

**Genomic access to monarch migration using TALEN and CRISPR/Cas9-mediated targeted mutagenesis**

Matthew J. Markert<sup>\*</sup>, Ying Zhang<sup>\*</sup>, Metewo S. Enuameh<sup>†</sup>, Steven M. Reppert<sup>‡</sup>, Scot A. Wolfe<sup>†</sup>, and Christine Merlin<sup>\*</sup>

<sup>\*</sup> Department of Biology and Center for Biological Clocks Research, Texas A&M University, College Station, TX, USA; <sup>†</sup> Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA; <sup>‡</sup> Department of Neurobiology, University of Massachusetts Medical School, Worcester, MA, USA.

Running title: Heritable genome targeting in a butterfly

Key words: TALENs, CRISPR, insect, germline targeting, clock genes.

To whom correspondence should be addressed:

Dr. Christine Merlin

Biological Sciences Building East, Room 118D

3258 TAMU

College Station, TX 77843-3258

979-862-2457

[cmerlin@bio.tamu.edu](mailto:cmerlin@bio.tamu.edu)

## ABSTRACT

The eastern North American monarch butterfly, *Danaus plexippus*, is an emerging model system to study the neural, molecular and genetic basis of animal long-distance migration and animal clockwork mechanisms. While genomic studies have provided new insight into migration-associated and circadian clock genes, the general lack of simple and versatile reverse-genetic methods has limited *in vivo* functional analysis of candidate genes in this species. Here, we report the establishment of highly efficient and heritable gene mutagenesis methods in the monarch butterfly using transcriptional activator-like effector nucleases (TALENs) and CRISPR-associated RNA-guided nuclease Cas9 (CRISPR/Cas9). Using two clock gene loci, *cryptochrome 2* and *clock (clk)*, as candidates, we show that both TALENs and CRISPR/Cas9 generate high-frequency non-homologous end joining-mediated mutations at targeted sites (up to 100%), and that injecting fewer than 100 eggs is sufficient to recover mutant progeny and generate monarch knockout lines in about three months. Our study also genetically defines monarch CLK as an essential component of the transcriptional activation complex of the circadian clock. The methods presented should not only greatly accelerate functional analyses of many aspects of the monarch biology, but are also anticipated to facilitate the development of these tools in other non-traditional insect species as well as the development of homology-directed knock-ins.

## INTRODUCTION

With their rich biology and their extraordinary diversity, butterflies and moths alike offer great opportunities to study the molecular and genetic basis of fascinating biological phenomena not manifested by conventional genetic model organisms. A striking example is the seasonal monarch butterfly's (*Danaus plexippus*) long-distance migration from its northern breeding range in southeastern Canada and northeastern USA to its overwintering roosting sites in central Mexico (URQUHART 1987). Despite its spectacular nature and an increasing public interest in its conservation, we still know very little about the genetic basis governing this behavior. With the availability of a high-quality reference genome (ZHAN *et al.* 2011), the monarch represents a unique model to decipher the genetic and epigenetic basis underlying animal long-distance migration and the associated timing mechanisms (REPPERT *et al.* 2015).

Northeastern American migratory monarch populations possess several generations of migratory and non-migratory forms, all equipped with the same core genetic makeup. In the fall, the migratory state is initiated in response to changing environmental factors predicting winter's arrival (i.e. decreasing day length and temperature; (REPPERT *et al.* 2010). Migrants exhibit an oriented flight guided by a sun compass that is time-compensated by circadian clocks in the antennae, allowing them to maintain a fixed flight direction throughout the day (PEREZ *et al.* 1997; MOURITSEN and FROST 2002; FROY *et al.* 2003; MERLIN *et al.* 2009; GUERRA *et al.* 2012). While a recent population genomic analysis of 101 *Danaus* genome sequences from migratory and non-migratory monarch populations identified a set of over 500 candidate genes associated with migration (ZHAN *et al.* 2014), understanding the role of these genes in the biology of

monarch migration will require further functional genomic approaches such as efficient and accessible reverse genetic tools to dissect gene function *in vivo*.

Engineered nucleases have proved to be powerful for genome editing in a wide range of model and non-model organisms (GAJ *et al.* 2013; CARROLL 2014). By inducing double-stranded breaks (DSBs) at specific genomic sites, they favor the introduction of small random insertions/deletions (indels) through imperfect non-homologous end joining repair, thereby allowing the generation of gene knockouts. In fact, zinc-finger nucleases have previously enabled heritable targeted mutagenesis *in vivo* of a monarch clock gene, *cryptochrome 2 (cry2)*, the main transcriptional repressor of the CLOCK (CLK):CYCLE (CYC) activation complex within the migration-relevant clock mechanism (MERLIN *et al.* 2013). Although the zinc-finger nucleases approach provided a proof-of-principle for the use of engineered endonucleases in butterflies and established reliable methods for nuclease delivery and mutation screening approaches, its efficiency failed to reach a standard enabling facile and rapid recovery of germline mutants. Notably, the recovery of a few germline mutants required the injection of thousands of eggs (MERLIN *et al.* 2013). Other engineered nucleases, including transcription activator-like effector nucleases (TALENs) and RNA-guided nucleases derived from clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 nuclease, have enabled efficient heritable targeted gene disruption in another lepidopteran, the silkworm *Bombyx mori* (MA *et al.* 2012; SAJWAN *et al.* 2013; WANG *et al.* 2013; LIU *et al.* 2014; WEI *et al.* 2014). CRISPR/Cas9 has also been recently reported to induce somatic mutations with high efficiencies in the swallowtail butterfly (LI *et al.* 2015). However, efficient germline transmission of Cas9-mediated mutations in a butterfly has yet to be reported.

Here, we report the generation of highly efficient, heritable, gene knockouts at two clock gene loci, *cry2* and *clk*, in the monarch butterfly using TALENs and CRISPR/Cas9. We demonstrate that for both target genes and nuclease platforms, injecting less than 100 eggs is sufficient to recover mutant progeny and generate monarch knockout lines in about three months. We also present the functional characterization of a monarch *clk* knockout, thereby establishing a novel reagent to study the role of circadian clocks in monarch migration. Our data provide new technological resources to rapidly establish genetic mutant lines in the monarch butterfly that will ultimately accelerate the quest to elucidate the genetic basis of their remarkable migration abilities.

## **MATERIAL AND METHODS**

### **Monarch butterfly husbandry**

Monarch butterflies were reared in the laboratory as previously described (MERLIN *et al.* 2013). Gravid females were placed in cages to lay eggs on potted tropical milkweed plants (*Asclepias curassavica*). Eggs were transferred to petri dishes containing milkweed leaves until hatched. Larvae were fed milkweed up to their second instar and then transferred onto individual cups of semi-artificial diet. Larvae were raised in Percival incubators at 25°C, 70% humidity, and under 15-hours light: 9-hours dark (LD) conditions. Adult monarchs were housed in glassine envelopes in the same lighting and temperature conditions, and were manually fed a 25% honey solution daily.

### **TALEN target selection and assembly**

TALEN target sites were selected within the 5' end of the *cry2* and *clk* coding sequences. The *cry2* TALEN target site was chosen to overlap with the previously

described *cry2* ZFN target site (MERLIN *et al.* 2013). Criteria for target site selection included the presence of an endogenous restriction endonuclease site for assessment of TALEN activity and genotyping of carriers of null alleles, and a 5' T within each monomer half-site. The standard Repeat Variable Di-residues (i.e. Ade = NI; Cyt = HD; Gua = NN; Thy = NG) were utilized for base recognition in the remaining sequence positions. The TALE monomers comprised between 16.5 and 19.5 modules with a 16-bp gap separating the TALEN target sites. TALE module arrays were constructed using a modified Voytas golden gate system (CERMAK *et al.* 2011) in which the TALE arrays were assembled into modified pJDS vectors containing the wild-type FokI nuclease domain (SANDER *et al.* 2011), as described previously (GUPTA *et al.* 2013).

### **gRNA design and construction**

The gRNA target sites were selected within the 5' end of the *cry2* and *clk* coding sequences using the Jack Lin's CRISPR/Cas9 gRNA finder (<http://spot.colorado.edu/~slin/cas9.html>). The *cry2* gRNA target site was chosen to overlap with the previously described *cry2* ZFN target site (MERLIN *et al.* 2013) and the TALEN target site. The gRNA expression vectors were constructed by inserting annealed synthetic oligomers into the DR274 plasmid from Addgene (HWANG *et al.* 2013) at the *BsaI* cleavage sites.

### **Synthesis of TALEN mRNA, Cas9 mRNA and sgrRNA**

*In vitro* transcription of TALEN mRNAs was performed using the mMessage mMachinE T7 kit (Ambion) using pJDS-TALEN constructs linearized with *PmeI* and phenol-chloroform purified, and were polyadenylated using the Poly(A) Tailing Kit (Ambion), according to the manufacturer's instructions. *In vitro* transcription of

*Streptococcus pyogenes* Cas9 mRNA was performed using the mMessage mMachine T3 transcription kit (Ambion) using the pCS2-nCas9n expression plasmid from Addgene (JAO *et al.* 2013) linearized with *XbaI* and purified with phenol-chloroform. The resulting capped PolyA mRNAs were purified by acid-phenol-chloroform extraction and resuspended in RNase-free water following isopropanol precipitation. sgRNAs were *in vitro* transcribed from sgRNA containing DR274 vectors linearized with *DraI* and purified by phenol-chloroform extraction, using T7 RNA polymerase (Promega) according to the manufacturer's instructions. RNAs were quantitated by spectrophotometry (NanoDrop 1000) and diluted in RNase free water to a final concentration of 0.5 µg/µl for *cry2* TALENs and *clk* TALENs, of 0.5 µg/µl for Cas9 and of 0.1 µg/µl for *cry2* and *clk* sgRNAs.

### **Egg microinjection**

To perform microinjections, eggs were collected as previously described (MERLIN *et al.* 2013) and injected within 20min of being laid. A solution containing TALEN mRNAs or Cas9 mRNA, sgRNAs and blue food coloring was loaded into a pulled borosilicate glass needle (World Precision Instruments, Inc.) attached to an IM 300 microinjector (Narishige), and injections were performed under a dissecting microscope. After injection, embryos were placed in an incubator at 25°C and 70% relative humidity. Developing embryos were transferred into individual small petri dishes until larvae hatched. Larvae were fed milkweed leaves until the second larval instar before being transferred onto semi-artificial diet.

### **Analysis of TALEN-induced mutations using restriction assays and mutant line establishment**

Surviving fourth instar larvae were screened for the presence of mutations at the targeted site by non-invasive genotyping as previously described (MERLIN *et al.* 2013). Genomic DNA from larval sensors was extracted using 0.01X proteinase K (Sigma) in lysis buffer (100 mM Tris pH 8.0, 25 mM NaCl, 1 mM EDTA). Fragments flanking the targeted regions were amplified by PCR using the primers in Table S1 and subjected to restriction fragment length polymorphism analysis using *EagI* for *cry2* or *NgoMIV* for *clk*. After resolution of the digestion reaction using agarose gel electrophoresis and visualization with EtBr staining, the frequency of NHEJ-mediated indels was estimated for each founder by quantification of ethidium bromide staining and densitometry of the resistant band relative to wild-type fragments using ImageJ 1.49. Fragments corresponding to mutated alleles were gel purified, subcloned and sequenced using the same primers. Larvae presenting a high degree of targeting in somatic cells were selected for further analysis and reared to adulthood. Adults were hand-paired in individual cages either with virgin wild-type or mosaic individuals of the opposite sex to establish lines. Eggs were collected from each line, and the hatched larvae were raised and screened for the presence of mutated alleles as described above.

#### **Analysis of CRISPR/Cas9-induced mutations using a Cas9 *in vitro* cleavage assay**

For *in vitro* cleavage assay, recombinant Cas9 protein was produced in *E. coli* using a pET28-Cas9 expression plasmid purchased from PNA Bio, Inc. The recombinant Cas9 protein was expressed in BL21 (DE3) competent cells induced with IPTG and purified, as previously described (KIM *et al.* 2014). Cas9 protein was purified over a Nickel column using Ni-NTA His-Bind Resin (Novagen) and analyzed by SDS-PAGE. For analysis of CRISPR-Cas9 induced mutations, PCR fragments flanking the targeted

regions were amplified with *cry2* and *clock* primers given in Table S1. PCR products were purified using 1.5X modified Sera-Mag Magnetic Speed-beads (Fisher) as described by Rohland and Reich (ROHLAND and REICH 2012) and resuspended in 10  $\mu$ l of water. The Cas9-based cleavage assay was performed as previously described (KIM *et al.* 2014), with slight modifications. Purified PCR products (150-200 ng) were incubated for 3h at 37°C with purified Cas9 protein (100 ng), sgRNA (100-300 ng) used for targeting, BSA (1  $\mu$ g/ $\mu$ l; New England Biolabs, NEB) and NEB Buffer 3 (1X). After incubation, 4  $\mu$ g of RNase A (Amresco) was added and reactions were incubated for an additional 2h at 37°C. Reactions were stopped using a 6X stop solution described in (KIM *et al.* 2014). Products were purified using 1.5X modified Sera-Mag Magnetic Speed-beads before being resolved with agarose gel electrophoresis and EtBr staining. The frequency of NHEJ-mediated indels was estimated as described for TALENs, and fragments corresponding to mutated alleles were gel purified and sequenced using the same primers used to generate the PCR amplicons. To assess germline targeting rates, larvae were selected and their progeny recovered as described above.

### **Eclosion behavior**

Eclosion behavior was performed as previously described with slight modifications (MERLIN *et al.* 2013). Monarch larvae were housed in LD 15:9 in a Percival incubator at 25°C and 70% humidity through pupation and pupae were transferred to constant darkness (DD) one or two days prior to the predicted time of adult eclosion. Eclosion behavior was monitored continuously over time using a digital video recorder and night vision security cameras (Night Owl Security L-45-4511) mounted inside the incubator. Eclosion data were analyzed and plotted as 1-hour bins.

## Real-time qPCR

Brains from adult monarch butterflies entrained to seven LD cycles were dissected under red light on the first day of transfer into DD. Dissections were performed in 0.5X RNA later (Ambion) to avoid RNA degradation, and brains free of eye photoreceptors were stored at -80°C until use. Total RNA was extracted using RNeasy Mini kit (Qiagen), treated with RQ1 Dnase (Promega), and random hexamers (Promega) were used to prime reverse transcription with Superscript II Reverse Transcriptase (Invitrogen), all according to the manufacturers' instructions. Quantifications of clock gene expression were done using SYBR Green-based real-time quantitative PCR assays with an ABI 7500 Fast Real-Time PCR system (Applied biosystems). PCR reactions were assembled by combining two master mixes: the first mix contained 5 µl of iTaq Universal SYBR® Green Supermix (Bio-Rad) and forward and reverse primers (5 pmol each) per reaction and the second mix contained approximately 75 ng of cDNA template and the water needed to bring each reaction to a final volume of 10 µl. Each reaction was assembled by consecutively aliquoting each mix into a PCR plate. The monarch *per*, *tim*, *clock* and control *rp49* primers were as follows (F, forward primer; R, reverse primer): monarch *perF*, AGTGAAGCGTCCCTCAAAACA; *perR*, TGGCGACGAGCATCTGTGT; monarch *timF*, GGACGGCAGCGGATACG; *timR*, CGCCGTTTCGCACACA; monarch *rp49F*, TGC GCAGGCGTTTAAAGG; *rp49R*, TTGTTTGATCCGTAACCAATGC; monarch *clockF*, CATACCGGTTATAGTCGCACAAGA; *clockR*, GATGGGTTTCGCTGCAACTG. The near 100% efficiency of each primer set was validated by determining the slope of Ct versus dilution plot on a 3 x 10<sup>4</sup> dilution series. Individual reactions were used to quantify

each RNA level in a given cDNA sample, and the average Ct from duplicated reactions within the same run was used for quantification. The data for each gene in a given condition (genotype and time of day) were normalized to *rp49* as an internal control, and normalized to the mean of one time point within a set for statistics.

### **Statistical analysis**

P-values were calculated using a Student's *t*-test and one-way or two-way ANOVAs using the QI Macros Statistical software for Excel.

## **RESULTS**

### **Knocking out monarch clock genes using TALENs**

To assess the capacity of TALENs to efficiently produce loss-of-function mutations in the monarch butterfly, we chose to target two clock genes: *cryptochrome 2* (*cry2*), the main transcriptional repressor of the monarch circadian clock (ZHU *et al.* 2008; MERLIN *et al.* 2013), and *clock* (*clk*), a component of the heterodimeric transcriptional activator complex of the clock (ZHU *et al.* 2008).

The *cry2* locus was selected as an ideal candidate for direct comparison of targeting efficiencies between different nuclease-mediated targeting technologies, including ZFNs (MERLIN *et al.* 2013). The TALE binding sites were therefore chosen to overlap with the binding sites of the ZFN pair previously used to target the second exon of *cry2* (Figure 1A). To test the efficiency of TALEN-mediated targeted mutagenesis at the *cry2* locus, embryos in early stages of development (i.e. 20 min post-oviposition) were injected with 0.5 µg/µl of mRNAs encoding the TALEN pair. Given that fertilization occurs at oviposition and the zygote nucleus is formed and starts dividing

soon after, injecting shortly after the eggs have been laid likely increases the chances to target the nuclei that will later give rise to germline precursors (MERLIN *et al.* 2013). TALEN-induced mutations (insertions or deletions; indels) were screened in surviving larvae using polymerase chain reaction (PCR)-based assays from clipped cuticular expansions, as previously described (MERLIN *et al.* 2013).

To facilitate the screening of indels at the targeted site and readily assess the level of mosaicism in potential founders, the spacer between the TALE binding sites was chosen to contain an endogenous restriction site, *EagI* (Figure 1A). PCR amplicons carrying indels that disrupt the enzyme recognition site become resistant to digestion, and the percentage or rate of NHEJ-mediated lesions can be estimated in somatic tissues by quantifying the proportion of uncleaved mutated fragments relative to cleaved wild-type ones (Figure 1B). Complete digestion of the PCR-amplified fragments from non-injected wild-type larvae was observed (Figure 1B). Digestion-resistant fragments were found in somatic tissue of ~65% of the larvae surviving injections (11/17) from a pool of 86 injected embryos. In these targeted individuals, the rates of mutagenic lesions were estimated to range from 5 to 74% (Figure 1B, Figure S1A and Table 1). This stands in striking contrast to the efficiency of ZFNs-mediated targeted mutagenesis previously reported at this locus, with only 4 somatic mutants recovered out of ~1200 injected eggs, albeit with comparable mosaicism levels (Table 1; (MERLIN *et al.* 2013)). Cloning and sequencing of the purified resistant PCR fragments from one individual confirmed the presence of frameshifting microdeletions at the targeted site (Figure 1C), similar to what has been previously described with ZFNs (MERLIN *et al.* 2013).

To determine whether TALEN-induced mutations could be transmitted through the monarch germline, we intercrossed two adult mosaic males to two adult mosaic females presenting the highest levels of mosaicism in somatic cells (two mating pairs; Figure 1B, red stars). From the 24 F<sub>1</sub> larvae that eclosed out of the 73 oviposited eggs, 1 carried a mutated *cry2* allele (~4 %; Table 1).

To expand our analysis and demonstrate that TALENs can be used efficiently to target other loci within the monarch genome, we next applied this technology to the *clk* locus. The choice of this target was motivated by the fact that a loss-of-function monarch *clk* knockout would constitute a complementary reagent to the already existing *cry2* monarch knockout for future functional studies on the circadian control of the monarch butterfly seasonal long-distance migration. Indeed, monarch *clk* has been shown *in vitro* to function in a manner similar to its *Drosophila* homologue as a transcriptional activator within the CLK- CYC heterodimeric complex, having therefore an opposite role to *cry2* (ZHU *et al.* 2008); (ALLADA *et al.* 1998).

A high quality TALEN site was identified in exon 3 of the monarch *clk* gene, with the two TALE binding sites flanking a spacer containing an endogenous restriction site, *NgoMIV*, which can be used to detect the presence of indels and estimate targeting frequencies, as described above (Figure 2A). Frameshift mutations at this site are predicted to result in the production of truncated, non-functional, proteins lacking the basic-helix-loop-helix-PER-ARNT-SIM domain required for the formation of the functional CLK:CYC heterodimer (ALLADA *et al.* 1998). To test the efficiency of TALEN-mediated targeted mutagenesis at the *clk* locus, embryos were injected with 0.5 µg/µl of mRNAs encoding the TALEN pair, and the presence of indels at the targeted site

was examined in surviving larvae by restriction digest. Approximately 19% of injected embryos (17/90) survived injection, and all exhibited mosaicism at the targeted site with frequencies of mutated lesions ranging from 3 to 86% (Figure 2B, Figure S1B and Table 1). Cloning and sequencing of the purified resistant PCR fragments from one individual confirmed the presence of frameshifting microdeletions at the targeted site (Figure 2C).

To determine the germline transmission rates of TALEN-induced mutations and establish a *clk* knockout line, we screened the progeny of two intercrossed mosaic males and two mosaic females, selected among the most highly targeted individuals (Figure 2B, red stars). Out of 68 F<sub>1</sub> larvae, ~52% carried at least one mutated allele, with 21 heterozygotes and 12 homozygotes. Sexing the corresponding adults revealed that all homozygous mutants were females, all heterozygous mutants were males, and wild-type butterflies were of both sexes. This suggested that *clk* is sex-linked and located on the Z sex chromosome (in lepidopterans, the female is the heterogametic (ZW) sex and the male is homogametic (ZZ); Figure 2D), adding to the list of sex-linked clock genes in Lepidoptera (GOTTER *et al.* 1999; ABE *et al.* 2010). To unambiguously determine that the homozygous mutants were resulting from sex-based segregation, and not from biallelic mutations transmitted from two mosaic parents, we sequenced the resistant fragments of four female and seven male F<sub>1</sub> progeny. Consistent with *clk* being linked to the Z chromosome, we found a single mutation per animal tested. All these mutations were microdeletions, and a 4bp-del mutant that would generate a truncated, non-functional, protein (Figure 2E) was chosen to establish a *clk* knockout line. Genetic evidence that monarch *clk* was Z-linked prompted us to test whether it would be dosage compensated between sexes, as sex chromosome dosage compensation in the lepidopteran clade is still

debated (WALTERS *et al.* 2015). In the absence of dosage compensation, *clock* expression in females would be half the expression level in males. We found no statistical differences in *clock* expression in wild-type monarch brains between sexes, suggesting complete dosage compensation in the monarch butterfly at least at this Z-linked locus (Figure 2F, Student's *t*-test,  $p=0.37$ ; Figure S2).

### **CRISPR/Cas9 mediate efficient clock gene knockouts in the monarch butterfly using single sgRNA and dual sgRNAs**

With a more straightforward design and faster construction time, the bacterial-derived CRISPR/Cas9 system has emerged as an alternative to TALENs for the generation of gene knockouts. We tested its efficiency in the monarch butterfly, focusing our study on the genes previously targeted with ZFNs and TALENs, *clk* and *cry2*, for direct comparison.

For targeting of *clk*, we co-injected monarch eggs with 0.1  $\mu\text{g}/\mu\text{l}$  of an *in vitro* transcribed single guide RNA (sgRNA) complementary to a 19-bp DNA sequence in the second exon, along with 0.5  $\mu\text{g}/\mu\text{l}$  of mRNA encoding Cas9 (Figure 3A). To screen for CRISPR/Cas9-mediated indels in surviving larvae and quantitatively estimate their occurrence rate in somatic cells, we used a Cas9-based *in vitro* cleavage assay modified from (KIM *et al.* 2014). Unlike the T7 endonuclease I and mismatch-sensitive surveyor assays (KIM *et al.* 2009), this assay is quantitative and substantially more cost effective than high-throughput sequencing (KIM *et al.* 2014). It relies on the sgRNA used for introducing *in vivo* mutations and a recombinant Cas9 protein to detect the presence of lesions via the loss of sensitivity of PCR fragments spanning the targeted site to Cas9. In

this assay, the sgRNA hybridizes with its target sequence on wild-type fragments but fails to do so with mutated target sequences, resulting in cleavage of the wild-type sequences only. The occurrence and frequency of NHEJ-mediated indels can then be estimated by quantifying the proportion of uncleaved PCR fragments relative to the total PCR product (cleaved and uncleaved; Figure 3B). Comparable to what we found with TALENs, ~20% of the injected embryos developed into healthy larvae (19/93; Table 1), and ~50% of those individuals presented indels at rates ranging from 3 to 28% (Figure 3C; Table 1). Cloning and sequencing of the resistant PCR fragments from two potential founders confirmed the presence of indels in somatic cells (Figure 3D). Interestingly, the same mutations were recovered in the progeny of the founders backcrossed to wild-type monarchs (Figure 3D and 3E), supporting the hypothesis that injecting very early into newly oviposited eggs is a robust strategy to achieve high germline transmission rates in the monarch, regardless of the genome targeting technology applied.

We next investigated whether CRISPR/Cas9 could be widely applicable for genome editing in the monarch butterfly by targeting an additional locus, *cry2* (Figure 4). While the use of a single sgRNA proved to be successful for generating small indels, inducing deletions larger than a few base-pairs with the simultaneous use of two sgRNAs would ultimately facilitate the screening of mutagenic lesions solely based on PCR-based fragment size differences (LIU *et al.* 2014). We thus co-injected newly oviposited monarch eggs with 0.1  $\mu\text{g}/\mu\text{l}$  of two *in vitro* transcribed sgRNAs respectively complementary to 19-bp DNA sequences in the second and third exons of *cry2*, along with 0.5  $\mu\text{g}/\mu\text{l}$  of mRNA encoding Cas9 (Figure 4A). Out of 88 injected embryos, 21.6 % larvae hatched and about 15% (13/88) developed into healthy larvae. Similar to our

findings at the *clk* locus, these hatching and survival rates were comparable to those obtained with TALENs injections (Table 1), suggesting that both TALENs and CRISPR/Cas9 are substantially less toxic than ZFNs to the butterflies. To simultaneously screen for both the presence of small indels and larger deletions in the 13 surviving mosaic larvae, we performed *in vitro* Cas9-based cleavage assays on 1403-bp PCR fragments using either sgRNA (sgRNA 1 and sgRNA2; Figure 4B). However, only results with the sgRNA2 can inform for both the presence of small indels and larger deletions as a deletion resulting from DSBs between the two sgRNAs would result in a fragment of similar size than the cleaved wild-type product by sgRNA1 (~ 1085-bp; Figure 4A). Nevertheless, we detected a surprisingly high frequency of targeted mutagenesis at the site targeted with sgRNA1 in 8 out of 13 founders G<sub>0</sub>, with rates ranging from 45 to 100% (Figure 4B, left panel) and lesions harboring a variety of small indels in somatic cells (Figure 4C).

At the site targeted with sgRNA2, we only detected mutagenic lesions in two of the 13 larvae genotyped. These lesions were however large deletions of ~350-bp (Figure 4B, right panel; founders #2 and #5). Cloning of the corresponding PCR fragments confirmed that they resulted in a 349-bp deletion between the two target sites (Figure 4B, Figure S3). Although these data suggested that the DNA cleavage mediated by sgRNA2 was likely inefficient, they also suggested that when two DSBs were introduced, the occurrence of large deletions was likely favored. Testing whether these large deletions were transmitted to the germline was hindered by the fact the butterflies did not survive to reproductive maturity. We were however able to provide evidence of germline transmitted indels generated with sgRNA1 in the two larvae that eclosed from two

backcrossed founders presenting the highest rates of somatic targeting (Figure 4C, Table 1). This low progeny number might be due to the fact that high targeting of monarch *cry2* negatively impacts egg production, as we previously observed that *cry2* knockout female monarchs develop few to no mature eggs (unpublished data).

### **Targeted mutagenesis of *clk* disrupts the monarch circadian clockwork**

The occurrence of a mammalian-like core component in the monarch (*cry2*) and the fact that mouse CLK is not required for central circadian clock function (DEBRUYNE *et al.* 2006) prompted us to characterize monarch CLK function *in vivo* to unambiguously determine whether it is essential for central clock function in the brain. To assess the effect of knocking out monarch CLK on circadian behavior, we examined the timing of butterfly pupal eclosion (i.e. the emergence of the adult from its pupal case), a robust and easily tractable behavior under the control of the brain circadian clock (FROY *et al.* 2003; MERLIN *et al.* 2013). We focused our analysis on female knockouts because the location of *clk* on the Z chromosome permitted the recovery of female knockouts from every backcross of heterozygote males to wild-type females. This strategy allowed for a rapid behavioral analysis of *clk* female knockouts (*clk<sup>out</sup>*) generated through three independent and consecutive backcrosses, thereby reducing the chances that the phenotype observed result from potential off-target effects. We found that the circadian timing of adult eclosion was disrupted in *clk* knockouts, while sibling wild-type and heterozygote butterflies eclosed rhythmically, exhibiting a peak of eclosion in the early subjective day at the populational level (Figure 5A; one-way ANOVA:  $p < 0.0002$ ; Tukey post-hoc test:  $clk^{+/+}$  vs.  $clk^{+/-}$ ,  $p > 0.05$ ;  $clk^{+/+}$  vs.  $clk^{out}$ ,  $p < 0.01$ ;  $clk^{+/-}$  vs.  $clk^{out}$ ,  $p < 0.01$ ).

To determine how the molecular clock was affected by CLK deficiency, we quantified by quantitative real-time PCR the expression of two core clock components, *period* and *timeless*, from wild-type and *clk<sup>out</sup>* homozygous monarch brains, the location of circadian clocks controlling insect eclosion behavior (ZHU *et al.* 2008). While *period* and *timeless* robustly cycled in brains of wild-type butterflies, their rhythms were abolished in brains of CLK<sup>out</sup> and their expression levels were constitutively low (Figure 5B; two-way ANOVA, interaction genotype × time: *per*,  $p < 0.0001$ ; *tim*,  $p < 0.01$ ). These behavioral and molecular data confirm that monarch CLK is functioning as an essential component of the CLK:CYC transcriptional activator complex of the monarch circadian clock, similar to its *Drosophila* homologue (ALLADA *et al.* 1998; MAHESH *et al.* 2014).

## DISCUSSION

The ability to harness efficient reverse-genetic approaches in butterflies is critical for these species to provide powerful model systems. Even though the use of artificial nucleases has proven to be successful for targeted mutagenesis in two butterfly species, the monarch butterfly and the swallowtail butterfly (MERLIN *et al.* 2013; LI *et al.* 2015), the rapid adaptation of these technologies to other lepidopterans has been hindered by an absence of straightforward and practical strategies to introduce targeted mutations and identify carriers in particular for genes without morphological phenotypes. Here, we demonstrate that TALENs and CRISPR/Cas9 mediate highly efficient, heritable, targeted mutagenesis at two genomic loci in the monarch butterfly. Our study provides a simple workflow to recover germline mutants in the progeny of a small number of mosaic butterflies obtained from less than 100 injected eggs and establish knockout lines in

approximately three months with minimal injection and screening efforts. In the process of expanding the monarch genetic toolbox, we also generated a monarch knockout for *clk* and defined its critical function as a transcriptional activator of the circadian clock used by monarchs during their migration, providing a valuable resource to further our understanding of the clock control of this spectacular behavior.

As previously reported, the use of TALENs and/or CRISPR/Cas9 for non-model insect genome editing confers several advantages over ZFNs: target selection is more flexible, their assembly is easier (CERMAK *et al.* 2011; JINEK *et al.* 2012), and their success in attaining high mutagenesis efficiencies is greater (MA *et al.* 2012; WATANABE *et al.* 2012; SAJWAN *et al.* 2013; WANG *et al.* 2013; LIU *et al.* 2014; WEI *et al.* 2014; GILLES *et al.* 2015; LI *et al.* 2015). Our study in the monarch shows that both technologies substantially increase targeting efficiencies in somatic cells by almost an order of magnitude in comparison to ZFNs, with rates as high as 100% in the case of CRISPR/Cas9. The targeting efficiencies we observed appeared however to be locus-dependent, with greater efficiencies of *cry2* targeting using CRISPR/Cas9 and *clk* targeting using TALENs. Whether these differences are due to different sgRNAs achieving different mutation rates, or to differences in local DNA accessibility remains unclear.

In spite of the broad interest in the use of CRISPR/Cas9 for targeted mutagenesis in non-model insects, the general lack of simple and inexpensive assays to quantitatively assess rates of targeting in individual mutants with no visible phenotype has limited the generation of knockout lines to only a few species (GRATZ *et al.* 2013; MA *et al.* 2014; WEI *et al.* 2014; AWATA *et al.* 2015; BASU *et al.* 2015; DONG *et al.* 2015; GILLES *et al.*

2015; KISTLER *et al.* 2015). The quantitative methods that we used, which rely on Cas9-based *in vitro* cleavage assays (KIM *et al.* 2014), not only permits the rapid identification of highly targeted founders, but does so at low cost, allowing us to introduce an additional screening step in our workflow. Selecting only a few of the most promising candidates for crosses is sufficient to recover germline mutants, which avoids the burden associated with breeding a large number of individuals and screening their progeny for mutant alleles. Similar strategies could be implemented to accelerate the use of CRISPR/Cas9-mediated genome editing in other non-model insect species. Compared to TALENs, CRISPR/Cas9 can also be more readily adapted for multiplex gene editing to engineer large genomic deletions or knockout several genes simultaneously (MA *et al.* 2014). While our study provides a proof-of-principle for CRISPR/Cas9-mediated multiplexing in the monarch butterfly, its success rate could be greatly enhanced by testing individually each sgRNA and selecting the most active ones in multiplex experiments. Importantly, because of its simplicity of use and efficacy, CRISPR/Cas9 offers the possibility of developing knock-in approaches in the monarch to tag protein coding genes and facilitate site-specific recombination and the creation of conditional alleles.

In conclusion, the methods we established for the rapid generation of heritable targeted mutations in the monarch butterfly should be widely applicable to other lepidopterans. Nuclease technology will also ultimately enable the systematic dissection of the monarch molecular clock mechanisms and the genetic basis of its long-distance migration. The threat to the North American monarch migration posed by ongoing climate change and habitat loss both across the major flyway in the US and at the

overwintering sites in Mexico has prompted national conservation efforts to protect and restore their habitat. Understanding the genetic system underlying the monarch migratory program may help strengthen current conservation strategies.

### **Acknowledgments**

We thank Lauren Beaver, Emily McKnight, Shayna Groves, Justin Vann, Tarika Khaladkar and Melanie Goodman for help with upkeep of the monarch husbandry, Chip Taylor and Ann Ryan (Kansas University) for providing us the semi-artificial diet, and Jerome Menet (Texas A&M University) for valuable discussions. We also thank Dr. Ty Kim (PNA Bio, Inc.) for generously sending us his pET28-Cas9 plasmid, and Dr. Steve Lockless (Texas A&M University) for guidance in purifying the recombinant Cas9 protein. The work was supported by start up funds from Texas A&M University and NSF IOS-1456985 to CM, NIH 1R21A1097978-01 to SAW and L. Zwiebel, and NIH GM086794 to SMR. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions: CM and SAW conceived the study, which was initiated when CM was a postdoctoral fellow in the lab of SMR. MJM, YZ, MSE and CM performed the experiments. MJM, YZ and CM analyzed the data. CM wrote the manuscript and all authors provided comments and edits.

## References:

- Abe, H., T. Fujii and T. Shimada, 2010 Sex chromosomes and sex determination in *Bombyx mori*., pp. 65-87 in *Molecular Biology and Genetics of the Lepidoptera*, edited by M. R. Goldsmith and F. Marec. CRC Press, Boca Raton, FL USA.
- Allada, R., N. E. White, W. V. So, J. C. Hall and M. Rosbash, 1998 A mutant *Drosophila* homolog of mammalian Clock disrupts circadian rhythms and transcription of period and timeless. *Cell* 93: 791-804.
- Awata, H., T. Watanabe, Y. Hamanaka, T. Mito, S. Noji *et al.*, 2015 Knockout crickets for the study of learning and memory: Dopamine receptor Dop1 mediates aversive but not appetitive reinforcement in crickets. *Sci Rep* 5: 15885.
- Basu, S., A. Aryan, J. M. Overcash, G. H. Samuel, M. A. Anderson *et al.*, 2015 Silencing of end-joining repair for efficient site-specific gene insertion after TALEN/CRISPR mutagenesis in *Aedes aegypti*. *Proc Natl Acad Sci U S A* 112: 4038-4043.
- Carroll, D., 2014 Genome engineering with targetable nucleases. *Annu Rev Biochem* 83: 409-439.
- Cermak, T., E. L. Doyle, M. Christian, L. Wang, Y. Zhang *et al.*, 2011 Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* 39: e82.
- Debruyne, J. P., E. Noton, C. M. Lambert, E. S. Maywood, D. R. Weaver *et al.*, 2006 A clock shock: mouse CLOCK is not required for circadian oscillator function. *Neuron* 50: 465-477.

- Dong, S., J. Lin, N. L. Held, R. J. Clem, A. L. Passarelli *et al.*, 2015 Heritable CRISPR/Cas9-mediated genome editing in the yellow fever mosquito, *Aedes aegypti*. *PLoS One* 10: e0122353.
- Froy, O., A. L. Gotter, A. L. Casselman and S. M. Reppert, 2003 Illuminating the circadian clock in monarch butterfly migration. *Science* 300: 1303-1305.
- Gaj, T., C. A. Gersbach and C. F. Barbas, 3rd, 2013 ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* 31: 397-405.
- Gilles, A. F., J. B. Schinko and M. Averof, 2015 Efficient CRISPR-mediated gene targeting and transgene replacement in the beetle *Tribolium castaneum*. *Development* 142: 2832-2839.
- Gotter, A. L., J. D. Levine and S. M. Reppert, 1999 Sex-linked period genes in the silkworm, *Antheraea pernyi*: implications for circadian clock regulation and the evolution of sex chromosomes. *Neuron* 24: 953-965.
- Gratz, S. J., A. M. Cummings, J. N. Nguyen, D. C. Hamm, L. K. Donohue *et al.*, 2013 Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. *Genetics* 194: 1029-1035.
- Guerra, P. A., C. Merlin, R. J. Gegear and S. M. Reppert, 2012 Discordant timing between antennae disrupts sun compass orientation in migratory monarch butterflies. *Nat Commun* 3: 958.
- Gupta, A., V. L. Hall, F. O. Kok, M. Shin, J. C. McNulty *et al.*, 2013 Targeted chromosomal deletions and inversions in zebrafish. *Genome Res* 23: 1008-1017.
- Hwang, W. Y., Y. Fu, D. Reyon, M. L. Maeder, S. Q. Tsai *et al.*, 2013 Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* 31: 227-229.

- Jao, L. E., S. R. Wente and W. Chen, 2013 Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc Natl Acad Sci U S A* 110: 13904-13909.
- Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna *et al.*, 2012 A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337: 816-821.
- Kim, H. J., H. J. Lee, H. Kim, S. W. Cho and J. S. Kim, 2009 Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res* 19: 1279-1288.
- Kim, J. M., D. Kim, S. Kim and J. S. Kim, 2014 Genotyping with CRISPR-Cas-derived RNA-guided endonucleases. *Nat Commun* 5: 3157.
- Kistler, K. E., L. B. Vosshall and B. J. Matthews, 2015 Genome engineering with CRISPR-Cas9 in the mosquito *Aedes aegypti*. *Cell Rep* 11: 51-60.
- Li, X., D. Fan, W. Zhang, G. Liu, L. Zhang *et al.*, 2015 Outbred genome sequencing and CRISPR/Cas9 gene editing in butterflies. *Nat Commun* 6: 8212.
- Liu, Y., S. Ma, X. Wang, J. Chang, J. Gao *et al.*, 2014 Highly efficient multiplex targeted mutagenesis and genomic structure variation in *Bombyx mori* cells using CRISPR/Cas9. *Insect Biochem Mol Biol* 49: 35-42.
- Ma, S., J. Chang, X. Wang, Y. Liu, J. Zhang *et al.*, 2014 CRISPR/Cas9 mediated multiplex genome editing and heritable mutagenesis of BmKu70 in *Bombyx mori*. *Sci Rep* 4: 4489.
- Ma, S., S. Zhang, F. Wang, Y. Liu, Y. Liu *et al.*, 2012 Highly efficient and specific genome editing in silkworm using custom TALENs. *PLoS One* 7: e45035.

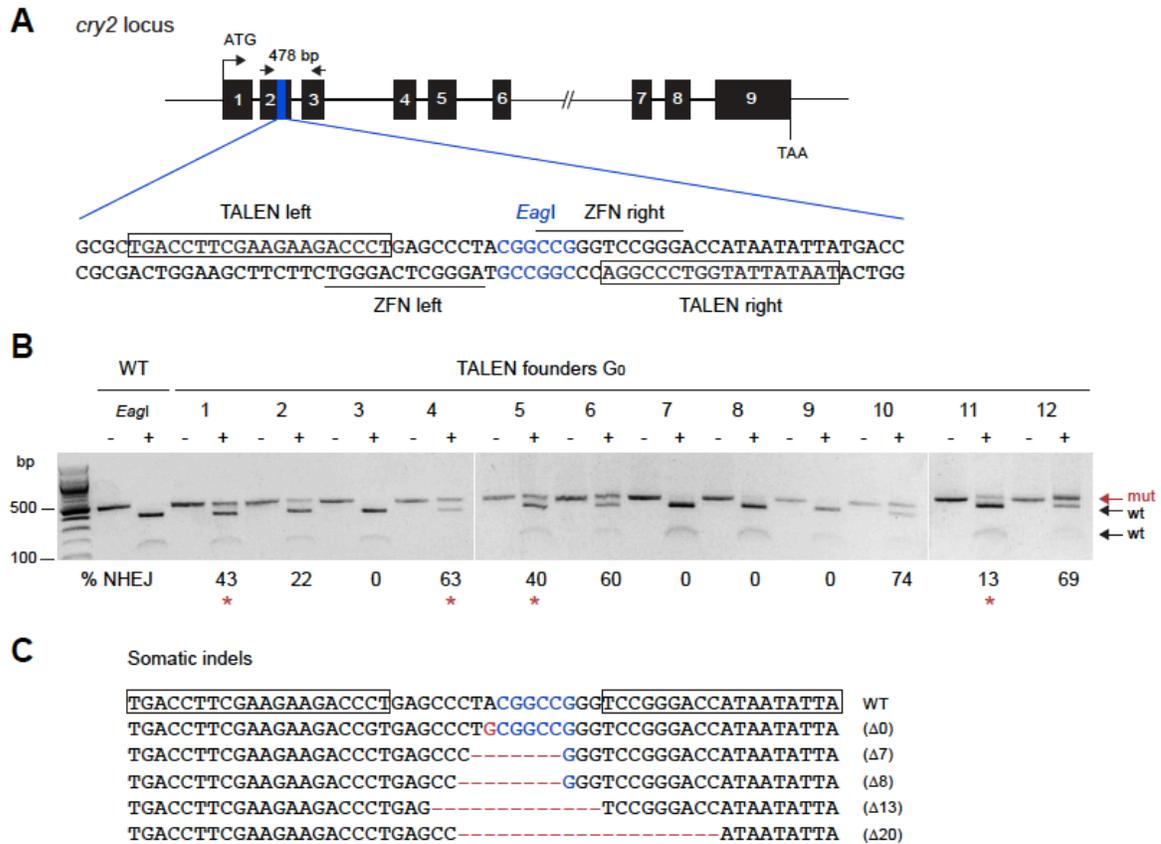
- Mahesh, G., E. Jeong, F. S. Ng, Y. Liu, K. Gunawardhana *et al.*, 2014 Phosphorylation of the transcription activator CLOCK regulates progression through a approximately 24-h feedback loop to influence the circadian period in *Drosophila*. *J Biol Chem* 289: 19681-19693.
- Merlin, C., L. E. Beaver, O. R. Taylor, S. A. Wolfe and S. M. Reppert, 2013 Efficient targeted mutagenesis in the monarch butterfly using zinc-finger nucleases. *Genome research* 23: 159-168.
- Merlin, C., R. J. Gegear and S. M. Reppert, 2009 Antennal circadian clocks coordinate sun compass orientation in migratory monarch butterflies. *Science* 325: 1700-1704.
- Mouritsen, H., and B. J. Frost, 2002 Virtual migration in tethered flying monarch butterflies reveals their orientation mechanisms. *Proc Natl Acad Sci U S A* 99: 10162-10166.
- Perez, S. M., O. R. Taylor and R. Jander, 1997 A sun compass in monarch butterflies. *Nature* 387: 29.
- Reppert, S. M., R. J. Gegear and C. Merlin, 2010 Navigational mechanisms of migrating monarch butterflies. *Trends in neurosciences* 33: 399-406.
- Reppert, S. M., P. A. Guerra and C. Merlin, 2015 Neurobiology of Monarch Butterfly Migration. *Annu Rev Entomol.*
- Rohland, N., and D. Reich, 2012 Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res* 22: 939-946.

- Sajwan, S., Y. Takasu, T. Tamura, K. Uchino, H. Sezutsu *et al.*, 2013 Efficient disruption of endogenous *Bombyx* gene by TAL effector nucleases. *Insect Biochem Mol Biol* 43: 17-23.
- Sander, J. D., J. R. Yeh, R. T. Peterson and J. K. Joung, 2011 Engineering zinc finger nucleases for targeted mutagenesis of zebrafish. *Methods Cell Biol* 104: 51-58.
- Urquhart, F. A., 1987 *The monarch butterfly : international traveler*. Nelson-Hall, Chicago, Ill.
- Walters, J. R., T. J. Hardcastle and C. D. Jiggins, 2015 Sex Chromosome Dosage Compensation in *Heliconius* Butterflies: Global yet Still Incomplete? *Genome Biol Evol* 7: 2545-2559.
- Wang, F., S. Ma, H. Xu, J. Duan, Y. Wang *et al.*, 2013 High efficiency system for construction and evaluation of customized TALENs for silkworm genome editing. *Mol Genet Genomics* 288: 683-690.
- Watanabe, T., H. Ochiai, T. Sakuma, H. W. Horch, N. Hamaguchi *et al.*, 2012 Non-transgenic genome modifications in a hemimetabolous insect using zinc-finger and TAL effector nucleases. *Nat Commun* 3: 1017.
- Wei, W., H. Xin, B. Roy, J. Dai, Y. Miao *et al.*, 2014 Heritable genome editing with CRISPR/Cas9 in the silkworm, *Bombyx mori*. *PLoS One* 9: e101210.
- Zhan, S., C. Merlin, J. L. Boore and S. M. Reppert, 2011 The monarch butterfly genome yields insights into long-distance migration. *Cell* 147: 1171-1185.
- Zhan, S., W. Zhang, K. Niitepold, J. Hsu, J. F. Haeger *et al.*, 2014 The genetics of monarch butterfly migration and warning colouration. *Nature* 514: 317-321.

Zhu, H., I. Sauman, Q. Yuan, A. Casselman, M. Emery-Le *et al.*, 2008 Cryptochromes define a novel circadian clock mechanism in monarch butterflies that may underlie sun compass navigation. PLoS Biol 6: e4.

FIGURE AND FIGURE LEGENDS

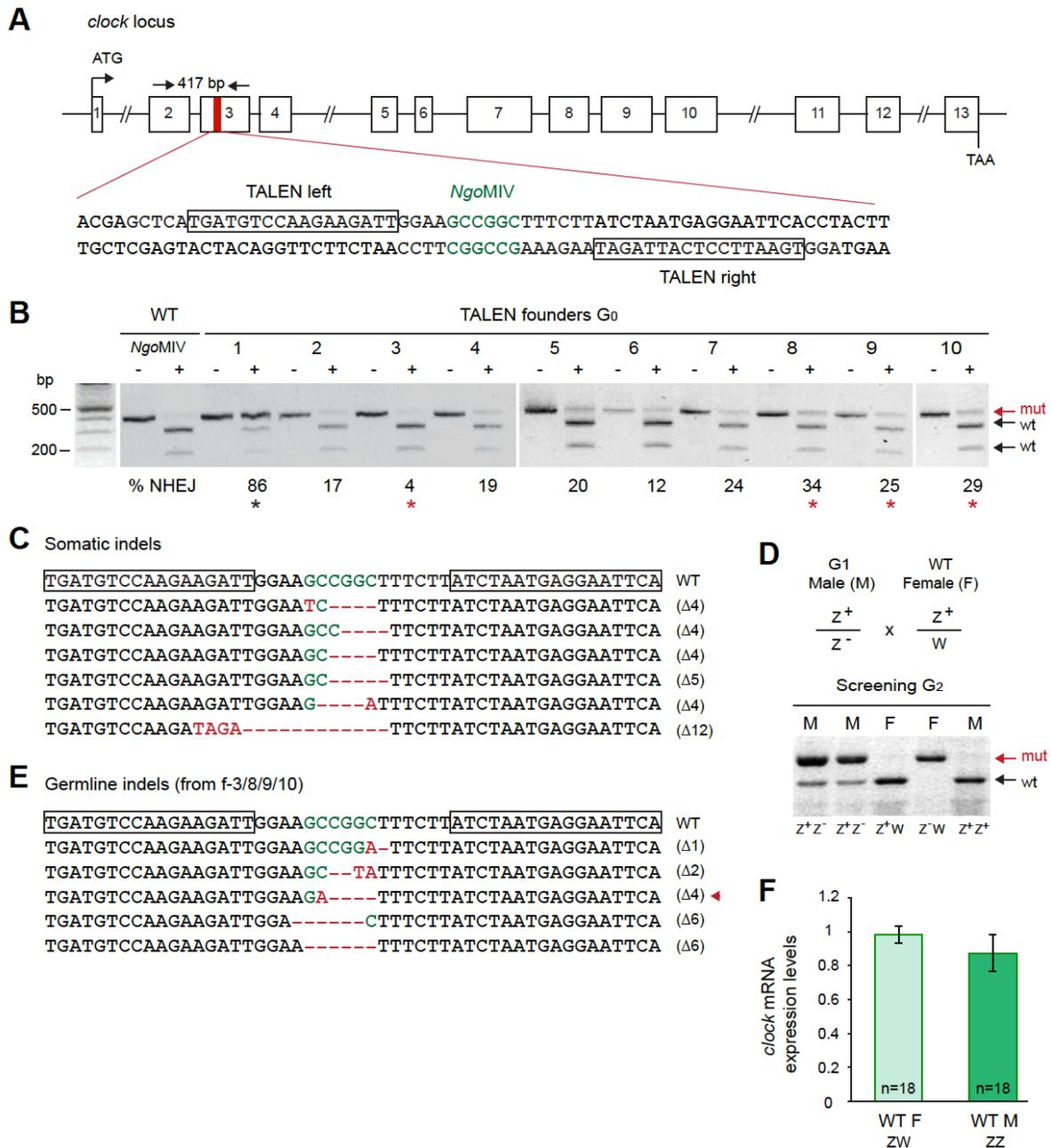
Figure 1



**Figure 1. Targeted mutagenesis induced by microinjection of *cry2* TALEN mRNAs into monarch butterfly embryos.** (A) Schematic diagram of the TALEN targeting sites. The black line depicts the genomic locus of monarch *cry2*, and the numbered black boxes represent the 9 exons of *cry2*. The blue region in exon 2 is expanded to provide the sequence that includes the TALEN binding sites (shown in boxes) and the previously described ZFN binding sites (underlined; (MERLIN *et al.* 2013)). The restriction site (*EagI*) in the spacer is shown in blue. Arrows on top represent the position of the primers used to amplify the 478-bp targeted region for analysis of mutagenic lesions in B (*cry2*

F1, *cry2* R1 in Table S1). (B) Targeted mutagenesis in representative founder G<sub>0</sub> butterflies validated by PCR and *EagI* digestion. For each founder, the PCR product was either not subjected (-) or subjected to *EagI* (+). Genomic amplicons carrying non-homologous end-joining (NHEJ)-induced mutations (mut; red arrow) leading to the loss of the *EagI* site are resistant to restriction enzyme digestion. The frequency of NHEJ-mediated indels provided under each founder was estimated by relative quantification of the resistant and digested products. Red stars, founders crossed to determine germline targeting rates. (C) TALEN-induced mutations in somatic cells of larvae surviving the injection. TALEN binding sites are boxed on the wild-type sequence. *EagI* site is in blue, and the positions of deletion and insertions are in red (dashes and letters, respectively).

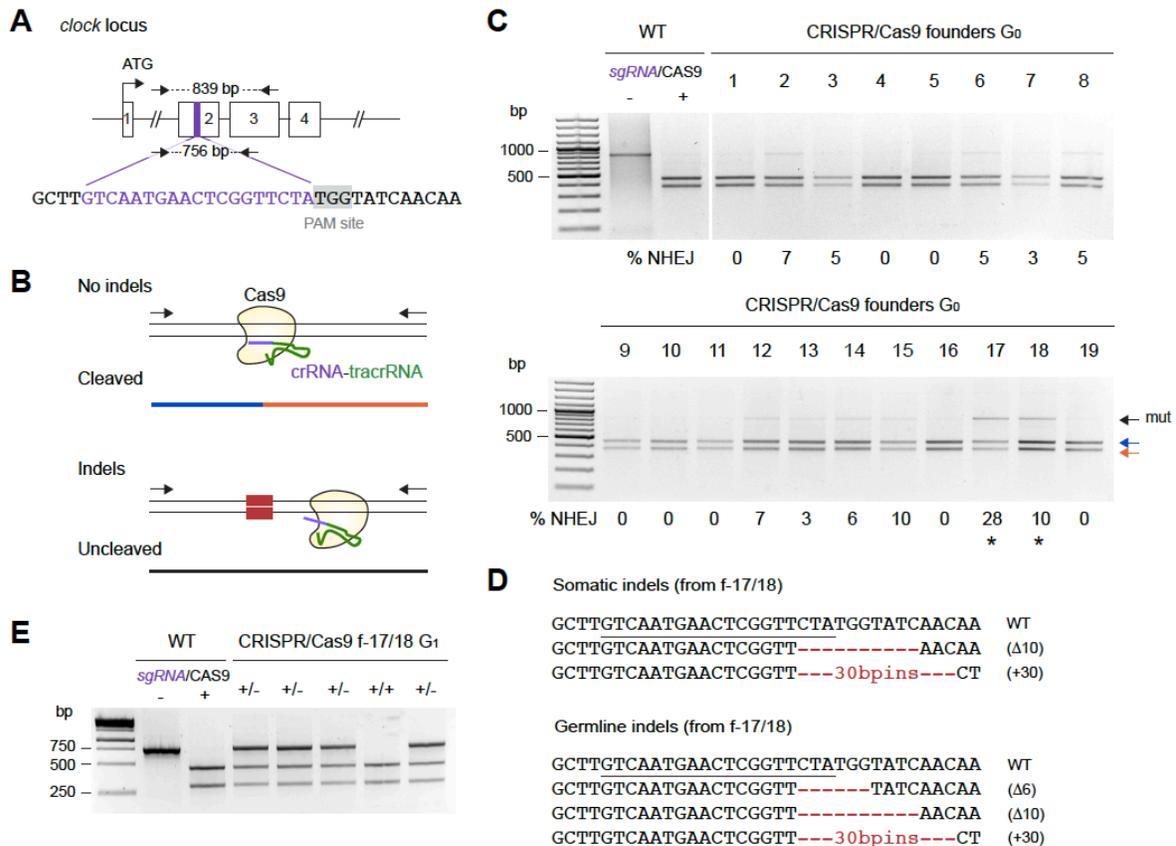
**Figure 2**



**Figure 2. TALEN-mediated targeted mutagenesis in somatic and germline cells of monarch butterfly *clock*.** (A) Schematic diagram of the monarch *clock* genomic locus and the TALEN targeting sites. The red region in exon 3 is expanded to provide the sequence that includes the TALEN binding sites (shown in boxes). The restriction site

(*NgomIV*) in the spacer is shown in green. Arrows on top represent the position of the primers used to amplify the 417-bp targeted region for analysis of mutagenic lesions in B (*clock* F1, *clock* R1 in Table S1). (B) Targeted mutagenesis in representative founder G<sub>0</sub> butterflies validated by PCR and *NgomIV* digestion. For each founder, the PCR product was either not subjected (-) or subjected to *NgomIV* (+). Genomic amplicons carrying NHEJ-induced mutations (mut; red arrow) leading to the loss of the *NgomIV* site are resistant to restriction enzyme digestion. Estimation of the frequency of NHEJ-mediated indels is provided under each founder. Red stars, founders crossed to determine germline targeting rates. Black star, highly targeted G<sub>0</sub> butterfly that did not survive to adulthood. (C) TALEN-induced mutations in somatic cells of larvae surviving the injection. TALEN binding sites are boxed on the wild-type sequence. *NgomIV* site is in green, and the positions of deletion and insertions are in red (dashes and letters, respectively). (D) The *clock* gene is located on the Z chromosome. Crosses between heterozygotes clock mutant (G<sub>1</sub>) and wild-type butterflies produce progeny that include heterozygote males (M) and homozygote females (F) for the mutated allele (mut; red arrow). Wt, wild-type siblings. (E) TALEN-induced mutations in the germline of founders 3, 8, 9 and 10. Red arrowhead, butterfly selected to establish a *clock* mutant line. (F) Relative expression level of *clock* in brains of wild-type monarch females (WT F) and males (WT M). Values are mean ± SEM of 18 animals, collected every three hours in triplicates over 24-hrs. *t*-test:  $p=0.37$ .

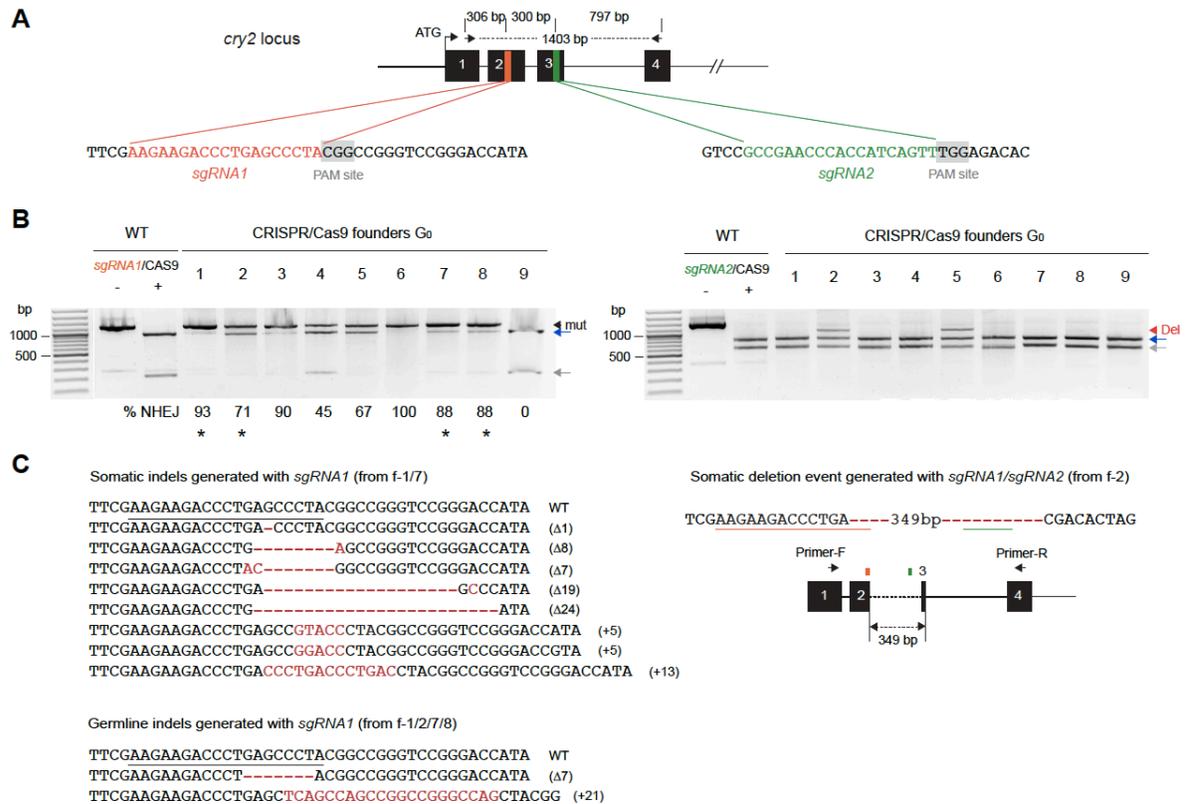
**Figure 3**



**Figure 3. Heritable CRISPR/Cas9-mediated targeted mutagenesis of monarch butterfly *clock*.** (A) Schematic of part of the monarch *clock* genomic locus containing the CRISPR/Cas9 target site. The purple region in exon 2 is expanded to provide the sequence targeted for genome editing by a single guide RNA (sgRNA, purple letters). The protospacer adjacent motif (PAM site, 5'-NGG-3') 3' of the target sequence is highlighted in gray. Arrows on top and bottom represent the positions of the primers used to amplify the 839-bp (*clock* F2, *clock* R1; Table S1) and 756-bp (*clock* F2, *clock* R2; Table S1) targeted regions for analysis of mutagenic lesions in C and E, respectively. (B) Diagram of the *in vitro* cleavage assay used to detect mutagenic lesions showing a PCR amplicon subjected to the sgRNA and the Cas9 protein. Blue and red lines correspond to

a wild-type (WT) genomic fragment cleaved by Cas9. Fragments with mutations induced by NHEJ at the site targeted by the sgRNA (red boxes) are uncleaved (black line). (C) Detection of mutagenic lesions at the *clock* locus in somatic cells of founder G<sub>0</sub> butterflies. For each founder, a PCR fragment was subjected to a Cas9-based *in vitro* cleavage assay. Blue and red arrows, WT fragments. Black arrow, amplicons carrying mutations (mut) at the site targeted by the sgRNA. Estimation of the frequency of NHEJ-mediated indels is provided under each founder. Black stars, founders selected for crosses to determine germline targeting rates. (D) CRISPR/Cas9-induced mutations in somatic and germline cells of founders 17 and 18. The sgRNA binding site is underlined on the wild-type sequence. Red dashes and red letters, deletion and insertions, respectively. (E) Genotyping of G<sub>1</sub> butterflies from founders 17 and 18 backcrossed to WT using the Cas9 *in vitro* cleavage assay. Heterozygote butterflies carrying the mutated allele are robustly discriminated by the presence of an additional uncleaved PCR fragment.

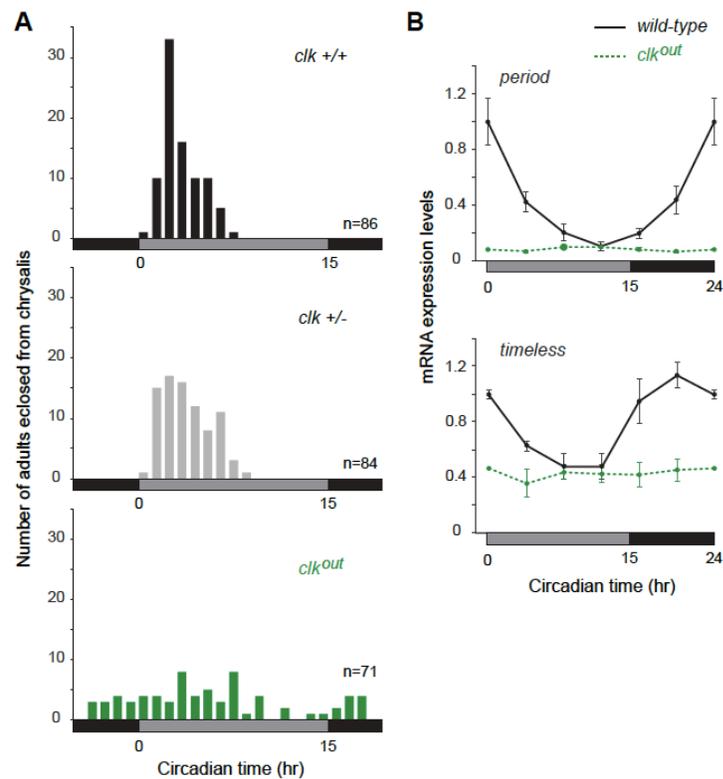
**Figure 4**



**Figure 4. Highly efficient CRISPR/Cas9-mediated targeted mutations and deletions of monarch butterfly *cry2* using dual sgRNAs.** (A) Schematic of part of the monarch *cry2* genomic locus containing the two CRISPR/Cas9 target sites. The orange and green regions in exons 2 and 3 are expanded to provide the sequences targeted for genome editing by each guide RNA (sgRNA). Protospacer adjacent motifs (PAM site, 5'-NGG-3') are highlighted in gray. Arrows on top represent the positions of the primers used to amplify the 1403-bp targeted region for analysis of mutagenic lesions in B (*cry2* F2, *cry2* R2 in Table S1). (B) Detection of mutagenic lesions at the *cry2* loci targeted by sgRNA1 (left) and sgRNA2 (right) in somatic cells of founder G<sub>0</sub> butterflies (9 out of 13 are shown). For each founder, a PCR fragment was subjected to a Cas9-based *in vitro*

cleavage assay with each of the sgRNAs. Black arrowhead, amplicons carrying mutations (mut). Estimation of the frequency of NHEJ-mediated indels is provided under each founder. Red arrowhead, amplicons carrying genomic deletions (del) at the sites targeted by the sgRNAs. Blue and gray arrows, cleaved wild-type PCR fragments. Black stars, founders selected for crosses to determine germline targeting rates. (C) *Left*, CRISPR/Cas9-induced mutations in somatic and germline cells with sgRNA1. Guide RNA binding site is underlined on the wild-type sequence. Red dashes and red letters, deletion and insertions, respectively. *Right*, Dual sgRNA-mediated genomic deletion in somatic cells. Positions of sgRNA1 and sgRNA2 are underlined on the sequence and represented by colored boxes on a schematic representation.

**Figure 5**



**Figure 5. CLOCK deficiency disrupts circadian behavior and the molecular clockwork.** (A) Profiles of adult eclosion in constant darkness (DD) of wild-type (<sup>+/+</sup>, black), heterozygous (<sup>+/-</sup>, gray) and homozygous mutant (<sup>out</sup>, green) siblings of a *clock* (*clk*) mutant line carrying a 4-bp deletion in exon 3, entrained to LD throughout their larval and pupal stages. Eclosion occurred on the first and second days of DD. Data from both days are pooled and binned in 1-hour intervals. Effect of genotype on eclosion time, one-way ANOVA:  $p < 0.0002$ ; Tukey post-hoc test: *clk*<sup>+/+</sup> vs. *clk*<sup>+/-</sup>,  $p > 0.05$ ; *clk*<sup>+/+</sup> vs. *clk*<sup>out</sup>,  $p < 0.01$ ; *clk*<sup>+/-</sup> vs. *clk*<sup>out</sup>,  $p < 0.01$ . Black horizontal bars, subjective night; gray horizontal bars, subjective day. (B) Circadian expression of *period* and *timeless* in brains of wild-type (black lines) and homozygous mutant (dashed green lines) siblings of the *clk* mutant line entrained to LD throughout their larval and pupal stages. Values are mean  $\pm$

SEM of three animals, except for circadian time 0 (CT0), which represents the mean of two animals. Interaction genotype  $\times$  time, two-way ANOVA: *per*,  $p < 0.0001$ ; *tim*,  $p < 0.01$ .

Box shading: gray, subjective day; black, subjective night.

**Table 1**

Table 1. ZFN-, TALEN- and CRISPR/Cas9-induced mutagenesis after mRNA microinjection in the monarch butterfly.						
ZFNs/TALENs/ CRISPRs	Number of injected eggs	Hatching rate, %	Live larvae, <i>n</i> (%)	Somatic mutants, <i>n</i> (%*)	Germline mutation rate, % (Mut crossed, <i>n</i> )	Reference
<i>cry2</i> ZFNs	1194	4.8	45 (3.8)	4 (8.9)	44.5 (2)	Merlin <i>et al.</i> , 2013
<i>cry2</i> TALENs	86	22	17 (19.8)	11 (64.7)	4.1 (4)	This study
<i>cry2</i> CRISPR (2 sgRNAs)	88	21.6	13 (14.8)	8 (61.5)	100 (2)	This study
<i>clock</i> TALENs	90	21.1	17 (18.9)	17 (100)	52.4 (4)	This study
<i>clock</i> CRISPR (1 sgRNA)	93	39.8	19 (20.4)	11 (57.9)	12.2 (2)	This study

\*: % Somatic mutants from live larvae

The percentage of live larvae corresponds to the number of larvae that survived to adulthood relative to the number of injected eggs. The percentage of somatic mutants (%\*) corresponds to the number of live larvae that presented any degree of somatic mosaicism. The germline mutation rate corresponds to the percentage of progeny carrying a mutated allele from the total of progeny tested.