University of South Bohemia

Faculty of Science

CRISPR/ CAS9

Genome editing in *Pyrrhocoris apterus*

Bachelor Thesis

Maly Bertolutti

Supervisor: dr Joanna Kotwica-Rolinska

Department of Molecular Biology and Genetics, Institute of Entomology,

Biology Centre, Czech Academy of Sciences

Co-Supervisor: Mgr. David Doležel, Ph.D.

České Budějovice (2019)

Bertolutti M., 2019: CRISPR/ CAS9 Genome editing in *Pyrrhocoris apterus*. Bachelor Thesis, in English. – 33 pages, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation:

The CRISPR/Cas9 technology is a cutting edge method to genetically modify specific targets in the genome, which allows to create mutants in non-model organisms. The aim of this project was to genetically modify the *Pyrrhocoris apterus* genomic sequence of the *cryptochrome2 gene* using the CRISPR/Cas9 technology. Specifically, this project aimed to create a mutant with removed or modified C-terminal part of the CRY2 protein, which should later allow for studying the role of this sequence in CRY2 function.

Declaration:

I hereby declare that I have worked on my bachelor's thesis independently and used only the sources listed in the bibliography.

I hereby declare that, in accordance with Article 47b of Act No. 111/1998 in the valid wording, I agree with the publication of my bachelor thesis, in full form to be kept in the Faculty of Science archive, in electronic form in publicly accessible part of the STAG database operated by the University of South Bohemia in České Budějovice accessible through its web pages. Further, I agree to the electronic publication of the comments of my supervisor and thesis opponents and the record of the proceedings and results of the thesis defence in accordance with aforementioned Act No. 111/1998. I also agree to the comparison of the text of my thesis with the Theses.cz thesis database operated by the National Registry of University Theses and a plagiarism detection system.

České Budějovice, 11.12.2019 ..……..

Maly Bertolutti

Part of thesis was already published in:
Kotwica-Rolinska J., Chodakova L., Chvalova D., Kristofova L., Fenclova I., Provaznik J., **Bertolutti M**., Wu B.C.H. & Dolezel D. (2019) CRISPR/Cas9 Genome Editing Introduction and Optimization in the Non-model Insect Pyrrhocoris apterus. *Frontiers in Physiology* **10**. https://doi.org/10.3389/fphys.2019.00891
Student's contribution in this publication included: design, production and efficiency testing of guide RNAs targeting C-terminal part of *Pyrrhocoris apterus* CRY2 protein.

Acknowledgements

First of all, I would like to thank and show my extreme gratitude to my supervisor, Joanna Kotwica-Rolinska, Ph.D., for her constant guide and support through each stage of the project. I am thankful and indebted to her for sharing expertise, and sincere and valuable guidance and encouragement extended to me. Thank you for your patience and all the help you gave me both during the experimental work and while writing the thesis.

I would like to pay my special regards to David Doležel, Ph.D., Head of the Laboratory, for creating a pleasant working environment and for co-supervising this project.

I wish to also thank all the other employers of the Laboratory of Molecular Chronobiology for whose assistance I am very grateful and for being always helpful.

Another special thank you goes to prof. RNDr. Libor Grubhofer, CSc., for helping students whenever was needed and for being an extremely caring representative figure of the Biological Chemistry program.

I would like to acknowledge the support my family and friends gave me. Without them I would not have made it through these years. And at last but not least, I would like to thank myself for pushing through all the university workload and finishing this thesis.

Abstract

CRISPR/Cas9 is a recent genome editing technology that permits relatively straightforward modifications of the genome. The Cas9 nuclease and short guide RNA (sgRNA) complex bind to the target sequence and causes a double-strand break, which will be repaired by the cell introducing mutations. The aim of this project was to use this technique in order to modify the circadian clock gene *cryptochrome 2* (*cry2*) in the *Pyrrhocoris apterus.* C-terminal part of the CRY2 protein was targeted, with intention to possibly modify but not destroy its function. Using bioinformatics tools conserved protein domains and predicted phosphorylation sites in the CRY2 protein were found. After predicting possible cut sites recognized by Cas9 protein and designing guide RNAs targeting sequence of *cry2* gene, sgRNAs for the targeted sequence were produced. Later, their on-target efficiency of the genome editing was tested *in vivo,* by embryo injections and screening mosaicism in the egg. After screening the injected eggs, it was confirmed that successful production of mosaics eggs was achieved. The sgRNA with the highest efficiency for mutation creation was identified and used in following studies. Large scale embryo injections with the chosen sgRNA allowed to create several heterozygous *P. apterus* lines with modified C-terminal tail of CRY2 protein. Further molecular and behavioural analysis of the mutant will allow to decipher particular roles of CRY2 protein in the linden bug time measuring system. The results achieved in this project are just a small part of the possibilities which CRISPR/Cas9 technology can offer, which is especially important for exploring biology of non-model organisms.

Table of Contents

1 **Introduction** 1

1.1 CRISPR/Cas 9 1

1.1.1 Occurrence in nature 1

1.1.2 Selection of the target location for specific gene editing 3

1.1.3 Target locations and guide design 4

1.1.4 Cas9 and guide RNA sources 5

1.2 Pyrrhocoris apterus 6

1.3 Circadian clock 7

1.4 Cryptochrome gene 10

2 **Aim** 11

3 **Materials and methods** 12

3.1 Conserved domain check 12

3.2 Protein sequence comparison 12

3.3 Phosphorylation sites 12

3.4 Mapping of *cry2* exons 12

3.5 Sequence verification 13

3.6 CRISPR/Cas9 cut sites prediction by CRISPOR 13

3.7 Single guide RNA production 14

3.8 *Pyrrhocoris apterus* 16

3.9 sgRNA efficiency testing 16

3.10 Heteroduplex Mobility Assay 17

3.11 *cry2* mutant production 17

4 **Results** 18

4.1 Conserved domains 18

4.2 Sequence comparison 19

4.3 Phosphorylation site 19
4.4 Exons and transcript 20
4.5 CRISPOR 21
4.6 Egg screening 23
5 **Discussion** 26
6 **Conclusion** 29
7 **Literature** 30

List of abbreviations

aa- amino acid

BLAST - Basic Local Alignment Search Tool

bp - base pair

CLK – Clock protein

CRISPR - Clustered Regularly Interspaced Short Palindromic Repeats

crRNA - CISPR RNA

cry1 - c*ryptochrome 1* gene

CRY1 - cryptochrome 1 protein

cry2 - *cryptochrome 2* gene

CRY1 cryptochrome 2 protein

CYC – Cycle protein

D - dark

DBS - Double Stranded Break

F1 - first filial generation

FAD - Flavin Adenine Dinucleotide

G0 - generation zero

HDR - Homologour Directed Repair

L - light

LD - long day

mRNA - messenger RNA

NCBI - National Center for Biotechnology Information

NHEJ - Non-homologous End Joining

PAGE– polyacrylamide gel electrophoresis

PAM - Protospacer Adjacent Motif

PCR - polymerase chain reaction

PER – Period protein

RNP - ribonucleoprotein-complex

SD - short day

sgRNA - short guide RNA

TIM – Timeless protein

tracRNA - trans-activating RNA

TTFL - transcription/translation feedback loop

1 Introduction

1.1 CRISPR/Cas 9

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) / Cas9 is a genome editing technology that enables to modify the genome by removing, adding or altering sections of the DNA sequence. It was discovered in a bacterial immune system and since then it has been adapted into a powerful tool for genomic research.

1.1.1 Occurrence in nature

Bacteria and archea have been able to incorporate in their own genome DNA sequences of past invaders, allowing them to recognize future invaders and thus creating an adaptive immune system for prokaryotes (Thurtle-Schmidt and Lo, 2018).

In order to do this, bacteria or archea gain cellular memory of the invader by integrating short sequences of the viral genome into their CRISPR locus as a spacer (Thurtle-Schmidt and Lo, 2018). From these spacers, they are transcribed into long RNA transcripts which are subsequently processed further into short CRISPR RNA (crRNA) (Hwang, et al 2013). crRNA will bind to a trans-activating RNA (tracRNA) which will results in a double stranded RNA complex. This complex will respectively bind onto a Cas9 protein leading it towards eventual invaders by recognition of a DNA sequence which is complementary to the crRNA. If the sequences are complementary, cleavage will occur. Cas9 needs however to bind onto a PAM sequence in the target genome, which must be near the targeted area (Jinek et al., 2012).

This finding has been adapted for making double-stranded breaks in eukaryotic cells and organisms and research is constantly being carried out (Thurtle-Schmidt and Lo, 2018).

In the laboratory, this CRISPR/Cas9 system is used for genome editing and it consists of two components:

1. An enzyme called Cas9 (CRISPR associated protein 9) that can cut two strands of DNA at specific locations and produces double-strand breaks (DSB)
2. A RNA molecule known as guide RNA (gRNA), which consists of a short, around 20 bases, pre-designed RNA sequence crRNA which will bind to the DNA and help Cas9 to cut the right part of the genome (also called “spacer”) and a longer RNA, tracRNA

(trans-activating crRNA), which forms a scaffold between crRNA and the Cas9 enzyme (Figure 1) (Thurtle-Schmidt, D. and Lo, T., 2018, Jinek et al., 2012)

The RNA strand is designed to match and bind a particular sequence in the DNA because the gRNA has bases complementary to those of the target DNA (Figure 1). Cas9 has to locate and bind to a common sequence in the genome called PAM (Protospacer Adjacent Motif) and after it has found the correct sequence, Cas9 can cut the DNA across both strands (Figure 1) (https://www.addgene.org/guides/crispr/; Jinek et al., 2012; Sun et al., 2017; Thurtle-Schmidt and Lo, 2018).

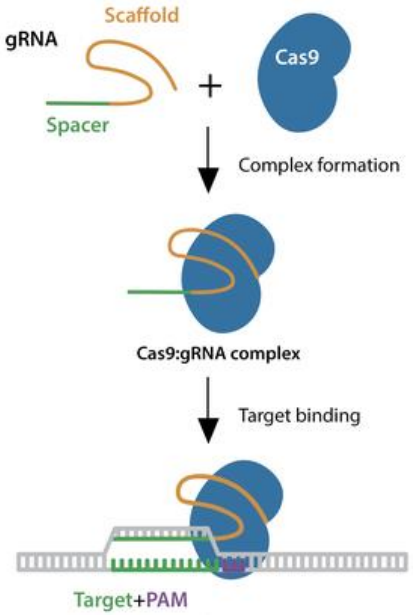

Figure 1. Scheme of the components of the CRISPR/Cas9 system. gRNA and Cas9 protein form ribonucleoprotein complex which bind to the target sequence in the genome. The target in the genome is recognized by the presence of the PAM sequence and the sequence complementary to the spacer. (Figure reproduced from (https://www.addgene.org/guides/crispr/)).

In Eukaryotes, after the DNA sequence is cut, the cell can repair the damage in two different ways: 1) non-homologous end joining (NHEJ) repair or 2) homologous-directed repair (HDR). In the first case, when repairing the double stranded-break, random mutations will be introduced. Since NHEJ is a highly error prone mechanism thus this mechanism of repair will result in creating either deletions or insertions (called “indels”) or substitutions in the genome (Figure 2) (Wyman and Kanaar, 2006).

By introducing an exogenous DNA strand flanked by the DNA sequence which has homology near the area where the double-break was made, the new DNA can be used by the cell to repair the break through HDR, inducing specific desirable mutations or inserts (Figure 2) (Thurtle-Schmidt, D. and Lo, T., 2018).

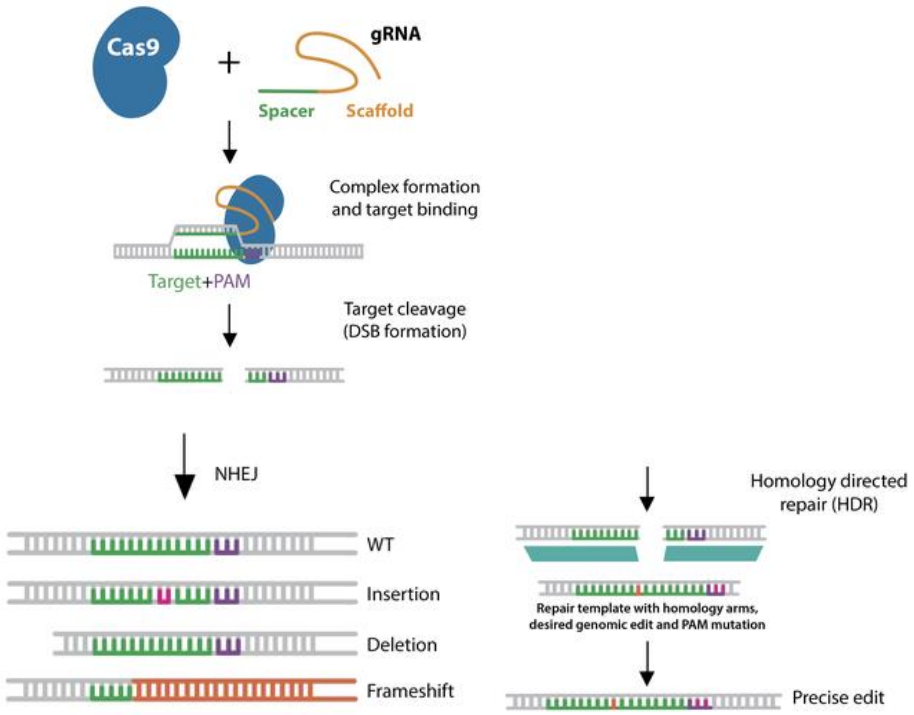

Figure 2. Double stranded brakes caused by CRISPR/Cas9 in the eukaryotic cells can be repaired by NHEJ mechanism introducing "indels" or one can introduce desired mutation by the HDR mechanism. (Figure reproduced from (https://www.addgene.org/guides/crispr/)).

1.1.2 Selection of the target location for specific gene editing

When using this system for genome editing, one of the most crucial things is guide RNA design. Target location depends on the desirable outcome of the planned experiment. By CRISPR/Cas9 method one can produce knock-out mutants in which the gene is made inoperative. Target sequences for guide RNAs are usually then chosen in the 5′ exons, because when targeting this area the mutations are expected to lead to full-gene inactivation. One can specifically target just particular domains of the protein – to establish the role of this domain in the protein function. Similarly, it is possible to attempt to target specific amino acids in order to study the role of post-translational modifications in the protein function. In order to introduce specific protein TAGs (for example polyHis-TAG, or Myc-TAG) helping to study localization or binding partners of studied protein, mostly N-terminal or C-terminal end of protein are chosen (https://www.addgene.org/guides/crispr/).

1.1.3 Target locations and guide design

PAM is a DNA sequence recognized by Cas proteins and is composed of 2 to 6 base pairs, and its length and specific sequence is dependent on CAS protein used for genome editing (Table 1) (http://crispor.tefor.net/; Ding et al., 2016; Jamal et al., 2016). For this reason it is especially important to identify the correct PAM sequences on the DNA strand, respective to the enzyme of choice.

The guide RNA is around 20 nucleotides long and needs to be complementary to the sequence in the genome which is located upstream of the PAM sequence. The specific length of the guide RNA depends also on the type of the Cas protein used in the genome editing. This sequence is important because the protein recognizes the PAM sequence and makes the cut between the 3rd and 4th base pairs upstream of the PAM in the targeted area specified by the guide RNA (Jinek et al., 2012).

The protein that has been used most frequently was SpCas9 which has a PAM sequence of 5'-NGG-3', where N can be either Adenine, Thymine or Guanine and guide RNA is 20 nucleotides long (Ding et al., 2016; Jamal et al., 2016; Thurtle-Schmidt and Lo, 2018).

Table 1. Different types of Cas proteins and their respective PAM sequences. (Table modified from (http://crispor.tefor.net/).

CAS PROTEIN	PAM	Origin
SpCas9	NGG	*Streptococcus pyogenes*
VRQR-SpCas9	NGA	Engineered from SpCas9
VRER-SpCas9	NGCG	Engineered from SpCas9
SthCas9- CRISPR1	N_2AGAA(A/T)	*Streptococcus thermophilus*
SthCas9- CRISPR3	NGGNG	*Streptococcus thermophilus*
NmCas9	N_4GMTT (M= A or C)	*Neisseria meningitidis*
SaCas9	N_2GRRT (R = A or G)	*Staphylococcus aureus*
KKH-SaCas9	N_2NRRT (R = A or G)	Engineered from SaCas9
BlatCas9	N_4CND	*Brevibacillus latesporus*
AsCpf1	TTTN	*Acidaminococcus sp. BV3L6*
LbCpf1	TTTN	*Lachnospiraceae bacterium ND206*

1.1.4 Cas9 and guide RNA sources

After sgRNA has been designed, it needs to be delivered to the cell together with Cas9 nuclease. There are different ways to deliver the Cas9 protein into the organisms: it can be introduced as 1) DNA, 2) mRNA or 3) protein (Thurtle-Schmidt and Lo, 2018).

1. When using the DNA method, an expression vector (plasmid) with the sgRNA and encoding the Cas9 protein need to be introduced into the organism (https://www.addgene.org/guides/crispr/; Dickinson et al., 2013). Advantages of this method are flexibility, cost effectiveness. Since plasmids are stable, overproduction of Cas9 and guide RNA can occur leading to cutting the genomic DNA in unspecific locations (off-targets) (https://www.mirusbio.com/applications/genome-editing-using-crispr-cas/overview).
2. Cas9 can be also delivered in a form of mRNA, which then will be translated into protein in the target cell. To have a successful expression of Cas9, it is necessary to synthesize mRNA *in vitro* with a 5' cap and a 3' Poly-A tail (https://www.addgene.org/guides/crispr/; Thurtle-Schmidt and Lo, 2018). Since mRNA is less stable, this method reduces risks of off-targeting and also of insertional mutagenesis with unwanted plasmid DNA (https://www.mirusbio.com/applications/genome-editing-using-crispr-cas/overview).
3. A third possibility is the introduction in the cell of a complex consisting of purified Cas9 protein and gRNA which will form a ribonucleoprotein-complex (RNP) (Lin et al., 2014) This method could be the most effective, however purifying Cas9 and ordering RNA is less cost efficient (Kistler et al., 2015; Kotwica-Rolinska et al., 2019).

As mentioned before, guide RNA consists of two parts crRNA and tracrRNA which needs to be delivered into the cell. They can be delivered as separate RNA molecules, or they can be produced in a form of single guided RNA (sgRNA) where crRNA sequence is fused with tracRNA sequence forming a single RNA molecule (Figure 3).

Both forms of described guide RNAs can be delivered to the cell as DNA template on an expression plasmid, or in synthetized final working form as RNA molecules (Hwang et al., 2013; Kistler et al., 2015).

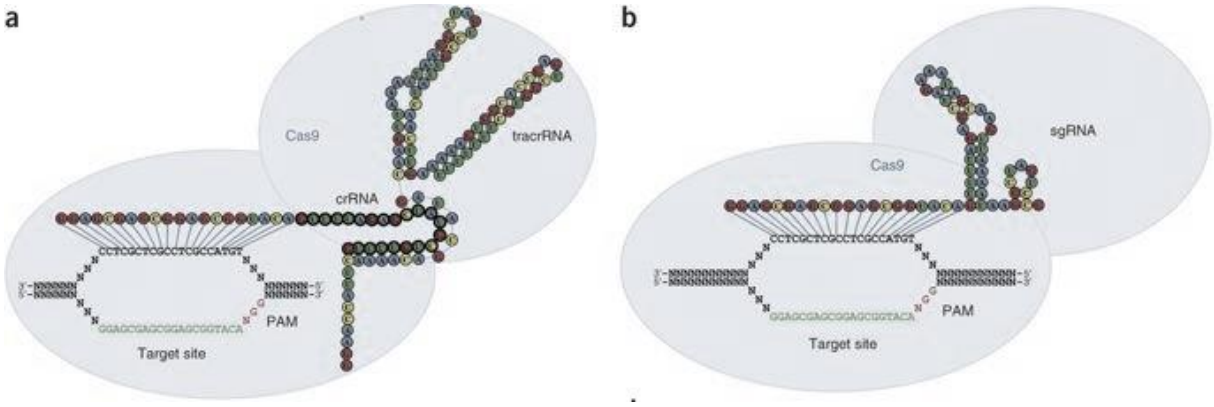

Figure 3. Different types of guide RNA used in genome modification studies. **a)** guide RNA molecule consisting of two separate RNAs (crRNA and tracrRNA), where crRNA finds the complementary target DNA strand located near a PAM sequence and tracrRNA then forms a complex tracrRNA:crRNA. On this scaffold Cas9 is bound and cleaves target genomic sequence near the PAM **b)** single guide RNA (sgRNA) works similarly, however guide RNA is delivered as one fused RNA molecule consisting of crRNA and tracrRNA (Figure reproduced from (Hwang et al, 2013)).

1.2 Pyrrhocoris apterus

The Laboratory of Molecular Chronobiology, where experiments for the thesis were performed, focuses on the *Pyrrhocoris apterus,* also known as linden bug, an insect from the order Heteroptera, family Pyrrhocoridae.

This non-model insect is distributed throughout all Europe, from the Atlantic coast to north-west China. It is one of the few species who lives even in temperate zones from the Palaearctic Region until the south of western Siberia (Socha, 1993). Its availability and easy laboratory breeding make it a useful species to work with. A big collection of field lines is available in our laboratory originated from different locations in Europe. It was shown that depending on the geographical area, these animals show different functioning of the circadian clock, which was illustrated by different speed of the internal clock in constant conditions (Pivarciova et al., 2016). The Roana line was used for this study.

In order to survive harsh environmental conditions e.g. winter or low temperatures, insects often enter diapause which is a period of metabolic and developmental arrest (Dolezel, 2015). Diapause begins before the onset of adverse weather due to changes in day- and night-length, perceived by photoperiodic timers (Dolezel, 2015). *P. apterus* overwinters in imago

stage and diapause is reflected by the lack of reproduction. Ovaries of *P. apterus* females are inhibited during diapause (Socha, 1993; Smykal et al., 2014) while male accessory glands decrease in size and their function repressed thus arresting reproduction (Urbanova et al., 2016). *P. apterus* enters diapause during short-day (SD) conditions (12L : 12D) and starts reproduction again with long-day conditions (LD) (18L : 6D) (Socha, 1993).

The main research topic in the laboratory focuses on the better understanding of the time measuring mechanisms mainly circadian and seasonal clocks using *P. apterus* as a model organism.

1.3 Circadian clock

Any animal exhibit circadian rhythms in physiology and behaviour which are controlled by a circadian clock. The majority of circadian rhythms have a period of circa 24 hours and are thought to originate from two transcription/translation feedback loops (TTFL) (Figure 4) (Bell-Pedersen et al., 2005; Tataroglu and Emery, 2015; Pritchett and Reddy, 2017).

This means that circadian rhythms help the animal having a temporal equilibrium with the surrounding environment. It is achieved by regulating expression of genes in a span of 24 hours and allowing organisms to improve their metabolic and physiological behaviour in response to the time of a day. Some examples of critical circadian rhythm outputs in insects include the time of day of locomotor activity, egg hatching, and the time of day of adult eclosion (Bell-Pedersen et al., 2005; Tataroglu and Emery, 2015).

The central working mechanism of the circadian regulation is the multiple interactions between the clock proteins. Based on *Drosophila* studies it was shown that core clock protein building TTFL is composed of four main clock proteins Clock (CLK), Cycle (CYC), Period (PER) and Timeless (TIM). The positive elements of the feedback loop - transcription factors: CLK and CYC bind together during the day and form CLK/CYC heterodimer which activates expression of negative elements of the feedback loop - *period* (*per*) and *timeless* (*tim)*. During afternoon and early night PER and TIM are translated, form PER/TIM heterodimer which is then, with some delay caused by additional posttranslational modifications, translocated to the nucleus. In the nucleus PER/TIM inhibits transcription of their own genes by blocking the transcriptional activity of CLK/CYC. During the day activation of the photosensitive cryptochrome 1 protein (CRY1) leads indirectly to TIM and following PER degradation, which in turn unblocks CLK/CYC heterodimer and starts the cycle again (Bell-Pedersen et al., 2005; Hardin, 2005; Peschel and Helfrich-Forster, 2011; Tataroglu and Emery, 2015). The

stability and activity of all clock proteins is regulated by daily regulated subsequent phosphorylation and specific kinases and phosphatases were shown to affect greatly the function of the circadian clock in *Drosophila*. In case of PER and TIM hyperphosporylated proteins are faster degraded, and actually CRY dependent degradation of TIM is also mediated by increased phosphorylation of TIM protein (Bell-Pedersen et al., 2005; Hardin, 2005; Peschel and Helfrich-Forster, 2011). This shows how the interactions between the clock proteins are extremely essential for the regulation of the circadian system. Therefore the length and stability of the circadian clock depends on the rate at which these proteins degrade or accumulate in the cell nucleus, where faster degradation means faster running clock.

Similar to *Drosophila*, four essential clock proteins in the nucleus guide the fundamental TTFL in mammals: there are two transcriptional activators, CLOCK and BMAL1 (homolog of Cycle), and two groups of repressors, PER (PER1, PER2 and PER3) and CRY (CRY1 and CRY2) (Partch, C. et al, 2014). As the name implies, CLOCK and BMAL1 activate the transcription of the two repressor proteins PER and CRY which will leave the nucleus towards the cytoplasm. These accumulate in the cytoplasm where they eventually heterodimerize and are transferred back to the nucleus restraining and suppressing their own transcription by inhibiting the activity of CLOCK and BMAL1 heterodimers (Pritchett, D and Reddy, A. 2017). The cycle can start again with a periodicity of around 24 hours after PER and CRY proteins degrade, stopping the inhibition of CLOCK:BMAL1. In contrast to *Drosophila*, mammalian CRY proteins are not photosensitive and do not play a role in synchronization to light cycles. Instead it seems that CRY proteins took over the role of *Drosophila* TIM as a negative element of the circadian clock (Figure 4) (Bell-Pedersen et al., 2005; Partch et al., 2014; Pritchett and Reddy, 2017).

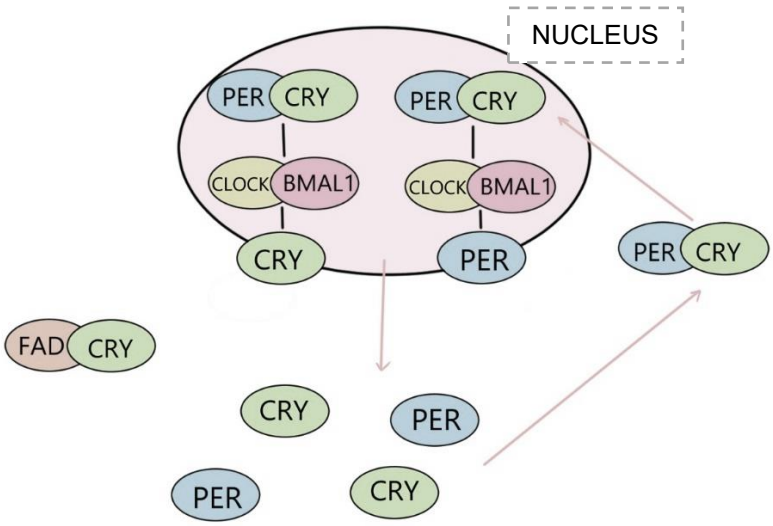

Figure 4. Simplified scheme of the mammalian circadian clock. The regulation of stability of the PER and CRY proteins has been found to be important for determining the periodicity of the circadian oscillation. Description in the text. (Figure modified from (Pritchett, D and Reddy, A. 2017)).

In insects other than *Drosophila* the genetic toolkit of circadian clock varies. In some species the same components as in *Drosophila* are found, in others mammalian CRY is essential and there are also species where circadian clock constitution resembles both, *Drosophila* and mammalian-based mechanisms (Yuan et al., 2007; Tomioka and Matsumoto, 2010; Ingram et al., 2012; Tomioka and Matsumoto, 2015). For example, in butterflies there are two types of CRY protein present - light sensitive *Drosophila* type 1, and light insensitive mammalian type CRY2. Interestingly they also possess functional TIM protein. In honeybees - only mammalian type CRY2 is present and TIM is absent in the genome showing the highest similarity to the mammalian circadian clock machinery (Yuan et al., 2007; Tomioka and Matsumoto, 2015). *P. apterus* possess only light insensitive mammalian type of CRY2 protein (Bajgar et al., 2013), but TIM is also present, however its function in the machinery of the circadian clock is still unclear (Kotwica-Rolinska at al., unpublished data).

Despite above described similarity of *P. apterus* clock with mammals, the protein sequences are not completely identical. Thus, it is also still not completely clear how the molecular basis of the circadian clock works and how exactly the essential proteins operate and their mechanism of action.

1.4 Cryptochrome gene

Cryptochromes are flavoproteins that act as blue light receptors and can be found in both animals and plants. They were originally discovered in *Arabidopsis thaliana* but further studies identified their presence in other organisms too (Yu et al., 2010).

Cryptochromes have two conserved domains: the Photolyase-Homologous Region (PHR) which can be found at the N-Terminal and the Flavin Adenine Dinucleotide (FAD) found at the C-terminus that appears to be extremely crucial for cryptochromes function and regulation. The cryptochrome family shows similarities to the photolyase family and is structurally similar to the FAD proteins (Figure 5). These proteins are activated by light and need to absorb blue light in order to accomplish their functions. Photolyases however use blue light in order to repair UV-induced DNA damage while cryptochromes are involved in circadian clock of animals and plants. Studies hypothesize that cryptochromes get excited by electrons of the flavin molecule and undergo phosphorylation to adopt an open conformation which will signal partner proteins to alter gene expression (Hoang et al., 2008; Yu et al., 2010).

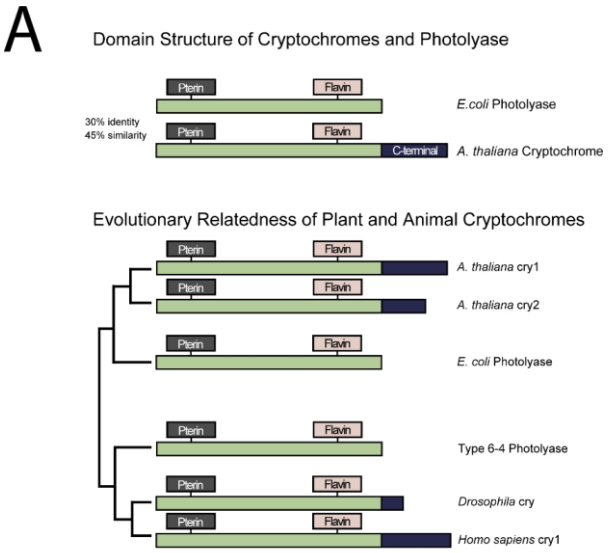

Figure 5. Comparison of protein domains in prokaryotic photolyase and eukaryotic cryptochrome protein (top picture). Comparison of protein domains in animal and plants cryptochromes (bottom picture). (Figure reproduced from (Hoang, N. et al 2008)).

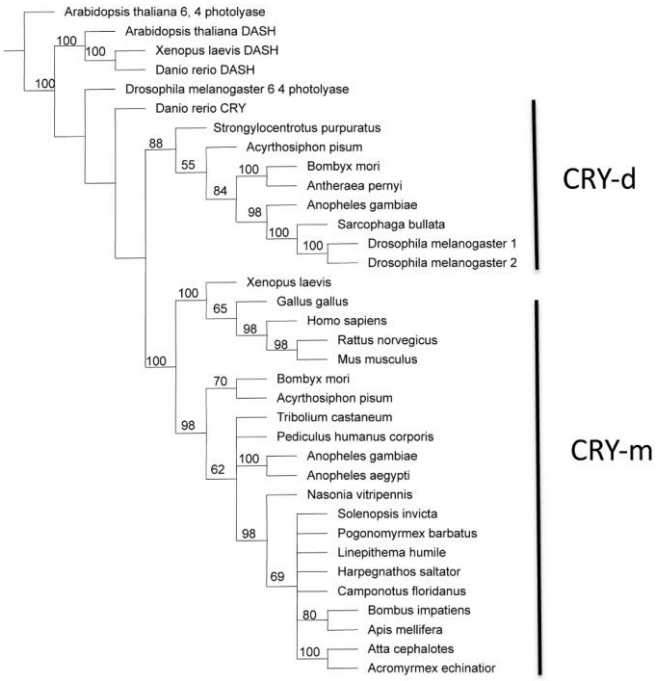

Figure 6. Parsimony tree of the cryptochrome gene family. Cryptochromes fall into two categories: Drosophila-like CRY (CRY-d) and mammalian-like CRY (CRY-m) (Figure reproduced from (Ingram, K. et al, 2012)).

Two types of *cry* gene families were found. They encode CRY1 (or CRY-d) type protein found in *Drosophila* or a vertebrate-like CRY2 (or CRY-m) protein (Figure 6.) (Yuan et al., 2007; Ingram et al., 2012). Previous investigations have shown that insect CRY1 is photosensitive, while CRY2 is photo-insensitive and acts as transcription repressor. Animals can either have only CRY1 (as in *Drosophila*), both CRY1 and CRY2 (as *Danaus plexippus* and *Culex pipiens*) or only CRY2 (as in *Homo sapiens* and *Mus musculus*) (Yuan et al., 2007; Ingram et al., 2012; Tomioka and Matsumoto, 2015).

2 Aim

The aim of this project was to genetically modify the *Pyrrhocoris apterus* genomic sequence of the *cryptochrome2 gene* using the CRISPR/Cas9 technology. Since *P. apterus cry2* knock-out line is already available in the laboratory (Kotwica-Rolinska et al., 2019), this project aimed to create a mutant with removed or modified C-terminal part of the CRY2 protein, which should later allow for studying the role of this sequence in CRY2 function.

3 Materials and methods

3.1 Conserved domain check

Conserved domains in the *P. apterus* CRY2 protein sequence (GenBank: AGI17567.1) were analysed by using *NCBI Conserved Domain* Search tool against cdd V3.database and optional parameters were set to default.

3.2 Protein sequence comparison

Using Geneious software (Geneious, Bimetters, Ltd.) *P. apterus* CRY2 sequence was aligned with cryptochrome protein sequences from other insects and vertebrates, including human, by the ClustalW Alignment with default settings. The list of compared organisms is as follows: *P. apterus* (AGI17567.1), cricket *Gryllus bimaculatus* (BAX56244.1), planthopper *Nilaparvata lugens* (AJY53622.1), mosquito *Culex pipiens* (AIW65407.1), honey bee *Apis mellifera* (NP_001077099.1), beetle *Triboleum castaneum* (EFA04537.1), butterfly *Danaus plexippus* (ABA62409.1), mouse *Mus musculus* (AAD46561.1), rat *Rattus norvegicus* (NP_596896.2) and human *Homo sapiens* (NP_001120929.1).

3.3 Phosphorylation sites

NetPhos 3.1 Server tool (http://www.cbs.dtu.dk/services/NetPhos/) specifically designed to predict phosphorylation sites in eukaryotic proteins was used in order to find where the phosphorylation sites in *P. apterus* CRY2 protein were located. Default settings were used allowing for prediction of phosphorylation of serine, threonine and tyrosine (Ingrell et al., 2007).

3.4 Mapping of *cry2* exons

In order to map the location of *cry2* exons in the transcript (and find the exon, where predicted phosphorylation sites were encoded), full length *cry2* cds (coding sequence) (GenBank ID: JX560423.1) was compared by the Basic Local Alignment Search Tool (BLAST) to the draft of the genome of *P. apterus,* which is available in the laboratory. BLAST was performed in Geneious 9.0 (Geneious, Biomatters Ltd., New Zealand) software with default setting.

3.5 Sequence verification

In order to verify the sequence of the target region:

1. Genomic DNA was isolated

One antenna of the bug was cut out and placed in the Eppendorf tube. Next DNA out if antenna was isolated by squishing in the 50 µl of squishing buffer (10 mM Tris-HCl pH=8.0, 1 mM EDTA, 25 mM NaCl, 200 µg/ml Proteinase K). Homogenate was incubated for 1 h at 37°C, and then Proteinase K was inactivated by incubation for 3 min at 95°C.

2. Fragment of genomic DNA was amplified by PCR.

For the PCR reaction 1ul of homogenate from the previous step, 12.5 µl of 2x TP 2x Master Mix (TopBio, Czech Republic), 0.5 µl of each 10 µM specific primers and 10.5 µl of nuclease-free water were added. PCR reaction was run with the following parameters: 3 min at 94°C, (94°C – 30 s, 58°C – 30 s, and 72°C – 30 s) for 30 cycles, followed by the final elongation at 72°C for 10 min.

Sequences of specific primers are as follows:

cry2 forward primer: 5’ - GCACTGATTATCCATTGCCTATGT - 3’

cry2 reverse primer: 5’ – GGACGTTTACACTGTTCGTGAG - 3’

3. PCR product was purified by the QIAquick PCR Purification Kit (Qiagen) according to instructions provided by manufacturer and sent for direct sequencing (Eurofins, Germany)

3.6 CRISPR/Cas9 cut sites prediction by CRISPOR

To find the cut sites in the targeted fragment of the *cry2* gene, we used the online CRISPOR tool (http://crispor.tefor.net/), which automatically predicts the cut sites, on-target efficiency and if the genome is known it also predicts off-targets (Maximilian Haeussler, 2019). While using CRISPOR software, in Step 1 the sequence encompassing the target region (-/+100 bp) was inserted. In the Step 2 “No Genome” option was selected, and in the Step 3 20bp-NGG PAM sequence (recognized by Cas9 from *Streptococcus pyogenes*) was chosen.

Based on the prediction 7 possible guides were found in the targeted region and 3 guide sequences were chosen for *cry2* gene modification.

3.7 Single guide RNA production

Single guide RNA (sgRNA) is composed of sequence specific crRNA fused to the common sequence of the tracrRNA. In order to produce sgRNAs previously published protocol (Kistler, 2015) was followed. The production of sgRNA consists of several steps

1. Non template PCR

For the production of the template for sgRNAs, so called non template PCR was used. The template is formed by amplification of the product of binding together two primers which have partially complementary sequence. Test tubes were prepared adding KAPA HiFi Fidelity Buffer (5X) (20 µl), KAPA dNTP Mix (2,5 mM each) (3 µl), KAPA HiFi DNA Polymerase (2 µl), Nuclease free water (65 µl) and 5 µl of each (10µM) CRISPR forward and (10µM) CRISPR reverse primer.

2. Primers

CRISPR Reverse primer is universal for all gRNAs and contains part of the **tracrRNA.** In ***italics*** *sequence complementary to part of the forward primer is marked*:
5'-**AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCT TATTTTAACTT*GCTATTTCTAGCTCTAAAAC*** – 3'

CRISPR Forward primer is a specific primer for a different templates production, because it contains sequence of the target sgRNA. It contains (starting from the 5' sequence: T7 promoter sequence, and **two guanines (GG)**, which are needed for *in vitro* transcription, 20nt long region specific for each gRNA (N20) followed by *sequence complementary to part of the reverse primer* which allows for binding to the common reverse primer
5'-GAAATTAATACGACTCACTATA **GG** (N20) *GTTTTAGAGCTAGAAATAGC* – 3'GCTAGAAATAGC 30

Sequences of forwards specific primers used for production 3 templates for sgRNA production are as follows:

cry2 188 forward 5' – GAAATTAATACGACTCACTATA**GG**(ACTTACAGTTAATGAAATTC)*GTTTTAGAGCTA GAAATAGC* - 3'

cry2 194 forward 5’ -
GAAATTAATACGACTCACTATA**GG**(ATGGTATAGTAAAAATTACC)*GTTTTAGAGCTA GAAATAGC* - 3’

cry2 196 forward 5’ –
GAAATTAATACGACTCACTATA**GG**(TTAATGAAATTCAGGCAGCC)*GTTTTAGAGCTA GAAATAGC* - 3’

PCR was run with parameters described for KAPA Hifi enzyme in provider’s instruction (KAPA Biosystems): 3 min at 94°C, (98°C – 15 s, 58°C – 30 s, 72°C – 15 s) for 35 cycles, elongation at 72°C for 5 min.

3. Gel purification

Samples were run on 2% agarose gel in 1x TAE (Tris Acetic acid-EDTA) buffer. After electrophoresis, piece of gel with sample was cut out and put into Eppendorf tube. DNA was the isolated from the gel with QIAquick Gel Extraction Kit (Qiagen), following manufacturer’s instruction. Concentration of the purified DNA was measured with the NanoDrop 2000 spectrophotometer (ThermoFisher Scientific).

4. In vitro transcription

RNA was produced from prepared templates by *in vitro* transcription. In vitro transcription was performed with MegaScript T7 Kit (Ambion). For the reaction 300 ng of the purified PCR template from the previous step, 10x reaction buffer (2 µl), rATP, rUTP, rGTP, rCTP mix (8 µl), T7 enzyme mix (2 µL) and nuclease free water to 20 µl were mixed together. Reaction was incubated for 2,5 h at 37°C. To destroy the DNA template 1 µL of Turbo DNAse was added and reaction was incubated again for 15 min at 37°C.

5. sgRNA purification.

sgRNAs were then purified by 3M sodium acetate and ethanol precipitation. 2 µl of 3M sodium acetate pH=5,5 and 60 µl of 100% ethanol were added. Precipitation was carried out at -20°C overnight. Afterwards samples were centrifuged 14 000 rpm for 30 min at 4°C. Pellet was washed with 75% ethanol, air dried and resuspenden in nuclease free water. sgRNAs were run on 1% agarose gel in TAE buffer, in order to check the quality.

sgRNA concentration was measured by the NanoDrop 2000 spectrophotometer (ThermoFisher Scientific). sgRNAs were diluted in nuclease free water to the concentration of 1mg/ml and stored at -80°C until further use.

3.8 *Pyrrhocoris apterus*

In this study laboratory established "Roana" line was used and the name indicates the original location of the insect line (Pivarciova et al., 2016). *P. apterus* bugs were reared at 25°C and were maintained from hatching under long day conditions (18 h light, 6 h dark) in jars supplemented with linden sees and water. A night before eggs collection, bugs were transferred to new jars and newly laid eggs were collected for injections after the period of 12h.

3.9 sgRNA efficiency testing

For the injections 5 µl of injection mixture was prepared. It was composed of 500 ng/ul of Cas9 protein (PNA Bio) diluted in nuclease free water and 200 ng/µl of sgRNA. Before injection Cas9 protein was mixed with the sgRNA and incubated on the bench for 30 min to obtain ribonucleoprotein complexes.

Eggs were prepared and injected by the supervisor.

After injections eggs were transferred to the incubator with light: dark regime 18h:6h and temperature 25 °C. On the next day, around 30 h post-injection, eggs were collected separately to the PCR tubes and DNA was isolated in squishing buffer, identically as described above. Eggs of the same age, which were not injected were used as wild type controls.

Homogenates congaing DNA from injected and control eggs served as templates for PCR, where the region targeted by all sgRNAs was located. 10 µl PCR reaction was composed of: 5 µl of 2x TP 2x Master Mix (TopBio, Czech Republic), 0,25 µl of each 10 µM specific primers and 3,5 µl of nuclease-free water. PCR was run with following parameters: 3 min at 94°C, (94°C – 30 s, 58°C – 30 s, and 72°C – 30 s) for 35 cycles, and final elongation at 72°C for 10 min. Specific primers were identical as described above for sequence verification.

3.10 Heteroduplex Mobility Assay

5 µl of the PCR reaction and 2,5 ml of the 1kb+ DNA ladder (ThermoFisher Scientific) were run on the 15% non-denaturing acrylamide gels. 10 ml of the acrylamide gel was composed of 5ml 19:1 Acrylamide/Bis-acrylamide solutions, 5 ml of 1xTBE buffer (45 mM Tris-borate and 1mM EDTA) and 160 µl 10% APS (*ammonium persulfate*) and 8µl TEMED.

Gels were run in the Bio-Rad Mini-PROTEAN Tetra Cell (Bio-Rad). Gels were run in 0.5xTBE buffer for 2,5h at 100V. After the run gels were stained for 10 min in the 1:10 000 GelRed nucleic acid stain solution in water (Biotium, USA). Afterwads gel were photographed by the by Gel Documentation System Smart3-EZ (VWR, Belgium).

3.11 *cry2* mutant production

221 eggs were injected with the injection mixture of 500 ng/µl of Cas9 protein (PNA Bio) diluted in nuclease free water and 200 ng/µl of sgRNA (cry2 Exon 10 194 R) which showed the highest efficiency of inducing mutations during sgRNA efficiency testing. Eggs were prepared and injected by the supervisor.

After injections eggs were left in the incubator with light: dark regime 18h:6h and temperature 25 °C until hatching (G0 generation) and then reaching adulthood. The surviving fertile adult G0 bugs were crossed to wild type partners from Roana line and produced F1 generation.

When the F1 generation reached adulthood, they were screened for occurrence of mutations in the targeted region. Because the development of *P. apterus* is relatively long, and the time to reach adult F1 generation exceeded the duration of the thesis, following steps in mutants' characterization were continued by researchers from Laboratory of Molecular Chronobiology. Shortly, genomic DNA was isolated from one antenna as described above, and the 1 µl of the homogenate served as template for the PCR. PCR composition and PCR parameters were identical as PCR performed on eggs squish during sgRNA efficiency test. Next, 2,5 µl of the PCR reaction was tested by Heteroduplex Mobility Assay with identical parameters as described above. For surviving mutant lines - fragment of *cry2* carrying mutation was PCR amplified and sent for sequencing. (Kotwica-Rolinska 2019).

4 Results

4.1 Conserved domains

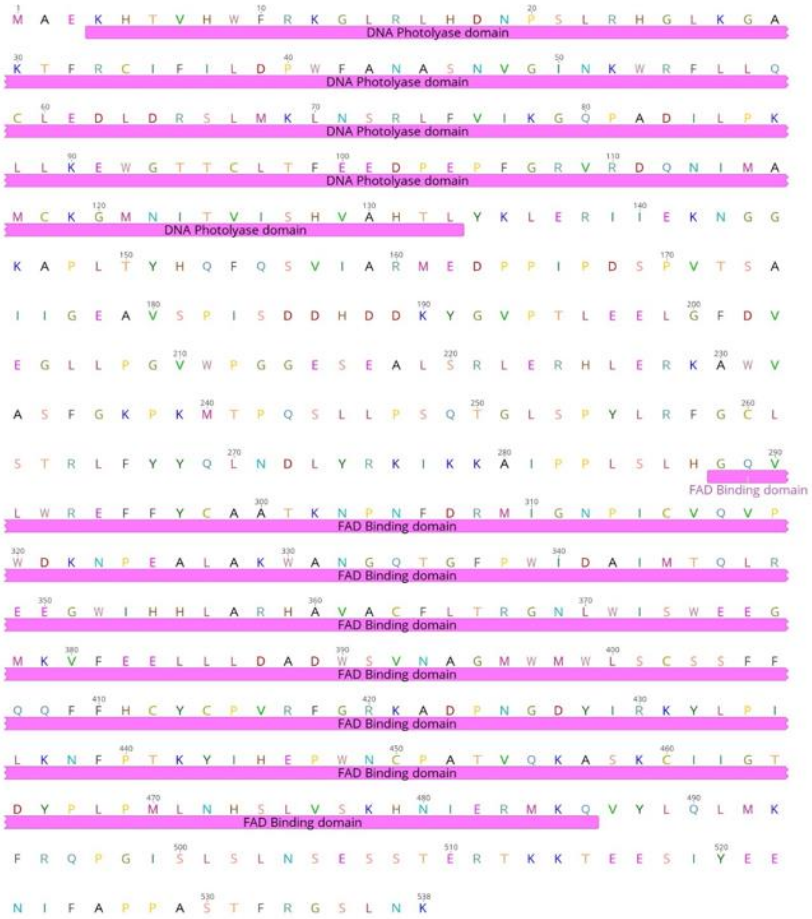

Figure 7. Sequence of the P. apterus CRY2 protein (GenBank: AGI17567.1) with two conserved domains found (pink).

P. apterus CRY2 protein is 538 amino acids (aa) long. The search with NCBI Conserved Domain showed two conserved domains in the protein sequence (Figure 7): *DNA Photolyase domain* (4aa-133aa) and the *FAD Binding domain* (288aa-486aa) located near the C-terminal part of the protein.

DNA photolyase domain is a conserved protein domain where photolyase and cryptochrome proteins are found. These proteins can absorb blue light in order to conduct different tasks. DNA photolyases are enzymes that can repair DNA damaged by UV light (Hoang et al., 2008; Yu et al., 2010). The second domain binds FAD coenzyme. FAD stands for Flavin Adenine Dinucleotide and they are proteins or coenzymes that are associated with a various array of proteins and are involved in several enzymatic reactions (Hoang et al., 2008; Yu et al., 2010)

4.2 Sequence comparison

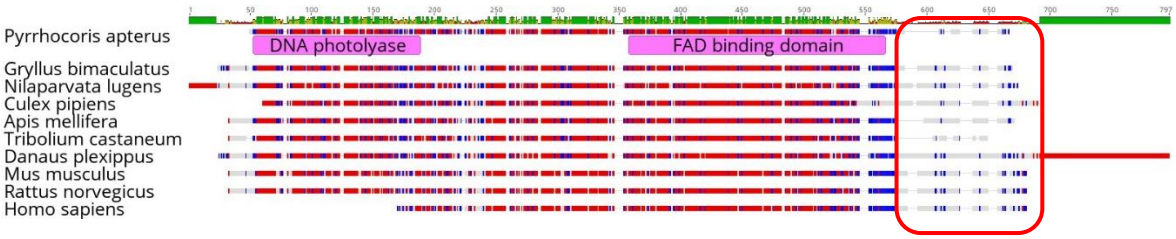

Figure 8. ClustalW alignment of CRY2 proteins originating from different organisms. Red, blue and grey boxes in sequences show presence of identical, similar (60-80% similarity) and different amino acids, respectively, in all compared species. The location of 2 conserved domains are marked with pink bars. The red rectangle marks the highly divergent C-terminal part of the protein.

By using ClustalW alignment CRY2 proteins from different insect species and three mammalian representatives were compared. It can be observed that the CRY protein is mostly highly conserved in all organisms, differing however in the end part of the sequence (highlighted by the red rectangle). The identical or very similar amino acid composition is found thorough almost whole CRY2 protein, excluding C-terminal part, which starts just downstream the FAD binding domain. In *P. apterus* this region consists of the last 43 aa of the CRY2 protein sequence.

4.3 Phosphorylation site

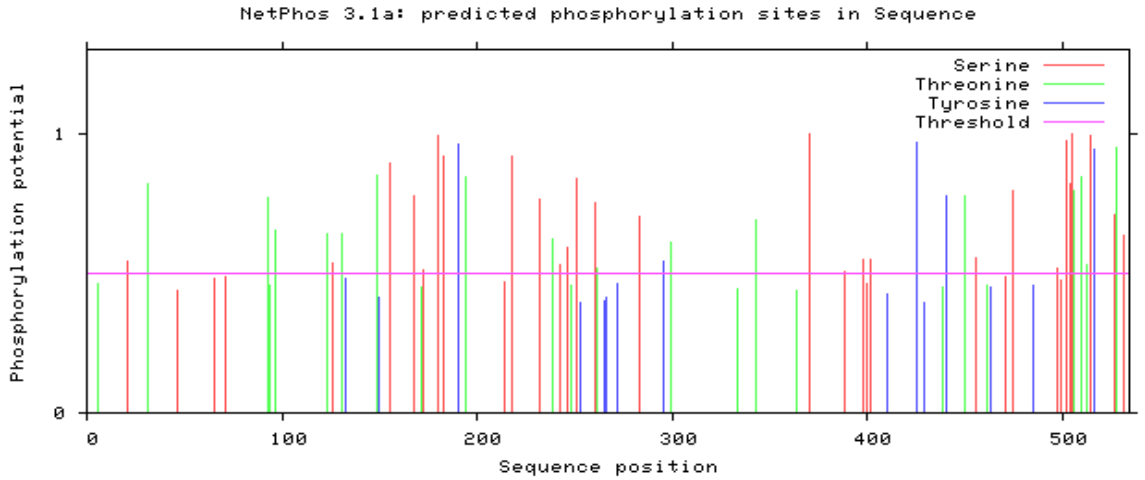

Figure 9. Graph of the predicted phosphorylation sites. Majority of phosphorylations are found in the middle and in the C-terminal end of the sequence. Each colour represents an amino acid which is predicted to be phosphorylated. The pink horizontal line is the threshold that specifies the reliability of the signals.

The C-terminal part of the CRY2 protein is not evolutionary conserved (Figure 8). Therefore it is hard to predict a role of this region in the CRY2 protein function. By using the Netphos3.1 tool it was found that in this region composed of 43 aa twelve predicted phosphorylation sites are found (500S, 505S, 507S, 508S, 509T, 512T, 515T, 518S, 520Y, 530S, 531T and 535S). The enrichment in the putative phosphorylation sites could imply, that this region could be important for CRY2 function in the circadian clock, where phosphorylation is known to affect the stability of circadian clock proteins, their activity as a transcription factors, or their cellular localization.

In order to analyse the function of the C-terminal part of the CRY2 protein a mutant with this region removed or changed would be ideal model.

4.4 Exons and transcript

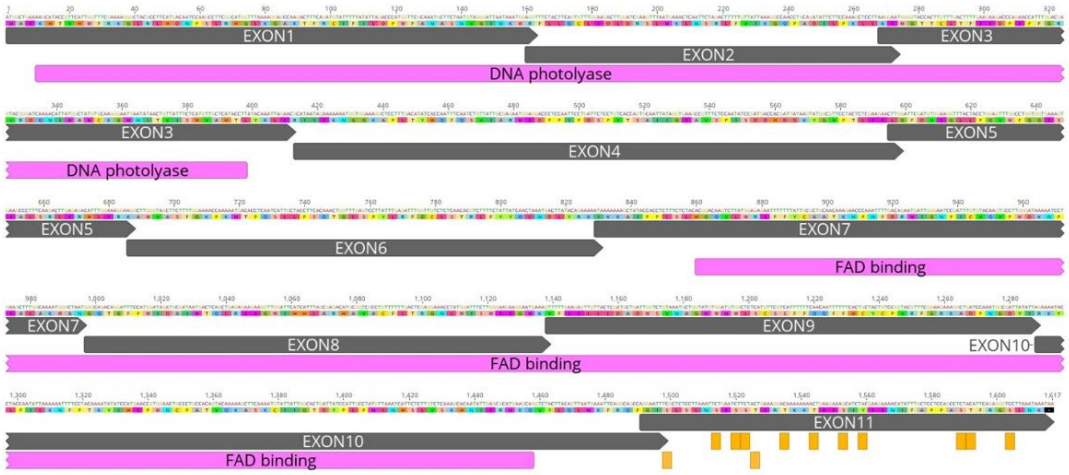

Figure 10. Coding sequence of the *P. apterus cry2* gene (1617 nucleotides). Below nucleotide sequence, the translated protein sequence is marked. Protein domains locations are marked in pink, exons are marked in grey and phosphorylation sites of the C - terminal part of the protein are marked in orange.

The coding sequence of CRY2 protein is 1617 nucleotides long including stop codon (TAA). The sequence of the transcript of the *cry2* gene was blasted against the genomic sequence of the *P. apterus* (available in the laboratory, data not published) and 11 exons were discovered (grey). It was found that all the predicted phosphorylation sites of the C-terminal part of the protein were located on the last exon and they are marked in orange (Figure 10).

BLAST search through the genome with *cry2* exons as a query allowed for finding also partial intronic sequences encompassing them (one example is shown in Figure 11).

4.5 CRISPOR

Table 2. Predicted CRISPOR cut sites and on-target efficiency.

CUT SITES	TARGET SEQUENCE	Guide and tracer from T7 promoter
Exon 10 188 F	ACTTACAGTTAATGAAATTCAGG	38
Exon 10 194 R	ATGGTATAGTAAAAATTACCTGG	25
Exon 10 196 F	TTAATGAAATTCAGGCAGCCAGG	30
Exon 11 202 R	AGTAGAAGATTCAGAATTTAAGG	37
Exon 11 228 F	ATTCTGAATCTTCTACTGAAAGG	35
Exon 11 276 R	CCTCTGAATGTAGAGGCTGGAGG	58
Exon 11 297 F	CTCCAGCCTCTACATTCAGAGGG	31

Through CRISPOR software it is possible to find guide sequences of an input DNA sequence. The website automatically predicts cut sites, on-target efficiency and if the genome is known it also predicts off-targets. For the mutant generation SpCas9 recognizing 5' –NGG -3' was chosen (Haeussler et al., 2016; Concordet and Haeussler, 2018). CRISPOR finds all possible PAM sequences on the leading and complementary strain and designs 20nt long guide RNA sequence upstream of the PAM sequence. Using CRISPOR software 3 targets for Cas9 in the exon 10 and 4 targets for Cas9 in the exon 11 were found (Table 2). The name of the particular gRNA shows in which exon the target is located and relative position of the target in the input sequence (around 400 nt of the genomic sequence was given as template) and F and R means location of the PAM sequence on the leading or complementary strain, respectively. The numbers in the column on the right show the relative predicted on-target efficiency score, where 100 is set by the program as the best score (Haeussler et al., 2016; Concordet and Haeussler, 2018).

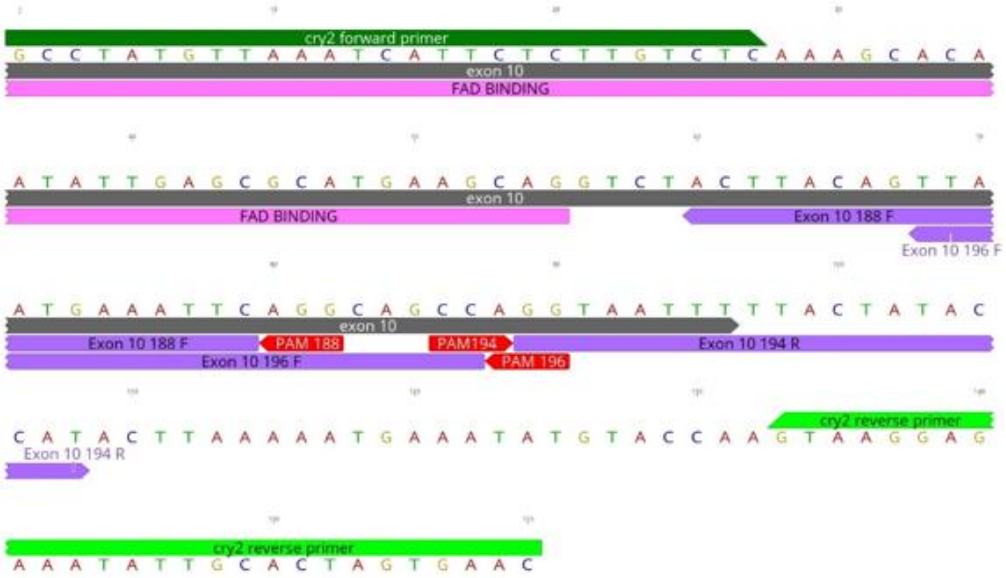

Fig. 11. Fragment of the genomic sequence with the fragment of the exon 10 of the *P. apterus cry2* gene (grey box) and following intronic region. PAM (red), sgRNA targets (violet), screening primers (green) and part of FAD binding domain (pink) are marked.

Figure 11 represents the genomic sequence of the fragment of exon 10 marked with the grey box followed by the partial sequence of the intron. Fragment of conserved FAD binding domain located on the exon 10 is marked in pink. Position of 3 PAM and guides RNA sequences (target sequences are listed in Table 2) is shown in red and violet, respectively. The PAM sequence 5' – TGG – 3' of the guide "Exon 10 194 R" is located on complementary strand. Predicted double stranded breaks made by the Forward primer is indicated by the dark green box while the reverse primer by the lighter green one. According to the literature, cut sites made by SpCas9 are predicted to be located between 3rd and 4th nucleotide upstream PAM sequence (https://www.addgene.org/guides/crispr/; Jinek et al., 2012; Thurtle-Schmidt and Lo, 2018).

Only 3 out of 7 designed guides were produced by non-template PCR and *in vitro* transcription. All of guides were targeting the exon 10, downstream of FAD binding domain (Figure 11). By this approach, in case of actual mutation we would expect that the C-end of the protein could be removed or modified, without affecting the conserved part known to be crucial for protein function.

4.6 Egg screening

In order to test actual efficiency of designed guides, *in vivo* test was employed. Eggs were injected with Cas9/ guide RNA mixture and 30h afterwards DNA was isolated and targeted region was amplified by PCR and analysed for the presence of mutations. Since indels occurrence is random and can happen independently in different cells of developing embryo, many different mutations were expected to be seen.

To check for mosaicism in the eggs Heteroduplex Mobility Assay was used (Zhu et al., 2014). The main mechanism of the Heteroduplex mobility assay is based on the fact that heteroduplex formed from the DNA with an indel and wild type DNA will migrate slower than a homoduplex DNA on the polyacrylamide gel. Therefore PCR amplifying the sequence of the target region in heterozygote will show on the gel three different bands 1) band formed out of homoduplex wild type DNA 2). band formed out of homoduplex of mutated DNA and 3) slower migrating heteroduplex band formed wild type and mutated DNA (Zhu et al., 2014) Since eggs were injected in syncytium stage (many cell nuclei without cell membranes formed), each nucleus can be cut and repaired in different manner. Therefore many different heteroduplexes can be formed, which are seen on the gel as additional, slower migrating bands than wild type PCR product. In general – the more bands are observed, the more breaks and repair events occurred, which points to higher efficiency of the guide RNA used. Most of the embryonic DNA was not mutated, which is shown by the thick band identical to the size of the *cry2* PCR fragment from wild type animals.

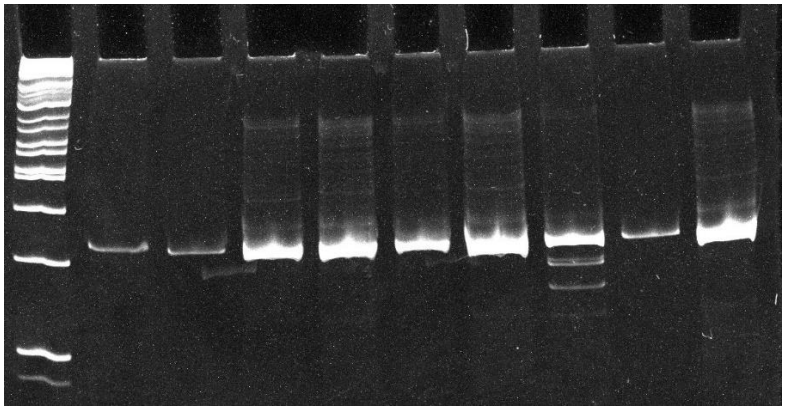

Figure 12. PCR gel egg with testing for Exon 10 188 F sgRNA (lanes from left: 1 - DNA ladder, 2 – wild type DNA – only one band is visible, 3-10 eggs injected with Cas9/sg RNA 188). Additional faster migrating bands (possible deletion) are seen in lane 8.

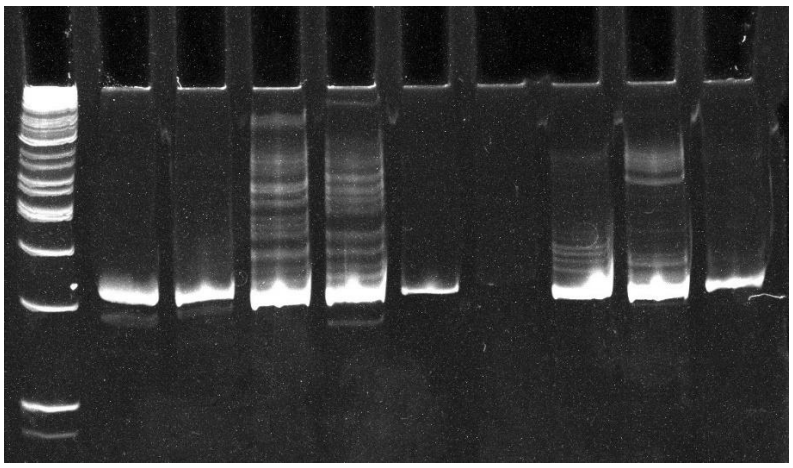

Figure 13. PCR gel with egg testing for Exon 10 194 R sgRNA. (lanes from left: 1 - DNA ladder, 2 – wild type DNA – only one band is visible, 3-10 eggs injected with Cas9/sg RNA 194). Many additional slower migrating bands (heteroduplexes indicating mutations) are seen in lanes 4,5,8,9.

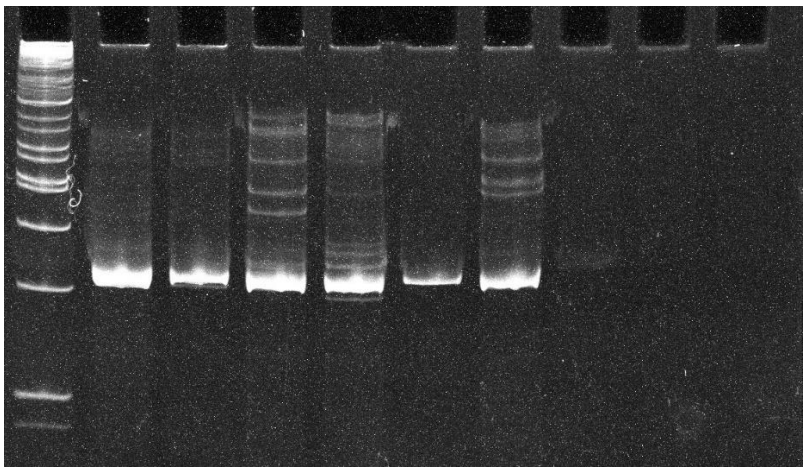

Figure 14. PCR gel with egg testing for Exon 10 196 F sgRNA. (lanes from left: 1 - DNA ladder, 2 – wild type DNA – only one band is visible, 3-7 eggs injected with Cas9/sg RNA 194). Many additional slower migrating bands (heteroduplexes indicating mutations) are seen in lanes 3,4 and 6.

Table 3. Summary of the *in vivo* sgRNA efficiency.

sgRNA	Eggs screened	Eggs fertilized	Mosaic
Exon 10 188 F	29	26	1
Exon 10 194 R	26	21	8
Exon 10 196 R	29	25	7

Similar number of eggs were injected for *in vivo* trial of selected guides' efficiency. Eggs were injected at the same time and were selected from the same cohort of eggs. Different guides are showing different efficiencies of ability to induce double stranded brakes which are then repaired by NHEJ mechanism. The most mutants showing mosaicisms were produced using sgRNA Exon 10 194R (38% eggs showed mosaicism), and Exon 10 196 F (28% eggs showed mosaicism) (Table 3).

Table 4. Efficiency of the mutant production

Eggs injected	**Eggs hatched**	**Survivors (G0)**	**Fertile adults (G0)**	***F1 screened***	***F1 heterozygotes***
221	81	31	21	*251*	*11*

Italics – results obtained in the laboratory by others (Kotwica-Rolinska 2019).

For the mutant production sgRNA with the highest efficiency was used – sgRNA Exon 01 194 R. The Table 4 shows the numbers of eggs injected. One third of injected eggs hatched and around 40% of hatched larvae reached adulthood. All surviving adults (G0 generation) were crossed separately with the wild type partners from Roana line. 21 of them had progeny (F1). All surviving adults (G0 generation) were crossed separately with the wild type partners from Roana line. 21 of them had progeny (F1). Because developmental time of *P. apterus* bugs is relatively long – around 1-1.5 month under laboratory conditions, therefore, further experiments – the screening of the F1 generation for heterozygous mutants detection was done by researchers of the Laboratory of the Molecular Chronobiology (Table 4 in italics). In total 11 heterozygous were found. Out of them 3 survived and the sequence of mutants was analysed. Two mutants resulted in identical deletion -2bp and one mutant shown deletion – 4 bp (Figure 12.) (Kotwica-Rolinska et al., 2019).

WT CTACTTACAGTTAATGAAATTCAGGCAG**CCA**GGTAATTTTTACTATACCA
-2bp CTACTTACAGTTAATGAAATTCAGGCAG**CCA**GG - - ATTTTTACTATACCA
-4bp CTACTTACAGTTAATGAAATTCAGGCAG**CCA**GG - - - - TTTTACTATACCA

Figure 12. Sequences of mutants obtained by CRISPR/Cas9 genome editing. Dashes mean deletion, underlined is a PAM sequence (data from Kotwica-Rolinska et al., 2019).

5 Discussion

All animals and insects possess a circadian- and some of them also a photoperiodic clock. These clocks help the organism to balance their day and have a temporal equilibrium with the environment surrounding them (Bell-Pedersen et al., 2005; Hardin, 2005; Partch et al., 2014). They are crucial for the correct metabolisms and behaviour of insects, however the exact molecular mechanism behind them is still unclear. Circadian rhythms are thought to originate from transcription/translation feedback loops (TTFL) which are regulated by the cell (Bell-Pedersen et al., 2005; Pritchett and Reddy, 2017) and are based on multiple interactions between the clock proteins. One of the most important proteins are the cryptochrome proteins, which in many insect species and in vertebrates act as repressors for the TTFL.

From comparison of CRY2 proteins in different organisms (Figure 8), it can be observed that the *P. apterus* CRY 2 protein has two highly conserved domains, DNA Photolyase domain and the FAD binding domain which are similar to other species. The amino acid composition is almost if not identical along the whole protein differing only in their C-terminal part. This indicates that these domains are crucial for the proper function of CRY proteins. Indeed mutation *P. apterus* CRY2 in DNA Photolyase domain produced with CRISPR/Cas9 (Kotwica-Rolinska et al., 2019) destroyed circadian rhythmicity of bugs (unpublished data). However, it can be seen that the C-Terminus differs among species and is highly variable (Figure 8) and it is not evolutionary conserved which implies that it is not crucial for the function of the protein. It is not easy to predict the role of this region, nonetheless in this area a lot of phosphorylation sites can be found (Figure 9). However, it was shown that site specific phosphorylation in the C-terminal part of the mouse CRY1 protein can alter the period of the circadian clock (Gao et al., 2013). Other research analysing altered circadian rhythms of both animals and humans show how important phosphorylation is when regulating the clock at the molecular level (Eide et al., 2002) and phosphorylation is known to affect the stability of circadian clock proteins, their activity as a transcription factors, or their cellular localization (Lee et al., 2014).

The aim of this study was to use CRISPR/Cas9 in order to successfully genetically modify the *Phyrrocoris apterus* genomic sequence of the c*ry2* gene. The main objective was obtaining mutants changing the stability of the CRY2, by either removing or modifying its

C-terminal without changing conserved CRY2 domains, making sure its function was not disrupted. This could be important for the better understanding of the circadian and photoperiodic clocks in non-model organisms.

Freely available online prediction tools like CRISPOR, which was used in this study, allow for easy design of guide RNAs targeting the specific place in the genome (Haeussler et al., 2016; Concordet and Haeussler, 2018). They also offer prediction of the efficiency of mutants' generation by particular sgRNA, so called on-target efficiency (Table 2). All of guide RNAs designed in this study had similar predicted efficiency, cry2 188 – 38, cry2 194 - 25 and cry2 196 - 30 (where the highest value is set as 100). To check if the on-target efficiency prediction algorithm offered by CRISPOR software is correct for choosing the best guide RNA, test of the actual efficiency of designed guides was performed. For injections, newly laid eggs were secured on tape and injected with CRISPR-Cas9 and sgRNA in the posterior part, because that's where the nucleus divides. The eggs were kept on a damp piece of paper in a Petri dish in order to maximize the chance of egg survival waiting for them to hatch (Figure 13).

Figure 13. Eggs in different stages. The eggs have an ovoid shape and the colour is usually white or off-white. However, near the end of the embryonic development, their colour turns yellow-red (Socha, 1993). Some larvae that have already hatched can be observed too.

30 hours afterwards, DNA was isolated and the targeted region was isolated and amplified by PCR in order to detect the presence of any mutation. In case of successful mutations in the *cry2* gene, no noticeable change to the outer appearance of the insect (e.g. colour change of body and/or eyes) will be present and thus a screening of the genome is necessary to analyse the changes in the sequence (Sun et al., 2017; Kotwica-Rolinska et al., 2019).

Indels caused by the double stranded breaks and following NHEJ repair can be detected through traditional PCR and gel electrophoresis on the non - denatured acrylamide gels. This method is mostly useful for detecting heterozygotes, since in heterozygotes, heteroduplex formed by the DNA with modifications and non-modified wild type DNA will migrate slower than homoduplex DNA on the gel and can be seen in the form of additional band (Zhu et al., 2014). The same principle can be applied for mosaicism screening, where one can expect appearance of many additional bands of heteroduplexes that do not appear on the wild type PCR product, if many cells were cut and repaired by NHEJ. Images of the different gels were compared to find the sgRNA that caused the most mutations. Analysis of the number of mosaics produced by different guide RNAs showed that in vivo on-target efficiency differs from values predicted by the CRISPOR software. In case of tested guides, the best was cry2 194 (38% eggs showing mosaicism), then cry2 196 (28% eggs showing mosaicism) and the last cry2 188 (3.8 % eggs showing mosaicism) and those values did not correspond to values predicted by CRISPOR software. Similar results were obtained in testing other guides RNA targeting different genes in *P. apterus* (Kotwica-Rolinska et al., 2019). Since it is extremely difficult to predict the efficiency of the mutations, the only way to see tangible results is to test the efficiency of particular guides *in vivo*.

The sgRNA with the highest efficiency was discovered to be sgRNA 194 which was subsequently used for final injections of the eggs. In total 221 eggs were injected of which 81 hatched, resulting in 31 survivors (Table 4). To see if mutations are passed to their progeny, all the surviving adults (G0 generation) were crossed with wild type partners from the Roana line.

Due to time restraints and the long developmental time of the insect (1 to 1.5 months), further screening and experiments were carried out by other researchers of the Laboratory of Molecular Chronobiology. It was shown that out of 251 screened F1 progeny, 11 heterozygotes were found. Only 3 of them survived and shown small deletions (-2bp and -4bp) near the predicted cut site (Kotwica-Rolinska et al., 2019). Those deletion imply that frame shift occurred in engineered protein. However to see any effect of those mutations on CRY2 function need to be established.

6 Conclusion

In conclusion, CRISPR/Cas9 is a relatively easy and practical method for gene engineering, although the process is time-demanding. *P. apterus* mutants were successfully generated and the production of sgRNA is not excessively time consuming: in fact, if working every day, it is possible to be able to inject G0 eggs even after two weeks from the start of the experiment. The real time limitation in the production of mutants is the generation time of chosen organism and its fertility and survival rate. The CRISPR/Cas9 is a technique that still needs some adjustments since it may be difficult to make extremely precise mutations due to low efficiency of homology directed-repair events and the results can be sometimes unpredictable. However, it is currently really useful to study non-model organisms (Sun et al., 2017) and scientists are starting to test it even on other organisms including humans (Liang et al., 2015). It is an extremely versatile approach that will definitely be frequently used in the future.

7 Literature

Bajgar, A., Jindra, M., and Dolezel, D. (2013). Autonomous regulation of the insect gut by circadian genes acting downstream of juvenile hormone signaling. Proceedings of the National Academy of Sciences of the United States of America 110(11), 4416-4421. doi: 10.1073/pnas.1217060110.

Bell-Pedersen, D., Cassone, V.M., Earnest, D.J., Golden, S.S., Hardin, P.E., Thomas, T.L., et al. (2005). Circadian rhythms from multiple oscillators: Lessons from diverse organisms. Nature Reviews Genetics 6(7), 544-556. doi: 10.1038/nrg1633

Concordet, J.P., and Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res 46(W1), W242-W245. doi: 10.1093/nar/gky354.

Dickinson, D.J., Ward, J.D., Reiner, D.J., and Goldstein, B. (2013). Engineering the Caenorhabditis elegans genome using Cas9-triggered homologous recombination. Nature Methods 10(10), 1028-+. doi: 10.1038/Nmeth.2641.

Ding, Y., Li, H., Chen, L.L., and Xie, K. (2016). Recent Advances in Genome Editing Using CRISPR/Cas9. Front Plant Sci 7, 703. doi: 10.3389/fpls.2016.00703.

Dolezel, D. (2015). Photoperiodic time measurement in insects. Current Opinion in Insect Science 7, 98-103. doi: 10.1016/j.cois.2014.12.002.

Eide, E.J., Vielhaber, E.L., Hinz, W.A., and Virshup, D.M. (2002). The circadian regulatory proteins BMAL1 and cryptochromes are substrates of casein kinase Iepsilon. J Biol Chem 277(19), 17248-17254. doi: 10.1074/jbc.M111466200.

Gao, P., Yoo, S.H., Lee, K.J., Rosensweig, C., Takahashi, J.S., Chen, B.P., et al. (2013). Phosphorylation of the cryptochrome 1 C-terminal tail regulates circadian period length. J Biol Chem 288(49), 35277-35286. doi: 10.1074/jbc.M113.509604.

Haeussler, M., Schonig, K., Eckert, H., Eschstruth, A., Mianne, J., Renaud, J.B., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol 17(1), 148. doi: 10.1186/s13059-016-1012-2.

Hardin, P.E. (2005). The circadian timekeeping system of Drosophila. Curr Biol 15(17), R714-722. doi: 10.1016/j.cub.2005.08.019.

Hoang, N., Schleicher, E., Kacprzak, S., Bouly, J.P., Picot, M., Wu, W., et al. (2008). Human and Drosophila cryptochromes are light activated by flavin photoreduction in living cells. Plos Biology 6(7), 1559-1569. doi: ARTN e16010.1371/journal.pbio.0060160.

http://crispor.tefor.net/.

https://www.addgene.org/guides/crispr/.

https://www.mirusbio.com/applications/genome-editing-using-crispr-cas/overview.

Hwang, W.Y., Fu, Y.F., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., et al. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature Biotechnology 31(3), 227-229. doi: 10.1038/nbt.2501.

Ingram, K.K., Kutowoi, A., Wurm, Y., Shoemaker, D., Meier, R., and Bloch, G. (2012). The molecular clockwork of the fire ant Solenopsis invicta. PLoS One 7(11), e45715. doi: 10.1371/journal.pone.0045715.

Ingrell, C.R., Miller, M.L., Jensen, O.N., and Blom, N. (2007). NetPhosYeast: prediction of protein phosphorylation sites in yeast. Bioinformatics 23(7), 895-897. doi: 10.1093/bioinformatics/btm020.

Jamal, M., Khan, F.A., Da, L., Habib, Z., Dai, J., and Cao, G. (2016). Keeping CRISPR/Cas on-Target. Curr Issues Mol Biol 20, 1-12.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science 337(6096), 816-821. doi: 10.1126/science.1225829.

Kistler, K.E., Vosshall, L.B., and Matthews, B.J. (2015). Genome Engineering with CRISPR-Cas9 in the Mosquito Aedes aegypti. Cell Reports 11(1), 51-60. doi: 10.1016/j.celrep.2015.03.009.

Kotwica-Rolinska, J., Chodakova, L., Chvalova, D., Kristofova, L., Fenclova, I., Provaznik, J., et al. (2019). CRISPR/Cas9 Genome Editing Introduction and Optimization in the Non-model Insect Pyrrhocoris apterus. Frontiers in Physiology 10. doi: ARTN 89110.3389/fphys.2019.00891.

Kotwica-Rolinska, J., Pivarciova, L., Vaneckova, H., and Dolezel, D. (2017). The role of circadian clock genes in the photoperiodic timer of the linden bug Pyrrhocoris apterus during the nymphal stage. Physiological Entomology 42(3), 266-273. doi: 10.1111/phen.12197.

Lee, E., Jeong, E.H., Jeong, H.J., Yildirim, E., Vanselow, J.T., Ng, F., et al. (2014). Phosphorylation of a Central Clock Transcription Factor Is Required for Thermal but Not Photic Entrainment. Plos Genetics 10(8). doi: ARTN e100454510.1371/journal.pgen.1004545.

Liang, P., Xu, Y., Zhang, X., Ding, C., Huang, R., Zhang, Z., et al. (2015). CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. Protein Cell 6(5), 363-372. doi: 10.1007/s13238-015-0153-5.

Lin, S., Staahl, B.T., Alla, R.K., and Doudna, J.A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. Elife 3, e04766. doi: 10.7554/eLife.04766.

Partch, C.L., Green, C.B., and Takahashi, J.S. (2014). Molecular architecture of the mammalian circadian clock. Trends Cell Biol 24(2), 90-99. doi: 10.1016/j.tcb.2013.07.002.

Peschel, N., and Helfrich-Forster, C. (2011). Setting the clock--by nature: circadian rhythm in the fruitfly Drosophila melanogaster. FEBS Lett 585(10), 1435-1442. doi: 10.1016/j.febslet.2011.02.028.

Pivarciova, L., Vaneckova, H., Provaznik, J., Wu, B.C.H., Pivarci, M., Peckova, O., et al. (2016). Unexpected Geographic Variability of the Free Running Period in the Linden Bug Pyrrhocoris apterus. Journal of Biological Rhythms 31(6), 568-576. doi: 10.1177/0748730416671213.

Pritchett, D., and Reddy, A.B. (2017). No FAD, No CRY: Redox and Circadian Rhythms. Trends Biochem Sci 42(7), 497-499. doi: 10.1016/j.tibs.2017.05.007.

Smykal, V., Daimon, T., Kayukawa, T., Takaki, K., Shinoda, T., and Jindra, M. (2014). Importance of juvenile hormone signaling arises with competence of insect larvae to metamorphose. Dev Biol 390(2), 221-230. doi: 10.1016/j.ydbio.2014.03.006.

Socha, R. (1993). Pyrrhocoris-Apterus (Heteroptera) - an Experimental-Model Species - a Review. European Journal of Entomology 90(3), 241-286.

Sun, D., Guo, Z.J., Liu, Y., and Zhang, Y.J. (2017). Progress and Prospects of CRISPR/Cas Systems in Insects and Other Arthropods. Frontiers in Physiology 8. doi: ARTN 60810.3389/fphys.2017.00608.

Tataroglu, O., and Emery, P. (2015). The molecular ticks of the Drosophila circadian clock. Curr Opin Insect Sci 7, 51-57. doi: 10.1016/j.cois.2015.01.002.

Thurtle-Schmidt, D.M., and Lo, T.W. (2018). Molecular biology at the cutting edge: A review on CRISPR/CAS9 gene editing for undergraduates. Biochemistry and Molecular Biology Education 46(2), 195-205. doi: 10.1002/bmb.21108.

Tomioka, K., and Matsumoto, A. (2010). A comparative view of insect circadian clock systems. Cell Mol Life Sci 67(9), 1397-1406. doi: 10.1007/s00018-009-0232-y.

Tomioka, K., and Matsumoto, A. (2015). Circadian molecular clockworks in non-model insects. Current Opinion in Insect Science 7, 58-64. doi: 10.1016/j.cois.2014.12.006.

Urbanova, V., Bazalova, O., Vaneckova, H., and Dolezel, D. (2016). Photoperiod regulates growth of male accessory glands through juvenile hormone signaling in the linden bug, Pyrrhocoris apterus. Insect Biochemistry and Molecular Biology 70, 184-190. doi: 10.1016/j.ibmb.2016.01.003.

Wyman, C., and Kanaar, R. (2006). DNA double-strand break repair: all's well that ends well. Annu Rev Genet 40, 363-383. doi: 10.1146/annurev.genet.40.110405.090451.

Yu, X., Liu, H., Klejnot, J., and Lin, C. (2010). The Cryptochrome Blue Light Receptors. Arabidopsis Book 8, e0135. doi: 10.1199/tab.0135.

Yuan, Q., Metterville, D., Briscoe, A.D., and Reppert, S.M. (2007). Insect cryptochromes: gene duplication and loss define diverse ways to construct insect circadian clocks. Mol Biol Evol 24(4), 948-955. doi: 10.1093/molbev/msm011.

Zhu, X., Xu, Y., Yu, S., Lu, L., Ding, M., Cheng, J., et al. (2014). An efficient genotyping method for genome-modified animals and human cells generated with CRISPR/Cas9 system. Sci Rep 4, 6420. doi: 10.1038/srep06420.