Bachelor Thesis

CRISPR/Cas9 gene editing in *Drosophila melanogaster*

University of South Bohemia in České Budějovice

Faculty of Science

České Budějovice, 2021

Konstantin Kaltenböck

Supervisor: David Doležel, PhD

Kaltenböck, K., 2021: “CRISPR/Cas9 gene editing in *Drosophila melanogaster*”, Bachelor of Science thesis, 30p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation

The aim of this thesis was to predict functional sites in the *timeless* gene of *Drosophila melanogaster*, to design fitting gRNAs for the identified sites to create transformation vectors for fly transformations and to induce CRISPR/Cas9 mediated target mutations by crossing gRNA expressing flies with Cas9 expressing flies.

Affirmation

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 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.

Date.....

České Budějovice

Signature .....

Konstantin Kaltenböck

Acknowledgement

I would like to sincerely thank my supervisor David Doležel, PhD for the chance he gave me to explore his scientific field. He was a great mentor and teacher and supported me with not only his expertise but also as an excellent motivator.

Furthermore, I want to direct some words towards Samarjeet Singh, PhD who found time for me during his own dissertation. He is not only an exceptional instructor in the lab but also a great companion.

I also want to thank Olga Bazalová, PhD for her patient guidance through the first part of the lab procedure. Her reassurance really helped me to get started with this thesis.

Contents

1. Abstract

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and their associated endonucleases (Cas) are prokaryotic adaptive immune systems that target the cleavage of nucleic acids. This target specific nucleic acid cleavage system, when introduced into eukaryotic organisms, provides an efficient method for genetic engineering. Targeted DNA cleavage is achieved by the introduction of a guide RNA to an organism expressing a Cas9 gene. In this thesis, potentially functionally essential sites in the *timeless* gene of *Drosophila melanogaster* were identified by the Biological sequence analysis tools NetNes and NetPhos and the investigation of the evolutionary rate of the loci through MUSCLE alignment to design gRNAs. Molecular cloning was performed to obtain vectors for potential fly transformation. In the second part of the experiment, transformed gRNA expressing flies were crossed with Cas9 expressing flies to obtain site specific deletion mutations at the gRNA targeted site in the used stocks of *Drosophila melanogaster.*

2. Aims of the thesis

- Prediction of candidate regions in TIMELESS protein worth of mutagenesis
- Construction of gRNA for CRISPR/Cas9 mediated fly mutagenesis
- Crossing of mutated *Drosophila melanogaster* strains to obtain homozygous mutant flies in the *timeless* gene
- Identification of mutants in the *timeless* gene

3. Abbreviations

bp	Base pair
Cas	CRISPR associated protein
CDK1	Cyclin-dependent kinase 1
CLK	Clock
cNLS	CDK dependent nuclear localization signals
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CYC	Cycle
DBT	Doubletime
MUSCLE	Multiple Sequence Comparison by Log-Expectation
NES	Nuclear Export Signals
PAM	Protospacer adjacent Motif
PCR	Polymerase Chain Reaction
PER	Period protein
per	*period* gene and mRNA
sgRNA	Single guide RNA
TIM	Timeless protein
tim	*timeless* gene and mRNA
tracrRNA	Trans-activating CRISPR RNA

4. Introduction

4.1. CRISPR/Cas System

Clustered regularly interspaced palindromic repeats (CRISPR) and their associated Cas proteins developed as an adaptive immune system against phages in prokaryotes. Upon a primary infection, short fragments of intruding foreign DNA sequence are integrated into the genome of the host and can be recognized and in the case of a new infection cleaved. The process can be characterized by 3 general steps. In the first step, the viral DNA is recognized a fragment, the protospacer incorporated. This has shown to frequently happen by the identification of a PAM motif which is a 2-5 base pair(bp) long motif that is located adjacent to a protospacer. In the second step, the newly incorporated protospacer locus is transcribed into a pre-CRISPR RNA which then is enzymatically cleaved by endoribonucleases yielding a mature CRISPR RNA (crRNA). In a third step, the crRNA binds to its complementary strand of viral DNA in the case of a new infection. Upon the formation of a complex with the Cas-protein, the viral DNA is then cleaved (Wiedenheft et al. 2012; Terns et Terns, 2011; reviewed in Jinek et al. 2012; reviewed in Gleditzsch et al. 2019).

Several CRISPR systems are required for this immune system. The different kinds of Cas-proteins have different functions. For instance, Cas1 and Cas2 are required for the integration of foreign DNA in the form of spacers. Then, these spacers then are transcribed and cleaved to yield a crRNA. Cas9 proteins are nucleases that form complexes with this crRNA to target the sequence that was integrated by Cas1 and Cas2 and cleave them. The Cas9 protein requires a PAM sequence downstream of the target to cleave it (Nuñez et al. 2014; reviewed in Gleditzsch et al. 2019).

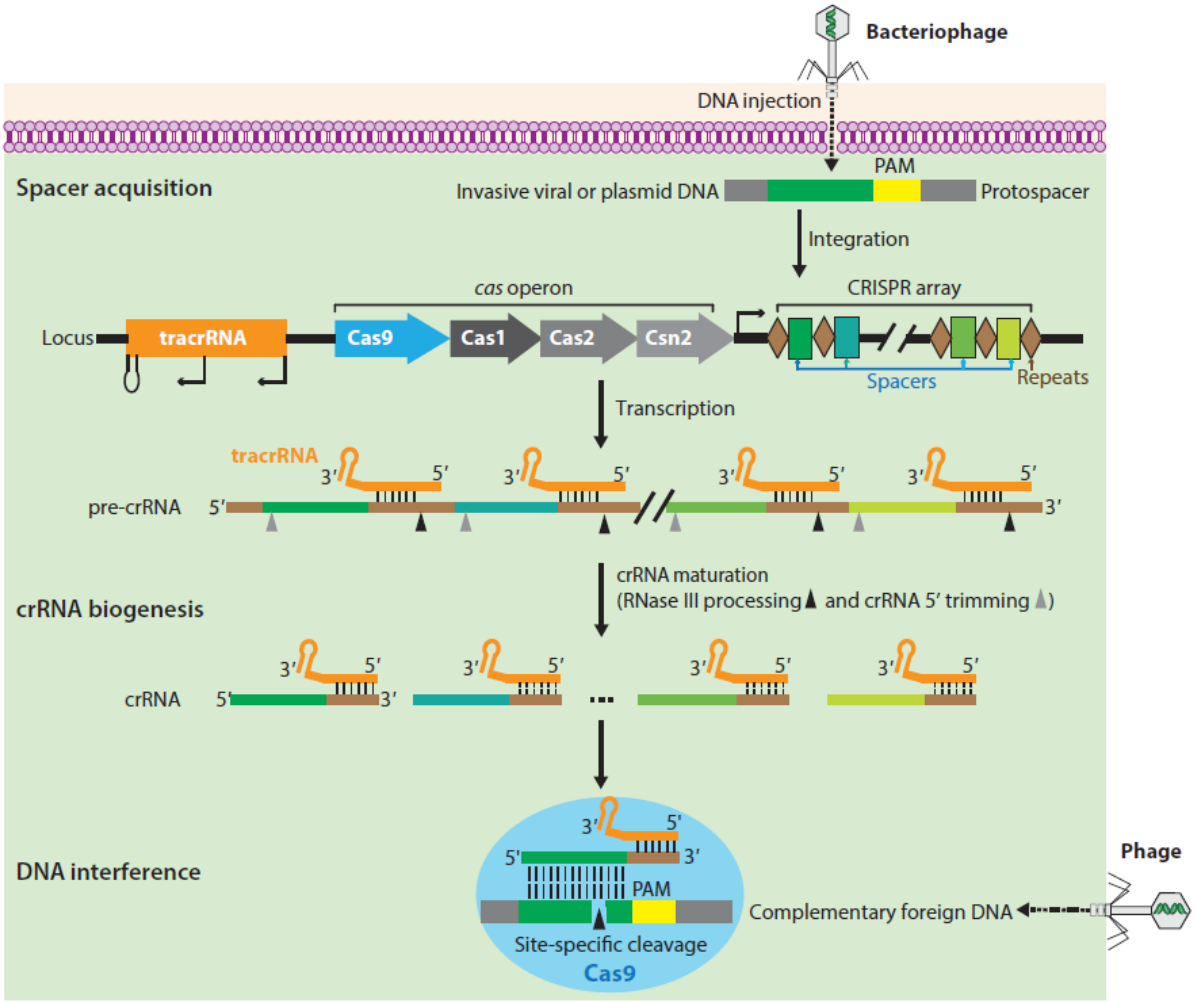

Figure 1: CRISPR/Cas system in 3 main steps: Spacer acquisition of viral DNA via PAM identification and integration by Cas1 and Cas2, crRNA biogenesis by transcription and crRNA maturation and DNA interference mediated site specific cleavage in the case of a sequential infection with the same foreign DNA (Jiang et Doudna, 2017).

4.2. CRISPR/Cas9 gene editing

The Cas9 protein is an endonuclease that targets a specific sequence upstream of a PAM motif and cleaves it. In its original purpose, this is used to silence foreign viral DNA. However, if a PAM sequence is present immediately downstream of a locus within a genomic DNA of the host, it is targeted by a Cas9 crRNA complex with an identic target sequence. A single guide RNA can be constructed for specific gene targeting. It consists of a 20 nucleotides long target sequence and a crRNA that forms a chimeric homologous pairing with a tracrRNA (Bhaya et al. 2011).

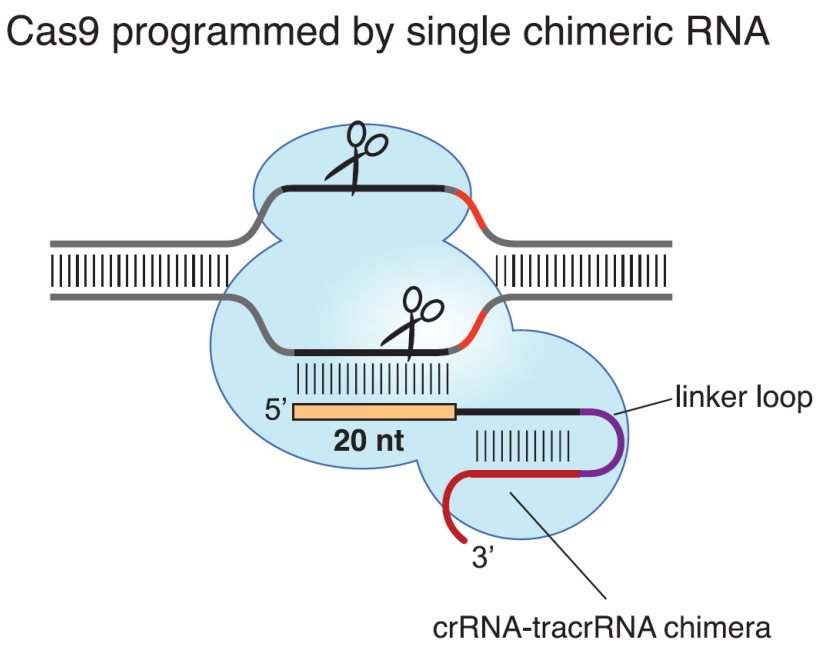

Figure 2: Cas9-sgRNA complex. A 20nt long target sequence is constructed into a sgRNA with a crRNA-tracrRNA chimera. The target sequence is pairing with its homologue in of the investigated DNA locus of the host. The PAM sequence downstream of the locus is shown in red, the cleavage site is indicated by scissors (Jinek et al. 2012).

4.3. CRISPR/Cas9 Gene editing in *Drosophila melanogaster*

The CRISPR/Cas9 system can be used to induce mutations in the genome of *Drosophila melanogaster* at specific loci with PAM motives immediately downstream. A site-specific

gRNA and a Cas9 expression are required within the organism. To obtain both characteristics, a fly expressing a Cas9 gene is crossed with a fly expressing the fitting gRNA. A gRNA with the sequence of interest can be digested into a vector with a marker mutation and an attB site. The phiC31 integrase then inserts the attB-containing plasmid into a specific attP site, so called docking site, located in *Drosophila* genome. When this integration happens in germline cells, a stable genomic mutation occurs. The transgenic flies then express a marker mutation and the gRNA (Kondo et Ueda 2013).

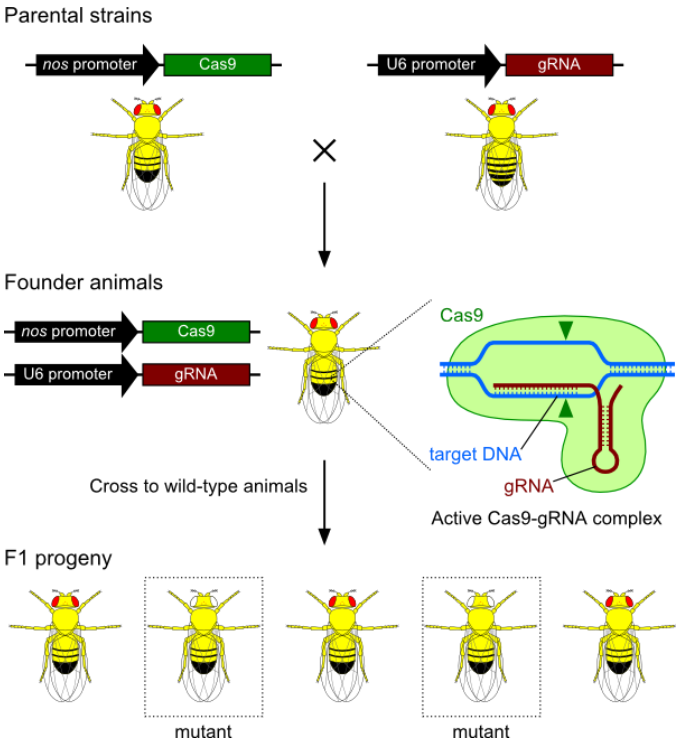

Figure 3: Crossing scheme of a CRISPR Cas9 gene editing system in *Drosophila melanogaster* (Kondo et Ueda, 2013).

When the transgenic flies are then crossed with flies that express a Cas9 gene their offspring is expressing an active Cas9-gRNA complex. This active complex leads to the continuous cleavage of the targeted locus. Therefore, cells are targeted in different developmental states which induces a mosaic genotype of the targeted locus (Oliver et al. 2015; reviewed in Mehravar et al. 2019).

4.4. Circadian clock

The circadian clock is the controlling mechanism of the physiological and behavioural rhythm of functions such as the eclosion activity, the locomotor activity and sleep patterns with a period of approximately 24 hours. This rhythm relies on a molecular transcription-based feedback loop of the proteins Period (PER) and Timeless (TIM) and their transcriptional factors Clock (CLK) and Cycle (CYC). The mRNAs of *period* and *timeless* regulate their own transcription in a negative feedback loop (Hardin et al. 1990; reviewed in Dubowy et Sehgal 2017).

The 2 transcription factors CLK and CYC both are characterized by a DNA binding domain and a protein dimerization domain. They reach their highest concentration levels around dawn and can heterodimerize and bind to the E-box sequences that promote the transcription of *timeless* and *period*. The translated TIM and PER proteins reach their peak concentrations during the night. They form PER-TIM dimers that bind to the CLK-CYC dimers inhibiting them from promoting TIM and PER translation. The promotion of TIM and PER is then halted until the PER-TIM dimers are no longer present in the nucleus. This is the case at around noon of the following day (Hardin et al. 1990; reviewed in Stanewsky, 2002).

An important factor in this loop is the delay of the PER-TIM dimer production after the CLK-CYC induced transcription promotion. This delay is caused by the destabilization of PER via the casein kinase DBT and additional kinases. In the meantime, TIM accumulates (Kloss, 1998). This destabilization process is active until the abundancy of TIM proteins is high enough to form dimers with the PER proteins before they react with DBT (Hardin et al. 1990; Stanewsky, 2002).

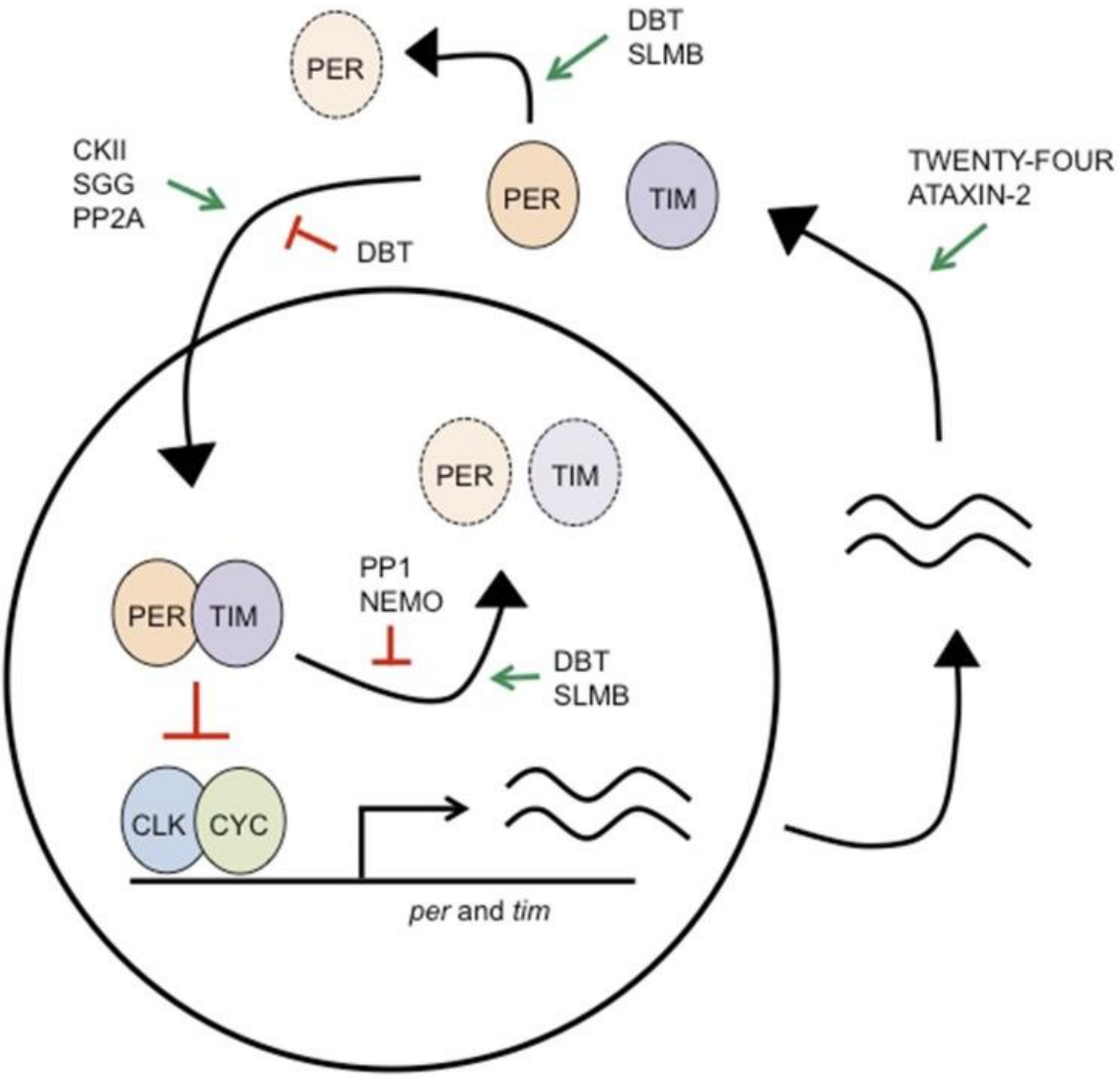

Figure 4: Self-regulation of TIM and PER in a feedback loop with CLK and CYC. The transcription of *tim* and *per* is self-regulated (translation of *per* in addition is also regulated by Twenty-four and Ataxin-2). This self-regulation is caused by the interaction of TIM and PER with their transcriptional activators CLK and CYC. The kinases doubletime (DBT) and NEMO delay the nuclear entry of PER and phosphorylate it in the nucleus which promotes degradation target recognition by SLMB (Dubowy et Sehgal 2017).

5. Materials and Methods

5.1. gRNA design

In this thesis, the timeless gene *tim* in *Drosophila melanogaster* was subject of investigation. To obtain loci in the *tim* gene of different species with a varying degree of relation to *Drosophila melanogaster*, the amino acid sequence of TIM of *Drosophila melanogaster* (NCBI Reference sequence NP_722912.3) has been aligned with the *tim* sequences of the following species using MUSCLE alignment in Geneious software: *Drosophila obscura* (XP_022222457.1), *Drosophila virilis* (O17482.1), *Drosophila ficusphila* (XP_017047591.1), *Musca domestica* (XP_019894051.1), *Aedes aegypti* (AAY40757.1), *Aedes albopictus* (AEX14536.1), *Ephestia kuehniella* (AJG06482.1), *Tribolium castaneum* (NP_001106936.1), *Pyrrhocoris apterus* (GFOX01093283.1), *Blattella germanica*(GDCR01026697.1), *Thermobia domestica* (BAL27710.1), *Halymorpha halys* (XP_014278196.1) and *Cimex lectularius* (XP_024082246.1; XP_014260027.1; GDEK01006033.1)

NetNES 1.1 (La Cour et al. 2004) was used to predict nuclear export signals in the TIM protein of *Drosophila melanogaster* and NetPhos 3.1 (Blom et al. 1999) to predict phosphorylation sites. Bipartite nuclear localization signals were predicted using cNLS Mapper (Kosugi et al. 2009). The predicted resulting sites of these tools were then investigated on the MUSCLE alignment and searched for the protospacer adjacent motif PAM. The obtained sites of interest have been used to design 2 complementary 24 bp long oligonucleotides with a common 20 bp long target sequence (Singh et al. 2019) (Ren et al. 2014).

5.2. Vector design

The general gRNA expression vector, pBFv-U6.2 (Kondo et Ueda, 2013) was used in the procedure. The vectors have been digested with the restriction enzyme BbsI .2 µl of BbsI were heated to 37°C, then 15 µl of plasmids in solution (101 µg/ml), 2 µl 10x NEBuffer and 1 µl nuclease free water were added. The reaction was incubated for 1 hour at 37°C (New England Biolabs, 2021).

The digested vectors were then electrophoresed in a 1% agarose in TAE (using the same procedure as in colony PCR) and the resulting gel was cut out at the site of product and purified with the QIAquick Gel Extraction Kit: the cut-out gel piece was incubated for 30 minutes at 50°C and then centrifuged at 13000 rpm for 1 minute. The supernatant was put into a QIAquick spin column and 3 times the volume of the supernatant of buffer QG was added. The solution was then centrifuged at 13000 rpm allowing the DNA to adsorb to the silica membrane of the column. The flow-through was discarded, 0.75 ml of buffer PE were added to the column and the mixture was again centrifuged for 1 minute at 13000 rpm. Then, the column was placed into a new microcentrifuge tube and 50 µl of Buffer EB was used to elute the DNA on the column surface. The mixture was centrifuged again for 1 minute at 13000 rpm which yielded the extracted, purified vector DNA in the flow-through solution (QIAGEN, 2001).

The 2 oligonucleotides have been annealed to yield 20 bp long gRNA encoding DNA with 4 bp long overhangs at both ends using Phusion HF Buffer and ligated into the digested vectors using T4 DNA ligase: 1 µl T4 ligase, 1 µl T4 DNA ligase buffer (10x), 3 µl vectors, and 5 µl gRNA were added together and cooled on ice. The reaction was then mixed by pipetting and incubated at room temperature for 10 minutes. Then the reaction was inactivated by heat shock at 65°C for 10 minutes. The mixture was then again cooled on ice (New England Biolabs, 2021).

5.2.1. Vector Transformation

The general gRNA expression vector pBFv-U6.2 (Kondo et Ueda, 2013) containing the gRNA of interest and an antibiotic resistance gene were then transformed into competent *Escherichia coli*:

Agar plates were prepared by heating LB Agar letting it cool to 50°C and adding ampicillin with a concentration of 100 µg/µl. After polymerization on the plates, 50 µl of X-gal were added and spread on the surface. 100 µl of the competent *Escherichia coli* were thawed on ice and added to an Eppendorf tube together with 5 µl of the gRNA expression vector. It was then left for 20 minutes on ice. Subsequently, the solution was heat shocked for 45 seconds at 42°C and cooled on ice for 2 minutes. Afterwards, 250 µl of SOC medium was added. The solution was then incubated in a shaker for 1 hour at 37°C, then the transformed cells were spread on the prepared agar plates and incubated overnight at 37°C.

5.2.2. Plasmid DNA isolation

The plasmid DNA of single colonies were then purified using High Pure™ Plasmid Isolation Kit (Roche, 2014):

The *Escherichia coli* colonies of the Agar plates were transferred and individually inoculated in 3 ml LB medium. 1.5ml of the LB media were centrifuged at 8000 rpm for 90 seconds at 4°C. After the supernatant was separated from the pelleted bacterial cells, the remaining 1.5ml of the LB media were centrifuged to the same tube at the same conditions. 250 µl of Suspension Buffer with RNase A were added to the resulting pellet and the mixture was incubated for 5 minutes. Subsequently, 250 µl of Lysis Buffer were added and the sample was mixed by inverting the container. After a 5-minute-long incubation at room temperature, 350 µl of pre-cooled Binding Buffer was added, the sample was then mixed by inversion and incubated on ice for 5 minutes. Afterwards, the sample was centrifuged at 4°C and maximal speed, the resulting supernatant was transferred into a filter tube of a collection tube and centrifuged for another minute at maximal speed. The resulting flow-through liquid was discarded and 700 µl of washing buffer II were added before the sample was centrifuged again at maximal speed for 1

minute. Subsequently the flow-through was then again centrifuged at full speed for 1 minute to remove the residual washing buffer II. The filter column was then transferred to a new collection tube together with 50 µl of elution buffer and the assembly was then centrifuged for 1 minute at full speed.

5.2.3. Colony PCR

To obtain a high concentration of isolated plasmid DNA for gel electrophoresis analysis, it was amplified by Polymerase chain reaction (PCR). For this reaction, a solution of 12.5 µl diluted PPP master mix 1 µl bacterial template, 1 µl of each forward and reverse primer and 9.5 µl water was prepared. This solution was then gently mixed and transferred into a PCR thermocycler and amplified in 35 cycles.

The resulting PCR product was then investigated by agarose gel electrophoresis. A 1% agarose gel was prepared by adding 0.45 g agarose to 45 ml TAE Buffer. The agarose was then dissolved by heating (to keep the initial concentration, boiling was prevented). 2 µl gel red were added to the solution, and it was poured into the apparatus. A comb was applied to the solution and it was left to cool down and solidify. The comb was then removed and TBE buffer added onto the gel. The samples and a 1 kb plus DNA ladder marker were loaded into individual wells and electrophoresed at 120 V (Top-Bio, 2017).

5.3. Fly transformation

For clarity of the entire procedure, I also describe here the Drosophila transgenesis, albeit I did not perform this experiment. Instead, pre-prepared founder flies with another mutation were used.

To integrate the expression vector pBFv-U6.2 into the attpP landing site of the germ line cells of the TBX-0003 strain *Drosophila melanogaster,* the embryos were injected with said vectors at their posterior end. Out of the surviving flies, individual lines of founder flies were established by crossing them with the Cas9 encoding *Drosophila melanogaster* strain CAS-0003 (Kondo et Ueda 2013).

5.4. Germ line screening

The founder flies were dissected and their gonads PCR-screened to investigate mosaicism of the targeted gene:

The female flies were fed with yeast 2 days prior to fatten up the ovaries. Prior to the dissection, the flies were anaesthetized on CO_2 fly pads. They were then individually grabbed at their lower thorax and submerged in a 1X PBS buffer using tweezers. A second pair of tweezers was used to pull at the lower abdomen to expose the internal organs. The gonads were then detached from the other organs and transferred into an Eppendorf tube (Wong et Schedl 2006).

The obtained gonads were screened by PCR:

To each dissected gonad, 50 µl of squishing buffer and 0.5 µl of Proteinase K was added and the gonad was completely squished and mixed. The sample was then incubated for 37°C for 30 minutes and subsequently heat shocked at 95°C for 1 minute. For the PCR, 2 µl of the sample were added to 2 µl H_2O, 5 µl master mix and 0.5 µl of each, forward and reverse primer. The target DNA was then amplified in PCR 35 cycles in a thermal cycler. The PCR product was then investigated by gel electrophoresis using a 1% agarose gel in TAE buffer and gel red to investigate mosaicism of the targeted gene (Singh et al. 2019).

5.5. Crossing scheme

The used founder flies expressed the mosaic gene mutations M^1 and M^2 of the gRNA targeted site on the second chromosome and an active Cas9-gRNA complex on the third chromosome. Virgin founder flies were collected and individually allowed to breed with virgin males expressing the balancer Cyo/Sco on the second chromosome. They were kept in vials in an incubator with fly diet at 20°C with an alternating 12-hour light and 12-hour dark photoperiod. This resulted in an F_1 progeny heterozygous on the second chromosome expressing a mosaic gene mutation and a balancer as seen in the following schematic summary:

$$\frac{M^1}{M^2};\frac{Cas9}{gRNA}\ x\ \frac{Sco}{Cyo};\frac{WT}{WT}\ \rightarrow\ \frac{M^1}{Cyo}, or \frac{M^2}{Cyo}, or \frac{M^1}{Sco}, or \frac{M^2}{Sco}$$

Heterozygous virgin flies of each F_1 stock were then collected and allowed to individually breed at beforementioned conditions. The phenotypical wild type resulting F_2 generation then expressed the homozygotic targeted mutated chromosome.

$$\frac{M^1}{Cyo} x \frac{M^1}{Cyo} \rightarrow \frac{M^1}{M^1} \; and \; \frac{M^2}{Cyo} x \frac{M^2}{Cyo} \rightarrow \frac{M^2}{M^2}$$

Each homozygous stock was then cultivated. Of each stock one fly was screened for a successful targeted mutation by PCR SDS-Page screening analogous to the gonad screening. However, in this screening the entire body was screened, and no dissection was performed (Singh et al. 2019).

6. Results

6.1. gRNA

MUSCLE alignment of the used *tim* sequences yielded the following 2 regions of interest:

In the first region, the cNLS Mapper prediction yielded a bipartite nuclear localization signal score of 7.3

```
                                  Drosophila melanogaster   N-TKPR----NKPRTIMSPMDKKELRRKKLVKRS
                                       Drosophila obscura   N-TKPR----NKQRLILSPLDKKELRRKKLVKRS
                                       Drosophila virilis   S-SKHR----SKQRIFAVPQDTKDLRRKKLVKRS
                                   Drosophila ficusphila   N-TKPR----NKPRTIMSPMDKKELRRKKLVKRS
                                         Musca domestica   S-SKSR----KGQRNVLTAIDKKELRRKKLVKRS
                                           Aedes aegypti   N-QKQRFNAANKSRNPTTIHEKKELRRKKLVKRS
                                        Aedes albopictus   N-QKQRFNAANKSRNPTTIHEKKELRRKKLVKRS
                                    Ephestia kuehniella   N-PKQRIN--NKLRPSVMLQDR---KRKKIVKRG
                                    Tribolium castaneum   F-QKSRYN--TGKSRAATALDKKELRRKKLVKRS
                                    Pyrrhocoris apterus   MIKKSRY---ANGKVSLTTQEKEELKRKKLMRRF
                                     Blatella germanica   M-QKVR----YSSKTNIQLQDKKEWRRKKLVKRS
                                    Thermobia domestica   MVQKAR-----PAKPIQTQQDKNESRRKKLVKRS
                                      Halyomorpha halys   MIQKSRF---LNCKTQLSLQDKAELKKKKLMKRF
Cimex lectularius timeless isoform X3   MLQKSRF---KDGKPTLTPQEKEEMRRKKLMKRS
Cimex lectularius timeless isoform X2   MLQKSRF---KDGKPTLTPQEKEEMRRKKLMKRS
Cimex lectularius timeless isoform X1   MLQKSRF---KDGKPTLTPQEKEEMRRKKLMKRS
```

NetPhos 3.1 predicted phosphorylation sites at the amino acids highlighted in red with the following output.

Position of amino acid in the *tim* sequence of *Drosophila melanogaster*	Amino Acid	NetPhos Score	Kinase
534	Threonine	0.837	*Protein kinase C*
545	Serine	0.992	unspecified
561	Serine	0.857	unspecified

A PAM motif can be found in the form of Valine in the 3’ direction of our chosen gRNA motif which is highlighted in yellow. The following primers were then designed to obtain the fitting gRNA.

gRNA	5’ GGAGCTTAGACGCAAAAAAC 3’
Forward primer	5’ CTTCGGAGCTTAGACGCAAAAAAC 3’

Reverse primer	3’ CCTCGAATCTGCGTTTTTTGCAAA 5’

In the second region of interest, the following MUSCLE alignment was obtained:

```
                    Drosophila melanogaster   AGESMDDGDYEDQRHRQLNEHGEED
                         Drosophila obscura   DVEDYEESDFILRQLQRAR-----Q
                         Drosophila virilis   AANYILEGPCSAQQPWSDCEMQEYK
                      Drosophila ficusphila   AGESMDDGDYEDQPRSQDD---DED
                            Musca domestica   KEAMANEEATTTPTKIKIE---PNK
                              Aedes aegypti   SIKVIEGVSCFVKHPSCIK---PVT
                           Aedes albopictus   SIKVIEGVSCFVKHPSCAK---SVS
                         Ephestia kuehniella   TKAAENKDKSGSVENI---------
                         Tribolium castaneum   ---------STETETS---------
                         Pyrrhocoris apterus   ---------TGKRKQS---------
                          Blatella germanica   ----SSDSMSGKQHMT---------
                         Thermobia domestica   YNKNIPKGDLEVSKKSQNK---SPK
                           Halyomorpha halys   -------KPAGKRKQS---------
Cimex lectularius timeless isoform X3   ---------TGGRKQS---------
Cimex lectularius timeless isoform X2   ---------TGGRKQS---------
Cimex lectularius timeless isoform X1   ---------TGGRKQS---------
```

To be noted is that this sequence region shows a high variety between the investigated species. A phosphorylation site at the amino acid highlighted in red with the following output is predicted by NetPhos 3.1. The gRNA motif is highlighted in yellow.

Position of amino acid in the *tim* sequence of *Drosophila melanogaster*	Amino Acid	NetPhos Score	Kinase
364	Thyrosine	0.982	*unspecified*

The gRNA and primers designed for this region were the following ones:

gRNA	5’ GCCGGCGAGTCCATGGACGA 3’
Forward primer	5’ CTTCGCCGGCGAGTCCATGGACGA 3’
Reverse primer	3’ CGGCCGCTCAGGTACCTGCTCAAA 5’

6.2. *E. coli* colony screening

To verify the ligation of the gRNAs into the used vector and its successful transformation into the *E. coli* colonies, a colony screening was performed. The restriction enzyme BbsI was used for the plasmids' digestion. The digested plasmids were investigated using agarose gel electrophoresis. Fragments of the size of ~ 760 bp were expected to have successfully incorporated the insert of interest.

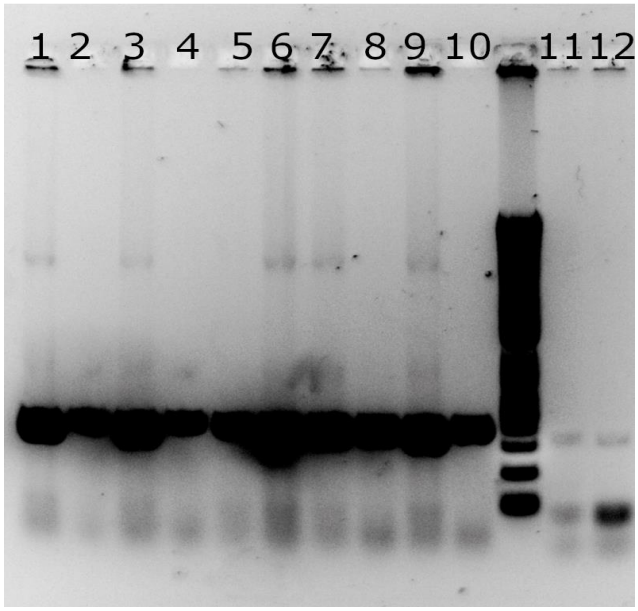

Figure 5: Colony gel electrophoresis of pBFv U6.2 plasmid inserted with the gRNA targeting the NetPhos region at the tim site 534- 561 digested by BbsI restriction enzyme. A 1 kb ladder was used as marker, the expected product size was at ~ 760 bp. The colonies 11 and 12 were sent for sequencing.

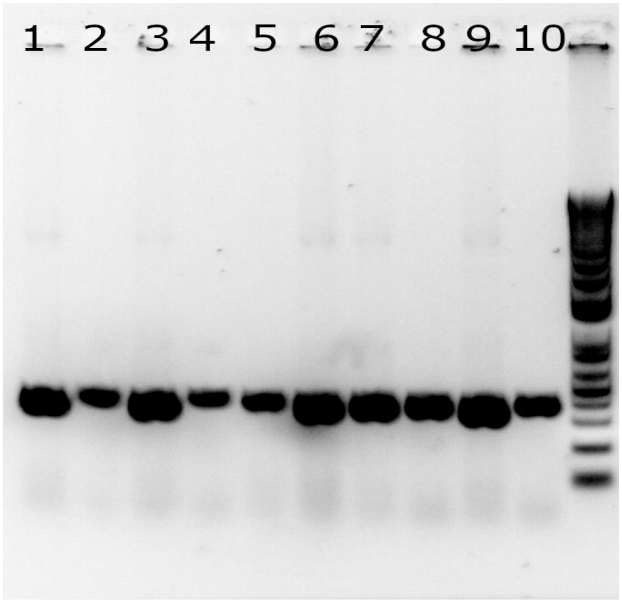

Figure 6: Colony gel electrophoresis of pBFv U6.2 plasmid inserted with the gRNA targeting the NetPhos region at the tim site 364 digested by BbsI restriction enzyme. A 1 kb ladder was used as marker, the expected product size was at ~ 760 bp. The colonies 4 and 8 were sent for sequencing.

6.3. Gonad PCR screening

The gonads of the injected G_0 flies were screened for heteroduplexes. These might occur during PCR when wild type and mutant amplicons anneal. As these DNA amplicons are not entirely complementary, they can form bubbles at the mismatching sites between the amplicons. This led to PCR products of different lengths which were separated by gel electrophoresis (Ota et al. 2013; Foster et al. 2019).

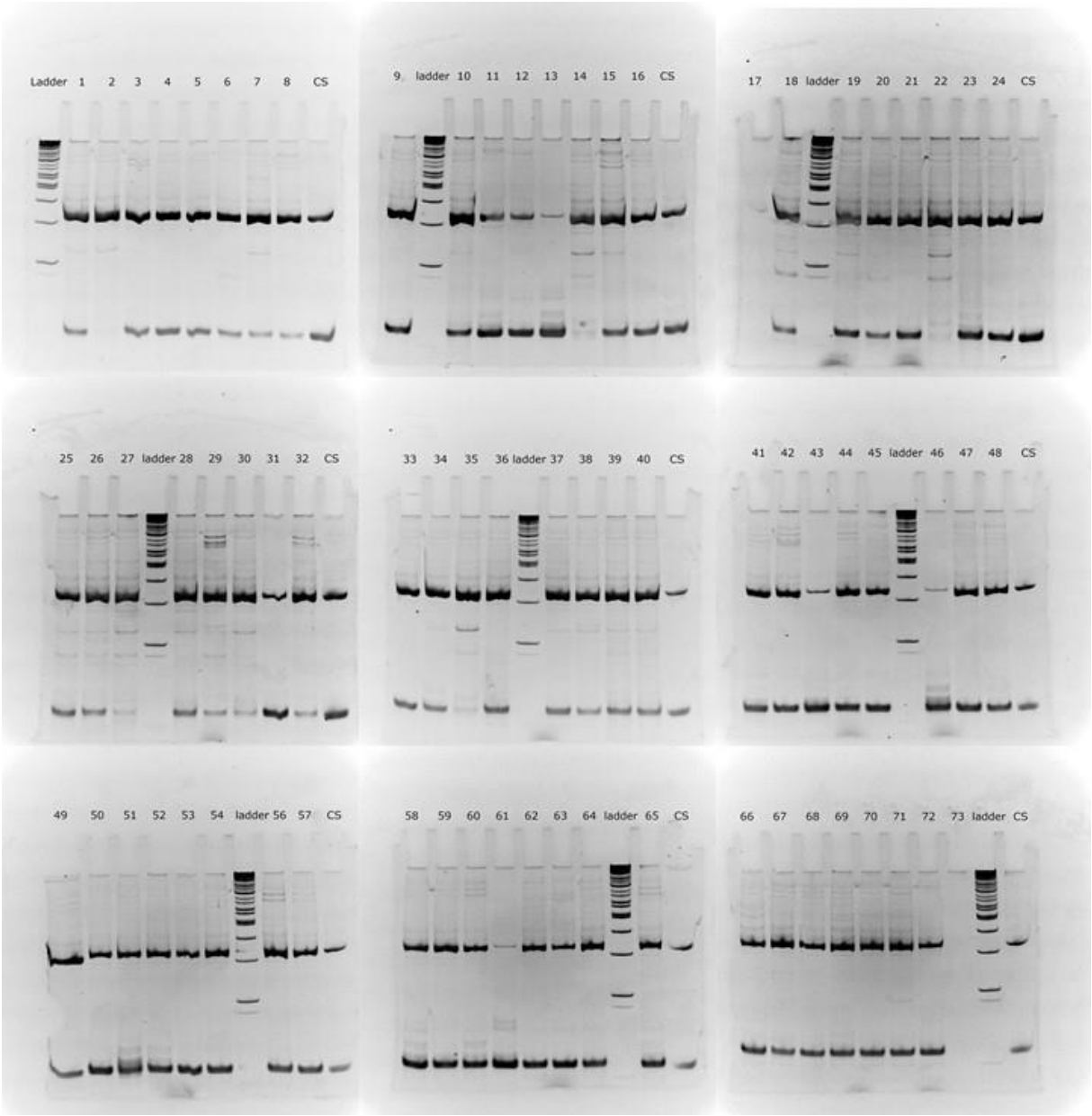

Figure 7: Heteroduplex mobility screening of the G_0 flies via gel electrophoresis. With a wildtype fly depicted as CS (Canton S). The G_0 flies number 3, 6, 7, 28, 29, 32, 35, 42, 48, 54,

56, 60, 61, 62, 68, 69 and 72 showed heteroduplexes. Their progeny was then crossed with the balancer fly.

6.4. Target mutation screening

The F_1 generation heterozygous flies were screened via electrophoresis to verify a successful mutation. PCR was performed with the primers of the constructed gRNA. Multiple bands of different lengths in the trans mutated stocks were used to identify them.

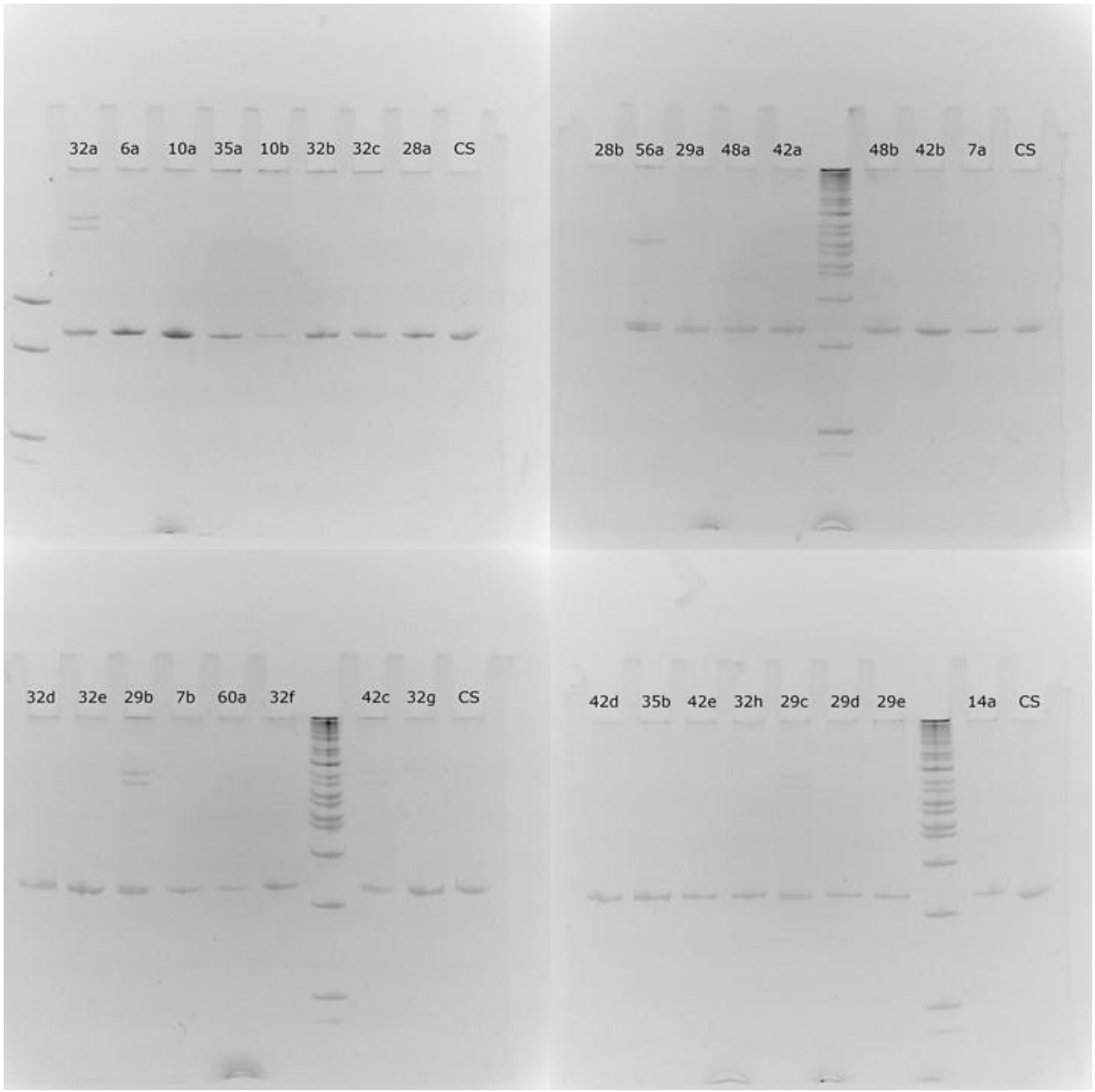

Figure 8: Heterozygote screening of the F_1 fly stocks via gel electrophoresis. With a wildtype fly depicted as CS (Canton S). The F_1 stocks are numbered after their G_0 transgenic parent flies. The individual stocks are described with letters. The stocks 32a, 56a, 29b and 29c show

multiple bands characteristic for heterozygotes. They were used to establish the F_2 homozygous stocks.

The F_2 generations were expected to have homozygous target mutation of our investigated sites at the *tim* gene. Electrophoresis was performed to verify the homozygous identity of the stocks. Then, the target mutations were verified via Sanger DNA sequencing (company SeqMe).

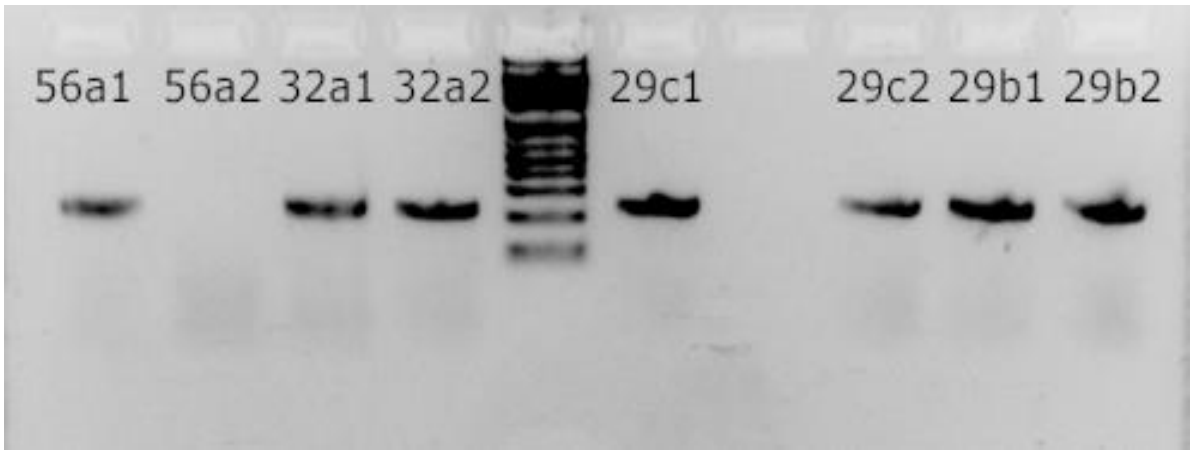

Figure 9: Screening of the PCR products of the gRNA primers of the homozygous stocks. The second PCR screening of stock 56a was unsuccessful. All screened stocks show single bands characteristic for homozygotes and were sent for Sanger sequencing (company SeqMe).

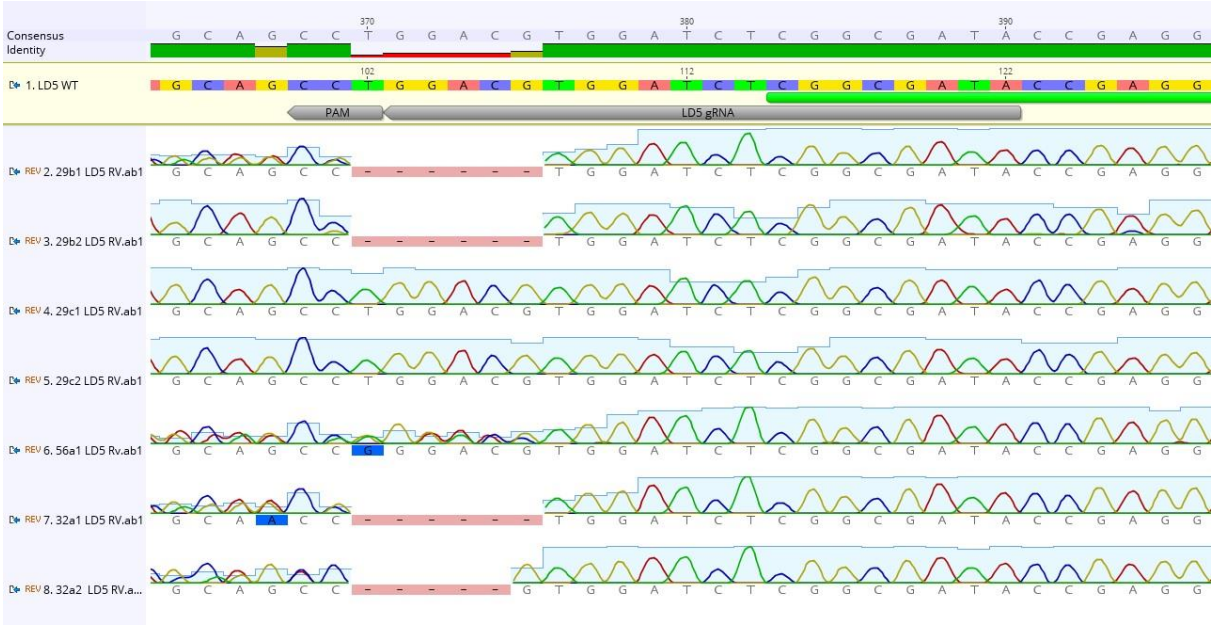

Figure 10: Results of the sequencing of the PCR products from figure 9 with chromatograms. A successful targeted deletion of 6 bp was found in the stocks 29b and 32a. The low quality in the chromatogram of the Guanine in downstream of the deletion site suggests a 6 bp deletion in stock 32a since 32a1 also shows a bp deletion.

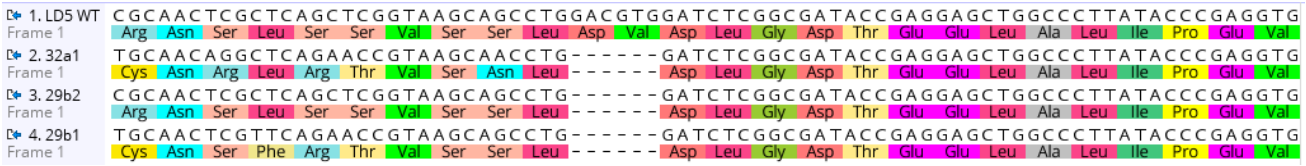

Figure11: Translated amino acid sequences and the corresponding DNA sequences of the wild type of fly and the 6 bp deletion mutation flies 32a1, 29b1 and 29b2. The mutations stayed in the reading frame and a deletion of an aspartic acid and a valine were deleted.

7. Discussion

The aim of this thesis was to obtain *Drosophila melanogaster* fly stocks with a targeted mutation of the *timeless* gene. For this purpose, the CRISPR Cas9 mechanism was used, by crossing a fly expressing a Cas9 gene with a fly that expresses a specifically designed gRNA (Kondo S, 2013).

A gRNA must contain a PAM sequence to be recognized as target for the Cas9. However, it is also important to predict which sites have a structural and functional impact on the translated protein. In this project, 2 gRNA sites were designed.

The first gRNA has a predicted bipartite nuclear localization signal score of 7.3 which indicates a partial transport of the protein into the nucleus through the interaction of the nuclear localization signal with transport receptor importins. A deletion in this nuclear localization signal region can change the cellular localization of the protein and therefore might have an impact on its function (Kosugi S., 2009). Within this region, also 3 possible phosphorylation sites were predicted. The amino acids with the uncharged side chains Serine and threonine, as well as tyrosine, which has a hydrophobic side chain can undergo this kind of modification. When an amino acid undergoes phosphorylation, the resulting phosphoamino acid is negatively charged and hydrophilic. A deletion of a phosphorylation site can therefore have a high impact on the post translational modification of the protein (Hunter, 2012). When compared with other species in a MUSCLE alignment, this region shows to be highly conserved which can be linked to a critical region for the function of the protein (Echave et al. 2016)

In the second designed gRNA presented in this thesis, similarly to the first gRNA, a phosphorylation site was predicted. Interestingly, this TIM region showed a very high variability among compared insect species. This high variability stands out and was therefore also chosen as a site of interest.

Due to time restrictions, the plasmids with our ligated gRNAs were not used for fly transformation. Already transformed flies expressing another gRNA were used to further investigate the practical mechanisms of CRISPR/Cas9 gene editing in *Drosophila melanogaster.*

In figure 7, a heteroduplex mobility screening of the G_0 fly's gonads is depicted. This screening method is used to identify mutations of PCR products. While a non-mutated site will be

present as a single strand, a mutated strand will have bands of several sizes. This is caused by the mismatching of the mutant amplicons with wild type amplicons that leads to the formation of bubbles between the amplicons during PCR. In our case, this was used to identify flies with transformed germline cells (Foster et al. 2019)

The following screening (figure 8) of the resulting F_1 generation was performed to investigate the successful mutation of the different fly stocks. Mosaic patterns of the PCR product were expected in our mutated flies (Mehravar et al 2019). This was verified in 4 of the cultivated stocks. Their offspring was then expected to be homozygous for the investigated gene. This was verified for all the mutated stocks by gel electrophoresis (figure 9).

The sequencing then showed that in 2 of the 4 homozygous F_2 stocks, deletions occurred downstream of the PAM sequence. The cleavage sites lie within the gRNA, therefore a targeted Cas9 cleavage and a resulting deletion occurred (Jinek et al. 2012). In both stocks, a 6 bp deletion at the same site occurred (figure 10) which caused the deletion of an aspartic acid and a valine in the translated protein (figure 11). It must be noted, that in stock 32a, the second sequencing yielded a 5 bp deletion with one more guanine downstream of the deletion site in comparison to the first sequencing. This would mean that a frameshift mutation occurred. However, the low quality of the chromatogram in comparison to the following chromatograms downstream in this measurement and the other sequencing highly suggest a 6 bp deletion.

The deletion occurred in the position 1150-1151 of the amino acid sequence of *Drosophila melanogaster* (NCBI Reference sequence NP_722912.3). Interestingly, phosphorylation sites were predicted for the serine at position 1147 (NetPhos score: 0.979, unspecified kinase) and the threonine at the position 1156 (NetPhos score: 0.559, kinase CK2) using NetPhos 3.1 (Blom et al. 1999).

The deletion in the close vicinity of this phosphorylation site potentially has an impact on the phosphorylation of the TIMELESS protein at this site (Singh et al. 2019). The phosphorylation of TIM is thought to be involved in its degradation (Naidoo et al. 1999; Rothenfluh et al. 2000). Furthermore, the behavioural rhythmicity of a fly has shown to correlate with delay nuclear accumulation of TIM. This accumulation can be delayed by mutations at CK2 mediated phosphorylation sites. The mutation therefore potentially causes a change in behavioural rhythmicity of the mutated flies (Top et al. 2016).

8. Conclusion

The aim of this bachelor project was to learn the basic procedure behind the usage of the prokaryotic immune system CRISPR/Cas9 as a tool for genetic engineering. The key result of this thesis was that a directed target mutation in the form of a deletion could be obtained. For CRISPR/ Cas9 gene editing, established well known lab procedures with high reproducibility need to be followed. In the first part of this project, a gRNA was designed using the readily available online tools NetNES and NetPhos. This gRNA was then successfully molecularly cloned, and the resulting vectors were usable for fly transformations. However, due to time restrictions, already transformed flies with different gRNA vectors were used. These were then crossed to obtain mutated flies at the targeted site of the *timeless* gene which could be verified by sanger sequencing. This project shows that CRISPR/Cas9 gene editing is a readily available and reliable technique in genetic engineering that can be reproduced even by students with no prior experience in this field.

9. References

Bhaya, D., Davison, M., & Barrangou, R. (2011). CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annual review of genetics*, *45*, 273-297.

Blom, N., Gammeltoft, S., & Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. Journal of molecular biology, 294(5), 1351-1362.

Dubowy, C., & Sehgal, A. (2017). Circadian rhythms and sleep in Drosophila melanogaster. Genetics, 205(4), 1373-1397.

Echave, J., Spielman, S. J., & Wilke, C. O. (2016). Causes of evolutionary rate variation among protein sites. Nature Reviews Genetics, 17(2), 109.

Foster, S. D., Glover, S. R., Turner, A. N., Chatti, K., & Challa, A. K. (2019). A mixing heteroduplex mobility assay (mHMA) to genotype homozygous mutants with small indels generated by CRISPR-Cas9 nucleases. MethodsX, 6, 1.

Gleditzsch, D., Pausch, P., Müller-Esparza, H., Özcan, A., Guo, X., Bange, G., & Randau, L. (2019). PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. RNA biology, 16(4), 504-517.

Hardin, P. E., et al. (1990). "Feedback of the Drosophila Period Gene-Product on Circadian Cycling of Its Messenger-Rna Levels." Nature 343(6258): 536-540.

Hunter, T. (2012). Why nature chose phosphate to modify proteins. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1602), 2513-2516.

Jiang, F., & Doudna, J. A. (2017). CRISPR–Cas9 structures and mechanisms. Annual review of biophysics, 46, 505-529.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816-821.

Kloss, B., Price, J. L., Saez, L., Blau, J., Rothenfluh, A., Wesley, C. S., & Young, M. W. (1998). The Drosophila clock gene double-time encodes a protein closely related to human casein kinase Iε. Cell, 94(1), 97-107.

Kondo, S., & Ueda, R. (2013). Highly improved gene targeting by germline-specific Cas9 expression in Drosophila. Genetics, 195(3), 715-721.

Kosugi, S., Hasebe, M., Tomita, M., & Yanagawa, H. (2009). Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proceedings of the National Academy of Sciences, 106(25), 10171-10176.

La Cour, T., Kiemer, L., Mølgaard, A., Gupta, R., Skriver, K., & Brunak, S. (2004). Analysis and prediction of leucine-rich nuclear export signals. Protein Engineering Design and Selection, 17(6), 527-536.

Mehravar, M., Shirazi, A., Nazari, M., & Banan, M. (2019). Mosaicism in CRISPR/Cas9-mediated genome editing. *Developmental biology*, *445*(2), 156-162.

Naidoo, N., Song, W., Hunter-Ensor, M., & Sehgal, A. (1999). A role for the proteasome in the light response of the timeless clock protein. Science, 285(5434), 1737-1741.

New England BioLabs. (2021, February 12). NEBcloner. NEB. http://nebcloner.neb.com/#!/

Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., & Doudna, J. A. (2014). Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. Nature structural & molecular biology, 21(6), 528

Oliver, D., Yuan, S., McSwiggin, H., & Yan, W. (2015). Pervasive genotypic mosaicism in founder mice derived from genome editing through pronuclear injection. PloS one, 10(6), e0129457.

Ota, S., Hisano, Y., Muraki, M., Hoshijima, K., Dahlem, T. J., Grunwald, D. J., ... & Kawahara, A. (2013). Efficient identification of TALEN-mediated genome modifications using heteroduplex mobility assays. Genes to Cells, 18(6), 450-458.

QIAGEN (2001). QIAquick Gel Extraction Kit. https://www.qiagen.com/us/resources/download.aspx?id=1426dbb4-da09-487c-ae01-c587c2be14c3&lang=en

Ren, X., Yang, Z., Xu, J., Sun, J., Mao, D., Hu, Y., ... & Ni, J. Q. (2014). Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in Drosophila. Cell reports, 9(3), 1151-1162.

Roche. (2014, August). *High Pure Plasmid Isolation Kit* (Version 10). Roche Diagnostics GmbH.

Rothenfluh, A., Young, M. W., & Saez, L. (2000). A TIMELESS-independent function for PERIOD proteins in the Drosophila clock. Neuron, 26(2), 505-514.

Singh, S., Giesecke, A., Damulewicz, M., Fexova, S., Mazzotta, G. M., Stanewsky, R., & Dolezel, D. (2019). New Drosophila circadian clock mutants affecting temperature compensation induced by targeted mutagenesis of timeless. Frontiers in physiology, 10, 1442.

Stanewsky, R. (2002). Clock mechanisms in Drosophila. Cell and tissue research, 309(1), 11-26

Terns, M. P., & Terns, R. M. (2011). CRISPR-based adaptive immune systems. Current opinion in microbiology, 14(3), 321-327.

Top-Bio. (2017). *Combi PPP Master Mix*. http://www.top-bio.com/pcr-master-mixes-46.html?combi-ppp-master-mix-1401

Top, D., Harms, E., Syed, S., Adams, E. L., & Saez, L. (2016). GSK-3 and CK2 kinases converge on timeless to regulate the master clock. Cell reports, 16(2), 357-367.

Wiedenheft, B., Sternberg, S. H., & Doudna, J. A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. Nature, 482(7385), 331-338.

Wong, L. C., & Schedl, P. (2006). Dissection of Drosophila ovaries. *Journal of visualized experiments: JoVE*, (1).