

A Reporter System Measuring the Efficiency of DNA Error Repair Mechanisms in Zebrafish

36. OTDK Biológia szekció

Written by:

Dávid Hodunov

Biologist MSc, 3rd grade

Supervisor:

Dr. Máté Varga

Associate Professor

ELTE Department of Genetics

Table of Contents

1. Summary	6
2. Introduction	7
3. Objectives of the Thesis	11
4. Materials and Methods	13
4.1. DNA template amplification and agarose gel electrophoresis	13
4.2. Purification of PCR products	14
4.3. Ligation of PCR amplicons into the cloning vectors	15
4.4. Cell Transformation and colony PCR reactions	16
4.5. Plasmid purification	17
4.6. Plasmid linearization by restriction digestion and the purification of linearized fragments	17
4.7. Ligation of restriction digested DNA fragments	18
5. Results and Discussion	20
5.1. Elongation process to create Sem36	20
5.2. Changing the BGH- to rbgl PolyA, synthesis of Sem36	22
6. Conclusion	26
7. Összefoglaló	27
8. References	28

List of Abbreviations

Amp	Ampicillin
BGH	bovine growth hormone
BGS	Baller-Gerold syndrome
BLM	Bloom syndrome protein
BS	Bloom syndrome
Cas	CRISPR-associated protein
CCMDC	CRISPR/Cas9-mediated DNA cleavage
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
dH ₂ O	distilled water
dpf	day(s) post- fertilization
DSB	DNA double stranded break
DSBR	DNA double stranded break repair
dsDNA	double-stranded DNA
eGFP	enhanced green fluorescent protein
gDNA	genomic DNA
GFP	green fluorescent protein
<i>gol</i>	<i>golden</i>
gRNA	guide RNA
HDR	homology directed repair

HR	homology recombination
IPTG	isopropylthio- β -galactoside
Ku	Ku70-Ku80 heterodimer
LB	Lysogeny broth/Luria-Bertani mediumLB medium
LV	lentivirus
mCherry	monomeric Cherry
mRNA	messenger RNA
mStrawberry	monomeric Strawberry
mutNHEJ	mutagenic non-homologous end joining
NHEJ	non-homologous end joining
PCR	polymerase chain reaction
polyA	polyadenylation signal
RAPA	RAPADILINO syndrome
<i>rbgl</i>	<i>rabbit β-globin</i>
RE	restriction endonuclease
RTS	Rothmund-Thomson syndrome
Sem	Semaphore Reporter system
SFII	(Platinum) Superfi II
sgRNA	single guide RNA
ssDNA	single-stranded DNA
SV40	simian virus 40

T2A	thosea asigna virus 2A
TALEN	transcription activator-like effector nuclease
TLR	TrafficLight Reporter system
WRN	Werner syndrome protein
WS	Werner syndrome
X-Gal	5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside
ZFN	zinc finger nuclease

Summary

DNA repair is a crucial feature of living cells as unrepaired DNA damage would lead to the accumulation of harmful mutations in the genome. It is perhaps unsurprising therefore, that deficiencies of the protein complexes involved in DNA repair usually result in genetic disorders with rather devastating symptoms.

In order to reveal the etiology of such diseases, it is crucial to use model organisms to study them and to combine genetic and biotechnological approaches that can help us to visualize and analyze the impairment of DNA repair.

My work is focused on developing a reporter system to measure DNA double stranded break repair (DSBR) accuracy in living zebrafish embryos. The “Semaphore” reporter consists of a donor construct encoding two fluorescent proteins (eGFP and mStrawberry) and homology arms that target it to the *golden (gol)* locus of zebrafish, encoding a crucial pigmentation gene. Efficient DSB at the *gol* locus is induced using CRISPR, while the outcome of DSBR can be visualized by the relative amount of “green” and “red” foci: homology directed repair (HDR) results in eGFP expression from the *gol* promoter, whereas the non-homologous end joining (NHEJ) events will be often marked by mStrawberry expression.

Introduction

The DNA double-stranded break repair (DSBR) is one of the main DNA repair systems in living cells against the chemical and physical threats that can be detrimental to the stability of the genome (Tatin et al., 2021). Genome maintenance requires multiple sets of complex networks that can focus on fundamental tasks, especially DNA integrity maintenance (Vitor et al., 2020). There are two dominant DSBR pathways that can restore broken ends after DSB, depending on the availability of the homologous DNA repair template. Non-homologous end joining (NHEJ), independent of the repair template, is the direct ligation of two broken ends. Firstly, the Ku70-Ku80 heterodimer (Ku) recognizes the DSB, and recruits various additional proteins including NHEJ polymerase, nuclease and ligase complexes that participate in resecting and adding nucleotides at the break site. However, this pathway can be error-prone because the small insertions, deletions and substitutions due to the lack of the template can cause mutations (Chapman et al., 2012; Lin & Musunuru, 2016; Chang et al., 2017; Vitor et al., 2020). Homology directed repair (HDR) promotes a more precise outcome because it requires the repair template. The template can be dsDNA or ssDNA oligonucleotide (Chapman et al., 2012; Lin & Musunuru, 2016; Vitor et al., 2020).

In cases of DSBs, the five human RecQ helicase orthologs (RECQL1, BLM (Bloom Syndrome protein), WRN (Werner syndrome protein), RECQL4 and RECQL5) play the roles of ‘guardians of the genome’ as they coordinate the repair mechanisms (Croteau et al., 2014; Lu & Davis, 2021). Notably, in the initial steps of homologous recombination (HR), BLM with EXO1 and/or DNA2 participates in the production of 3'-overhangs; BLM can unwind the dsDNA by binding to the 3' ssDNA, and that interaction leads to the endonuclease activity of DNA2 where an extended 3' ssDNA is generated (Nimonkar et al., 2011; Croteau et al., 2014; Lu & Davis, 2021). During NHEJ, the exonuclease activity of WRN is necessary for the resection of DNA ends for XRCC4/LIG4-mediated ligation, which requires an interaction between WRN and XRCC4/LIG4 (Croteau et al., 2014; Lu & Davis, 2021).

A multitude of syndromes have been associated with the impairment of genome maintenance and many of them are results of deficiencies in protein complexes of the DNA repair pathways. Five of these syndromes – Werner (WS), Bloom (BS), Rothmund-Thomson (RTS), RAPADILINO, and Baller-Gerold (BGS) syndromes – are autosomal-recessive disorders and they show devastating clinical features, including spectrum of cancer development, short

lifespan, high risk of diabetes, or reduced fertility in the cases of WS and BS (Siitonen et al., 2009; Cunniff et al., 2017; Oshima et al., 2017; Lu & Davis, 2021).

To investigate the effect of RecQ helicases on the ratio of HDR to NHEJ events, the Fish Genetics Research Group at ELTE initially focused on the implementation of a reporter system that can be applicable *in vivo*, called TrafficLight Reporter system (TLR) developed by Certo et al. (2011). As this experimental tool is the precursor of the Semaphore system, it is necessary to briefly focus on its use by highlighting three of the results of the Certo lab and the intended usage of TLR by the Fish Genetics Research Group.

TLR consists of three sequence blocks fused after a ubiquitously expressing promoter: an in-frame (+1) enhanced Green Fluorescent Protein (*eGFP*) cassette containing an early STOP codon, a T2A self-cleaving peptide sequence and finally the sequence encoding a monomeric red fluorescent protein (mCherry) (Figure 1). Due to the cleavage by targeted nuclease activity at the STOP codon, the repair could occur through HDR or NHEJ. HDR requires the exogenous functional *eGFP* template and produce a functional eGFP protein. Without the template the creation of frameshift indels at the STOP codon leads to a functional mCherry in one third of the cases. Flow cytometry was applied to sort cells expressing eGFP or mCherry.

Certo and colleagues used this system to analyze the repair pathway choice through flow cytometry by integrating a single copy of TLR with a target site for the homing endonuclease I-SceI to the genome of HEK293T cells (HEK293T TLR-Sce cells), and then expressing the I-SceI with or without the donor template. The presence/transduction of both the nuclease and the donor template lentiviruses resulted in mCherry⁺ or eGFP⁺ cells (Certo et al., 2011).

The Fish Genetics Research Group created zebrafish strains carrying the TLR (Certo et al., 2011) in attempt to study the effect of RecQ helicases on genome repair. In this specific case the system consisted of a *ubiquitin* promoter (*ubi*), an in-frame (+1) *eGFP* cassette containing an early STOP codon, a T2A self-cleaving peptide sequence and *mStrawberry* (instead of *mCherry*). The STOP codon in the eGFP coding sequence was targetable also with CRISPR/Cas9 and at the induced DSB the repair would have occurred through HDR or NHEJ (Figure 1). The construct was inserted into the wild-type genome by the transgenesis system based on the Tol2 transposon (Kawakami, 2007)).

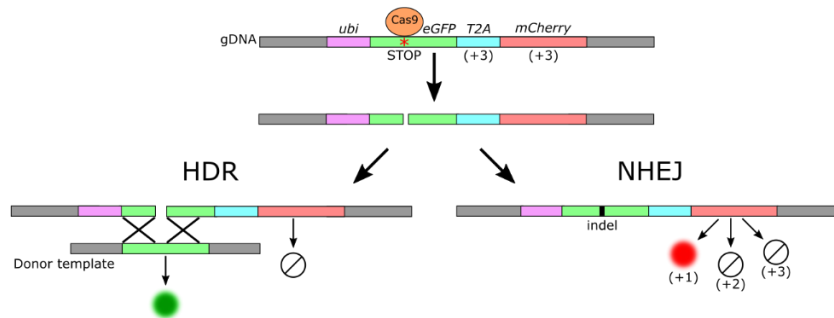
Unfortunately, preliminary tests on the transgenic line utilizing various methods of DNA digestion (CRISPR/Cas9, ZFN and restriction endonuclease cleavage) proved to be

unsuccessful (M. Varga and T. Annus, personal communication) therefore the group opted to pivot towards a novel alternative technique, which we call Semaphore (Figure 1).

Similar to TLR, Semaphore also utilizes two fluorophores in order to detect the ratio of NHEJ and HDR events, however unlike TLR it is plasmid-based. In our construct an *mStrawberry* cassette is followed downstream by *eGFP* which is flanked by sequences homologous to the zebrafish *golden* (*slc24a5*) locus, a region that can be targeted with high efficiency using CRISPR (Jao et al, 2013). Upstream and downstream of the whole sequence are *golden* CRISPR targets, thus when coinjected with a corresponding Cas9 RNP these sites will also be cleaved along with the genomic DNA. Should the occurring DSB be repaired via HDR using the homology arms that flank the *eGFP* cassette, its sequence will be integrated into the genome in frame and under the regulation of the *golden* promoter, green fluorescence will be observable. However, if the entirety of the construct is ligated into the genome via NHEJ, red fluorescence will be observable due to mStrawberry expression in one sixth (accounting for both forward and reverse orientation) of the integration events. In these cases, *eGFP* expression will be inhibited by STOP codons downstream of mStrawberry in all three possible frames (Figure 1).

Two advantages should be highlighted comparing the Semaphore approach to the TLR. One of them is that crossing the strains is unnecessary when conducting experiments as this method is not reliant on stable transgene carrying strains. The other is that the flanking sequences and target sites can be changed to suit the needs of the researchers working on a given project allowing the researchers to easily change and compare different target systems as theoretically any CRISPR-targetable locus within the genome could be studied with Semaphore.

The TrafficLight reporter system



The Semaphore reporter system

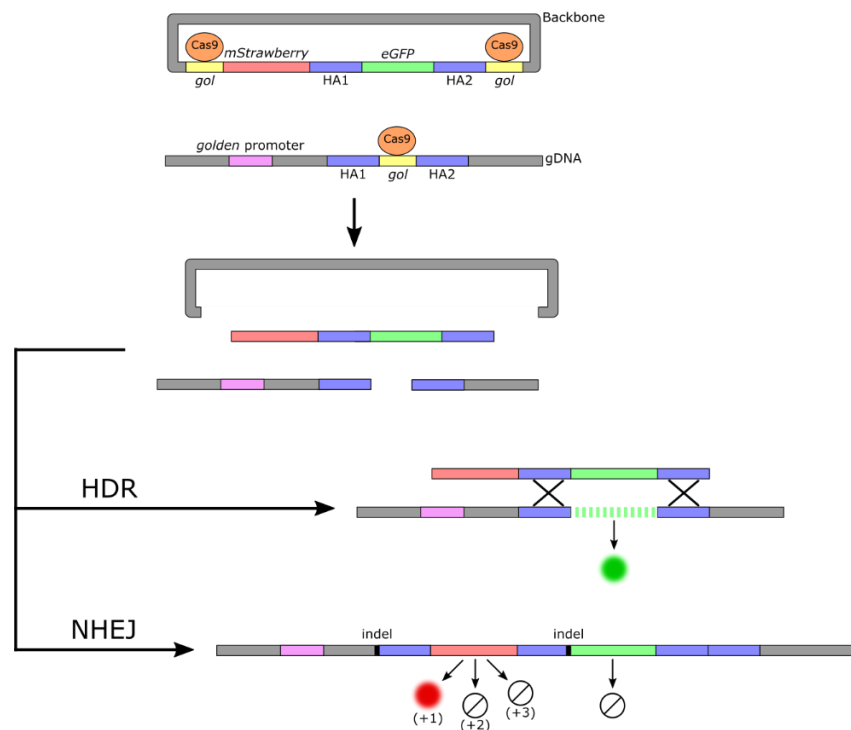


Figure 1. Comparison of the TrafficLight reporter system (top) and the Semaphore reporter system (bottom). TrafficLight is a genomically integrated system regulated by a ubiquitin promoter (ubi). Downstream of ubi the eGFP sequence contains an early STOP codon which hinders its expression and is followed by a self-cleaving peptide sequence (T2A) and a frame shifted mCherry cassette. Upon coinjection with a functional eGFP template and a Cas9 ribonucleoprotein (RNP) targeting the early STOP, repair may occur in two ways. Either through homology directed repair (HDR), whereby the eGFP cassette is restored leading to green fluorescence, or through non-homologous end joining (NHEJ) whereby the eGFP remains dysfunctional, however due to insertions and deletions during the process, in one third of the cases mCherry will be in frame and red fluorescence will be observable. Semaphore is a plasmid-based system, that when coinjected with Cas9 RNP targeting the golden locus (gol) is excised from the plasmid. As the genomic gol locus will also be digested, it may be repaired via HDR utilizing the homology arms (HA1 and HA2), or via NHEJ, thereby integrating eGFP or the whole construct into the genome respectively. As with the TrafficLight system, green fluorescence would indicate HDR, while in one sixth of the NHEJ cases (one third of forwardly oriented integration events) red

fluorescence will be visible. STOP codons in all three frames downstream of mStrawberry would inhibit eGFP expression. (gDNA: genomic DNA.)

Objectives of the thesis

The initially proposed development of the Semaphore system (Figure 3) can be subdivided into four steps as shown in Figure 2.

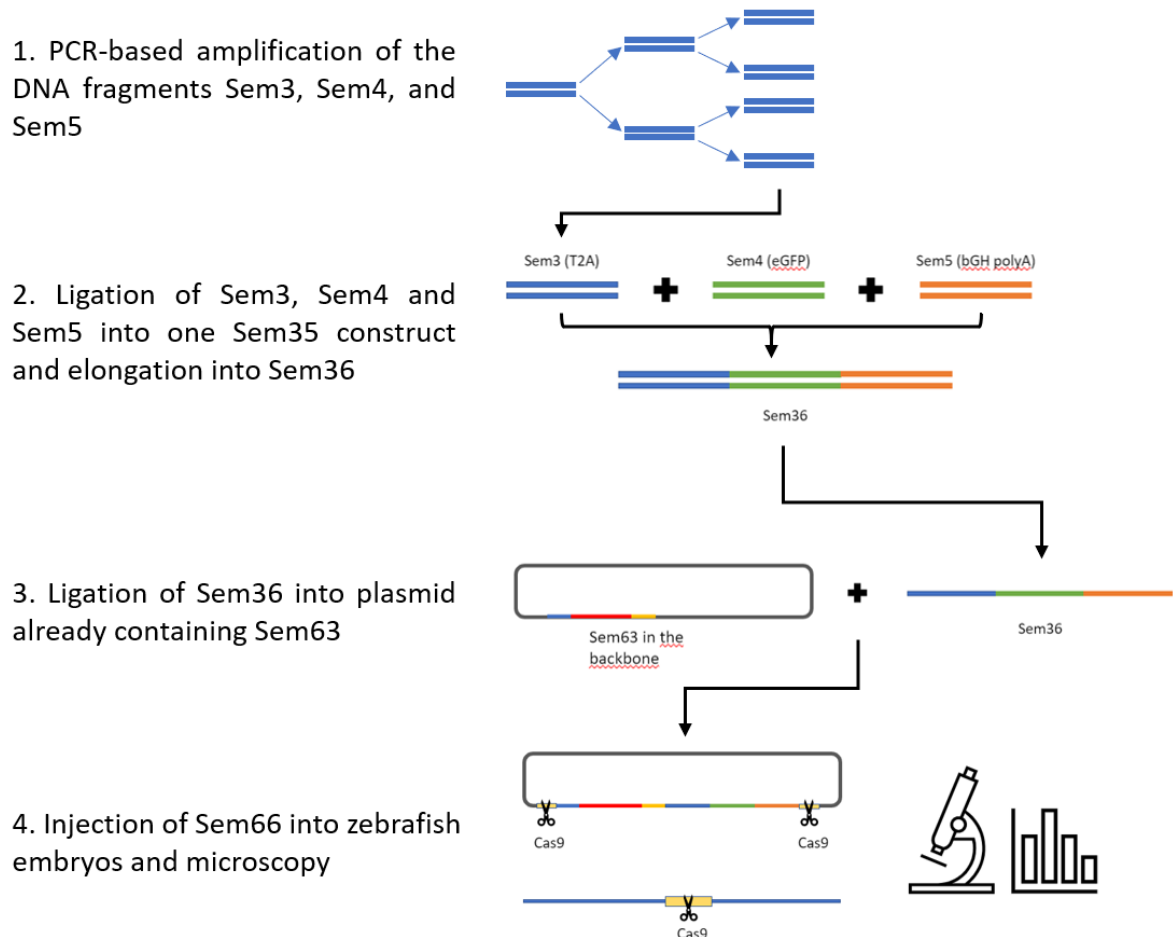


Figure 2. The initial development plan of the Semaphore system.

The upstream half of the construct has already been assembled and ligated into a pBS-KS(+) plasmid before my involvement in the project. During my research, my major aim was to PCR-amplify and clone all the remaining DNA fragments of the reporter and finally to assemble them in the right order. These fragments were numbered according to their positions in the sequence (Figure 3). Double digits in the names of ligated products denote the upstream end (first digit) and the downstream end (second digit) of a given fragment. *Sem66* refers to the whole sequence.

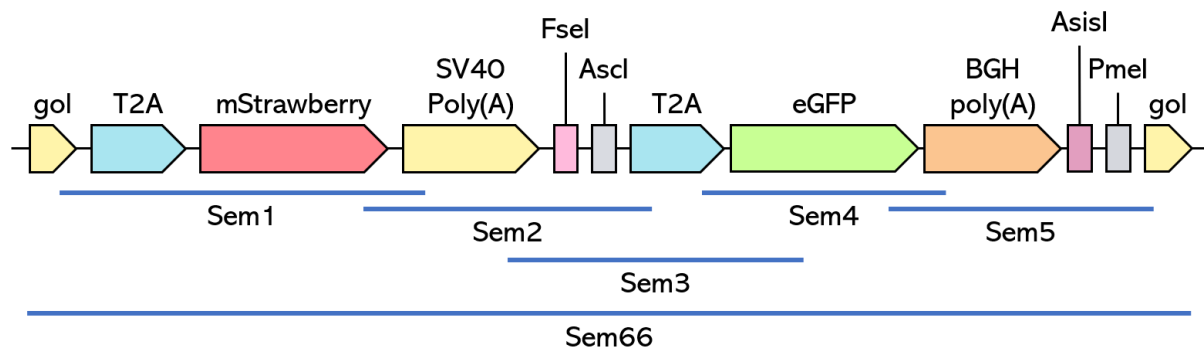


Figure 3. The Semaphore system with its components. Our construct is flanked by CRISPR target sequences of the golden locus (gol), in which the DSBR can be induced by Cas9 ribonucleoprotein (RNP) (Jao et al., 2013). (BGH: bovine growth hormone.)

Materials and Methods

DNA template amplification and agarose gel electrophoresis

Different DNA polymerases – Q5 (NEB, M0491S), DreamTaq (Thermo Scientific, EP0702), Platinum™ SuperFi II (Thermo Scientific, 12361010) and Phusion™ Plus Green PCR Master Mix (Thermo Scientific, F632X) – optimized for polymerase chain reactions (PCRs) were tested during my work in accordance with their respective manufacturer's protocols (Table 1).

PCR reactions were assembled on a Bioer Technology LifeECO Thermal Cycler instrument as follows (Table 2 and Table 3).

Table 1. Materials required for different PCR reactions.

Materials	Q5	DreamTaq	Platinum SuperFi II
	Volumes (μl)		
dH ₂ O	10.75	12.56	12.2
Buffer	5	4	4
Forward primer (10 μM)	1.25	0.4	1
Reverse primer (10 μM)	1.25	0.4	1
dNTP (10 mM)	0.5	0.4	0.4
DNA template	1	1	1
DNA polymerase	0.25	0.24	0.4
Total volume	25	20	20

Table 2. Q5 PCR program in accordance with its manufacturer's protocols.

Q5 PCR program						
	Initial Denaturation	Denaturation	Annealing	Extension	Final Extension	End
Temperature (°C)	98.0	98.0	72.0	72.0	72.0	10.0
Time (mm:ss)	00:30	00:10	00:10	00:30	02:00	∞
35 cycles						

Table 3. Platinum SuperFi II PCR program in accordance with its manufacturer's protocols.

Platinum SuperFi II PCR program						
	Initial Denaturation	Denaturation	Annealing	Extension	Final Extension	End
Temperature (°C)	98.0	98.0	60.0	72.0	72.0	10.0
Time (mm:ss)	00:30	00:10	00:10	00:32	05:00	∞
				35 cycles		

Overall, in our hands, the use of Platinum SuperFi II was the most successful in the template elongation and amplification.

Purification of PCR products

Approximately 5 µl of every PCR reaction was used in agarose gel electrophoresis (90-120 V for 20-30 minutes) using 1% agarose gels.

If a single fragment of the expected size was observed, I purified it using the EZ-10 Spin Column PCR Products Purification Kit (Bio Basic, BS664) following the manufacturer's protocol.

1. I added 5 volumes of the Buffer B3 to the PCR product, and then the solution was transferred to the EZ-10 column.
2. I left it for incubation for 2 minutes and centrifuged at 10 000 rpm for 2 minutes.
3. After centrifugation, I removed the flow-through from the column, 750 µl of Wash Solution was added to the membrane, and the column with the new solution was centrifuged at 10 000 rpm for 2 minutes.
4. Then I repeated this washing procedure.
5. The column was spun at 10 000 rpm for one minute to remove any residual Wash Solution.
6. Then the column was transferred from the collection tube to a new Eppendorf tube and 30 µl of dH₂O was added.
7. After that, the centrifugation occurred at 10 000 rpm for 2 minutes to elute the DNA.
8. The purified DNA was stored at -20°C.

If multiple fragments could be observed after amplification, I ran the leftover of the PCR in a 1% agarose gel and isolated the product with the expected size. In this procedure, I used EZ-10 Spin Column DNA Gel Extraction Kit (Bio Basic, BS654).

1. The DNA fragment was excised from the gel with a clean, sharp scalpel. Then the gel slice was weighed and transferred to a 1.5 ml microcentrifuge tube.
2. I added 400 μ l of Binding Buffer II for each 100 mg of gel weight. Incubation occurred at 50°C-60°C for 10-15 minutes and the mixture was shaken occasionally until agarose was completely dissolved.
3. The mixture was added to the EZ-10 column and let stand for 2 minutes. It was centrifuged at 10 000 rpm for 2 minutes and the flow-through was discarded.
4. 750 μ l of Wash Solution was added, and the centrifugation occurred at 10 000 rpm for one minute. The solution was discarded.
5. I repeated this previous step.
6. I centrifuged at 10 000 rpm for an additional minute to remove any residual Wash Buffer.
7. The column was placed in a clean 1.5 ml microcentrifuge tube. 30 μ l of dH₂O was added to the center of the column and incubation occurred at room temperature for 2 minutes. The sample was centrifuged at 10 000 rpm for 2 minutes to elute DNA.
8. The purified DNA was stored at -20°C.

Instruments for measuring DNA concentration:

- BioTek Epoch Microplate Spectrophotometer
- BioTek Gen5 software

Purified PCR fragments were sent for Sanger sequencing so I could make sure that a.) the right amplicon was created and b.) no novel mutations were introduced during the amplification. Sanger sequencing was performed by EurofinsGenomics Ltd.

Ligation of PCR amplicons into the cloning vectors

PCR products amplified using DreamTaq polymerase (thus having an A overhang) were ligated into the pGEM®-T Easy Vector System (Promega, A1360) in accordance with the manufacturer's protocol.

Blunt ended PCR products of high-fidelity polymerases were ligated into pBS-KS(+) plasmids linearized with EcoRV.

Cell Transformation and colony PCR reactions

Ligated products were heat-shock transformed into *Escherichia coli* DH5 α cells. Samples were spread on LB plates containing 200 $\mu\text{g/ml}$ X-Gal (Thermo Scientific, R0404), 1 mM IPTG (Invitrogen, 15529019) and 100 $\mu\text{g/ml}$ ampicillin (Sigma-Aldrich, A9518), and incubated overnight at 37 °C.

To test if the new colonies contained the expected plasmids, I performed colony PCR reactions (Table 4, Table 5). In brief, I sampled colonies from the agar plate using sterile toothpicks and suspended these samples in 50 μl dH₂O. As template, 1-2 μl of this suspension was used in PCR reactions.

Table 4. Materials for colony PCR.

PCR master mix for Colony PCR (10 reactions)	Volume (μl)
dH ₂ O	106
Buffer	12.5
M13 forward and reverse primers (10 μM)	2.5
dNTP (10 mM)	2.5
DreamTaq polymerase	1.5
Total	125

Table 5. The colony PCR program.

Colony PCR program – DreamTaq Green Master Mix PCR						
	Initial denaturation	Denaturation	Annealing	Extension	Final extension	End
Temperature (°C)	95.0	95.0	68.0	72.0	72.0	10.0
Time (mm:ss)	05:00	00:30	00:30	00:20	05:00	∞
32 cycles						

If amplicons of the right size were observed, tubes of 5 ml of LB medium were inoculated with the respective colony samples and were incubated in a shaking incubator overnight at 220 rpm and 37°C.

Plasmid purification

To purify plasmid DNA from overnight grown bacterial cultures, I used the NucleoSpin Plasmid EasyPure Kit (Macherey-Nagel, 740727.250) according to the manufacturer's protocol.

1. 1.5 ml of the culture was transferred to an Eppendorf tube. Then it was centrifuged at 12 300 x g (12.3 RCF) for a minute – supernatant was discarded. To centrifuge 3 ml of the sample, this step was performed 2 times.
2. 150 µl of Buffer A1 was added to the sample, and was vortexed to resuspend cells completely.
3. 250 µl of Buffer A2 was added to the sample, and was inverted 5 times. It was incubated for 2 minutes at room temperature to lyse cells.
4. 350 µl of Buffer A3 was added, and the sample was inverted until lysate had turned colorless. It was centrifuged for 3 minutes at 12 300 x g to pellet precipitate.
5. NucleoSpin Plasmid EasyPure Columns were put into Collection tubes.
6. The supernatant was loaded onto the spin column. It was centrifuged for 30 s at 2000 x g and the flow-through was discarded.
7. 450 µl of Buffer AQ was added to the spin column. It was centrifuged at 12 300 x g (12.3 RCF) for a minute. The collection tube was discarded.
8. The spin column was put into a new 1.5 ml microcentrifuge tube (Eppendorf tube). 50 µl of dH₂O was added onto the membrane and incubation occurred for 1 minute at room temperature. It was centrifuged for 1 min at 12 300 x g (12.3 RCF).
9. The purified plasmid DNA was stored at -20°C.

Plasmid linearization by restriction digestion and the purification of linearized fragments

Purified plasmids were digested with the respective restriction enzymes according to the manufacturer's protocols. The efficacy of the digestions was tested on 1% agarose gels.

Linearized fragments were purified using standard phenol:chloroform:isoamyl (25:24:1) purification and precipitated with 3 M, pH 5.2 sodium acetate. Samples were dissolved in 30 μ l of H₂O and stored at -20 °C.

1. dH₂O was added to a final volume of 200 μ l.
2. 200 μ l of phenol:chloroform:isoamyl alcohol (25:24:1) was added and vortexed for roughly 20 seconds.
3. It was centrifuged at 12 000 rpm or maximal rpm for 5 minutes.
4. The dH₂O with the DNA as supernatant was transferred into a fresh tube. The amount was approximately 183 μ l.
5. The volume of sodium acetate (3 M, pH 5.2) corresponded to 10% of the pipetted supernatant; the volume of 100% ethanol to the 250% of the supernatant. For example, if the supernatant was approximately 183 μ l, 18.3 μ l of sodium acetate ($183 \mu\text{l} \times 0.1$) and 457.5 μ l of 100% ethanol ($183 \mu\text{l} \times 2.5$) would be added to the supernatant.
6. Incubation occurred for a minimum 1 hour at -80 °C.
7. It was centrifuged at 12 000 rpm at 4 °C for 30 minutes.
8. The supernatant was removed from the pellet. It is important to note that the pellet was not seen, so one needs to know the position of the microfuge tube during the centrifugation and beware of pipetting the supernatant.
9. 150 μ l of 4 °C 70% ethanol was added. It was centrifuged at 12 000 rpm at 4 °C for 2 minutes.
10. I removed the supernatant and let the pellet air dry for 5-10 minutes at RT.
11. The pellet was dissolved in 30 μ l of dH₂O by pipetting up and down.
12. The sample was stored at -20 °C.

Ligation of restriction digested DNA fragments

Ligation reactions were performed using three methods. For the first half of the construct (Sem63) the NEBuilder kit (E2621, New England BioLabs) was tested as a viable option, however in light of successes with more affordable approaches this was eventually abandoned.

As the NEBuilder method necessitated the use of segments with complementary segments with their neighboring sequences and thus we designed our primers with this in mind, we tested an unconventional way of ligation with relative success. This PCR-based approach utilized two template molecules with complementary regions mixed in a molar ratio of 1:1.

PCR was performed without primers for the first 10 cycles, where the two templates essentially acted as primers themselves. Then we added the forward primer for the upstream segment and the reverse primer for the downstream segment in order to allow amplification of the “ligated” product.

Conventional ligation was performed using T4 DNA Ligase (Thermo Scientific, EL0011) according to the manufacturer’s protocol in 1:5 to 1:7 vector:insert ratios and transformed to DH5 α cells.

1. The sample for ligation reaction was prepared according to the protocol of T4 DNA Ligase (Thermo Scientific, EL0011).
2. Then it was incubated at 22°C for 10 minutes.
3. 5 μ l of the sample was transformed to the DH5 α cells.

Results and Discussion

Elongation process to create *Sem36*

At first, I wanted to be sure that I was putting together the right sequences one after the other. Before multiple ligation steps, PCR-based amplification of the DNA fragments (*Sem3* as T2A self-cleaving peptide sequence, *Sem4* as *eGFP*, *Sem5* as *BGH PolyA*) was necessary for the construction of Semaphore. Their gel results and the Sanger sequencing confirmed successful cloning.

After the creation, isolation and purification of the *Sem3*, *Sem4* and *Sem5* sequences, I intended to ligate them in subsequent PCR-amplification steps to create the *Sem36* sequence.

Firstly, I performed a PCR reaction using the overlapping *Sem3* and *Sem4* products as templates. (For these reactions we used mainly Q5 DNA polymerase.) After several optimization steps (e.g. excluding H₂O from the PCR mix to increase the volume of the template) I successfully created and isolated the intermediate *Sem34* product (Figure 4).

Next, I ligated the *Sem5* to *Sem34*. Initially, instead of the expected 1kb size fragment we obtained strong bands at 800 bp and 200-300 bp. After several trouble-shooting steps I decided to gel-isolate a fainter band of the expected size (Figure 4). Sanger sequencing of the purified product confirmed that it contained the expected sequence.

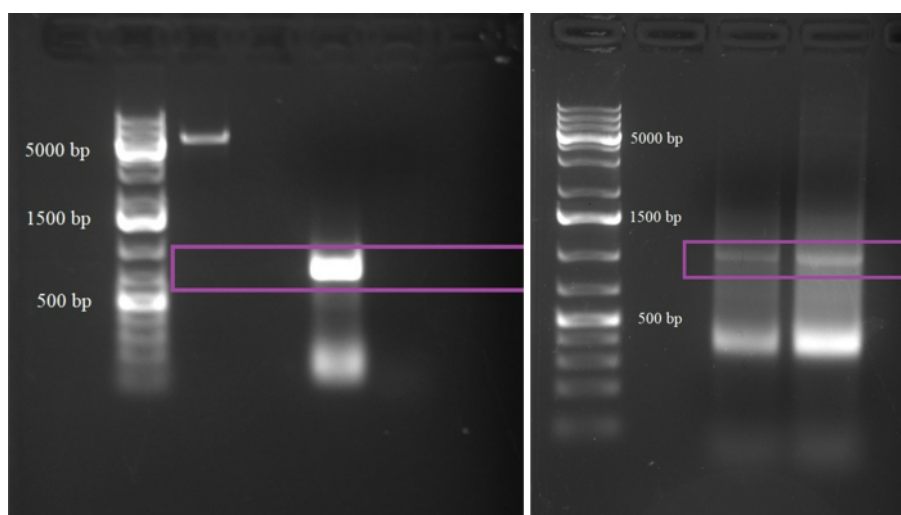


Figure 4. Gel photos of *Sem34* at 800 bp (on the left) and *Sem35* product at 1000 bp (on the right). The ladder in each picture was GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific, SM1333). The length of the desired amplicon is indicated by a rectangle with purple frame. On the left picture we can see that the strong band at 800 bp shows the *Sem34*, so we continued the experiment with PCR purification kit. On the right picture however, two bands were present at 1000 bp and 300-400 bp, so the former ones as *Sem35* were isolated by applying EZ-10 Spin Column DNA Gel Extraction Kit (Bio Basic, BS654).

In order to save the desired insert for further amplification without the need for repeated ligation, I cloned the putative *T2A-eGFP* (*Sem34*) fragment in a pBS-KS(+) vector. The sequencing results confirmed that the amplification of *T2A-eGFP* was successful. As previous ligation efforts were unsuccessful, I set out to amplify *Sem34* with a reverse primer that introduced a ClaI restriction site to allow conventional ligation down the line (Table 6). Similarly, ClaI cleavage site was to be upstream of *BGH polyA* (*Sem5*) with a forward primer (Table 6). I amplified *Sem34* with the mentioned primers and attempted to introduce it into *pSem63*, thus creating *pSem64*, however despite my efforts, the sequencing results indicated a perplexing situation: the *Sem34* sequence was successfully ligated into the plasmid, but with the upstream Semaphore sequences – previously present in the plasmid – curiously missing. Although the reason for that result is still unknown, we hypothesize this novel deletion was due to microhomologies in the cases of the T2A sequences and/or polyA-s, leading to bacterial recombination (Figure 5). In addition, in order to check the reliability of pBS-KS(+), we performed digestion methods with other enzymes, especially XhoI and FastDigest XhoI - the *pSem63* showed bands with different sizes, indicating that the vector for hosting any insert for further amplification processes cannot be trusted as it harbored mutations.

Table 6. Primers for attaching the sequence of *Sem5* to *Sem34*.

Sem4-ClaI-R1	GATCGATTCACTTGTACAGCTCGTCCATGC
Sem5-ClaI-F1	GCATCGATCTGTGCCTTCTAGTTGCCAGC

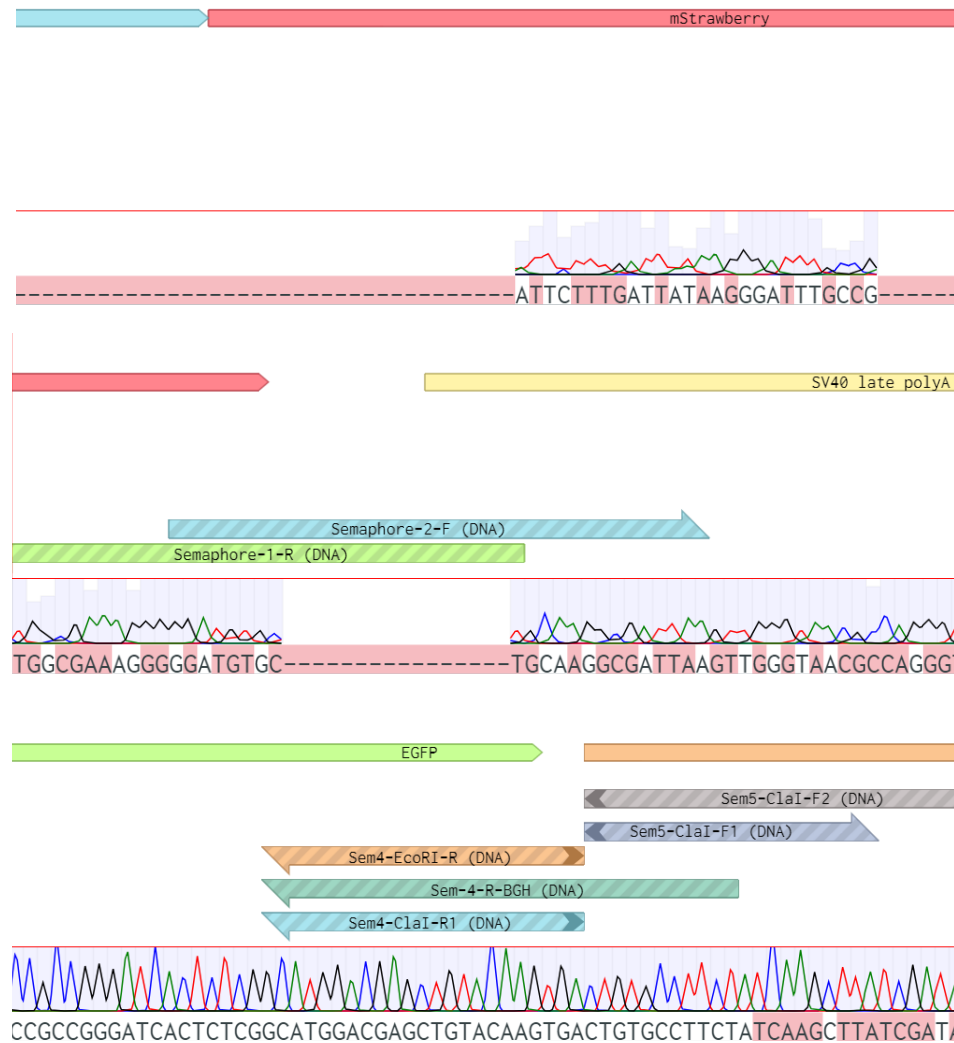


Figure 5. Sanger sequencing result of pSem64. Lack of peaks and low signal strength (short peaks) show that the host vector including the mStrawberry and SV40 PolyA may be lost. However, the ligated sequence with eGFP and BGH PolyA was present by much taller and distinctive peaks.

Changing the BGH- to rbgl PolyA, synthesis of *Sem36*

After several failed ligation attempts, we decided to have the *Sem36* segment commercially synthesized. Although less economical, this approach presented the opportunity to add features to our system which would enhance its utility. During the planning process we also decided to replace the *BGH polyA* with a *rabbit beta globin (rbgl) polyA* (Figure 6). This latter sequence is significantly shorter, but also results in efficient maturation of the transcripts once injected into zebrafish (Addgene.org; Tabor et al., 2018). Additionally, we included restriction endonuclease cleavage sites between the major components of the

construct (Figure 6). The modularity achieved this way will allow for simpler editing in future studies, should the need for such arise.

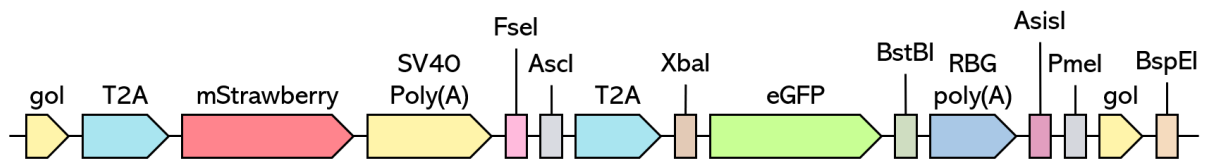


Figure 6. The Semaphore construct with the improved downstream half. (gol: golden locus CRISPR target sequence, RBG: rabbit beta-globin.)

Because the pGEM-T Easy Vector has proven to be reliable and efficient in hosting a given desired insert, the *Sem63* was transferred from the pBS-KS(+)-*Sem63* to pGEM-T Easy Vector (resulting in pGEM-T-*Sem63*). It is important to note that there are thymine overhangs at the pGEM-T Easy Vector, so the transfer of *Sem63* was secured by its PCR-amplification with Phusion Plus Green Master Mix, followed by its ‘DreamTaq-elongation’ with the addition of dATPs (1nM). Its Sanger sequencing confirmed that the *Sem63* was successfully ligated to the pGEM-T (Figure 7).

After linearization of pGEM-T-*Sem63* with *AscI* and *SpeI*, I had the opportunity to excise the newly obtained *Sem36* from the carrier plasmid with the same restriction enzymes and insert it into the cloning vector using T4 DNA ligation, resulting in *pSem66*. At least three of the samples (bands at 2 kb) on the gel photo from colony PCR demonstrated the successful ligation (Figure 7).

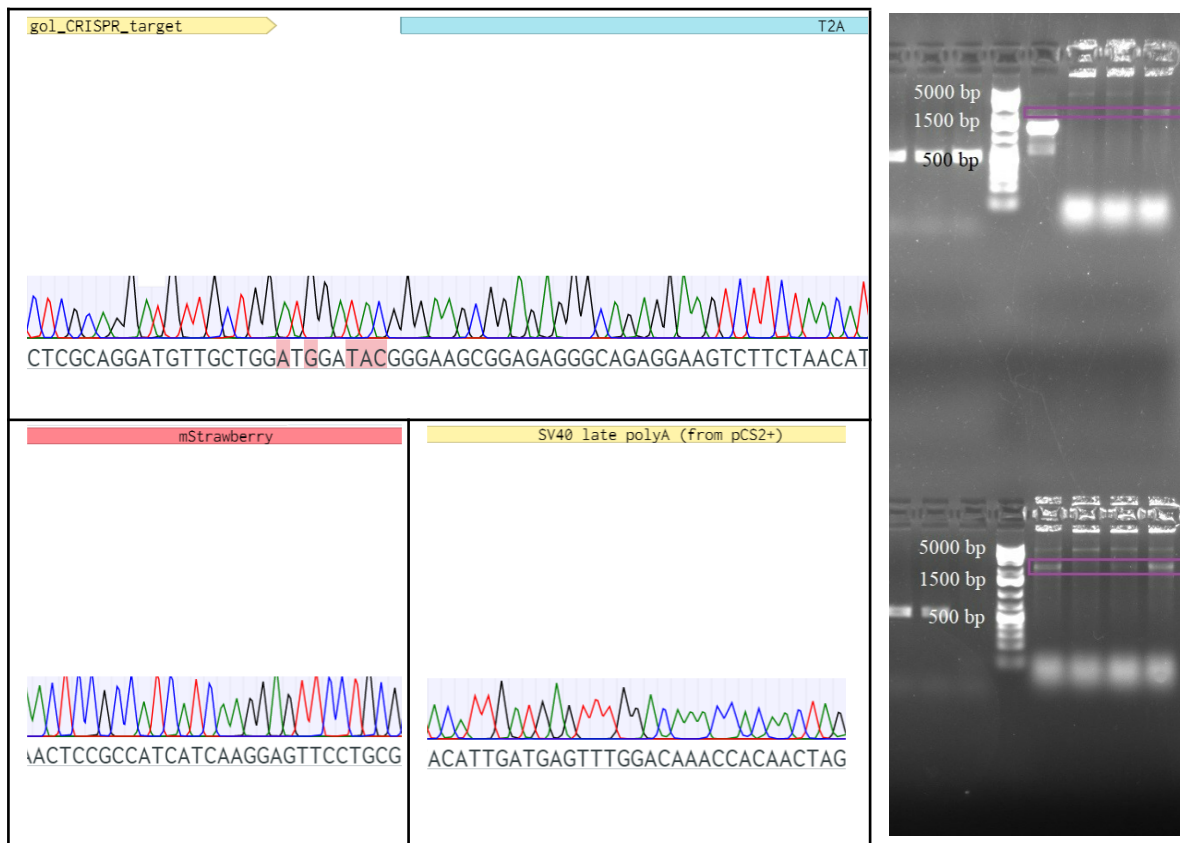


Figure 7. Sanger sequencing result of pGEM-T-Sem63 (on the left) and colony PCR gel photo of pSem66 (on the right). The Sem63 sequence includes the sequences of T2A-self-cleaving peptide, mStrawberry and the SV40 PolyA. At each component we can see well-formed, distinctive peaks, so the ligation procedure was successful. On the right gel picture, the ladders are GeneRuler 1 kb Plus DNA Ladders (Thermo Scientific, SM1333) and the bands at 2000 bp showed the successful T4 DNA ligation of Sem36 to the linearized pGEM-T-Sem63 – these bands are indicated by the rectangles with purple frames.

My future plans include verifying the success of the ligation via Sanger sequencing and integrating the *gol* homology arms into the sequence. Following these preparations injections may begin in order to test the utility of the system. One possible way of validating our approach would be to treat larvae with compounds such as the SCR7 (Srivastava et al., 2012) or RS-1 (Jayathilaka et al., 2008) that are known to shift the HDR:NHEJ ratio and checking if the fluorescent signals match our expectations. I also intend to analyze injected embryos of various genotypes (e.g., wild-types and RecQ mutants) in multiple developmental stages. Eventually we plan to improve the upstream half of the construct as well with additional restriction sites, following which we could check the efficiency of the system using other genomic regions by changing the CRISPR target sequences and we could also seek out potential collaborators utilizing other model organisms as well. It is our hope that this system

will be a valuable and widely applicable tool in future studies investigating genome integrity maintenance.

Conclusion

In my research, I focused on the creation of the Semaphore reporter system that is plasmid-based and transgenic strain independent. These properties indicate the novelty of our research in the Fish Genetics Research Group at ELTE. Although the whole construct is finally inserted into a reliable plasmid pGEM-T (pGEM-T-*Sem66*, simply *pSem66*), its development and optimization took more time and alternative solutions than we expected.

Firstly, I succeeded in the PCR-based amplification of DNA fragments *Sem3*, *Sem4* and *Sem5* that are required to build the remaining second half of the whole construct. However, their integration into one construct in order to elongate into *Sem36* required optimization and troubleshooting steps. It was especially true to the integration of *Sem5* to *Sem34* because I obtained strong bands at different sizes at each ‘PCR-attempt’, so the given band of 1 kb had to be gel-isolated (Figure 4). Although saving *Sem34* for further amplification was successful by cloning it in a pBS-KS(+) vector, its ligation into another pBS-KS(+) plasmid containing the *Sem63* resulted in an unexpected situation: the upstream Semaphore sequences were missing. How it occurred is unknown, but possible reasons have been raised e.g., microhomologies, or the backbone plasmid was mutated. During the whole Semaphore-creation process, we had to perform changes in design strategy like having the *Sem36* segment commercially synthesized and replacing the *BGH polyA* (*Sem5*) with a *rabbit beta globin (rbgl) polyA*.

It is still important to note that choosing reliable and efficient vectors in hosting a given desired insert must be crucial when we want to expand it with new DNA fragments (e.g., homology arms). After the transfer of the *Sem63* from pBS-KS(+) to pGEM-T Easy Vector for hosting the insert in a safer way and the linearization of pGEM-T-*Sem63*, I ensured and succeeded in the integration of the newly obtained *Sem36* into the vector, resulting in *pSem66*.

Összefoglaló

Kutatómunkám során a Semaphore riporter rendszer létrehozására összpontosítottam, amely plazmid alapú és transzgenikus törzs független. Ezek a tulajdonságok jelzik az ELTE Halgenetikai Kutatócsoportjában végzett kutatásunk újszerűségét. Bár a teljes konstrukciót végül egy megbízható pGEM-T plazmidba (pGEM-T-*Sem66*, egyszerűen *pSem66*) illesztettük be, annak kifejlesztése és optimalizálása a vártnál több időt és alternatív megoldást igényelt.

Először is sikerült PCR segítségével felszaporítani a *Sem3*, *Sem4* és *Sem5* DNS-fragmentumokat, melyek a teljes konstrukció fennmaradó második felének felépítéséhez szükségesek. Azonban ezek integrálása egy közös konstrukcióba, a *Sem36*-ba, optimalizálási és hibakeresési lépéseket igényelt. Különösen igaz volt ez a *Sem5 Sem34*-be történő integrálására, mivel minden egyes PCR-alapú amplifikációnál eltérő méretű, erős sávokat kaptam. Ezért végül a kívánt, 1 kb nagyságú sávot gélből izoláltam (4. ábra). Bár a *Sem34* fragmenst sikeresen klónoztam egy pBS-KS(+) vektorba, amikor az innen restrikciós emésztéssel kivágott fragmenst a *Sem63*-at tartalmazó másik pBS-KS(+) plazmidba próbáltam ligálni, azzal kellett szembesülnöm, hogy a Semaphore konstrukció (korábban már sikeresen klónozott) upstream szakaszai rendre hiányoztak. A jelenség pontos okát nem tudtuk megfejteni, de a felmerült lehetséges okok közül megemlíteném a mikrohomológiát mutató szakaszok közti rekombinációt, vagy a pBS-KS-*Sem63* plazmid mutációját. A teljes Semaphore-alkotási folyamat során változtatásokat kellett végrehajtanunk a tervezési stratégiában, például a *Sem36* szegmenst kereskedelmi úton szintetizáltattuk, és a *BGH polyA* szakaszt (*Sem5*) *rbgl polyA*-val helyettesítettük.

Tapasztalataink arra engednek következtetni, hogy a későbbiek folyamán is nagyon fontos lesz a megfelelő vektorok kiválasztása, amelyek megbízhatóan és hatékonyan képesek befogadni egy-egy adott inszertet. Ezt mindenképpen figyelembe kell venni, amikor az alap plazmidunkat új DNS-fragmentumokkal (pl. homológ karok) kívánjuk bővíteni. A *Sem63* pBS-KS(+)-ból a pGEM-T Easy Vektorba történő átvitele és a pGEM-T-*Sem63* linearizálása után biztosítottam és sikeresen integráltam az újonnan kapott *Sem36*-ot a vektorba, aminek eredményeképpen a *pSem66*-ot kaptam.

References

- Addgene.org, <https://www.addgene.org/108355/>, date of access: 23th of September, 2022.
- Certo, M. T., Ryu, B. Y., Annis, J. E., Garibov, M., Jarjour, J., Rawlings, D. J., & Scharenberg, A. M. (2011). Tracking genome engineering outcome at individual DNA breakpoints. *Nature Methods*, 8(8), 671–676. <https://doi.org/10.1038/nmeth.1648>
- Chang, H., Pannunzio, N. R., Adachi, N., & Lieber, M. R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature reviews. Molecular Cell Biology*, 18(8), 495–506. <https://doi.org/10.1038/nrm.2017.48>
- Chapman, J. R., Taylor, M. R., & Boulton, S. J. (2012). Playing the end game: DNA double-strand break repair pathway choice. *Molecular Cell*, 47(4), 497–510. <https://doi.org/10.1016/j.molcel.2012.07.029>
- Chudakov, D. M., Matz, M. V., Lukyanov, S., & Lukyanov, K. A. (2010). Fluorescent proteins and their applications in imaging living cells and tissues. *Physiological Reviews*, 90(3), 1103–1163. <https://doi.org/10.1152/physrev.00038.2009>
- Croteau, D. L., Popuri, V., Opresko, P. L., & Bohr, V. A. (2014). Human RecQ helicases in DNA repair, recombination, and replication. *Annual Review of Biochemistry*, 83, 519–552. <https://doi.org/10.1146/annurev-biochem-060713-035428>
- Cunniff, C., Bassetti, J. A., & Ellis, N. A. (2017). Bloom's Syndrome: Clinical Spectrum, Molecular Pathogenesis, and Cancer Predisposition. *Molecular Syndromology*, 8(1), 4–23. <https://doi.org/10.1159/000452082>
- Dreyfus, M., & Régnier, P. (2002). The poly(A) tail of mRNAs: bodyguard in eukaryotes, scavenger in bacteria. *Cell*, 111(5), 611–613. [https://doi.org/10.1016/s0092-8674\(02\)01137-6](https://doi.org/10.1016/s0092-8674(02)01137-6)
- Eki, R., She, J., Parlak, M., Benamar, M., Du, K. P., Kumar, P., & Abbas, T. (2020). A robust CRISPR-Cas9-based fluorescent reporter assay for the detection and quantification of DNA double-strand break repair. *Nucleic Acids Research*, 48(21), e126. <https://doi.org/10.1093/nar/gkaa897>
- Goodwin, E. C., & Rottman, F. M. (1992). The 3'-flanking sequence of the bovine growth hormone gene contains novel elements required for efficient and accurate polyadenylation. *The Journal of Biological Chemistry*, 267(23), 16330–16334.

- Hans, H., & Alwine, J. C. (2000). Functionally significant secondary structure of the simian virus 40 late polyadenylation signal. *Molecular and Cellular Biology*, 20(8), 2926–2932. <https://doi.org/10.1128/MCB.20.8.2926-2932.2000>
- Hisano, Y., Sakuma, T., Nakade, S., Ohga, R., Ota, S., Okamoto, H., Yamamoto, T., & Kawahara, A. (2015). Precise in-frame integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. *Scientific Reports*, 5, 8841. <https://doi.org/10.1038/srep08841>
- Irion, U., Krauss, J., & Nüsslein-Volhard, C. (2014). Precise and efficient genome editing in zebrafish using the CRISPR/Cas9 system. *Development (Cambridge, England)*, 141(24), 4827–4830. <https://doi.org/10.1242/dev.115584>
- Jayathilaka, K., Sheridan, S. D., Bold, T. D., Bochenska, K., Logan, H. L., Weichselbaum, R. R., Bishop, D. K., & Connell, P. P. (2008). A chemical compound that stimulates the human homologous recombination protein RAD51. *Proceedings of the National Academy of Sciences of the United States of America*, 105(41), 15848–15853. <https://doi.org/10.1073/pnas.0808046105>
- Kim, J. H., Lee, S. R., Li, L. H., Park, H. J., Park, J. H., Lee, K. Y., Kim, M. K., Shin, B. A., & Choi, S. Y. (2011). High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PloS One*, 6(4), e18556. <https://doi.org/10.1371/journal.pone.0018556>
- Kuhar, R., Gwiazda, K. S., Humbert, O., Mandt, T., Pangallo, J., Brault, M., Khan, I., Maizels, N., Rawlings, D. J., Scharenberg, A. M., & Certo, M. T. (2014). Novel fluorescent genome editing reporters for monitoring DNA repair pathway utilization at endonuclease-induced breaks. *Nucleic Acids Research*, 42(1), e4. <https://doi.org/10.1093/nar/gkt872>
- Lanoix, J., & Acheson, N. H. (1988). A rabbit beta-globin polyadenylation signal directs efficient termination of transcription of polyomavirus DNA. *The EMBO Journal*, 7(8), 2515–2522. <https://doi.org/10.1002/j.1460-2075.1988.tb03099.x>
- Li, M., Zhao, L., Page-McCaw, P. S., & Chen, W. (2016). Zebrafish Genome Engineering Using the CRISPR-Cas9 System. *Trends in Genetics : TIG*, 32(12), 815–827. <https://doi.org/10.1016/j.tig.2016.10.005>

- Li, W., Zhang, Y., Han, B., Li, L., Li, M., Lu, X., Chen, C., Lu, M., Zhang, Y., Jia, X., Zhu, Z., Tong, X., & Zhang, B. (2019). One-step efficient generation of dual-function conditional knockout and geno-tagging alleles in zebrafish. *eLife*, 8, e48081. <https://doi.org/10.7554/eLife.48081>
- Lin, J., & Musunuru, K. (2016). Genome engineering tools for building cellular models of disease. *The FEBS Journal*, 283(17), 3222–3231. <https://doi.org/10.1111/febs.13763>
- Liu, Z., Chen, O., Wall, J., Zheng, M., Zhou, Y., Wang, L., Vaseghi, H. R., Qian, L., & Liu, J. (2017). Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Scientific Reports*, 7(1), 2193. <https://doi.org/10.1038/s41598-017-02460-2>
- Lu, H., & Davis, A. J. (2021). Human RecQ Helicases in DNA Double-Strand Break Repair. *Frontiers in Cell and Developmental Biology*, 9, 640755. <https://doi.org/10.3389/fcell.2021.640755>
- Meeker, N. D., Hutchinson, S. A., Ho, L., & Trede, N. S. (2007). Method for isolation of PCR-ready genomic DNA from zebrafish tissues. *BioTechniques*, 43(5), 610–614. <https://doi.org/10.2144/000112619>
- National Center for Biotechnology Information (2022). PubChem Bioassay Record for AID 1159543, Source: Jackson Lab, Gurdon Institute, University of Cambridge. Retrieved November 7, 2022 from <https://pubchem.ncbi.nlm.nih.gov/bioassay/1159543>.
- Nimonkar, A. V., Genschel, J., Kinoshita, E., Polaczek, P., Campbell, J. L., Wyman, C., Modrich, P., & Kowalczykowski, S. C. (2011). BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. *Genes & development*, 25(4), 350–362. <https://doi.org/10.1101/gad.2003811>
- Oshima, J., Sidorova, J. M., & Monnat, R. J., Jr (2017). Werner syndrome: Clinical features, pathogenesis and potential therapeutic interventions. *Ageing Research Reviews*, 33, 105–114. <https://doi.org/10.1016/j.arr.2016.03.002>
- Ranjha, L., Howard, S. M., & Cejka, P. (2018). Main steps in DNA double-strand break repair: an introduction to homologous recombination and related processes. *Chromosoma*, 127(2), 187–214. <https://doi.org/10.1007/s00412-017-0658-1>

- Roidos, P., Sungalee, S., Benfatto, S., Serçin, Ö., Stütz, A. M., Abdollahi, A., Mauer, J., Zenke, F. T., Korbel, J. O., & Mardin, B. R. (2020). A scalable CRISPR/Cas9-based fluorescent reporter assay to study DNA double-strand break repair choice. *Nature Communications*, 11(1), 4077. <https://doi.org/10.1038/s41467-020-17962-3>
- Siitonen, H. A., Sotkasiira, J., Biervliet, M., Benmansour, A., Capri, Y., Cormier-Daire, V., Crandall, B., Hannula-Jouppi, K., Hennekam, R., Herzog, D., Keymolen, K., Lipsanen-Nyman, M., Miny, P., Plon, S. E., Riedl, S., Sarkar, A., Vargas, F. R., Verloes, A., Wang, L. L., Kääriäinen, H., ... Kestilä, M. (2009). The mutation spectrum in RECQL4 diseases. *European Journal of Human Genetics : EJHG*, 17(2), 151–158. <https://doi.org/10.1038/ejhg.2008.154>
- Srivastava, M., Nambiar, M., Sharma, S., Karki, S. S., Goldsmith, G., Hegde, M., Kumar, S., Pandey, M., Singh, R. K., Ray, P., Natarajan, R., Kelkar, M., De, A., Choudhary, B., & Raghavan, S. C. (2012). An inhibitor of nonhomologous end-joining abrogates double-strand break repair and impedes cancer progression. *Cell*, 151(7), 1474–1487. <https://doi.org/10.1016/j.cell.2012.11.054>
- Tabor, K. M., Smith, T. S., Brown, M., Bergeron, S. A., Briggman, K. L., & Burgess, H. A. (2018). Presynaptic Inhibition Selectively Gates Auditory Transmission to the Brainstem Startle Circuit. *Current Biology : CB*, 28(16), 2527–2535.e8. <https://doi.org/10.1016/j.cub.2018.06.020>
- Tatin, X., Muggiolu, G., Sauvaigo, S., & Breton, J. (2021). Evaluation of DNA double-strand break repair capacity in human cells: Critical overview of current functional methods. *Mutation Research. Reviews in Mutation Research*, 788, 108388. <https://doi.org/10.1016/j.mrrev.2021.108388>
- Tsouroula, K., Furst, A., Rogier, M., Heyer, V., Maglott-Roth, A., Ferrand, A., Reina-San-Martin, B., & Soutoglou, E. (2016). Temporal and Spatial Uncoupling of DNA Double Strand Break Repair Pathways within Mammalian Heterochromatin. *Molecular Cell*, 63(2), 293–305. <https://doi.org/10.1016/j.molcel.2016.06.002>
- van de Kooij, B., Kruswick, A., van Attikum, H., & Yaffe, M. B. (2022). Multi-pathway DNA-repair reporters reveal competition between end-joining, single-strand annealing and

homologous recombination at Cas9-induced DNA double-strand breaks. *Nature Communications*, 13(1), 5295. <https://doi.org/10.1038/s41467-022-32743-w>

Vítor, A. C., Huertas, P., Legube, G., & de Almeida, S. F. (2020). Studying DNA Double-Strand Break Repair: An Ever-Growing Toolbox. *Frontiers in Molecular Biosciences*, 7, 24. <https://doi.org/10.3389/fmolb.2020.00024>

Acknowledgements

I would like to thank my thesis supervisor, Dr. Máté Varga for his professional help during the preparation of my thesis and for the opportunity to prepare my MSc thesis at the Department of Genetics of ELTE.

I am grateful to Tamás Annus, who provided his professional experience and help in the preparation of my thesis.

I would also like to thank Dr. Tibor Vellai, Head of the Department, for giving me the opportunity to work and prepare my thesis at the Department of Genetics.

I would like to thank my family and friends who supported me spiritually despite all the difficulties.