UPPSALA
UNIVERSITET

UPTEC X 16 006
Examensarbete 30 hp
Maj 2016

Development of a CRISPR system for gene knockouts in Diplomonad parasites

Viktor Törnblom

UPPSALA UNIVERSITET

Degree Project in Molecular Biotechnology

Masters Programme in Molecular Biotechnology Engineering,
Uppsala University School of Engineering

UPTEC X 16 006	**Date of issue 2016-05-10**

Author

Viktor Törnblom

Title (English)

Development of a CRISPR system for gene knockouts in Diplomonad parasites

Abstract

This thesis describes the attempt to implement the CRISPR-Cas9 system in the two diplomonad pathogens: *Giardia intestinalis* and *Spironucleus salmonicida*. Two approaches of the implementation were done: first the Cas9 was expressed, followed by expression of the guide RNA. Additionally, to extend the understanding of how small nuclear RNAs are processed in *Giardia intestinalis*, two putative RNA processing enzymes were expressed and localized within the cells. The results show expression of Cas9 in both *G. intestinalis* and *S .salmonicida*. However, the Cas9 does not translocate to the nuclei, which is crucial for a functional gene knockout system. The approach where the guide RNA is expressed show a successful expression. In the experiments with the two putative RNA processing enzymes expression is observed in both cases, where one of the enzymes is localized to the nuclei.

Keywords

Diplomonads, *G. intestinalis*, *S. salmonicida*, CRISPR-Cas9, snRNA processing

Supervisors

Staffan Svärd
Uppsala University

Scientific reviewer

Magnus Lundgren
Uppsala University

Project name	Sponsors
Language **English**	Security
ISSN 1401-2138	Classification
Supplementary bibliographical information	Pages **29**

Biology Education Centre Biomedical Center Husargatan 3, Uppsala
Box 592, S-751 24 Uppsala Tel +46 (0)18 4710000 Fax +46 (0)18 471 4687

Development of a CRISPR system for gene knockouts in Diplomonad parasites

Viktor Törnblom

Populärvetenskaplig sammanfattning

Livet på jorden kan delas in i tre större grupper, domäner, nämligen bakterier, arkéer och eukaryoter. Eukaryota organismer utmärker sig genom att ha en mer komplex inre cellorganisation i jämförelse med de två övriga grupperna, de innehar exempelvis en cellkärna i vilken arvsmassan, DNAt, förvaras. Till den eukaryota domänen tillhör exempelvis djur, växter, svampar och mindre organismer av vilka många är sjukdomsalstrande.

En spännande grupp bland eukaryoter är diplomonaderna. Organismer inom den här gruppen är encelliga och har 4 flagellpar, vilka är piskliknande utskott som de använder sig av för att röra sig framåt. En utmärkande egenskap bland diplomonaderna är att de har två stycken cellkärnor. Många av arterna inom den här gruppen är sjukdomsalstrande, exempel på två av dessa är *Giardia intestinalis* och *Spironucleus salmonicida*.

Giardia intestinalis är en tarmparsit som finns utbredd i stora delar av världen. Personer som smittas med parasiten får i många fall kraftig diarré och magsmärtor. Parasiten sprids via föda och vatten kontaminerad med avföring innehållandes cystformen av *Giardia*. Insjuknande är vanligt i U-länder med bristande hygien. Barn är speciellt utsatta.

Spironucleus salmonicida är en fiskparasit om vilken man inte vet mycket om. Den infekterar bland annat lax och har orsakat stor dödlighet och stora ekonomiska förluster bland laxfarmer i norra Norge. Idag finns inga metoder för att behandla eller förebygga dessa infektioner.

I dagsläget genomförs studier på dessa två parasiter för att bland annat ge en bättre förståelse av deras patologi och även ge en inblick i eukaryoters evolutionära utveckling.

Ett vanligt sätt att studera organismer och geners funktioner är genom så kallade ”gene knockouts” vilket innebär att man stänger av specifika gener för att sedan se hur organismen påverkas av modifikationen. I dagsläget finns ingen effekt teknik för att sådana studier i varken *Giardia intestinalis* eller *Spironucleus salmonicida*.

Ett genediteringsverktyg som nyligen utvecklats och visat sig vara mycket effektivt är CRISPR-Cas9 tekniken. Tekniken har implementerats i flertalet organismer samt många olika mänskliga celltyper i vilka lyckade gene knockouts har genomförts. I detta examensarbete beskrivs tillvägagångs av försök till att även implementera den här tekniken i de två diplomonada parasiterna *Giardia intestinalis* och *Spironucleus salmonicida*.

Examensarbete 30 hp Civilingenjörsprogrammet Molekylär bioteknik

Uppsala universitet, maj 2016

CONTENT

ABBREVIATIONS

ADI	-	arginine deiminase
CAS	-	CRISPR associated (genes)
CRISPR	-	clustered regularly interspaced short palindromic repeats
CRISPRi	-	CRISPR interference
crRNA	-	CRISPR RNA
CWP-1	-	cyst wall protein 1
dCas9	-	dead Cas9
DSB	-	double stranded breaks
G6Piso	-	glucose-6-phosphate isomerase
GUS	-	beta-glucuronidase
HDR	-	homology directed repair
MRO	-	mitochondrial related organelles
NLS	-	nuclear localization signal
NHEJ	-	non-homologous end joining
OCT	-	ornithine carbamoyltransferase
PAM	-	protospacer-adjacent motif
saCas9	-	*Staphylococcus aureus* Cas9
snRNA	-	small nuclear RNA
spCas9	-	*Streptococcus pyogenes* Cas9
sgRNA	-	single guided RNA
TALEN	-	transcription activator like effector nucleases
tracrRNA	-	trans-activating crRNA
VPS	-	variable surface proteins
ZFN	-	zinc-finger nucleases

1 INTRODUCTION

1.1 PROJECT DESCRIPTION

The main focus of this degree project was to implement the CRISPR-Cas9 gene editing system in the two diplomonad species: *Giardia intestinalis* and *Sprironucleus salmonicida*. A functional gene editing system would greatly benefit the research of these organisms since there is currently no effective way to perform gene knockout experiments.

The implementation attempt was conducted in two steps: first by expressing the Cas9 nuclease and determining its localization within the cells, subsequently by expression of the guide RNA and estimating the level of expression.

A properly processed guide RNA is crucial for a functional CRISPR-Cas9 system and in order to extend the understanding of how these small nuclear RNAs are processed in *Giardia intestinalis* two putative RNA processing enzymes were expressed and their localization within the cells were determined.

1.2 DIPLOMONADS

The eukaryotic domain of life contains a great diversity of organisms, from unicellular protists to plants and animals. They differ from the prokaryotes by containing their genetic material within an enclosed nucleus and by having a more complex intracellular organisation and organelle composition (Katz 2012).

One group of organisms that has been of great interest while studying the origin of the eukaryotes is the diplomonad group in the order diplomonadida. This is a group of single cellular flagellated protozoans that is classified within the supergroup Excavata.

Diplomonads are a group of anaerobic or microaerobic protozoans containing species living as free-living cells as well as symbionts, either commensal or parasitic. For example, the pathogen *Giardia intestinalis* infects humans and many other species. A characteristic feature of the diplomonads is that each cell contains two nuclei.

Diplomonads lack several organelles and other characteristic features present in the majority of eukaryotes, like peroxisomes, mitochondria and a typical Golgi apparatus (Ankarklev *et al.* 2010, Thompson & Monis 2012). This led to speculations of whether this group of protists emerged in the beginning of the eukaryotic evolution, therefore diplomonads were for a long time considered biological relics (Ankarklev *et al.* 2010, Birky 2010). However due to advances in molecular phylogeny and the discovery of organelles that were thought to be missing, the idea of diplomonads being remains from the emergence of the eukaryotes is being questioned.

Mitochondrial related organelles (MRO) have been discovered in diplomonad species. Examples of MROs are hydrogenosomes, present among species of the genus *Spironucleus* and *Trepomonas*, and mitosomes which are found in species of the genus *Giardia*. These organelles have lost their original function as ATP producers through oxidative respiration and are thought to be the result of reductive evolution (Jerlström-Hultqvist et al. 2013). Hydrogenosomes produce energy in an anaerobic environment through an alternative pathway generating hydrogen as a biproduct. Mitosomes on the other hand have lost all energy generating traits as well as all genomic material typical for

mitochondria. One feature of the mitosome is the maturation of iron-sulphur protein, which is an important role of the mitochondria as well (Jerlström-Hultqvist *et al.* 2013).

This group of protozoans shows the great diversity of the eukaryotes and studying them will provide a better understanding of the basic fundamentals of eukaryotes. More knowledge of the pathogenicity of the parasitic species could help developing future treatments.

1.2.1 *Giardia*

The most studied group of diplomonad so far is *Giardia*, since it is used as a model organism for this group of protozoans (Svärd *et al.* 2003). *Giardia* is a parasitic organism that causes the disease *Giardia*sis and it infects a diversity of animals; like mammals, birds and reptiles. The transmission of *Giardia* occurs through contaminated food and water. Moreover human infections occur in both developed and in developing countries.

The symptoms of *Giardia*sis may vary from non-symptomatic to severe diarrhea associated with abdominal pain and nausea. In most cases the infections are self-limiting, but chronic infections do occur (DuPont 2013). The prevalence of *Giardia*sis is very high in many developing countries and young children are the most affected. Longer intestinal malabsorption because of re-infections and chronic infections may result in growth or cognitive retardation among these children (Halliez & Buret 2013).

The *Giardia* cell has a droplet shaped body consisting of two nuclei and four pairs of flagella (Fig. 1). On the ventral side of the cell is an attachment organelle located, this organelle is called the ventral disc. The median body is an organelle located in the centre of the cell. The function of the median body is not fully understood, but when stained it appears like a smile on the cell.

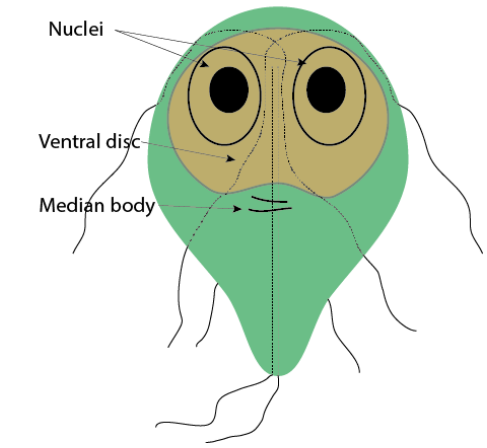


Figure 1. The morphology of the *Giardia* trophozoite. The figure shows a *Giardia* trophozoitic cell and the distribution of the two nuclei and organelles. The ventral adhesive disc, used for attachment to epithelial cells, are located on the ventral side of the cells. The distribution of the four flagella pairs and the location of the median body can also be observed.

Several morphological different species of *Giardia* have been found infecting a variety of animal species. One of these is *Giardia intestinalis* which in turn has been divided into eight morphological indistinguishable genotypes or assemblages (A-H). Two of the assemblages, A and B, have been found infecting humans. Large genomic differences between the assemblages suggest that these should be considered different species (Jerlström-Hultqvist *et al*. 2010)

The cell cycle of *Giardia* consists of two main stages: the vegetative stage as trophozoites and the infectious stage as cysts. The cysts are highly infectious and only between 10 – 25 cysts are required for the initiation of an infection (Thompson & Monis 2012). After cysts have been ingested, the excystation process starts in the stomach; the low pH is believed to trigger this process (Ankarklev *et*

al. 2010). During the excystation the cysts differentiate into trophozoites through a short lived stage called excyzoitic stage. The excyzoite undergoes cytokinesis two times, resulting in four trophozitic cells. The trophozoites enter the small intestines where colonization of duodenum and jejunum is initiated. The attachment of the trophozoites to the intestinal mucosa is mediated by an attachment organelle called the ventral adhesive disc. *Giardia* reproduces asexually through binary fission, however there are indications of sexual or parasexual recombination events (Birky 2010). An infection may last for some days to several months and cysts start to appear in the feces after 3 days to 3 months after infection depending on the host organism (Thompson & Monis 2012). The encystation occurs in the posterior small intestines. Factors inducing the encystation in vivo are not fully understood, but the presence of bile salts and cholesterol might be important factors. After excretion the cysts remain infectious, if the conditions are right, for approximately two months. Trophozoites excreted will not survive outside the host (Thompson & Monis 2012).

There are very few virulence factors known and it is not fully understood how *Giardia* avoids detection of the adaptive immune system of the host. One way of avoiding detection is believed to be through antigenic variation. The surface of the trophozoites is covered with variable surface proteins (VPS) which are proteins that alter periodically after a certain numbers of generations, thus avoiding recognition of antibodies produced by the host (Ankarklev *et al.* 2010).

1.2.2 Spironuclues

Another genus of diplomonads is *Spironucleus*. This group of organisms also contains several parasitic species, as well as some commensal living species. Most of the species are infecting fish, while some are found causing infections in birds and mice (Andresson *et al.* 2007).

The specific species of interest for this project is *Spironucleus salmonicida*. This parasite has been found causing severe systemic infections in Atlantic salmon, Chinook salmon and Arctic char with high mortality rate. *S. salmonicida* is thus of great concern within aquaculture industry. There are examples in northern Norway where outbreaks in fish farms caused mass mortality followed by great economic loss (Jørgensen *et al. 2011)*. So far there is no drug treatment available (Xu *et al.* 2014).

At the moment there is barely anything known about how transmission occurs and other virulence mechanisms of *Spironucleus*.

Recent advances in genomics have shed some light on the life style of this mysterious parasite. In a study performed by Xu *et al.* (2014) the genome of *S. salmonicida* was compared to the better studied *Giardia intestinalis*. The study shows that *S. salmonicida* possesses more genes encoding enzymes involved in metabolic processes as well as genes involved in handling oxidative-stress compared to *Giardia*. This indicates that *S. salmonicida* is more capable of handling environmental fluctuations. Homologous genes to cyst wall proteins were also found, this indicates that transmission between hosts occurs through encystated cells, as in *Giardia*.

1.3 CRISPR-CAS9

1.3.1 History of eukaryotic gene editing

During the 1970s the first endogenous modifications were performed by homologous recombination technology. The first knock out and gene exchange experiments became possible. However these

recombination events occur very infrequently and screening assays are time consuming (Hsu *et al.* 2014).

It was later discovered that the frequency of recombination events could be increased by creating double stranded breaks (DSB) at the target sequence and at the same time providing a DNA sequence homologous to this sequence. This technique utilizes the cell's own DNA repair system. There are two main pathways for cellular DNA repair: homology directed repair (HDR) and non-homologous end joining (NHEJ) (Burma *et al. 2006)* .These two pathways can be used for different gene editing objectives. The HDR requires a homologous repair template in order to repair a DSB. By supplying a designed repair template this pathway can be used for incorporation of new genes into the genome. The NHEJ on the other hand does not require any repair template, but is very error prone and often results in insertion or deletion mutations (indels). If the DSB occurs within a gene an indel will cause a frame shift consequently resulting in a non-functional gene. A gene directed DSB together with the NHEJ-pathway could therefore be used in gene knockout experiments (Hsu *et al.* 2014).

There are several techniques developed for inducing sequence specific DSBs. Two of the most used ones are zinc-finger nucleases (ZFNs) and transcription activator like effector nucleases (TALENS), and now also the recently discovered CRISPR-Cas9 system. ZFN and TALEN are engineered nucleases, composed of two domains: a DNA binding domain and a non-sequence specific DNA cleavage domain. The DNA binding domains are composed of series of sequence binding proteins, either zinc-finger proteins or TALE proteins, and by arranging these proteins in a certain sequence the nuclease can be programmed to target the DNA in a sequence specific manner (Gaj *et al.* 2013). These two techniques have shown to be powerful gene editing tools, however they have several drawbacks (Gupta & Musunuru 2014). The design and production of the DNA binding domains are time consuming and expensive and there are difficulties for non-specialists to implement the techniques. The fact that the nucleases bind through protein-DNA interaction makes the specificity and off-target cleavage an issue. The CRISPR-Cas9 system on the other hand binds to the target sequence through Watson-Crick base pairing, increasing the specificity compared to the other techniques. This system is composed of the Cas9 nuclease which is guided to its target sequence by a designed guideRNA. The guideRNA is the only part that needs to be redesigned when changing the target gene and this makes the CRISPR-Cas9 system a faster, easier and less expensive alternative to ZFN and TALEN. It also enables multiplexing by introducing several different guideRNAs at the same time. (Gupta & Musunuru 2014, Hsu et al. 2014).

1.3.2 Biological function of CRISPR

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) is naturally a part of a prokaryotic adaptive immune system protecting the cells from viral infections.

The CRISPR locus is organized in a very characteristic way. It is composed of an array of repetitive palindromic sequences, repeats, interspaced with so called spacer sequences. The spacer sequences are derived from viral DNA, so called protospacers, and are incorporated into the CRISPR array after previous viral infections. Adjacent to the CRISPR array are the CRISPR associated (Cas) genes located (Fig. 2). This is a cluster of genes encoding for proteins involved in the processing and maintenance machinery of the CRISPR-Cas immune system (Hsu *et.al* 2014).

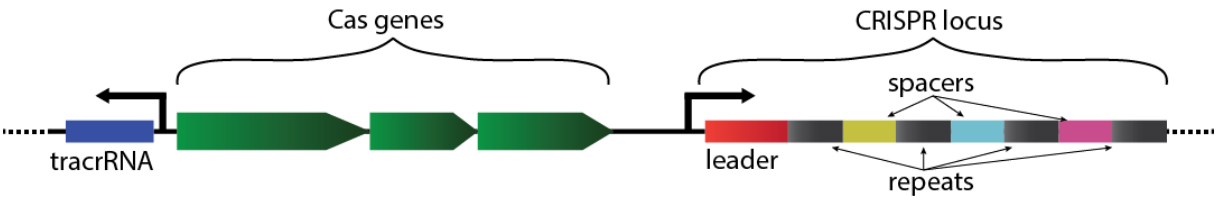


Figure 2. The type II CRISPR locus and adjacent genes. The CRISPR locus is composed of spacers and repeats. The whole locus is transcribed as one segment and is regulated by one promoter and a leader strand located at the 5'-end of the locus. Upstream the CRISPR locus are the Cas genes located, one of them encodes the Cas9 enzyme. Adjacent to the Cas gens are the gene encoding the tracrRNA, which together with the crRNA (encoded by the spacers) binds to the Cas9 enzyme.

The composition of the Cas proteins varies between different species and at the moment three distinct variants are known: type I, type II and type III. These variants are distinguished in the way the CRISPR-array transcript is processed and in the number of proteins required for the DNA interference during a viral infection. The type I and type III systems require a complex of several Cas proteins for DNA interference, in contrast to the type II system that only requires one, Cas9 (Hsu *et.al* 2014). The bacteria *Streptococcus pyogenes* maintains the type II system and is the organism from which the Cas9 as a gene editing tool is derived.

In order for the Cas9 enzyme to target and bind to a foreign piece of viral DNA three parts are required: the Cas9 protein, CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA is a 20 nucleotide long RNA sequence processed and derived from the spacer sequences of the CRISPR-array transcript. The crRNA is complementary to the target DNA. The tracrRNA facilitates the processing of the crRNA and together they form a RNA duplex, tracrRNA:crRNA. The structure of this duplex enables the binding to the Cas9 enzyme. In order to facilitate the use of Cas9 as a gene editing tool the tracrRNA:crRNA duplex has been engineered as a single guide RNA (sgRNA) (Doudna *et al*, 2014).

A crucial feature of the CRISPR-Cas system to avoid self-targeting of the spacers within the organisms owns genome is the protospacer-adjacent motif (PAM). The PAM is a short nucleotide sequence located upstream the target sequence (Fig. 3) and is necessary for initial binding of the Cas9. This PAM sequence is not present within the organism's own genome and hence the self-targeting is prevented. The PAM sequence recognized by the Cas9 enzyme varies depending from which organism the enzyme originates. For example the PAM recognized by the Cas9 originated from *Streptococcus pyogenes* is 5'-NGG. This has to be taken into consideration when designing the sgRNA which has to target a sequence adjacent to a PAM sequence (Ran *et al.* 2013).

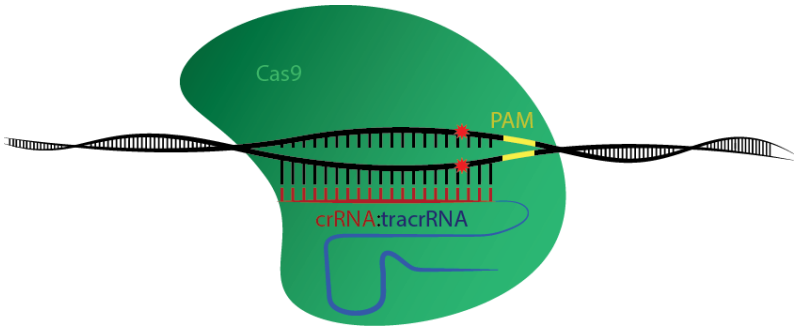


Figure 3. Cas9 DNA interference. The Cas9 enzyme is guided to the target sequence, resulting in a double stranded break (red stars). The crRNA:tracrRNA RNA complex guides the Cas9 nuclease to the target, which is located next to the PAM sequence.

1.3.3 Implementations of Cas9 as a gene editing tool

As mentioned earlier the CRISPR-Cas9 system has been implemented as a tool for editing and replacing genomic DNA sequences. Since the Cas9 enzyme was discovered it has been modified and optimized in several different ways. Codon optimization for usage in different organisms has been performed as well as fusion with nuclear localization signals (NLSs) enabling translocation to the nuclei in eukaryotes (Terns 2014).

The Cas9 enzyme contains two nuclease domains (HNH and RuvC-like domains) that are responsible for cutting either one of the strands of the targeted DNA sequence, creating the DSB. By mutating one of the domains, inactivating its nuclease activity, Cas9 enzymes performing single strand cuts, or nicks, has been created (Doudna & Charpentier 2014). Using this nicking technique has shown to favour the HDR-pathway compared to the NHEJ repair pathway, hence useful when performing gene insertion experiments (Terns 2014).

By mutating both the nuclease domains, Cas9 enzymes with no nuclease activity have been created, also called dead cas9 or dCas9. These modified enzymes have shown to be useful in many applications. The so called CRISPR interference (CRISPRi) platform has been used for knocking down gene expression by blocking transcription initiation or elongation (Terns 2014). dCas9 fused with transcription factors can be used to increase gene expression by guiding the transcription factors to promoters or enhancers of specific genes (Perez-Pinera *et al.* 2013). Epigenetic regulation is another interesting application of the guidable dCas9. The regulation is done by fusing dCas9 to epigenetic effector domains, which conducts chromatin remodelling through DNA methylation or histone modifications (Terns 2014, Yao *et al.* 2015).

One possible implementations of the CRISPR-Cas9 system in the future is as a tool in gene therapy, a tool to modify the human genome in order to prevent and treat diseases. The most promising way to deliver the Cas9 and guide RNA to the target cells is by using viral vectors and the adeno-associated virus (AVV) is a possible candidate for this task (Gori *et al* 2015). One limitation of the *Streptococcus pyogenes* Cas9 (spCas9) is its gene size, ~4.2kb, since the AVV has a restricted cargo size of ~4.5kb. When including the guide RNA coding cassette there is little room for additional regulatory elements (Ran *et al* 2015). In order to tackle this problem smaller orthologues to spCas9 have been studied. The gene editing efficiency of *Staphylococcus aureus* Cas9 (saCas9), with a 3.2kb gene, has shown to be similar to the efficiency of spCas9 (Ran *et al* 2015), hence might be more suitable when the gene

size is an issue. The saCas9 was also used in this project. The smaller sized enzyme could possibly facilitate the translocation of the translated protein to the nuclei of *Giardia* and *Spironucleus*.

1.3.4 Potential targets for Cas9 in *Giardia*

In this project sgRNAs were designed to target two different genes in *Giardia*: the gusA gene encoding β – glucuronidase (GusA), which is a part of the GUS reporter system, and the gene encoding arginine deiminase (ADI), which is a potential virulence factor in *Giardia*.

The GusA protein catalyses the cleavage of a variety of β – glucuronides and it is naturally present in *E. coli* (Jefferson 1989) This protein has been implemented as a reporter system in several different organisms, including *Giardia* (Müller *et al.* 2009). The most common substrate used in GusA assays is the X-Gluc substrate (5-bromo-4-chloro-3-indolyl glucuronide) which makes the cells expressing GusA to turn blue *(*Jefferson 1989). A substrate previously used in *Giardia* was p-nitrophenyl-glucuronide, and the GusA activity was measured spectrophotometrically (Müller *et al.* 2009). After introducing the *gusA* gene to the *Giardia* genome a successful gene knockout, using Cas9, would work as a proof of concept.

The ADI enzyme plays several important roles in metabolic processes as well as in mechanisms protecting the trophozoites from the host immune system (Ankarlev *et al.* 2010, Touz *et al.* 2010). L-arginine serves as a main energy source in *Giardia* during anaerobic conditions and ADI plays a crucial role in the metabolic process. Released ADI also reduces the L-arginine in the surroundings of the cell, preventing the production of nitric oxide, NO. NO is produced by the host cells and acts as an antimicrobial agent in the innate immune system of the host (Stadelmann *et al.* 2012). ADI has also shown to be a part in the regulation of the antigenic variation of the trophozoites (Touz *et al.* 2010).

1.4 SMALL NUCLEAR RNA PROCESSING IN *GIARDIA*

The most common way of expressing the non-protein coding sgRNA gene is usually by using the RNA polymerase III promoter for the U6 spliceosome RNA gene (Jacobs *et al.* 2014). The U6 promoter is also utilized in this project for sgRNA expression in *Giardia*. To get a functional sgRNA it is crucial to get a well-defined 5’ and 3’-end of the transcribed sgRNA, a proper processing is therefore of importance.

The U6 spliceosome RNA is involved in pre-mRNA processing and it is an example of a small nuclear RNA (snRNA) (Marz *et al.* 2008). The processing and transcription of the snRNA genes vary compared to protein coding genes. These genes lack the typical promoter, they do not contain any introns, they do not undergo polyadenylation and they lack an open reading frame (ORF). Upstream the snRNA genes is a promoter composed of a distal sequence element (DSE) and a proximal sequence element (PSE) and downstream these genes is located a 3’-end box, which determines the formation of the 3’-end (Chen and Wagner 2010).

It is not fully understood how the expression and processing of the snRNAs are attained (Chen and Wagner 2010). In this project two potential snRNA processing enzymes were expressed and their localization within the cell was determined. If the enzymes were localized to the nucleus this would strengthen the possibility of the proteins being involved in the processing of the sgRNAs and snRNAs in *Giardia*.

2 RESULT AND DISCUSSION

The attempt to implement the CRISPR-Cas9 system in two diplomonad species, *Giardia intestinalis* and *Spironucleus salmonicida*, was done using two main approaches, first by expression of the Cas9 enzyme in trophozoites and secondly by expression of the guideRNA. These are the two component necessary for the CRISPR-Cas9 system and for an optimal and stable gene knockout system both of the parts would be expressed by the cells, preferably with the genes integrated into the chromosomes. However if any of the two parts are stably expressed the additional part could possibly be prepared *in vitro* and transfected into the nuclei.

In order to further investigate the mechanism behind the processing of the guideRNA in *Giardia*, two putative RNA processing enzymes were expressed and their localization within the cell was determined.

2.1 EXPRESSION OF CAS9 IN *GIARDIA* AND *SPIRONUCLEUS*

In the first approach, where the Cas9 enzyme were episomally expressed, different levels of expressions could be observed depending on the promoter regulating the gene (Fig. 4 and 5). However, in order to utilize the Cas9 it is crucial that the enzyme is successfully translocated to the nuclei. No distinct localization to the nuclei could be observed in any of the transfectants. The enzyme seems to be equally distributed in the cytoplasm of the trophozites. The same expression pattern is observed in both *Giardia* and *Spironucleus.*

For expression in *Giardia* three different promoters were tested, Cyst Wall Protein 1 promoter (CWP-1p), Glucose-6-phosphate isomerase promoter (G6Piso-p) and the ornithine carbamoyltransferase promoter (OCTp). Different levels of expression were estimated by observing the intensity of the fluorescence between the isolates. The strongest expression was observed from the inducible CWP-1p after induction for 2 hours (Fig. 4, M). Expression was also observed in the control samples where there was no induction of the promoter, though to a lower extent and probably due to leakage of the promoter. (Fig. 4, E) Of the three promoters the weakest expression of Cas9 was observed using the G6Piso-p (Fig. 4, Q) while an intermediate expression was observed using the OCTp (Fig. 4, A).

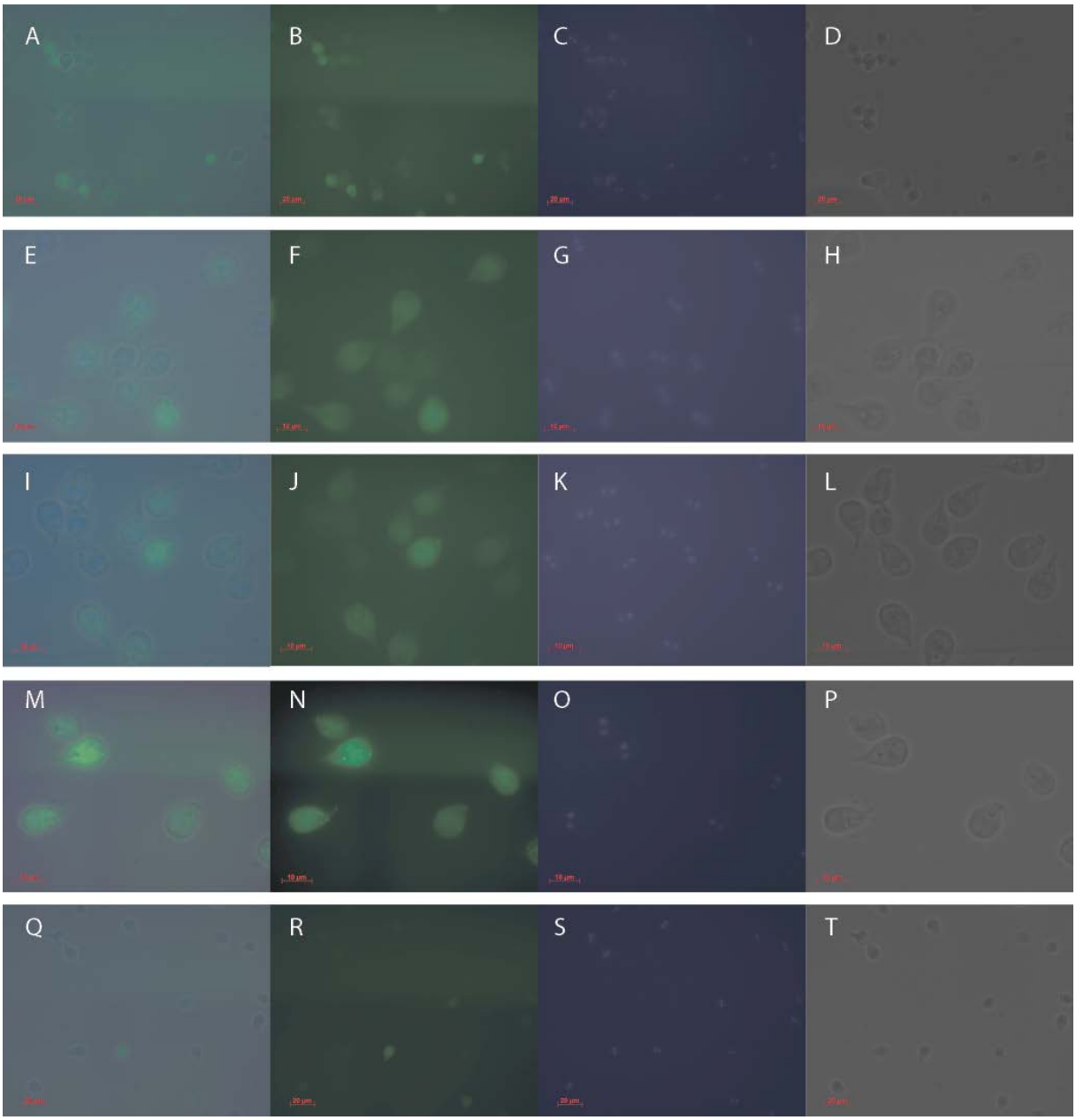


Figure 4. Expression and localization of Cas9, using different promoters, in *Giardia intestinalis*. *Giardia* trophozites were fixed and permeabilized on microscope slides and stained with anti-HA mouse antibodies conjugated to Alexa Flour 488. Three different promoters were tested: OCTp (A-D), CWP-1p (E-P) and G6Piso-p (Q-T). The CWP-1 promoter is induced through encystation and were incubated in encystation media for different time periods: 0 hours (control) (E-H), 1 hour (I-L) and 2 hours (M-P). Images displays: the phase contrast (D,H,L,P,T), DAPI staining (C,G,K,O,S), anti-HA antibody staining (B,F,J,N,R) and a fused image (A,E,I,M,Q).

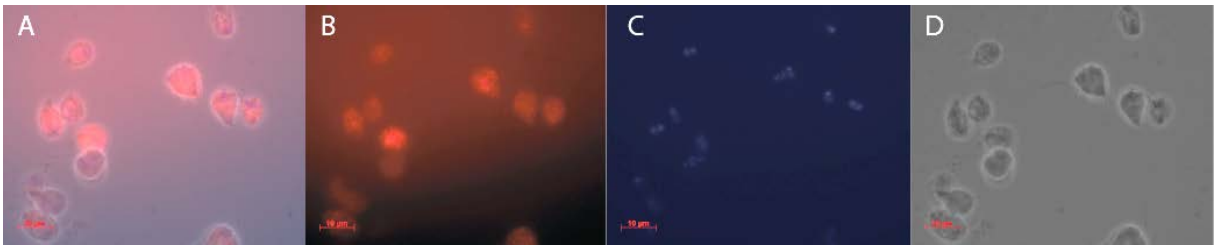


Figure 5. Expression and localization of spCas9 in *Spironucleus salmonicida*. The trophozites were fixed and permeabilized on microscope slides, followed by staining with a primary CRISPR/Cas9 mouse antibody and a subsequent staining with a secondary goat anti-mouse antibody conjugated to Alexa Fluor 594. The images displays a fused image (A), the Cas9- Alexa Fluor 594 (B), DAPI (C) and the phase contrast (D).

Additional to the immunofluorescence assay the expression of Cas9 was also determined using Western Blot. For the *Giardia* samples weak bands could be observed on the gel, which corresponds to the expected size of Cas9, 160kDa. The band is observed for all the samples except for the G6Piso promoter sample (Fig. 6, left image). A very distinct band, corresponding to the size of Cas9, could also be observed in the *Spironucleus* western blot (Fig. 6, right image). Interestingly, additional and very strong bands can be observed for all the Cas9 samples regulated with the CWP-1 promoter, even in the non-induced sample. The bands corresponds to proteins with the size of about 55 kDa and 40 kDa. These smaller protein fragments could possibly be the result of partly degraded Cas9 proteins, where chunks of the protein have been removed.

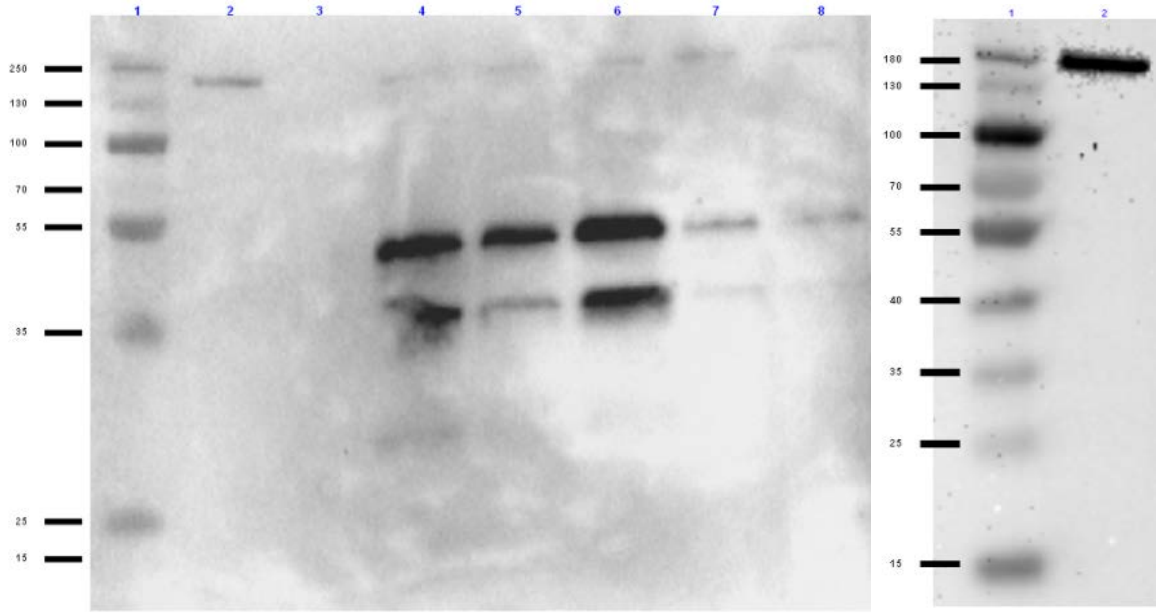


Figure 6. Expression and size determination of spCas9 in *Giardia intestinalis* and *Spironucleus salmonicida*. Left image: Western blot assay performed on *Giardia* with isolates expressing Cas9 using different promoters: OCT (lane 2), G6Piso (lane 3), CWP-1 (lane 4-8). The CWP-1 promoter is induced by encystation and were incubated in encystation media for different time periods: 2 hour (lane 4), 1 hours (lane 5) and 0 hours (control) (lane 6). One sample encysted for either 1 or 2 hours were also incubated in TYDK media overnight (1 hour: lane 7, 2 hours: lane 8). Right image: Western blot assay with a *Spironucleus* sample (lane 2).

The reason the enzyme did not translocate to the nuclei may be due to several factors. The enzyme might be too large in size in order to enter the nuclei or the NLS used in the project might be deficient or not strong enough in *Giardia* and *Spironucleus*. The knowledge of the mechanism and control of the transport between the cytoplasm and the nuclei among these organisms is scarce.

Considering the size, *Spironucleus* trophozoites were also transfected with episomes encoding the 120kDa large saCas9 enzyme, which is smaller than the regular 160 kDa spCas9. The saCas9 was successfully expressed, however, the results resemble the ones observed with the regular spCas9: expression of the enzyme which is distributed to the cytoplasm and no distinct localization in the nuclei (Fig. 7).

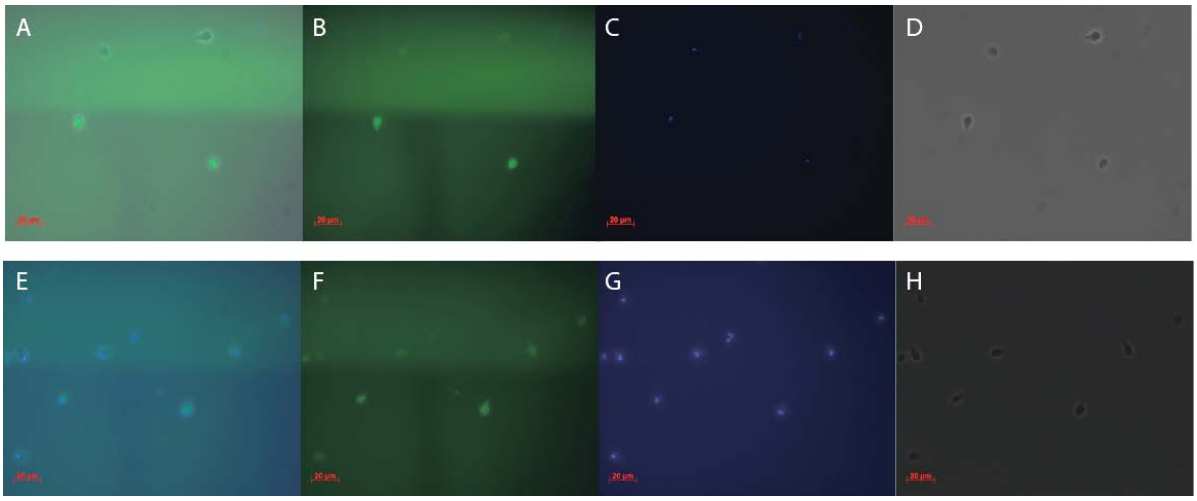


Figure 7. Expression of and localization of saCas9 in *Spironuclues salmonicida*. *Spironucleus* trophozites were fixed and permeabilized on microscope slides and stained with anti-HA mouse antibodies conjugated to Alexa Flour 488. The expression and localization of saCas9 was determined. Two different constructs were used: one including a NLS in both the N- and C-terminal of the saCas9 enzyme (A-D) and one with the NLS in only the N-terminal (E-H). Images shows the phase contrast (D, H), DAPI staining(C, G), the Cas9- Alexa Fluor 488 (B, F) and a fused image (A, E).

Another possible reason of the failed nuclear translocation is that the used NLS is insufficient for nuclei translocation in the trophozoites. In this project the SV40 large T antigen NLS was fused with the Cas9 enzyme, in the N-terminal as well in both the ends for one of the saCas9 constructs used in *Spironucleus* (Fig. 7, E). This NLS is widely used for protein nuclei translocation in a various eukaryotes and a SV40 NLS tagged to GFP has been successfully translocated to the nuclei in *Giardia* (Elmendorf *et al.* 2000). However, this NLS might not be optimal to use in these organisms and therefore usage of alternative NLS signals should be considered.

2.2 EXPRESSION OF GUIDERNA IN *GIARDIA*

Four different constructs were prepared, two encoding for targeting of the GusA reporter gene and two for the in *Giardia* endogenous ADI gene. The trophoziotes transfected with the guideRNAs targeting the ADI gene did not survive the antibiotic selection pressure, hence the transfection failed. The trophozoites transfected with the GUS targeting guideRNAs did on the other hand survive and the level of transcribed guideRNA within the trophozoites was analysed using qPCR.

Figure 8 shows the expression of the two GUS guideRNAs compared to the expression of tryptophanyl-tRNA-synthetase, which work as a reference gene. The figures show a strong expression of both the guideRNAs compare to the reference. There is also a huge difference between the levels of expression of the two guideRNAs, even though they are regulated by the same promoter. This is probably because the secondary structure of the two RNA molecules varies in stability, with Gus-2 being the more stable one.

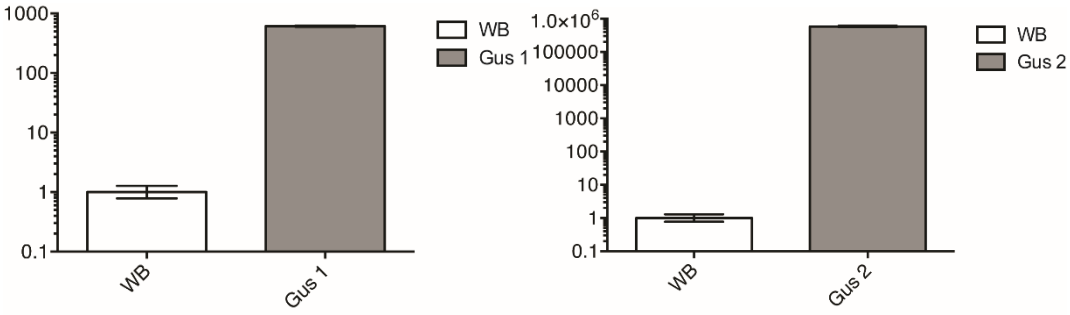


Figure 8. guideRNA expression in *Giardia intestinalis.* Two guideRNA targeting the GUS-reporter gene were episomally expressed in *Giardia*. The level of expression was determined with qPCR and compared to the expression of tryptophanyl-tRNA-synthetase (WB). The bars represents the standard error of the mean.

The target of these two guiderRNAs, the gusA-reporeter gene, was to be integrated into the chromosome of *Giardia*. The potential knockout of this gene would work as a proof of concept for Cas9 in this organism. The integrational plasmid with the gusA-reporeter gene was prepared and trophozoites were transfected. However, the transfectants did not survive the subsequent antibiotic selection pressure and more transfection attempts will need to be performed.

The results show that expression of guideRNA under the regulation of the U6 promoter in *Giardia* is possible.

2.3 EXPRESSION OF SRNA PROCESSING ENZYMES IN *GIARDIA*

In addition to the expression of the Cas9 and the guideRNAs two putative snRNA processing enzymes were expressed, GL50803_8026 and GL50803_4498. The cellular localization of the two proteins can be observed in figure 9. The GL50803_8026 show distinct localization to the nuclei as well as to some extent to the cytoplasm (Fig. 9, A), while the GL50803_4498 solely localizes to the cytoplasm (Fig. 9, E).

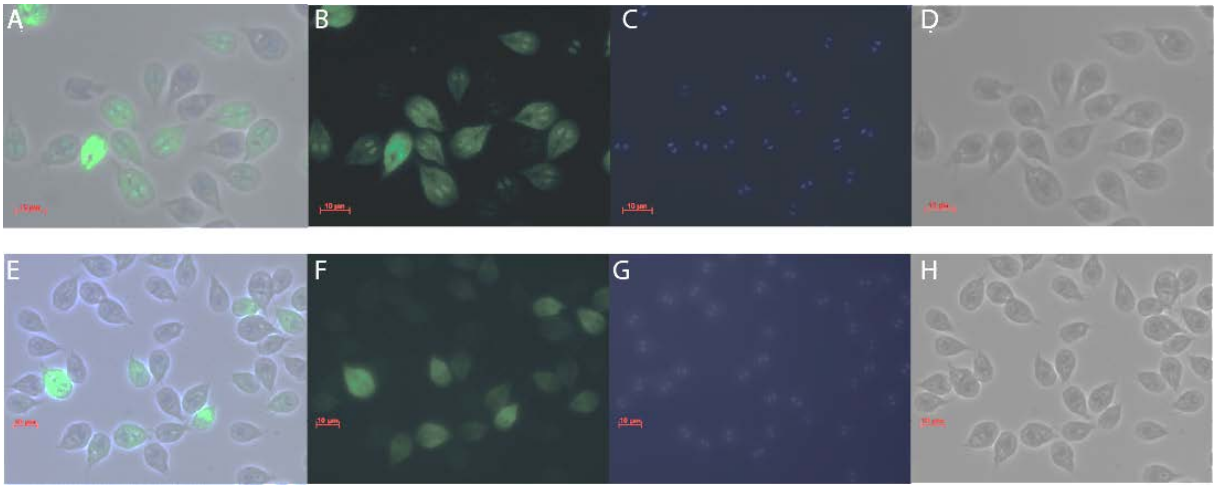


Figure 9. Expression of and localization of putative RNA processing proteins in *Giardia intestinalis*. *Giardia* trophozites were fixed and permeabilized on microscope slides and stained with anti-HA mouse antibodies conjugated to Alexa Flour 488. The expression and localization of two putative RNA processing enzymes were determined, GL50803_8026 (A-D) and GL50803_4498 (E-H). Images shows: the phase contrast (D, H), DAPI stain (C, G), the Cas9- Alexa Fluor 488 (B, F) and a fused image (A, E).

The two proteins contain domains that are putative homologs to RNA processing enzymes in other organisms.

GL50803_4498 contains a domain which is a putative homolog to the YSH1 protein (Morrison *et al.* 2007). This protein is an endoribonuclease and is a part of the mRNA cleavage and polyadenylation specificity complex in yeast and is required for proper processing of the 3'-end of precursor mRNA (pre-mRNA) (Jenny *et al.* 2015). Processing of pre-mRNA, 5' and 3'-end modification and splicing, is conducted in the nuclei. Hence, the pre-mRNA processing proteins should be translocated to the nuclei and the GL50803_4498 does not. It is possible that the GL50803_4498 protein has another functionality than the YSH1 protein and is involved in some processes located to the cytoplasm. Another explanation could be that the epitope tag fused with the protein at the C-terminal, enabling localization, have disrupted the NLS of the protein thus prohibiting nuclear translocation.

The other protein, GL50803_8026, have a domain which seems to be a homolog to the integrator complex subunit 11 (Morrison *et al.* 2007). This complex is involved in the 3'-end processing of small nuclear RNAs (snRNA), which differs from the processing of pre-mRNAs (Egloff *et al*. 2008). The GL50803_8026 protein clearly translocates to the nuclei and this strengthens the possibility of it being involved in a similar function in *Giardia*. It has also been suggested that integrator proteins may be involved in other cellular processes (Chen and Wagner, 2010), which could explain why the protein to some extent localizes to the cytoplasm.

The guideRNAs are being regulated and processed by the U6 promoter and associated 3'-end box. The U6 spliceosomal RNA is a snRNA and do not undergo polyadenylation. This makes the GL50803_8025 protein the stronger candidate for being involved in the guideRNA processing compared to the GL50803_4498 protein.

3 CONCLUSION

The results obtained in this project shows that further work needs to be done in the attempts to implement the CRISPR-Cas9 system in these two diplomonads.

Progress was done in the sense of a successful expression of the guideRNA in Giardia. This enables a possible gene knockout approach where the Cas9 protein is transformed into trophozoites expressing the guideRNA. During this project the guideRNAs where expressed episomally, a more steady expression could possibly be achieved by integrating the guideRNA and its cassette onto the chromosome. An issue with this approach is the question how to select for the trophozoites successfully transfected with Cas9 enzyme. A method for positive selection would be needed.

The Cas9 was expressed in both *Giardia* and *Spironucleus*. Unfortunately, no specific localization of the protein could be observed in any of the attempts. The smaller Cas9 derived from *Staphylococcus aureus* were expressed in *Spironucleus* to examine whether the smaller size would facilitate the nuclei translocation. However, the results resembles the ones seen with the conventional Cas9 derived from *Streptococcus pyogenes*. Inadequate NLS is another possible explanation. Trials where native *Giardia* and *Spironucleus* NLSs are fused with the Cas9 protein could possibly be the solution to this translocation problem.

In addition to the CRISPR-Cas9 system implementation attempts, two putative guideRNA processing enzymes were expressed. The guideRNA is processed in the nuclei and a proper processing of the guideRNA is crucial for its functionality. One of the proteins did localize to the nuclei and contain a domain homologous to a snRNA processing enzyme, thus is a strong candidate. Constructs with Cas9 fused to the predicted NLS of this protein are at the moment being prepared.

A functional CRISPR-Cas9 system in *Giardia* and *Spironucleus* would be extremely useful in the research of these organisms and I believe a successful implementation will be seen in the near future.

4 MATERIAL AND METHODS

4.1 CLONING & CONSTRUCT DESIGNS

Two vectors were used for episomal expression: the pPac-MCS vector in *Giardia* and the pSpiro-PAC vector (Jerlström-Hultqvist *et al.* 2012) in Spironucleus.

One vector was prepared to enable integration of genes into the *Giardia* chromosome, NptII-vInteg.

The vectors used for episomal expression contains a resistance gene for puromycin (pac), for selection of trophozoites, a resistant gene for ampicillin (amp), for selection in *E. coli* and a sequence encoding a 3xHA epitope tag. A 3xHA epitope tag was inserted in the 3' end of the inserts and utilized in immunofluorescence assays and western blots.

The MSC region of the *Giardia* pPAC vector was modified to facilitate further cloning. The MCS sequence was amplified from the original pPac vector, subsequently cloned into the same vector, using the PacI and SacI restriction sites flanking the MCS fragment. The modified vector contains two additional restriction sites, XbaI and SpeI (supplementary fig).

The Giadria integration vector was altered by removing an existing pac resistance cassette and replacing it with a neomycin-phosphotransferase (NtpII) resistance gene, allowing selection for trophozoites transfected with both the expression vector as well as the integration vector. The NptII gene was amplified from plasmid DNA using primers introducing an XbaI restriction site at the 5'-end and an EcoRV restriction site at the 3'-end, subsequently cloned into the integration vector.

4.1.1 spCas9 constructs

The *Streptococcus pyogenes* Cas9 (spCas9) gene was amplified from plasmid DNA (pSpCas9(BB)-2A-GFP(PX458), addgene, plasmid #48138) using primers with sequences including restriction sites flanking the spCas9 gene; a BamHI restriction site in the 5'-end and a NotI restriction site in the 3'-end. The forward primer also contained a part of a nuclear localization signal (NLS), which was codon optimized to include the BamHI restriction site, the other part of this NLS is included in the revers primer of the amplified promoter sequence.

The following promoters sequences were amplified from the genomic DNA extracted from *Giardia* (isolate WB): ornithine carbamoyltransferase (OCT) promoter, cyst wall protein 1 (CWP-1) promoter and Glucose-6-phosphate isomerase (G6Piso) promoter. The annexin 3 promoter used in the Spironucleus construct were amplified from plasmid DNA. The primers introduced a PacI restriction site in the 5'-end of the promoter and BamHI restriction site in the 3'-end.

The spCas9 gene promoters were subsequently cloned into either the pPAC-MCS or the pSpiro-PAC vector.

4.1.2 saCas9 constructs (*Spironucleus*)

The Staphylococcus aureus cas9 (saCas9) gene was amplified from plasmid DNA (pX600-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA, addgene, plasmid #61592). A BamHI restriction site was introduced upstream the gene and downstream the gene either a NotI or an ApaI restriction site. The different restriction sites downstream the gene allows either inclusion (NotI) or exclusion (ApaI) of the 3xHA epitope tag. The amplified genes were subsequently cloned into the pSpiro-PAC vector.

4.1.3 GusA reporter gene constructs and genome integration vector

The GusA gene was amplified from genomic DNA from *E. coli* and fused with a β-giardin promoter and a β-giardin 3'-UTR, which were amplified from genomic *Giardia* WB DNA. The assembly of the sequences were done by overlapping extension PCR. The restriction sites PacI and HpaI were introduced, in following order, upstream the promoter and SpeI and XbaI downstream the 3'UTR sequence.

The GusA gene together with its flanking sequences was supposed to be ligated into the pNptII-vInteg vector using the PacI and XbaI restriction sites. However this was not possible, since the order of the restriction sites in the pNptII-vInteg vector were in reverse order (5'-XbaI-PacI-3'). The problem was solved by cloning the sequence into the original pPac vector followed by reamplification of the gene using new primers changing the order of restriction sites to XbaI, HpaI upstream the promoter and SpeI, PacI downstream the 3'-UTR, subsequently cloned into the pNptII-vInteg vector.

4.1.4 guideRNA constructs

A plasmid containing the guideRNA cassette was ordered from Eurofin. The cassette contains a U6 sRNA promoter, tracrRNA part and an U6 termination/processing motif. It was cloned into the pPac-MCS vector using the restriction sites: SpeI and XbaI, flanking the cassette (Fig. 10). Two inverted BpiI

restriction site located between the promoter and trancrRNA sequence allows the insertion of guide RNAs.

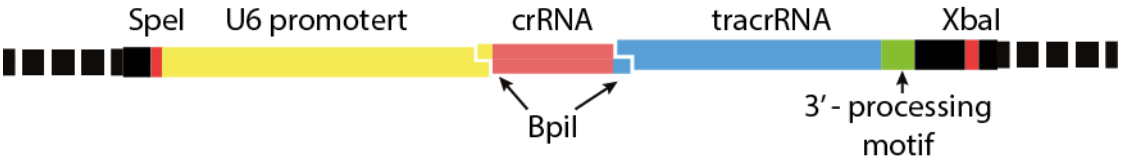


Figure 10. Illustation of the guideRNA cassette. The image illustrates the guideRNA cassette after insertion of the crRNA (guideRNA). The crRNA is cloned into the cassette using two inverted BpiI restriction sites. The cassette may in turn be cloned in to any other vector using a SpeI and XbaI restriction site. The expression is conducted by the U6 spliceosomal promoter and the 3'-end defined by the 3'-processing motif. During expression the crRNA is fused with the tracrRNA (forming the sgRNA).

The guideRNAs were ordered as single stranded oligo nucleotide sequences (Sigma-Aldrich) which were ligated together creating overhangs enabling ligation into the guideRNA cassettes. Four different guideRNA sequences were designed: two targeting the *Giardia* endogenous ADI gene and two targeting the gusA reporter gene.

Suitable guideRNA sequences were designed through an online guidRNA design tool (Peng *et al.* 2015), which allows you to find sequences with a PAM, located downstream the target sequence and limiting the possible off-target cleavages within the genome.

4.1.5 sRNA processing enzyme constructs

The genes encoding the putative sRNA processing enzymes were PCR-amplified from genomic *Giardia* WB DNA. The two following genes were PCR amplified: GL50803_4498 and GL50803_8026 (Morrison *et al.* 2007), using primers introducing an EcoRI restriction site at the 5'-end and a NotI restriction site at the 3'-end. Upstream the genes an OCT promoter sequence was insetered. The OCT sequence was PCR-amplified with primers introducing a PacI restriction site at the 5'-end and a EcoRI restriction site at the 3'-end. The parts were cloned into pPAC-MCS vector.

4.2 CULTURING OF CELLS, MEDIA PREPARATION, TRANSFECTION, ENCYSTATION

4.2.1 Growth conditions and media preparation

The *Giardia* cells were grown in 10 ml culture tubes (Nunc) in TYDK media at 37°C. 250 ml of TYDK media is prepared accordingly: 7.5 g peptone, 2.5 g glucose, 0.5 g NaCl, 0.05 g L-ascorbic acid, 0.25 g K_2HPO_4, 0.15 g KH_2PO_4, 0.5 g L-cysteine, 2.5 ml ferric ammonium citrate solution (0.11 g in 50 ml) is dissolved in ~150 ml H_2O and additional H_2O is added to 220 ml. The pH of the media is adjusted to 6.8 and then filtrated (0.45 µm filter). 25 ml bovine serum (heat inactivated) and 2.5 ml filter sterilized bile (0.25 g in 20 ml, 0.45 µm filter) is added

The *Spironucleus* cells were cultivated in LYI media and grown at 16°C in 10ml culture tubes (Nunc). 250 ml of LYI media is prepared accordingly: 0.25 g K_2HPO_4, 0.15 g KH_2PO_4, 0.25 g NaCl, 6.25 g yeast extract, 1.25 g liver digest, 0.25 g L-cysteine, 2.5 g glucose, 0.05 g ascorbic acid, 2.5 ml ferric ammonium citrate solution (2.28 mg/ml) is dissolved in ~150 ml H_2O followed by addition of H_2O to 218.75 ml. The pH was raised to 6.8 by adding proper amount of NaOH (4 M). The media was filtered

(0.45 µm filter) and aliquoted into 50 ml falcon tubes, 43.75 ml in each tube, and frozen down. Before usage the tubes are thawed and 1.25 ml Diamond Vitamin Tween80 Solution 40x and 5 ml bovine serum (heat inactivated) is added to the tube (50 ml total).

Depending on selection marker the cells were grown under selective drug pressure. Either 50 µg/ml puromycin, cells maintaining the pac gene, or 100 µg/ml gentamicin (G418), cells maintaining the NptII gene, was used.

4.2.2 Transfection by electroporation

Confluent cells were washed once and then resuspended in 1 ml culture media. All samples were kept on ice. 300µl sample aliquots were added to cold 4 mm gap electroporation cuvettes. 20-40 µg plasmid DNA were added to the cuvettes and the samples were electroporated using following settings: 350 V for *Giardia* or 320 V for *Spironucleus*, 960 µF and 800 Ω (Bio-Rad). The cuvettes were kept on ice for 15 min and then samples were transferred to fresh 10ml culture media, without any selection drug. The tubes were incubated overnight followed by addition of selective drug.

4.2.3 Encystation

For the *Giardia* cells containing constructs with cell wall protein 1 promoters (CWPp-1) transcription was induced by letting the trophozoites initiate encystation. This was performed by incubating confluent trophozoites in encystation media. The encystation media differs from regular TYDK media (see above) by containing 10x more bile and the pH is set to 7.8. Cells were incubated in encystation media and collected at different time points, 1 h and 2 h. After the incubation 3.33 ml of culture is passed to 6.66 ml 37^{o}C TYDK media and incubated for 1 h and an additional sample were incubated overnight. After the incubation the samples were prepared for IMF and western blot.

4.3 IMMUNOFLUORESCENCE MICROSCOPY

Cells were fixated at microscope slides with 2% PFA for *Giardia* cells or 3% formaldehyde for *Spironucleus* cell and incubated for 20 min at 37 ^{o}C. The slides were kept in a humidity chamber during the incubations to avoid evaporation. The fixative was removed by washing two times with PBS. 20µl 0.1 M glycine in PBS was added to quench the remaining fixative and incubated for 10 min at room temperature. The quencher was removed by washing x2 with PBS. The cells were thereafter permeabilized by adding 20µl 0.1% TritionX-100 in PBS and incubated for 30 min at room temperature. The droplets were removed and the cells were blocked by adding 20 µl 2% BSA + 0.05% TritonX-100 in PBS. The cells were blocked for 1 h at room temperature or at 8^{o}C (fridge) overnight. The block solution was removed and 20µl antibody solution diluted in antibody block solution (2% BSA+0.05% TritonX-100 in PBS) was added and incubated for 1-2 h at room temperature. For proteins tagged with 3xHA epitope tags anti-HA mouse antibodies conjugated to Alexa Flour 488 (1:250) (Nordic Biosite, #901509) were used and for Cas9 proteins without 3xHA epitope tags a CRISPR/Cas9 mouse antibody (1:500) (Epigentek, #A-9000) were used as primary antibody. For the samples stained with the Cas9 primary body, samples were washed 6 times or more using 0.1% TritonX-100 in PBS followed by addition of goat anti-mouse secondary antibody conjugated to Alexa Fluor 594 (1:800) (ThermoFisher, # A11005) and incubation at room temperature for 1 hour. Thereafter the washing step was repeated another 6 times. The washing solution was removed and the slides were dried in dark for 1 min. 3-4µl VectraShield+DAPI was added and coverslips were put on top of the slide. The coverslips were fixed by adding nail-polish in the corners, in order to let the nail-polish set the slide is incubated for 10min in dark. The slides were viewed using a Zeiss Axioplan

2 fluorescence microscope. Images were processed using the AxioVision LE release 4.8.2.0 or Zen 2011 v7.0.0.285 (Carl Zeiss GmbH).

4.4 WESTERN BLOT

Samples were prepared by resuspending pelleted trophozoites in 1 ml PBS and keeping them on ice. The OD_{600}-value was determined by blanking against PBS alone. The cells were pelleted by centrifuging samples (6000 rpm, 15 min, 4°C) and the supernatant was discarded. The pellet was resuspended in 1x Laemmli sample buffer (4 ml 10% (w/v) SDS, 2.5 ml 80% glycerol, 1.2 ml 1M Tris-Cl (pH 6.8), 0.2 ml 1% Bromophenol blue, 0.1 ml ddH_2O, 8 ml PBS), right before using the buffer 200 mM DTT was added to the buffer. The volume of buffer to add depended on the measured OD_{600}-value as each sample should contain 0.05 OD units / 10 µl after addition of the 1x Laemmli sample buffer. Samples were boiled for 10 min. After this step samples were stored at -20°C.

A precasted gel (Bio-Rad, Mini-PROTEAN TGX gel) was put into an electrophoresis tank and the tank was filled with 1x running buffer (10x running buffer: 30.3 g Tris, 144.2 g Glycine, 100 ml 10% SDS, fill H_2O to 1 L), starting by filling the inner chamber and letting it overflow until the right level is reached in the outer tank. 10 µl marker (Page ruler plus prestained, #26619, or Page ruler prestained, #26616,ThermoFisher) and 5 -15 µl sample were added to proper well. The electrophoresis was started by applying 80 volt for 10 min, followed by 130 volt until the bands of the samples reach the bottom of the gel.

The gel was transferred to a Polyvinylidene Fluoride (PVDF) membrane (Bio-Rad**)** and placed between layers of filter papers and pads and soaked in transfer buffer (1X running buffer with 20% methanol). The transfer was done by applying 100V for 1.5 h at 4°C with continuous stirring. The membrane was washed in washing solution (TBS with 0.05% tween20) followed by transfer to blocking solution (washing buffer with 5% non-fat dry milk). The membrane was blocked overnight on a rocker at 4 °C. Thereafter rinsed with washing buffer 3 times, with 5min incubation between each washing, followed by 2 h incubation in antibody solution (antibodies diluted in washing buffer with 1% Bovine serum albumin). Rat anti-HA High Affinity antibodies (Roche, # 11867423001), diluted 1:1500, and CRISPR/Cas9 mouse antibodies (Epigentek, #A-9000), diluted 1:1000, were used as primary antibodies. The membrane was washed one more time and incubated for 1 h with secondary antibodies diluted in blocking solution. Goat anti-rat-HRP secondary antibodies (ThermoFisher, #31470), diluted 1:10'000, were used for the anti-HA primary antibodies and α-mouse-HRP (DAKO, #P0161), diluted 10'000, were used for CRISPR/Cas9 primary antibodies. The membrane was washed one more time and blots were developed using the BioRad ECL kit.

4.5 QPCR

4.5.1 RNA extraction

Confluent cells were spun down 5 min, 4°C, 750 rpm *Sprionucleus*, 2000 rpm *Giardia* and pellets were resuspended in 10 ml PBS and spun once again. The pellets were thereafter resuspended and pipetted up and down in 1 ml TRIzol to lyse the cells. The TRIzol cell solutions were transferred to 1.5 ml Eppendorf tubes and incubated for 10 min in room temperature. 0.2 ml chloroform was added to each sample and the tubes are vigorously shaken for 15 seconds followed by incubation for 2-3 min at RT. The tubes were spun 12000xg for 15min at 4°C and the aqueous phases were transferred to a new 1.5 Eppendorf tube. 0.5ml of isopropanol was added to the aqueous phases, inverted a few

times and incubated for 10 min at room temperature. The tubes were spun 12000xg for 10min at 4°C followed by removal of the supernatant. The pellets were washed with 1 ml ice cold 75% ethanol and spun 7500xg for 5 min at 4°C. The ethanol was removed and the pellets were left to dry for 5-10 min. Pellets were resuspended in 30µl DEPEC water and stored at -80°C.

4.5.2 cDNA synthesis

The extracted RNA was converted to cDNA using the ThermoFisher scientific cDNA synthesis kit (#K1632).

4.5.3 RT-qPCR

Sample mixtures were prepared, 1 reaction containing: 1 µl cDNA and 10 µl SYBR Green qPCR Master Mix (ThermoFisher, #K0221). A dilution series of the cDNA was done, 1:10, 1:100 and 1:1000. Assay mixtures were prepared, 1 reaction containing: 2 µl forward and reverse primer mix (3µM) and 7 µl DNase free ddH_2O. The tryptophanyl-tRNA-synthetase gene was used as a reference gene. 11 µl of the sample mixture was added to the corresponding well of the qPCR plate (company), followed by addition of 9 µl of assay mixture and the RT-qPCR was performed (Applied Biosystem).

5 ACKNOWLEDGMENT

First of all, I would like to thank Staffan Svärd for the opportunity to work with this exciting project and introducing me to the fascinating world of *Giardia.* I would also like to give a special thanks to Ásgeir Ástvaldsson who has been a great support by helping me out and answering all my questions during my year at the lab. I am thankful for all the ideas and suggestions given by Jon Jerlström-Hultqvist through the project, for all the good comments and questions during my final presentation given by my scientific reviewer Magnus Lundgren and my friend and student opponent Dries Boers. I am also very grateful for the comments and excellent feedback given by my friend Javier Aguirre when reviewing this thesis. Finally I would like to thank the whole Giardia group and everyone at the microbiology department.

6 BIBLIOGRAPHY

Andersson, J. O., Sjögren, A. M., Horner, D. S., Murphy, C. a, Dyal, P. L., Svärd, S. G., … Roger, A. J. (2007). A genomic survey of the fish parasite Spironucleus salmonicida indicates genomic plasticity among diplomonads and significant lateral gene transfer in eukaryote genome evolution. *BMC Genomics*, *8*, 51. http://doi.org/10.1186/1471-2164-8-51

Birky, C. W. (2010). *Giardia* Sex? Yes, but how and how much? *Trends in Parasitology*, *26*(December), 70–74. http://doi.org/10.1016/j.pt.2009.11.007

Burma, S., Chen, B. P. C., & Chen, D. J. (2006). Role of non-homologous end joining (NHEJ) in maintaining genomic integrity. *DNA Repair*, *5*(9-10), 1042–1048. http://doi.org/10.1016/j.dnarep.2006.05.026

Chen, J., & Wagner, E. J. (2010). snRNA 3′ end formation: the dawn of the Integrator complex. *Biochemical Society Transactions*, *38*(4), 1082–1087. http://doi.org/10.1042/BST0381082

Doudna, J. a., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science*, *346*(6213), 1258096–1258096. http://doi.org/10.1126/science.1258096

DuPont, H. L. (2013). *Giardia*: Both a harmless commensal and a devastating pathogen. *Journal of Clinical Investigation*, *123*(6), 2352–2354. http://doi.org/10.1172/JCI69932

Egloff, S., O'Reilly, D., & Murphy, S. (2008). Expression of human snRNA genes from beginning to end. *Biochemical Society Transactions*, *36*(Pt 4), 590–4. http://doi.org/10.1042/BST0360590

Elmendorf, H. G., Singer, S. M., & Nash, T. E. (2000). Targeting of proteins to the nuclei of *Giardia* lamblia. *Molecular and Biochemical Parasitology*, *106*(2), 315–319. http://doi.org/10.1016/S0166-6851(99)00219-4

Gaj, T., Gersbach, C. a, & Barbas, C. F. (2013). ZFN, TALEN and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology*, *31*(7), 397–405. http://doi.org/10.1016/j.tibtech.2013.04.004

Gori, J. L., Hsu, P. D., Maeder, M. L., Shen, S., Welstead, G. G., & Bumcrot, D. (2015). Delivery and Specificity of CRISPR/Cas9 Genome Editing Technologies for Human Gene Therapy. *Human Gene Therapy*, *26*(7), 443–451. http://doi.org/10.1089/hum.2015.074

Gupta, R. M., & Musunuru, K. (2014). Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *The Journal of Clinical Ivestigation*, *124*(10), 4154–4161. http://doi.org/10.1172/JCI72992

Halliez, M. C. M., & Buret, A. G. (2013). Extra-intestinal and long term consequences of *Giardia* duodenalis infections. *World Journal of Gastroenterology*, *19*(47), 8974–8985. http://doi.org/10.3748/wjg.v19.i47.8974

Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*, *157*(6), 1262–1278. http://doi.org/10.1016/j.cell.2014.05.010

Jacobs, J. Z., Ciccaglione, K. M., Tournier, V., & Zaratiegui, M. (2014). Implementation of the CRISPR-Cas9 system in fission yeast. *Nature Communications*, *5*, 5344. http://doi.org/10.1038/ncomms6344

Jefferson, R. a. (1989). The GUS reporter gene system. *Nature*, *342*(6251), 837–838. http://doi.org/10.1038/342837a0

Jenny, A., Minvielle-sebastia, L., Preker, P. J., Keller, W., Jenny, A., Minvielle-sebastia, L., & Preker, P. J. (2015). Sequence Similarity Between the 73-Kilodalton Protein of Mammalian CPSF and a Subunit of Yeast Polyadenylation Factor I, *274*(5292), 1514–1517.

Jerlström-Hultqvist, J., Ankarklev, J., & Svärd, S. G. (2010). Is human *Giardia*sis caused by two different *Giardia* species? *Gut Microbes*, *1*(6), 379–382. http://doi.org/10.4161/gmic.1.6.13608

Jerlström-Hultqvist, J., Einarsson, E., & Svärd, S. G. (2012). Stable transfection of the diplomonad parasite Spironucleus salmonicida. *Eukaryotic Cell*, *11*(11), 1353–1361. http://doi.org/10.1128/EC.00179-12

Jerlström-Hultqvist, J., Einarsson, E., Xu, F., Hjort, K., Ek, B., Steinhauf, D., … Svärd, S. G. (2013). Hydrogenosomes in the diplomonad *Spironucleus salmonicida*. *Nature Communications*, *4*, 2493. http://doi.org/10.1038/ncomms3493

Jørgensen, A., Torp, K., Bjørland, M. a, & Poppe, T. T. (2011). Wild arctic char *Salvelinus alpinus* and trout *Salmo trutta*: hosts and reservoir of the salmonid pathogen *Spironucleus salmonicida* (Diplomonadida; Hexamitidae). *Diseases of Aquatic Organisms*, *97*(1), 57–63. http://doi.org/10.3354/dao02404

Katz, L. a. (2012). Origin and Diversification of Eukaryotes. *Annual Review of Microbiology*, *66*(1), 411–427. http://doi.org/10.1146/annurev-micro-090110-102808

Perez-Pinera, P., Kocak, D. D., Vockley, C. M., Adler, A. F., Kabadi, A. M., Polstein, L. R., … Gersbach, C. a. (2013). RNA-guided gene activation by CRISPR-Cas9–based transcription factors. *Nature Methods*, *10*(10), 973–976. http://doi.org/10.1038/nmeth.2600

Marz, M., Kirsten, T., & Stadler, P. F. (2008). Evolution of Spliceosomal snRNA Genes in Metazoan Animals. *Journal of Molecular Evolution*, *67*(6), 594–607. http://doi.org/10.1007/s00239-008-9149-6

Morrison, H. G., McArthur, A. G., Gillin, F. D., Aley, S. B., Adam, R. D., Olsen, G. J., … Sogin, M. L. (2007). Genomic minimalism in the early diverging intestinal parasite *Giardia* lamblia. *Science (New York, N.Y.)*, *317*(5846), 1921–1926. http://doi.org/10.1126/science.1143837

Müller, J., Nillius, D., Hehl, A., Hemphill, A., & Müller, N. (2009). Stable expression of Escherichia coli beta-glucuronidase A (GusA) in *Giardia* lamblia: application to high-throughput drug susceptibility testing. *The Journal of Antimicrobial Chemotherapy*, *64*(6), 1187–91. http://doi.org/10.1093/jac/dkp363

Peng *et al*. 2015. *EuPaGDT: a web tool tailored to design CRISPR guide RNAs for eukaryotic pathogens*. [ONLINE] Available at: http://grna.ctegd.uga.edu/. [Accessed 18 December 15].

Ran, F. A., Hsu, P. P. D., Wright, J., Agarwala, V., Scott, D. a, & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, *8*(11), 2281–308. http://doi.org/10.1038/nprot.2013.143

Ran, F. A., Cong, L., Yan, W. X., Scott, D. a., Gootenberg, J. S., Kriz, A. J., … Zhang, F. (2015). In vivo genome editing using Staphylococcus aureus Cas9. *Nature*, *520*(7546), 186–190. http://doi.org/10.1038/nature14299

Sander, J. D., & Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnology*, *32*(4), 347–55. http://doi.org/10.1038/nbt.2842

Stadelmann, B., Merino, M. C., Persson, L., & Svärd, S. G. (2012). Arginine consumption by the intestinal parasite *Giardia* intestinalis reduces proliferation of intestinal epithelial cells. *PloS One*, *7*(9), e45325. http://doi.org/10.1371/journal.pone.0045325

Svärd, S. G., Hagblom, P., & Palm, J. E. D. (2003). *Giardia* lamblia -- a model organism for eukaryotic cell differentiation. *FEMS Microbiology Letters*, *218*(1), 3–7.

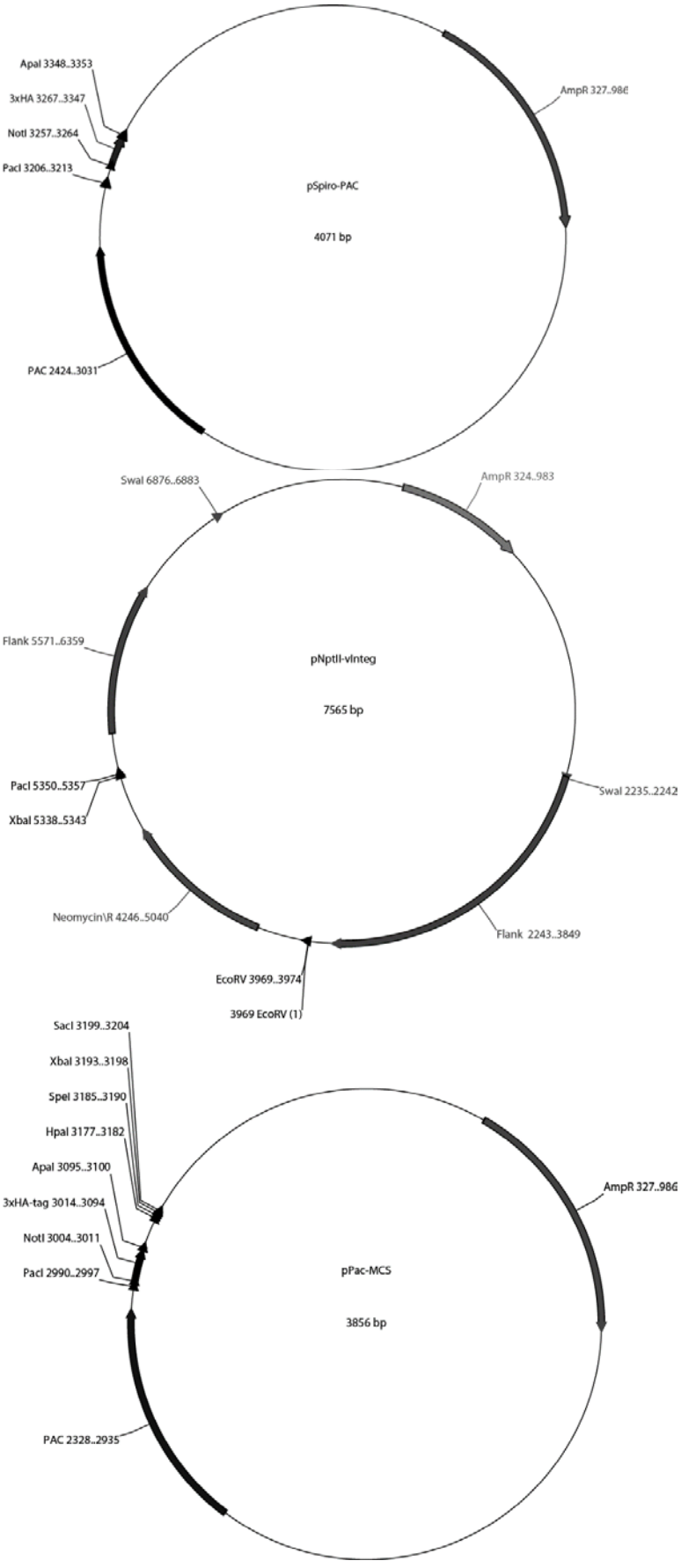
Terns, R. M., & Terns, M. P. (2014). CRISPR-based technologies: prokaryotic defense weapons repurposed. *Trends in Genetics*, *30*(3), 111–118. http://doi.org/10.1016/j.tig.2014.01.003

Thompson, R. C. A., & Monis, P. (2012). *Giardia—From Genome to Proteome*. *Advances in Parasitology* (1st ed., Vol. 78). Elsevier Ltd. http://doi.org/10.1016/B978-0-12-394303-3.00003-7

Touz, M. C., Rópolo, A. S., Rivero, M. R., Vranych, C. V., Conrad, J. T., Svard, S. G., & Nash, T. E. (2008). Arginine deiminase has multiple regulatory roles in the biology of *Giardia* lamblia, 2930–2938. http://doi.org/10.1242/jcs.026963

Xu, F., Jerlström-Hultqvist, J., Einarsson, E., Astvaldsson, A., Svärd, S. G., & Andersson, J. O. (2014). The genome of Spironucleus salmonicida highlights a fish pathogen adapted to fluctuating environments. *PLoS Genetics*, *10*(2), e1004053. http://doi.org/10.1371/journal.pgen.1004053

Yao, S., He, Z., & Chen, C. (2015). CRISPR/Cas9-Mediated Genome Editing of Epigenetic Factors for Cancer Therapy. *Human Gene Therapy*, *26*(7), 463–471. http://doi.org/10.1089/hum.2015.067


ApaI 3348..3353
3xHA 3267..3347
NotI 3257..3264
PacI 3206..3213
AmpR 327..986
pSpiro-PAC
4071 bp
PAC 2424..3031
SwaI 6876..6883
AmpR 324..983
Flank 5571..6359
pNptII-vInteg
7565 bp
PacI 5350..5357
XbaI 5338..5343
SwaI 2235..2242
Neomycin\R 4246..5040
Flank 2243..3849
EcoRV 3969..3974
3969 EcoRV (1)
SacI 3199..3204
XbaI 3193..3198
SpeI 3185..3190
HpaI 3177..3182
ApaI 3095..3100
3xHA-tag 3014..3094
NotI 3004..3011
PacI 2990..2997
AmpR 327..986
pPac-MCS
3856 bp
PAC 2328..2935

MCS primers	
Forward:	TATTTAATTAAGAATTCGCGGCCGCCCT
Reverse:	TATGAGCTCTCTAGACCACTAGTATGTTAACCCCGCATTTCTGGGGAGAC
NptII primers	
Forward:	TATGATATCAGTGCGCCGTCTGAGCCG
Reverse:	TATTCTAGACTCGAGTGAGGGCTGAAGTGAATATTTACC
spCas9 amplification	
Forward:	TATGGATCCATGGAGTCCCAGCAGCCGAC
Reverse:	TATGCGGCCGCCAGTCGCCTCCCAGCTGAG
Promoter amplification	
OCTp	
Forward:	TATTTAATTAAGTCAGCTTCATTTTTATTATCGAATAACATCATTG
Reverse:	TATGGATCCCGACCTTACGCTTTTTCTTGGGAGCCATTTTAATTTTCAGCCTCTACTGTAGAGCGTTT
CWP-1p	
Forward:	TATTTAATTAAGATTCAAGGACATGGGTCGATGT
Reverse:	TATGGATCCCGACCTTACGCTTTTTCTTGGGAGCCATCCCTGATATTTTATTTCTGTGTTTCTTGATCTG
G6Piso	
Forward:	TATTTAATTAACTAGTCTGGGATTTTCCGGATTTCA
Reverse:	TATGGATCCCGACCTTACGCTTTTTCTTGGGAGCCATCCTAGAAGATTTAAATTTAGGAAGGCCCTGTA
Anx3p	
Forward:	TATTTAATTAAGAATGACGACAGCGCGTATTAAAGAAT
Reverse:	TATGGATCCCGACCTTACGCTTTTTCTTGGGAGCCATTTTTTGCGGAAAGTATAGATGAATTTTATTAATAAACC
GusA assembly	
GusA	
Forward:	ATTTGAGCCCAAGTTAAGAAAATAAAAGGATGTTACGTCCTGTAGAAACCCCA
Reverse:	CTAAGTAAATGATTTACTGCAGGCGCTTATTGTTTGCCTCCCTGCTGC
5´-UTR	
Forward:	TATTTAATTAAGTTAACGTATGCAGCACTCACAGAGAGATG
Reverse:	TGGGGTTTCTACAGGACGTAACATCCTTTTATTTTCTTAACTTGGGCTCAAAT
3´-UTR	
Forward:	GCAGCAGGGAGGCAAACAATAAGCGCCTGCAGTAAATCATTTACTTAG
Reverse:	TATTCTAGAACTAGTGTGCTCAACTGCAACCACTACTTC

2nd GusA primers for exchange of restriction sites	
Forward:	ATATTCTAGAGTTAACGTATGCAGCACTCACAG
Reverse:	ATATTTAATTAAACTAGTGTGCTCAACTGCAACC
Guide RNA oligos	
ADI 1	
oligo 1:	CGCGGC GAAGTACGAGCTCCATGCCAC GT
oligo 2:	TAAAAC GTGGCATGGAGCTCGTACTTC GC
ADI 2	
oligo 1:	CGCGGC GCCTTGATGAAGTCCACTGT GT
oligo 2:	TAAAAC ACAGTGGACTTCATCAAGGC GC
GUS1	
oligo 1:	CGCGGC GCAAAAAACTCGACGGCCTGT GT
oligo 2:	TAAAAC ACAGGCCGTCGAGTTTTTTGC GC
GUS2	
oligo 1:	CGCGGC GCGTTGGCGGTAACAAGAAA GT
oligo 2:	TAAAAC TTTCTTGTTACCGCCAACGC GC
sRNA processing enzyme construct primers	
GL50803_4498	
Forward:	TATGAATTCATGCAGAAGAGGGTGGGCA
Reverse:	ATAGCGGCCGCGCCGGGAGGAGTGTTCGCAC
GL50803_8026	
Forward:	TATGAATTCATGCCGGGAGTACAGAACTG
Reverse:	ATAGCGGCCGCGCCTTGAAGCAGAGGCCGTTATC
OCT promoter	
Forward:	TATTTAATTAAGTCAGCTTCATTTTTATTATCGAATAACATCATTG
Reverse:	ATAGAATTCTTTAATTTTCAGCCTCTACTGTAGAGCGTTT
qPCR primers	
GUS1	
Forward:	CGACGGCCTGTGTTTTAGAG
Reverse:	CGACTCGGTCCCACTTTTTC
GUS2	
Forward:	TGGCGGTAACAAGAAAGTTTTAGA
Reverse:	GACTCGGTCCCACTTTTTCAA
Cas9 sequencing primers	
1-Cas9-R	CCAGGGCCAGATAGATCA
2-Cas9-F	TGGCCTACCACGAGAAGT
2-Cas9-R	GATGCTGCCGTTGTCGAA
3-Cas9-F	CCAGCCAGGAAGAGTTCTA
3-Cas9-R	CGTCGGACTTCAGGAAAT

4-Cas9-F	AAAGTGATGAAGCAGCTGA
4-Cas9-R	GTACTTAGTGTTCATCCG
5-Cas9-F	GAGCGAACTGGATAAGGC
5-Cas9-R	CTCGAACAGGGAGTACTT
6-Cas9-F	GAAAGAGCTGCTGGGGAT
pPAC sequencing primers	
Forward:	CACCGTGGGCTTGTACTC
Reverse:	GTAAAACGACGGCCAGT
Integration vector sequencing primers	
Forward:	TCATGCTGGAGTTCTTCG
Reverse:	TTTTCTTGAGTAGCCACC
pSpiro sequencing primers	
Forward:	TATCCATGGTGGGAAGTGGTGACCAAATC
Reverse:	GTAAAACGACGGCCAGT