bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A practical guide to CRISPR/Cas9 genome editing in Lepidoptera

Linlin Zhang and Robert D. Reed

*Department of Ecology and Evolutionary Biology, Cornell University, 215 Tower Rd., Ithaca, NY 14853-7202, USA

Corresponding Authors: lz355@cornell.edu, robertreed@cornell.edu

Abstract

CRISPR/Cas9 genome editing has revolutionized functional genetic work in many organisms and is having an especially strong impact in emerging model systems. Here we summarize recent advances applying CRISPR/Cas9 methods in Lepidoptera, with a focus on providing practical advice on the entire process of genome editing from experimental design through to genotyping. We also describe successful targeted GFP knock-ins that we have achieved in butterflies. Finally, we provide a complete, detailed protocol for producing targeted long deletions in butterflies.

Introduction

The order Lepidoptera represents a tenth of the world's described species and contains many taxa of significant scientific and agricultural importance. Despite major interest in this group, however, there has been a frustrating lack of progress in developing routine approaches for manipulative genetic work. While the last decade has seen examples of transgenesis and targeted knockouts using methods like transposon insertion[1], zinc-finger nucleases[2,3], and TALENs[4,5], especially in the silk moth *Bombyx mori*, these approaches have resisted widespread application due to their technically challenging nature. Lepidopterans in general pose several unusual challenges relative to other insect groups. One major challenge for transgenic work is that many lepidopterans are sensitive to inbreeding, and in some species it can be difficult to maintain experimental lines without special effort. Another challenge is that lepidopterans appear to have an unusual resistance to RNAi[6,7], a method that has dramatically

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

accelerated work in other groups of both hemimetaboulous and holometablous insects. Given the history of challenges facing manipulative genetics in Lepidoptera, it is with great excitement that over the last few years we have seen an increasing number of studies that demonstrate very high efficiency of CRISPR/Cas9-mediated genome editing in this group. Our own lab began experimenting with genome editing for our butterfly wing pattern work in 2014, and we and our collaborators have now successfully edited over 15 loci across six species, generating both targeted deletions and insertions. The purpose of this review is to briefly summarize the current state of this fast moving field and to provide practical advice for those who would like to use this technology in their own work.

Published examples of Cas9-mediated genome editing in Lepidoptera

Between 2013 and early 2017 we identify 22 published studies applying CRISPR/Cas9 methods in Lepidoptera (Table 1). The earliest published reports of Cas9-mediated genome editing in Lepidoptera, from 2013 and 2014, all describe work done in *B. mori*[8-10] – an experimental system that benefits from a large research community that had already developed efficient methods for injection, rearing, and genotyping. To our knowledge Wang et al. (2013)[8] represents the first published report of Cas9-mediated genome editing in Lepidoptera, and set three important precedents. First, they established the protocol that has been more or less emulated by most following studies, where single guide RNAs (sgRNAs) are co-injected with Cas9 mRNA into early stage embryos. Second, they demonstrated that it is possible to co-inject dual sgRNAs to produce long deletions. In this respect, the 3.5kb deletion they produced was an important early benchmark for demonstrating the possibility of generating long deletions in Lepidoptera. Third, they showed that deletions could occur in the germ line at a high enough frequency to generate stable lines.

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

After Wang et al. (2013)[8] one of the next most important technical advancements came from Ma et al (2014)[9], who showed that knock-ins could be achieved using a donor plasmid to insert a DsRed expression cassette using ~1kb homology arms. Following this, Zhu et al. showed successful epitope tagging of *BmTUDOR-SN* gene by CRISPR/Cas9 mediated knock-in in *Bombyx* cells[11]. Unfortunately, to our knowledge, these remain the only two examples of lepidopteran knock-ins outside of the new data we present below. The first example of Cas9 genome editing in a species besides *B. mori* was described by Li et al. (2015)[12], who produced deletions in three genes in the swallowtail butterfly *Papilio xuthus*. This was an important case study because it showed the general approach used by Wang et al. (2013)[8] in *B. mori* could be transferred to other species and still retain the same level of high efficiency. Two more notable technical advancements include the production of an 18kb deletion in *B. mori* by Zhang et al. (2015)[13] – the longest deletion we know of in Lepidoptera, and much longer than anything in *Drosophila* reports we have seen – and the direct injection of recombinant Cas9 protein instead of Cas9 mRNA[14,15], which was an important improvement to the protocol that significantly simplifies the genome editing workflow.

Through our lab's research on wing pattern development we have tried most of the methods described in the studies cited above, and we have gained significant experience in porting these protocols across species. We now perform targeted long deletions routinely and with a fairly high throughput. As of the end of 2016, we and our colleagues have successfully applied this general approach in six butterfly and two moth species (*Vanessa cardui, Junonia coenia, Bicyclus anynana, Papilio xuthus, Heliconius erato, Agraulis vanillae*, *B. mori*, and *Plodia interpunctella*), with each species requiring only minor modifications to physical aspects of egg injection protocol. As we describe below, we have also successfully achieved knock-in epitope tagging similar to Zhu et al.[11], although our efficiency levels

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

remain similarly low. In this report we outline the approach that we have found to be the most time- and cost-efficient and transferable between species (Fig. 1A).

Experimental Design

Deletions: Loss-of-function deletion mutations can be generated by non-homologous end joining (NHEJ) following double strand breaks (DSBs). Both small indel (single cleavage) and long deletion knockout strategies (co-injection of two sgRNAs) have been employed in Lepidoptera (Table 1). Our lab currently favors long deletions using dual sgRNAs because it facilitates rapid screening and genotyping of mutants using PCR and regular agarose gel electrophoresis. Small indels produced by single cleavages are too small to detect easily using normal agarose gels. Dual sgRNA deletions, however, can be tens, hundreds, or thousands of base pairs long and are easy to identify in gels. sgRNA target sites can be easily identified simply by scanning the target region for $GGN_{18}NGG$ or $N_{20}NGG$ motifs on either strand using the CasBLASTR web tool (http://www.casblastr.org/). In our experience the relative strandedness of sgRNAs does not appear to have a significant effect on the efficiency of double sgRNA long deletion experiments. If a reference genome is available candidate sgRNA sequences should be used for a blast search to confirm there are not multiple binding sites that may produce off target effects. The injection mix we typically use is 200 ng/µl Cas9 and 100 ng/µl of each sgRNA – this will tend to give larger effects and is suitable for less potentially lethal loci. For targets that may result in more deleterious effects we recommend decreasing the amount of Cas9/sgRNA mix and injecting later in embryonic development to induce fewer and smaller clones. We have been able to induce mosaic mutants (e.g. Fig. 1B) using as low as 20 ng/µl Cas9 and 50 ng/µl of each sgRNA in different butterfly species.

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Insertions: CRISPR/Cas9-induced site specific DSBs can be precisely repaired by homology directed recombination repair (HDR). The HDR pathway can replace an endogenous genome segment with a homologous donor sequence, and can thus be used for knock-in of foreign DNA into a selected genomic locus. To our knowledge there are only two published examples of this approach in Lepidoptera, both of which are in *B. mori* [9,11]. To test the feasibility of this approach in butterflies we sought to insert an in-frame EGFP coding sequence into to the *V. cardui dopa decarboxylase* (*Ddc*) locus using a donor plasmid containing the EGFP coding sequences and homologous arms matching endogenous sequences flanking the Cas9 cut sites (Fig. 2A). As shown in Fig. 2B, EGFP fluorescence was detected in clones in the mutant caterpillars. As well, PCR analysis with primers flanking the 5′ and 3′ junctions of the integration shows a clear band in mutants, but not in wild type (Fig. 2C). Our results show that donor DNA with ~500bp homology arms is sufficient for precise in-frame knock-ins. Compared to NHEJ mediated high efficient knockout (69% in the case of *V. cardui Ddc* deletions knockouts[14]), the rate of HDR mediated targeted integration is low, at ~3% in our current trials. It has been shown that knocking out of factors in the NHEJ pathway can enhance the HDR pathway and increase gene targeting efficiency in *Bombyx*[9,11]. Some Cas9 mediated homology-independent knock-in approaches have shown higher efficiency rates in zebrafish[16] and human cell lines[17], suggesting NHEJ repair may provide an alternate strategy to improve incorporation of donor DNA in Lepidoptera.

Embryo Injection

When adapting CRISPR/Cas-9 genome editing to a new species, the greatest technical challenges we face typically lie in optimizing the injection protocol. The main reason for this is that the eggs of different species can be quite different in terms of how difficult they are to puncture with a glass needle, and also how they react to mechanical injection, especially in terms of internal pressure and post-injection backflow.

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Injection Needles: Proper needle shape is critical for achieving successful egg injections in Lepidoptera. In our experience some taxa like *Heliconius spp.* have very soft, easy-to-inject eggs that present very few problems, and are more robust to variation in needle shape. Many lepidopterans, however, have thick-shelled eggs that have high internal pressure. The key challenge for these eggs is to pull needles that are strong enough to puncture tough eggshells, but are not so wide as to weaken pressure balance or destroy embryos. For instance, needles that are too long and narrow can break easily when used on tough eggshells and will clog at a high frequency. Conversely, needles that have a very wide diameter will tend to have problems with pressure loss and backflow. We recommend the needle shape shown in Fig. 3A which is characterized by a short rapid taper to a fine point. We have found that this shape provides enough strength to puncture fairly tough eggshells, yet is relatively resistant to clogging and pressurization problems. Our initial attempts at pulling needles like this with a traditional gravity needle puller failed. We currently pull our needles using a velocity sensitive Sutter P-97 programmable needle puller, which works very well for crafting nuanced needle shapes. We currently prefer to use Sutter Instrument 0.5mm fire polished glass capillary needles (Sutter BF-100-50-10) and 3mm square box heating filaments (Sutter FB330B). Although settings will vary by instrument and filament, we use a single cycle program on our puller with parameters HEAT 537, PULL strength 77, VELOCITY (trip point) 16, and TIME mode (cooling) 60. Among these parameters, the HEAT value has to be adjusted relative to the RAMP value, which is specific to certain instruments – different pullers can produce slightly different needle shapes even with the same parameter setting. We provide our settings as a starting point for other users to work towards optimizing production of needles with a steep taper and a large orifice as shown in Fig. 3A.

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Egg Treatment: Egg treatment is different among eggs from different taxa. For butterflies with soft eggs such as *Heliconius*, *Agraulis*, and *Danaus*, freshly collected eggs can be immediately arranged on double-sided adhesive tape on a microscope slide (Fig. 3B) and injected. For those eggs with A relatively soft chorion but high pressure, like *V. cardui*, collected eggs should be arranged on a slide, and then kept in a desiccation chamber for 15min before injection. We use a simple sealed petri dish filled with desiccant for this purpose. For species with thick-shelled eggs like *J. coenia*, we recommend that collected eggs be dipped in 5% benzalkonium chloride (Sigma-Aldrich, St Louis, MO, USA) for 90s to soften the chorion, and then washed in water for 2min before mounting on microscope slide. We also tried treatment with 50% bleach solution to soften eggs, however this significantly reduced the hatch rate. Softened eggs can then be dried in a desiccation chamber for 15min and injected.

Injection Timing: In all published cases we are aware of injections of sgRNA and Cas9 (either mRNA or recombinant protein) were completed between 20m and 4 hours after oviposition, when embryos are presumed to be in an acellular syncytial state. Most of our injecting experience has been in eggs 1-3 hours old. Although we have not rigorously quantified this effect, after extensive work with pigmentation genes in *V. cardui* we found that injecting earlier (e.g. at 1 hr) typically produces more and larger mutant clones compared to injection performed later (e.g. at 4 hr). This is consistent with previous studies that have found a higher deletion frequency when embryos are injected at earlier versus later stages[12]. Thus, for most of our deletion experiments we aim to inject ~1-2 hours after oviposition. If we expect that deletion of the locus will have a strongly deleterious or embryonic lethal effect, we will begin by injecting at 2-4 hours to decrease the magnitude of somatic deletions.

Egg Injection: The key concern during injection is to minimize damage as much as possible. An optimum angle for needle insertion is about 30-40 degrees in our experience. We prefer to use a

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Narishige MM-3 micromanipulator for full three-dimensional control of the needle during injection. In the butterfly species we have worked with the location of the injection does not seem to have a major impact on editing efficiency, although in *V. cardui* we get a slightly higher survival rate by injecting into the side near the base of the egg. Proper positive balance pressure is critical for successful injection. Users should adjust balance pressure to a point where the needle is just able to retain the solution. Prior to any egg injection, adjust the injection pressure and time to ensure the flowing droplet is visible when pressing the injector's footswitch. We have worked extensively with two different injectors: a Harvard Apparatus PLI-100 Pico-Injector and a Narishige IM 300 Microinjector. In our experience PLI-100 Pico-Injector has better sensitivity in terms of balance pressure, which is very important for species with high pressure eggs like *V. cardui* and *J. coenia*. The IM 300 does not perform as well on these eggs. The other two injectors we know of that also work well for butterfly eggs are Eppendorf FemtoJet microinjector and Drummond Nanoject III. In our hands, we use 10 psi injection pressure and 0.5 psi balance pressure for soft-shelled eggs with the Narishige IM 300 injector, and 20 psi injection pressure and 0.8 psi balance pressure for *V. cardui* and *J. coenia* eggs with the PLI-100 Pico-Injector. After injection we maintain slides with the injected eggs in a petri dish and move larvae to their rearing containers immediately upon emergence.

Interpreting somatic mosaics

While several studies have been published that describe the germ line transmission of edited alleles in *B. mori*, thus far most studies in Lepidoptera have focused on interpreting deletion phenotypes in G_0 somatic mosaics. Maintaining edited genetic lines is necessary for looking at the homozygous effects of specific edited alleles, and will also be essential for a future generation of more sophisticated knock-in studies. Maintaining edited lines presents a few challenges in Lepidoptera, however. First, the deletion phenotypes of many interesting genes would likely be embryonic lethal. For example, our lab

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

has thus far been unsuccessful in efforts to produce living larvae with *wingless* or *Notch* coding region deletions, which is unsurprising because these genes are known to be essential for early embryonic development in insects. For loci like these we can confirm deletions by PCR and sequencing, but all embryos with deletions die before or shortly after hatching. Second, based on our experience with inbreeding attempts in *Heliconius spp.*, *V. cardui*, and *J. coenia*, and through discussions with colleagues working in other systems, it is clear that many Lepidoptera are sensitive to inbreeding and lines will die out quickly unless fairly large stocks are kept. Large stocks then make it more difficult to identify individuals with specific genotypes. So while maintaining lines is certainly possible in many lab-adapted species, it can require a fair amount of time and resources.

Because of the challenges posed by the embryonic lethality of many target genes, along with the difficulty of maintaining and genotyping edited lines, most of our attention has focused on analysis of mosaic G_0 phenotypes. One obvious advantage of focusing on somatic mosaics is that data can be collected in a single generation. Another advantage is that the phenotypic effects of lesions are limited to the subset of cell lineages (clones) hosting deletion alleles, thus reducing the deleterious effects of many deletions. Because of their clear phenotypes, knockout work on melanin pigmentation genes has allowed a very useful visual demonstration of the nature of somatic mosaicism in injected animals. Our work on eight pigmentation genes across several butterfly species[18] has allowed us some general insights into work with mosaics. First, as described above, we found a loose association between the number and size of clones and the timing of injection, where earlier injections with higher concentrations of Cas9/sgRNA complexes tend to produce larger clones. We have not attempted to quantify this effect, but across replicated experiments our tentative conclusion is that this is a real and consistent phenomenon. This is important because it gives rough control over the strength of a phenotype, and can thus be important for trying to get small non-deleterious clones for an otherwise

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

lethal gene. Conversely, by injecting at very early stages to knock out minimally pleiotropic genes we could produce animals with very large clones, such as entire wings.

One challenge of working with somatic mosaic lies in detecting and interpreting more subtle phenotypes. Most of the phenotypes published to date, such as loss of wing pattern features like eyespots or production of discolored patches, are fairly obvious and/or far outside the range of natural variation. Without having a dramatic phenotype or independent clone boundary marker, however, minor or highly localized effects can be difficult to differentiate from natural variation. It is possible that quantitative image analysis approaches could address this issue, although we are unaware of published examples of this. In our own work we have relied on two main criteria to validate putative deletion phenotypes: (1) Replicates, which, of course, are useful for increasing confidence (we typically aim for a minimum of three, although the number of replicates required to make a particular inference is somewhat arbitrary and there is no standard), and (2) asymmetry, which is perhaps the most powerful criterion for inferring more subtle deletion phenotypes. Because natural variation is ordinarily symmetrical, strongly asymmetric phenotypes are best explained by left/right variation in clonal mosaics.

Genotyping

To validate that genome editing is occurring as expected at the appropriate locus, it is necessary to perform genotyping on experimental animals. We have found that the simplest and most robust genotyping approach is to design PCR primers flanking deletion sites, and then to compare PCR product sizes between wild type and experimental animals. We recommend that genotyping amplicons cover less than 1.5kb and be at least 100bp outside of the closest sgRNA site to allow proper band size resolution and detection of large deletions. This approach works best for long deletions produced by

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

double sgRNAs – indels induced by repair of a single Cas9 cut site will usually be too small to detect by PCR alone. For this reason, our lab always uses double sgRNAs to produce deletions. These PCR products can also be cloned and sequenced for further validation, as well as to better characterize the diversity and nature of deletion alleles. If a single sgRNA is used it is likely that deletion alleles will need to be sequenced to confirm lesions. To genotype insertions PCR primers flanking the insertion site may be used similarly, or one may also use a primer inside the transgene (e.g. Fig 2C).

A current challenge in genotyping edited animals is the lack of tools to rigorously confirm specific deletion alleles in specific cell populations. First, there is the physical problem of isolating a population of cells representing a single pure clone. To our knowledge, this has not been done in insects outside of using transgenic cell sorting methods[19]. Even carefully dissected presumptive clones cannot be assumed to be pure clonal cell populations. Indeed, to our knowledge there is not yet a practical method developed to firmly associate specific alleles with specific phenotypes. This challenge also makes it difficult to decisively confirm whether a clone is monoallelic (i.e. has a single edited allele) vs. biallelic (i.e. has two edited alleles), thus making it difficult to infer dosage effects without additional information. Therefore, even though some previous studies present DNA sequences of edited alleles isolated from tissues including cells with deletion phenotypes (e.g. whole embryos), none of these studies rigorously associate individual alleles with specific clones because it cannot be ruled out that the tissue samples maybe have contained multiple monoallelic or bialleleic clones. A second challenge for genotyping specific clones is that some tissues of special interest, such as adult cuticle structures, including wing scales, do not have genomic DNA of sufficient quality to permit straightforward PCR genotyping, especially for longer amplicons. Thus, even if methods become available for isolating specific clone populations, there will still be limitations when dealing with some tissue types. Given the challenges outlined above, readers should understand that most genotyping to date should be seen

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

as a validation of the experimental approach (editing accuracy and efficiency), and not necessarily as completely decisive confirmation that specific alleles underlie a certain clone phenotype.

Future prospects

CRISPR/Cas9 genome editing is rapidly revolutionizing genetic work in Lepidoptera, as it is across all of biology. It is now fairly straightforward to quickly and cheaply induce long, targeted deletions in virtually any species that can be reared in captivity. Published reports to date have focused on producing deletions in gene coding regions, however we anticipate there will be significant interest in also applying long deletion approaches to test the function of non-coding regulatory regions, especially now that cis-regulatory elements can be functionally annotated with high resolution thanks to methods like ChIP-seq[20]. Pilot work shown here and elsewhere also shows that targeted insertions are possible as well, thus promising even further developments on the near horizon such as protein tagging, reporter constructs, and tissue-specific expression constructs. Right now the main challenge with knock-in strategies is the relatively low efficiency rate, although newer technologies such as non-homologous end joining mediated knockin[16] promise to dramatically improve this. Perhaps the most exciting thing about CRISPR-associated genome editing approaches, though, is the straightforward portability of the technology between species. This is truly an exciting time to be a comparative biologist.

Acknowledgments

We thank Arnaud Martin for extensive discussions about regarding genome editing methods in butterflies, Joseph Fetcho for early assistance with needle pulling and injection procedures, and Katie Rondem for helpful comments on the manuscript. This work was funded by United States National Science Foundation grant DEB-1354318

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1 Tamura, T. *et al.* Germline transformation of the silkworm *Bombyx mori* L. using a piggyBac transposon-derived vector. *Nature Biotechnology* **18**, 81-84 (2000).

2 Takasu, Y. *et al.* Targeted mutagenesis in the silkworm *Bombyx mori* using zinc finger nuclease mRNA injection. *Insect Biochemistry and Molecular Biology* **40**, 759-765 (2010).

3 Merlin, C., Beaver, L. E., Taylor, O. R., Wolfe, S. A. & Reppert, S. M. Efficient targeted mutagenesis in the monarch butterfly using zinc-finger nucleases. *Genome Research* **23**, 159-168 (2013).

4 Takasu, Y. *et al.* Efficient TALEN construction for *Bombyx mori* gene targeting. *PLoS One* **8**, e73458 (2013).

5 Markert, M. J. *et al.* Genomic access to monarch migration using TALEN and CRISPR/Cas9-mediated targeted mutagenesis. *G3: Genes| Genomes| Genetics* **6**, 905-915 (2016).

6 Terenius, O. *et al.* RNA interference in Lepidoptera: an overview of successful and unsuccessful studies and implications for experimental design. *Journal of insect physiology* **57**, 231-245 (2011).

7 Kolliopoulou, A. & Swevers, L. Recent progress in RNAi research in Lepidoptera: Intracellular machinery, antiviral immune response and prospects for insect pest control. *Current Opinion in Insect Science* **6**, 28-34 (2014).

8 Wang, Y. *et al.* The CRISPR/Cas system mediates efficient genome engineering in *Bombyx mori*. *Cell Research* **23**, 1414-1416 (2013).

9 Ma, S. *et al.* CRISPR/Cas9 mediated multiplex genome editing and heritable mutagenesis of *BmKu70* in *Bombyx mori*. *Scientific Reports* **4**, 4489 (2014).

10 Wei, W. *et al.* Heritable genome editing with CRISPR/Cas9 in the silkworm, *Bombyx mori*. *PloS One* **9**, e101210 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

12 Li, X. *et al.* Outbred genome sequencing and CRISPR/Cas9 gene editing in butterflies. *Nature communications* **6** (2015).

13 Zhang, Z. *et al.* Functional analysis of Bombyx Wnt1 during embryogenesis using the CRISPR/Cas9 system. *Journal of insect physiology* **79**, 73-79 (2015).

14 Zhang, L. & Reed, R. D. Genome editing in butterflies reveals that *spalt* promotes and *Distal-less* represses eyespot colour patterns. *Nature communications* **7**, doi:10.1038/ncomms11769 (2016).

15 Perry, M. *et al.* Molecular logic behind the three-way stochastic choices that expand butterfly colour vision. *Nature* **535**, 280-284, doi:10.1038/nature18616 (2016).

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

17 He, X. *et al.* Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic acids research*, gkw064 (2016).

18 Zhang, L. *et al.* Genetic basis of melanin pigmentation in butterfly wings. *Genetics*, 102632, doi:10.1534/genetics.116.196451 (2017).

19 Böttcher, R. *et al.* Efficient chromosomal gene modification with CRISPR/cas9 and PCR-based homologous recombination donors in cultured *Drosophila* cells. *Nucleic Acids Research* **42**, e89-e89 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

20 Lewis, J. J., van der Burg, K. R., Mazo-Vargas, A. & Reed, R. D. ChIP-Seq-Annotated *Heliconius erato* genome highlights patterns of cis-regulatory evolution in Lepidoptera. *Cell Reports* **16**, 2855-2863 (2016).

21 Liu, Y. *et al.* Highly efficient multiplex targeted mutagenesis and genomic structure variation in *Bombyx mori* cells using CRISPR/Cas9. *Insect Biochemistry and Molecular Biology* **49**, 35-42 (2014).

22 Ling, L. *et al.* MiR-2 family targets awd and fng to regulate wing morphogenesis in *Bombyx mori. RNA Biology* **12**, 742-748 (2015).

23 Li, Z. *et al.* Ectopic expression of ecdysone oxidase impairs tissue degenration in *Bobmyx mori*. *Proceedings of the Royal Society B.* **282,** 20150513 (2015).

24 Xin, H. h. *et al.* Transcription factor Bmsage plays a crucial role in silk gland generation in silkworm, *Bombyx mori*. *Archives of insect biochemistry and physiology* **90**, 59-69 (2015).

25 Zeng, B. *et al.* Expansion of CRISPR targeting sites in *Bombyx mori*. *Insect Biochemistry and Molecular Biology* **72**, 31-40 (2016).

26 Khan, S. A., Reichelt, M. & Heckel, D. G. Functional analysis of the ABCs of eye color in *Helicoverpa armigera* with CRISPR/Cas9-induced mutations. *Scientific Reports* **7** (2017).

27 Beldade, P. & Peralta, C. M. Developmental and evolutionary mechanisms shaping butterfly eyespots. *Current Opinion in Insect Science* **19**, 22-29 (2017).

28 Bassett, A. R., Tibbit, C., Ponting, C. P. & Liu, J.-L. Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. *Cell Reports* **4**, 220-228, doi:10.1016/j.celrep.2013.06.020 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. Comparison of published CRISPR genome editing studies in Lepidoptera

Reference	Knockin/knockout	Konckout Strategy	Delivery	Gene targeted	Genotyping	Mosaic (%)
Wang Y. et al., 2013 [8]	knockout	Small indels & long deletions	Cas9 mRNA & sgRNA	*BLOS2*	PCR analysis	94%-100%
Ma S. et al., 2014 [9]	knockout/-in	Small indels	Plasmid of Cas9 & gRNA	*ku70*	T7E1	16.8% – 30.3%
Wei W. et al., 2014 [10]	knockout	Small indels & long deletions	Cas9 mRNA & sgRNA	*ok/KMO/TH/tan*	sequencing	16.7%–35.0%
Liu Y. et al., 2014 [21]	knockout	Small indels	Plasmid of Cas9 & gRNA	*Th/re/fl/yel-e/kynu/e*	T7E1	5.7%-18.9%
Zhu L, et al., 2015 [11]	knockout/-in	Small indels	All-in-one vector	*Ku70/Ku80/Lig IV/XRCC4/XLF*	T7E1	20%-80%
Ling L, et al., 2015 [22]	knockout	Long deletions	Cas9 mRNA & sgRNA	*awd/fng*	PCR analysis	40%-61%
Li Z, et al., 2015 [23]	knockout	Long deletions	Cas9 mRNA & sgRNA	*EO*	PCR analysis	60.60%
Xin H, et al., 2015 [24]	knockout	Small indels	Cas9 mRNA & sgRNA	*sage*	sequencing	46.67%
Zhang Z, et al., 2015 [13]	knockout	Long deletions	Cas9 mRNA & sgRNA	*wnt1*	T7E1	42.5%-90.6%
Zeng B, et al., 2016 [25]	knockout	Small indels	Cas9 mRNA & sgRNA	*GFP/BLOS2*	T7E1	NA
Khan S. et al., 2017 [26]	knockout	Small indels	Cas9 mRNA & sgRNA	*w/st/bw/ok*	T7E1	NA
Li X, et al., 2015 [12]	knockout	Small indels & long deletions	Cas9 mRNA & sgRNA	*Abd-B/e/fz*	T7E1	18.33%-90.85%
Zhang L., et al., 2017	knockout	Long deletions	Cas9 protein & sgRNA	*y*	NA	83.33%
Perry M., et al., 2016 [15]	Knockout	Long deletions	Cas9 protein & sgRNA	*y/ss*	sequencing	NA
Zhang L. & Reed R., 2016 [14]	knockout	Long deletions	Cas9 protein & sgRNA	*Ddc/Dll/spalt*	PCR analysis	51.7%-56%
Zhang L., et al., 2017	knockout	Long deletions	Cas9 protein & sgRNA	*y/b/e/yel-d*	PCR analysis	42%-80%
Perry M., et al., 2016 [15]	knockout	Long deletions		*y/ss*	sequencing	NA
Zhang L. & Reed R., 2016 [14]	knockout	Long deletions	Cas9 protein & sgRNA	*Dll/spalt*	PCR analysis	33%-41%
Zhang L., et al., 2017	knockout	Long deletions	Cas9 protein & sgRNA	*Ple/e*	NA	27%-42%
Zhang L., et al., 2017	knockout	Long deletions	Cas9 mRNA/protein & sgRNA	*y/e*	T7E1	31%-72%
Beldade P. & Peralta C., 2017[27]	knockout	NA	NA	*Ddc/e*	NA	NA
Markert M., et al., 2016 [5]	knockout	Small and long deletions	Cas9 mRNA & sgRNA	*cry2/clk*	Cas9 in vitro cleavage	57.9%-61.5%

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Legends

Figure 1. Timeline and example outcome of G_0 CRISPR/Cas9 mosaic knockout experiments in butterflies. (A) Overview and timeline of mutant generation by CRISPR/Cas9 injection to butterfly embryos. (B) Example of larval and adult wing somatic mosaic phenotypes resulting from knockout of the melanin pigmentation gene *Ddc*.

Figure 2. Knock-in epitope tagging of the *Ddc* gene in *V. cardui*. (A) Schematic overview of the *Ddc* locus and donor construct consisting of homology arms, EGFP coding region, and genotyping primers. PAM regions are marked by yellow, cut sites are marked by yellow arrowhead, and genotyping primers are marked by red arrows. (B) Strong mosaic EGFP expression in a knock-in caterpillars visualized by fluorescent microscopy. (C) PCR analysis demonstrates using the primers in (A) show the insertion of EGFP into the *Ddc* coding region.

Figure 3. Needle shape and egg arrangement for butterfly embryo injections. (A) The injection needle shape we prefer has a steep taper and a relatively large orifice. Here a preferred needle is shown next to a *Heliconius* egg. (B) An example of arranging *Heliconius* eggs on double-sided tape on a microscope slide just before injection.

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

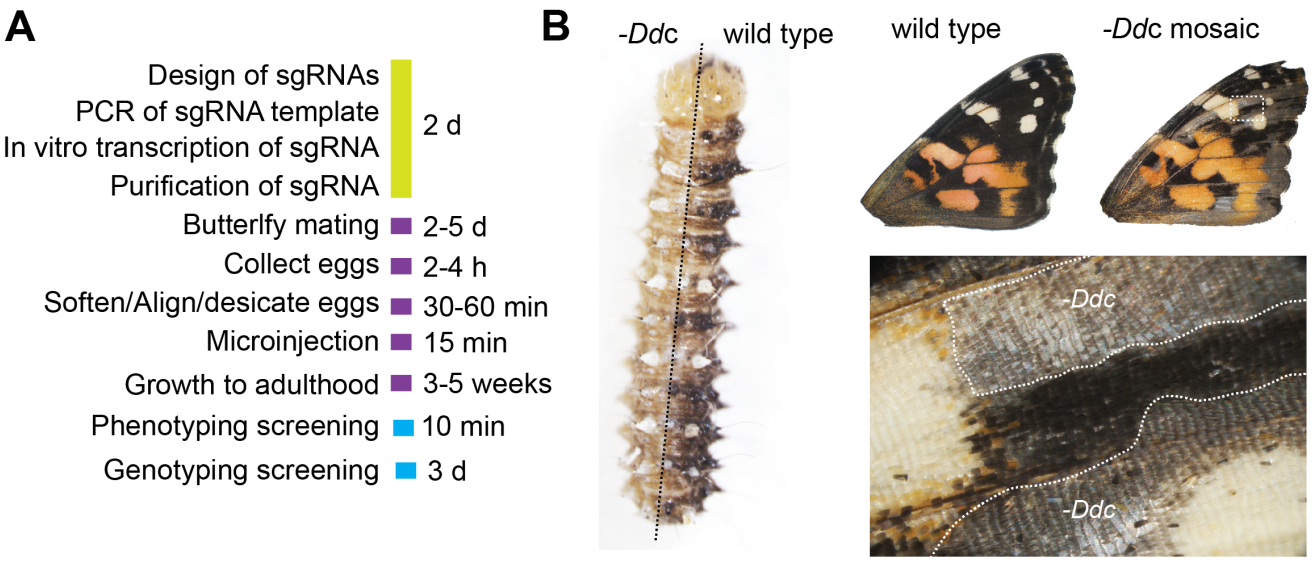

Fig. 1

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

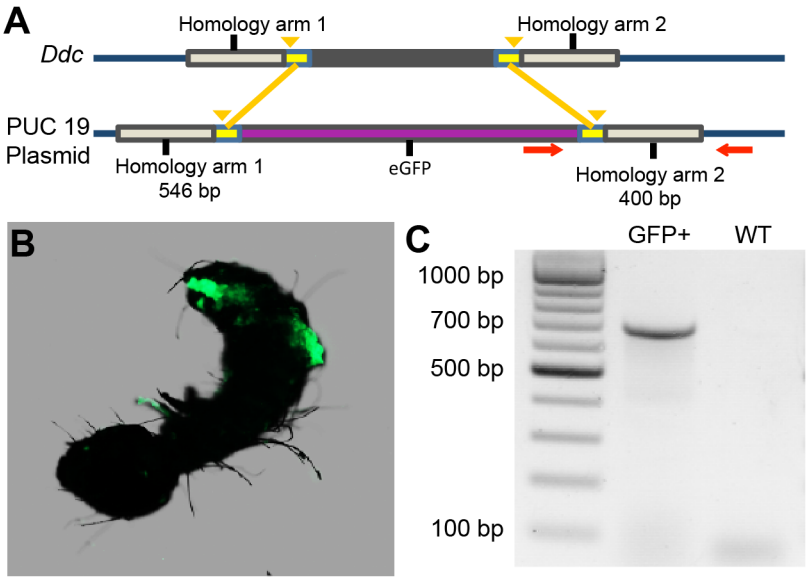

Fig. 2

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

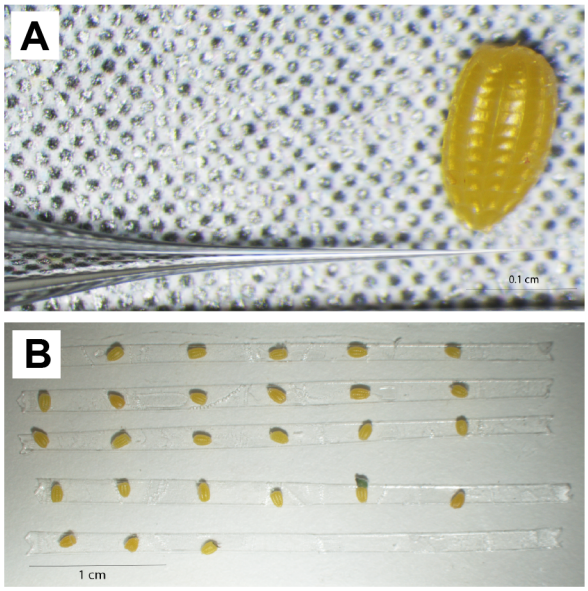

Fig. 3

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Appendix: A detailed example of CRISPR/Cas9 genome editing in the painted lady butterfly *V. cardui*

The following procedure provides guidelines to generate genomic deletions in the butterfly *V. cardui* using the CRISPR/Cas9 nuclease system. This protocol includes a specific example of the Reed Lab's work deleting the melanin pigmentation pathway gene *Ddc* as previously reported[14].

1. Target design

No genome reference was available for *V. cardui* when we first began our experiment, so we used a transcriptome assembly[18] to identify sequences of the *Ddc* coding region. Primers GCCAGATGATAAGAGGAGGTTAAG and GCAGTAGCCTTTACTTCCTCCCAG were designed to amplify and sequence the target region of the genome, and exon-intron boundaries were inferred by comparing genomic and cDNA sequences. We recommend designing target sites at exons because they are more conserved than introns and therefore provide more predictably consistent matches between sgRNAa and genomic targets. We design sgRNAs by scanning for $GGN_{18}NGG$ or $N_{20}NGG$ pattern on the sense or antisense strand of the DNA. Target sequences GGAGTACCGTTACCTGATGA**AGG** and **CCT**CTCTACTTGAAACACFACCA (PAM sequences underlined) were designed to excise a region of 131 bp spanning the functional domains of the DDC enzyme. sgRNA oligos containing T7 promoter, target sequences and sgRNA backbone were synthesized by a commercial supplier (Integrated DNA Technologies, Inc.). Of note, the PAM sequence is not included in the CRISPR forward primer.

CRISPR forward oligos:

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ddc sgRNA1:

GAAATTAATACGACTCACTATA**GGGATCAGCTTTCGTCTGCC**GTTTTAGAGCTAGAAAT

AGC

Ddc sgRNA2: GAAATTAATACGACTCACTATA**GGAGTACCGTTACCTGATGA**GTTTTA

GAGCTAGAAATAGC

CRISPR universal oligo:

AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTG

CTATTTCTAGCTCTAAAAC

2. sgRNA production

2.1 sgRNA template generation

- With the oligos generated in the preceding step, use High-Fidelity DNA Polymerase PCR Mix (NEB, Cat No. M0530) to generate the template for each sgRNA with CRISPR forward and reverse oligos. We recommend using DEPC-free nuclease-free water (Ambion, Cat No. AM9938).

PCR Reaction		PCR program
High-Fidelity DNA Polymerases PCR	50 µl	
Mix	5 µl	98 °C 30s
CRISPR forward oligo (10 µM)	5 µl	35 cycles (98 °C 10s; 60 °C
CRISPR universal oligo (10 µM)	40 µl	30s; 72 °C 15s)
Nuclease free water		72 °C 10 min

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

	4 °C hold

- Purify the PCR reaction with MinElute PCR purification kit (Qiagen, Cat No. 28004) following the kit instructions and eluting in 15 µl nuclease free water.
- Dilute 1µl of this reaction with 9 µl nuclease free water, then run on a gel and a fluorometer (e.g. Qubit) to confirm purity, integrity, fragment length, and yield. It is also possible to use gel extraction at this stage if non-specific products are present.
- The expected size should be around 100bp, and the expected yield should be around 200 ng/ul.

2.2 in vitro transcription (IVT)

- Generate sgRNAs by in vitro transcription of the sgRNA PCR template using the T7 MEGAscript kit (Ambion, Cat. No. AM1334). When producing and handling RNA, it is important to wear gloves, and clean equipment and benches with detergent prior to use to avoid RNAase contamination. Pipette tips with filters can also be beneficial to prevent contamination from pipettes.

IVT Reactions		Incubation and purification
ATP	2 µl	
CTP	2 µl	37 °C overnight incubation
GTP	2 µl	Add 1 µl Turbo DNAse and incubate for 15 min
UTP	2 µl	at 37 °C;
10 X Reaction	2 µl	Add 115 µl ddH_2O and 15 µl ammonium acetate
Buffer	2 µl	stop solution.

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Template	2 µl
T7 Enzyme Mix	Up to 20 µl
Nuclease free water	

- Extract sgRNA by adding 150 µl phenol:chloroform:isoamyl alcohol (25:24:1) at pH 6.7 (Sigma-Aldrich, Cat No. P2069), and vortex thoroughly for 30 s.
- Separate phases by centrifugation at 10,000 x g for 3 min at room temperature, and remove the upper phase to a fresh tube.
- Precipitate the RNA by addition of an equal volume (150 µl) of cold isopropanol (Sigma-Aldrich, Cat No. I9516).
- Mix thoroughly, and incubate at -20 °C for greater than 2 hours (can be left overnight).
- Collect RNA by centrifugation at 17,000 x g for 30 min at 4 °C.
- Wash pellet twice in 0.5 ml room temperature fresh made 70% ethanol, centrifuging at 17,000 x g for 3 min at 4 °C between each wash.
- Remove the remaining liquid and dry RNA pellet for 3 min at room temperature.
- Resuspend in 30 ul nuclease-free water.
- Measure concentration on a Qubit. The expected concentration should be around 2 µg/ul; sgRNAs can be stored at -80 °C.
- MEGAclear™ Transcription Clean-Up Kit (Ambion, Cat No. AM1908) also works very well for sgRNA purification.

3. Cas9 production

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cas9 is typically provided by injection of a plasmid, mRNA, or recombinant protein. We have tried both Cas9 mRNA and protein injections, and both yield similarly efficient mutation rates in butterflies. However, we recommend using commercially available Cas9 protein (PNA Bio, Cat No. #CP01) because it is more stable than Cas9 mRNA and is easier and faster to use.

- Cas9 mRNA is generated by in vitro transcription of the linearized MLM3613 (Addgene plasmid 42251) plasmid template. The mMessage mMachine T7 kit (Ambion, Cat No. AM1344) is used to perform in vitro transcription with T7 RNA polymerase, followed by in vitro polyadenylation with the polyA tailing kit (Ambion, Cat No. AM1350). An Agilent Bioanalyzer, or similar instrument, should be used to check the size and integrity of Cas9 mRNA. Note that Cas9 mRNA can show some degree of degradation yet still produce fairly efficient results.

4. Egg injection and survivor ratio calculation

- Collect eggs for 2-4 hours by placing a host plant leaf into the butterfly cage.
- For thick chorion eggs (e.g. *J. coenia*), dip eggs in 5% benzalkonium chloride for 90 s.
- Cut double-sided tape into several thin strips and fix them to a glass slide.
- Using a paintbrush line the eggs onto the double-sided tape.
- For high-pressure eggs (e.g. *V. cardui* or *J. coenia*) place the slide in a desiccation chamber for 15 min before injection.
- Mix Cas9 and CRISPR sgRNAs prior to microinjection.

Injection Mix	Incubation

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cas9 mRNA or protein (1 μg/μl)	1μl	Incubate on ice for 20 min
CRISPR sgRNA1 (375 ng/μl)	1μl	
CRISPR sgRNA2 (375 ng/μl)	1μl	
Nuclease free water	2μl	

- Break the closed tip of the needle with an optimum angle about 30-40 degree.
- Load the needle with 0.5 μl injection mix by capillary action, or by using by Eppendorf™ Femtotips microloader tips (Eppendorf, Cat No. E5242956003).
- One-by-one inject the eggs with the injector.
- Generally, higher amounts of sgRNA and Cas9 protein will increase mutation rate and decrease egg survival (hatch rate).

5. Genotyping for modification

- In order to investigate the efficiency of CRISPR/Cas9 mediated *Ddc* knockout, we randomly surveyed 81 first instar caterpillars. DNA was extracted according to Bassett et al. [28] to confirm CRISPR/Cas9 lesions. Generally, place one caterpillar in a PCR tube and mash the caterpillar for 30 s with a pipette tip in 50 μl of squishing butter (10 mM Tris-HCl, pH 8.2, 1mM EDTA, 25 mM NaCl, 200 μg/ml proteinase K). Incubate at 37 °C for 30 min, inactivate the proteinase K by heating to 95 °C for 2 min and store in -20 °C for PCR genotyping. Genotyping can also be done with adult butterfly leg DNA by using proteinase K in digestion buffer. We typically use QIAamp DNA mini Kit (Qiagen, Cat No. 51304) for DNA extraction when genotyping from muscle tissue.
- Design genotyping primers outside of the target region. For *Ddc*, genotyping forward (GCTGGATCAGCTATCGTCT) and reverse primers (GCAGTAGCCTTTACTTCCTCCCAG)

bioRxiv preprint doi: https://doi.org/10.1101/130344; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were designed and gave a 584bp PCR fragment in the wild type individuals.

- Mix PCR reagents. PCR fragments containing two sgRNA target sites are expected to produce smaller mutant bands than wild type.

PCR Reaction for Genotyping		PCR program
Taq DNA Polymerases PCR Mix (NEB)	12.5µl	98°C 1 min
Genotyping F Primer (10 µM)	1µl	35 cycles (98°C 10s; 55°C
Genotyping R Primer (10 µM)	1µl	30s; 72°C 40s)
DNA template for *V. cardui*	1µl	72°C 10 min
Nuclease free water	9.5µl	4°C hold

- Recover mutant bands by gel-extraction using Minelute gel extraction kit (Qiagen, Cat No. 28604).
- Ligate recovered DNA fragment to T4 vector for TA-cloning using a TA cloning kit (Invitrogen, Cat No. K202020).
- Extract plasmid with mutant DNA fragment using QIAprep Miniprep Kit (Qiagen, Cat No. 27104).
- Sequence plasmids and align mutant sequences to wild type sequences to confirm deletions. (Figure 1a in Zhang & Reed, 2016)[14].