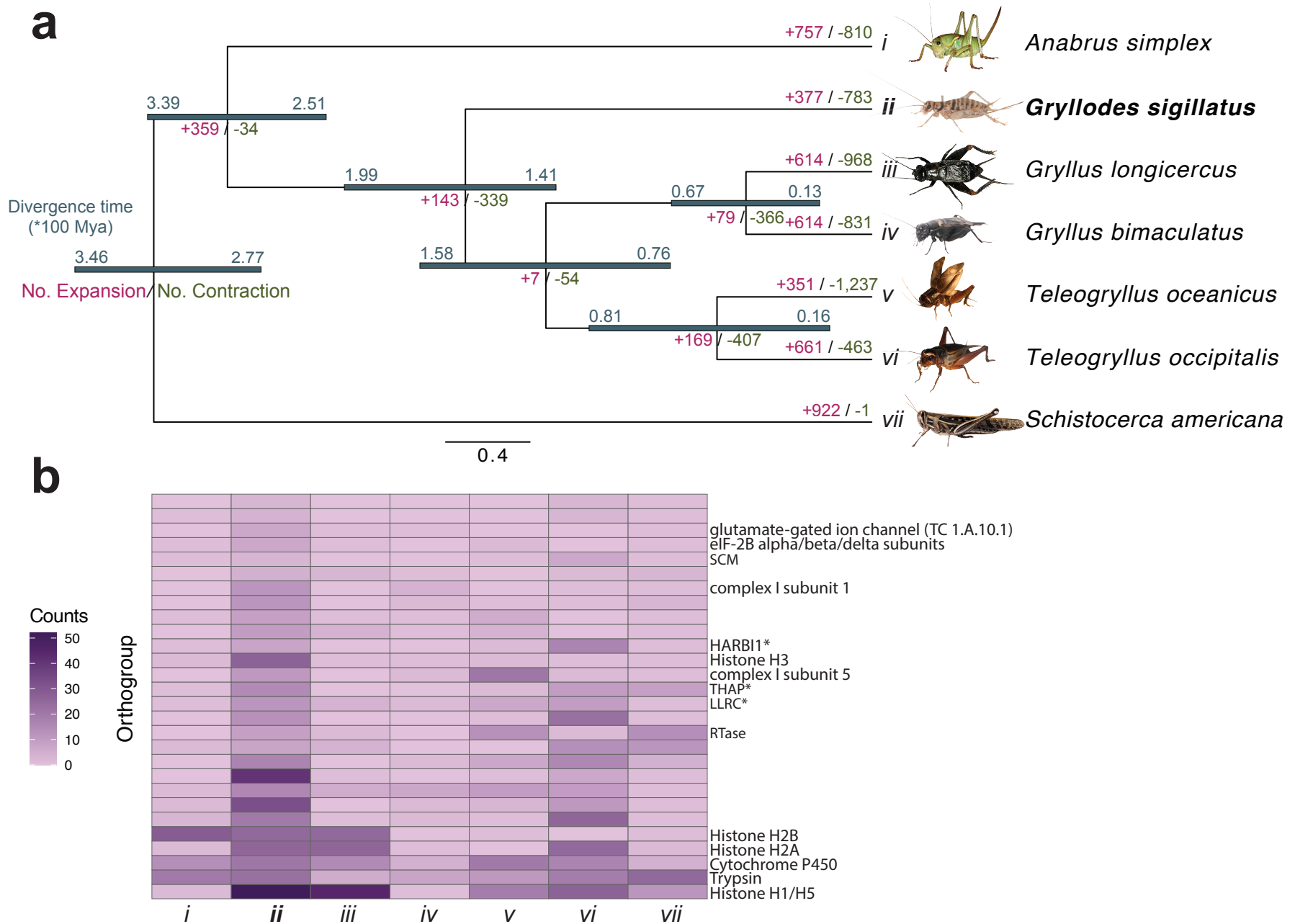
You, J.-X., Li, N., Ren, B.-Z., Yuan, H.-B., & Shang, L.-N. (2007). The study on karyotypes of five Grylloidea species (Orthoptera: Grylloidea) in Northeast China. *Zootaxa*, *1611*(1), Article 1. https://doi.org/10.11646/zootaxa.1611.1.5

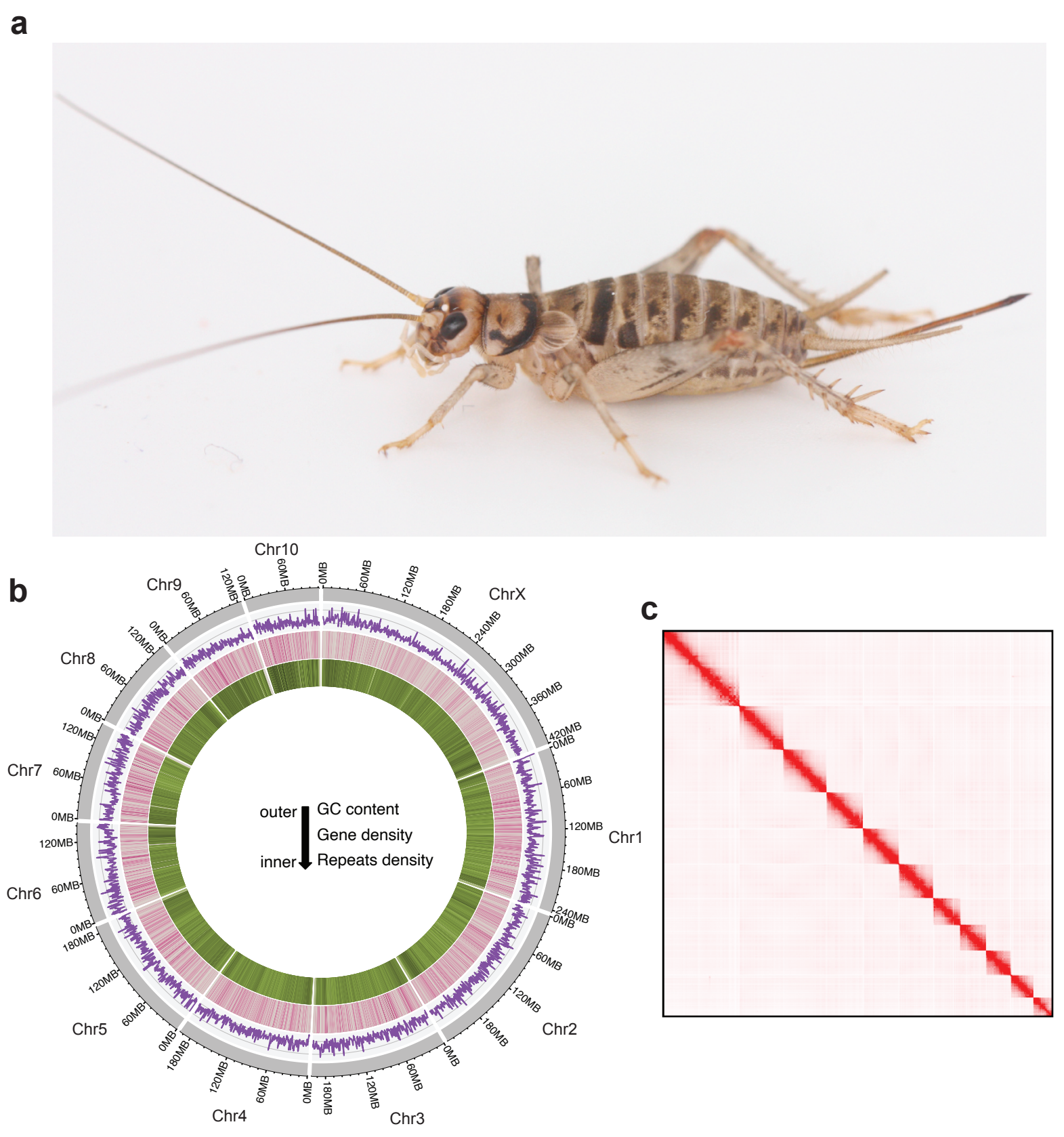
Yusuf, L. H., Pascoal, S., Moran, P. A., & Bailey, N. W. (2024). Testing the genomic overlap between intraspecific mating traits and interspecific mating barriers. *Evolution Letters*, *8*(6), 902–915. https://doi.org/10.1093/evlett/qrae042

Zhang, X., Rayner, J. G., Blaxter, M., & Bailey, N. W. (2021) Rapid parallel adaptation despite gene flow in silent crickets. *Nature Communications* *12*(1), 50. https://doi.org/10.1038/s41467-020-20263-4.

Zhang, X., Blaxter, M., Wood, J. M. D., Tracey, A., McCarthy, S., Thorpe, P., Rayner, J. G., Zhang, S., Sikkink, K. L., Balenger, S. L., & Bailey, N. W. (2024). Temporal genomics in Hawaiian crickets reveals compensatory intragenomic coadaptation during adaptive evolution. *Nature Communications*, *15*(1), 5001. https://doi.org/10.1038/s41467-024-49344-4

Zhou, C., McCarthy, S. A., & Durbin, R. (2023). YaHS: Yet another Hi-C scaffolding tool. *Bioinformatics*, *39*(1), btac808. https://doi.org/10.1093/bioinformatics/btac808


a
Divergence time
(*100 Mya)
No. Expansion/No. Contraction
3.46
2.77
3.39
2.51
+359 / -34
1.99
1.41
+143 / -339
1.58
0.76
+7 / -54
0.67
0.13
+79 / -366
0.81
0.16
+169 / -407
+757 / -810
+377 / -783
+614 / -968
+614 / -831
+351 / -1,237
+661 / -463
+922 / -1
i
ii
iii
iv
v
vi
vii
Anabrus simplex
Gryllodes sigillatus
Gryllus longicercus
Gryllus bimaculatus
Teleogryllus oceanicus
Teleogryllus occipitalis
Schistocerca americana
0.4
b
Counts
50
40
30
20
10
0
Orthogroup
glutamate-gated ion channel (TC 1.A.10.1)
eIF-2B alpha/beta/delta subunits
SCM
complex I subunit 1
HARBI1*
Histone H3
complex I subunit 5
THAP*
LLRC*
RTase
Histone H2B
Histone H2A
Cytochrome P450
Trypsin
Histone H1/H5
i
ii
iii
iv
v
vi
vii


a
b
c
Chr10
Chr9
ChrX
Chr8
Chr7
Chr1
Chr6
Chr5
Chr2
Chr4
Chr3
0MB
60MB
120MB
180MB
240MB
300MB
360MB
420MB
outer
inner
GC content
Gene density
Repeats density