# Establishment of CRISPR/Cas9-based knock-in in a hemimetabolous insect: targeted gene tagging in the cricket *Gryllus bimaculatus*

Yuji Matsuoka[1,3*], Taro Nakamura[2,4*], Takahito Watanabe[1,5], Austen A. Barnett[2,6], Sumihare Noji[7], Taro Mito[1*], Cassandra G. Extavour[2,8*]

1. Department of Life Systems, Institute of Technology and Science, the University of Tokushima Graduate School, 201 Minami-Jyosanjima-cho, Tokushima City, 770-8506, Japan
2. Department of Organismic and Evolutionary Biology, 16 Divinity Avenue, Cambridge MA 02138, USA
3. Department of Biological Sciences, National University of Singapore, 14 Science Drive 4 117543, Singapore
4. Current address: National Institute for Basic Biology, Nishigonaka 38, Myodaiji, Okazaki 444-8585, Aichi, Japan
5. Bio-Innovation Research Center, Tokushima University, 2272-2 Ishii, Ishii-cho, Myozai-gun, Tokushima 779-3233, Japan
6. Current address: DeSales University, 2755 Station Avenue, Center Valley, PA 18034, USA
7. Tokushima University, 2-14 Shinkura-cho, Tokushima City, 770-8501, Japan
8. Department of Molecular and Cellular Biology, 16 Divinity Avenue, Cambridge MA 02138, USA

*** Correspondence to:** Yuji Matsuoka dbsyuji@nus.edu.sg
Taro Nakamura taro@nibb.ac.jp
Taro Mito mito.taro@tokushima-u.ac.jp
Cassandra G. Extavour extavour@oeb.harvard.edu

**Running Title:** CRISPR/Cas9 knock-ins in a cricket

**Abstract**

Studies of traditional model organisms like the fruit fly *Drosophila melanogaster* have contributed immensely to our understanding of the genetic basis of developmental processes. However, the generalizability of these findings cannot be confirmed without functional genetic analyses in additional organisms. Direct genome editing using targeted nucleases has the potential to transform hitherto poorly-understood organisms into viable laboratory organisms for functional genetic study. To this end, here we present a method to induce targeted genome knock-out and knock-in of desired sequences in an insect that serves as an informative contrast to *Drosophila*, the cricket *Gryllus bimaculatus*. The efficiency of germ line transmission of induced mutations is comparable to that reported for other well-studied laboratory organisms, and knock-ins targeting introns yields viable, fertile animals in which knock-in events are directly detectable by visualization of a fluorescent marker in the expression pattern of the targeted gene. Combined with the recently assembled and annotated genome of this cricket, this knock-in/knock-out method increases the viability of *G. bimaculatus* as a tractable system for functional genetics in a basally branching insect.

**Keywords**: CRISPR/Cas9, Orthoptera, genome editing, Hox genes, *Ultrabithorax*, *abdominal-A*

## Introduction

In what is often called the "post-genomic era," (Wainberg et al., 2021), massive advances in nucleic acid sequencing chemistry over the last two decades have given scientists access to greater volumes of gene sequence data than ever before (Kulski, 2016; Papageorgiou et al., 2018). However, this wealth of genomic information has highlighted two major gaps in our understanding of gene function and evolution. First, comparative genomic data and increased taxon sampling in functional genetic, developmental and cellular biology have revealed that the biology of many traditional laboratory model organisms is not representative of the broader clades to which they belong (Goldstein and King, 2016). Second, our ability to deduce accurately gene function from sequence data is limited to those genes that display high sequence and structural conservation (Ashburner et al., 2000; Consortium et al., 2020), and tools for manipulating gene function have been developed for only a small fraction of organisms (Russell et al., 2017). Addressing these problems calls for both increased taxon sampling and development of techniques to enable targeted alteration of gene function in understudied organisms. Here we address both of these issues by developing a method for targeted genome editing, including both knock-out and knock-in editing, in a basally branching insect model organism, the cricket *Gryllus bimaculatus*.

*G. bimaculatus* is an emerging model organism in a variety of fields of biology (Horch et al., 2017a; Kulkarni and Extavour, 2019). Ease of husbandry (Horch et al., 2017b), detailed developmental staging tables (Donoughe and Extavour, 2016), established gene expression analysis methods (Horch et al., 2017b), and an assembled and annotated genome (Ylla et al., 2021) make this cricket a highly amenable hemimetabolous laboratory model system (Kulkarni and Extavour, 2019). The developmental biology of *G. bimaculatus* has been of special interest, as in contrast to the ontogenetically derived model *Drosophila melanogaster*, many aspects of cricket embryogenesis are thought to resemble putative ancestral developmental modes of insects (Davis and Patel, 2002). For example, the function of several axial patterning genes has been analyzed and compared to that of their *D. melanogaster* homologues, revealing that the gene regulatory networks governing axial patterning have undergone considerable evolutionary change across insects (Matsuoka et al., 2015; Mito et al., 2007; Mito et al., 2008). *G. bimaculatus* is also used for the analysis of gene function in tissue regeneration (Bando et al., 2013; Mito et al., 2002; Nakamura et al., 2007), as this cricket can regenerate amputated organs, including legs and antennae, after several rounds of juvenile molts. In addition, *G. bimaculatus* is also used for the analysis of gene function and neuronal circuits in neuronal activity, including learning, memory

and circadian clocks (Hedwig and Sarmiento-Ponce, 2017; Matsumoto et al., 2018; Mizunami and Matsumoto, 2017; Tomiyama et al., 2020). Moreover, several species of cricket are being farmed as a new food source for humans because of their high protein and nutrient content (Huis et al., 2013).

Genome editing techniques using artificial nucleases were previously established in *G. bimaculatus* (Watanabe et al., 2012). However, construction of artificial nucleases is laborious. More recently, use of the clustered regulatory interspaced short palindromic repeat (CRISPR)/associated Cas9 nuclease (CRISPR/Cas9) has emerged and been verified as an efficient tool for genome editing in several arthropod species (Gilles et al., 2015; Gratz et al., 2013; Kistler et al., 2015; Li et al., 2015; Martin et al., 2016; Zhang et al., 2017). Here we briefly review this technique, investigate and demonstrate its utility for targeted genome modification in *G. bimaculatus*.

The CRISPR system is thought to have its ancestry in adaptable immune mechanisms used by many bacteria, which protect the genome from invasion by viruses (Amitai and Sorek, 2016). The type II CRISPR system from *Streptococcus pyogenes* was adopted for sequence-specific introduction of double strand breaks in other organisms (Jinek et al., 2012). The CRISPR/Cas9 system is comprised of the Cas9 nuclease and a short guide RNA (sgRNA), which consists of a Cas9 interaction site and a target recognition site. By changing the sequence of the target recognition site, in principle any sequence in the genome can be targeted. However, in practice there are several constraints on experimental design in this system to optimize the function and specificity of sgRNAs. For example, a protospacer adjacent motif (PAM) located in the 3' downstream region of the target sequence is needed for interaction of sgRNA with Cas9 nuclease (Jinek et al., 2012). Recently, several additional factors that can impact the specificity and efficiency of sgRNA in several systems have been reported, including the number and positions of mismatches, and GC content (Hsu et al., 2013).

In the CRISPR/Cas9 system, sgRNAs recruit Cas9 nuclease to the target sequence, and Cas9 then introduces a double strand break (DSB) at the target sequence. The presence of DSBs triggers the activity of the cell's DNA repair machinery, non-homologous end joining (NHEJ), or homology directed repair (HDR). NHEJ is an error-prone machinery, such that insertions or deletions can be generated at the break point (Branzei and Foiani, 2008). By utilizing artificial nucleases to trigger NHEJ, we previously succeeded in generating mutant lines in *G. bimaculatus* (Watanabe et al., 2012). HDR, however, would be offer a more precise repair machinery, as the

break is repaired through use of a homologous template. By supplying a donor template containing sequence homologous to the target, in principle a desired donor sequence can be integrated into the genome though HDR. Such gene knock-ins, while highly desirable for detailed analysis of the function of genomic regions, are more difficult to achieve than gene knock-outs because of the low efficiency of HDR in eukaryotes (Hagmann et al., 1998), and although success with HDR has been reported in some insects including silk moth (Ma et al., 2014; Zhu et al., 2015), multiple mosquito species (Gantz et al., 2015; Hammond et al., 2016; Kistler et al., 2015; Purusothaman et al., 2021) and mosquito cell lines (Rozen-Gagnon et al., 2021)and the beetle *Tribolium castaneum* (Gilles et al., 2015), our attempts at HDR-based knock-in techniques have never succeeded in *Gryllus* (unpublished observations). Recently, an efficient gene knock-in method through NHEJ was developed in the zebrafish *Danio rerio* (Auer et al., 2014). In this method, both the genome and the donor vector are cleaved *in vivo*, then the terminal genomic and donor sequences are combined through NHEJ. The method is efficient and can integrate longer constructs into the genome than knock-ins achieved through HDR (Auer et al., 2014). Bosch and colleagues (Bosch et al., 2019) subsequently reported that this knock-in strategy also works in *D. melanogaster*.

Here, we present evidence that CRISPR/Cas9 system functions efficiently in *G. bimaculatus*, and that the efficiency of targeted gene disruption is much higher than that achieved using artificial nucleases. We demonstrate the utility of this technique for developmental biology by performing functional analysis of the *G. bimaculatus* orthologues of the Hox genes *Ultrabithorax* (*Gb-Ubx*) and *abdominal-A* (*Gb-abd-A*). Furthermore, using a donor vector containing an autonomous expression cassette, we demonstrate that gene knock-in by a homology-independent method works efficiently in *G. bimaculatus*. We show that this homology-independent gene knock-in method can be applied to identify mutant individuals simply by detecting marker gene expression. Efficient targeted genome editing, now including both knock-out and knock-in techniques, will pave the way for making this cricket a much more sophisticated model animal for functional genetic laboratory studies.

## Results

### *Targeted mutagenesis of the* Gb-lac2 *locus*

To determine whether the CRISPR/Cas9 system was functional in the cricket, we first tried to perform a targeted gene knock-out of the *laccase 2* (*Gb-lac2*) gene (Table 1), which regulates

tanning of the arthropod cuticle following molting (Arakane et al., 2005). We chose this gene because of its easily detectable loss of function phenotype, and because we had previously successfully generated stable mutant lines for this gene by using artificial nucleases (Watanabe et al., 2012). sgRNA target sites were first determined using the ZiFit online tool (Sander et al., 2010), and then we chose the target sequence from among these candidate sequences based on the number of mismatches relative to the other sequences on the genome (< 3 mismatches in the whole sgRNA sequence) and GC content (70 ± 10% for the whole sgRNA sequence). Based on these criteria, we designed sgRNAs against the fifth exon of *Gb-lac2*, which is close to the target regions of the previous artificial nuclease experiment (Fig. 1D; Table 2) (Watanabe et al., 2012).

We co-injected 0.5 μg/μl sgRNA and 0.5 μg/μl of Cas9 mRNA into 128 fertilized cricket eggs within 1-3 h after egg laying (AEL) (Table 3). Five days after injection, we evaluated the frequency of mutant alleles in individual eggs using the Surveyor™ nuclease assay (Qiu et al., 2004); see Materials and Methods for detailed mechanism and procedure). For sgRNA #1, we detected cleaved fragments of the expected sizes in all X examined eggs (Fig.1C), and for sgRNA #2, we detected these fragments in 13 out of 29 eggs examined (Table 3).

We observed mosaic pigmentation of the cuticle in 92% of $G_0$ hatchlings that emerged from the individual eggs injected with sgRNA#1, and in 44% of $G_0$ hatchlings that emerged from individuals injected with sgRNA #2, consistent with Cas9-mediated interruption of the *Gb-lac2* gene in some, but not all, somatic cells of the $G_0$ hatchlings raised these hatchlings to adulthood (Fig.1A). We crossed these $G_0$ adults with wild type crickets of the opposite sex, to determine the efficiency of germ line transmission of the Cas9-induced *Gb-lac2* mutations to the $G_1$ generation. We found that 77% of the mosaic $G_0$s injected with sgRNA#1, and 20% of those injected with sgRNA#2, transmitted the mutation to their offspring (Fig.1B). To determine the nature of the *Gb-lac2* Cas9-induced alleles, we isolated genomic DNA from each line and analyzed the sequence of the *Gb-lac2* locus. We found that several different types of indel mutation were introduced at the target locus (Fig.1D). These results indicate that this CRISPR/Cas9-mediated genome editing system is functional in the cricket.

*Targeted mutagenesis of the* Gb-Ubx *locus via knock-out*

To compare phenotypes obtained with targeted gene disruption to those obtained with the RNA interference (RNAi) method that has hitherto been the most common method of performing functional genetics in this cricket (Mito and Noji, 2008), we used the CRISPR/Cas9 system to

perform functional analyses of the *G. bimaculatus* ortholog of the Hox gene *Ultrabithorax* (*Gb-Ubx*) (Table 1). RNAi-induced phenotypes for *Gb-Ubx* have been previously examined in developing abdominal segments (Barnett et al., 2019; Matsuoka et al., 2015), providing a basis for comparison with CRISPR-induced mutants. We designed sgRNA for a sequence within an exon upstream of the homeodomain (Fig.2A), and co-injected 0.5 µg/µl of this sgRNA and 1 µg/µl Cas9 mRNA into 167 fertilized cricket eggs within 1-3 h AEL. Seven days following injection, we extracted genomic DNA and performed the Surveyor™ assay to determine the efficiency of gene targeting. We found that *Gb-Ubx* mutations had been induced in all examined eggs (n=16) (Fig.2B). The remaining 151 injected $G_0$ embryos gave rise to ten adults, which we backcrossed to wild type adults of the opposite sex. We randomly chose approximately 30 $G_1$ eggs from each of the ten $G_0$ crosses, extracted genomic DNA from the pooled embryos, and performed the SURVEYOR Surveyor™ assay. We found that six out of ten $G_0$ crickets transmitted *Gb-Ubx* mutations to the next generation. We selected one of the $G_1$ *Gb-Ubx*$^{CRISPR}$ lines, which had a frame-shift mutation in the *Gb-Ubx* locus, for further phenotypic analysis. These *Gb-Ubx*$^{CRISPR}$ mutants displayed two different classes of phenotype: (1) *Contraction of the T3 leg*. Wild type *G. bimaculatus* adults have large, conspicuous T3 jumping legs. However, heterozygous mutants had smaller T3 legs than wild type (Fig.2C), and homozygous mutants obtained in the $G_2$ generation had T3 legs that were even smaller, almost the same size as T1/T2 legs (Fig.2C). These phenotypes were in good correspondence with those previously observed for *Ubx* RNAi in the cricket *Acheta domestica* (Mahfooz et al., 2007). (2) *Transformation of the A1 appendage*. Wild type *G. bimaculatus* germ band stage embryos possess two appendage-like organs on the A1 segment called the pleuropodia (Rathke, 1844; Wheeler, 1892). Instead of the pleuropodia present in wild type adults, the appendage outgrowths on the T1 segment of homozygous mutants were transformed towards leg-like structures (Fig.2D). This phenotype matches that previously observed in *Gb-Ubx* RNAi embryos (Barnett et al., 2019). In addition, in late embryos it was evident that an ectopic tergite was formed on the A1 segment of these mutants (Fig.2C). *Gb-Ubx*$^{CRISPR}$ heterozygous mutants are fertile but homozygous mutants were lethal. Therefore, to maintain this line, *Gb-Ubx*$^{CRISPR}$ heterozygous mutants were crossed to each other, and we performed the SURVEYOR Surveyor™ assay to isolate heterozygous mutants.

To confirm whether the production of Gb-Ubx protein was indeed disrupted by these CRISPR-induced mutations, we performed immunostaining with the “UbdA” monoclonal antibody FP6.87 (Kelsh et al., 1994), which recognizes both Ubx and Abd-A proteins, and was

previously reported to cross-react in multiple *Gryllus* species including *G. bimaculatus* (Barnett et al., 2019; Mahfooz et al., 2004). In wild type embryos, the UbdA antibody revealed the expected combination of the Gb-Ubx and Gb-Abd-A expression patterns (Barnett et al., 2019; Mahfooz et al., 2004) (Fig.2D). In *Gb-Ubx*$^{CRISPR}$ embryos, however only the Gb-Abd-A expression domain was detected, while Gb-Ubx expression was clearly absent (Fig.2D), suggesting that the CRISPR-induced *Gb-Ubx* mutations interfered with Gb-Ubx protein production. To address the possibility of off-target CRISPR-mediated gene disruptions, we compared the abdominal appendage phenotypes of *Gb-Ubx*$^{CRISPR}$ embryos with those of *Gb-Ubx*$^{RNAi}$ embryos. Both types of embryos displayed similar phenotypes, namely moderate outgrowth of leg-like structures on the A1 segment, which normally generates pleuropodia rather than walking legs (Fig.2D). Taken together, these results suggest that the CRISPR/Cas9 system induced mutations specifically into the *Gb-Ubx* locus, which disrupted *Gb-Ubx* function.

*In-depth analysis of mutagenesis profile for the CRISPR/Cas9 system in* G. bimaculatus

To optimize the genome editing procedure, we wished to evaluate how the timing of injection affected NHEJ mutagenesis. For detailed assessment of this mutagenesis, we therefore performed in-depth analysis of the CRISPR mutants using next generation sequencing.

Our previous study had revealed early cellular dynamics during cricket embryogenesis (Nakamura et al., 2010), allowing us to assess whether specific mutagenesis events were correlated with cellular behaviors during early development. As in *D. melanogaster* (Foe and Alberts, 1983)early mitotic divisions in *G. bimaculatus* embryos are syncytial, meaning that mitosis takes place without cytokinesis, resulting in multiple energids (nuclei surrounded by aqueous cytoplasm but lacking a unique lipid bilayer) within a single cell membrane (Donoughe and Extavour, 2016; Nakamura et al., 2010; Sarashina et al., 2003). To evaluate whether and how the timing of injection affected mutagenesis outcomes, we chose four early embryonic time points following the one-hour embryo collection period, as follows (Supplementary Fig.2A): (1) At the 1h injection time point, energids start to migrate from the center of the egg to the cortex. (2) At the 3h injection time point, energids continue to become distributed throughout the yolk, accompanied by mitotic cycles. (3) At the 5h injection time point, energids have become nearly uniformly distributed throughout the egg cortex and begin tangentially oriented nuclear division. (4) The 9h injection time point is one to eight hours before cellularization (Donoughe and Extavour, 2016). We co-injected 0.5 µg/µl of the *Gb-Ubx* sgRNA, the Gb-lac2 sgRNA or an sgRNA targeting abdominal-A (*Gb-abd-A*); see

"*Knock-in of donor vector sequence at the* Gb-abd-A *locus*" below) described above, and 1 µg/µl of Cas9 mRNA into the eggs at each of these time point. At 5d AEL, genomic DNA was isolated from three individual embryos for each injection time point and used for amplicon sequencing. We examined the on-target site and the single highest predicted potential off-target site for each of the *Gb-Ubx* and *Gb-abd-A* genes. For each sample we performed amplicon sequencing with three replicates.

For *Gb-Ubx* on-target genome disruptions, we found that the rate of NHEJ-induced mutations decreased with the age of the embryo at injection (Supplementary Fig.2B). The same trend was also observed for *Gb-abd-A* on-target mutations (Supplementary Fig.2D; see section "*Targeted mutagenesis of the* Gb-abd-A *locus via knock-in"* below). This result is well correlated with the phenotypic severity observed in the $G_0$ hatchlings emerging from the *Gb-lac2*$^{CRISPR}$ embryos. *Gb-lac2*$^{CRISPR}$ embryos injected at the two earlier time points (1h and 3h) gave rise to hatchlings with broad mutated white patches of cuticle (Supplementary Fig.2H). In contrast, the embryos injected at 5h showed milder phenotypes (Supplementary Fig.2H) and the embryos injected at 9h showed only little detectable phenotype (Supplementary Fig.2H). For both *Gb-abd-A* and *Gb-Ubx* genes, the rate of NHEJ-induced mutations at the studied off-target site was less than 1.3% for all injection time points (Supplementary Fig.2C and 2E), suggesting that off-target effects may be minimal in this system.

*Knock-in of donor vector sequence at the* Gb-Ubx *locus*

In addition to targeted sequence deletions, targeted sequence knock-in is a highly desirable technique that would expand our ability to understand the functions of genomic regions of interest. We had previously attempted to achieve targeted gene knock-ins through homology-dependent repair, but this method has not worked in *Gryllus* in our hands to date (data not shown). In a homology-independent knock-in method reported for *D. rerio* and *D. melanogaster* (Auer et al., 2014; Bosch et al., 2019), both genome and donor vector are cleaved *in vivo*, then the cut ends of genome and donor vector are combined through NHEJ. This method is more efficient than the homology-dependent method, potentially because NHEJ is highly active throughout the cell cycle in eukaryotes (Hagmann et al., 1998). However, due to the nature of NHEJ, the orientation of integration of the donor vector sequence cannot be controlled. In addition, indel mutations are generated at the junction point. To try to circumvent these issues, which might otherwise prevent functional knock-in, we generated a donor vector containing an autonomous expression cassette

comprising the *Gryllus actin* (*Gb-act*) promoter followed by the *eGFP* coding sequence (Nakamura et al., 2010). As a sgRNA recognition site, we included a partial *DsRed* gene sequence (Auer et al., 2014), which is native to the coral *Discosoma sp.* (Baird et al., 2000) and not present in the cricket genome. We predicted that successful knock-in of this donor sequence into the genome would result in GFP expression being driven by the *Gb-act* promoter regardless of the insert's orientation or any potential induced indel mutations. To try to further increase the utility of this tool to facilitate identification of targeted gene disruptions, we targeted knock-in of the donor sequence to an exon of the target gene, which we anticipated would result in disruption of target gene function. Our goal was to be able to identify such successfully knocked-in individuals by detectable GFP expression in the known expression domains of the target gene.

We targeted the *Gb-Ubx* locus for this targeted knock-in strategy and used the same sgRNA as that used for the knock-out experiment described above (see section "*Targeted mutagenesis of the Gb-Ubx locus via knock-out*", Fig.2A; Fig.3A). We co-injected 50 ng/µl of sgRNA for *Ubx* locus, 50 ng/µl of sgRNA for donor vector, 100 ng/µl of Cas9 mRNA, and 100 ng/µl of donor vector into fertile cricket eggs. By seven days after injection, four out of 85 injected embryos (4.7%) showed mosaic GFP expression in the T3 trunk and leg (Fig.3B). Of the 85 injected embryos, 18 individuals (21.2%) grew to adulthood. We crossed them individually with wild type counterparts of the opposite sex and evaluated GFP expression in their offspring. One out of these 18 $G_0$ crickets (5.6%) produced embryos with GFP expression in a pattern identical to that of *Gb-Ubx* (Barnett et al., 2019; Matsuoka et al., 2015; Zhang et al., 2005) (Fig.3B). The GFP expression was detectable through the eggshell even at late embryonic stages. At adult stages, knock-in crickets showed detectable GFP expression in the hind wing and T3 legs (Supplementary Fig.3A).

To confirm the integration of donor sequence into the genome, we performed PCR and sequence analysis. We designed specific primers for each 5' and 3' junction point (Fig.3A). All XX examined embryos showed the expected amplicon size for both junctions (Fig.3D), suggesting that at least two copies of the donor vector fragment were integrated into the genome. Sequence analysis further confirmed the integration of the donor plasmid into the genome, and that indel mutations were generated at each junction (Fig.3E). We also performed copy number estimation by real-time quantitative PCR. The expression level of GFP was normalized to the expression level of the endogenous *orthodenticle* gene, which is known to have only one copy in the genome (Nakamura et al., 2010; Ylla et al., 2021). The results of this analysis indicated that four copies of

the donor plasmid were integrated into the genome (Supplementary Fig.4A).

To determine whether the function of the target gene was indeed disrupted by this Knock-in/Knock-out strategy, we examined GFP expression and morphology in $G_1$ *Gb-Ubx*$^{CRISPR-KI}$ embryos. Among the $G_1$ *Gb-Ubx*$^{CRISPR-KI}$ embryos, we found they displayed one of two different intensities of GFP expression (Fig.3C). Some crickets showed weak GFP expression and displayed no detectable morphological abnormalities (Fig). The crickets with strong GFP expression, however, had smaller T3 legs and formed leg-like structures rather than pleuropodia on the A1 segment (Fig). These phenotypes, which were the same as those observed in the *Gb-Ubx* homozygous mutant (Fig.2C), suggested that the weak GFP expression crickets may be heterozygous mutants, and the strong GFP expression crickets may be homozygous mutants.

*Knock-in of donor vector sequence at the* Gb-abd-A *locus*

To confirm the efficiency and utility of this method, we next chose the Hox gene *Gb-abdominal-A* (*abd-A*) as a target (Table 1). We designed sgRNAs for the sequence within the exon just upstream of the homeodomain (Fig.4A). We co-injected 50 ng/μl of sgRNA for the *Gb-abd-A* locus, 50 ng/μl of sgRNA for the donor vector, 100 ng/μl of Cas9 mRNA, and 100 ng/μl of donor vector into fertilized cricket eggs.

Of 38 injected $G_0$ embryos, five showed mosaic GFP expression in the abdomen (Fig.4B). Four $G_0$ adults were individually backcrossed with wild type counterparts of the opposite sex to obtain multiple $G_1$ crickets. We obtained one stable transgenic line, in which GFP expression in $G_2$ embryos was similar to the previously documented expression pattern of *Gb-abd-A* transcript (Barnett et al., 2019; Matsuoka et al., 2015; Zhang et al., 2005) (compare Fig.4B with Supplementary Fig.1C). In a replicate injection experiment, we obtained a second such transgenic line (Table 3). PCR and sequence analysis confirmed that one of the two lines contained the plasmid fragment in the sense orientation, and the second line contained the plasmid fragment in the antisense orientation (Fig.4C). Copy number estimation analysis results suggested that a single plasmid fragment was integrated into the genome in each line (Supplementary Fig.4).

In the *Gb-abd-A*$^{KI-exon}$ lines, GFP expression was detectable in nymphs even through the cuticle FIG. We further detected GFP expression in adult male and female internal organs. In wild type females, a pair of ovaries, each comprising hundreds of ovarioles, is located in the anterior abdomen (Nandchahal, 1972). Mature eggs are stored in an egg chamber at the posterior of each ovariole, and eggs are subsequently moved posteriorly through the oviduct. The posterior end of

the oviduct is connected to the uterus, where fertilization takes place, located at the base of the ovipositor (Supplementary Fig.5C''). In *Gb-abd-A*$^{KI\text{-}exon}$ mutant females, GFP expression was detected in the posterior portion of the oviduct (compare Supplementary Fig.5B' and B'' to Supplementary Fig.5A'). In *Gb-abd-A*$^{KI\text{-}exon}$ males, ubiquitous GFP expression was detected throughout the testis (compare Supplementary Fig.5E' and E'' to Supplementary Fig.5D'). The observed GFP expression in females is reminiscent of the expression pattern of *D. melanogaster abd-A* in the developing female genital disc (which gives rise to the somatic reproductive structures including the oviduct in this fruit fly (Epper, 1983; Sánchez and Guerrero, 2001), and in the adult oviducts (Foronda et al., 2006). *abd-A* expression has not, to our knowledge, been detected in the *D. melanogaster* male genital disc (Freeland and Kuhn, 1996), which gives rise to male somatic reproductive structures. However, high-throughput sequencing data from the modENCODE project do report *abd-A* expression in the adult *D. melanogaster* testis (Brown et al., 2014).

*Targeted insertion of an expression cassette into an intron of the* Gb-abd-A *locus*

Kimura and colleagues (Kimura et al., 2014) demonstrated that in *D. rerio* the homology independent method could be applied for trapping endogenous enhancer activity by inserting a donor sequence containing an expression cassette into the 5'UTR of genes of interest. We aimed to apply this technique to *G. bimaculatus* by attempting to knock-in a donor vector into the intronic region of *Gb-abd-A* (Fig.5A).

We co-injected an sgRNA against an intron of *Gb-abd-A*, together with all other relevant reagents as described above (sgRNA against the donor vector, donor vector, and Cas9 mRNA). Of 100 injected eggs, two eggs showed mosaic expression of GFP in the abdomen (Fig.5B). When the donor sequence was inserted into an exon in the previous experiment (see "*Knock-in of donor vector sequence at the* Gb-abd-A *locus*" above), GFP expression was accompanied by a phenotype of ectopic leg-like structure development on abdominal segments (Fig.5B, B', B''), as previously observed in *Gb-abd-A* RNAi experiments (Barnett et al., 2019). However, when the plasmid fragment was inserted into an intron, the region expressing GFP did not generate ectopic leg-like structures (Fig.5C, C', C''). This apparent absent or minimal loss of function phenotype in the intron knock-in embryos might explain the relatively high survival rate of the intron-targeted $G_0$ embryos (28% of injected $G_0$ embryos survived to adulthood) compared to that of the exon-targeted $G_0$ embryos (4.9% to 8.5% of XX injected embryos survived to adulthood; Table 3).

We obtained one *Gb-abd-A*$^{KI\text{-}intron}$ line, in which we confirmed that one donor vector sequence was integrated into the target region in a forward orientation (Fig). We carefully inspected the morphology of the *Gb-abd-A*$^{KI\text{-}intron}$ adult crickets to assess whether potential post-embryonic functions of the target gene were affected by insertion of the donor sequence into an intron. $G_1$ heterozygous *Gb-abd-A*$^{KI\text{-}intron}$ females did not show the supernumerary ovipositors observed in XX adults FIG, and they laid eggs normally (data not shown). For further confirmation, we examined the morphology of $G_2$ homozygous *Gb-abd-A*$^{KI\text{-}intron}$ mutants. Approximately 25% of examined $G_2$ eggs showed strong GFP expression, which we interpret is likely indicative of a homozygous mutant. All of these strong-GFP $G_2$ embryos generated leg-like structures on the abdomen (Fig.5E, E', E''), suggesting that the function of the target gene was somewhat affected in the homozygous condition, unlike in the mosaic condition exhibited by $G_0$ embryos FIG.

**Discussion**

In the present study, we demonstrated that targeted knockout and knock-in by using CRISPR/Cas9 system works efficiently in the cricket *G. bimaculatus*. We performed functional analysis of CRISPR/Cas9-induced mutations in the Hox genes *Gb-Ubx* and *Gb-abd-A* during embryogenesis and at post-embryonic stages. We found that the cleavage efficiency of the CRISPR/Cas9 system was much higher than that previously reported for artificial nucleases in this cricket ((Watanabe et al., 2012; Watanabe et al., 2014)). We demonstrated that gene knock-in via a homology-independent method is effective in this cricket, and successfully applied it to functional analysis of Hox genes by knocking a donor sequence into an exon of the target gene to disrupt the function of the target gene (knock-in/knock-out). In addition, we succeeded in trapping endogenous gene activity using this method and revealed a number of new expression domains that had not been previously observed with traditional methods (Barnett et al., 2019; Matsuoka et al., 2015; Zhang et al., 2005)(Barnett et al., 2019; Matsuoka et al., 2015; Zhang et al., 2005). This homology-independent method is technically simpler than the homology-dependent methods, as the donor plasmid does not need to be newly made for each target region.

*CRISPR/Cas9 system vs RNA interference*

Delivering the proper amount of genome editing constructs at the proper time is important for efficient outcomes. For example, the embryos injected around 3h AEL, at which energids are distributing, were highly efficiently affected, but both detectable mosaic phenotypes (of *Gb-lac2*

crispants) and the rate of NHEJ decreased with later injections at blastoderm stages (Supplementary Fig.2H). Furthermore, the efficiency of knock-in was much lower than that of knockout FIG. Thus, we suggest that further optimization of delivery conditions can be achieved by using pigmentation genes as an index.

In our study, reproducibility and severity of phenotypes generated with the CRISPR/Cas9 were greater than those obtained with RNAi FIG. We speculate that the efficiency of RNAi-mediated knockdown may be influenced by when and at what levels the target gene is expressed. In the case of *Gb-abdA* and *Gb-Ubx*, these genes are expressed much later in development (stage) than the stage at which we perform injections (stage). Correspondingly, much higher concentrations of dsRNA have proven necessary to produce even mild phenotypes (5-6 μg/μl; Fig.2D), than those typically used for RNAi against most genes in this cricket (1-2 ug/ul; e.g. Donoughe et al., 2014). In contrast, early indel mutations generated by genome editing techniques resulted in clearly detectable, severe phenotypes FIG. Several studies have demonstrated that genome editing techniques can sometimes be adequate for functional analysis of target genes in mosaic $G_0$ individuals (e.g. Daimon et al., 2015; Martin et al., 2016; Matsuoka and Monteiro, 2018). However, it is often difficult or impossible to unambiguously identify mutant cells in such mosaics. In this regard, establishment and maintenance of stable mutant lines as performed herein, allows for less ambiguous phenotypic analysis in this cricket.

Although the CRISPR/Cas9 system is efficient, it offers little to no conditionality, which can complicate study of the many genes that act pleiotropically during development (Minelli, 2016). For example, in the case of *Gb-abd-A*, the gene acts to repress leg formation in the abdomen at embryonic stages (Supplementary Fig.6), while at adult stages, it regulates proper development of female genitalia (Supplementary Fig.5). Likely because of this latter phenotype, we were unable to obtain homozygous *Gb-abdA*$^{CRISPR}$ animals. To overcome this problem, sophisticated genetic methods, like balancer chromosomes (Miller et al., 2019), will need to be developed in the future. In this regard, RNAi offers more options for conditional control of gene function. By controlling the timing of injection of dsRNA, target gene activity can be knocked down at any desired developmental stage in *G. bimaculatus* (Dabour et al., 2011; Nakamura et al., 2008; Takahashi et al., 2009). Thus, while the CRISPR/Cas9 system is a powerful new tool for gene function analyses, RNAi remains a useful technique for this system.

*Application of homology independent knock-in method for functional analysis of endogenous genes*

Homology-independent knock-in methods will expand our ability to analyze the function of target genes in this hemimetabolous insect model. Here, we demonstrated one such application, the KI/KO method, which allows for isolation of mutants without PCR-based genotyping. When analyzing mutant phenotypes, affected individuals must typically be distinguished either by their morphology or by molecular methods to detect changes in target gene product levels or functions. In the case of the *Gb-Ubx* mutant, we would have needed to distinguish subtle differences in the T3 and A1 embryonic segments (Fig.2C), requiring destructive sampling. Moreover, antibodies against target genes may not be routinely available in many cases. The KI/KO method allows us to distinguish mutant individuals based on maker gene expression. Even heterozygous and homozygous mutants can sometimes be distinguished based on the intensity of marker gene expression. A similar strategy was employed in mosquitos via HDR (McMeniman et al., 2014). In the present study, we could easily identify the GFP expression resulting from the KI/KO event because it matched the previously characterized expression pattern for *Gb-Ubx* (Barnett et al., 2019; Matsuoka et al., 2015; Zhang et al., 2005). However, for target genes with previously uncharacterized expression domains, analysis may be more complex.

The promoter used in all expression cassettes herein is the same one used in a previous study to drive ubiquitous constitutive expression (Nakamura et al., 2010). Nevertheless, all knocked-in lines showed a spatially and temporally restricted GFP expression pattern like that of the target gene. We speculate that the promoter in the expression cassette acts as a minimal promoter, and that the observed GFP expression resulted from trapping endogenous enhancer activity. The GFP expression was not caused by the fusion to the endogenous gene product, since both the line containing an inverted orientation of the donor sequence, and the knock-in line targeting an intronic region, showed similar GFP expression patterns. To enhance the usefulness of this method, identification and use of a ubiquitous and strong promoter could in principle drive exogenous marker gene expression in the whole embryo without being subject to positional effects.

A remarkable feature of the homology-independent knock-in method is the length of sequence that can be integrated. In case of KI through HDR, a few kb of sequence can be integrated into the genome in arthropods (Gilles et al., 2015; McMeniman et al., 2014). In this study, through NHEJ, at least six kb of plasmid sequence was integrated into the genome. Furthermore, in some cases, four copies of plasmid sequence were integrated in tandem into the genome. In this case,

we speculate that first the donor plasmids were digested and combined via NHEJ, and then the combined fragment was knocked in into the genome via NHEJ, suggesting that homology-independent knock-in might be able to integrate several tens of kb of sequence into the genome. A recent study showed that a 200 kb BAC vector could be integrated into a rodent genome through a similar strategy (Yoshimi et al., 2016). This method might therefore be used for direct functional comparison of genomic regions by exchanging homologous regions between related species of interest.

The efficiency of knock-in through NHEJ is high, but to make improve its feasibility as a technique for functional genetic analysis in this cricket, future studies may be able to further enhance efficiency by optimizing at least one of three parameters, as follows: (1) *Enhance expression cassette copy number*. Empirically, $G_0$ crickets showing mosaic expression tend to transmit their knocked-in transgene to their offspring. To increase the efficiency of obtaining knock-in lines, future efforts should therefore focus on increasing the number of mosaic marker gene expression cassettes in $G_0$ embryos. Based on the results of our copy number estimation, there was variation in the number of vectors integrated into the genome, and the GFP expression level appeared correlated with the number of inserts. We also observed that some expression cassettes seemed to show higher expression levels than others and were therefore easier to detect in mosaic $G_0$s. For example, the transgenic line containing one copy of the donor sequence within a *Gb-Ubx* intron had weaker GFP expression than the *Gb-abd-A* intronic knock-in line that also possessed one copy of the donor sequence FIG. We speculate that the *Gb-abd-A* locus might be able to drive higher gene expression than the *Gb-Ubx* locus. Inclusion of inducible expression elements such as a heat shock promoter or a modified Gal4/UAS system might help to enhance the activity of the expression cassette. (2) *Supply sgRNA and/or Cas9 nuclease via plasmid.* Recently, homology-dependent gene knock-in in the beetle *Tribolium T. castaneum* was reported (Gilles et al., 2015). In this protocol, both sgRNA and Cas9 nuclease were supplied from a donor plasmid. We speculate that consecutive production of Cas9 and sgRNA might contribute to the high efficiency of knock-ins reported in that study (Gilles et al., 2015) . In human cultured cells, the half-life of sgRNA drops to near-background levels by 4h after nucleofection, and earlier introduction of Cas9 mRNA attenuates sgRNA degradation because Cas9 protein functions to protect sgRNAs from degradation (Hendel et al., 2015). Alternatively, using Cas9 protein instead of mRNA might also improve efficiency in *G. bimaculatus*. (3) *Introduce insulator sequences into the donor cassette*. Positional effects might also in principle prevent the full potential activity of

the expression cassette. In this study, our vector plasmid contained insulators of the sea urchin *Hemicentrotus pulcherrimus* arylsulfatase gene (Takagi et al., 2011) on either side of the expression cassette (see Materials and Methods), but we nonetheless detected while GFP expression in a pattern matching that of the target gene. To achieve effective insulation, future studies might evaluate several different combinations of insulator orientations, which can affect insulator activity (Tchurikov et al., 2009). Alternatively, other insulators such as that of the *gypsy* retrotransposon (Modolell et al., 1983), might be additional options for future optimization (Carballar-Lejarazú et al., 2013).

In conclusion, we provide evidence that the CRISPR-Cas9 system works well in the cricket *G. bimaculatus*. In depth analysis of CRISPR-Cas9-induced mutations revealed optimized injection timing. In addition, we succeeded in the targeted functional knock-in of exogenous sequences into the genome through NHEJ.

## Materials and Methods

### *Cricket husbandry*

All adult and juvenile *Gryllus bimaculatus* were reared in plastic cages at 26–30 °C and 50% humidity under a 10-h light, 14-h dark photoperiod. They were fed on artificial fish food (Tetra) or Purina cat food (item #178046). For microinjections, fertilized eggs were collected on a wet kitchen towel in a plastic dish and incubated at 28 °C as previously described (Barry et al., 2019; Watanabe et al., 2017).

### *Construction of sgRNA vectors*

For designing sgRNA, target sequences were designed with the ZiFit online tool (Sander et al., 2007; Sander et al., 2010). From the suggested candidates, we selected target sequences based on the number of mismatches (> 3 mismatches) and GC content (around 70%), as per Ren and colleagues (Ren et al., 2014). Off-target sites were designed using CasOT (Xiao et al., 2014). We modified the pDR274 vector (Addgene plasmid #42250) to expand its utility (the $GGN_{18}NGG$ sequence was present in the original pDR274, while in the modified vector, a $GN_{19}NGG$ sequence was used). Two synthetic oligonucleotides (5'-ATAG-$N_{19}$-3' and 5'-AAA-$N_{20}$) were annealed and inserted into the Bsa I site of the modified pDR274 vector. We confirmed insertion by Sanger sequence analysis.

524

525 *Synthesis of sgRNA and mRNA*

526 For sgRNA synthesis, the template for *in vitro* transcription was digested from the vectors
527 generated as described above with DraI. For Cas9 mRNA synthesis, the template for *in vitro*
528 transcription was digested from pMLM3613 (Addgene catalogue #42251) with PmeI. Both sgRNA
529 and Cas9 mRNA were *in vitro* transcribed using mMESSAGE mMACHINE T7 Kit (Life
530 Technologies catalogue #AM1344), and) and purified by ethanol precipitation. For the Cas9
531 mRNA, we attached a poly-A tail by using a poly-A tailing Kit (Life Technologies catalogue
532 #AM1350). The concentration of synthesized RNAs was estimated by NanoDrop and gel
533 electrophoresis.

534

535 *Construction of donor plasmids*

536 The *eGFP*bait-2A-RFP donor plasmid was generated in a pUC57 vector by commercial artificial
537 composition (GeneScript). The *DsRed*bait-G'act-*eGFP* donor plasmid was generated based on the
538 *eGFP*bait-2A-RFP donor plasmid. First, 2A-RFP was digested using BglII and NotI. *Gb-act-eGFP*
539 was also digested from a pXL-BacII- G'act-*eGFP* vector and ligated to generate the *eGFP*bait-
540 G'act-*eGFP* vector. Then, we digested this *eGFP*bait vector using BglII and SacI. We amplified
541 *DsRed*bait with primers (5' to 3') DsRed_fwd: GCTCAGATCTCTTGGAGCCGTACTGGAAC,
542 and DsRed-rev: GTACGAGCTCCATCACCGAGTTCATGCG. The amplicon was ligated to
543 generate the *DsRed*bait-G'act-*eGFP* donor plasmid. The *DsRed*bait-2×Ars_rev-G'act-*eGFP*-
544 2×Ars_fwd donor plasmid was generated based on the *DsRed*bait-G'act-*eGFP* donor plasmid. The
545 Ars insulator sequence ArsInsC from *H. pulcherrimus* (Takagi et al., 2011)was amplified from an
546 ArsInsC-containing plasmid (kind gift of Naoaki Sakamoto, Hiroshima University, Japan) and
547 integrated on either side of the expression cassette in the donor plasmid.

548

549 *Microinjection*

550 Cas9 mRNA, sgRNA, and donor vectors were injected into 2-5h AEL cricket eggs. Cricket eggs
551 were aligned in a groove 0.7 mm deep and 0.7 mm wide, made with 2% agarose in 1x phosphate-
552 buffered saline (PBS) using a custom mold as previously described (Barry et al., 2019; Watanabe
553 et al., 2017) and filled with 1xPBS. Needles for injection were made by pulling glass capillaries
554 with filament (Narishige catalogue # GD-1) with a pipette puller (Sutter Instrument catalogue # P-
555 1000IVF), using the following pulling program: (1) x3 Heat; 858, Pull; 0, Velocity; 15, Time; 250,

Pressure; 500, and (2) x1 Heat; 858, Pull; 80, Velocity; 15, Time; 200. To minimize the invasiveness of the injection, the tips of the pulled needles were sharpened and ground to a 20° angle by using a Micro Grinder (Narishige catalogue # EG-400). Approximately 5 nl of solution was injected into eggs with a Micro Injector (Narishige catalog # IM300). After injection, eggs were moved to a fresh Petri dish and submerged in fresh 1xPBS containing 50 U/ml penicillin and 50 µg/ml streptomycin (15070-063, Thermo Fisher), and incubated at 28°C. During the incubation period, the 1xPBS with penicillin and streptomycin was replaced every day. We observed fluorescent protein expression at the stages when the target gene was known to be expressed. Genomic DNA was extracted from 7 d AEL eggs and adult T3 legs and used for insertion mapping and sequence analyses. After 2 days of incubation, injected cricket eggs were moved to wet filter paper in a fresh Petri dish for hatching.

*Detection of indel mutations*

After Cas9 nuclease digests a target sequence, the disrupted sequence is repaired by either the NHEJ or the HDR cell machinery. To confirm a KO mutation, we searched for errors repaired by the NHEJ pathway, which sometimes induces or deletes nucleotides at the digested site during the repair process. Since disruption or repair are unlikely to take place identically in all cells of an injected $G_0$ embryo, $G_0$ animals are expected to contain heterogeneous sequences at the CRISPR targeted site, and thus to be heterozygous for a putative Cas9-induced indel. To confirm the activity of the sgRNAs, the SURVEYOR Surveyor™ Mutation Detection Kit (Transgenomic,) was used. This assay relies on a “surveyor” nuclease that can recognize and digest a heteroduplex DNA structure. First, genomic DNA was extracted from whole eggs or part of the T3 leg by a phenol chloroform method as previously described (Barry et al., 2019; Watanabe et al., 2017). Subsequently, approximately 200 bp of the targeted region was amplified by PCR from genomic DNA (Table 4). PCR conditions were optimized to reduce non-specific amplification or smearing. To create the putative heterogeneous DNA structure for the nuclease assay, the PCR product was heated to 98°C for five minutes, and then re-annealed by gradually cooling down to 30°C. Half of the PCR product was digested with the SURVEYOR Surveyor™ nuclease, and the other half was used as a negative control and incubated without the nuclease. Digestion was confirmed by agarose gel electrophoresis. For sgRNAs that yielded indels in the target sequence, digest of the PCR product by the Surveyor™ nuclease is expected to produce split fragments around the CRISPR targeted site relative to the negative control; the latter should not be digested by the nuclease and

thus should remain intact and run at the same size as the original amplicon. Positive PCR products were extracted from the gel, purified with the QIAquick Gel Extraction Kit (Qiagen catalogue #28506), and sub-cloned into the pGEM-Teasy vector (Promega catalogue #A1360) using TA-cloning. The vectors were used for Sanger sequence analysis.

*Amplicon sequence analysis*

After a 1h egg collection, eggs were incubated for the desired length of time at 28ºC. We co-injected 0.5 µg/µl sgRNA and 1 µg/µl Cas9 mRNA into fertilized cricket eggs after each of these incubation periods. Five days after injection, genomic DNA was extracted individually from three individual eggs from each of the four tested injection times; the latter analysis was performed in biological triplicate. Amplicon sequence analysis was performed by using MiSeq (Illumina), and the preparation of DNA libraries and sequencing reactions were performed according to the manufacturer's instructions. We read ~10,000 reads for on-target regions and ~50,000 reads for off-target regions. The assembly of output paired end reads was performed by using CLC Genomic Workbench (CLC Bio, QIAGEN Digital Insights). The relative proportions of reads containing indels and substitutions in the individual eggs were calculated with the online-tool CRISPResso (Pinello et al., 2016). We used the Integrative Genomic Viewer (Broad Institute) for investigation of the distribution of indels and substitutions (Thorvaldsdóttir et al., 2013).

*Insertion mapping*

Genomic DNA was extracted from GFP-positive eggs of each line. Due to the specifics of this knock-in method, two types of insertion of vector fragment (sense and antisense orientations) would be expected to occur; we therefore performed PCR using primers designed against either side of the putative junction. PCR was performed using target region-specific (upstream or downstream of sgRNA recognition site) and donor vector-specific primers (sequence within *eGFP* for forward integration and M13Fw for reverse integration). Primer sequences are listed in Table 4. Positive PCR products were extracted from the gel, purified by using the QIAquick Gel Extraction Kit (Qiagen catalogue #28506), and sub-cloned into the pGEM-Teasy vector (Promega catalogue #A1360) using TA-cloning. The vectors were used for Sanger sequence analysis.

*Embryo fixation, whole mount* in situ *hybridization, and immunohistochemistry*

Embryos were dissected in 1xPBS and fixed with 4% paraformaldehyde PFA in 1xPBS + 0.1% Tween (PBT) for 1h at 4℃. The fixed embryos were dehydrated stepwise in 25%, 50%, 75%, and 100% methanol in 1xPBT with five minutes per wash. The dehydrated embryos were stored in 100% methanol at -30℃. Whole-mount *in situ* hybridization with digoxigenin (DIG)-labeled antisense RNA probes was performed as previously described (Niwa et al., 2000; Zhang et al., 2005). Immunohistochemistry was performed as follows: Fixed embryos were rehydrated stepwise in 75%, 50%, and 25% solutions of methanol/ PBT and finally in 100% PBT for five minutes in each solution. After blocking with 1% bovine serum albumin (BSA) (Thermo Fisher) in PBT for one hour at room temperature, embryos were incubated with an anti-UbdA antibody FP6.87 (Kelsh et al., 1994) (Developmental Studies Hybridoma Bank) diluted 1:200 in 1% BSA/PBT overnight at 4℃. After washing with PBT three times, embryos were incubated in 1% BSA/PBT for one hour at room temperature, and then incubated with Alexa Fluor 488-conjugated Goat Anti-mouse IgG(H+L) (Invitrogen catalogue #A32723) diluted 1:400 in 1% BSA/PBT for one hour at 4℃. After washing the embryos with PBT once for 10-60 minutes, embryos were counter-stained with DAPI (Sigma catalogue #10236276001) stock solution 1mg/mL diluted 1:1000 in PBT for ten minutes, and then washed with PBT two times for 10-60 minutes per wash. PBT was then substituted with 25% and 50% glycerol/PBT to clear embryos for microscopy.

*Copy number estimation by using quantitative RT-PCR*

To estimate the number of plasmid fragments integrated into the genome via NHEJ events, we performed quantitative RT-PCR using total RNA from individual 5 day-old embryos of wild type, *Gb-Ubx*$^{\text{KI-exon}}$, and *Gb-abd-A*$^{\text{KI-exon}}$ lines, and compared the level of expression of the inserted gene with that of *Gb-otd*, which is present in single copy in the *G. bimaculatus* genome (Nakamura et al., 2010; Ylla et al., 2021). Total RNA was extracted from embryos using ISOGEN (Nippon-Gene catalog #315-02504). After treatment with DNaseI (Invitrogen catalog #AM2224), RNA was reverse transcribed to cDNA using SuperScriptIII reverse transcriptase (Invitrogen catalog #12574026). Real-time quantitative PCR was performed using the power SYBR Green PCR Master Kit (Applied Biosystems catalogue #4368577) and an ABI 7900 Real Time PCR System (Applied Biosystems) as described previously (Nakamura et al., 2008). Primer sequences are listed in Table 4. The level of *eGFP* was normalized to the level of *Gb-otd*, which suggested that the *Gb-Ubx*$^{\text{KI-exon}}$ line likely contained four copies of the plasmid fragment, and *Gb-abd-A*$^{\text{KI-exon}}$ line likely contained a single copy of the plasmid fragment.

**Acknowledgements**

We thank N. Sakamoto for the Ars plasmid, and the members of the Extavour and Mito labs for discussion.

**Competing Interests**

No competing interests declared.

**Funding3)**

This work was supported by National Science Foundation award number IOS-1257217 and funds from Harvard University to CGE, an Overseas Research Fellowship (grant # 693) from the Japan Society for the Promotion of Science (JSPS) to TN, a Grant-in-Aid for Young Scientists B (grant # JP16K21199, JP26870415) from the JSPS to TW, and a Grant-in-Aid for Scientific Research (B) (grant # JP26292176) from the JSPS to TM.

**Data availability**

Not applicable.

668 **Figure Legends**

669

670 **Figure 1. *Gb-laccase2* knock-out $G_0$ and $G_1$ phenotypes.**

671 (A) Cuticle of adult wild type (WT) *G. bimaculatus* is uniformly dark brown or black. Head and
672 anterior thorax shown in dorsal view. (A') *Gb-lac2* gene somatic mutagenesis in $G_0$ animals can
673 be detected by the presence of white spots of cuticle (asterisks). (B) Representative cuticle
674 phenotypes of $G_2$ homozygous mutant nymphs (B') at one day after hatching. Control first instar
675 nymphs (B) have dark melanized cuticle by one day after hatching, while homozygous mutants
676 (B') showed homogeneous pale brown cuticle even at one day after hatching. Scale bar:1 mm. (C)
677 T7 endonuclease I assay result. The control experiment, in which the PCR product was amplified
678 from the genome of a wild type control individual, did not produce any band of the expected size
679 after T7 endonuclease I treatment (white arrowhead). In contrast, the PCR product amplified from
680 the genome of CRISPR reagent-injected animals included small fragments of the expected size
681 after T7 endonuclease I treatment (black arrowhead). (D) Sequence analysis of *Gb-lac2* mutant
682 alleles #6-2 and #1-10 induced by CRISPR/Cas9 system. Top row: wild-type sequences; green:
683 Protospacer Adjacent Motif (PAM) sequence; red: target sequence; and arrowheads: predicted
684 double strand break site. Asterisks at right show induced frame-shift mutations.

685

686 **Figure 2. Knock-out vs knock-down phenotype of *Gb-Ubx*.**

687 (A) Schematic diagram of *Ubx* locus. White boxes: exons; red box: homeodomain; black
688 arrowhead: sgRNA target site. (B) Surveyor$^{TM}$ assay with G0 eggs. Plus (+) indicates the PCR
689 products digested by Surveyor$^{TM}$ nuclease; minus (-) indicates the PCR products with no digestion
690 (no nuclease added). L = ladder. (C) Phenotype of heterozygous and homozygous *Gb-Ubx*$^{CRISPR}$
691 mutant stage X embryos. Anterior is to the left here and in all other figures. The size of the T3 leg
692 (region between white arrowheads) was decreased mildly and severely in heterozygous and
693 homozygous mutants, respectively. Anterior white arrowhead marks the posterior end of T3
694 segment. Posterior white arrowhead marks the junction of femur and tibia. Asterisk indicates the
695 A1 segment. (D) Ubx/Abd-A (UbdA) protein expression pattern in homozygous *Gb-Ubx*$^{CRISPR}$
696 stage X embryos and *Gb-Ubx*$^{RNAi}$ stage X embryos. In *Gb-Ubx*$^{CRISPR}$ embryos, only the T3/A1
697 Gb-Ubx expression domain was undetectable. (E) In stage X *Gb-Ubx*$^{RNAi}$ embryos, UbdA protein
698 expression was undetectable in the T3 leg (L3) but was still detected in the A1 segment. Scale bar:
699 100 μm. Pp: pleuropodium. Embryonic staging as per (Donoughe and Extavour, 2016).

**Figure 3. Exonic knock-in/Knock-out against *Gb-Ubx*.**

(A) Scheme of knock-in experiment against *Gb-Ubx*. White box: exons; red box: homeodomain; black arrowhead: sgRNA target site. Donor vector contains bait sequence (yellow box) and expression cassette with the following elements: *Gryllus actin* promoter (blue arrow) followed by *eGFP* coding sequence (green box), and flanking *Ars* insulators (gray arrow). After co-injection of the donor vector with sgRNA for the donor vector, sgRNA for the genomic target site, and Cas9 mRNA, two patterns of insertion are predicted to occur due to NHEJ. (B) GFP expression in *Gb-Ubx*$^{KI}$ $G_0$ and $G_1$ stage X embryos. 4.7% of $G_0$ embryos showed mosaic GFP expression in the T3 legs (Table 3). In $G_1$ stage X embryos, the GFP expression pattern was identical to that of the previously reported expression pattern of *Gb-Ubx* (Barnett et al., 2019; Matsuoka et al., 2015; Zhang et al., 2005). (C) GFP expression in heterozygous and homozygous $G_2$ mutants. *Gb-Ubx*$^{KI}$ homozygous mutants showed strong GFP expression, and also showed phenotypes characteristic of *Gb-Ubx*$^{CRISPR}$ mutants, including shortened T3 legs and formation of leg-like structures on the A1 segment (FIG). Asterisks in (B) and (C) mark position of embryonic head. (D) Assessment of knock-in event by using PCR and Sanger sequencing. We designed PCR primers specific for each putative junction (black arrows flanking shaded areas in a and b of panel (A)), and all three $G_2$ individual mutant animals assayed showed bands of the expected size for each junction, suggesting that this *Gb-Ubx*$^{KI}$ line has at least two copies of donor vector insert. (E) Sequence analysis using the same primers indicated in (A) and (D) for genotyping confirmed that several deletions were generated due to the NHEJ events at each junction. Blue: genomic sequence; red: PAM sequence; pink: deleted nucleotides. Scale bar: 200 μm in (B) and (C). Embryonic staging as per (Donoughe and Extavour, 2016).

**Figure 4. Exonic knock-in/Knock-out against *Gb-abd-A*.**

(A) Scheme of knock-in experiment against *Gb-abd-A*. White box: exons; red box: homeodomain; black arrowhead: sgRNA target site. We used the same donor vector construct as that used in the experiment against *Gb-Ubx* (Fig. 3), substituting a *Gb-abd-A* exon-specific sgRNA. Two patterns of insertion are predicted to occur due to NHEJ. (B) Expression of GFP in $G_0$ and $G_1$ *Gb-abd-A*$^{KI}$ embryos. 10.6% of $G_0$ *Gb-abd-A*$^{KI}$ embryos showed mosaic GFP expression in the abdomen of stage X embryos at seven days after injection (Table 3). In $G_1$ *Gb-abd-A*$^{KI}$ stage X embryos, the expression pattern of GFP was identical to the previously reported expression pattern of *Gb-abd-*

*A* (Barnett et al., 2019; Matsuoka et al., 2015; Zhang et al., 2005). (C) Assessment of knock-in event by using PCR and Sanger sequencing. Genomic DNA was extracted from the seven day $G_2$ *Gb-abd-A*$^{KI}$ stage X embryos and used as a PCR template. We designed PCR primers specific for each putative junction (black arrows and shaded regions a, b and c, d in panel (A)). The expected amplicon size was detected for each junction. (D) Sequence analysis using primers indicated in (A) and (C) for genotyping confirmed that multiple deletions or insertions ere generated due to the NHEJ events at each junction. Blue: genomic sequence; red: PAM sequence; pink: deleted or inserted nucleotides. Scale bar: 200 μm in (B) and (C). Embryonic staging as per (Donoughe and Extavour, 2016).

**Figure 5. Intronic knock-in against *Gb-abd-A*.**

(A) Scheme of knock-in experiment targeted to a *Gb-abd-A* intron. White boxes: exons; red box: homeodomain; black arrowhead: sgRNA target site. We used the same donor vector construct as that used in the experiment against *Gb-Ubx* (Fig. 3), substituting a *Gb-abd-A* intron-specific sgRNA. Two patterns of insertion are predicted to occur due to NHEJ. (B) Expression pattern of GFP in $G_0$ and $G_2$ stage X KI embryonic abdomen. In *Gb-abd-A*$^{KI\text{-}exon}$ embryos, patchy GFP expression was accompanied by ectopic phenotypic leg-like structures (arrowheads; compare with Fig.4B). In *Gb-abd-A*$^{KI\text{-}intron}$ embryos, patchy GFP expression was observed but embryos did not show the ectopic abdominal appendage phenotype of Gb-abd-A$^{KI\text{-}exon}$ embryos. *Gb-abd-A*$^{KI\text{-}intron}$ $G_2$ heterozygous embryos show abdominal GFP expression corresponding to the known pattern of embryonic *Gb-abd-A* transcripts (Barnett et al., 2019; Matsuoka et al., 2015; Zhang et al., 2005), and the embryos did not show any morphological abnormality. *Gb-abd-A*$^{KI\text{-}intron}$ $G_2$ homozygous embryos generated ectopic leg-like structures on the abdomen (white arrowheads) as observed in $G_0$ *Gb-abd-A*$^{CRISPR}$ embryos (compare with (B')). Scale bar: 500 μm. Embryonic staging as per (Donoughe and Extavour, 2016).

## Tables

### Table 1. *G. bimaculatus* genes disrupted by targeted genome modification

| *Gene* | *Functions* | *Tissue distribution* | *Phenotype of Knockout/Knockdown cricket* | *Refs* |
|---|---|---|---|---|
| *Laccase 2* | Phenol oxidase, cuticle tuning (sclerotization and pigmentation) | ND | *Lac2* knock-out nymphs show defect in pigmentation. | (Watanabe et al., 2014); this study |
| *Ultrabithorax* | Enlargement of T3 leg<br>Identification of A1 pleuropodia | T3 and A1 segment | - *Ubx* knock-out embryos show partial transformation of A1 pleuropodia into T3 thoracic leg, and of T3 thoracic leg into T2 thoracic leg.<br>- Embryonic lethal. | This study |
| *abdominal-A* | Repression of leg formation in abdomen | Abdomen | - *abd-A* knock-out embryos show generation of leg-like structures on the abdomen.<br>- *abd-A* knock-out nymphs show fusion of abdominal segments.<br>- *abd-A* knock-out female adults generated ectopic ovipositors and had defects in oviducts and uterus attachment. | This study |

**Table 2. sgRNA sequences used in this study**

| ***sgRNA name*** | ***sgRNA sequence (5'->3') (<u>Bold underline</u> indicates PAM sequence)*** |
|---|---|
| *Gb-lac2* exon | GGGGTCCTGGCCCGGGTTGA**<u>CGG</u>** |
| *Gb-Ubx* exon | GGGTAGAAGGTGTGGTTGGC**<u>GGG</u>** |
| *Gb-Ubx* intron | GGACTGGCCACGCTCCAAGG**<u>AGG</u>** |
| *Gb-abd-A* exon | GGGGCAAGGCTCACCCGTGA**<u>TGG</u>** |
| *Gb-abd-A* intron | GCTCGCGGTGTTTTACGGCT**<u>GGG</u>** |

**Table 3. Efficiency of CRISPR/Cas9-mediated genome editing in *G. bimaculatus***

| | *Shown in Figure* | *# eggs injected* | *# injected embryos with GFP expression or phenotype* | *# embryos developed by 7d AEL (% of injected embryos)* | *# hatched nymphs (% of embryos developed by 7d AEL)* | *# fertile adults (% of nymphs hatched)* | *% injected embryos yielding fertile adults* | *# fertile adults showing germ line transmission* | *% fertile adults showing germ line transmission* | *% injected embryos yielding fertile adults showing germline transmission* |
|---|---|---|---|---|---|---|---|---|---|---|
| *Gb-Lac2*$^{CRISPR}$ KO | 1 | 128 | nd | nd | nd | 18 (nd) | 14.1% | 14 | 77.8% | 10.9% |
| *Gb-Ubx*$^{CRISPR}$ KO | 2, S1, S2 | 167 | nd | nd | nd | 10 (nd) | 59.9% | 6 | 60.0% | 3.6% |
| *Gb-Ubx*$^{KI}$ exon | 3, S3, S4 | 85 | 4 (4.7%) | 58 (68.2%) | 30 (51.7%) | 25 (83.3%) | 29.4% | 1 | 4.0% | 1.2% |
| *Gb-abd-A*$^{KI}$ exon *1st trial* | 4, S2, S3, S4, S5, S6 | 47 | 5 (10.6%) | 38 (80.8%) | 9 (23.7%) | 4 (44.4%) | 8.5% | 1 | 25.0% | 2.1% |
| *Gb-abd-A*$^{KI}$ exon *2nd trial* | n/a | 41 | 0 (0.0%) | 36 (87.8%) | 22 (61.1%) | 2 (9.1%) | 4.9% | 1 | 50.0% | 2.4% |
| *Gb-abd-A*$^{KI}$ intron | 5, S5, S6 | 100 | 2 (2.0%) | 77 (77.0%) | 47 (61.0%) | 28 (59.6%) | 28.0% | 1 | 3.6% | 1.0% |
| *Gb-Ubx* KI intron | S7 | 73 | 5 (6.8%) | 62 (84.9%) | 59 (95.2%) | 22 (37.3%) | 30.1% | 2 | 9.1% | 2.7% |

| | | | | |
|---|---|---|---|---|
| Control (Donoughe and Extavour, 2016) | n/a | 42 | 40 (95.2%) | 28 (70.0%) |
| Control (Ewen-Campen et al., 2012) | n/a | 78 | | 64 (82.0%) |

**Table 4. Primers used in this study.**

| ***Primer name*** | ***Primer seq (5'->3')*** |
|---|---|
| *Gb-Ubx* exon genotype KO Fw | CGTTTGTGAAACGTATGGCCCGTTA |
| *Gb-Ubx* exon genotype KO Rv | GTCCCTGGGCTCCTGGAACACG |
| *Gb-Ubx* exon genotype KI genome Fw | AACACGTGCTCCCTCAACTC |
| *Gb-Ubx* exon genotype KI genome Rv | TGAAACGTATGGCCCGTTAT |
| *Gb-Ubx* exon genotype KI vector 5' Rv | GTCGCATGCTCCTCTAGACTCG |
| *Gb-Ubx* intron genotype Fw | GCAGAACCGTTTCATGAATGT |
| *Gb-Ubx* intron genotype Rv | ATTCTCGCCCTTATGCAGAG |
| *Gb-abd-A* exon genotype Fw | CCGATTCCATGGTGAACTA |
| *Gb-abd-A* exon genotype Rv | AGAACGGAACGCAGTGAGTTAG |
| *Gb-abd-A* exon genotype vector 3' | GAACTTCAGGGTCAGCTTGC |
| *Gb-abd-A* exon genotype vector 5' | CACAAGGCACAAATGCTCGT |
| *Gb-abd-A* intron Fw | CGGATCTATTCGGCCATTT |
| *Gb-abd-A* intron Rv | TCAAACGGATCTTCCTCTCG |
| *Gb-otd* qPCR Fw | CATTCACGTCTCCGCCATAC |
| *Gb-otd* qPCR Rv | GCTCCATCAACAGGCAAACA |
| *eGFP* qPCR Fw | CAGAAGAACGGCATCAAGGT |
| *eGFP* qPCR Rv | GGGTGCTCAGGTAGTGGTTG |

## References

**Amitai, G. and Sorek, R.** (2016). CRISPR–Cas adaptation: insights into the mechanism of action. *Nat Rev Microbiol* **14**, 67–76.

**Arakane, Y., Muthukrishnan, S., Beeman, R. W., Kanost, M. R. and Kramer, K. J.** (2005). Laccase 2 is the phenoloxidase gene required for beetle cuticle tanning. *P Natl Acad Sci Usa* **102**, 11337–11342.

**Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., et al.** (2000). Gene Ontology: tool for the unification of biology. *Nat Genet* **25**, 25–29.

**Auer, T. O., Duroure, K., Cian, A. D., Concordet, J.-P. and Bene, F. D.** (2014). Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res* **24**, 142–153.

**Baird, G. S., Zacharias, D. A. and Tsien, R. Y.** (2000). Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proc National Acad Sci* **97**, 11984–11989.

**Bando, T., Ishimaru, Y., Kida, T., Hamada, Y., Matsuoka, Y., Nakamura, T., Ohuchi, H., Noji, S. and Mito, T.** (2013). Analysis of RNA-Seq data reveals involvement of JAK/STAT signalling during leg regeneration in the cricket Gryllus bimaculatus. *Development* **140**, 959–964.

**Barnett, A. A., Nakamura, T. and Extavour, C. G.** (2019). Hox genes limit germ cell formation in the short germ insect Gryllus bimaculatus. *Proc National Acad Sci* **116**, 16430–16435.

**Barry, S. K., Nakamura, T., Matsuoka, Y., Straub, C., Horch, H. W. and Extavour, C. G.** (2019). Injecting Gryllus bimaculatus Eggs. *J Vis Exp*.

**Bosch, J. A., Colbeth, R., Zirin, J. and Perrimon, N.** (2019). Gene Knock-Ins in Drosophila Using Homology-Independent Insertion of Universal Donor Plasmids. *Genetics* **214**, genetics.302819.2019.

**Branzei, D. and Foiani, M.** (2008). Regulation of DNA repair throughout the cell cycle. *Nat Rev Mol Cell Bio* **9**, 297–308.

**Brown, J. B., Boley, N., Eisman, R., May, G. E., Stoiber, M. H., Duff, M. O., Booth, B. W., Wen, J., Park, S., Suzuki, A. M., et al.** (2014). Diversity and dynamics of the Drosophila transcriptome. *Nature* **512**, 393–399.

**Carballar-Lejarazú, R., Jasinskiene, N. and James, A. A.** (2013). Exogenous gypsy insulator sequences modulate transgene expression in the malaria vector mosquito, Anopheles stephensi. *Proc National Acad Sci* **110**, 7176–7181.

**Consortium, T. G. O., Carbon, S., Douglass, E., Good, B. M., Unni, D. R., Harris, N. L., Mungall, C. J., Basu, S., Chisholm, R. L., Dodson, R. J., et al.** (2020). The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res* **49**, D325–D334.

**Dabour, N., Bando, T., Nakamura, T., Miyawaki, K., Mito, T., Ohuchi, H. and Noji, S.** (2011). Cricket body size is altered by systemic RNAi against insulin signaling components and epidermal growth factor receptor. *Dev Growth Differ* **53**, 857–869.

**Daimon, T., Uchibori, M., Nakao, H., Sezutsu, H. and Shinoda, T.** (2015). Knockout silkworms reveal a dispensable role for juvenile hormones in holometabolous life cycle. *Proc National Acad Sci* **112**, E4226–E4235.

**Davis, G. K. and Patel, N. H.** (2002). SHORT, LONG, AND BEYOND: Molecular and Embryological Approaches to Insect Segmentation. *Annual Review of Entomology* **47**, 669–699.

**Donoughe, S. and Extavour, C. G.** (2016). Embryonic development of the cricket Gryllus bimaculatus. *Dev Biol* **411**, 140–156.

**Donoughe, S., Nakamura, T., Ewen-Campen, B., Green, D. A., Henderson, L. and Extavour, C. G.** (2014). BMP signaling is required for the generation of primordial germ cells in an insect. *Proc National Acad Sci* **111**, 4133–4138.

**Epper, F.** (1983). The evagination of the genital imaginal discs ofDrosophila melanogaster. *Wilhelm Roux's Archives Dev Biology* **192**, 275–279.

**Ewen-Campen, B., Srouji, J. R., Schwager, E. E. and Extavour, C. G.** (2012). oskar Predates the Evolution of Germ Plasm in Insects. *Curr Biol* **22**, 2278–2283.

**Foe, V. E. and Alberts, B. M.** (1983). Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis. *Journal of Cell Science* **61**, 31–70.

**Foronda, D., Estrada, B., Navas, L. de and Sánchez-Herrero, E.** (2006). Requirement of abdominal-A and Abdominal-B in the developing genitalia of Drosophila breaks the posterior downregulation rule. *Development* **133**, 117–127.

**Freeland, D. E. and Kuhn, D. T.** (1996). Expression patterns of developmental genes reveal segment and parasegment organization of D. melanogaster genital discs. *Mech Develop* **56**, 61–72.

**Gilles, A. F., Schinko, J. B. and Averof, M.** (2015). Efficient CRISPR-mediated gene targeting and transgene replacement in the beetle Tribolium castaneum. *Development* **142**, 2832–2839.

**Goldstein, B. and King, N.** (2016). The Future of Cell Biology: Emerging Model Organisms. *Trends Cell Biol* **26**, 818–824.

**Gratz, S. J., Cummings, A. M., Nguyen, J. N., Hamm, D. C., Donohue, L. K., Harrison, M. M., Wildonger, J. and O'Connor-Giles, K. M.** (2013). Genome Engineering of Drosophila with the CRISPR RNA-Guided Cas9 Nuclease. *Genetics* **194**, 1029–1035.

**Hagmann, M., Bruggmann, R., Xue, L., Georgiev, O., Schaffner, W., Rungger, D., Spaniol, P. and Gerster, T.** (1998). Homologous Recombination and DNA-End Joining Reactions in

Zygotes and Early Embryos of Zebrafish (Danio rerio) and Drosophila melanogaster. *Biol Chem* **379**, 673–682.

**Hedwig, B. and Sarmiento-Ponce, E. J.** (2017). Song pattern recognition in crickets based on a delay-line and coincidence-detector mechanism. *Proc Royal Soc B* **284**, 20170745.

**Hendel, A., Bak, R. O., Clark, J. T., Kennedy, A. B., Ryan, D. E., Roy, S., Steinfeld, I., Lunstad, B. D., Kaiser, R. J., Wilkens, A. B., et al.** (2015). Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat Biotechnol* **33**, 985–989.

**Horch, H. Wi., Mito, T., Popadic, A., Ohuchi, H. and Noji, S.** (2017a). *The Cricket as a Model Organism, Development, Regeneration, and Behavior*.

**Horch, H., Liu, J., Mito, T., Popadić, A. and Watanabe, T.** (2017b). Protocols in the Cricket. In *The Cricket as a Model Organism* (ed. Horch, H.), Mito, T.), Popadic, A.), Hideyo), Ohuchi), and Noji, S.), pp. 327–370.

**Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., et al.** (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827–832.

**Huis, A. van, Itterbeeck, J. V., Klunder, H., Mertens, E., Halloran, A., Muir, G. and Vantomme, P.** (2013). *Edible insects: Future prospects for food and feed security*. Rome: Food and Agriculture Organization of the United Nations.

**Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpentier, E.** (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821.

**Kelsh, R., Weinzierl, R. O. J., White, R. A. H. and Akam, M.** (1994). Homeotic gene expression in the locust Schistocerca: An antibody that detects conserved epitopes in ultrabithorax and abdominal-A proteins. *Dev Genet* **15**, 19–31.

**Kimura, Y., Hisano, Y., Kawahara, A. and Higashijima, S.** (2014). Efficient generation of knock-in transgenic zebrafish carrying reporter/driver genes by CRISPR/Cas9-mediated genome engineering. *Sci Rep-uk* **4**, 6545.

**Kistler, K. E., Vosshall, L. B. and Matthews, B. J.** (2015). Genome Engineering with CRISPR-Cas9 in the Mosquito Aedes aegypti. *Cell Reports* **11**, 51–60.

**Kulkarni, A. and Extavour, C. G.** (2019). Chapter 8: The Cricket Gryllus bimaculatus: Techniques for Quantitative and Functional Genetic Analyses of Cricket Biology. In *Evo-Devo: Non-model Species in Cell and Developmental Biology* (ed. Twozydlo, W.) and Bilinski, S. M.), pp. 183–216.

**Kulski, J. K.** (2016). Next Generation Sequencing - Advances, Applications and Challenges.

**Li, X., Fan, D., Zhang, W., Liu, G., Zhang, L., Zhao, L., Fang, X., Chen, L., Dong, Y., Chen, Y., et al.** (2015). Outbred genome sequencing and CRISPR/Cas9 gene editing in butterflies. *Nat Commun* **6**, 8212.

**Ma, S., Chang, J., Wang, X., Liu, Y., Zhang, J., Lu, W., Gao, J., Shi, R., Zhao, P. and Xia, Q.** (2014). CRISPR/Cas9 mediated multiplex genome editing and heritable mutagenesis of BmKu70 in Bombyx mori. *Sci Rep-uk* **4**, 4489.

**Mahfooz, N. S., Li, H. and Popadić, A.** (2004). Differential expression patterns of the hox gene are associated with differential growth of insect hind legs. *P Natl Acad Sci Usa* **101**, 4877–4882.

**Mahfooz, N., Turchyn, N., Mihajlovic, M., Hrycaj, S. and Popadić, A.** (2007). Ubx Regulates Differential Enlargement and Diversification of Insect Hind Legs. *Plos One* **2**, e866.

**Martin, A., Serano, J. M., Jarvis, E., Bruce, H. S., Wang, J., Ray, S., Barker, C. A., O'Connell, L. C. and Patel, N. H.** (2016). CRISPR/Cas9 Mutagenesis Reveals Versatile Roles of Hox Genes in Crustacean Limb Specification and Evolution. *Curr Biol* **26**, 14–26.

**Matsumoto, Y., Matsumoto, C. S. and Mizunami, M.** (2018). Signaling Pathways for Long-Term Memory Formation in the Cricket. *Front Psychol* **9**, 1014.

**Matsuoka, Y. and Monteiro, A.** (2018). Melanin Pathway Genes Regulate Color and Morphology of Butterfly Wing Scales. *Cell Reports* **24**, 56–65.

**Matsuoka, Y., Bando, T., Watanabe, T., Ishimaru, Y., Noji, S., Popadić, A. and Mito, T.** (2015). Short germ insects utilize both the ancestral and derived mode of Polycomb group-mediated epigenetic silencing of Hox genes. *Biol Open* **4**, 702–709.

**McMeniman, C. J., Corfas, R. A., Matthews, B. J., Ritchie, S. A. and Vosshall, L. B.** (2014). Multimodal Integration of Carbon Dioxide and Other Sensory Cues Drives Mosquito Attraction to Humans. *Cell* **156**, 1060–1071.

**Miller, D. E., Cook, K. R. and Hawley, R. S.** (2019). The joy of balancers. *Plos Genet* **15**, e1008421.

**Minelli, A.** (2016). Evo-Devo and Phylogenetics. In *Evolutionary Developmental Biology*, pp. 1–12.

**Mito, T. and Noji, S.** (2008). The Two-Spotted Cricket Gryllus bimaculatus: An Emerging Model for Developmental and Regeneration Studies. *Cold Spring Harb Protoc* **2008**, pdb.emo110-pdb.emo110.

**Mito, T., Inoue, Y., Kimura, S., Miyawaki, K., Niwa, N., Shinmyo, Y., Ohuchi, H. and Noji, S.** (2002). Involvement of hedgehog, wingless, and dpp in the initiation of proximodistal axis formation during the regeneration of insect legs, a verification of the modified boundary model. *Mech Develop* **114**, 27–35.

**Mito, T., Kobayashi, C., Sarashina, I., Zhang, H., Shinahara, W., Miyawaki, K., Shinmyo, Y., Ohuchi, H. and Noji, S.** (2007). even-skipped has gap-like, pair-rule-like, and segmental functions in the cricket Gryllus bimaculatus, a basal, intermediate germ insect (Orthoptera). *Dev Biol* **303**, 202–213.

**Mito, T., Ronco, M., Uda, T., Nakamura, T., Ohuchi, H. and Noji, S.** (2008). Divergent and conserved roles of extradenticle in body segmentation and appendage formation, respectively, in the cricket Gryllus bimaculatus. *Dev Biol* **313**, 67–79.

**Mizunami, M. and Matsumoto, Y.** (2017). Roles of Octopamine and Dopamine Neurons for Mediating Appetitive and Aversive Signals in Pavlovian Conditioning in Crickets. *Front Physiol* **8**, 1027.

**Modolell, J., Bender, W. and Meselson, M.** (1983). Drosophila melanogaster mutations suppressible by the suppressor of Hairy-wing are insertions of a 7.3-kilobase mobile element. *Proc National Acad Sci* **80**, 1678–1682.

**Nakamura, T., Mito, T., Bando, T., Ohuchi, H. and Noji, S.** (2007). Molecular and Cellular Basis of Regeneration and Tissue Repair. *Cell Mol Life Sci* **65**, 64.

**Nakamura, T., Mito, T., Miyawaki, K., Ohuchi, H. and Noji, S.** (2008). EGFR signaling is required for re-establishing the proximodistal axis during distal leg regeneration in the cricket Gryllus bimaculatus nymph. *Dev Biol* **319**, 46–55.

**Nakamura, T., Yoshizaki, M., Ogawa, S., Okamoto, H., Shinmyo, Y., Bando, T., Ohuchi, H., Noji, S. and Mito, T.** (2010). Imaging of Transgenic Cricket Embryos Reveals Cell Movements Consistent with a Syncytial Patterning Mechanism. *Curr Biol* **20**, 1641–1647.

**Nandchahal, N.** (1972). Reproductive organs of Gryllodes sigillatus (Walker) (Orthoptera: Gryllidae). *J Nat Hist* **6**, 125–131.

**Niwa, N., Inoue, Y., Nozawa, A., Saito, M., Misumi, Y., Ohuchi, H., Yoshioka, H. and Noji, S.** (2000). Correlation of diversity of leg morphology in Gryllus bimaculatus (cricket) with divergence in dpp expression pattern during leg development. *Development* **127**, 4373–4381.

**Papageorgiou, L., Eleni, P., Raftopoulou, S., Mantaiou, M., Megalooikonomou, V. and Vlachakis, D.** (2018). Genomic big data hitting the storage bottleneck. *Embnet J* **24**, e910.

**Pinello, L., Canver, M. C., Hoban, M. D., Orkin, S. H., Kohn, D. B., Bauer, D. E. and Yuan, G.-C.** (2016). Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat Biotechnol* **34**, 695–697.

**Qiu, P., Shandilya, H., D'Alessio, J. M., O'Connor, K., Durocher, J. and Gerard, G. F.** (2004). Mutation detection using Surveyor nuclease. *Biotechniques* **36**, 702–707.

**Rathke, H.** (1844). Entwickelungsgeschichte der Maulwurfsgrille (Gryllotalpa vulgaris). *Archiv für Anatomie, Physiologie und Wissenschaftliche Medicin* 27–37.

**Ren, X., Yang, Z., Xu, J., Sun, J., Mao, D., Hu, Y., Yang, S.-J., Qiao, H.-H., Wang, X., Hu, Q., et al.** (2014). Enhanced Specificity and Efficiency of the CRISPR/Cas9 System with Optimized sgRNA Parameters in Drosophila. *Cell Reports* **9**, 1151–1162.

**Russell, J. J., Theriot, J. A., Sood, P., Marshall, W. F., Landweber, L. F., Fritz-Laylin, L., Polka, J. K., Oliferenko, S., Gerbich, T., Gladfelter, A., et al.** (2017). Non-model model organisms. *Bmc Biol* **15**, 55.

**Sánchez, L. and Guerrero, I.** (2001). The development of the Drosophila genital disc. *Bioessays* **23**, 698–707.

**Sander, J. D., Zaback, P., Joung, J. K., Voytas, D. F. and Dobbs, D.** (2007). Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic Acids Res* **35**, W599–W605.

**Sander, J. D., Maeder, M. L., Reyon, D., Voytas, D. F., Joung, J. K. and Dobbs, D.** (2010). ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Res* **38**, W462–W468.

**Sarashina, I., Shinmyo, Y., Hirose, A., Miyawaki, K., Mito, T., Ohuchi, H., Horio, T. and Noji, S.** (2003). Hypotonic buffer induces meiosis and formation of anucleate cytoplasmic islands in the egg of the two-spotted cricket Gryllus bimaculatus. *Dev Growth Differ* **45**, 103–112.

**Takagi, H., Inai, Y., Watanabe, S. -i, Tatemoto, S., Yajima, M., Akasaka, K., Yamamoto, T. and Sakamoto, N.** (2011). Nucleosome exclusion from the interspecies-conserved central AT-rich region of the Ars insulator. *J Biochem* **151**, 75–87.

**Takahashi, T., Hamada, A., Miyawaki, K., Matsumoto, Y., Mito, T., Noji, S. and Mizunami, M.** (2009). Systemic RNA interference for the study of learning and memory in an insect. *J Neurosci Meth* **179**, 9–15.

**Tchurikov, N. A., Kretova, O. V., Moiseeva, E. D. and Sosin, D. V.** (2009). Evidence for RNA synthesis in the intergenic region between enhancer and promoter and its inhibition by insulators in Drosophila melanogaster. *Nucleic Acids Res* **37**, 111–122.

**Thorvaldsdóttir, H., Robinson, J. T. and Mesirov, J. P.** (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178–192.

**Tomiyama, Y., Shinohara, T., Matsuka, M., Bando, T., Mito, T. and Tomioka, K.** (2020). The role of clockwork orange in the circadian clock of the cricket Gryllus bimaculatus. *Zoological Lett* **6**, 12.

**Wainberg, M., Kamber, R. A., Balsubramani, A., Meyers, R. M., Sinnott-Armstrong, N., Hornburg, D., Jiang, L., Chan, J., Jian, R., Gu, M., et al.** (2021). A genome-wide atlas of co-essential modules assigns function to uncharacterized genes. *Nat Genet* 1–12.

**Watanabe, T., Ochiai, H., Sakuma, T., Horch, H. W., Hamaguchi, N., Nakamura, T., Bando, T., Ohuchi, H., Yamamoto, T., Noji, S., et al.** (2012). Non-transgenic genome modifications in a hemimetabolous insect using zinc-finger and TAL effector nucleases. *Nat Commun* **3**, 1017.

**Watanabe, T., Noji, S. and Mito, T.** (2014). Gene knockout by targeted mutagenesis in a hemimetabolous insect, the two-spotted cricket Gryllus bimaculatus, using TALENs. *Methods* **69**, 17–21.

**Watanabe, T., Noji, S. and Mito, T.** (2017). Genome Editing in Animals, Methods and Protocols. *Methods Mol Biology* **1630**, 219–233.

**Wheeler, W. M.** (1892). On the appendages of the first abdominal segment of embryo insects. *Transactions of the Wisconsin Academy of Sciences, Arts and Letters* **8**, 87–140.

**Xiao, A., Cheng, Z., Kong, L., Zhu, Z., Lin, S., Gao, G. and Zhang, B.** (2014). CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics* **30**, 1180–1182.

**Ylla, G., Nakamura, T., Itoh, T., Kajitani, R., Toyoda, A., Tomonari, S., Bando, T., Ishimaru, Y., Watanabe, T., Fuketa, M., et al.** (2021). Insights into the genomic evolution of insects from cricket genomes. *Communications Biology* **in press**,.

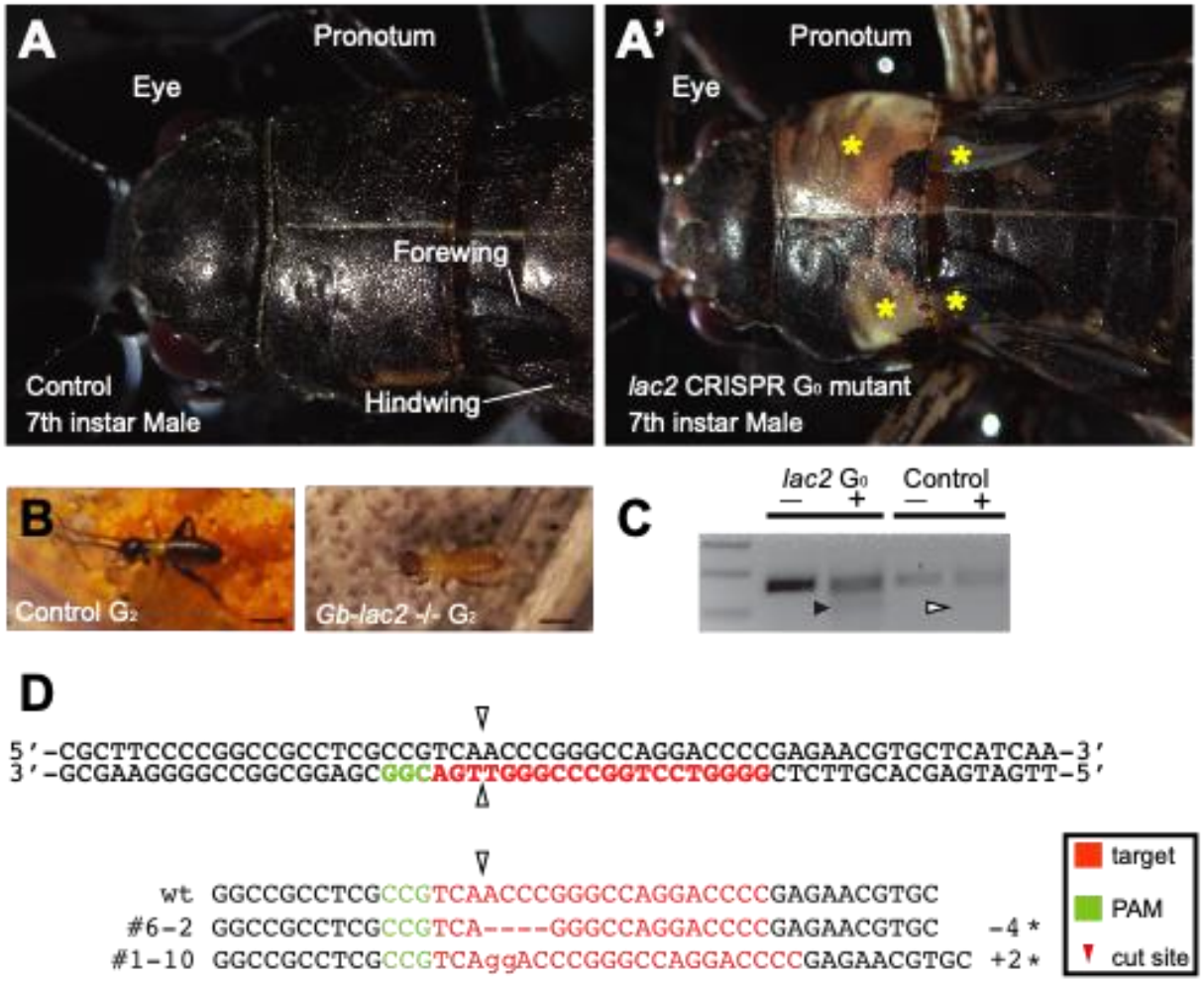
**Yoshimi, K., Kunihiro, Y., Kaneko, T., Nagahora, H., Voigt, B. and Mashimo, T.** (2016). ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes. *Nat Commun* **7**, 10431.

**Zhang, H., Shinmyo, Y., Mito, T., Miyawaki, K., Sarashina, I., Ohuchi, H. and Noji, S.** (2005). Expression patterns of the homeotic genes Scr, Antp, Ubx, and abd-A during embryogenesis of the cricket Gryllus bimaculatus. *Gene Expr Patterns* **5**, 491–502.

**Zhang, L., Martin, A., Perry, M. W., Burg, K. R. L. van der, Matsuoka, Y., Monteiro, A. and Reed, R. D.** (2017). Genetic Basis of Melanin Pigmentation in Butterfly Wings. *Genetics* **205**, 1537–1550.

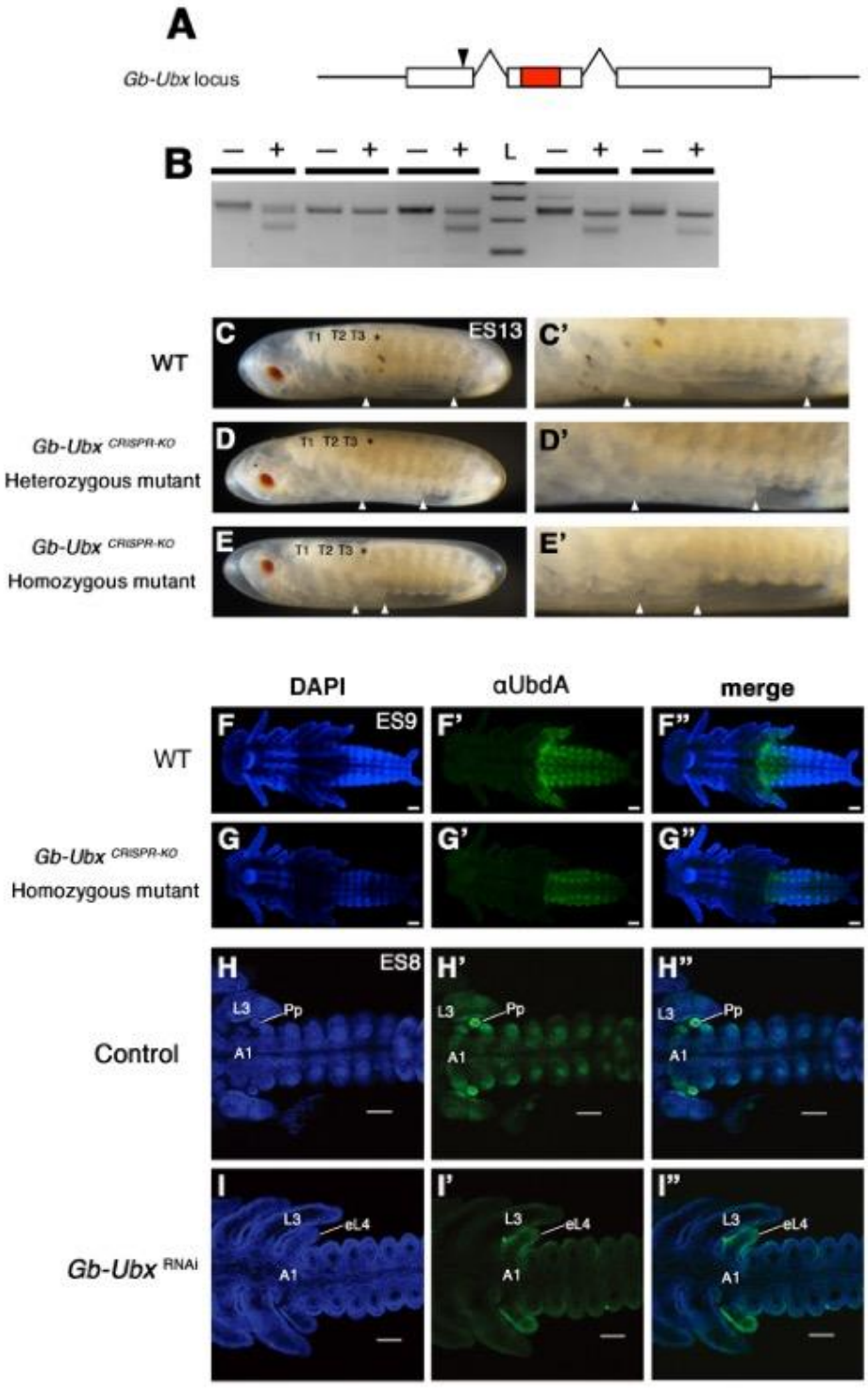
**Zhu, L., Mon, H., Xu, J., Lee, J. M. and Kusakabe, T.** (2015). CRISPR/Cas9-mediated knockout of factors in non-homologous end joining pathway enhances gene targeting in silkworm cells. *Sci Rep-uk* **5**, 18103.

**Figure 1**



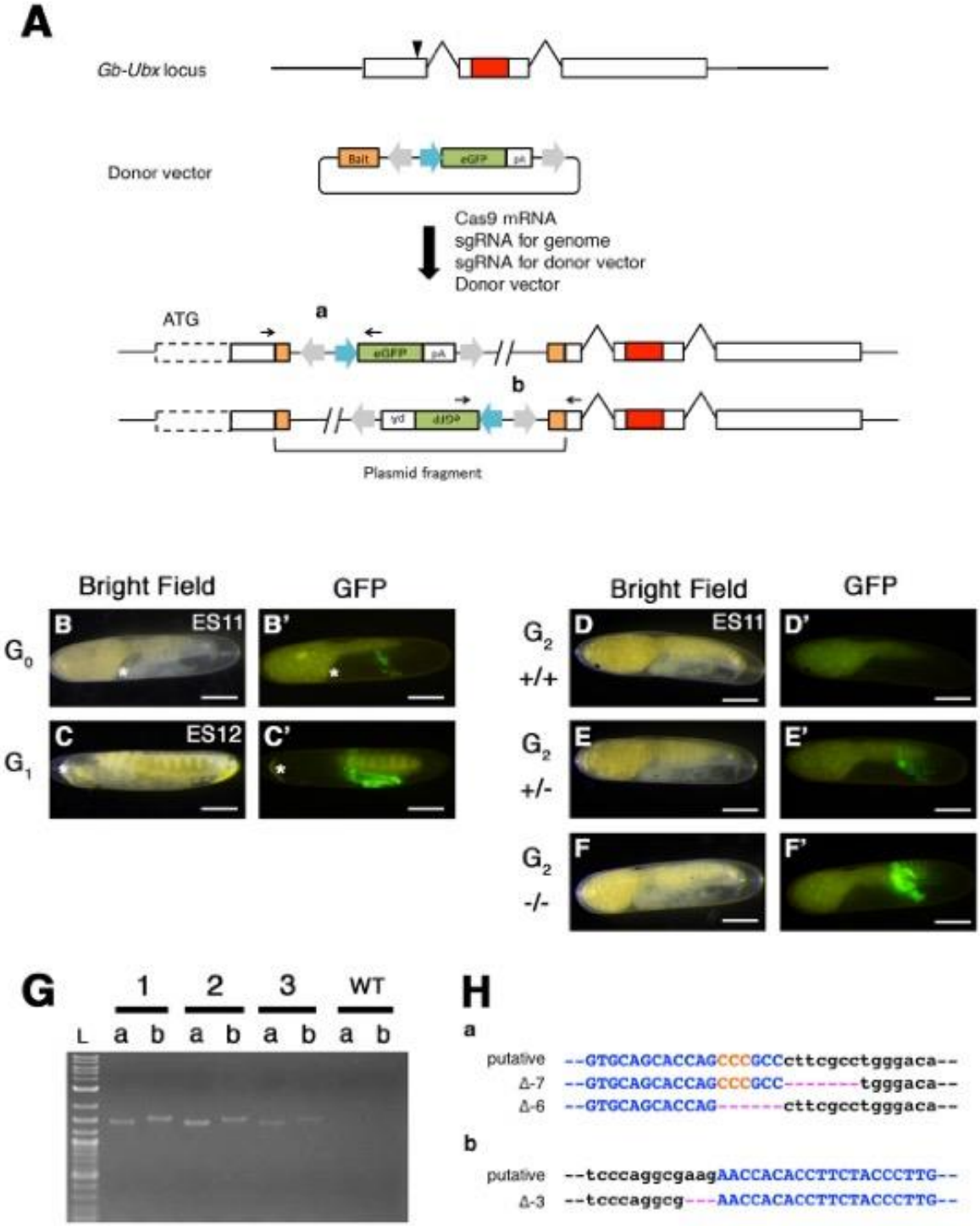
253 **Figure 2**

254




255

256 **Figure 3**




257
258

**Figure 4**

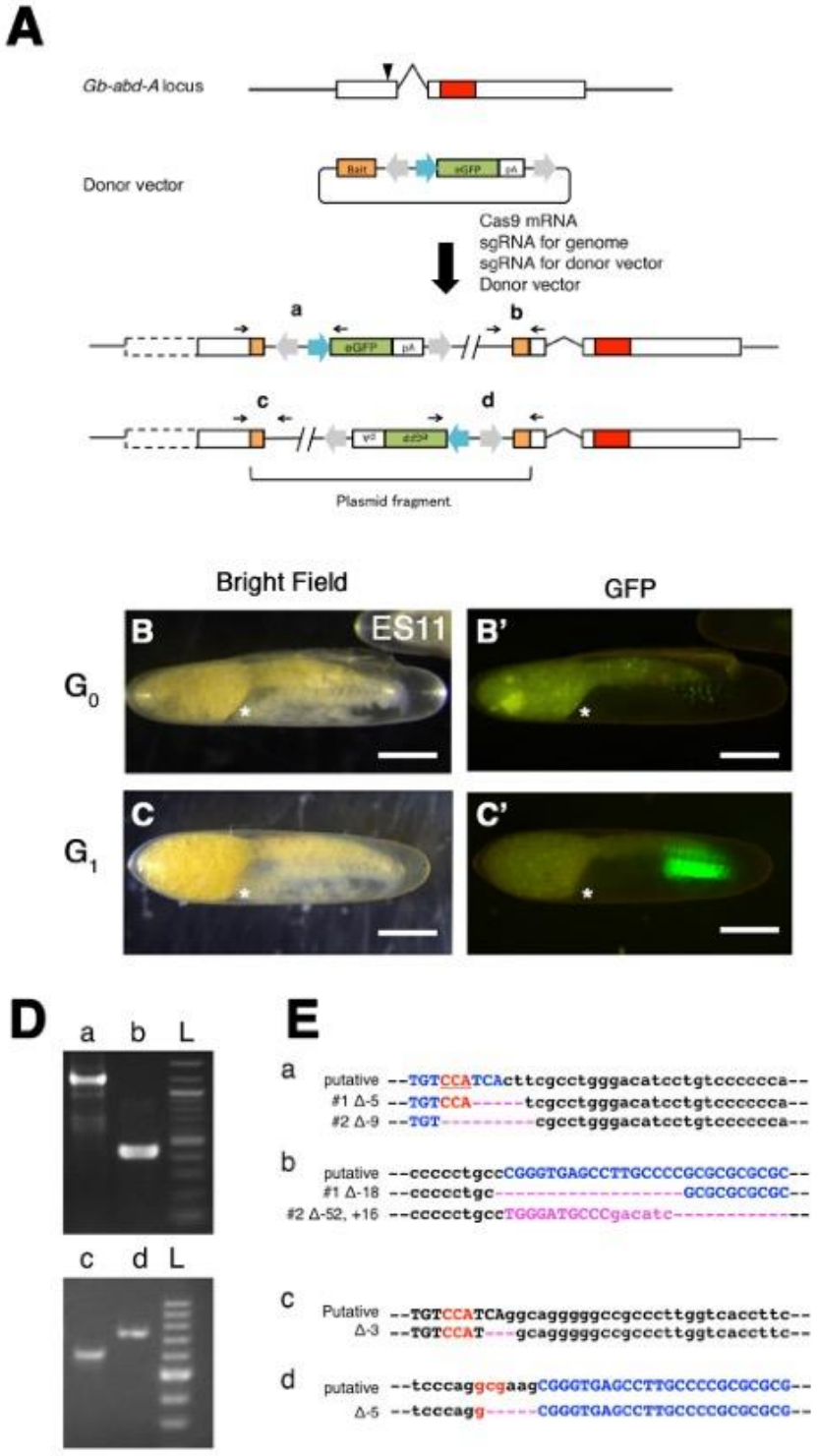
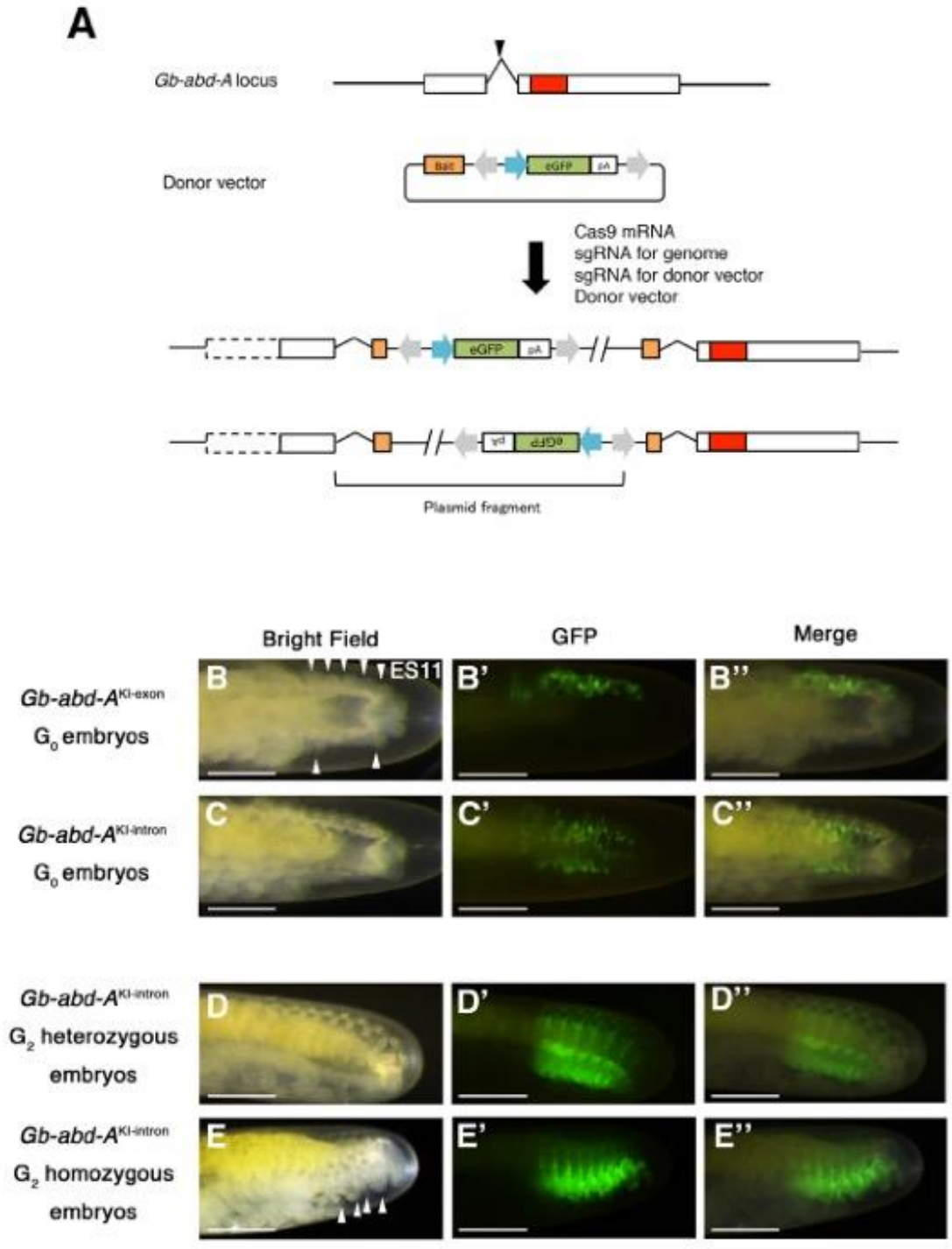


**Figure 5**



# Supplementary Information

**Establishment of CRISPR/Cas9-based knock-in in a hemimetabolous insect: targeted gene tagging in the cricket *Gryllus bimaculatus***

Yuji Matsuoka, Taro Nakamura, Takahito Watanabe, Austen A. Barnett, Taro Mito, Sumihare Noji and Cassandra G. Extavour

The Supplementary Information for this study consists of the following:

- Supplementary Results and Discussion
- Supplementary Figure Legends S1 through S7
- Supplementary References
- Supplementary Figures S1 through S7

## Supplementary Results and Discussion

*Analysis of* Gb-Ubx$^{CRISPR}$ *embryo phenotype*

To understand the genetic basis of the T3 phenotypes in the *Gb-Ubx$^{CRISPR}$* homozygous mutant, we examined the embryonic expression pattern of the leg patterning genes *wingless* (*Gb-wg*), *decapentaplegic* (*Gb-dpp*), *Distal-less* (*Gb-Dll*), and *dachshund* (*Gb-dac*). Wild type T3 legs express *Gb-Ubx* (Supplementary Fig.1A) and have T3 leg-specific expression of some leg-patterning genes. Namely, *Gb-wg* is expressed on the ventral side of each segment (Niwa et al., 2000) and on the dorsal side of the femur in the T3 leg, but not in the T1 or T2 legs (Supplementary Fig.1A). In *Gb-Ubx$^{CRISPR}$* mutants, dorsal T3 femoral expression of *Gb-wg* was undetectable, consistent with a transformation of this appendage towards an anterior fate (Supplementary Fig.1A). *Gb-dpp* is expressed in several dorsal and ventral spots in wild type developing limbs, and in the T3 leg, *Gb-dpp* is expressed in five circumferential bands (Niwa et al., 2000) (Supplementary Fig.1A). In contrast, in *Gb-Ubx$^{CRISPR}$* embryos *Gb-dpp* expression in the T3 leg resembled the expression pattern observed in wild type T1 or T2 legs (Supplementary Fig.1A). *Gb-Dll* is expressed strongly in the presumptive tarsus of developing wild type T1, T2 and T3 legs (Supplementary Fig.1A). In the tibia, however, *Gb-Dll* is differentially expressed between T1/T2 and T3 legs, having in T3 a sharply defined border between a distal domain of high expression and a proximal domain of lower expression, but lacking such a border and displaying moderate proximal expression as well as strong distal expression in T1 and T2 (Niwa et al., 1997) (Supplementary Fig.1A). In *Gb-Ubx$^{CRISPR}$* embryos, *Gb-Dll* expression in the presumptive tibia of the T3 leg lacked a strong boundary between high and low tibial expression and resembled that of wild type T1 or T2 legs (Supplementary Fig.1A). Taken together, the T3 leg-specific pattern of multiple leg-patterning genes was absent due to the disruption of the *Gb-Ubx* locus. These results suggest that the T3 segment in *Gb-Ubx$^{CRISPR}$* mutants acquired characteristics of the T2 segment, consistent with a homeotic transformation loss-of-function phenotype of *Gb-Ubx*.

The A1 appendage, called the pleuropodium, is a transient A1 appendage structure only observed during mid embryogenesis (Rathke, 1844; Wheeler, 1892), and is thought to be involved in secreting hatching enzymes (Konopová et al., 2020; Slifer, 1937). In wild type *G. bimaculatus*, *Gb-Ubx* is strongly expressed in this organ at early embryonic stages (Barnett et al., 2019; Matsuoka et al., 2015; Zhang et al., 2005), as are *Gb-Dll* and *tramtrack* (*Gb-ttk*) (Barnett et al., 2019) (Supplementary Fig.1A, Supplementary Fig.1B). In *Gb-Ubx$^{CRISPR}$* mutants, A1 appendages

appeared as small, twisted leg-like structures (Supplementary Fig.1A) that lacked *Gb-ttk* expression (Supplementary Fig.1B) and ectopically expressed *Gb-wg*, *Gb-dpp, Gb-Dll*, and *Gb-Dac* in patterns like those observed in wild type thoracic legs (Supplementary Fig.1A). We speculate that knock-out of *Gb-Ubx* may cause misexpression of other Hox genes and contribute to the *Gb-Ubx*$^{CRISPR}$ phenotype. To test this hypothesis, we examined the expression patterns of *Antennapedia* (*Antp*) and *abdominal-A* (*abd-A*) in *Gb-Ubx*$^{CRISPR}$ embryos. In wild type embryos *Gb-Antp* is not expressed in the pleuropodia, while *Gb-abd-A* is expressed in the posterior half of the A1 to A9 segments but not in the pleuropodia (Barnett et al., 2019; Matsuoka et al., 2015; Zhang et al., 2005). In *Gb-Ubx*$^{CRISPR}$ embryos, the expression pattern of *Gb-abd-A* was unchanged but *Gb-Antp* was misexpressed in the A1 appendages. We interpret these expression patterns as evidence that *Gb-Ubx* represses *Gb-Antp* expression in the pleuropodia, and the interruption of the *Gb-Ubx* locus causes pleuropodia to be transformed towards thoracic leg identity, which may be induced by misexpression of *Gb-Antp.*

*Analysis of* Gb-abd-A$^{CRISPR}$ *mutants*

We detected several different types of phenotype in the $G_0$ and $G_1$ *Gb-abd-A*$^{KI\text{-}exon}$ mutants (Supplementary Fig.5 and 6). (1) Supernumerary leg-like structures on the abdomen; (2) fusion of cuticle segments on abdomen (Supplementary Fig.6S) (3) Additional ovipositors in the heterozygous females. Supernumerary leg-like structures were observed on the abdomen of mosaic $G_0$ embryos and nymphs (Supplementary Fig.6C, D, G, H, K, L, O, P, and T). To elucidate the genetic mechanism underlying the production of the structure, we examined the expression patterns of leg patterning genes in $G_0$ *Gb-abd-A*$^{CRISPR}$ embryos.

We found that *Gb-wg* was ectopically expressed along the ventral side of the supernumerary leg-like structures, like in wild type developing limbs, generated on the A8 and A9 (Supplementary Fig.6C and D), and the pattern was retained at later stages (Supplementary Fig.6G and H). *Gb-Dll* was ectopically expressed in the supernumerary leg-like structures in the A2-A9 segments (Supplementary Fig.6K and L), which is similar to the expression pattern observed in the thoracic legs. As the embryo developed, *Gb-Dll* expression was no longer detected in the supernumerary leg-like structures in the A2 throughA7 segments, but it retained detectable in the supernumerary leg-like structures in the A8 and A9 segments (Supplementary Fig.6O and P). Taken together, misexpression of *Gb-Dll* led to the generation of leg-like structures on the early embryonic abdomen, while at later stages, the only structures expressing *Gb-wg* retained the

expression of *Gb-Dll*. The strongly affected embryos did not hatch, suggesting that the low hatching rate observed in these $G_0$ embryos (Table 3) might be due to this phenotype. Similar appendage-like structures were also observed on the abdomen of the nymphs (Supplementary Fig.6T).

Additional ovipositors were detected in $G_0$ and $G_2$ heterozygous and homozygous *Gb-abd-A*$^{KI\text{-}exon}$ females (fig). To evaluate potential effects on internal abdominal organs, we dissected adult males and females. The ovipositor and the subgenital plate are located on the A8 segment in wild type females (Supplementary Fig.6U and V). In *Gb-abd-A*$^{KI\text{-}exon}$ heterozygous females, the subgenital plate on the ventral side of the A9 was absent; instead, additional ovipositors were detected (compare Supplementary Fig.6U to Supplementary Fig.6R). *Gb-abd-A*$^{KI\text{-}exon}$ heterozygous females did not lay eggs. We dissected *Gb-abd-A*$^{KI\text{-}exon}$ heterozygous females to assess potential effects on internal organs and found that many eggs were stored inside the egg chamber (not shown), suggesting that egg production was not affected. We found that the posterior tip of the oviducts had not fused with the uterus (Supplementary Fig.6B' and B''). In *D. melanogaster iab-4* mutants, one of the *cis*-regulatory regions of *abd-A* affects gonadal development in females (Cumberledge et al., 1992). In severe cases, these mutants failed to form a gonad, and in mild cases, the gonad was formed but the ovary and oviduct junctions were not attached, and thus egg transfer from ovary to the uterus was blocked (Cumberledge et al., 1992). Taken together, these results suggest that the function of *Gb-abd-A* in the development of the adult ovary, specifically in ensuring appropriate joining of the uterus to the oviducts to allow egg release, is conserved between these *D. melanogaster* and *G. bimaculatus*.

*A putative genetic mechanism for defining leg and pleuropodia identity in* G. bimaculatus

CRISPR/Cas9 system allowed us to produce efficient and severe mutants for *Gb-Ubx* and *Gb-abd-A*, which gave us the opportunity to investigate leg patterning in *G. bimaculatus*. The genetic mechanism for leg specification and patterning has been well studied in *D. melanogaster*, and two different mechanisms for regulating *Dll* expression are important for determining the segments that possess ventral appendages. One of the mechanisms is the activation of *Dll* gene by the segment polarity genes *wingless* (*wg*) and *engrailed* (*en*) (Cohen, 1990). Although *wg* and *en* are expressed in all segments, *Dll* expression is excluded from the abdominal segments in *D. melanogaster*. The repression of *Dll* in the abdomen is achieved by the Hox genes *Ubx*, *abd-A*, and *Abd-B* (Vachon et al., 1992). Lewis and colleagues (2000) addressed the relationship between

*en* and Hox genes in the beetle*, Tribolium castaneum*. In that study, *T. castaneum abd-A* mutants showed ectopic pleuropodia throughout the abdomen resulting from ectopic *Dll* expression, and *en* expression was expanded to the distal region in the ectopic pleuropodia, indicating that *abd-A* represses both *Dll* and *en* expression in the abdomen of this beetle. In our cricket study, *Gb-abd-A*$^{CRISPR}$ mutants produced ectopic pleuropodia in the abdomen, and also produced leg-like structures on A8 and A9 FIG, which was not observed in *T. castaneum abd-A* mutants (Lewis et al., 2000). In both structures *Gb-Dll* was ectopically expressed, and the leg-like structures also expressed *Gb-wg*. We speculate that in A8-A9, disruption of *Gb-abd-A* leads to the misexpression *Gb-Dll*, and that *Gryllus Abdominal-B* (*Gb-Abd-B*), which is expressed in A8 andA9, activates *Gb-wg* expression and results in the production of ectopic partial thoracic legs. Similar misexpression of *Gb-wg* was observed in the ectopic A1 legs of *Gb-Ubx*$^{CRISPR}$ embryos. These results suggest that both *Gb-Ubx* and *Gb-abd-A* repress *Gb-wg* expression in the abdomen. We speculate that *Gb-Dll* expression alone is not sufficient, and that *Gb-wg* expression is necessary, for thoracic leg development.

*Potential roles of* Gb-abd-A *in insect genitalia development*

The genitalia have been considered as serially homologous to the ventral appendages, since several leg-patterning genes contribute to genitalia development in *D. melanogaster* (Estrada and Sanchez-Herrero, 2001). *D. melanogaster Abd-B* mutants transform male and female genitalia towards leg fate, and the transformation is accompanied by the ectopic expression of *Dll* and *Dac*, which are normally required in the legs (Estrada and Sanchez-Herrero, 2001) . *D. melanogaster abd-A* also contributes to the development of the genitalia in female, since some *D. melanogaster abd-A* mutants lack or showing abnormality in the ovaries (Foronda et al., 2006). However, the additional ovipositors generated in cricket *Gb-abd-A*$^{CRISPR}$ mutants are reminiscent of the phenotype of *D. melanogaster Abd-B* loss of function mutants, but the ectopically generated appendage in the cricket was not a leg. We speculate that *Gb-abd-A* represses some genes related to the outgrowth of appendages, potentially including *Gb-Dll*, during development of the inner genitalia. When the activity of *Gb-abd-A* is reduced or disrupted, the generated outgrowth might be defined as an ovipositor by the action of *Gb-Abd-B*. *Gb-abd-A*$^{CRISPR}$ mutant females showed disconnection of the uterus and oviduct (Supplementary Fig.5B). Similar phenotype observed in *Drosophila iab-4* mutant (Cumberledge et al., 1992). The function of *Gb-abd-A* in the development of genitalia may therefore be conserved among hemimetabolous and holometabolous insects. We

detected GFP expression in the male genitalia of *Gb-abd-A*$^{CRISPR}$ heterozygous mutants, and in the terminal filament of ovaries in *Gb-Ubx*$^{KI\text{-}exon}$ heterozygous mutants FIG. Neither of these male or female mutants showed any abnormality in fertility or external morphology FIG.

**Supplementary Figure Legends**

**Supplementary Figure S1. Expression pattern of limb patterning genes in *Gb-Ubx*$^{CRISPR}$ homozygous mutants.**

Expression pattern of limb patterning genes in wild type and *Gb-Ubx*$^{CRISPR}$ homozygous mutants. (A-A”’) In wild type (WT) embryos, the expression pattern of *Gb-Dll* was different between the T1/T2 and T3 limb buds of stage X embryos. All limb buds share an expression domain at the tip of the limb bud, but in the proximal regions, the T3 limb buds have stronger femoral expression pattern compared to the T1/T2 limb buds. In the pleuropodia (A1), *Gb-Dll* is expressed uniformly and weakly. (B-B”’) In *Gb-Ubx*$^{CRISPR}$ homozygous mutants, T3 limb buds did not show the femoral expression domain. An ectopic fourth leg (eL4) developed on the A1 segment and showed *Gb-Dll* expression at the distal tip, like the thoracic limb buds. (C-C”’) In wild type embryos, *Gb-wg* was expressed along the ventral side of appendages, and also expressed in the dorsal femur in the T3 limb bud but not in the T1 or T2 limb buds. (D-D”’) In *Gb-Ubx*$^{CRISPR}$ homozygous mutants, the dorsal expression domain of *Gb-wg* in the T3 limb bud was undetectable. In eL4, *Gb-wg* expression was only observed at the proximal region. (E-E”’) In wild type embryos, *Gb-Dac* was expressed in the prospective femur and tibia of limb buds, but not in the pleuropodia. (F-F”’) In *Gb-Ubx*$^{CRISPR}$ homozygous mutants, the expression pattern in the thoracic limb buds was not affected, but the eL4 on A1 showed *Gb-Dac* gene expression like that observed in limb buds. (G) In wild type limb buds, *Gb-dpp* is expressed in each limb segment, and also in a broad expression domain in the femur and tibia of T3 but not T1 or T2 limb buds (arrowheads). (F-F”’) (H-H”’, I) In *Gb-Ubx*$^{CRISPR}$ homozygous mutants, the T3-specific expression femur and tibia domain of *Gb-dpp* was undetectable. (H”’) In eL4, *Gb-dpp* was expressed as in the distal tip region of wild type limb buds. (J-K) Expression pattern of the pleuropodia marker gene *tramtrack* (*ttk*) in wild type and *Gb-Ubx*$^{CRISPR}$ homozygous mutants. (J, J’) *Gb-ttk* was expressed in the developing A1 proleg in wild type embryos. (K, K’) In *Gb-Ubx* homozygous mutants, *Gb-ttk* expression was undetectable in the A1 appendage. (L-O) Expression pattern of other Hox genes in wild type and *Gb-Ubx*$^{CRISPR}$ homozygous mutants. (M-M’) In wild type embryos *Gb-Antp* was expressed in the thoracic limb buds, but not in the pleuropodia. (L-L’) In *Gb-Ubx*$^{CRISPR}$ homozygous mutants, *Gb-Antp* was ectopically expressed in eL4. (N-N’, O-O’) The *Gb-abd-A* expression pattern was not affected in *Gb-Ubx*$^{CRISPR}$ homozygous mutants. Scale bar: 200 μm. Embryonic staging as per (Donoughe and Extavour, 2016).

**Figure S2. In-depth analysis of the feature of mutagenesis by CRISPR/Cas9 system in the crickets.**

(A) Schematic of injection timeline. After a 1h egg collection and 1h subsequent incubation, eggs were further incubated for 1h, 3h, 5h and 9h, respectively. (B, C) NHEJ mutation rate (indels) at both on-target and off-target sites at the *Gb-Ubx* locus. The NHEJ mutation rate decreased with increasing age of injection and was <1.3% at the studied off-target site. (D, E) NHEJ mutation rate (indels) at both on-target and off-target sites at the *Gb-abd-A* locus NHEJ mutation rate was decreased with increasing age of injection and was <1.2% at the studied off-target site. (F, G) Pattern of NHEJ mutations at each injection time point in the sequence around the CRISPR target site. Black bar indicates the site targeted for CRISPR/Cas9-induced double-stranded breaks. The pattern of NHEJ mutations was similar for all injection ages, but the rate decreased with increasing age of injection. (H-H'") Cuticular phenotype of mosaic $G_0$ hatchlings resulting from injection with Gb-Lac2 sgRNA (Fig. 1) at each tested injection time point.

**Supplementary Figure S3. GFP expression in the KI crickets.**

(A, A') GFP expression in $G_1$ *Gb-Ubx*$^{KI}$ heterozygous male nymphs was visible in the hindwings, which are on the T3 segment. (B, B') *Gb-Ubx*$^{KI}$ heterozygous adults also showed GFP expression in T3 legs. (C, C') GFP expression in *Gb-abd-A*$^{KI}$ nymphs. (C,C') In $G_0$ mosaic *Gb-abd-A*$^{KI}$ nymphs, GFP expression was detected in patchy cells of the abdominal cuticle. (D, D') In $G_1$ *Gb-abd-A*$^{KI}$ nymphs, GFP expression was observed in the cuticle of A1 to A8, which is similar to the previously reported embryonic expression pattern of *Gb-abd-A* (Barnett et al., 2019; Matsuoka et al., 2015; Zhang et al., 2005). (E,E') Higher magnification of region boxed in white in (D, D').

**Supplementary Figure S4. Copy number estimation by using quantitative RT-PCR.**

Total RNA was extracted from individual $G_2$ heterozygous *Gb-abd-A*$^{KI\ -exon}$ and *Gb-Ubx*$^{KI\ -exon}$ embryos and used for synthesis of complementary DNA, which was used as a template for quantitative RT-PCR. *eGFP* expression (A) was normalized (B) with the expression level of the *Gb-orthodenticle* (*Gb-otd*) gene (C), which is known to be present in single copy in the genome (Nakamura et al., 2010; Ylla et al., 2021). The results may indicate that the *Gb-abd-A*$^{KI\ -exon}$ line has one copy of the plasmid insertion, and that the *Gb-abd-A*$^{KI\ -exon}$ line has four copies of the

plasmid insertion. (D-D') GFP expression detected in each KI mutant. Scale bar: 500 μm. Embryonic staging as per (Donoughe and Extavour, 2016).

s

**Supplementary Figure S5. GFP expression in genitalia of KI crickets.**

(A, A') Ventral posterior view of wild type (WT) adult female abdomen, indicating cerci (ci) and single ovipositor (ov). (B, B') In *Gb-abd-A*$^{\text{KI-exon}}$ females, we detected GFP expression, visible through the ventral abdominal cuticle, in the posterior regions of oviducts. (B") Further dissection revealed that the oviducts were not fused with the uterus. (C, C') In *Gb-abd-A*$^{\text{KI-intron}}$ females, GFP expression was visible through the ventral abdominal cuticle in a domain brighter and broader than that detected in *Gb-abd-A*$^{\text{KI\_exon}}$ females, and (C") the oviducts were fused with the uterus. (D-F) Adult male viscera in ventral view made visible by removing the cuticle. (D"-F") Single testes dissected out of the abdominal cavity. In (E, E', E") *Gb-abd-A*$^{\text{KI-exon}}$ and (F, F', F") *Gb-abd-A*$^{\text{KI-intron}}$ $G_2$ adult males, ubiquitous GFP expression was observed in the testis. (G, G', G") *Gb-Ubx*$^{\text{KI-exon}}$ females showed GFP expression at the anterior tip of the ovaries. (H, H') In *Gb-Ubx*$^{\text{KI-exon}}$ males, no GFP expression was detected in the testis. Scale bar: 2 mm.

**Supplementary Figure S6. Phenotypes in *Gb-abd-A*$^{\text{CRISPR}}$ mutants.**

(A, B) Expression pattern of *Gb-wg* in ES9 wild type (WT) embryos. (C, D) Expression pattern of *Gb-wg* in ES9 *Gb-abd-A*$^{\text{CRISPR}}$ $G_2$ embryos. An ectopic thoracic leg-like pattern of *Gb-wg* was observed in the ectopic leg-like structures generated on A8 and A9. (E, F) Expression pattern of *Gb-wg* in ES13 WT embryos. (G, H) Expression pattern of *Gb-wg* in ES13 *Gb-abd-A*$^{\text{CRISPR}}$ $G_2$ embryos. *Gb-wg* expression was still detected in the ectopic leg-like structure on A8. (I, J) Expression pattern of *Gb-Dll* in ES9 WT embryos. (K, L) Expression pattern of *Gb-Dll* in ES9 *Gb-abd-A*$^{\text{CRISPR}}$ $G_2$ embryos. *Gb-Dll* was expressed in all ectopic leg-like structures generated on the abdomen, but only in A8 did *Gb-Dll* show and expression pattern similar to the pattern in developing thoracic limb buds. (M, N) Expression pattern of *Gb-wg* in ES13 WT $G_2$ embryos. (O, P) Expression pattern of *Gb-Dll* in ES13 *Gb-abd-A*$^{\text{CRISPR}}$ $G_2$ embryos. *Gb-Dll* expression was only detected in the leg-like structure on the A8. (Q, R) WT nymph abdomen. (S) Approximately 10% of $G_0$ *Gb-abd-A*$^{\text{CRISPR}}$ nymphs showed fusion of abdominal cuticle in some segments (white brackets). (T) Approximately 10% of $G_0$ *Gb-abd-A*$^{\text{CRISPR}}$ nymphs showed the leg-like structures on the abdomen. (U) $G_1$ *Gb-abd-A*$^{\text{KI-exon}}$ nymphs showed ectopic ovipositors (arrowheads). (V) GFP expression in $G_1$ *Gb-abd-A*$^{\text{KI -exon}}$ nymphs. (W) *Gb-abd-A*$^{\text{KI -exon}}$ nymphs never developed

ectopic ovipositors. (X) GFP expression in $G_1$ *Gb-abd-A*$^{KI\text{-}intron}$ nymphs. Scale bars: 500 μm in (A) through (P); 2 mm in (Q) through (T). Embryonic staging as per (Donoughe and Extavour, 2016).

**Supplementary Figure S7. Knock-in against *Gb-Ubx* intronic region.**

(A) Scheme of knock-in experiment targeted to a *Gb-Ubx* intron. White boxes: exons; red box: homeodomain; black arrowhead: sgRNA target site. We used the same donor vector construct as that used in the experiment against *Gb-Ubx* (Fig. 3), substituting a *Gb-Ubx* intron-specific sgRNA. Two patterns of insertion are predicted to occur due to NHEJ. (B, B') Expression pattern of GFP in $G_2$ *Gb-Ubx*$^{KI\text{-}intron}$ stage X embryos. (C, C') Expression pattern of GFP in $G_2$ *Gb-Ubx*$^{KI\text{-}exon}$ stage X embryos. *Gb-Ubx*$^{KI\text{-}intron}$ embryos (B') showed shorter legs than $G_2$ *Gb-Ubx*$^{KI\text{-}exon}$ embryos (C'). Scale bars: 500 μm. Embryonic staging as per (Donoughe and Extavour, 2016).

**Supplementary References**

**Barnett, A. A., Nakamura, T. and Extavour, C. G.** (2019). Hox genes limit germ cell formation in the short germ insect Gryllus bimaculatus. *Proc National Acad Sci* **116**, 16430–16435.

**Cohen, S. M.** (1990). Specification of limb development in the Drosophila embryo by positional cues from segmentation genes. *Nature* **343**, 173–177.

**Cumberledge, S., Szabad, J. and Sakonju, S.** (1992). Gonad formation and development requires the abd-A domain of the bithorax complex in Drosophila melanogaster. *Development* **115**, 395–402.

**Donoughe, S. and Extavour, C. G.** (2016). Embryonic development of the cricket Gryllus bimaculatus. *Dev Biol* **411**, 140–156.

**Estrada, B. and Sanchez-Herrero, E.** (2001). The Hox gene Abdominal-B antagonizes appendage development in the genital disc of Drosophila. *Development* **128**, 331–339.

**Foronda, D., Estrada, B., Navas, L. de and Sánchez-Herrero, E.** (2006). Requirement of abdominal-A and Abdominal-B in the developing genitalia of Drosophila breaks the posterior downregulation rule. *Development* **133**, 117–127.

**Konopová, B., Buchberger, E. and Crisp, A.** (2020). Transcriptome of pleuropodia from locust embryos supports that these organs produce enzymes enabling the larva to hatch. *Front Zool* **17**, 4.

**Lewis, D. L., DeCamillis, M. and Bennett, R. L.** (2000). Distinct roles of the homeotic genes Ubx and abd-A in beetle embryonic abdominal appendage development. *Proc National Acad Sci* **97**, 4504–4509.

**Matsuoka, Y., Bando, T., Watanabe, T., Ishimaru, Y., Noji, S., Popadić, A. and Mito, T.** (2015). Short germ insects utilize both the ancestral and derived mode of Polycomb group-mediated epigenetic silencing of Hox genes. *Biol Open* **4**, 702–709.

**Nakamura, T., Yoshizaki, M., Ogawa, S., Okamoto, H., Shinmyo, Y., Bando, T., Ohuchi, H., Noji, S. and Mito, T.** (2010). Imaging of Transgenic Cricket Embryos Reveals Cell Movements Consistent with a Syncytial Patterning Mechanism. *Curr Biol* **20**, 1641–1647.

**Niwa, N., Saitoh, M., Ohuchi, H., Yoshioka, H. and Noji, S.** (1997). Correlation between Distal-less Expression Patterns and Structures of Appendages in Development of the Two-Spotted Cricket, Gryllus bimaculatus. *Zool Sci* **14**, 115–125.

**Niwa, N., Inoue, Y., Nozawa, A., Saito, M., Misumi, Y., Ohuchi, H., Yoshioka, H. and Noji, S.** (2000). Correlation of diversity of leg morphology in Gryllus bimaculatus (cricket) with divergence in dpp expression pattern during leg development. *Development* **127**, 4373–4381.

**Rathke, H.** (1844). Entwickelungsgeschichte der Maulwurfsgrille (Gryllotalpa vulgaris). *Archiv für Anatomie, Physiologie und Wissenschaftliche Medicin* 27–37.

**Slifer, E. H.** (1937). The Origin and fate of the Membranes surrounding the Grasshopper Egg; together with some Experiments on the Source of the Hatching Enzyme. *Journal of Cell Science* **79**, 493–506.

**Vachon, G., Cohen, B., Pfeifle, C., McGuffin, M. E., Botas, J. and Cohen, S. M.** (1992). Homeotic genes of the bithorax complex repress limb development in the abdomen of the Drosophila embryo through the target gene Distal-less. *Cell* **71**, 437–450.

**Wheeler, W. M.** (1892). On the appendages of the first abdominal segment of embryo insects. *Transactions of the Wisconsin Academy of Sciences, Arts and Letters* **8**, 87–140.

**Ylla, G., Nakamura, T., Itoh, T., Kajitani, R., Toyoda, A., Tomonari, S., Bando, T., Ishimaru, Y., Watanabe, T., Fuketa, M., et al.** (2021). Insights into the genomic evolution of insects from cricket genomes. *Communications Biology* **in press**,.

**Zhang, H., Shinmyo, Y., Mito, T., Miyawaki, K., Sarashina, I., Ohuchi, H. and Noji, S.** (2005). Expression patterns of the homeotic genes Scr, Antp, Ubx, and abd-A during embryogenesis of the cricket Gryllus bimaculatus. *Gene Expr Patterns* **5**, 491–502.

**Figure S1**

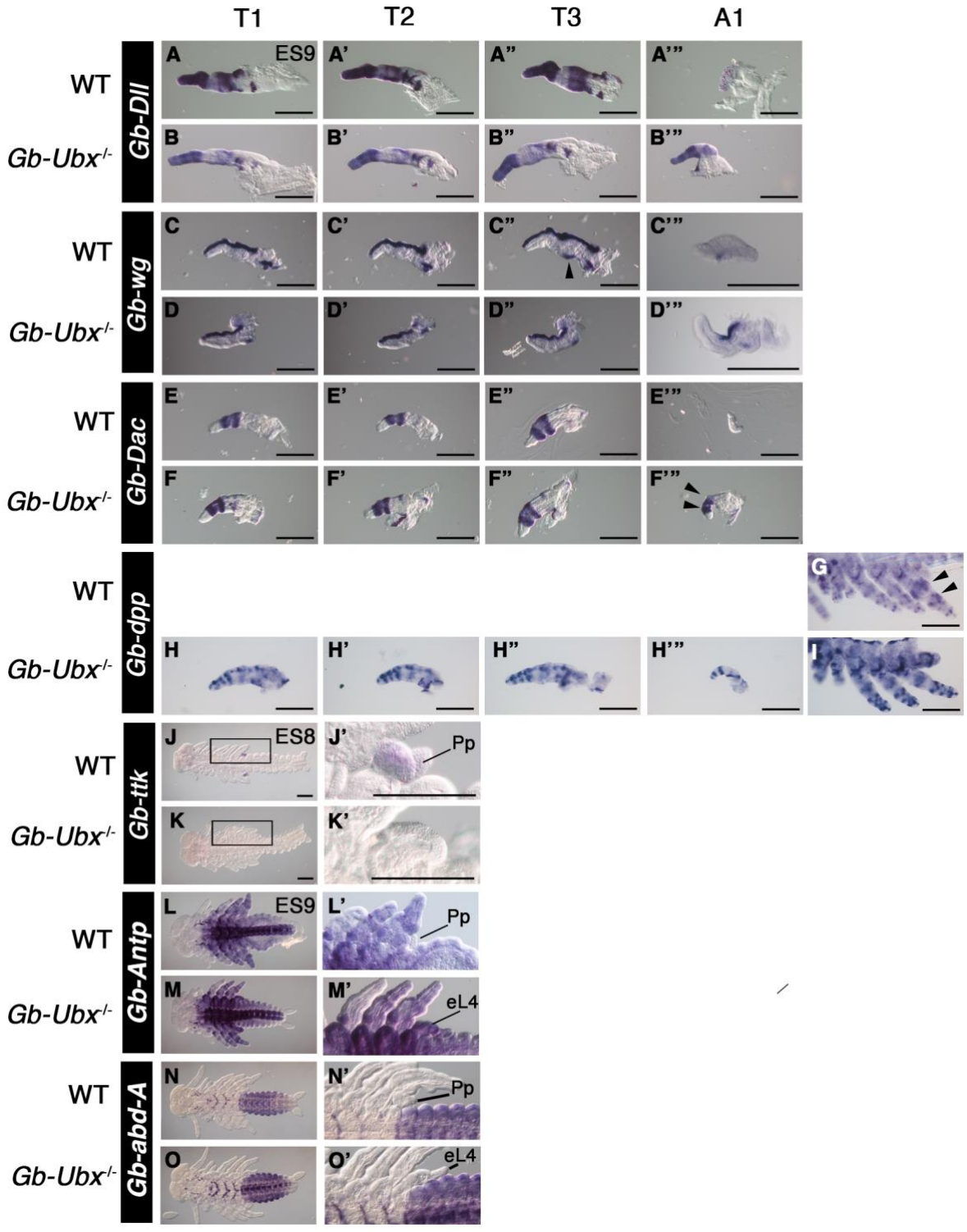


**Figure S2**

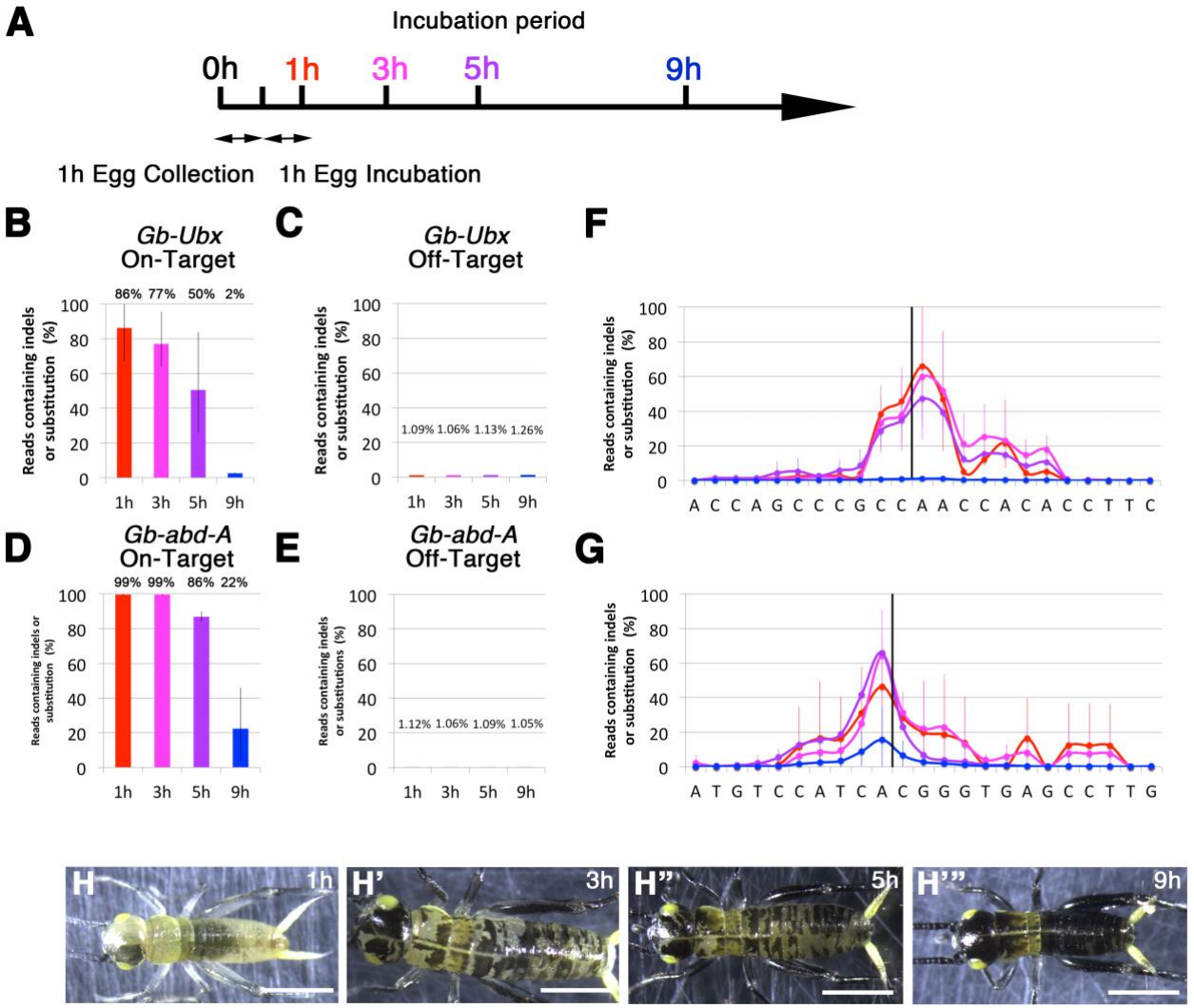

A
Incubation period
0h
1h
3h
5h
9h
1h Egg Collection
1h Egg Incubation
B
Gb-Ubx
On-Target
86% 77% 50% 2%
C
Gb-Ubx
Off-Target
1.09% 1.06% 1.13% 1.26%
F
D
Gb-abd-A
On-Target
99% 99% 86% 22%
E
Gb-abd-A
Off-Target
1.12% 1.06% 1.09% 1.05%
G
Reads containing indels or substitution (%)
1h 3h 5h 9h
A C C A G C C C G C C A A C C A C A C C T T C
A T G T C C A T C A C G G G T G A G C C T T G
H 1h
H' 3h
H'' 5h
H''' 9h

**Figure S3**

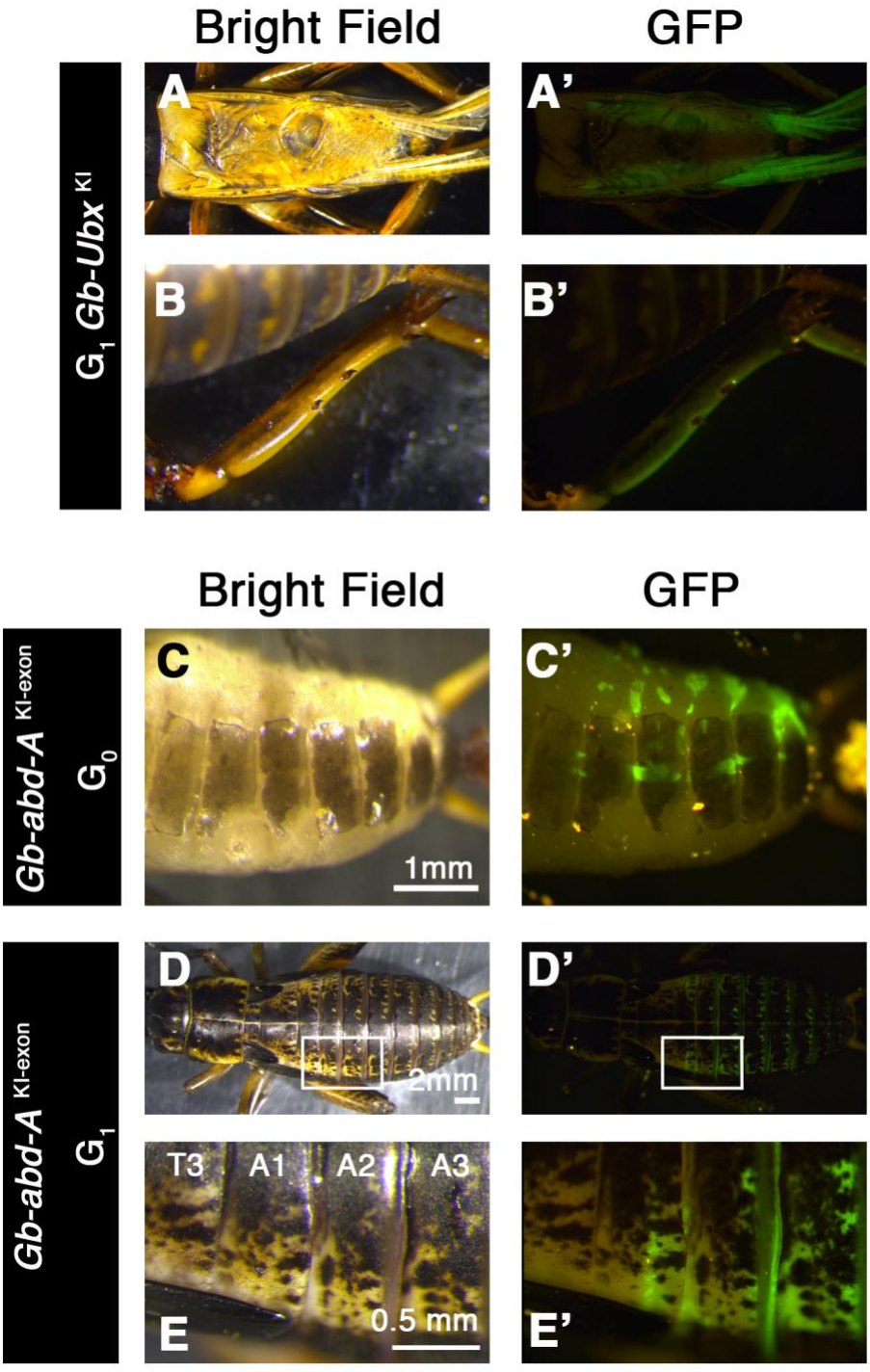


**Figure S4**

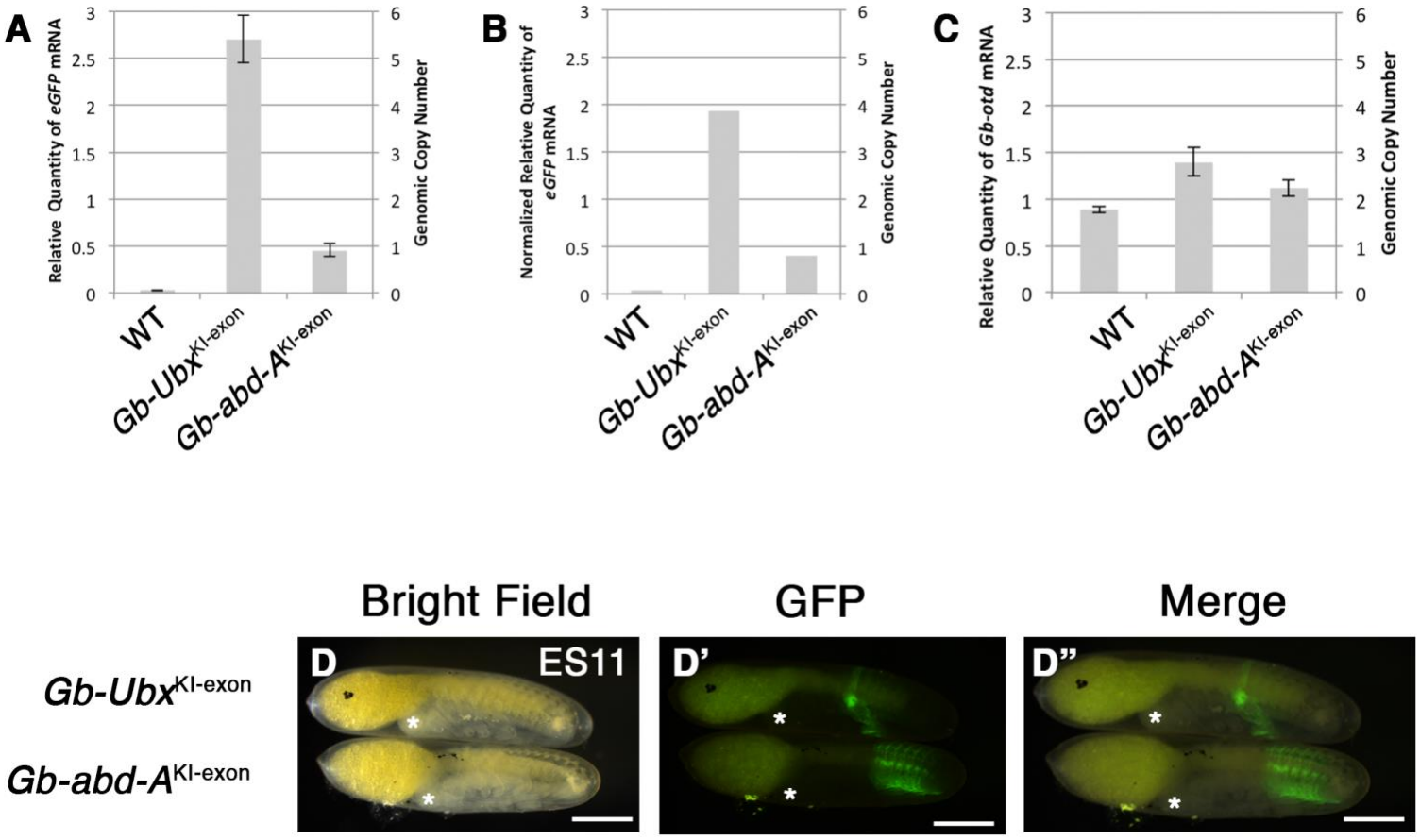


**Figure S5**

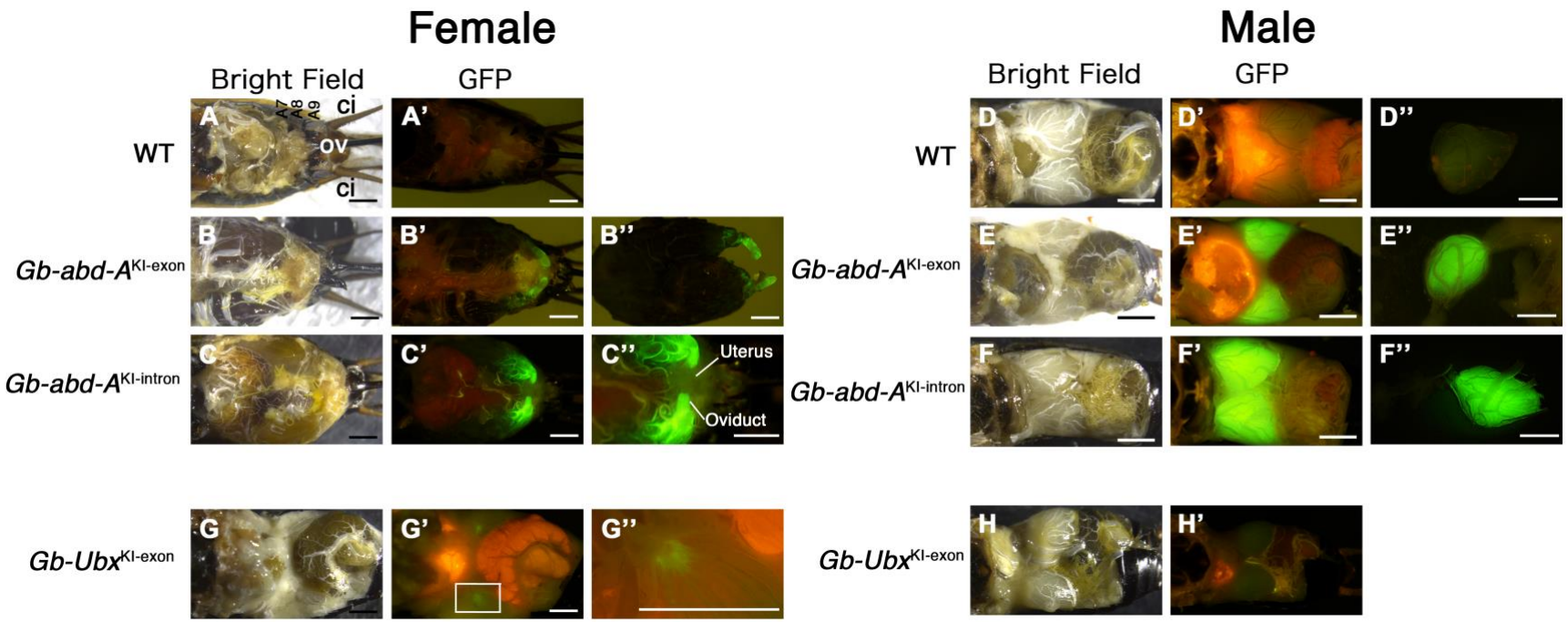
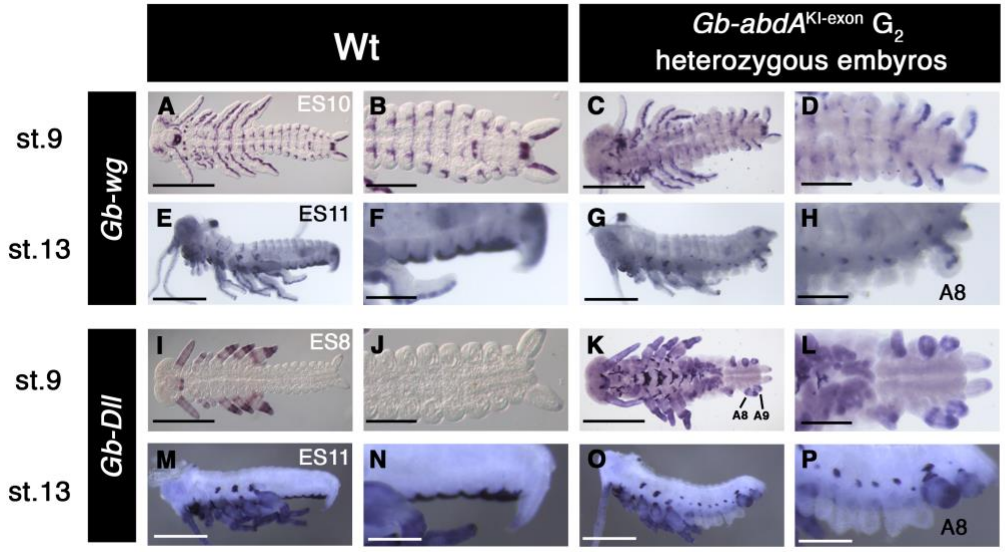


**Figure S6**


Wt
Gb-abdA^KI-exon G2 heterozygous embyros
st.9
st.13
Gb-wg
Gb-Dll
A
ES10
B
C
D
E
ES11
F
G
H
A8
I
ES8
J
K
A8 A9
L
M
ES11
N
O
P
A8

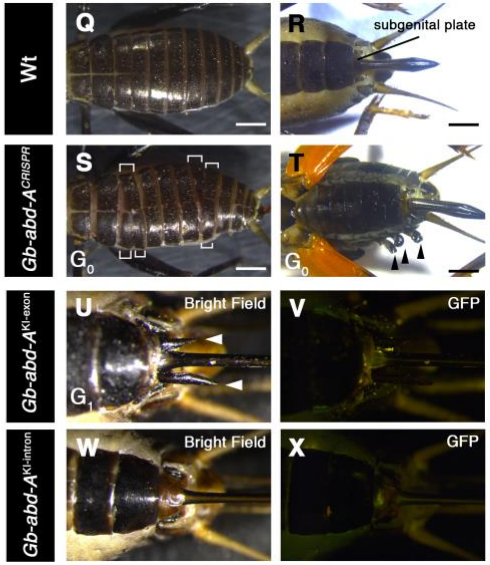

Wt
Gb-abd-A^CRISPR
Gb-abd-A^KI-exon
Gb-abd-A^KI-intron
Q
R
subgenital plate
S
G0
T
G0
U
Bright Field
G1
V
GFP
W
Bright Field
X
GFP

**Figure S7**