AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis dissertation.

26. 5. 2015
Date

Weinzierl Astrid
Signature

Danksagung

Als Erstes möchte ich meinen Eltern, Frieda und Karl, für ihre persönliche und finanzielle Unterstützung nicht nur während meiner Masterarbeit sondern während meines gesamten Studiums danken!

Besonderer Dank gilt Prof. Anton Glieder für die wertvolle Betreuung während der gesamten Masterarbeit. Danke, dass du mich in dieser aufregenden Zeit meines Studiums begleitet und mir dieses Projekt ermöglicht hast!

Ein riesiges Danke geht an Thomas Vogl. Dafür, dass du mich immer ermutigt hast, für unzählige Diskussion, für die Korrekturen sämtlicher Arbeiten, für dein Interesse und Verständnis.

Herzlich bedanken möchte ich mich auch bei Anna Neumayer, Anna Hatzl und Christian Schmid. Danke, dass ihr mich bei der Laborarbeit unterstützt habt, nachdem die Masterarbeit schon beendet war, um möglichst viele Daten zu sammeln.

Des Weiteren möchte ich mich ausnahmslos bei der gesamten Glieder-Gruppe bedanken. Einerseits für die fachliche Unterstützung, aber auch für die letzten zwei Jahre, die wir gemeinsam mit allen Höhen und Tiefen erlebt haben. Es ist schön Freunde wie euch zu haben.

Auch möchte ich allen Studienkollegen und Freunden für die tolle gemeinsame Zeit danken.

Ein weiteres Dankeschön geht an meine Schwester Karin.

Der größte Dank gebührt meinem Freund Thomas, der mir immer zu Seite gestanden ist und mit dem ich schon so vieles zusammen gemeistert habe.

Weiters möchte ich mich für die Finanzierung, die diese Masterarbeit möglich gemacht hat, bedanken: The research leading to these results has received funding from the Innovative Medicines Initiative Joint Undertaking project CHEM21 under grant agreement n°115360, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution.

Abstract

Over the last decades *Pichia pastoris* (*Komagataella phaffi*) has been extensively used for the production of recombinant proteins and more recently also for metabolic engineering and synthetic biology. However, the introduction of targeted genome modifications is uttermost challenging in *P. pastoris*. In contrast to *Saccharomyces cerevisiae*, where short homologous sequences of ~50 bp are sufficient to target correct integration efficiently, even much longer homologous sequences are integrated predominately randomly in *P. pastoris*. Frequently gene replacement by homologous recombination occurs in *P. pastoris* only in 0.1% up to 30% of all transformants depending on the target locus and the size and the design of the donor fragment. Methods for targeted genome engineering in *P. pastoris*, which allow a generation of knockout strains or to perform gene deletion studies in a short time, are highly sought after. In this work, the genome editing methods CRISPR-Cas9 and TALENs were evaluated for their usability to introduce targeted genome modifications in the methylotrophic yeast.

As a first step a CRISPR-Cas9 vector was developed based on the features that had been used for CRISPR-Cas9 mediated genome engineering in *S. cerevisiae*. This construct was not functional in *P. pastoris*. Step by step the single elements of this vector were exchanged and combined in a successive manner. These features include various (codon optimized) DNA sequences of Cas9, different gRNA sequences, RNA Polymerase III and RNA Polymerase II promoters (in combination with ribozymes) for the expression of the gRNAs and RNA Polymerase II promoters for the expression of Cas9. In addition a variety of nuclear localization signals, autonomously replicating sequences and selection markers were characterized. Only three of approximately 74 constructs tested enabled efficient gene targeting in *P. pastoris*, namely those bearing a RNA polymerase II promoter used for the expression of ribozyme flanked gRNAs in combination with a human codon optimized Cas9 sequence. This study resulted in a CRISPR-Cas9 system with 83% to 94% targeting frequency for the test locus *GUT1* (*glycerol kinase 1*). Subsequently, to show that the developed system can be used to generate knockout strains, co-transformation experiments with various homologous donor cassettes were performed. Additionally interesting genomic loci were targeted (*AOX1*, *MXR1*, *MPP1* and *PRM1*) and CRISPR-Cas9 was used to introduce mutations simultaneously in the *AOX1* and *GUT1* locus. Also synthetic TALENs had been purchased, but they have never been tested in *P. pastoris*, due to severe cloning problems caused by the highly repetitive nature of the single TALE-repeats.

Kurzfassung

Pichia pastoris (*Komagataella phaffi*) wird seit Jahrzenten für die Produktion von rekombinanten Proteinen verwendet und seit kurzem auch in den Bereichen Metabolic Engineering und Synthetische Biologie. Das Einbringen gezielter Mutationen in das Genom von *P. pastoris* ist allerdings schwierig zu bewerkstelligen. Im Gegensatz zur Bäckerhefe *Saccharomyces cerevisiae*, wo kurze homologe Sequenzen (~50 bp) für die ortsspezifische Integration reichen, werden in *P. pastoris* auch viel längere Sequenzen zufällig integriert. Abhängig von dem Integrationsort und dem Design und der Größe der Donor Fragmente können Abschnitte im Genom nur in 0,1% bis zu 30% der Transformanten durch homologe Rekombination entfernt werden (Gene Replacement). Methoden, die es ermöglichen gezielt Mutationen im Genom von *P. pastoris* einzubringen, um Knockout Stämme in kürzester Zeit herzustellen, sowie Deletionsstudien durchzuführen, sind von großer Notwendigkeit. Im Zuge dieser Masterarbeit, wurden die Methoden CRISPR-Cas9 und TALEN für zielgerichtetes „Genome Engineering" in *P. pastoris* getestet.

Basierend auf einem CRISPR-Cas9 System für die gezielte Modifikation von Sequenzen im Genom von *S. cerevisiae* wurde ein Vektor für *P. pastoris* entwickelt. Allerdings war das Konstrukt nicht funktionell. Stück für Stück wurden einzelne Elemente von diesem Vektor getauscht. Dazu gehören unterschiedliche (Codon-optimierte) DNA Sequenzen von Cas9, unterschiedliche gRNA Sequenzen, RNA Polymerase III und RNA Polymerase II Promotoren (in Verbindung mit Ribozymen) für die Expression der gRNAs und unterschiedliche Promotoren für die Expression von Cas9. Zusätzlich wurden eine Reihe an Kernlokalisierungssequenzen, autonom replizierender Sequenzen und Selektionsmarker charakterisiert. Nur drei von 74 Konstrukten ermöglichten das Einbringen von Modifikationen mit hohen Raten, nämlich jene, die einen RNA Polymerase II Promoter für die Expression von mit Ribozymen flankierter gRNAs in Kombination mit einer für *Homo sapiens* Codon optimierten Cas9 Sequenz enthielten. Im Laufe dieser Masterarbeit wurde ein CRISPR-Cas9 System mit 83% bis 94% Mutationshäufigkeit für den Testlocus *GUT1* (*gylcerol kinase 1*) entwickelt. Anschließend, um zu zeigen, dass dieses CRISPR-Cas9 System für die Erzeugung von Knockout Stämmen verwendet werden kann, wurden Co-Transformationsexperimente mit unterschiedlichen Donor Kassetten durchgeführt. Zusätzlich wurde das CRISPR-Cas9 System genützt, um Mutationen in verschiedenen Genen (*AOX1*, *MXR1*, *MPP1* und *PRM1*) einzubringen und um Multiplexing (*AOX1* und *GUT1*) durchzuführen. Auch wurden synthetische TALENs bestellt, die jedoch nie in *P. pastoris* angewendet wurden, da die hohe Repetitivität der einzelnen TALE-Einheiten Probleme beim Klonieren verursachte.

Table of contents

1. INTRODUCTION 1
Targeted genome engineering for custom designed organisms 1
Non programmable genome engineering 2
Site–specific recombinases (Cre/loxP, FLP/FRT) 2
Transposons 4
RNA interference (RNAi) and translational repression 5
Prokaryotic genome engineering 8
Group II introns - Targetrons 8
Recombineering 9
Universally applicable genome editing strategies 10
Homing endonucleases 10
Zinc finger nucleases 11
TALE nuclease technologies 13
CRISPR–Cas9 technology 14
2. OBJECTIVES 19
Background and concepts of this thesis 20
Significance and selection of features for CRISPR-Cas9 mediated targeted genome engineering in *P. pastoris* 20
The nuclear import of Cas9 20
The influence of different autonomously replicating sequences (ARS) and antibiotic resistance markers on protein production and plasmid maintenance 24
Cas9 based targeted genome engineering 25
The expression of gRNAs 26
Ribozymes in CRISPR-Cas9 applications 28
Genomic target loci for testing CRISPR-Cas9 based genome engineering 29
TALEN mediated genome engineering in *P. pastoris* 31
3. MATERIALS AND METHODS 32
Media, buffers and kits 32
Protocols and assays 33
DNA isolation from *E. coli* (plasmid) and *P. pastoris* (genomic DNA) 33
Ligation reactions and Gibson Assembly 34

Cloning into the pJET1.2 blunt vector 34
PCR procedures 34
Restriction endonuclease reactions 35
Transformation protocols 35
96-deep well plate cultivation of *P. pastoris* strains expressing eGFP under the control of P_{AOX1} 36
96-deep well plate cultivation and screening with of *P. pastoris* transformants with ARS plasmids 36
Measurement of eGFP and sTomato fluorescence levels 37
Nuclear staining and fluorescence microscopy 37
Codon optimization 37
gRNA design 38
Strains and plasmids 38
Construction of NLS plasmids 38
Construction of sTomato-ARS plasmids 39
Construction of donor DNA cassettes and control strains 40
Construction of CRISPR-Cas9 plasmids with RNA Polymerase III promoters 40
Construction of CRISPR-Cas9 plasmids bearing a bidirectional RNA Polymerase II promoter and ribozymes 42
*Pp*Cas9 fusion constructs 43
Additional CRISPR-Cas9 targets 44
Multiplexing 44
Experimental work on TALEN vectors 45
E.coli strains and plasmids 46
P. pastoris strains 51
4. RESULTS AND DISCUSSION 53
The identification of potent nuclear localization signals 53
The effect of autonomously replicating signals and different antibiotic resistance genes on heterologous protein production 58
Genomic target loci for the evolution of CRISPR-Cas9 in *P. pastoris* 62
CRISPR-Cas9 constructs with RNA polymerase III promoters 65
CRISPR-Cas9 constructs with RNA Polymerase II promoter and ribozymes 69
CRISPR-Cas9 genome editing in *P. pastoris* CBS 7435 *ku70* 71

CRISPR-Cas9 and donor DNA co-transformation in *P. pastoris* CBS 7435 WT and *P. pastoris* CBS 7435 *ku70* 73
Plasmid maintenance and curing 77
Characterization of the developed CRISPR-Cas9 system 78
Additional CRISPR-Cas9 target loci 78
Multiplexing 79
Synthetic TALENs to target *GUT1* 80
5. SUMMARY AND OUTLOOK 82
6. BIBLIOGRAPHY 85
7. ABBREVIATIONS 107
8. LIST OF FIGURES 110
9. LIST OF TABLES 111
10. APPENDIX 112
GUT1 DNA sequence 112
Supplementary data 113
CRISPR-Cas9 vector design 115
List of materials 117
Chemicals 117
Instruments and devices 118
Enzymes 121
Software and webtools 122
Primer and gBlocks 123

1. INTRODUCTION

Targeted genome engineering for custom designed organisms

The introduction to genome engineering techniques was written as part of the book chapter "Genome engineering and DNA assembly" with the intention of being published in "Key methods for synthetic biology" (Springer, preliminary publication date: 2015).

Synthetic biology is described as engineering–driven building of organisms and biological entities with beneficial functions [1], [2]. In the last decade the interest of biological engineering shifted from individual genes to entire genomes. For basic research and industrial applications the generation of rationally designed organisms cannot be achieved solely by random mutagenesis and selection, but requires advanced genome engineering techniques [3]–[5].

Genome engineering refers to any technology and method for genome-scale modifications of organisms [3]. Molecular engineering tools to introduce random modifications and techniques for targeted genome engineering, which address a defined, desired genomic locus, are available. Some methods modify the genome in a self-sufficient manner, whereas others rely on the host cellular machinery. Genome editing is the collective term for methods taking advantage of the DNA repair mechanisms of the host cell. Host mediated DNA repair is achieved by two different strategies and depends on the nature of the DNA strand break. In case of a double strand break the repair is either triggered by non-homologous endjoining (NHEJ) [6] or by homologous recombination (HR) [7]. Single strand breaks are solely repaired by HR. Both mechanisms allow the introduction of genome modifications and alteration of genomic conditions. In its natural role HR promotes the exchange of endogenous DNA sequences, but for genome engineering applications the mechanism is hijacked to exchange genetic information between endogenous sequences and an artificially constructed exogenous DNA fragment [8]. NHEJ rejoins DNA ends without the requirement of a homologous template and is often accompanied by substitutions, insertions and deletions (indels) of nucleotides in the targeted region [9]. Targeted single and double strand breaks induced DNA repair increases the rate of homologous recombination by several orders of magnitude (up to 4000–fold in yeast) [10]–[13]. Therefore systems introducing breaks at programmable positions in the genome are highly sought after.

Early genome engineering approaches showed low efficiency rates and relied on random integration or on a limited number of predefined genomic sites. Nowadays a vast range of prokaryotic and eukaryotic genome engineering techniques (Table 1) facilitate the generation of gene knockouts, gene-delivery and the introduction of gene displacements or chromosomal rearrangements [3], [14].

In the first section of this chapter I discuss several genome engineering techniques by comparing their capabilities and limitations. Genome engineering techniques range from site–specific endonucleases (Cre/loxP, flippases) to mobile genetic elements such as transposons or ribozymes (Group II Introns). Methods such as recombineering are only applicable for prokaryotes, whereas others are suitable for all domains of life. One of these techniques is genome editing, which takes advantage of engineered nucleases to cut a defined genome locus and thereby introduces a disruption and sequence change [15]. These nucleases possess either a specific affinity for a defined DNA sequence (Homing endonucleases), are fused to DNA binding domains such as transcription activator-like effectors (TALEs) and zinc finger domains or they use a short guide RNA to specifically cut a DNA sequence of interest (CRISPR-Cas9).

Non programmable genome engineering

Site–specific recombinases (Cre/loxP, FLP/FRT)

Site–specific recombinases catalyse DNA cleavage reactions between two identical recognition sites to inverse, excise or integrate a DNA fragment (Figure 1A) [16]. In contrast to programmable systems such as CRISPR-Cas9 [17] or TALENs [18], the recognition sites of site–specific recombinases cannot be altered (non programmable). The site specific Cre (**C**auses **re**combination [19]) recombinase from bacteriophage P1 and the yeast derived flippase (FLP) are the most extensively studied recombinases and are widely used for genome engineering [20]. Cre recombination occurs between two consensus 34 bp DNA recognition sites, called loxP sites and its applications date back to the early 1980s [21]. However, cytotoxic effects due to off-target endonucleolytic activity with pseudo recombination sites in some eukaryotes shifted the interest to the FLP recombinase for eukaryotic site-specific recombination [22], [23]. FLP is naturally encoded on the 2 µ plasmid of *Saccharomyces cerevisiae* and promotes recombination between two identical, minimal 34 bp FLP recombination sites (FRT sites) [24]. Cre/FLP recombination does not depend on any supporting host cell factors such as topoisomerases or the DNA

replication machinery. Moreover the recombinases can enter the eukaryotic nucleus and perform recombination unimpaired by the chromatin structure [25]. Recombination between two recognition sites organized in the same orientation leads to the excision of the flanked DNA, whereas recombination between target sites arranged in opposite orientations triggers the inversion of an intervening DNA fragment. If the recognition sites are located on different chromosomes translocation events can take place [26]. Incorporation of DNA sequences only occurs, if the target site is present on the donor vector and in the targeted genomic locus (Figure 1A): After transformation of the recombinase expression vector and a donor plasmid into the cells, the recombinase cuts the genomic locus at the target site and also linearizes the donor vector. The donor sequence is incorporated at the target locus accompanied by a duplication of the FRT/loxP sites. A major disadvantage of this system is the requirement for pre-integration of the recognition site at the target site. Therefore, site specific recombinases cannot be used to introduce targeted mutations in any desired locus [27], except for the case strains with specifically introduced binding parts had been generated before. Nevertheless, efficient bacterial genetic engineering was performed by combining the Cre/loxP recombinase and mobile group II introns. The introns were used to deliver the loxP sites to a specific genomic locus enabling the genomic modification by the site specific recombinase [28]. Another disadvantage is that the recognition sites of the recombinases are identical before and after the recombination, which facilitates the excision of the integrated fragment [27]. This drawback had been overcome by the use of poisoned half sites, which cannot be cleaved by the enzyme upon recombination [29].

Recombinase-mediated cassette exchange (RMCE) [30], [31] is an integrase based technique used for the nondisruptive insertion of a DNA cassette into a precharacterized genomic locus. A genomic destination called landing pad contains a selection marker cassette flanked by recombinase recognition sites. A circular donor vector containing an analogous cassette encoding the gene of interest is transformed into the host cells and is used to replace the resident cassette by the aid of the recombinase. RMCE was successfully used for predictable expression of heterologous genes in cell cultures and for the systematic generation of transgenic animals in a selection marker free environment [32].

The same site-specific recombinases show also a great promise for selection marker recycling (*e.g.* in *S. cerevisiae* [33] or *Pichia pastoris* [34]). Typically a knockout cassette, which contains a sequence identical to the 5’ region of the target, a selection marker cassette gene flanked by two

FRT/loxp sites and a sequence identical to a region 3' of the target, is transformed into the host and integration takes place by the cellular homologous recombination machinery. After verification of a positive transformant the selection marker can be selectively excised by a site specific recombinase [33].

Transposons

Transposons are mobile genetic elements, which change their position in the host genome by a self–mediated mechanism called transposition. Many naturally occurring DNA transposons consist of a transposase gene and its promoter flanked by terminal inverted repeats (TIR) [35]. The TIR of DNA transposons function as DNA binding sites for the transposase, which on the one hand excises the transposon from one genomic locus and on the other hand cuts at a genomic integration site and mediates the reintegration of the transposon [35]. In contrast to DNA transposons, which replicate mainly by this cut and paste strategy, RNA transposons are transcribed into RNA, reverse transcribed into DNA and then inserted into a new position applying a copy and paste mechanism [36]. The integration sites of transposons are either close-to-random (two bp recognition) or occur at specific recognition sites or genomic hot spots [35], [37]. Therefore transposons have been widely used for random mutagenesis in prokaryotic and eukaryotic host systems (e.g. [38]–[42]). For genome engineering purposes a DNA sequence of interest is cloned between the TIRs on a plasmid vector from where it gets excised and further integrated in the chromosome by the transposase, expressed from a separate expression plasmid (Figure 1B) [35]. A collection of transposable elements with different integration preferences, cargo capacities and species-specificities are available [43]. Widely used transposons include the commercially available, mariner-type transposon Sleeping Beauty [44], the Ac/Ds system described by the Nobel Prize winner Barbara McClintock [45], the synthetically reconstructed Frog Prince [46] and the artificial piggyBAC transposon [47], [48]. The integration site of these transposons range from AT dinucleotides to palindromic consensus sequences similar to restriction enzymes [35]. Transposases with higher site specificity have been created by fusions to zinc finger (ZF) [49], [50] and TALE DNA binding domains [51]. Although transposase fusion proteins have a high affinity for their intended target region, they also show off-targeting activity with a prevalent danger of dysregulating endogenous genes [49].

Transposition was one of the first genome editing tools used for for insertional somatic and germline transgenesis in mammalian cell lines [52]–[54] . Moreover transposons are often used

as a delivery system for DNA as a less oncogenic alternative to viral vectors. In contrast to their viral counterparts, transposons can be maintained as plasmid DNA. Also, packing capacities are higher and the transfection protocols are less labour intensive and time consuming [35]. One of the major advantages of the transposition is its independency from cellular repair pathways, the stage of the cell cycle and the cell type [55].

RNA interference (RNAi) and translational repression

RNA interference (RNAi)-induced targeted gene knockdown by small interfering RNAs (siRNA) or microRNAs (miRNAs) is a rapid and inexpensive technique primarily applied in higher eukaryotes (Figure 1C). In contrast to the previously described genome engineering methods, regulation takes place at the posttranscriptional level. It offers an alternative to genome engineering methods by taking advantage of short non-coding RNAs (miRNAs and siRNAs), which guide an RNA-induced silencing complex (RISC) to bind complementary sequences in the messenger RNA (mRNA). miRNAs and siRNAs derive from structurally different precursor molecules, but they are processed in a similar manner. Binding of the miRNA-RISC complex to the target mRNA decreases protein output; however, the molecular details and timing of how mRNA degradation and translational repression each contribute to this effect are still a matter of scientific debate. Several studies show that miRNAs can function as siRNAs and *vice versa* and that the mechanism of choice is highly dependent on the degree of complementarity of the RNA target [56]–[59].

The RNAi mechanism is naturally protecting the genome against mobile genetic elements such as transposons or viruses [60] and is used as an important mechanism for regulating gene expression [61]. In metazoan RNAi the primary siRNA transcripts are cleaved by a ribonuclease III termed Drosha (in animals, DcI1 in plants) into 70–80 bp precursor-miRNAs (pre-miRNAs) [62]. The double stranded pre-miRNAs are then exported to the cytoplasm and processed by the multidomain ribonuclease Dicer into double stranded siRNA. The 21-23 bp siRNAs have a phosphorylated 5’ terminus and a two nt 3’ overhang, which is required for the recognition of the siRNA by RISC [63], [64]. Dicer passes the ds siRNA to the RISC, where the DNA duplex is unwound and the passenger strand, which is the strand whose 5’ terminus has the thermodynamically more stable end of the duplex, gets degraded by the nuclease called Ago2 (in human, fly) [65], [66]. The single stranded guide RNA (guide strand) is incorporated in the RISC

and guides the complex for sequence specific mRNA cleavage, which occurs in the region complementary to the siRNA [67].

In contrast to siRNAs, miRNA are naturally synthesized from a 60-70 nt transcript, which folds into a stem loop precursor. These precursor molecules are processed by the nucleic Drosha and the cytosolic Dicer, similarly to siRNAs, into ds miRNAs of ~22 bp length [68]–[71]. The mature miRNA is incorporated into a ribonucleoprotein (RNP) complex comparable to the RISC [72]. miRNAs bind with mismatches and bulges to the target and cause direct translational repression or mRNA destabilization [73]. A short miRNA region designated as "seed", which comprises the positons 2-7 or 2-8 of the mature miRNA, shows perfect base pairing to the mRNA's "seed match" sequence [74]. In general miRNAs and siRNAs use similar mechanisms for the repression of mRNA expression and the cleavage of mRNA. Although the binding sites for miRNAs in animals lie almost only in the 3'UTR of the target gene, any position on the mRNA is mechanistically sufficient for miRNA binding and translation repression [75].

In addition to the naturally occurring mi/siRNAs synthetic ss and ds RNAs, 21–22 nucleotide siRNAs and short haipin RNAs (shRNAs) have been successfully used to induce RNA interference and specific gene silencing in numerous organisms [63], [76]–[80]. These molecules are either endogenously expressed *in vivo* or synthesized *in vitro* and delivered into the host organism [81]. Transfected RNA is prone to degradation, whereas the stable integration of RNA expressing vectors ensures a persistent gene silencing [79], [82].

Creating a clean, confirmed gene knockout is a tedious procedure in higher eukaryotes. RNAi is a popular method for posttranscriptional loss of function studies due to is simplicity, efficiency and reduction of cost compared to genome engineering techniques. siRNAs and miRNAs for RNAi can be quickly prepared and introduced into the host organism or cell, where they are used to perform functional studies and generate double- or triple-loss-of-function effects independently from molecular genetic technologies [83]. Nevertheless RNAi is accompanied by off-targeting effects, since it can activate dsRNA responsive cellular pathways resulting in a dysregulation of host cell genes [84]. In addition, off-target binding to similar mRNA sequences alters the endogenous protein production levels and thereby limits the applicability of RNAi for basic research and clinical therapy [85], [86]. Thus it is recommended to confirm RNAi effects with two or more dsRNAs/siRNAs/shRNAs targeting different regions in the gene of interest [83].

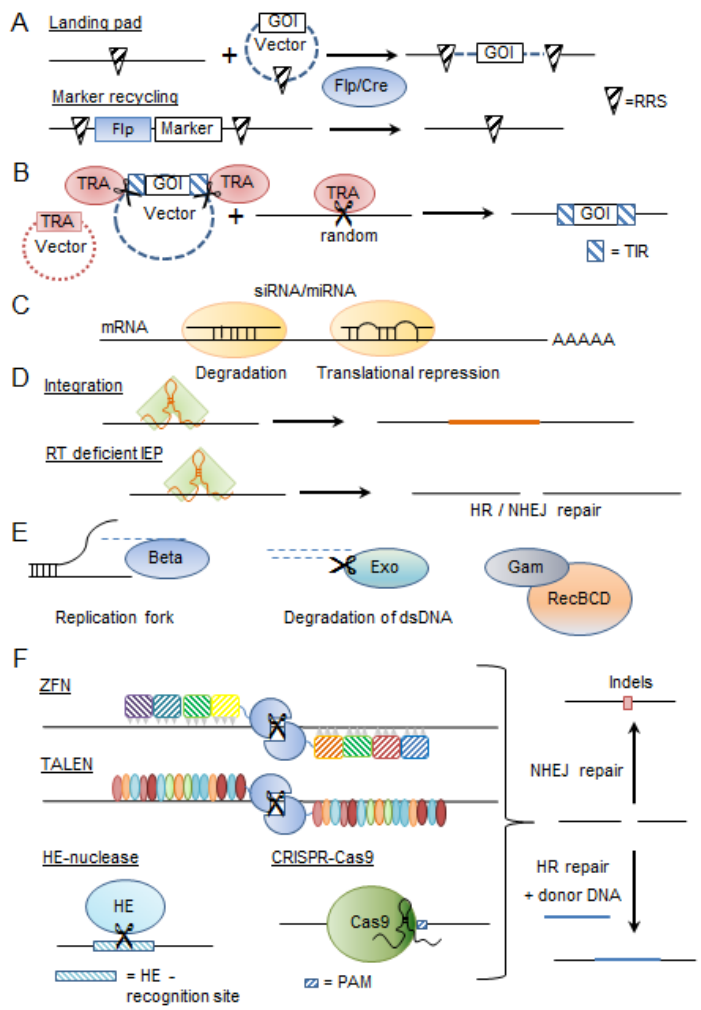


Figure 1: Genome engineering technologies. A – Site specific recombination: The specific recognition site (landing pad) is integrated in the host genome. The flippase/Cre integrase cuts at the recognition site in the genome and on the vector and mediates the intregation of the gene of interest. For selection marker recycling the flippase/Cre integrase is used to remove the selection marker. Some recombination cassettes bear the Flp/Cre under an inducible promoter. Upon induction the Flp gene and the resistance marker gene are exicised. **B – Transpostion:** A vector expressing the transposase gene and a vector bearing the GOI flanked by two transposon recognitio sites (TIR) are transformed in to the cell. The transposase excise the GOI and mediates its integration into the host genome. The intregration occurs in a close to random manner. Recognition sites range from dinucleotides to 4-8 bp palindromic sites. **C – RNAi:** Short, processed miRNA and siRNAs guide the RNA-induced silencing complex (RISC) to bind complementary sequences on the mRNA. The degree of complementarity causes the degradation or the translational repression of the mRNA.

D – Group II introns: The intron encoded protein (IEP) and the intron RNA form a complex, which scans the DNA. First IEP recognizes 5 bp and starts unwinding of the DNA, if basepairing between the RNA and a complementary DNA sequence can take place, the intron RNA is integrated in the genome by the IEP. In case a reverse transcriptase deficient IEP is used the RNA is not integrated in the genome, however a ds break is introduced, which has to be repaired by the cellular homologous recombination or non – homologous endjoining machinery. **E – Recombineering:** The three enzymes Beta, Exo and Gam are involved in lamdba red Recombineering. Beta is a ssDNA binding protein, which anneals the ss donor molecule to the complementary lagging strand near the replication fork. If ds donor DNA molecules are transformed into the host cell one strand is degraded by Exo. Gam inhibits the RecBCD nuclease, which is responsible for the degradation of dsDNA. **F – Genome editing tools:** Zinc finger modules, of which each recognizes three DNA bases, are combined to recognize a DNA sequence of interest. A monomer of the catalytic domain of the *Fok*I nuclease is fused to the zinc finger array. If two arrays bind adjacent DNA sequences, *Fok*I can dimerize and introduce a strand break. TALE repeats recognize a single DNA base. The repeats are combined to a TALE – array and a *Fok*I monomer is fused to the DNA binding domain. If two TALE monomers can bind in near proximity *Fok*I can dimerize and introduce a ds break. Homing endonucleases recognize a 12-40 bp recognition site and introduce a ds break. CRISPR-Cas9 takes advantage of a short guide RNA. The Cas9 nuclease scans the DNA and starts unwinding the strands at the PAM. In case of base pairing between the gRNA and the complementary DNA strand, Cas9 is able introduce a ds break. The strand breaks introduced by genome editing tools can be repaired by homologous recombination, if a homologous donor DNA fragment is present, or by the non-homologous endjoining mechanism. Here, a homologous template is not required and insertion or deletions (indels) can be introduced.

Prokaryotic genome engineering

Genome engineering tools such as Group II introns and recombineering are limited to use in prokaryotes. Although Group II introns have been tested in eukaryotes the efficiency rates are very low. Recombineering has not been applied in eukaryotes, because of the substantial differences in DNA replication.

Group II introns - Targetrons

Group II introns are naturally found in bacterial genomes as well as in *organellar* genomes of some eukaryotes [87] and have become commercially available for targeted genome engineering in prokaryotes ('Targetrons'). The most widely used Targetron is the *Lactococcus lactis* Ll.LtrB group II intron, which consists of a multifunctional intron encoded protein (IEP) and an intron RNA that can be reprogrammed for a desired locus [88]. IEP has reverse transcriptase, RNA splicing and DNA endonuclease activities.

Group II introns are mobile genetic elements that insert themselves into specific genomic sites by a catalytic mechanism called retrohoming, where the target sequence is recognized by specific base pairing with the intron RNA (Figure 1D) [89]. The IEP and the intron form a RNA-protein complex (RNP), which scans DNA and specifically recognizes the target site by interactions of the IEP and the intron. The IEP recognizes a small number of fixed nucleotides (<5) similar to the PAM (protospacer adjacent motif, CRISPR-Cas9) and triggers the unwinding of the DNA, which enables the intron RNA to specifically base pair to the 14–16 bp target sequence [90]. These Group II intron recognition sequences are large enough to be unique even in a complex genome [91]. Thereafter, the intron RNA inserts into one strand of DNA by a reverse splicing mechanism and the IEP cleaves the opposite strand and uses the cleaved 3' end as a primer for reverse transcription of the intron RNA. Subsequently, the host cell repair enzymes integrate the intron cDNA into the genomic target [92]–[94]. Group II introns can be used for two types of DNA modifying reactions in the host genome: On the one hand the site specific integration of the intron RNA and on the other hand the introduction of double strand breaks by reverse–transcriptase deficient IEPs, which are repaired by the NHEJ/HR machinery of the host cell [95]. The integration reaction can also be applied to deliver additional genes within the intron as cargo into the target locus [96], [97].

Insertion efficiencies with antibiotic selection reached almost 100%, whereas without selection disruption frequencies of 1–80% have been obtained [98]. The main advantages of group II introns are the ease of retargeting and the highly site-specific integration. Although they have been primarily used for the modification of prokaryotic organisms (e.g. [95], [96]), a few applications in eukaryotes were reported [99]. Drawbacks for the use in higher organisms are the low efficiency rates, the requirement of high Mg^{2+} concentrations for IEP to be active and the chromatin structure of eukaryotic target DNA, which may impede access of the RNP complex [100]. However, the injection of additional Mg^{2+} noticeably improved the site-specific integration and double strand break induced recombination [99]. With increased efficiency rates, group II intron based systems may also become a valuable tool for eukaryotic genome engineering. They combine the low off-targeting rates similar to TALENs and Zinc-finger nucleases and the flexible reprogramming comparable to CRISPR-Cas9.

Recombineering

Recombineering is used for engineering of bacterial chromosomes by bacteriophage proteins, which promote gene replacement of a linear single or double stranded DNA substrate upon introduction in the host cells [101]. Lambda Red is the most commonly used recombineering system amongst several types reported [102], [103]. Lambda Red recombineering involves three enzymes, Gam, Exo and Beta, of which Beta is the key factor in the recombination process. Beta is a ssDNA binding protein that anneals the ss donor molecule to the complementary lagging strand near the replication fork, which leads to a permanent integration after one round of replication. Exo, a 5' -> 3' exonuclease, is only required for dsDNA integration for the generation of a single strand intermediate. Gam inhibits the endogenous RecBCD nuclease, which is responsible for the degradation of dsDNA, and thereby additionally increases the efficiency of dsDNA integration (Figure 1E) [104], [105]. Lamdba Red integration is independent from endogenous host cell recombination enzymes and requires ss or ds donor substrate sequences with 35 bp to 50 bp identities to the target sequence [103]. Lambda red recombineering can be applied for gene replacements, deletions, insertions, inversions, and point mutations in prokaryotic genomes. The development of recombineering systems is reminiscent of the efficient homologous recombination in *S. cerevisiae* upon transformation with linear DNA fragments [106]. In prokaryotes Lambda red mediated integration of oligonucleotides takes place with frequency rates of ~$5x10^{-4}$ recombinants per viable cell, depending on the oligo length and

concentration [107]. This efficiency is too low for markerless gene insertion and the simultaneous modification of multiple genes [108]. Recombination frequencies have been increased by mutating the host recombination machinery, deleting endogenous nucleases, varying promoters and copy numbers of the Lambda Red genes or engineering the Lambda Red proteins [102], [108]–[110].

Multiplex Automated Genome Engineering (MAGE) by Lambda Red mediated gene replacement has been used for large scale reprogramming and evolution of cells. A pool of oligos with degenerate sequences is repeatedly introduced into cells, which harbor the Lambda Red recombineering mechanism. These oligos are incorporated depending on the similarity to genomic sequences in more than 30% of the *Escherichia coli* cell population every 2–2.5 h. This technique enables an automated, simultaneous modification of different genomic loci in a fast and reliable manner [111]. MAGE was used for instance to insert hexa-histidine sequences into genes coding for the entire translation machinery in *E. coli in vivo*. His-tag purification and reconstitution of the translation machinery allowed fully recovered *in vitro* protein synthesis [112]. Moreover it was applied to insert multiple T7 promoters into several genomic operons to optimize biosynthesis of aromatic amino acid derivatives [113] and for genome wide codon replacement [114].

Strains that were modified by MAGE introduced or that harbor other interesting mutations can easily be combined by Conjugative Assembly Genome Engineering (CAGE). This technique is used to facilitate the large scale assembly of genomes by merging genes of a donor and a recipient strain by bacterial conjugation [114].

Universally applicable genome editing strategies

Homing endonucleases

Homing endonucleases are rare cutting endonucleases, which cleave dsDNA larger than 12 bp and thereby trigger the HR and NHEJ repair mechanisms of the cell [115], [116]. They are encoded by introns, self-splicing inteins and free-standing open reading frames, which is described by the prefixes ‘I-‘ for intron, ‘PI’- for protein insert and ‘F-‘ for freestanding in the nomenclature of homing nucleases [116]–[118]. Intron encoded homing nucleases are considered as selfish genetic elements, because they recognize DNA sequences, which are identical to the sequences up and downstream of the nuclease-encoding intron and thereby promote their own

duplication and perpetuation of genetic information without providing advantages for the host organism [119]. The homing process itself is initiated by a double strand break performed by the homing endonuclease (Figure 1F) [120]. The homing sites, which are specifically recognized by the nucleases, are 12 to 40 bp long [116], [121] and appear, assuming a random sequence organization, once per $7x10^{10}$ bp for an 18 bp sequence. Accordingly, an 18 bp homing site occurs statistically once in a mammalian sized genome [122]. In order to survive, homing nucleases are more tolerant to mutations in the recognition sequence and off-targeting. In contrast to restriction enzymes, minor aberrations to the recognition sequence lead to a decrease in the activity and do not completely abolish DNA cutting [123], [124]. Genetic and protein engineering techniques increased the quantity of available homing nucleases by fusing preexisting nuclease domains to create protein chimeras and by directed evolution to specifically alter the DNA binding domain [125]–[128]. The prefix H- for Hybrid or E- for engineered was introduced to describe these synthetically modified enzymes [129], [130].

Zinc finger nucleases

Zinc finger nucleases (ZFN) are synthetically engineered enzymes, which contain a zinc finger DNA binding domain fused to the non-specific cleavage domain of the type IIS restriction enzyme *Fok*I [131]. The zinc finger binding domain was originally found in the *Xenopus laevis* transcriptions factor IIIA (TFIIIA), which contains tandem-arranged sequences of 30 amino acids (aa) including two pairs of cysteine and histidine residues. Each 30 aa unit binds one Zinc atom using the cysteine and histidine residues and thereby forms a structure reminiscent of a finger, which specifically contacts three DNA bases [132], [133]. Many different zinc finger motifs have been identified with binding affinities towards DNA, RNA, proteins and membrane associated ZF-domains [134]. The *Fok*I nuclease consists of a DNA binding and a cleavage domain. The cleavage domain cuts DNA non-specifically when separated from the DNA binding domain, which is responsible for the specific DNA recognition [135]. The *Fok*I nuclease is only active as a dimer. Cleavage by ZFNs is achieved, if two zinc-finger domains are designed to bind adjacent DNA sequences with a spacer region of ~6 bp in between (Figure 1F) [136, 137], upon dimerization of *Fok*I . Several studies show that the distance between the two ZF binding sites (~6 bp) and the size of the linker, which connects the nuclease and the DNA binding domain, are interdependent on each other [138, 139]. The zinc finger DNA binding domain can be individually designed to bind a sequence of interest by combining single zinc finger modules or

by altering the binding specificity of a given module by exchanging single aa residues [135]–[137]. Typically three to six single zinc fingers of known specificity are linked together to specifically recognize a contiguous DNA sequence. Two three-finger arrays recognize an 18 bp DNA sequence, which is in theory long enough to address a unique locus in a mammalian genome [138]. Until now for nearly all of the 64 possible nucleotide triplets specific ZF domains have been developed, which can be linked together to target a desired genomic sequence [136], [139]–[141]. Not all artificially designed fingers in a ZFN show equal binding affinities towards the target sequence. Thus some ZFNs have affinities for related sequences other than the supposed target and others completely fail to bind the target sequence. To overcome these obstacles, multiple pairs of ZFNs are tested for a single target gene and the number of fingers is often increased to improve the specificity of ZFN binding [138]. In order to reduce the levels of unwanted, mutagenic NHEJ, Zinc-finger nickases (ZFNickases) were created by the inactivation of the catalytic domain of one ZFN monomer. After binding of the ZFN to the target site, *Fok*I dimerizes and introduces a nick in the dsDNA [142]. Further improvements include engineered ZFN variants, which only cleave DNA, when paired as a heterodimer and thereby reduce off-targeting caused by the homodimerization of ZFNs by more than 40% [143] and hyperactivated nuclease variants [144]. A highly efficient *Fok*I nuclease derivate “Sharkey”, which had been created by multiple rounds of cycling mutagenesis and DNA shuffling, showed a more than 15-fold increase in activity compared to the conventional *Fok*I cleavage domain and a three- to six-fold improvement in targeted ZFN mutagenesis [145].

The first genome modification with ZFN was described more than 15 years ago and since then it has remained the most effective and versatile genome editing technique for many years [138]. ZFN technology for the first time allowed programmable genome engineering by combining small, single protein modules. Drawbacks of ZFN include the laborious and time-consuming design of the DNA binding domain and the inaccessibility of some genomic target sites for ZFN cleavage due to the chromatin status. ZFN *de novo* design requires screening of various ZF–libraries, which are often not publicly available [146]. Nowadays TALENs have been widely applied as alternative to ZFNs, providing some advantage: TALENs can target a larger range of sequences, are easier to design, and display higher rates of cleavage compared to ZFN [146], [147].

TALE nuclease technologies

Transcription activator like effector (TALE) mediated genome engineering takes advantage of the DNA binding domain of bacterial effector proteins, which influence the expression of plant genes to aid bacterial infection after injection into host cells. For genome engineering purposes the DNA binding domain of TALEs can be redesigned to bind a desired genomic sequence. The most common TALE applications are TALE nucleases (TALENs), which provide a powerful tool to introduce mutations in the genome of the host organism, and TAL-transcription factors (TAL-TFs), which are used to enhance or silence the expression of a gene of interest [55], [148].

Plant pathogen mediated promoter activation by bacterial TALE proteins was first described in 2007 [149], [150]. Until now more than 100 members of the TALE family have been described, all consisting of similar structural elements [151], [152]: C-terminal nuclear localization signals (NLSs), an N-terminal type III secretion and membrane translocation signal, a transcriptional activation domain (AD) and a DNA binding domain of 1.5 to 33.5 (in most cases 17.5) highly similar repeated domains (Figure 1F) [150], [153], [154]. Each repeat consists of 30–42 nearly identical amino acids, where only the 12^{th} and the 13^{th} amino acid are highly variable. These two amino acids, called "repeat variable diresidues (RVDs)", recognize a single DNA base in the target DNA. Some highly specific RVDs recognize unique DNA bases, whereas others differentiate purine or pyrimidine bases. There exist also non-specific repeats, which bind any DNA base [155], [156]. RVDs are further classified into weak, intermediate and strong according to their efficiency to bind a DNA base [157]. TALE–DNA recognition and activity are influenced by the number and position of DNA–repeat mismatches, the number of repeats present and the specificity of the single repeats [158]. Certain genomic regions are most probably inaccessible for TALE binding because of the condensed chromatin status [159].

TAL effectors for specific gene activation were used with native or heterologous transcriptional activation domains [160]. TALE fusions with VP16 and the VP64 transcription activators of the herpes simplex virus successfully activated transcription in human cells [161], [162]. TAL-TF mediated gene activation can lead to a more than 20-fold increase of the transcription level [160].

Similar to ZFNs, TAL effectors have been fused to nucleases to introduce targeted double strand breaks for genome engineering. *Fok*I is the most prominent nuclease used for TALEN research [18], [163], [164]. Since *Fok*I is only functional as a dimer, two specific TAL effectors are

required to bind opposing sites of the target DNA in a certain distance, which allows *Fok*I to dimerize and to introduce the strand break. The optimal length of this spacer region is between 10 to 30 bp and depends on the TALEN architecture and the number of repeats [18], [164]–[166]. The requirement for two TALE domains, which bind in close proximity and enable the dimerization of *Fok*I, causes minimal off-target effects and thereby reduces the risk of cytotoxicity [167].

One of the most laborious and time consuming aspects is the assembly of TALE repeats, because of their highly repetitive nature. Several hierarchical ligation strategies have been described and optimized by different research groups in terms of time efficiency, specificity and overall convenience [168]–[170]. A ligation independent method for high throughput TALE assembly takes advantage of 10–30 bp long, non-palindromic, single stranded overhangs, which anneal in a highly specific manner [171]. Automated assembly makes TALENs fast and easily available, but is accompanied by relatively high costs [172]. Independently from the assembly method for reprogramming TALENs, retargeting CRISPR-Cas9 can be achieved in a much faster and cheaper way. The large size of TALENs as well as ZFNs may limit their delivery by size restricted vectors [55]. Forthcoming research may also include TALE fusion proteins with several domains such as methyl- and acetyltransferases, deacetylases or deaminases [173].

CRISPR–Cas9 technology

Since the discovery of clustered regularly interspaced short palindromic repeats (CRISPR) in 2007 ([174], [175]) and the RNA-guided DNA nuclease Cas9 in 2012 ([17], [176]), the number of publications has exploded suggesting a new standard strategy for genome editing. CRISPR-Cas9 uses a short, non-coding, ‘guide’ RNA, to direct a nuclease to specifically cut a DNA sequence of interest. In contrast to above described genome editing techniques, the genomic target locus can be changed by simply varying a 20 bp sequence of the guide RNA, instead of cumbersomely reprogramming DNA binding domains, giving researchers high flexibility and an accelerated throughput rate [177].

Natural function in adaptive immunity in bacteria

This method originates from bacteria and archaea, where it provides an adaptive immunity against invading nucleic acids such as phage or plasmid DNA [178], [179]. The term CRISPR describes a genomic locus consisting of multiple, short, palindromic sequences (typically 24–37

bp), which are interspaced by short sequences of foreign origin called spacers (27–72 bp) [180], [181]. Usually more than one CRISPR locus is found in a bacterial or archaeal genome [182]. These CRISPR loci are flanked by a 300–500 bp leader sequence on one side and a set of CRISPR associated (Cas) genes are located adjacent to the CRISPR array. The function of the leader sequence is suggested to on the one hand to promote the transcription of the CRISPR array and on the other hand to play an important role in the recognition and acquisition of incoming spacers [183]. Until now there were three CRISPR–Cas systems identified (Type I, II, III), which are further divided into various subtypes [184], [185].

CRISPR-Cas9 based targeted genome editing

Engineered CRISPR-Cas9 nowadays serves as one of the most promising, advanced methods in targeted genome editing. In contrast to other CRISPR types, which require multiple proteins, the endonuclease Cas9 from the type II A CRISPR system is solely responsible for DNA recognition, binding and cleavage. Cas9, originally isolated from *Streptococcus pyrogenes,* uses two separate nuclease domains each of which cuts one DNA strand. Deletion of the catalytic activity of one nuclease domain leads to nicking of the DNA (Cas9 nickase) [17], [186]. For genome engineering purposes the Cas9 endonuclease and a short artificially designed guide RNA (gRNA) are co-expressed in the host cells (Figure 1F). The gRNA is expressed from a single transcript, which consists of an 80 bp structural part and a 20 bp variable part. The structural part of the gRNA folds into a stem loop, whereas the variable part binds to the complementary target DNA. Upon binding of Cas9 to the DNA, the strands get unzipped at a sequence called protospacer adjacent motif (PAM) and basepairing of the gRNA to the complementary DNA strand takes place (Figure 2). Sequence specific cleavage of Cas9 is dependent on the PAM as well as on the complementarity between the gRNA and the DNA target [17]. The CRISPR-Cas9 target regions and the design of gRNAs are restricted to the PAM sequence, since DNA unwinding takes only place at the PAM. *Streptococcus pyrogenes* Cas9 tolerates the nucleotides NGG and NAG as PAM [187], but PAM sequences differ among the various CRISPR types and Cas proteins [188], [189].

Multiplexed genome engineering can easily be achieved by expressing several gRNAs on a single vector [190], [191]. CRISPR-Cas9 mediated transcription regulation (CRISPRi) was achieved by the use a catalytically inactive Cas9 mutant, which is fused to transcriptional activation and repression domains in eukaryotes [192], a strategy similar to TAL-TFs.

A main drawback of the CRISPR-Cas9 system is the high off-targeting rate. Cas9 off target sites can harbor up to five mismatches and the cleavage rates are similar to the on-target region [193], [194]. Single-base mismatches up to 11 bp upstream of the PAM completely abolish DNA cleavage, whereas mismatches further upstream retain Cas9 activity [190]. Undesired off-targeting can be reduced by the use of the Cas9 nickase and truncated gRNAs [195], [196]. Dimeric Cas9–*Fok*I fusion proteins reduce the likelihood that a suitable target site will occur more than once in a mammalian-sized genome. These dimeric RNA-guided FokI nucleases (RFN) combine the ease of reprogramming of the CRISPR-Cas9 system and precise targeting similar to TALEN or ZFN [197].

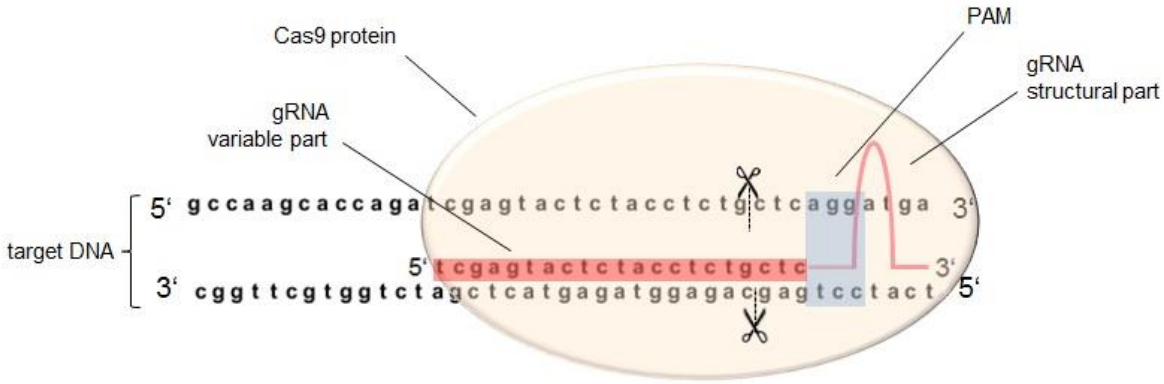


Figure 2: Essential parts of the CRISPR-Cas9 system. The nuclease Cas9 introduces a double strand break upon base pairing of the gRNA to the complementary DNA sequence. The cleavage position is located 3 bp upstream of the PAM.

Genome engineering techniques for synthetic biology are used for a broad range of applications starting from basic research use in biotechnology up to medical applications. Their huge potential is confirmed by a rapidly growing list of organisms and targeted genes. However, continuous improvements and diversification of currently available techniques are required to overcome their limitations and bottlenecks. Future prospects might include further improvements of existing technologies by increasing their specificities, reducing toxicity effects as well as finding novel technologies and combinatorial methods. The collection of methods for genome engineering mentioned here may serve as a guideline to identify the most appropriate method to address a certain application of interest. Due to their outstanding characteristics compared to other genome modification tools in terms of the flexibility of available target sites, the ease of retargeting and their potential to introduce targeted modifications, CRISPR-Cas9 and TALENs are the most promising tools for targeted genome engineering in eukaryotes. The focus of this master thesis lied on establishing these methods for enhanced genome engineering in *P. pastoris*.

Table 1: Detailed comparison of genome engineering methods

Genome engineering tool	Capacities	Target binding principle	Target length	Multiplexing	Targeting frequency	Programmable*	Off targeting	Advantages	Drawbacks	Improvements/ Combinatorial methods	Applications	Refer-ences
Site – specific recombinases	Integration, excision, inversion and translocation relying on pre-integrated recognition sites Selection marker recycling Landing pads and RMCE	Recombinase recognizes a specific recognition site	34 bp	-	High	No, the recombinase is restricted to the recognition sequence	Minor effects	Not dependent on host cell (co-) factors or DNA replication	Recombination only occurs at pre-existing recognition sites	Combined with Group II introns in prokaryotes [28]	Prokaryotes Eukaryotes	[16], [19], [24], [31], [33]
Transposons	(Close to) random mutagenesis Used for gene delivery as alternative to viral vectors Random mutagenesis	Transposons excise sequences, which are flanked by defined recognition sequences (TIR) Non – specific (re)integration	Depending on transposase either random, dinucleotides, or short palindromic consensus sequences	-	Low	Moderate	Wildtype enzymes integrate close to random; Fusion proteins to ZF and TALEN DNA binding domains also show high off targeting	Collection of transposons with differences in cargo capacity, integration preferences and species specificity Useful for random mutagenesis,	Due to high off targeting rate little suitable for targeted genome engineering	Fusions to Zinc finger DNA binding domains [49],[50] and TALE DNA binding domains allow targeted genome engineering [51]	Prokaryotes Eukaryotes	[35], [37], [38], [39], [40], [41], [42], [52], [53], [54]
Group II introns	Targeted integration Induction of ds breaks and host cell repair (HR, NHEJ)	Specific protein and RNA recognition	Predefinded 5 bp are required by the protein and 14 – 16 bp RNA binding sequence	-	Variable	High, but the target sites are limited to protein recognition sites (5 bp)	No, minor effects	Ease of retargeting	Low efficiencies in eukaryotes	Combined with transposons [28] Reverse transcriptase deficient Group II introns with nuclease activity	Prokaryotes	[95], [96], [97]
Recombineer-ing systems	*In vivo* cloning technique Insertions, deletions, point mutations, gene replacement	ss donor DNA fragment is annealed to complementary sequence during DNA replication	35 – 50 bp (length of donor fragment)	Yes, MAGE, [111] CAGE [114]	High	High	No	Independent from host cell recombination	Negative effect on host cell replication, Limited size of the inserted DNA, Dependent on host cell replication	MAGE, [111] CAGE, [114]	Prokaryotes	[101], [107], [108]
RNAi, Translational repression	Knock down of genes at posttranscriptional level	RNA guided protein complex binds to mRNA	21 – 23 bp	Yes	High	High	High, binding to similar RNA sequences	Fast and easy reprogrammable	No knock – outs, incomplete knock – downs, dsRNA can active various cellular pathways	-	Higher eukaryotes	[63], [76], [77], [78], [79], [80]

Homing endonucleases	Introduction of ds break triggers host cell repair (HR, NHEJ)	Protein recognizes specific DNA target site	12 – 40 bp	-	High	Low	Minor effects		Limited capacity of available target sites, engineering for new target sites is complicated and time consuming	Fusions of preexisting nuclease domains and specifically altered DNA binding domains [125], [126], [127], [128].	Prokaryotes Eukaryotes	[115], [122], [123], [124]
Zinc finger nucleases	Induction of ds or ss breaks and host cell repair (HR, NHEJ)	Assembly of single protein modules, of which each recognizes three DNA bases	2 x 18 bp + spacer (~6 bp)	-	Variable, depends on target locus	Moderate	Low - medium	Focus of research for many years, findings were quickly adopted to TAL technologies	Complicated reprogramming requires protein assembly and engineering	ZF – nickase [147], Heterodimeric ZFN [148], Hyperactive nuclease variants [149], [150]	Prokaryotes Eukaryotes	[141], [144], [145], [146]
TALEN	Induction of ds or ss breaks and host cell repair (HR, NHEJ)	Assembly of protein modules, of which each recognizes one DNA base	2 x 17 bp + spacer (~15 bp)	-	Variable, depends on target locus	High	No, minor effects	Almost every locus is targetable	Large protein size Moderate construction time	Fusions to transcription activation and repression domains (TAL-TF) [165],	Prokaryotes Eukaryotes	[170], [173], [174], [175]
CRISPR-Cas9	Induction of ds or ss breaks and host cell repair (HR, NHEJ)	RNA guided protein complex binds to DNA	20 bp + PAM (3 bp)	Yes	Variable, depends on target locus	High, restricted to PAM	High	Fast and easily reprogrammable	Large protein size, several gRNAs tested for one locus	*Fok*I– Cas9 fusion protein [201]	Prokaryotes Eukaryotes	[17], [181], [194]

* altering of the target site

2. OBJECTIVES

The creation of knockout strains in *P. pastoris* is currently a laborious and time-extensive procedure, which may require screening of up to several thousand transformants to obtain the desired mutation variant ([198], personal communication Ahmad M.). In contrast to *S. cerevisiae*, where targeting frequencies of 100% are easily achievable due to highly efficient HR, gene replacement events in *P. pastoris* occur with a frequency of 0.1% up to 30% depending of the target locus and the length of the homologous fragments [199]–[201]. Targeted single and double strand break induced DNA repair, however, increases the rate of homologous recombination by several orders of magnitude (up to 4000–fold in *S. cerevisiae*) [10]–[13]. Therefore systems introducing breaks at programmable positions in the *P. pastoris* genome are of great demand.

The aim of this master thesis was to identify genome engineering tools, which enable the fast and convenient generation of *P. pastoris* knockout and deletion strains. CRISPR-Cas9 and TALEN are nowadays the most promising tools for the modification of eukaryotic genomes. However, none of these techniques has been tested in *P. pastoris* yet. The targeting specificity of Cas9 nuclease can easily be reprogrammed by changing a 20 bp sequence of the gRNA, which allows researchers high flexibility and accelerated throughput rate. The thesis should result in a CRISPR-Cas9 system, which is functional in *P. pastoris* and can be used for enhanced knockout strain generation, strain characterization, to facilitate site specific integration of long DNA fragments into the genome and to replace specific genomic DNA regions efficiently.

Because of the highly repetitive nature of the TAL repeats TALEN assembly is very difficult and reprogramming requires enhanced skills in the field of protein engineering. Nevertheless it is of scientific interest to show, if TALEN based genome engineering is possible in *P. pastoris* and if the targeting rates differ compared to CRISPR-Cas9 induced mutation frequencies.

Background and concepts of this thesis

Significance and selection of features for CRISPR-Cas9 mediated targeted genome engineering in *P. pastoris*

The aim of the thesis was the construction and evolution of a CRISPR-Cas9 based system for targeted genome engineering in *P. pastoris*. In *Saccharomyces cerevisiae*, DiCarlo *et al*. [202] could simply employ previously reported parts (promoters, NLSs) from earlier CRISPR-Cas9 publications to prove that this technique is applicable. In *P. pastoris* barely any parts such as NLSs or RNA Polymerase III promoters have been described yet. These elements are however required for efficient of use of CRISPR-Cas9, as outline below.

Therefore my goal was to screen all essential parts of the vector and thereby to identify bottlenecks and key factors in order to develop a system for improved genome engineering in *P. pastoris*. These features include various (codon optimized) Cas9 DNA sequences, different gRNA sequences, RNA Polymerase III and RNA Polymerase II promoters for the expression of the gRNAs and RNA Polymerase II promoters for the expression of Cas9. Moreover I identified a variety of nuclear localization signals (NLSs), which are efficiently recognized in *P. pastoris* and I compared various autonomously replicating sequences (ARSs) and selection markers for their influence on plasmid maintenance and protein production.

The nuclear import of Cas9

The chapters describing NLSs in P. pastoris were prepared in shortened form as a manuscript "A toolbox of endogenous and heterologous nuclear localization sequences for the methylotrophic yeast Pichia pastoris" for submission in the journal FEMS Yeast Research (submission date: July, 2015).

NLSs are required for the import of proteins in the nucleus of eukaryotes [203]. Many proteins from bacteria or bacteriophages used for basic studies in molecular biology to generate synthetic genetic circuits or for genome editing applications such as Cas9 require efficient NLSs to function in eukaryotes. *P. pastoris* is a widely used expression platform for heterologous protein production, but molecular tools such as NLSs are limited.

NLSs are required for the active transport of proteins, which are bigger in size than 40 kDa through nuclear pore complexes (NCP) [203], [204]. These NCPs are located in the nuclear

envelop, which separates the cytoplasm from the nuclear compartment in eukaryotic cells. NCPs allow the passive diffusion of small proteins (< 40 kDa), metabolites and ions, but restrict the diffusion of large proteins, which do not contain a functional import signal [205]. The best characterized nuclear transport system is the classical nuclear import pathway, where soluble carrier proteins called karyopherins (Kap) mediate the transport of the protein cargo into the nucleus. Lange *et al.* suggested that in *S. cerevisiae* 45% of the all cellular proteins have the potential to enter the nucleus via the classical nuclear import pathway [206]. In *S. cerevisiae* Kap60 (also known as importin α) recognizes classical NLSs in the cytoplasm and further interacts with Kap95 (importin β). Kap95 drives the transport of the trimeric complex (Kap60 + Kap95 + protein) through the NCP into the nucleus by specifically interacting with the NCP. In the nucleus RanGTP binds to Kap95 and thereby causes the dissociation of the transport complex. Kap60 is recycled back into cytoplasm by the aid of the export receptor Msn5p, which is complexed with RanGTP [207], [208]. The export of Kap95 is mediated by the carrier protein XPO1 (formerly CRM1), which is also a key player in nuclear export signal mediated (NES) protein and mRNA export [209]. A classical NLS, which is recognized by Kap60, consists either of one (monopartite NLS) or two clusters (bipartite NLS) of basic amino acids [210], [211]. A prototype monopartite NLS is represented by the Simian Virus 40 (SV40) large T antigen NLS [212] with the sequence PKKKRKV. A consensus sequence for monopartite NLSs was identified, where a lysine residue is followed by two other basic residues resulting in a K(K/R)X(K/R) motif [213] [214]. Whereas two cluster NLSs comprise two basic amino acids followed by a spacer region of ten to twelve non-defined amino acids and a monopartite-like, basic stretch of five amino acids, of which three must show basic characteristics. Bipartite NLS are exemplified by the *Xenopus leavis* derived nucleoplasmin NLS with the sequence KRPAATKKAGQAKKKK [210], [211], [215].

Classical nuclear import and NLS in P. pastoris

In *P. pastoris* solely the well characterized SV40 NLS has been applied for the import the prokaryotic bacteriophage T7 RNA polymerase in the nucleus by purpose [216]. Additional NLSs, which are functional in *P. pastoris*, have been identified haphazardly by different groups, when trying to improve heterologous protein production. Yang *et al*. [217] obtained eight-fold higher expression levels for the production of secreted human topoisomerase I after they removed the NLS. Gradoboeva *et al*. [218] described that a nuclear localization signal provided the

translocation of a recombinant bovine gamma interferon into the *P. pastoris* nucleus, whereas the absence of this sequence led to the cytoplasmic accumulation of the recombinant protein. None of these NLSs has been further characterized for the nuclear import of heterologous proteins.

Heterologous NLSs from literature

During this thesis I tested various heterologous NLSs (Table 2) for the nuclear import of an enhanced green fluorescent reporter protein (eGFP) in *P. pastoris*. These sequences have been characterized in various higher eukaryotes and several yeasts for the nuclear import of proteins. The NLS of the large SV40 T antigen [212] has been successfully used in *S. cerevisiae* [219], *Schizosaccharomyces pombe* [220] and *P. pastoris* [216]. The translocation signal of the human Myc protein (*Hs*Myc) contains a monopartite consensus motif similar to SV40-NLS and has been characterized for the import of heterologous proteins in various mammalian cell lines [221], [222]. Crystallographic analysis revealed that *Hs*Myc interacts with a key factor of the *S. cerevisiae* nuclear import pathway [214]. The NLS of the *X. leavis* derived nucleoplasmin protein (*Xl*Nuc) is the most prominent bipartite classical NLS and has been successfully used in *S. cerevisiae* [214] and *S. pombe* [223]. The non-classical NLS of the *S. cerevisiae* transcription repressor *Sc*Matα2, has been successfully used for the import of an *E. coli* beta-galactosidase to the *S. cerevisiae* nucleus [224]. The functionality of import sequences, such as the NLS of the *S. cerevisiae* transcription factor *Sc*SWI5, is regulated by a phosphorylation dependent mechanism [225], leaving the question if such a NLS is also functional in a non-natural host.

Table 2: Heterologous NLSs used in this thesis.

NLS - Internal name	**Protein name**	**Source**	**NLS - Protein sequence**	**DNA sequence codon optimized for *P. pastoris***	**References**
SV40	**SV40 large T antigen**	Simian Virus 40	PKKKRKV	CCAAAGAAGAAAAGAAAAGTT	[212]
*Sc*Matα2	**Matα2**	*S. cerevisiae*	KIPIK	AAGATTCCAATTAAG	[224]
*Hs*Myc	**c-Myc**	*H. sapiens*	PAAKRVKLD	CCAGCTGCTAAGAGAGTTAA GTTGGAT	[221]
*Xl*Nuc	**Nucleoplasmin**	*X. leavis*	KRPAAATKKAGQAKKKK	AAGAGACCTGCTGCTGCCAC TAAGAAAGCAGGGCAAGCTA AGAAGAAGAA G	[214]
*Sc*SWI5	**SWI5**	*S. cerevisiae*	KYENVVIKRSPRKRGRPRK	AAGAAGTACGAAAACGTTGTTATC AAGAGATCCCCAAGAAAGAGAGG UAGACCAAGAAAA	[225]

Endogenous NLSs

I also tried to identify endogenous NLSs, which might have beneficial import characteristics compared to the heterologous import sequences. I selected five NLSs of putative nuclear proteins in the genome sequence of *P. pastoris* [226], according to the published consensus motifs. The *P. pastoris* (*Pp*) proteins were used as queries for a BLAST search in *S. cerevisiae*, demonstrating that they are homologs of the *S. cerevisiae* nuclear proteins Nob1p, Sda1p, Set7p and Uba1p [227]. The detailed function of these proteins is listed in Table 3. The putative NLSs of these proteins contain the bipartite consensus motif of NLSs for the classical import pathway and are located on the C-terminus. Additionally, I also included a NLS from the *P. pastoris* homolog of *Sc*SWI5.

Table 3: *P. pastoris* endogenous NLSs. NLSs of putative nuclear proteins were selected in the genome sequence of *P. pastoris* [226], according to the published consensus motifs [228]. The *P. pastoris* (*Pp*) proteins were used as queries for a BLAST search in *S. cerevisiae*, demonstrating that they are homologs of the *S. cerevisiae* nuclear proteins.

Abbreviation	**Genbank accession number**	***S. cerevisiae* homolog**	**Nuclear import motif**	**Homologous protein function in *S. cerevisiae* (from Saccharomyces Genome Database** [227]**)**	**p-Blast results**
*Pp*Nob1	XM_002493229	**Nob1p**	KGRRANASKKKK	Protein involved in proteasomal and 40S ribosomal subunit biogenesis; required for cleavage of the 20S pre-rRNA to generate the mature 18S rRNA;	Query coverage: 98% E value: 2e-114 Identity: 42%
*Pp*Sda1	XM_002490388	**Sda1p**	KQKVLRAHIDKQKKKGH	Protein required for actin organization and passage through Start; highly conserved nuclear protein; required for actin cytoskeleton organization; plays a critical role in G1 events;	Query coverage: 100% E value: 0.0 Identity: 55%
*Pp*Set7	XM_002491385	**Set7p**	KRKLEEEEGSKRNKRIKG	Ribosomal lysine methyltransferase; specific for monomethylation of Rpl42ap and Rpl42bp (lysine 55); Location nucleus	Query coverage: 98% E value: 3e-96 Identity: 39%
*Pp*Uba1	XM_002490958	**Uba1p**	KRPLEIEQEETYSKRKKSTI	Subunit of heterodimeric nuclear SUMO activating enzyme E1 with Aos1p; activates Smt3p (SUMO) before its conjugation to proteins (sumoylation), which may play a role in protein targeting; essential for viability	Query coverage: 32% E value: 1e-62 Identity: 61%
*Pp*Swi5	XP_002489440	**Swi5p**	KKFVRNHDLRRHKKK	Transcription factor that recruits Mediator and Swi/Snf complexes; activates transcription of genes expressed at the M/G1 phase boundary and in G1 phase; required for expression of the HO gene controlling mating type switching; localization to nucleus occurs during G1;	Query coverage: 21% E value: 1e-33 Identity: 52%

The influence of different autonomously replicating sequences (ARS) and antibiotic resistance markers on protein production and plasmid maintenance

Upon introducing the targeted genomic modification, the CRISPR-Cas9 system is no longer required in the cell and becomes an unwanted burden possibly interfering with cellular processes. In *P. pastoris*, usually linear plasmids for expressing recombinant proteins are transformed into the cells and get stably integrated in the genome. However, the integration of the CRISPR-Cas9 system is not desired. ARSs, which allow plasmids to be maintained in an episomal form, offer an alternative to the genomic integration of exogenous DNA [229]. Plasmids, containing an ARS, are not stable in the absence of selective pressure and get rapidly lost. In *P. pastoris* for several decades the PARS1 and PARS2 sequences [230] had been used. Recently ARSs have been mapped by deep sequencing methodology [231]. Also a pan ARS sequence for use in different yeasts has been reported to be functional in *P. pastoris* [232]. Yet these ARS sequences have not been directly compared for their influence on protein expression and plasmid maintenance.

Mechanistically, DNA replication in eukaryotes starts at replication origins by the recognition and binding of the Origin Recognition Complex (ORC) in an ATP dependent reaction. The multisubunit complex recruits initiator proteins that unwind the DNA and priming and elongation are conducted using the replication origin as a starting point [233], [234]. ARSs are defined as DNA sequences that serve as replication origins on extrachromosomal plasmids in yeast [235]. A common feature of all yeast ARSs is the ARS consensus sequence (ACS), which can be used to predict replication origins and ARSs [236], [237]. Although the consensus motif is universally present the sequence differs among single yeast species. The ASC is flanked by essential elements, which are involved in replication factor binding. Mutations in the ACS lead to loss of function, whereas mutations in the essential flanking elements can lead to a reduction in ARS efficiency and affect the binding efficiency of the ORC [238], [239].

In *S. cerevisiae* 100 bp can already function as ARSs [240], [241], whereas in *P. pastoris* the known essential functional regions have at least 200 bp [231]. The first ARSs described for *P. pastoris* are PARS1 and PARS2 [230], which allow the autonomous replication of plasmids for more than 50 generations. Regardless of their sequence similarity to *S. cerevisiae* ARSs PARS1 and PARS2 did not show activity in baker's yeast. In all budding yeast species A/T – rich origins are predominant, however in *P. pastoris* approximately one third of all origins contain a G/C-rich motif. G/C-rich ARS are found to be connected to early replicating origins and tend to be

promoter-associated, contrary to A/T-rich ARSs, which have termination activity. A detailed study of the location, structure and dynamics of *P. pastoris* replication origins was profiled by Liachko *et al*. [231]. This group also described an artificial ARS "panARSOPT", which is active in ten species of budding yeast [232].

In the course of the thesis, I compared three ARSs in order to identify a sequence, which is most suitable for the vectors bearing a CRISPR-Cas9 system. I tested the well-characterized, A/T-rich PARS1, the universally applicable panARSOPT and the G/C-rich "B1739" ARS reported by Liachko *et al*., judging from the colony size in their publication, on plasmids expressing a red fluorescence protein variant reporter gene (sTomato) under the constitutive *GAP* promoter in order to identify differences in the expression levels of the heterologous protein and in plasmid maintenance. In addition to the influence of ARSs, the protein production levels and the cell viability are also influenced by the selection marker (personal communication Thomas Vogl). Thus I tested two selection markers, either Zeocin or Geneticin, in combination with the different ARS-plasmids.

Cas9 based targeted genome engineering

CRISPR-Cas9 systems for genome engineering employ the nuclease Cas9, which is beside the gRNA the only factor required for DNA recognition, binding and cleavage [242]. The streptococcal Cas9 sequence (*Sp*Cas9) was the first Cas9 sequence efficiently used for genome engineering in bacteria [17]. The *Sp*Cas9 sequence was codon optimized for genome editing in e.g. human cells [243], mice [244], or zebrafish [245]. A human-codon optimized version of *Sp*Cas9 was also successfully used in *S. cerevisiae* [202].

Codon optimization can have a positive effect on the production of heterologous proteins in *P. pastoris* [246], [247], since the native sequences from other organisms may show a different codon usage bias. However, high level overexpression especially of transcription factors and proteins interacting with DNA can have a negative influence on the cell [248], [249]. Thus, it is beneficial to start with a broad range of sequences to identify one, which is functional in the desired host. In this thesis I tested the wild-type *Sp*Cas9 sequence and two codon optimized versions of the *Sp*Cas9 gene. The human-codon optimized *Sp*Cas9 sequence (*Hs*Cas9) [243] was successfully used in various human cell lines [243] as well as in *S. cerevisiae* [202]. Moreover I optimized the *Sp*Cas9 for expression in *P. pastoris* (*Pp*Cas9). The sequences differ in the GC-

content and the codon bias, they contain various secondary structure motifs and might be translated at different rates.

The expression of gRNAs

The gRNAs for CRISPR-Cas9 applications consist of a 20 bp variable sequence and an 80 bp structural component (stgRNA), which folds into a stem loop like structure [17]. In order to get the gRNA correctly incorporated into Cas9, gRNA transcripts should neither be processed nor directed to the cytoplasm. Additional (extraneous) RNA sequences 5' or 3' of the gRNA sequence might also have a negative influence on the functionality and the targeting rate. Transcripts of RNA polymerase II (generating mRNAs) contain a 5'-7-methylguanosine cap and a 3' polyA tail. The 5'-7-methylguanosine cap is responsible for ribosome recruitment and mRNA translation, whereas the polyA sequence is necessary for transcription termination. Both 5' cap and polyA tail are important for the export of the mRNA to the cytoplasm [250], [251]. Thus, alternative RNA polymerase promoters are usually better suited for the expression of gRNAs.

RNA Polymerase III promoters, which naturally direct the expression of 5S rRNAs, tRNAs and small non-coding RNAs [252], have been used for CRISPR-Cas9 applications in various eukaryotes. In contrast to RNA Polymerase II the multisubunit RNA polymerse III is missing a domain called CTD, which is responsible for the recruitment of CAPing and polyadenylation enzymes [253], [254]. RNA Polymerase III promoters have been divided into three classes, depending on the organization of the regulatory sequence elements. Type 1 and type 2 promoters are gene-internal (intragenic), whereas type 3 promoters, which have only been reported in higher eukaryotes, are located upstream of the transcription start (extragenic) [252]. Similarly, transcripts of the RNA polymerase I (all rRNAs except the 5S rRNA) are neither CAPed nor polyadenylated [255] and promoters, which are recognized by RNA polymerase I, would also be applicable for gRNA expression. However, little information about RNA polymerase I promoters is currently available. Another disadvantage of using RNA Polymerases I promoters is that the RNA polymerase I requires a complex termination sequence similar to RNA Polymerase II, whereas the RNA Polymerase III terminates after encountering a short polyT stretch [255], [256].

The extragenic RNA polymerase III promoter *U6* is most frequently used for CRISPR-Cas9 gRNA expression in higher eukaryotes [190], [196], [243]. In *S. cerevisiae* various intragenic RNA polymerase III promoters have been tested, which show great differences in the gRNA

expression levels and the targeting efficiency. The RNA polymerase III promoter *SNR52* has been used for several CRISPR-Cas9 applications in yeast and human cells [192], [202] and for the expression of exogenous tRNAs in *S. cerevisiae* [257]. Interestingly, the promoter was functional for gRNA expression in the diploid strain S288C, but did not trigger gRNA expression in the polyploid strain ATCC4124 [258]. Ryan *et al*. [258] tested several tRNA sequences as gRNA promoters and they observed that the expression and further the targeting efficiency was depending on the stain context. The tRNA promoter $tRNA^{Pro}$ enabled efficient gene targeting in the polyploid strain ATCC4124, but was inactive in the diploid strain S288C, whereas the opposite effect was reported for the tRNA promoter $tRNA^{Tyr}$. The targeting efficiencies at the *URA3* locus of almost 100% were obtained with tRNA promoters in the diploid S288C strain, whereas the RNA polymerase III promoters *SNR6*, *SCR1* and *RPR1* hardly directed functional gRNA expression. Farzadfard *et al*. [259] successfully tested the endogenous *RPR1* promoter [260] for the expression of gRNAs for CRISPR mediated transcription regulation in *S. cerevisiae* W303. The *SUP4* and the *RPR1* promoter from *S. cerevisiae* were used for the expression of short 5S rRNA in *S. cerevisiae* [261].

In the course of this thesis I tested ten RNA polymerase III promoters for the expression of non-coding gRNAs. The regulatory sequence elements, which are recognized by the RNA Polymerase III in yeast are located in the gene sequence. The gene sequences used as promoters encode for a various RNAs spanning from rRNAs, tRNAs or RNAs, which are associated with riboproteins. In order to achieve an expression of functional gRNAs, the transcripts have to remain in the nucleus or have to be processed in a way that the gRNA is cleaved off prior to the translocation to the cytoplasm. Assuming the nuclear localization of the gRNA, its conformation must not be influenced by 5’ and 3’sequences in a way that Cas9 and the non-coding RNA are unable to assemble correctly. I started with a broad range of seven exogenous and three endogenous RNA polymerase III promoters. The *SNR52*, *RPR1* and *SCR1* promoters have either been characterized in literature for the expression of various short RNAs or have been successfully used for CRISPR-Cas9 genome engineering. I additionally tested homologues of the *S. cerevisiae SN52* and the *RPR1* promoters from different yeast species [262]. Due to the lack of basic research on RNA Polymerase III promoters in *P. pastoris* I placed three different endogenous tRNA sequences (P_{LYS}, P_{MET}, P_{SER}) and one exogenous tRNAs (P_{ASN}) as promoters 5’ to the gRNA sequence similar as described by Wang *et al*. [257]. Considering the various factors influencing

the functionality of gRNA and the lack of information available, identifying a RNA polymerase III promoter system was one of the most critical factors for establishing a CRISPR-Cas9 system in *P. pastoris*.

Ribozymes in CRISPR-Cas9 applications

An alternative approach for the expression of gRNAs is the use of RNA polymerase II promoters in combination with ribozymes. Ribozymes are a diverse group of catalytic RNA molecules, which catalyze RNA and DNA cleavage or ligation reactions or peptide bond formation within the ribosome [264], [265]. Gao and Zhao [263] expressed gRNAs, which are flanked by the Hammerhead ribozyme (HH, [266]) at the 5' end and the hepatitis delta virus ribozyme (HDV, [267]) on the 3' end, in *S. cerevisiae* (Figure 3). Both ribozymes induce sequence specific cleavage of a phosphodiester bond [268]. The primary transcript undergoes a ribozyme mediated cleavage resulting in a mature 100 bp gRNA. HH is a 5' self-cleaving enzyme, which stays attached to the 3' product, whereas the 3' self-cleaving enzyme HDV ribozymes remains bound to the 5' end of the RNA. In contrast to the HDV ribozyme six nucleotides of 5' cleaving HH ribozyme have to be altered according to the variable 5' sequence of the gRNA. Ideally, the function of the 5' cleaving ribozyme would be independent from the adjacent RNA sequence, but such ribozymes have not been reported in literature yet [263]. Similarly, Nissim *et al.* [269] successfully tested the RNA polymerase II promoter *CMV* for the expression gRNAs flanked the HH- and HDV ribozyme in human cells. Ryan *et al.* [258] applied the HDV ribozyme in combination with RNA polymerase III promoters. Here the ribozyme was fused 5' to the gRNA sequence for removing the promoter sequence (= tRNA gene). Although the ribozyme stayed attached to the gRNA CRISPR-Cas9 targeting efficiencies of up to 100% were reached when using a tRNA promoter.

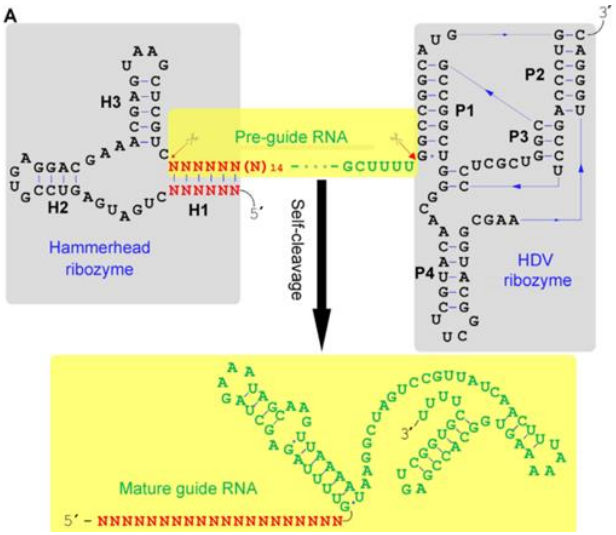


Figure 3: gRNA flanked by the 5' cleaving HH and the 3' cleaving HDV ribozyme. The variable 20 bp part of the gRNA is colored in red and the 80 bp structural part is coloured in green. Six bases of the HH ribozyme (red) have to be adapted according to the variable gRNA sequence to permit ribozyme cleavage (taken from [263]).

In addition to the various constructs containing an RNA polymerase III promoter for the expression of the gRNA, I also expressed gRNAs flanked by the HH and the HDV ribozyme as described by Gao and Zhao [263] under the control of the RNA polymerase II histone promoter *HTX1*.

Genomic target loci for testing CRISPR-Cas9 based genome engineering

CRISPR-Cas9 can be either used to introduce insertion or deletions (indels) or to induce HR-dependent cell repair [270]. A vector, bearing Cas9 and a gRNA, is transformed in the host cells, where the nuclease will introduce a double strand break, upon base pairing of the gRNA and to the complementary DNA sequence. The strand break will be repaired either by NHEJ (without donor fragment) or by HR (requiring a homologous donor fragment) [271].

In principle CRISPR-Cas9 can target any DNA sequence, which harbors the PAM, and modifications in almost any genomic loci such as CDSs, promoter or terminator regions can be introduced. However, the implementation or optimization of CRISPR-Cas9 requires target loci, which allow easy discrimination between active and inactive transformants.

Ideally a genomic target locus should be easy to screen for with a quickly realizable assay or the knockout/deletion mutants should display a distinct phenotype upon CRISPR-Cas9 treatment. Moreover the locus should be targetable with homologous knockout cassettes independently from CRISPR-Cas9 applications, in order to determine the targeting efficiency rate and to prove that it is accessible for the genome editing nuclease.

DiCarlo *et al.* [202], who were the first to describe CRISPR-Cas9 genome engineering in *S. cervisiae*, targeted the negative selection marker gene *CAN1*, which encodes for a plasma membrane arginine permease. In case that indels had been introduced by the CRISPR-Cas9 system in the *CAN1* CDS, the toxic arginine analogue canavanine was not transported in the cell, whereas transformants, which contained an intact arginine permease, imported the toxic analogue. In order to test CRISPR-Cas9 directed HR repair they transformed a strain, bearing a nonsense mutation in the *ADE2* gene, with a CRISPR-Cas9 vector and a donor fragment, which contained the correct *ADE2* sequence. *ADE2* is essential for the adenine biosynthesis and mutants accumulate a red pigment [272]. Several other groups integrated a GFP expression cassette in the genome and targeted the GFP-CDS with CRISPR-Cas9 (e.g. in *S. cerevisiae* [263], in mice [244] or human cells [243]). Ryan et al. [258] targeted the *URA3* and *LYP1* loci in *S. cerevisiae*. *URA3*

encodes for a Orotidine-5'-phosphate (OMP) decarboxylase, which is involved in pyrimidine biosynthesis and *URA3* deficient strains are unable to grow on 5-fluoro-orotic acid [273], [274]. *LYP1* encodes for a lysine permease and selection is performed similarly as described for *CAN1* mutants [275].

In *P. pastoris* various auxotrophy markers have been characterized and a large collection of auxotrophic strains is available [276]. *AOX1* knockout strains -called Mut^S- display a reduced growth on methanol [277]. Mut^S strains have been widely used in industry for the production of diverse proteins, since higher production rates for several proteins can be achieved compared to the WT (Mut^+) strains [278]. The HR based integration of linear homologous knockout cassettes depends on the sequence and length of the homologous arms. Integration rates for the *AOX1* locus vary in literature between 2% [199] up to 50% (Mudassar A., personal communication). *P. pastoris ADE1* knockouts accumulate a red pigment similar to *S. cerevisiae* mutants. However, the *ADE1* locus is difficult to target with regular knockout cassettes [199], [279].

I selected different endogenous genes (*AOX1*, *GUT1*, *ADE1* and *ADE2*), which exhibit distinct phenotypes when deleted and I tried to replace to these genes with homologous knockout cassettes containing a Zeocin selection marker. The locus, which showed the best characteristics in terms of the ease of screening and reasonable knockout rates applying homologous cassettes in various *P. pastoris* strains, was selected as test locus for the implementation of CRISPR-Cas9. I transformed the knockout cassettes in the *P. pastoris* CBS 7435 WT as well as in the CBS 7435 *ku70* knockout strain. KU70p is one of the keyplayers in the NHEJ-repair and significantly increased HR-rates can be obtained in CBS 7435 *ku70* compared to the wild-type, although the total number of transformants decreases [199].

TALEN mediated genome engineering in *P. pastoris*

TALENs contain a DNA binding domain of single repeats that can be engineered to bind a desired DNA sequence and the catalytic domain of the *Fok*I nuclease. Two TALEN have to bind on adjacent DNA sequences in a way that *Fok*I can dimerize and introduce a double strand break [18]. TALEN mediated genome engineering was performed in cultured cells as well as in intact eukaryotic organisms [55], [164]. They have been successfully applied in *S. cerevisiae* for gene disruption and DNA insertion experiments. Li *et al*. [164] targeted three native gene sequences, *URA3, LYS2* and *ADE2,* in *S. cerevisiae* and obtained high rates of site-specific gene disruption and gene replacements by HR. Christian *et al*. [163] measured TALEN activity with an in vivo yeast assay, where they introduced a disrupted *lac*Z gene reporter gene on a target plasmid in the cells. TALEN induced HR-repair led to the restoration of a functional *lac*Z gene.

The elaborate design and construction of TALENs, which target a specific genomic locus, is a serious disadvantage in contrast to CRISPR-Cas9, which can easily be reprogrammed by exchanging the variable 20 bp sequence of the gRNA. However CRISPR-Cas9 off-targeting rates are higher compared to TALENs, which introduce a strand break at a unique genomic site. Thus, TALEN offer an alternative for *P. pastoris* platform strain generation, where unpredictable off-targeting is not desired.

3. MATERIALS AND METHODS

All chemicals, enzymes, technical devices and the lab equipment used during this master thesis are listed in the Appendix.

Media, buffers and kits

Stock solutions:

10x YNB: 134 g/L of yeast nitrogen base.

10x PPB (1M K_2PO_4 buffer), pH 6: 30 g/L K_2HPO_4 and 118 g/L KH_2PO_4. The pH was adjusted to 6 with concentrated KOH.

500x Biotin: 10 mg Biotin dissolved in 50 ml dH_2O. The solution was filter sterilized and stored at 4°C.

10x Dextrose: 220 g/L of glucose monohydrate.

BEDS, pH 8.3 (1L): 10 mM bicine-NaOH, 30 ml ethylene glycol, 50 ml DMSO, 1 M sorbitol. The solution was filter sterilized and stored at -20°C.

1M DTT: 1.54 g DTT dissolved in 10 mL ddH_2O. The solution was filter-sterilized and 1 mL aliquots were stored at -20°C.

50x TAE buffer (1L): 242 g/L of TRIS base and 100 mL 0.5M of EDTA, 57.1 mL acetic acid (conc.).

Antibiotics:

100 mg/mL Zecoin stock: 100 µg/mL for *P. pastoris*, 25 µg/mL for *E. coli*

100 mg/mL Geneticin stock: 300 µg/mL for *P. pastoris*

100 mg/mL Kanamycin stock: 100 µg/mL for *E. coli*

100 mg/mL Ampicillin stock: 100 µg/mL for *E. coli*

Cultivation media:

The amounts correspond to the preparation of one liter.

YPD: 10 g yeast extract, 20 g peptone, 15 g agar (only for plates)

LB: 5 g yeast extract, 10 g tryptone, 5 g NaCl

BMD 1%: 50 ml 10x Dextrose, 200 ml 10x PPB, 100 ml 10x YNB, 2 ml 500x Biotin

BMM2: 10 mL MeOH, 200 mL 10x PPB, 100 mL 10x YNB, 2 mL 500x Biotin

BMM10: 50 mL MeOH, 200 mL 10x PPB, 100 mL 10x YNB, 2 mL 500x Biotin

BMG1: 2.5 ml glycerin, 200 mL 10x PPB, 100 mL 10x YNB, 2 mL 500x Biotin

BMD10: 50 ml 10x Dextrose, 200 mL 10x PPB, 100 mL 10x YNB, 2 mL 500x Biotin

SOC-media: 3.46 g glucose, 20 g tryptone, 5 g yeast extract, 0.58 g NaCl, 2 g $MgCl_2$, 0.16 g KCl, 2.46 g $MgSO_4$

Other chemicals and buffers

1x PBS, pH 7.4: 8 g/L NaCl, 0.2 g/L KCl, 1.44 g/L Na_2HPO_4, 0.24 g/L KH_2PO_4

dNTP mix: 2 mM dATP, 2 mM dTTP, 2 mM dCTP, 2 mM dGTP

Yeast lysis buffer, pH 8 (200ml): 4 mL Triton X-100, 20 ml 10% SDS, 4 ml 0.5M NaCl, 0.4 ml 0.5M EDTA, 2 ml 1M Tris-HCl, 196 mL dH2O

Kits

GeneJET Plasmid Miniprep Kit, Fisher Scientific GmbH, Vienna, Austria

Wizard SV Gel and PCR Clean-Up System, Promega GmbH, Mannheim, Germany

CloneJET PCR Cloning Kit, Fisher Scientific GmbH, Vienna, Austria

Protocols and assays

DNA isolation from *E. coli* (plasmid) and *P. pastoris* (genomic DNA)

E. coli plasmid DNA isolation was performed with the GeneJET Plasmid Miniprep Kit. The cells were streaked out on selective LB-agar plates and the plates were incubated at 37°C overnight. Cell material was directly removed from the plates with a toothpick and processed according to the manufacturer's guidelines. The plasmid DNA was either directly used or stored at -20°C. *P. pastoris* genomic DNA was isolated according to the Bust'n'Grab protocol from Harju *et al.* [280] with minor amendments. Liquid nitrogen was used in the cell lysis step instead of dry ice-ethanol. The air tried pellets were resuspended in 50 µl dH20 and I did not perform RNase treatment. The genomic DNA was stored at -20°C. The primers P14332 and P14252 were used for the amplification of the *GUT1* locus, when a NHEJ-mediated mutation had been introduced

with CRISPR-Cas9 and P10040 and P12485 were used for the amplification of the *GUT1* locus, when donor DNA fragments had been co-transformed. P09523 was used for sequencing.

Ligation reactions and Gibson Assembly

The ligation reactions were made with the T4 DNA ligase according to the manufacturer's protocol. 50 ng vector DNA and insert DNA corresponding to a molar ratio of 1:3 were used for the ligation reaction. The reaction was kept at 21°C for 15 to 30 min. Thereafter heat inactivation of the ligase was performed at 75°C for 5 min. 2-4 µL of the ligation reaction were used for transformation.

For Gibson Assembly [281] 50 ng vector DNA and insert DNA in an equimolar ratio were mixed in a total volume of 5 µl ddH_2O and added to 15 µl of the Gibson Assembly master mix. The reaction mixture was incubated at 50°C for one hour. Three µl of the Gibson Assembly mixture were used for the *E. coli* transformation.

Cloning into the pJET1.2 blunt vector

The CloneJET PCR Cloning Kit was used to clone PCR products and gBlocks into the pJET1.2blunt vector. The ligation was performed according to the manufacturer's instruction. The ligation mixture was incubated at room temperature for 15 to 30 min and 2-4 µl were used for the transformation of 80 µl electrocompetent *E. coli* cells.

PCR procedures

Standard PCR

Standard PCR reactions during this thesis were performed with Phusion High Fidelity DNA Polymerase using 25 ng vector DNA or 125 ng genomic DNA as templates. dNTPs in a final concentration of 200 µM and 20 pmol of the fw-primer as well as of the rv-primer were added to the PCR reaction mixture according to the manufacturer's protocol. The annealing temperature was chosen to be 2 °C below the primer DNA melting temperature. The elongation time was calculated in dependence of the expected PCR product length, considering a processivity of the Phusion High Fidelity DNA Polymerase of one kb per 10-30 sec. In case of PCR products larger than three kb an elongation time of one min per kb was used. If the PCR did not yield any product under standard conditions using either HF or GC buffer different annealing temperatures were tested (58°C, 65°C and 72°C). The PCR products were purified from a 1% preparative

agarose gel and further used for cloning or as donor DNA fragments for the *P. pastoris* transformation.

Overlap extension-PCR (OE-PCR)

The OE-PCR was performed with Phusion High Fidelity DNA Polymerase with cycle parameters as described for the standard PCR. The method was used to combine small linear fragments (gBlocks, PCR products) into one large fragment. The first round of the PCR (20 cycles) was performed without external primers using five ng of the largest fragment and the other fragments in an equimolar ratio in a total volume of 50 µl in order to create the initial oePCR template fragment. In the second round 10 nmol dNTPs, 0.25 µl fresh polymerase and 40 pmol primers were added to the reaction mixture and another PCR (25 cycles) was performed.

colonyPCR (cPCR)

cPCR was used to screen for the presence or absence of an insert fragment of *E. coli* transformants. The GoTaq G2 DNA Polymerase was used and the PCR was performed according to the manufacturer's instructions. Cell material from transformation plates was added with a pipette tip to the reaction mix and the initial heating step was set to five min in order to break up the cells. The PCR products were resolved in a 1% control agarose gel and in the case that a visible band with the correct size was obtained the *E. coli* transformant was streaked out for minilysat preparation and the plasmid DNA was send for sequencing.

Restriction endonuclease reactions

The restriction enzymes used during this thesis were obtained, either from Fisher Scientific or New EngIand Biolabs (NEB). In general, 1 µl enzyme was used to digest 1 µg DNA. Restriction digests using Fisher Scientific FastDigest enzymes were incubated for 1h at 37°C and digestion reactions using regular Thermo Fisher Scientific and NEB enzymes were incubated overnight. Enzyme inactivation was only performed for regular Thermo Fisher Scientific and NEB enzymes according to the manufacturers.

Transformation protocols

E. coli TOP10F' competent cells were prepared according to Seidman *et al*. [282] and stored at -80°C. Two to four µl from the Gibson or ligation reaction were mixed with 80 µl competent cells and transferred into an electroporation cuvette. The transformation was performed with 2.5 kV/25 µF/200 Ω using Bio-Rad Gene Pulser System. After the transformation the cells were regenerated

in 1 ml SOC-media at 37°C and 550 rpm for 45 min. After the regeneration step the cells were plated out on selective LB-media and incubated overnight at 37°C.

The preparation and transformation of *P. pastoris* CBS 7435 WT and *P. pastoris* CBS 7435 *ku70* competent cells were performed according to the condensed protocol described by Lin-Cereghino *et al*. [283]. The cells were freshly prepared for every transformation event. The amount of DNA added to the competent cells depends on the fragment size and form of the vector. Typically 10 to 100 ng (for CRISPR-Cas9 plasmids) of circular DNA and 1 µg to 3 µg of linear DNA were used for the transformation. The transformation was performed with the BioRad gene pulser system at 2 kV/25 µF/200 Ω and the cells were regenerated in 1 ml 1:2 mixture of 1M sorbitol and YPD media for two hours. Then the transformants were plated out on selective YDP-agar plates and incubated for two to three days at 28°C.

96-deep well plate cultivation of *P. pastoris* strains expressing eGFP under the control of P_{AOX1}

The screening of randomly chosen clones was performed in 96-deep well plates similar as described by Weis *et al.* [284]. Cells were cultivated in 250 µl BMD1 for approximately 60 h, followed by addition of 250 µl BMM2 and of 50 µl BMM10 after 12 h growth on methanol. The *P. pastoris* strains, which were used for the microscopy experiments, were cultivated accordingly: The cells were grown for 60 hours in BMD1, induced with 250 µl BMM2 and grown for four hours. Subsequently 0.5% glucose (50 µl BMD10) was added and cultivation was prolonged at 28°C and 320 rpm for additional four hours. Thereafter the cells were either directly used for the nuclear staining and fluorescence microscopy or stored at 4°C overnight.

96-deep well plate cultivation and screening with of *P. pastoris* transformants with ARS plasmids

Randomly chosen transformants were cultivated in 96-deep well plates with 250 µl selective YDP-media for 48 hours. The transformants were further used for fluorescence measurements (strains expressing sTomato under the P_{GAP}) or cultivated on different carbon sources (CRISPR-Cas9 strains targeting *GUT1*). A metallic stamp was used to transfer CRISPR-Cas9 transformants from the 96-deep well plates to BMD1 and BMG1 agar square plates. The square plates were incubated at 28°C for three days. Pictures were taken during the time course of the incubation.

Measurement of eGFP and sTomato fluorescence levels

Ten µl cell suspension from the 96-deep well plates were added to 190 µl ddH_20 in a Nunc MicroWell 96-Well Optical-Bottom plate. The optical density at 600 nm, the sTomato fluorescence with 554 nm excitation and 581 nm emission wavelength and the eGFP fluorescence with 488 nm excitation and 507 nm emission wavelength were determined by the SynergyMX platereader.

Nuclear staining and fluorescence microscopy

P. pastoris cells were cultivated in 96-deep well plates as described previously. 500 µl cell culture were removed from the 96-DWP and transferred into an Eppendorf tube (Eppendorf AG, Hamburg, Germany). 500 µl PBS buffer, pH 7.4, 25 µl 1 M DTT and 5 µg Hoechst 33258 were added and the cells were stained for 2 hour, 28°C, at 1000 rpm shaking. I observed an increased membrane permeability of Hoechst 33258 in contrast to DAPI (Sigma-Aldrich Handels GmbH, Austria), which did not enter the *P. pastoris* nucleus without a harsh pretreatment of the cells (e. g. when 70% EtOH was added instead of DTT to the staining mixture). The cells were observed with a Leica DM LB bright field and fluorescence microscope (Leica Mikrosysteme, Austria). For the nuclear stain images (DAPI, Hoechst 33258) the filter D (2) was used with 355-425 nm excitation wavelengths and 470 nm suppression wavelength. The eGFP-fluorescence was observed with the filter I3 (3) using excitation wavelengths between 450 and 490 nm and 515 nm suppression wavelength. The exposure time for eGFP images was set to 630 ms, 39.6 ms for Hoechst 33258 and 7.55 µs for the bright field images. The images were colored and merged with the freeware GNU Image Manipulation Program (Gimp) 2.8.

Codon optimization

The DNA sequences of the heterologous NLSs and the *S. pyogenes* Cas9 sequence (Accession number: WP_023610282) have been codon optimized for the expression in *P. pastoris*. NLSs were optimized manually using a gene optimization table [285]. The Gene Designer software from DNA2.0 Inc., Menlo Park, CA, USA, was used for the optimization of the Cas9 gene. The most common restriction sites, pentameric A/T sequences, and cryptic splicing motifs were excluded. The local GC-content was balanced (30% minimal – 70% maximal) over the whole sequence using the Mobyle portal-freak software. The GeneBee tool was used to make secondary structure predictions of the mRNA. RNA secondary structures with free energy higher than -20 kcal/mol were resolved in order to avoid a reduction of the ribosome processivity. The Cas9

sequence was ordered in five gBlocks (P14452, P14453, P14352, P14353, P14355) and the gBlocks were combined by OE-PCR.

gRNA design

The CDS of the gene of interest was entered in the gRNA design tool from DNA2.0. The programme lists a series of possible gRNA sequences. gRNAs, which bind downstream of the startcodon up to the middle of the DNA sequence were selected. The sequences were blasted against the *P. pastoris* CBS 7435 genome and in case that the sequence occurred more than once in the genome (100% off target region), it was rejected. In principle any sequence, which harbors the motif GG can be selected as gRNA. Homodimers and internal hairpin structures should be avoided. The gRNAs, which were selected to target the *GUT1* locus are listed in Table 4. The DNA sequence of the *GUT1* gene showing the binding sites of the gRNAs can be found in the Appendix.

Table 4: gRNAs to target the *GUT1* locus in *P. pastoris* CBS 7435

gRNA	Location	Sequence
GUT1-gRNA1	39 bp upstream of the start codon	CTAGTTGCTACCATCGATAT
GUT1-gRNA2	125 bp upstream of the start codon	CGAGTACTCTACCTCTGCTC
GUT1-gRNA3	256 bp upstream of the start codon	TGCAATTTCCTCAGCCAGGC
GUT1-gRNA4	429 bp upstream of the start codon	GTTGTTTGGTCCAAGAAGAC
GUT1-gRNA5	876 bp upstream of the start codon	GTCCCACACTTGGAGTCTAT

Strains and plasmids

The plasmid maps of all constructs, which were made during this master thesis are provided in a genbank format on the accompanying USB stick. Plasmids generated during the thesis are stored in my internal strain collection with internal name starting with the capital letter A and a number e.g. A1 and on filter paper. Interesting *E. coli* and *P. pastoris* strains had been transferred to the culture collection of the institute (IMBT). Strains obtained from this collection, which were used during the thesis are denoted with IMBT and the corresponding number e.g. IMBT 6070.

Construction of NLS plasmids

The design of the plasmids used for the expression of the eGFP-NLS fusion constructs is based on *P. pastoris* pPpT4_S vector (IMBT 6070) [199]. The codon optimized DNA sequences of the SV40, the *Hs*Myc, the *Xl*Nuc and the *Sc*Matα2 NLSs were fused either N- or C-terminally to the eGFP CDS with PCR-primers ((P14190/P14204), (P14205/P14203), (P14191/P14204),

(P14206/P14203), (P14192/P14204), (P14193/P14203), (P14207/P14204), (P14208/P14203)), which contain overhangs to the vector. A control strain was made, containing the eGFP CDS without a NLS (P14203/P14204). The pPpT4_S was digested with *Eco*RI and *Not*I and the PCR-fragments were cloned into the vector with Gibson Assembly. Vectors containing the *Pp*Nob1, *Pp*Sda1, *Pp*Set7, *Pp*Uba1, *Pp*Swi5 and *Sc*Swi5 NLSs were constructed accordingly: The control vector (pPpT4_S-eGFP-Zeo, A9) bearing the eGFP CDS without a NLS was cut with *Sal*I and *BamH*I. The enyzmes cut in the eGFP sequence and 3' of the AOX1TT terminator. gBlocks were designed containing the missing part of the eGFP CDS, the NLS, the terminator and overhangs to the vector backbone. For the N-terminal *Sc*SWI5 fusion construct the pPpT4_S-eGFP-Zeo vector was linearized with *Nhe*I and a gBlock containing the NLS and overhangs to the vector was designed. All gBlocks were cloned into the corresponding vector by Gibson Assembly.

Construction of sTomato-ARS plasmids

The plasmids contain the sTomato reporter gene under the control of the constitutive P_{GAP}, an ARS and either a Zeocin or a Geneticin resistance marker cassette. The starting vector pPpT4_GAP (IMBT 6072) was digested with *Eco*RI and *Not*I. The sTomato gene was amplified with to primer pairs ((P14196/P14199), (P14198/P14197)). P14198 and P14199 contain an overlapping sequence with a silent mutation in order to remove a *Pst*I site, which occurs in the CDS sequence. The two sTomato fragments were cloned in the vector backbone by Gibson Assembly. The resistance marker cassette of the pPpT4_GAP-sTom-Zeo (A10) was removed with *Xba*I and *Pst*I. The pPpKan_S (IMBT 6064) was digested with *Xba*I and *Pst*I and the kanamycin marker cassette was then cloned into the pPpT4_GAP-sTom-Zeo backbone to obtain pPpT4_GAP-sTOM-Kan (A23).

The PARS1 sequence was amplified from *P. pastoris* CBS 7435 gDNA with primers (P14200/P14201) and the panARSOPT ARS was ordered as a gBlock (P14259) with overhangs to the vector. The vectors bearing sTomato under the control of P_{GAP} and different selection marker cassettes (A10, A23) were linearized with *Pst*I and the PARS1 PCR fragment respectively the panARSOPT gBlock were cloned in the pPpT4_GAP-sTom-Zeo and pPpT4_GAP-sTOM-Kan plasmids with Gibson assembly. In order to introduce the B1739 ARS, the pPpT4_GAP-sTom-Zeo and pPpT4_GAP-sTOM-Kan plasmids (A10, A23) were cut with *Swa*I and *Pst*I to remove a 1000 bp fragment. This fragment was amplified with a large primer

containing the B1739 sequence from the vector template pPpT4_GAP-sTom-Zeo (P14202/P14210) and recloned into the digested vectors with Gibson assembly.

Construction of donor DNA cassettes and control strains

GUT1, *AOX1*, *ADE2* and *ADE*1 knockout cassettes are based on the plasmid pPpKC1 (IMBT 6602). The flippase was removed from the pPpKC1 plasmid (pPpKC1-2, A22) and the new vector was digested with *Pci*I and *Bgl*II. Roughly 1000 bp regions 5' upstream and 3' downstream of the target gene were selected to complement a *Swa*I cutting site and amplified from *P. pastoris* CBS 7435 genomic DNA with overhangs to the vector backbone ((*GUT1*: P14359/P14252, P14253/P14360), (*AOX1*: P14357/P14248, P14249/P14358), (*ADE2*: P14361/P14256, P14257/P14362), (*ADE1*: P14562/P14563, P14564/P14565)). The PCR fragments were joined by OE-PCR and cloned into vector backbone with Gibson Assembly. For the *P. pastoris* transformation the plasmids were *Swa*I-linearized and one to three µg DNA were added to the competent cells. The transformants were remarkably plated on YPD-plates with 25 µg/ml Zeocin and not 100 µg/ml Zeocin as usual.

A second type of donor DNA cassette was made, which contains 1000 bp regions 5' upstream and 3' downstream of the target gene directly fused together. The 5' upstream and the 3' downstream regions of the *ADE1*, *AOX1* and *GUT1* loci were amplified from *P. pastoris* genomic DNA ((*ADE1*: P14618/P14619, P14620/P14621), (*AOX1*: P14622/P14623, P14624/P14625), (*GUT1*: P14626/P14627, P14628/P14629)). The PCR fragments were joined by OE-PCR and cloned into the pJET1.2 blunt vector. For the *P. pastoris* transformation the donor cassettes were amplified from the pJET1.2 blunt vector with the outer primers used for the OE-PCR.

Construction of CRISPR-Cas9 plasmids with RNA Polymerase III promoters

The pPpT4_GAP (IMBT 6072) was linearized with *Pst*I and PARS1 sequence was cloned in the vector as described for the sTomato-ARS plasmids. The RNA Polymerase III promoters RP1, RP2, RP3, RP4, RP5, RP6 and RP10 were amplified from *S. cerevisiae* and *P. pastoris* genomic DNA. The RNA Polymerase III promoters RP7, RP8 and RP9 were ordered as gBlocks (Table 5).

Table 5: RNA Polymerase III promoters used in this Master thesis.

Short identifier	Promoter	Source	Cloning
RP1	*SNR52*	*S. cerevisiae*	Amplified from genomic DNA (P14267/P14268)
RP2	*RPR1*	*S. cerevisiae*	Amplified from genomic DNA (P14269/P14270)
RP3	*SUP4*	*S. cerevisiae*	Amplified from genomic DNA (P14277/P14278)
RP4	P_{LYS}	*P. pastoris*	Amplified from genomic DNA (P14271/P14272)
RP5	P_{MET}	*P. pastoris*	Amplified from genomic DNA (P14273/P14274)
RP6	P_{SER}	*P. pastoris*	Amplified from genomic DNA (P14275/P14276)
RP7	*SNR52*	*K. lactis*	Ordered as gBlock P14319
RP8	*RPR1*	*S. pombe*	Ordered as gBlock P14320
RP9	P_{ASN}	*P. angusta*	Ordered as gBlock P14321
RP10	*SCR1*	*S. cerevisiae*	Amplified from genomic DNA (P14296/P14297)

Vectors were made in a way that the variable part of the gRNA (20 bp) can be easily exchanged by the use of primers: The pPpT4_GAP-PARS1 plasmid (A24) was cut with *Pci*I and *Swa*I and the vector backbone, the RNA polymerase III promoter and a gBlock, which contains the structural gRNA, the *SUP4* RNA Polymerase III terminator and 600 bp suffer fragment flanked by two *Xho*I sites (P14354), were assembled by Gibson Assembly. The vectors containing the different RNA Polymerase III promoters and the gBlock P14354 were cut with *EcoR*I and *Not*I and the *P. pastoris* codon optimized Cas9 was cloned downstream of P_{GAP} (A77-A85).

In order to clone the 20 bp variable sequence of the GUT1-gRNAs into the vectors (A86-A122), which bear a RNA Polymerase III promoter, the structural gRNA sequence, the *SUP4* RNA Polymerase III terminator and Cas9 under the control of P_{GAP}, a *Xho*I RE-digest was performed. The structural gRNA sequence, the RNA Polymerase III terminator and stuffer segment were amplified with primers, which contain the variable part of the gRNA and the PCR fragment were recloned in the vector by Gibson Assembly (P14454-14492/P14483). The strains A86-A122 contain the RNA Polymerase III promoters expressing gRNAs, which target the *GUT1* locus and the *P. pastoris* codon optimized Cas9 under the control of P_{GAP}.

Additional Cas9 sequences were ordered from Addgene (Cambridge, MA, USA). The pMJ806 (A127, IMBT 7434) bears the *S. pyogenes* Cas9 sequence published by Jinek *et al*. [17] and the

p414-TEF1p-Cas9-CYC1t (A128, IMBT 7435) contains the *H. sapiens* codon optimized *S. pyogenes* Cas9 sequence published by DiCarlo *et al.* [202]. The *P. pastoris* codon optimized Cas9 sequence was removed from the vectors A86 – A122 by an *Eco*RI and *Not*I double digest. The *H. sapiens* codon optimized Cas9 and the *S. pyogenes* Cas9 were amplified from the template vectors using primer pairs P14683/P14684 and P14685/P14686 respectively and assembled with the vector backbones via Gibson Assembly.

Construction of CRISPR-Cas9 plasmids bearing a bidirectional RNA Polymerase II promoter and ribozymes

The plasmids bear the bidirectional histone promoter P_{HTX1}, which directs the expression of Cas9 and the gRNA flanked by ribozymes. The pPpT4_GAP-PARS1-Zeo (A24) was cut with *Swa*I and *Not*I to remove P_{GAP}. First a vector was made bearing the P_{HTX1} flanked by two terminators (DAS1TT and AOX1TT). The AOX1TT is a feature of the existing pPpT4_GAP-PARS1-Zeo vector (A24). The DAS1TT (P14682/P14784) was amplified from the vector TV0975 (dissertation Thomas Vogl, Synthetic biology to improve protein expression in *Pichia pastoris*, Graz University of Technology, 2015) and the P_{HTX1} (P14785/P14786) was amplified from the TV0254 vector. The vector backbone and the P_{HTX1} and DAS1TT PCR fragments were assembled with Gibson Assembly resulting in the vector pPpT4_pHTX1-PARS1-Zeo (A168, Figure 4).

Subsequently, the three different Cas9 sequences were cloned in the pPpT4_pHTX1-PARS1-Zeo. The *P. pastoris* codon optimized Cas9 sequence was amplified from the vector A86 (P14787/P14788), the *H. sapiens* codon optimized Cas9 sequence was amplified from the vector A132 (P14787/P14788) and the *S. pyogenes* Cas9 sequence was amplified from the vector A150 (P14831/P14832). The pPpT4_pHTX1-PARS1-Zeo vector was linearized with *Eco*RI and Gibson Assemblies with the linearized backbone vector and the different Cas9 sequences were performed.

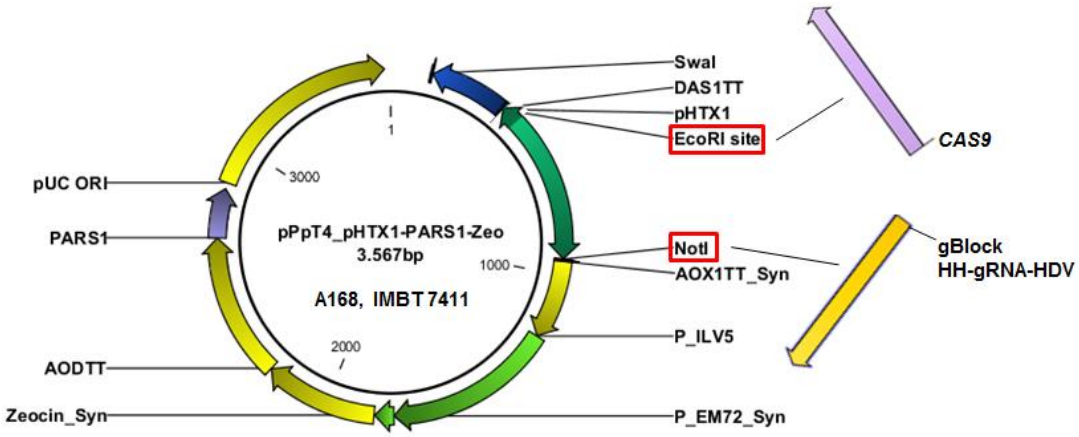


Figure 4: Design of CRISPR-Cas9 vectors bearing a bidirectional RNA Polymerase II promoter and ribozymes. Various *CAS9* genes were cloned in the pPpT4_pHTX1-PARS1-Zeo plasmid (A168, IMBT 7411) bearing the bidirectional *HTX1* promoter flanked by two terminators. Subsequently a gBlock containing the HH ribozyme, the gRNA, the HDV ribozyme and overhangs to the vector was cloned into the *CAS9* bearing vectors (A169, A170, A171) by Gibson Assembly.

The pPpT4_pHTX1-PARS1 plasmids bearing the different Cas9 sequences (A169, A170, A171) were linearized with *Not*I. Different gRNAs, targeting the GUT1 locus, were ordered on gBlocks (P14835, P14836, P14837). The gBlocks contain an overhang to the promoter, the Hammerhead ribozyme, the variable gRNA, the structural gRNA, the HDV ribozyme and an overhang to the *AOX1* terminator. For each gRNA the sequence of the Hammerhead ribozyme has to be changed in order to base pair with the gRNA [263]. The vector backbones and the various gBlock fragments were assembled with Gibson Assembly. A detailed description of the design of the gBlock bearing the gRNA as well as instructions how to design CRISPR-Cas9 plasmids can be found in the Appendix.

*Pp*Cas9 fusion constructs

In order to verify the expression of the Cas9 protein and its localization fusions with eGFP were made. A CRISPR-Cas9 vector, which contains *Pp*Cas9 under the control of P_{GAP} and the RNA Polymerase III promoter *SNR52* for the expressing of the gRNA (A66) was linearized with *Eco*RI. The enzymes cuts at the 5' end of the *Pp*Cas9 sequence. The eGFP sequence was amplified from the vector A9 with overhangs to *Pp*Cas9 and P_{GAP} (P14789/P14838). The eGFP PCR fragment and the vector backbone were assembled by Gibson Assembly. *P. pastoris* transformants of the eGFP-*Pp*Cas9 vector were cultivated for 36 hours in YDP-Zeocin media and stored at 4°C for one day. Then fluorescence microscopy experiments were performed.

For the HIS-Cas9 fusion construct the CRISPR-Cas9 vector A66 was cut with *Eco*RI and *Not*I to remove the PpCas9 sequence. The *Pp*Cas9 was amplified from A66 using a primer, which contains the HIS-Tag sequence and recloned into the vector backbone with Gibson assembly (P14830/P14380). It was initially planned to use the HIS-fusion construct for localization experiments, but it was never used for transformation of *P. pastoris*.

Additional CRISPR-Cas9 targets

gRNAs to introduce mutations in the *AOX1*, *MPP1*, *PRM1* and *MXR1* locus were ordered on gBlocks (P15076 to P15087). The pPpT4_pHTX1-PARS1-HsCas9 (A171, IMBT 7414) was linearized using *Not*I and the gBlocks were cloned into the vector using Gibson Assembly.

Multiplexing

For multiplexing experiments a plasmid bearing *Hs*Cas9 and gRNA3 to target *GUT1* flanked by two ribozymes (A179, pPpT4_pHTX1-hsCas9-RZ-GUT1-gRNA3-RZ) was linearized with *Swa*I. The HHF2 promoter was amplified from the plasmid TV257 (dissertation Thomas Vogl, Synthetic biology to improve protein expression in *Pichia pastoris*, Graz University of Technology, 2015) using different primer pairs (P15005/P15006, P15005/P15007, P15005/P15008). The resulting PCR fragments had different overhangs for Gibson Assembly. Three gBlocks were ordered bearing either AOX1-gRNA1, AOX1-gRNA2 or AOX1-gRNA3 and the *DAS1* terminator. The promoter fragments and the corresponding gBlocks, containing the gRNA and a terminator, were cloned into the vector backbone.

Experimental work on TALEN vectors

Six vectors containing a monomeric TALEN were purchased from Genecopoeia, Rockville, USA. The vectors were obtained in frozen *E. coli* stocks. The company does not offer expression vectors for *P. pastoris*. Thus the TALEN, which target the *P. pastoris GUT1* locus were sent in a mammalian expression vector. Two vectors have to be co-transformed in the mammalian host to express one TALEN pair. The Genecopoeia vectors are listed in Table 6.

Table 6: Mammalian TALEN expression vectors containing a binding domain to target *GUT1* in *P. pastoris*. The vectors were obtained from Genecopoeia, Rockville, USA. The TALEN-CDS was planned to be recloned in *P. pastoris* expression vectors.

Vector name	Internal strain collection number	Resistance marker	IMBT strain collection number
OTN000349-3-B-a-R	A184	Amp	7428
OTN000349-3-B-b-R	A185	Amp	7429
OTN000349-3-B-c-R	A186	Amp	7430
OTN000349-3-B-a-L	A187	Amp	7431
OTN000349-3-B-b-L	A188	Amp	7432
OTN000349-3-B-c-L	A189	Amp	7433

The Genecopoeia vectors were sent for sequencing using P14867 and P14868 as primers. The TALEN-CDSs was PCR amplified (P14901/P14902, P14903/P14904) and cloned into a pJET1.2 vector.

E.coli strains and plasmids

All plasmids generated in this thesis are listed in Table 7. *E. coli* TOP10F' cells bearing the plasmids were stored as 30% glycerol stocks in my internal strains collection and four µl plasmid DNA of the various strains were stored on filter paper. Interesting strains were transferred to the strain collection of the institute (IMBT).

Table 7: *E. coli* strains generated during the master thesis

Host strain	Plasmid	Internal strain collection number	Resistance marker	IMBT strain collection number
Nuclear localisation signals				
TOP 10 F'	pPpT4_S-nNLS-SV40-eGFP-Zeo	A1	Zeo	7351
TOP 10 F'	pPpT4_S-nNLS-c-myc-eGFP-Zeo	A2	Zeo	7352
TOP 10 F'	pPpT4_S-nNLS-nucleoplasmin-eGFP-Zeo	A3	Zeo	7353
TOP 10 F'	pPpT4_S-nNLS-trMata2-eGFP-Zeo	A4	Zeo	7354
TOP 10 F'	pPpT4_S-cNLS-SV40-eGFP-Zeo	A5	Zeo	7350
TOP 10 F'	pPpT4_S-cNLS-c-myc-eGFP-Zeo	A6	Zeo	7355
TOP 10 F'	pPpT4_S-cNLS-nucleoplasmin-eGFP-Zeo	A7	Zeo	7356
TOP 10 F'	pPpT4_S-cNLS-trMata2-eGFP-Zeo	A8	Zeo	7357
TOP 10 F'	pPpT4_S-eGFP-Zeo	A9	Zeo	-
TOP 10 F'	pPpT4_S-cNLS-PpNob1-eGFP-Zeo	A12	Zeo	-
TOP 10 F'	pPpT4_S-cNLS-PpSda1-eGFP-Zeo	A13	Zeo	-
TOP 10 F'	pPpT4_S-cNLS-PpSet7-eGFP-Zeo	A14	Zeo	7358
TOP 10 F'	pPpT4_S-cNLS-PpUba1-eGFP-Zeo	A15	Zeo	7359
TOP 10 F'	pPpT4_S-cNLS-PpSwi5-eGFP-Zeo	A16	Zeo	-
TOP 10 F'	pPpT4_S-cNLS-ScSwi5-eGFP-Zeo	A17	Zeo	7392
TOP 10 F'	pPpT4_S-nNLS-ScSwi5-eGFP-Zeo	A183	Zeo	7393
Autonomously replicating signals				
TOP 10 F'	pPpT4_GAP-sTom-Zeo	A10	Zeo	-
TOP 10 F'	pPpT4_GAP-sTOM-PARS1-Zeo	A25	Zeo	7394
TOP 10 F'	pPpT4_GAP-ARS-panARSOPT-sTOM-Zeo	A21	Zeo	7395
TOP 10 F'	pPpT4_GAP-ARS-B1739-sTOM-Zeo	A20	Zeo	7396

TOP 10 F'	pPpT4_GAP-sTOM-Kan	A23	Gen	-
TOP 10 F'	pPpT4_GAP-sTOM-PARS1-Kan	A26	Gen	-
TOP 10 F'	pPpT4_GAP-sTOM-panARS-OPT-Kan	A27	Gen	-
TOP 10 F'	pPpT4_GAP-sTOM-B1739-Kan	A28	Gen	
TOP 10 F'	pPpT4_GAP-PARS1-Zeo	A24	Zeo	7397
Knockout cassettes				
TOP 10 F'	pPpKC-1-2-flippase removed	A22	Zeo	7398
TOP 10 F'	pPpKC1-AOX1-Gibson-Zeo (1/4)	A51	Zeo	7401
TOP 10 F'	pPpKC1-ADE2-Gibson-Zeo (1/4)	A52	Zeo	-
TOP 10 F'	pPpKC1-GUT1-Gibson-Zeo (1/4)	A53	Zeo	7406
TOP 10 F'	pPpKC1-ADE1-Gibson-Zeo (1/4)	A123	Zeo	7407
TOP 10 F'	pJET-5+3UTR_AOX1	A129	Amp	7408
TOP 10 F'	pJET-5+3UTR_ADE1	A130	Amp	7409
TOP 10 F'	pJET-5+3UTR_GUT1	A131	Amp	7410
CRISPR-Cas9 constructs with RNA Polymerase III promoters and *P. pastoris* codon optimized Cas9				
TOP 10 F'	pPpT4_GAP-PARS1-RP1-Stuffer-Cas9-Zeo	A66	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP1-Stuffer-Cas9-gRNA1-GUT1-Zeo	A86	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP1-Stuffer-Cas9-gRNA2-GUT1-Zeo	A87	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP1-Stuffer-Cas9-gRNA3-GUT1-Zeo	A88	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP1-Stuffer-Cas9-gRNA4-GUT1-Zeo	A89	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP1-Stuffer-Cas9-gRNA5-GUT1-Zeo	A90	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP1-Stuffer-Cas9-gRNA1-ADE2-Zeo	A91	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP1-Stuffer-Cas9-gRNA4-ADE2-Zeo	A92	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP1-Stuffer-Cas9-gRNA5-ADE2-Zeo	A93	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP1-Stuffer-Cas9-gRNA6-ADE2-Zeo	A94	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP1-Stuffer-Cas9-gRNA7-ADE2-Zeo	A95	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP2-Stuffer-Cas9-gRNA2-GUT1-Zeo	A96	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP2-Stuffer-Cas9-gRNA3-GUT1-Zeo	A97	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP2-Stuffer-Cas9-gRNA4-GUT1-Zeo	A98	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP3-Stuffer-Cas9-gRNA2-GUT1-Zeo	A99	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP3-Stuffer-Cas9-gRNA3-GUT1-Zeo	A100	Zeo	-

TOP 10 F'	pPpT4_GAP-PARS1-RP3-Stuffer-Cas9-gRNA4-GUT1-Zeo	A101	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP4-Stuffer-Cas9-gRNA2-GUT1-Zeo	A102	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP4-Stuffer-Cas9-gRNA3-GUT1-Zeo	A103	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP4-Stuffer-Cas9-gRNA4-GUT1-Zeo	A104	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP5-Stuffer-Cas9-gRNA2-GUT1-Zeo	A105	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP5-Stuffer-Cas9-gRNA3-GUT1-Zeo	A106	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP5-Stuffer-Cas9-gRNA4-GUT1-Zeo	A107	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP6-Stuffer-Cas9-gRNA2-GUT1-Zeo	A108	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP6-Stuffer-Cas9-gRNA3-GUT1-Zeo	A109	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP6-Stuffer-Cas9-gRNA4-GUT1-Zeo	A110	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP7-Stuffer-Cas9-gRNA2-GUT1-Zeo	A111	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP7-Stuffer-Cas9-gRNA3-GUT1-Zeo	A112	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP7-Stuffer-Cas9-gRNA4-GUT1-Zeo	A113	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP8-Stuffer-Cas9-gRNA2-GUT1-Zeo	A114	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP8-Stuffer-Cas9-gRNA3-GUT1-Zeo	A115	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP8-Stuffer-Cas9-gRNA4-GUT1-Zeo	A116	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP9-Stuffer-Cas9-gRNA2-GUT1-Zeo	A117	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP9-Stuffer-Cas9-gRNA3-GUT1-Zeo	A118	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP9-Stuffer-Cas9-gRNA4-GUT1-Zeo	A119	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP10-Stuffer-Cas9-gRNA2-GUT1-Zeo	A120	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP10-Stuffer-Cas9-gRNA3-GUT1-Zeo	A121	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP10-Stuffer-Cas9-gRNA4-GUT1-Zeo	A122	Zeo	-
CRISPR-Cas9 constructs with RNA Polymerase III promoters and *H. sapiens* codon optimized Cas9				
TOP 10 F'	pPpT4_GAP-PARS1-RP1-Stuffer-hsCas9-gRNA2-GUT1-Zeo	A132	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP1-Stuffer-hsCas9-gRNA3-GUT1-Zeo	A133	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP1-Stuffer-hsCas9-gRNA4-GUT1-Zeo	A134	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP2-Stuffer-hsCas9-gRNA2-GUT1-Zeo	A135	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP2-Stuffer-hsCas9-gRNA3-GUT1-Zeo	A136	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP2-Stuffer-hsCas9-gRNA4-GUT1-Zeo	A137	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP3-Stuffer-hsCas9-gRNA2-GUT1-Zeo	A138	Zeo	-

TOP 10 F'	pPpT4_GAP-PARS1-RP3-Stuffer-hsCas9-gRNA3-GUT1-Zeo	A139	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP3-Stuffer-hsCas9-gRNA4-GUT1-Zeo	A140	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP4-Stuffer-hsCas9-gRNA2-GUT1-Zeo	A141	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP4-Stuffer-hsCas9-gRNA3-GUT1-Zeo	A142	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP4-Stuffer-hsCas9-gRNA4-GUT1-Zeo	A143	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP5-Stuffer-hsCas9-gRNA2-GUT1-Zeo	A144	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP5-Stuffer-hsCas9-gRNA3-GUT1-Zeo	A145	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP5-Stuffer-hsCas9-gRNA4-GUT1-Zeo	A146	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP6-Stuffer-hsCas9-gRNA2-GUT1-Zeo	A147	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP6-Stuffer-hsCas9-gRNA3-GUT1-Zeo	A148	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP6-Stuffer-hsCas9-gRNA4-GUT1-Zeo	A149	Zeo	-
CRISPR-Cas9 constructs with RNA Polymerase III promoters and *S. pyogenes* Cas9				
TOP 10 F'	pPpT4_GAP-PARS1-RP1-Stuffer-spCas9-gRNA2-GUT1-Zeo	A150	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP1-Stuffer-spCas9-gRNA3-GUT1-Zeo	A151	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP1-Stuffer-spCas9-gRNA4-GUT1-Zeo	A152	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP2-Stuffer-spCas9-gRNA2-GUT1-Zeo	A153	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP2-Stuffer-spCas9-gRNA3-GUT1-Zeo	A154	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP2-Stuffer-spCas9-gRNA4-GUT1-Zeo	A155	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP3-Stuffer-spCas9-gRNA2-GUT1-Zeo	A156	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP3-Stuffer-spCas9-gRNA3-GUT1-Zeo	A157	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP3-Stuffer-spCas9-gRNA4-GUT1-Zeo	A158	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP4-Stuffer-spCas9-gRNA2-GUT1-Zeo	A159	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP4-Stuffer-spCas9-gRNA3-GUT1-Zeo	A160	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP4-Stuffer-spCas9-gRNA4-GUT1-Zeo	A161	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP5-Stuffer-spCas9-gRNA2-GUT1-Zeo	A162	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP5-Stuffer-spCas9-gRNA3-GUT1-Zeo	A163	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP5-Stuffer-spCas9-gRNA4-GUT1-Zeo	A164	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP6-Stuffer-spCas9-gRNA2-GUT1-Zeo	A165	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP6-Stuffer-spCas9-gRNA3-GUT1-Zeo	A166	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP6-Stuffer-spCas9-gRNA4-GUT1-Zeo	A167	Zeo	-

CRISPR-Cas9 constructs with RNA Polymerase II promoter and ribozymes

TOP 10 F'	pPpT4_pHTX1-PARS1-Zeo	A168	Zeo	7411
TOP 10 F'	pPpT4_pHTX1-PARS1-ppCas9-Zeo	A169	Zeo	7412
TOP 10 F'	pPpT4_pHTX1-PARS1-spCas9-Zeo	A170	Zeo	7413
TOP 10 F'	pPpT4_pHTX1-PARS1-hsCas9-Zeo	A171	Zeo	7414
TOP 10 F'	pPpT4_pHTX1-ppCas9-RZ-GUT1-gRNA2-RZ	A172	Zeo	-
TOP 10 F'	pPpT4_pHTX1-ppCas9-RZ-GUT1-gRNA3-RZ	A173	Zeo	-
TOP 10 F'	pPpT4_pHTX1-ppCas9-RZ-GUT1-gRNA4-RZ	A174	Zeo	-
TOP 10 F'	pPpT4_pHTX1-spCas9-RZ-GUT1-gRNA2-RZ	A175	Zeo	-
TOP 10 F'	pPpT4_pHTX1-spCas9-RZ-GUT1-gRNA3-RZ	A176	Zeo	-
TOP 10 F'	pPpT4_pHTX1-spCas9-RZ-GUT1-gRNA4-RZ	A177	Zeo	-
TOP 10 F'	pPpT4_pHTX1-hsCas9-RZ-GUT1-gRNA2-RZ	A178	Zeo	7425
TOP 10 F'	pPpT4_pHTX1-hsCas9-RZ-GUT1-gRNA3-RZ	A179	Zeo	7426
TOP 10 F'	pPpT4_pHTX1-hsCas9-RZ-GUT1-gRNA4-RZ	A180	Zeo	7427
Tagged Cas9				
TOP 10 F'	pPpT4_GAP-PARS1-RP1-Stuffer-GFP-Cas9-Zeo	A181	Zeo	-
TOP 10 F'	pPpT4_GAP-PARS1-RP1-Stuffer-HIS-Cas9-Zeo	A182	Zeo	-

P. pastoris strains

Table 8 contains some of the *P. pastoris* strains generated during this master thesis. The *P. pastoris* CBS 7435 (IMBT strain collection 3132) and the *P. pastoris* CBS 7435 *ku70* (IMBT 3499) were used as host strains. Interesting strains were transferred to the strain collection of the institute (IMBT).

Table 8: *P. pastoris* strains generated during this master thesis

Host strain	Plasmid	Internal strain collection number	Resistance marker	IMBT strain collection number
Nuclear localisation signals				
CBS 7435	pPpT4_S-nNLS-SV40-eGFP-Zeo K-4E	A29	Zeo	-
CBS 7435	pPpT4_S-nNLS-c-myc-eGFP-Zeo K-11G	A30	Zeo	-
CBS 7435	pPpT4_S-nNLS-nucleoplasmin-eGFP-Zeo K-12C	A31	Zeo	-
CBS 7435	pPpT4_S-nNLS-trMata2-eGFP-Zeo K-9E	A32	Zeo	-
CBS 7435	pPpT4_S-cNLS-SV40-eGFP-Zeo K-9E	A33	Zeo	-
CBS 7435	pPpT4_S-cNLS-c-myc-eGFP-Zeo K-11E	A34	Zeo	-
CBS 7435	pPpT4_S-cNLS-nucleoplasmin-eGFP-Zeo K-10D	A35	Zeo	-
CBS 7435	pPpT4_S-cNLS-trMata2-eGFP-Zeo K-11D	A36	Zeo	-
CBS 7435	pPpT4_S-eGFP-Zeo K-1A	A37	Zeo	-
CBS 7435	pPpT4_S-cNLS-Pichia nuclear protein 1-eGFP-Zeo K-8G	A38	Zeo	-
CBS 7435	pPpT4_S-cNLS-Pichia nuclear protein 2-eGFP-Zeo K-4F	A39	Zeo	-
CBS 7435	pPpT4_S-cNLS-Pichia nuclear protein 3-eGFP-Zeo K-10G	A40	Zeo	-
CBS 7435	pPpT4_S-cNLS-Pichia nuclear protein 4-eGFP-Zeo K-10A	A41	Zeo	-
CBS 7435	pPpT4_S-cNLS-Pichia nuclear protein 5-eGFP-Zeo K-6F	A42	Zeo	-
CBS 7435	pPpT4_S-cNLS-SWI5-eGFP-Zeo K-6A	A43	Zeo	-
CBS 7435	pPpT4_S-nNLS-SWI5-eGFP-Zeo	A204	Zeo	-
Autonomously replication signals				
CBS 7435	pPpT4_GAP-sTOM-PARS1-Zeo K-5D	A44	Zeo	-
CBS 7435	pPpT4_GAP-sTOM-panARSOPT-Zeo K-12G	A45	Zeo	-

CBS 7435	pPpT4_GAP-sTOM-B1739-Zeo K-5B	A46	Zeo	-
CBS 7435	pPpT4_GAP-sTOM-PARS1-Kan K-2D	A47	Gen	-
CBS 7435	pPpT4_GAP-sTOM-panARS-OPT-Kan K-11A	A48	Gen	-
CBS 7435	pPpT4_GAP-sTOM-B1739-Kan K-1B	A49	Gen	-
CBS 7435	pPpT4_GAP-sTOM-Zeo K-3C	A69	Zeo	-
CBS 7435	pPpT4_GAP-sTOM-Kan K-8C	A70	Gen	-
Knockout cassettes (the strains contain a Zeocin marker cassette)				
CBS 7435	pPpKC1-AOX1-Gibson-Zeo (1/4)	A54	Zeo	7436
CBS 7435	pPpKC1-GUT1-Gibson-Zeo (1/4) K1-E8	A56	Zeo	7437
KU70	pPpKC1-ADE1-Gibson-Zeo (1/4) K1	A124	Zeo	7438
CRISPR-Cas9 GUT1 deletion strains				
CBS 7435	pPpT4_pHTX1-hsCas9-RZ-GUT1-gRNA2-RZ K4D4	A205	-	-
CBS 7435	pPpT4_pHTX1-hsCas9-RZ-GUT1-gRNA3-RZ K5E5	A206	-	-
CBS 7435	pPpT4_pHTX1-hsCas9-RZ-GUT1-gRNA4-RZ K6G7	A207	-	-

4. RESULTS AND DISCUSSION

The identification of potent nuclear localization signals

In Figure 5 the design of the plasmids used for the expression of the eGFP-NLS fusion constructs is shown. The exogenous and endogenous NLSs were directly fused to the eGFP CDS as described in the material and methods section. All DNA and protein sequences of the NLSs used in this study are listed in Table 2 and Table 3. The eGFP-NLS fusion proteins were expressed under the control of the exceptionally strong, methanol inducible promoter of the *alcohol oxidase 1* gene (P_{AOX1}) [286]. P_{AOX1} is repressed by glucose and can be strongly induced with methanol, when glucose is depleted.

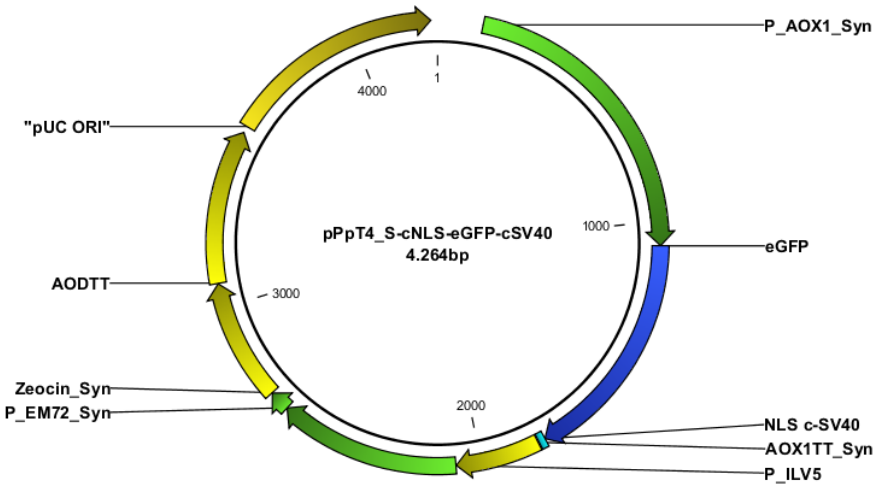


Figure 5: Plasmid map of NLS vectors. The vectors are based on our in house pPpT4_S plasmid. The eGFP reporter gene is expressed under P_{Aox1}. The plasmids contain a Zeocin resistance cassette. The NLS was either fused N- or C-terminally to the CDS of the reporter gene.

The transformants were cultivated for 60 hours in glucose containing media and methanol induction was performed for 24 hours. The fluorescence of the cells under the microscope was very intense and it was not possible to localize subcellular compartments, because of the high levels of intracellular eGFP. Thus, I modified the cultivation protocol in order to reduce eGFP expression. The cells were induced for four hours with methanol containing media, subsequently glucose was added to repress P_{AOX1} and the cultivation was prolonged for additional four hours (Figure 6). Thereby the production of eGFP ceased and the reporter protein could accumulate in

the nucleus due to additional time for sorting. The glucose addition could have also caused autophagy, which led to the degradation of non-essential cytoplasmic proteins including eGFP.

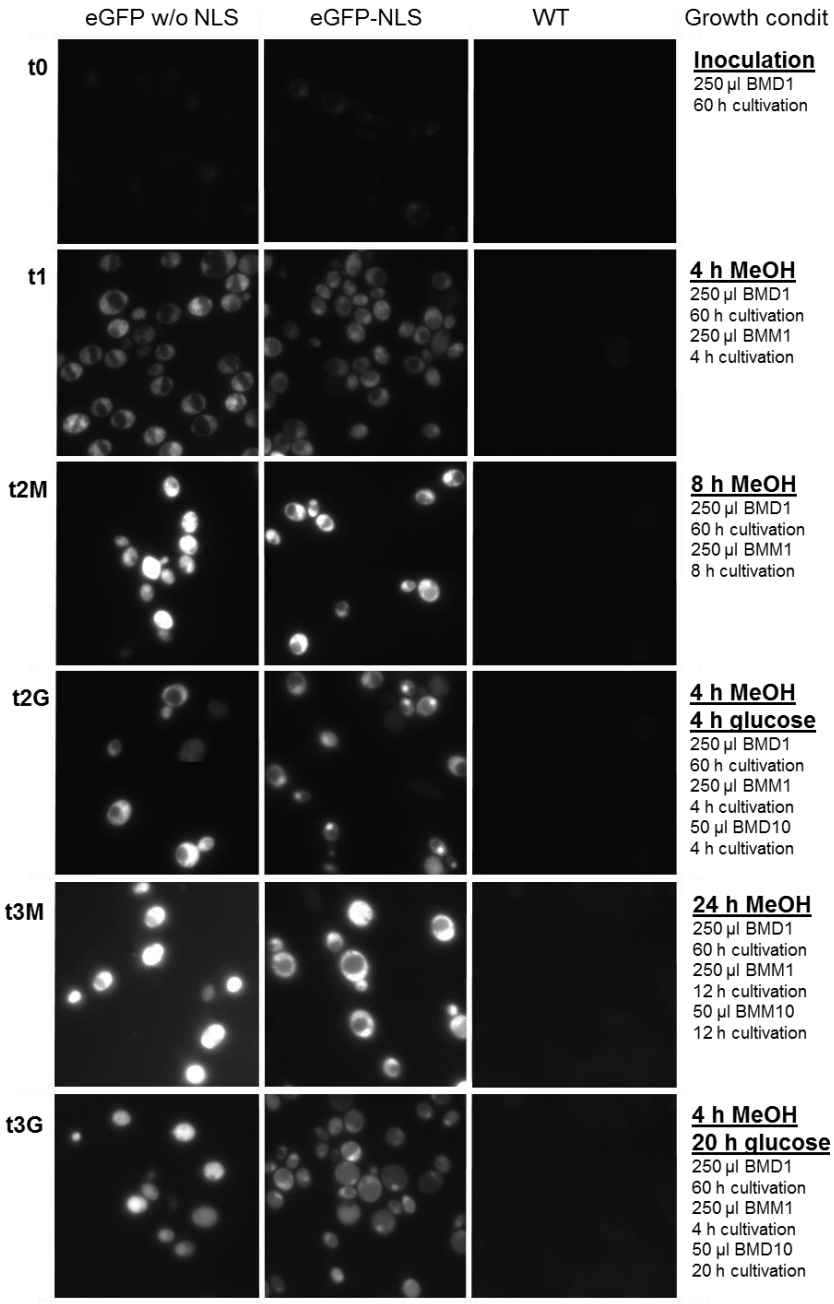


Figure 6: eGFP fluorescence obtained with different cultivation protocols. Fluorescence microscopy images of the *P. pastoris* CBS 7435 strains expressing GFP fused to a representative functional NLS (*Pp*Set7 C-terminally fused to eGFP), eGFP without a NLS (w/o eGFP) and the wildtype strain during the time course of different cultivation conditions. All cells were inoculated in a 96-DWP containing 250 µl BMD1 media and grown for 60 h (t0). After 60 h an induction step with methanol was performed and the cells were grown for four hours (t1). Then the cells were either induced with glucose containing media (50 µl BMD10, G) or kept shaking, omitting the induction step (M). After 4 hours cultivation fluorescence images of the differently cultivated cells were taken (t2M, t2G). The cells, which had been solely induced with methanol, were repeatedly induced with methanol (50 µl BMM10) after 12 hours. Twenty-four hours after the first methanol induction step images of cells, which were grown on methanol (24 h MeOH) and of cells, which were grown on methanol and glucose, were taken.

The localization of the eGFP-NLS fusions was observed with a Leica DM LB fluorescence microscope and the eGFP levels were measured with a fluorescence spectrometer (Figure 7). The nuclear localization of eGFP was confirmed by Hoechst 33258 staining. The microscopy images (Figure 7A and B) show that the NLS of the SV40 large T antigen, the *Xl*Nuc, *Hs*Myc and *Sc*SWI5 protein were able to direct the reporter protein to the nucleus, when they were fused either N- or C-terminally to the eGFP. Only the C-terminal application of the *Sc*Matα2 NLS did not promote eGFP translocation. The *Sc*Matα2 NLS was also ineffective, when fused C-terminally to a recombinant protein in mammalian cells [287]. In its natural role the *Sc*Matα2 NLS is not located close to the termini but rather in the middle of the protein sequence and its functionality may depend on the adjacent protein context. The fact that heterologous NLSs are recognized very efficiently in *P. pastoris* underlines the finding that the nuclear import pathway is well conserved among different eukaryotic species [288] and that nuclear targeting will not become a limiting factor for an efficient CRISPR-Cas9 system for *P. pastoris*.

N-terminal SV40 NLS-eGFP fusions were transported to the nucleus, but the fusion protein was present in low amounts in the cells. We assume that the N-terminal fusion of this NLS affected the stability of the protein, indicated by low fluorescence levels (Figure 7A and B). This is remarkable, since the SV40 NLS is the most commonly used sequence for the nuclear import of heterologous proteins and the finding suggests that the NLSs may have an influence on the application of interest.

Also for the *Pp*Set7- and *Pp*Uba1–fusion constructs relatively low fluorescence levels were measured. Here, bright nuclei and little amount of cytoplasmic eGFP can be observed under the microscope, indicating that most of the eGFP was directed to the nucleus. Thus, the NLS *Pp*Set7 and *Pp*Uba1 appear to be very effective for the nuclear import of the heterologous protein. The NLSs *Pp*Nob1, *Pp*Sda1 and *Pp*Swi5 did not promote translocation, although the non-functional endogenous NLSs are similarly organized as the functional sequences *Pp*Set7 and *Pp*Uba1 (a basic aa cluster followed by a spacer and a second cluster of basic aa). This finding implies that additional residues, which occur in the protein sequence, are most probably involved in the translocation in context of the native proteins.

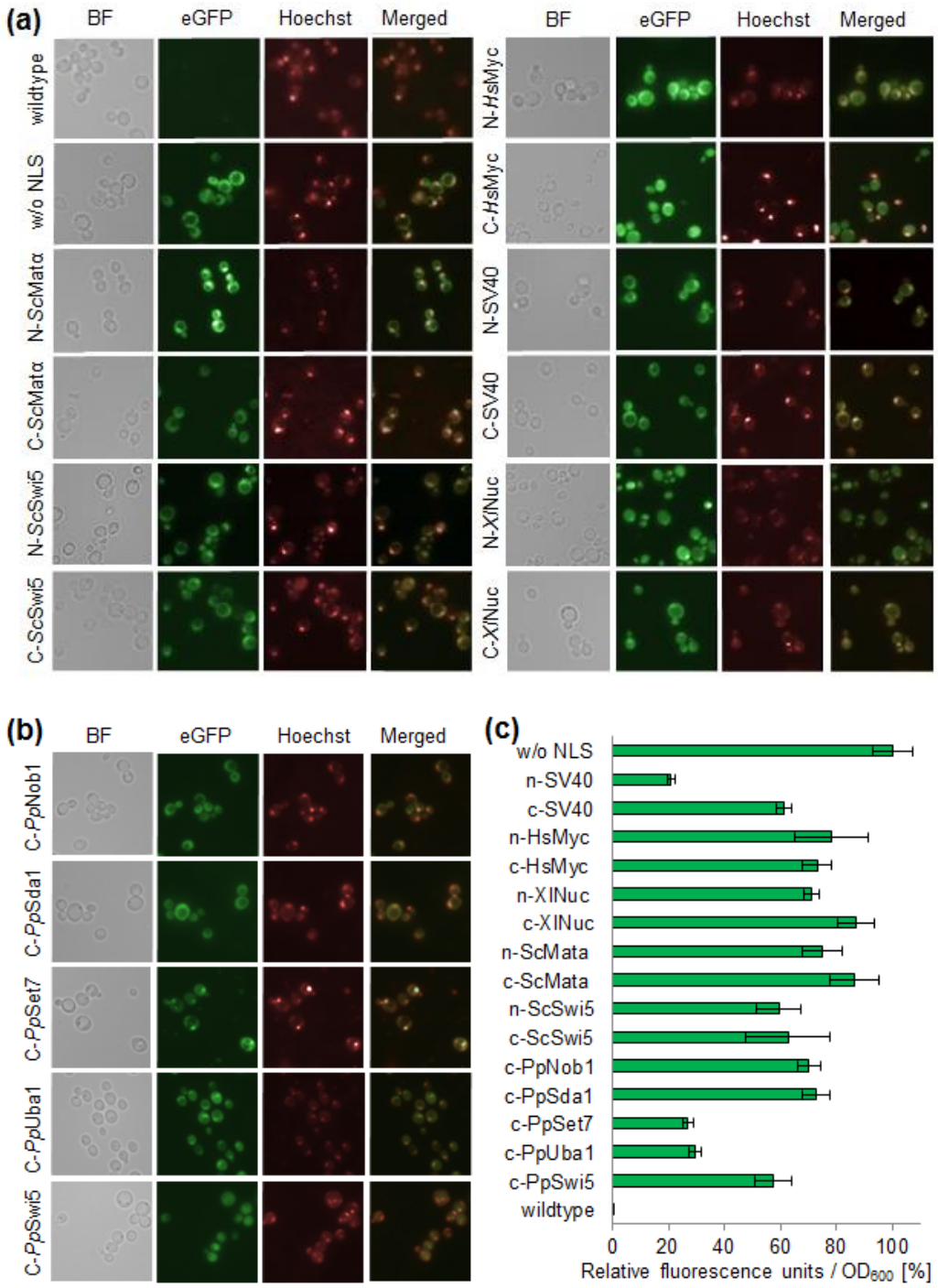


Figure 7: Characterization of heterologous and endogenous NLSs in *P. pastoris*. (a) Heterologous NLSs of *S. cerevisiae* (left side of the panel) and from higher eukaryotes and a eukaryotic virus (right side of the panel) were fused to a eGFP reporter gene and transformed in *P. pastoris*. The nuclei were stained with Hoechst 33258. The translocation of the fusion proteins was observed by fluorescence microscopy. Bright field images (BF) and fluorescence images using different filters are shown (Hoechst, eGFP). As a control the eGFP gene was expressed without a NLS. The CBS 7435 wildtype strain does not express eGFP. (b) Endogenous NLSs from *P. pastoris* nuclear proteins were fused to the eGFP reporter gene and analyzed as described in panel (a). (c) Quantitative fluorescence spectroscopy measurements of eGFP-NLS fusion constructs. The relative fluorescence levels of the eGFP-NLS fusion constructs to the control (GFP localized in the cytoplasm) are shown. The strains were measured according to the protocol outlined in the materials and methods section. Mean values and standard deviations of biological 6-fold replicates are shown.

All fusion constructs affect the expression levels of the reporter gene compared to the control, where eGFP (no NLS fusion) was produced into the cytosol (Figure 7C). The addition of the NLSs might trigger protein degradation or have an effect on the conformation, which may reduce signal strength. The reduction of the signal strength can also be caused by the translocation of reporter protein to the nucleus, where eGFP is surrounded by an additional layer, the nuclear envelop, which might impede the strength of the fluorescence signal.

It is favorable to test more than one NLS for the import of a heterologous protein of interest, because the efficiency of a particular NLS may vary depending on the protein to which it is fused [219]. The NLSs reported here constitute on the one hand a useful resource specifically for the *P. pastoris* community. On the other hand, the findings appear also relevant for other yeast systems: NLSs can show drastic differences in functionality ranging from efficient nuclear targeting even to detrimental effects on the protein of interest. Therefore our work may provide a useful resource for researchers, who want to establish similar toolboxes of NLSs in other host organisms.

For the nuclear import of Cas9 the SV40 sequence was fused C-terminally to the CDS. In order to prove that the 160 kDa protein is transported to the nucleus a fusion construct was made, which contains eGFP N-terminally and the SV40 sequence C-terminally fused to Cas9. Figure 8 shows that eGFP-Cas9-c-SV40 is transported to the nucleus. Cleavage of the fusion protein did not take place, because otherwise eGFP would have accumulated in the cytoplasm similar to the eGFP-control. Other NLSs have not been tested yet for the import of Cas9 and it remains of future interest, if they have a beneficial effect for CRISPR-Cas9 genome editing.

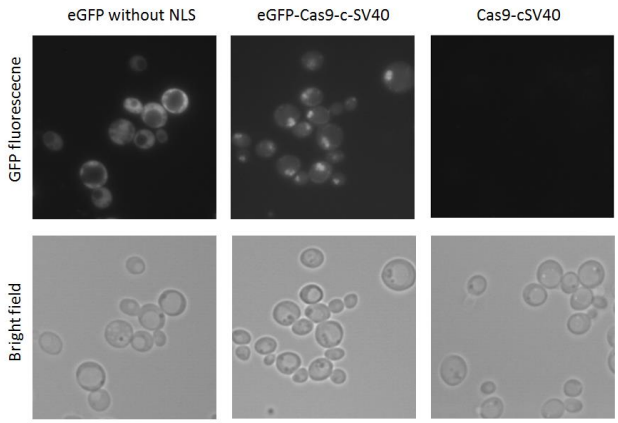


Figure 8: The nuclear import of *Pp*Cas9. Bright field and eGFP fluorescence images of the eGFP-Cas9-cSV40 and various controls were captured. As a control the eGFP, which did not contain a NLS, was expressed. The reporter protein accumulates in the cytoplasm. The Cas9-SV40 construct contains the NLS c-SV40 C-terminally fused to the *P. pastoris* codon optimized Cas9 sequence. No fluorescence can be detected. eGFP-Cas9-SV40 contains the eGFP N-terminally and the SV40 sequence C-terminally fused to the *Pp*Cas9. The protein was translocated to the nucleus.

The effect of autonomously replicating signals and different antibiotic resistance genes on heterologous protein production

Providing the CRISPR-Cas9 vector with an ARS allows the plasmid to get lost after the genome modification has been performed. At this stage the stable maintenance of the CRISPR-Cas9 plasmid would cause an additional and unwanted burden for the generated mutants, which might further influence protein production capabilities, cell growth and fitness.

The influence on protein production levels and plasmid stability of three different ARSs was tested to identify a sequence, which is most suitable for vectors bearing CRISPR-Cas9. I compared the well characterized PARS1 sequence [230] and a novel ARS "panARSOPT", which was described to be active in ten species of budding yeast [232]. This sequence is highly interesting for its application in shuttle vectors. The third sequence "B1739" was derived from a GC-rich origin of replication, which had been selected by comparing the colony growth of ARS transformants from the publication [231], however its application for expression vectors has not been described yet. Comparing the G/C-rich and the A/T-rich PARS1 is of fundamental interest in the understanding of the two classes of replication origins. Figure 9 shows a plasmid map of the various ARS plasmids.

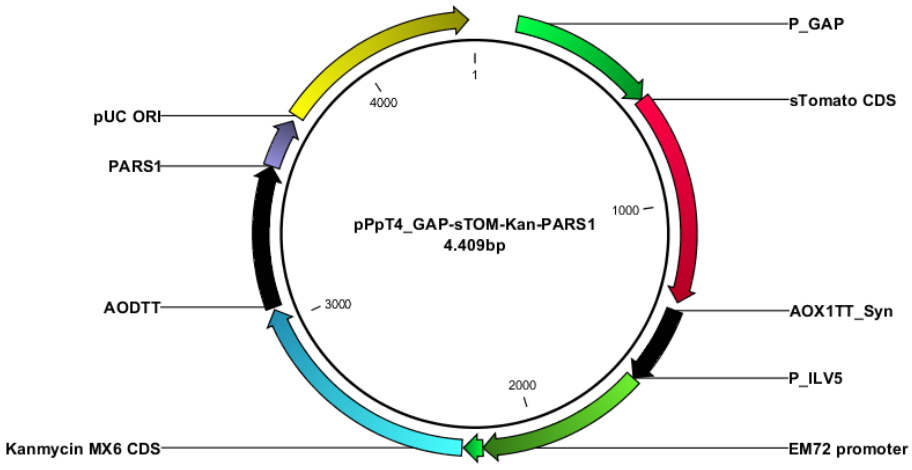


Figure 9: Plasmid map of ARS vectors. The vectors are based on our in house T4-plasmid. The sTomato reporter gene is expressed under P_{GAP}. The plasmids either contain a Kanamycin (Geneticin) or a Zeocin resistance gene. The respective ARS (PARS1, panARSOPT or B1739) was inserted upstream of the pUC ORI. All other features of the vector remain constant.

The plasmids were transformed into *P. pastoris* CBS 7435 WT and randomly chosen transformants were cultivated in YDP media supplemented with antibiotics. Four representative transformants of each construct were selected and grown in 96-deep well plates containing YPD- and YDP-media with antibiotics for 36 hours. Thereafter the sTomato fluorescence was measured. The relative fluorescence units are shown in Figure 10. Initially the fluorescence was measured after 60 hours growth on glucose as C-source, where ~3-fold higher expression levels than after 36 hours of cultivation were obtained. In order to reduce the dilution error and not to exceed the linear range of the spectrophotometer the incubation time was reduced.

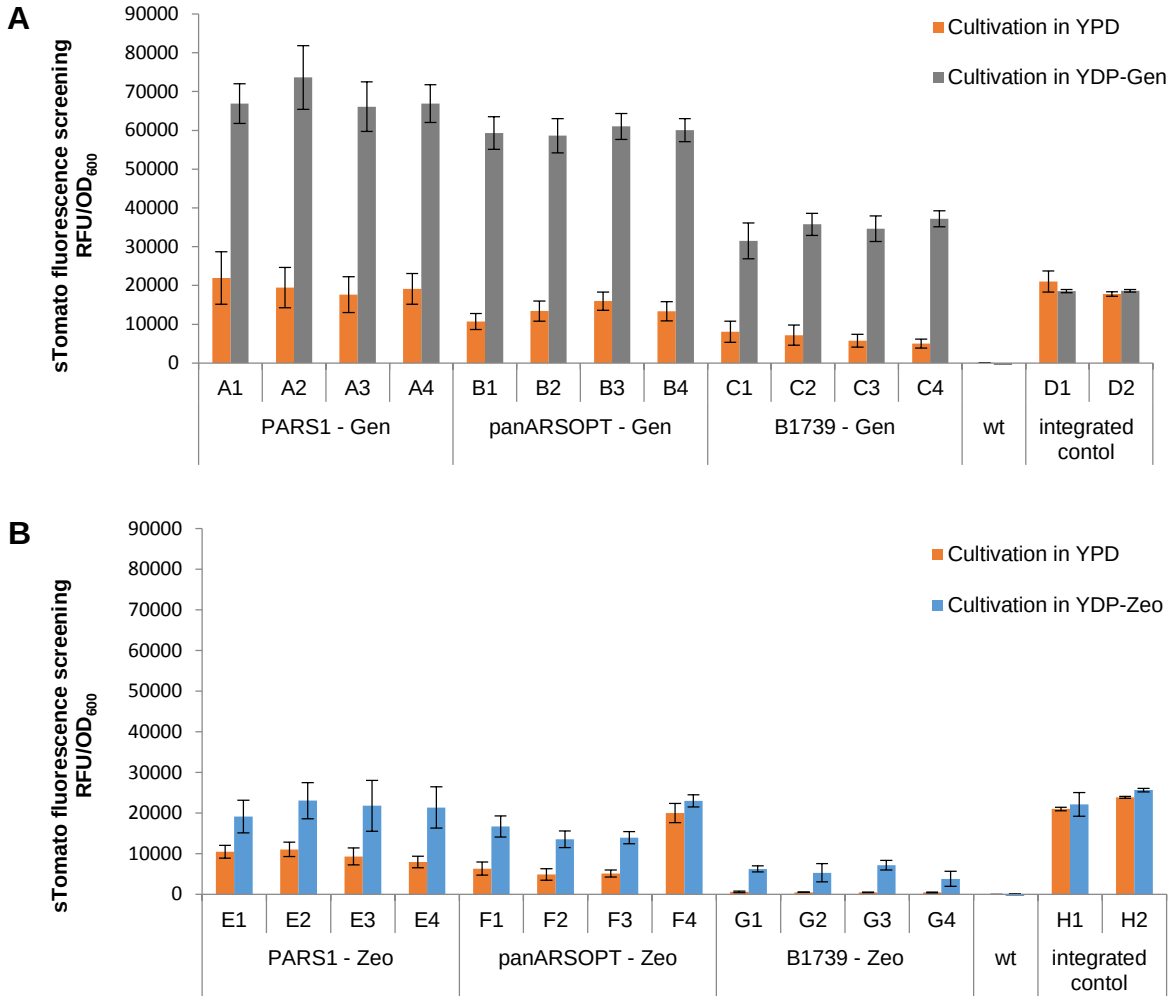


Figure 10: sTomato fluorescence levels of ARS-strains cultivated in media supplemented with Geneticin (A) or Zeocin (B) and antibiotic free media. ARS-plasmids containing the sTomato reporter gene under P_{GAP} were transformed in *P. pastoris* CBS 7435 WT. The plasmids bear either a Zeocin or a Geneticin resistance marker cassette. Single colonies of four representative clones (1, 2, 3, 4) of each construct were inoculated in 96-deep well plates containing YDP-media and YDP-media supplemented with antibiotics for 36 hours (Zeocin: 100 µg/ml or Geneticin: 300 µg/ml). Mean values and standard deviatons of biological seven-fold replicates are shown. **A:** Fluorescence levels of strains bearing a Geneticin resistance marker cassette normalized for cell growth. *P. pastoris* CBS 7435 WT was inoculated as wildtype control. Integrated control: Single colonies of two transformants (D1,D2) bearing a linear plasmid, which did not contain an ARS, integrated in the genome. **B:** Fluorescence levels of strains bearing a Zeocin resistance marker cassette normalized for cell growth. The "panARSOPT" transformant F4 integrated the plasmid in the genome.

Comparing the different ARS plasmids the highest expression levels were obtained with PARS1, using either Geneticin or Zeocin as selection marker. DNA replication can more easily initiate at origins, which are rather AT-rich, in contrast to GC-rich sequences that have a higher melting temperature. Most probably the AT-rich ARS plasmids are present in higher copy numbers in the cells. Similar to Liachko et al. [232] I was able to show that panARSOPT is functional in *P. pastoris*, however approximately 13% less sTomato fluorescence protein was produced with stains bearing a Geneticin cassette and 30% less reporter protein was expressed with a Zeocin cassette in contrast to the strains containing PARS1-plasmids.

The protein expression levels of the different ARS constructs were 3.5-7-fold higher, when Geneticin was used as selection marker. Geneticin, also known as G418, decreases the translation accuracy by inhibiting the elongation step [289], whereas Zeocin intercalates with the DNA and induces ds breaks [290]. The toxicity of the antibiotic apparently correlates with the protein production levels. When using Geneticin in combination with ARSs more protein can be produced in the same time.

Interestingly Zeocin selection based strains show also lower fluorescence levels than Geneticin selection based strains even in the absence of selective conditions (for PARS1 ~20,000 RFU/OD_{600} on YPD for strains bearing an ARS plasmid with Geneticin as selection marker, and only ~10,000 RFU/OD_{600} on YPD for strains bearing an ARS plasmid with Zeocin as selection marker). These effects may hint a negative effect of the expression of the Zeocin resistance gene on ARS plasmids. Upon genomic integration of the controls no clear difference between Zeocin and Geneticin were noticeable.

As controls linear plasmids, expressing sTomato under P_{GAP}, were integrated into the host genome. Additionally the control plasmids were transformed in a circular form, but no colonies were obtained after transformation (Appendix S 1). This indicates that neither the T4-plasmid nor the sTomato CDS bear a sequence, which is recognized as ARS. The sTomato levels of the control strains are similar, when either Zeocin or Geneticin was used as selection marker in YDP media and media with antibiotics.

All ARS-plasmids were unstable in the absence of selection pressure. When the ARS-strains were grown in YPD media a general reduction in the protein production was observed.

In order to further characterize plasmid maintenance of the strains streak-out experiments were performed. Single colonies of each of the transformants used for the fluorescence measurements were streaked out on YDP agar plates as well as on YPD-plates with antibiotics. After two days single colonies from the different agar plates were again streaked out on plates with and without antibiotics (Figure 11). Cells from the YPD-plates were not able to grow on YPD plates containing antibiotics, whereas cells from YPD plates supplemented with antibiotics were able to grow on antibiotics-plates. All ARS vectors were unstable in the absence of selection pressure and it was possible to generate selection marker-free strains.

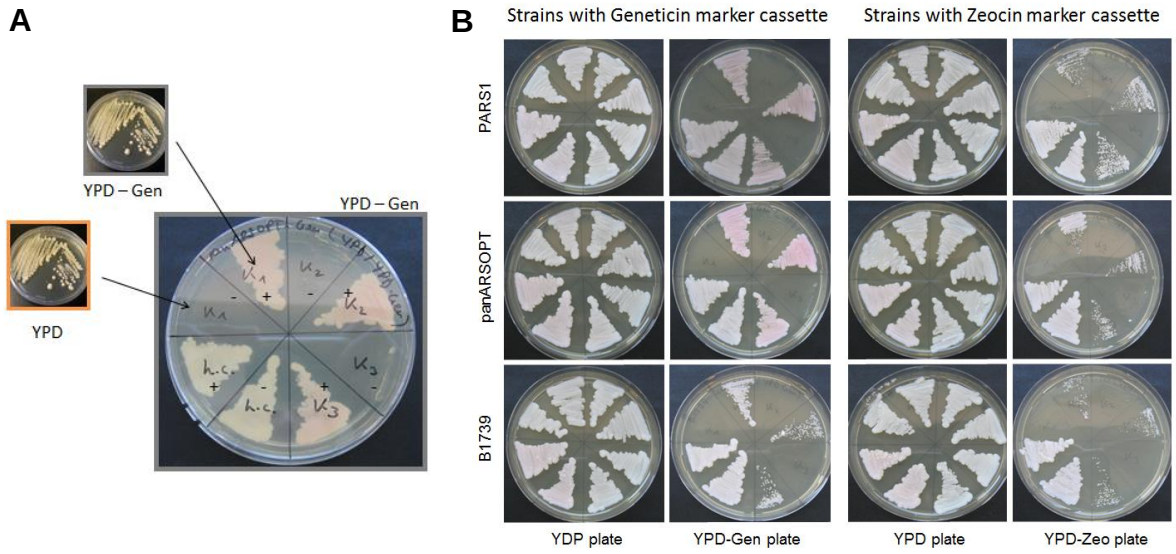


Figure 11: Streak outs of strains expressing the sTomato gene on ARS-vectors. Single colonies of the strains bearing a pPpT4-ARS-sTOM vector were streaked out on YDP plates with and without antibiotics (Zeocin: 100 µg/ml or Geneticin: 300 µl/ml). The plasmids contain either a Zeocin or a Geneticin resistance marker cassette. **A:** The scheme for the streak outs is shown. Strains, which incorporated the linear plasmid without an ARS in the genome, grew on a plate containing antibiotics, independently from previous cultivation conditions (L.c. – linear control, where the expression cassette was integrated into the genome). Strains bearing ARS vectors could only grow on plates supplemented with antibiotics, if they had been precultivated on antibiotics-plates. Three transformants (1, 2, 3) per construct were streaked out. **B:** Streak outs of strains with different ARS-plasmids containing a marker cassette and controls on YDP plates and YDP plates with antibiotics (Zeocin or Geneticin). The strains had been grown on plates with and without selective pressure before (Zeocin: 100 µg/ml or Geneticin: 300 µl/ml).

All three ARS sequences were recognized in *P. pastoris*. The highest expression levels were obtained, when PARS1 was used as plasmid replication origin and similar to the other ARSs, strains can be easily cured from plasmids again. Thus, the PARS1 sequence was used on the CRISPR-Cas9 vectors. Zeocin was chosen as selection marker, since Sturmberger L. had problems in generating knockouts strains, when using Geneticin as selection marker. Nevertheless, plasmids bearing PARS1 and a Geneticin resistance cassette are interesting for small-scale, high level protein production (e.g. for crystallization experiments), where the prize for antibiotics is of minor importance.

Genomic target loci for the evolution of CRISPR-Cas9 in *P. pastoris*

The evolution or optimization of CRISPR-Cas9 requires target loci, which allow easy distinction between active and inactive transformants. An ideal genomic target locus should be easy to screen by applying a simple assay or the CRISPR-Cas9 mutants should display a distinct phenotype. I tried to knockout various genes, which cause to an auxotrophic phenotype when deleted with homologous cassettes in order to identify a locus that can be used for the implementation of CRISPR-Cas9.

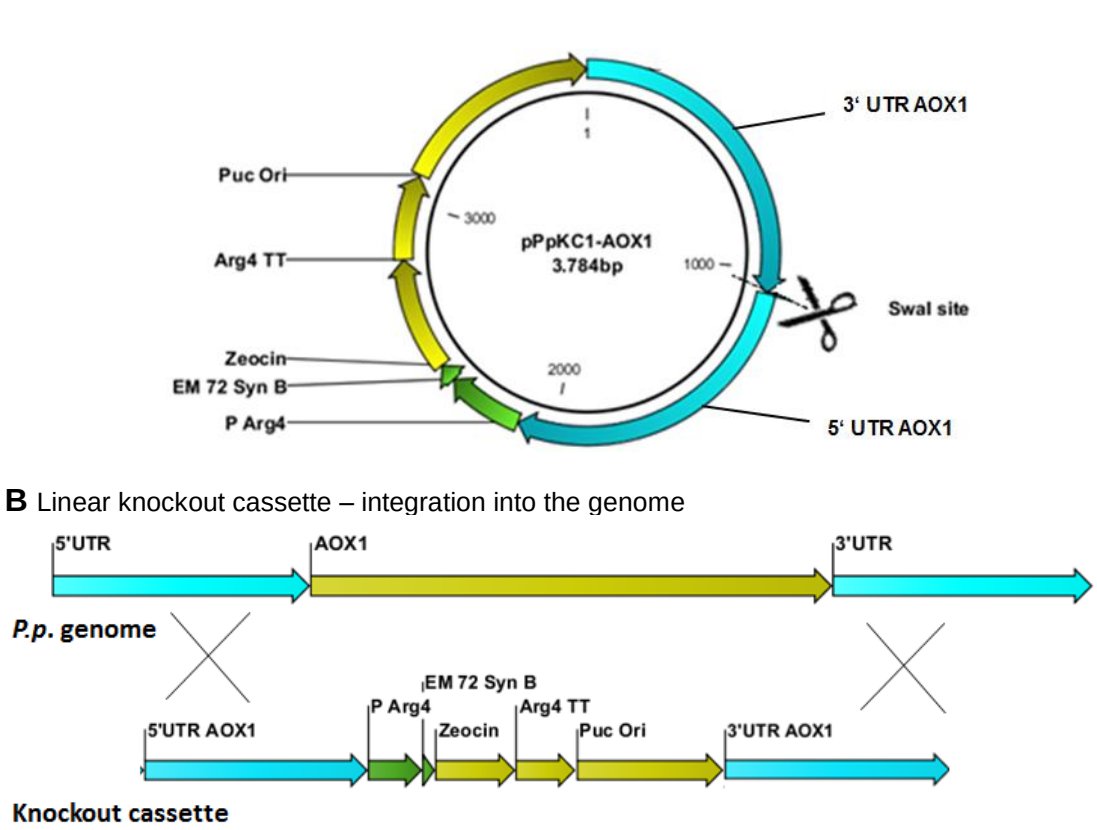


Figure 12: General design of cassettes used to knock out various auxotrophy genes in *P. pastoris* CBS 7435 WT and CBS 7435 *ku70*. The knockout cassette consists of a Zeocin marker cassette flanked by two homologous arms. The arms contain a 1000bp region upstream respectively downstream of the target gene. A cassette to knock out *AOX1* is shown. The elements are not drawn in scale. **A:** Circular vector map of the pPpKC1-AOX1. *Swa*I is used to linearize the cassette. **B:** Integration of the linear knockout cassette into the *P. pastoris* genome.

Originally I planned to design cassettes, which are based on the pPpKC1 plasmid from M. Ahmad. The plasmid contains the flippase under the control of P_{AOX1} and a Zeocin marker cassette flanked by two FRT sites. Five prime and 3' homologous regions of the target gene are selected to complement a *Swa*I cutting site, when they are fused to together by OE-PCR. The FRT flanked marker and flippase cassettes and the 5' and 3' homologous fragment are ligated.

For the *P. pastoris* transformation the plasmid is cut with *Swa*I resulting in a linear knockout cassette with two homologous arms.

However I faced problems, when trying to clone the homologous insert fragment into the vector backbone. I obtained up to seven bands of various sizes after isolated plasmid DNA (IMBT 6602) had been separated on an agarose gel. Also a RE-digest of the vector resulted in inexplicable bands. After I removed the flippase from the vector, the underlying problems were rectified, indicating that the flippase was active in *E. coli* and cut the vector at the FRT-sites.

Thus I used only the Zeocin cassette from the pPpKC1 plasmid and added an approximately 1000 bp 5' upstream and a 1000 bp 3' downstream fragment of the target gene, which complement a *Swa*I site, by Gibson Assembly. I also removed the FRT-sites by RE-digest of the vector, in order to avoid problems with Gibson Assembly due to the identical sequences. For the *P. pastoris* transformation, the plasmid was digested with *Swa*I resulting in a linear knockout cassette, which contains a Zeocin marker cassette flanked by homologous arms (Figure 12). Accordingly, knockout cassettes to target *AOX1*, *GUT1*, *ADE1* and *ADE2* were generated.

One µg of linearized template DNA was transformed into *P. pastoris* CBS 7435 WT as well as into the CBS 7435 *ku70* knockout strain. The transformants were plated on YDP-ager plates containing 25µg/ml Zeocin. A reduced Zeocin concentration is required, because the marker cassette contains a weak promoter (P_{ARG4}) for the expression of the Zeocin resistance gene (personal communication M. Ahmad and L. Sturmberger). Random transformants were picked, cultivated in 96-DWPs containing YDP-Zeocin for 60 h and stamped on minimal plates containing various carbon sources. The transformation plates of *ADE1* and *ADE2* knockout plasmids were incubated at room temperature for two weeks.

In *P. pastoris* wild type strains the HR-rates occur with a frequency of 0.1% up to 30% depending on the target locus and the design of the knockout cassettes [199]. I was able to knockout *AOX1* (9%) and *GUT1* (5%), but I did not obtain red-colored *ADE1* or *ADE2* deficient strains. Interestingly, the recombination efficiencies were also very low for most of the target genes in the CBS 7435 *ku70*, although recombination frequencies up to 100% were reported with the *KU70* knockout strain (Table 9) [199]. The design and the length of the homologous arms of the donor cassette are essential factors to influence the transformation efficiency. Usually the recombination efficiencies can be increased, when using larger homologous cassettes and when

the 5' and 3' end of the cassette are rather AT-rich (personal communication M. Ahmad). Most probably I would have obtained different recombination rates, if I would have systematically tested various lengths of the homologous arms and if I would have taken in consideration the GC-content. The weak promoter of the Zeocin resistance gene and the reduced Zeocin concentration for selection might have also influenced the HR efficiency rates, since strand breaks caused by Zeocin intercalation can be used as starting point for the integration of knockout cassettes.

Table 9: HR frequencies in *P. pastoris* CBS 7435 WT and CBS 7435 *ku70* obtained after the transformation with various homologous cassettes.

Target gene	**CBS 7435 WT** **Auxotrophic / number of transformants screened**	**CBS 7435 *ku70*** **Auxotrophic / number of transformants screened**
GUT1	20/357 (5%)	80/152 (53%)
AOX1	23/252 (9%)	19/ 217 (9%)
ADE2	0/>1000 (0%)	0/>1000 (0%)
ADE1	0/>1000 (0%)	2/>1000 (0.2%)

CRISPR-Cas9 constructs with RNA polymerase III promoters

Transcripts of the RNA polymerase III neither contain a 5'-7-methylguanosine cap nor are polyadenylated. The polyadenylation signal and the 5' cap lead to a translocation of the mRNA to the cytoplasm [250], [251]. However the gRNA has to remain in the nucleus in order to assemble with Cas9. Promoters, which are recognized by the RNA polymerase III have been described in various CRISPR-Cas9 publications and they are the obvious choice for the expression of gRNAs in *P. pastoris*. In *P. pastoris*, to the best of my knowledge, no information on endogenous RNA polymerase III promoters is available. Therefore I started with a broad range of endogenous and exogenous promoter sequences to increase the chances to identify a promoter, which directs the expression of functional gRNAs.

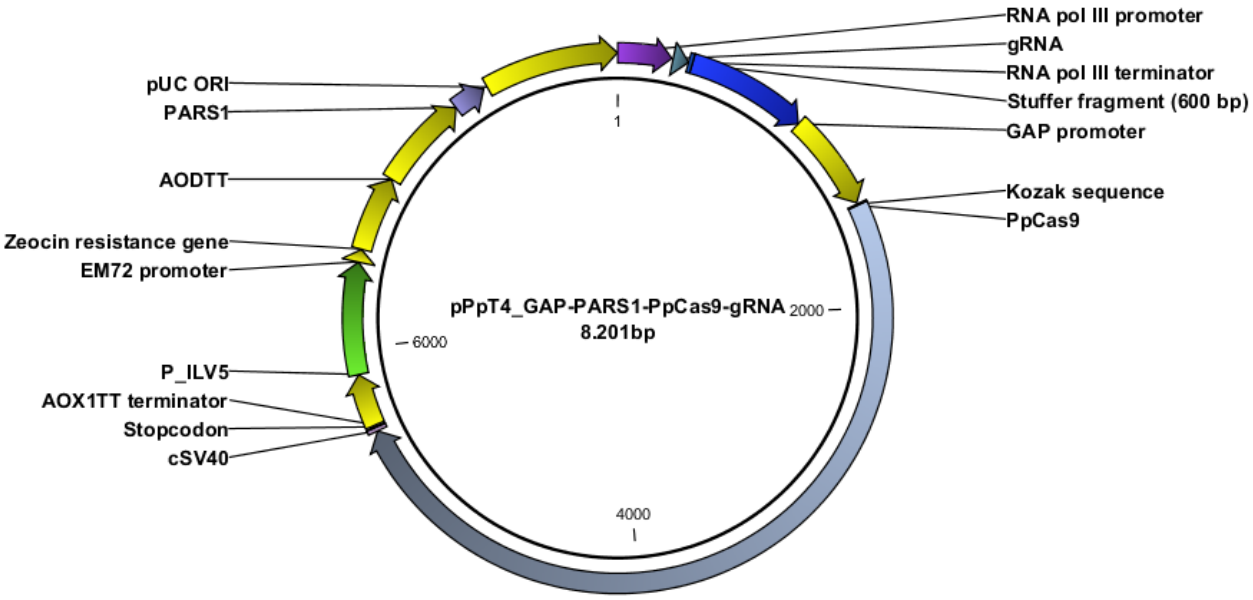


Figure 13: Plasmid map of a CRISPR-Cas9 vector with RNA polymerase III promoter for the expression of gRNAs. The vector is based on the pPpT4_GAP-PARS1 plasmid, which was generated during this thesis. Cas9 is expressed under the control of P_{GAP}. The gRNA expression cassette was cloned downstream of the pUC ORI.

A plasmid map of a typical CRISPR-Cas9 vector bearing a RNA polymerase III promoters is shown in Figure 13. In total ten different RNA polymerase III promoters were tested for the expression of gRNAs, which bind in the *GUT1* locus. *GUT1* encodes for a glycerol kinase in *P. pastoris* and gene-deficient strains show a reduced growth behavior on glycerol compared to the wild type [199]. All constructs are based on the pPpT4_GAP-PARS1 vector, which bears PARS1 as ARS and the constitutive *GAP* promoter for the expression of Cas9. The RNA polymerase III

promoter directs the expression of the gRNA. In total 65 different CRISPR-Cas9 plasmids containing RNA polymerase III promoter were made during this thesis.

The plasmids contain one out of three different Cas9 sequences (*Pp*Cas9, *Hs*Cas9 and *Sp*Cas9) under the constitutive P_{GAP}. Six different RNA polymerase III promoters (*SNR52*, *RPR1* and *SUP4* from *S. cerevisiae*, tRNA promoters: P_{LYS}, P_{MET}, P_{SER} from *P. pastoris*) were tested in combination with each of the Cas9 sequences. The four heterologous RNA polymerase III promoters P_{ASN} from *P. angusta*, *SNR52* from *K. lactis*, *SCR1* from *S. cerevisiae* and *RPR1* from *S. pombe* were only tested in combination with *Pp*Cas9. Each of the RNA polymerase III promoter – Cas9 combinations was tested with at least three out of five gRNAs.

I transformed 100 ng circular plasmid DNA into competent *P. pastoris* cells. Random colonies were picked and grown in YDP-media supplemented with Zeocin. After two days the transformants were stamped on minimal plates containing either glucose or glycerol as C-source. In case that Cas9 and gRNA are functional the complex assembles and introduces a double strand break in the *GUT1* locus, which is repaired by the NHEJ mechanism of the cell. Thereby a frameshift mutation can be introduced and the cells are not able to produce *glycerol kinase-1*. The mutants display reduced growth on glycerol.

However I was not able to generate a single strain showing growth deficiencies on glycerol plates, when using one of the plasmids bearing the *Pp*Cas9 and an endogenous or heterologous RNA polymerase III promoter expressing different gRNAs. Thus, I exchanged the *Pp*Cas9 sequence from the vectors to the *Sp*Cas9 and *Hs*Cas9 sequence. Transformants bearing the *Sp*Cas9 did not show any growth deficiencies on glycerol plates similar to *Pp*Cas9 expressing cells. Interestingly, cells bearing the *Hs*Cas9 showed a reduced growth behavior on glycerol after 24 h incubation, when combined with one of the *P. pastoris* endogenous tRNA promoters P_{LYS}, P_{MET} and P_{SER} (independently from the gRNA used, Figure 14). Cells expressing the *Hs*Cas9 in combination with exogenous RNA polymerase III promoters did not display a retarded growth. However strains expressing the *Hs*Cas9 and *P. pastoris* tRNA promoters still showed an enhanced growth on glycerol compared to the control strain, which bears a homologous cassette containing a Zeocin resistance cassette integrated into the *GUT1* locus. Presumably, the *GUT1* gene was not deleted in all the cells and mixed population were obtained containing *gut1*- cells and cells, which did not contain a mutation. Cells with an intact *GUT1* gene have a growth

advantage and were able to overgrow *gut1*- cells. Possibly functional gRNAs were present at very at low levels.

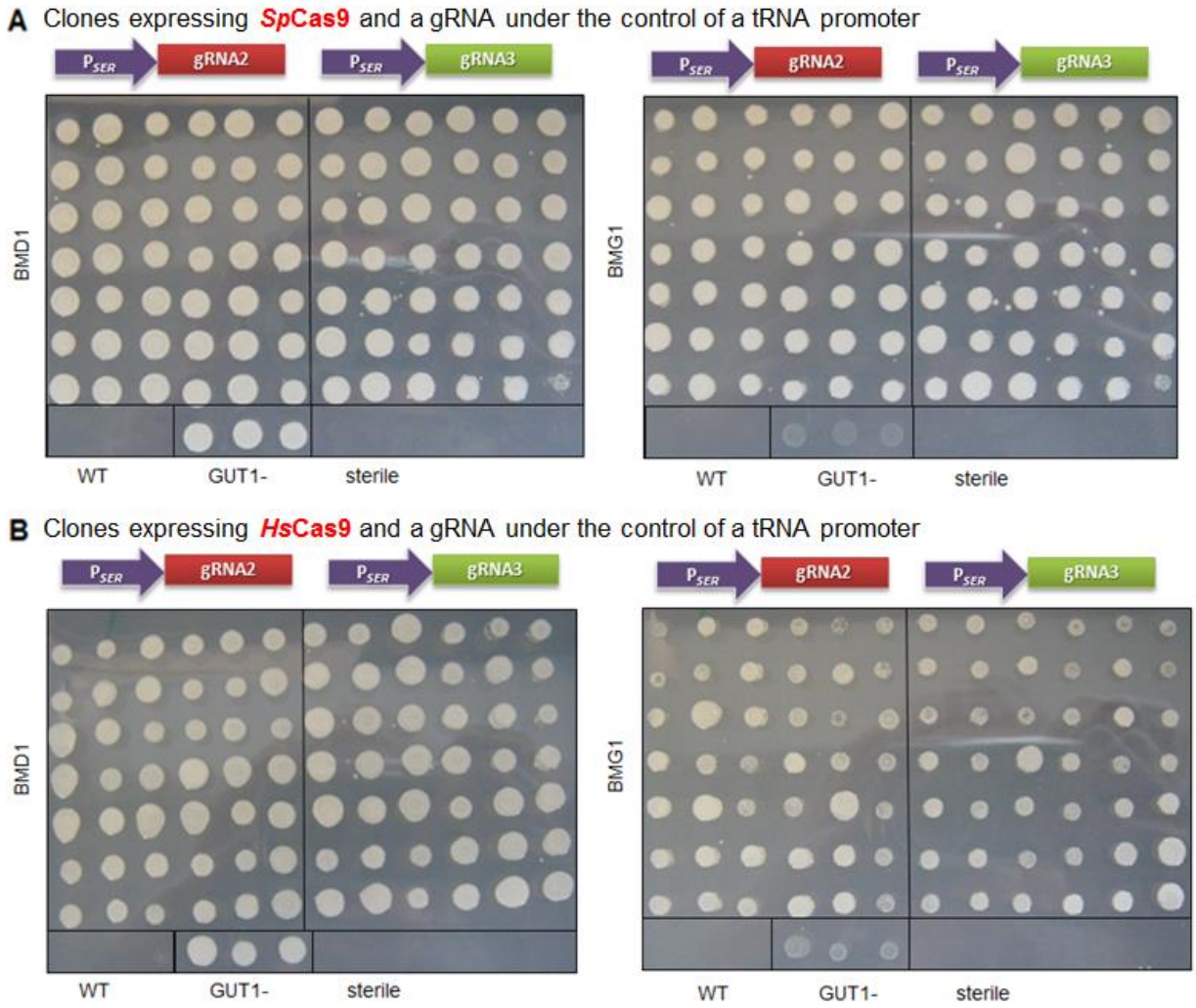


Figure 14: *P. pastoris* CBS 7435 transformants bearing CRISPR-Cas9 plasmids with RNA polymerase III promoters. Random transformants were picked and inoculated in 96-deep well plates containing YPD-Zeo media. The tranformants were stamped on minimal media plates with either glucose (BMD1) or glycerol (BMG1) as C-source. The pictures were taken after 24 hours incubation time. CBS 7435 WT was inoculated, but the strain does not contain a Zeocin resistance cassette and should not be able to grow. The *GUT1*- control strain was generated during this thesis. It contains a Zeocin cassette, which was integrated in the *GUT1* locus. **A:** The transformants express the *Sp*Cas9 sequence under P_{GAP}. The endogenous tRNA RNA polymerase III promoter P_{SER} was used for the expression of the gRNA targeting *GUT1*. Forty-two transformants bearing a vector containing gRNA2 and 42 transformants bear a vector containing the gRNA3. The sequences of the gRNAs are shown in the materials and methods section. **B:** The transformants express the *Hs*Cas9 sequence under P_{GAP}. The endogenous tRNA RNA polymerase III promoter P_{SER} was used for the expression of the gRNA targeting *GUT1*. Forty-two transformants bear a vector containing gRNA2 and 42 transformants bear a vector containing gRNA3.

Streak outs were performed with various transformants expressing the *Hs*Cas9 and *P. pastoris* tRNA promoters on glucose minimal media (BMD1) to obtain single colonies. Eight single colonies of each construct were then restreaked on BMG1 and BMD1 plates, but I was not able to isolate a clone with growth deficiency on glycerol similar to the *GUT1*- control. It remains

unclear, if indeed mixed populations were obtained or if the combination of *Hs*Cas9 and the overexpression of gRNAs, tRNAs or tRNA-gRNA fusions had a negative influence, when the cells grow on glycerol. Interestingly, the "mixed populations" were not obtained, when the cells were grown on glucose. Perhaps a stronger expression construct using a *P. pastoris* derived NLS for the translocation of Cas9 in combination with Geneticin as selection marker might have improved the results.

The *S. cerevisiae* derived *SNR52*, *RPR1*, *SUP4* and *SCR1* in combination with either *Sp*Cas9, *Pp*Cas9 and *Hs*Cas9 did not promote Cas9-directed DNA cleavage. Also the expression of heterologous RNA polymerase III promoters *RPR1* from *S. pombe*, *SNR52* from *K. lactis* and P_{ASN} from *P. angusta* did not effect growth on glycerol. Only the *P. pastoris* derived tRNA promoters P_{LYS}, P_{MET} and P_{SER} in combination with the *Hs*Cas9 sequence influenced the growth behavior on glycerol, whereas on glucose containing media normal growth was observed. The effect was not obtained, when other RNA polymerase III promoters were tested in combination with *Hs*Cas9. This finding indicates that the promoter, which is used for the expression of the gRNAs and also the Cas9 DNA sequence are the main bottlenecks for the functional application of CRISPR-Cas9. In order to avoid RNA polymerase III promoter based expression influences, as a next step ribozyme flanked gRNA expression was performed.

CRISPR-Cas9 constructs with RNA Polymerase II promoter and ribozymes

Ribozymes are self-splicing RNA elements, which can be used to remove 5' and 3' sequences, which are added in the course of the transcription, from gRNAs [264], [265]. I tested the 5' cleaving HH ribozyme in combination with 3' cleaving HDV ribozyme for the expression of gRNAs to target *GUT1*. Using ribozyme makes RNA polymerase III promoters redundant. Thus, I tested a bidirectional RNA polymerase II promoter, P_{HTX1}, for the expression of the ribozyme flanked gRNA and Cas9. Histone promoters are typically active during S-phase, when cell division takes place [291], whereas P_{GAP} drives strong constitutive gene expression [292]. The use of the growth regulated promoter can reduce cellular stress, which is caused by immediate gene expression after the transformation event. P_{HTX1} and P_{GAP} showed similar strength with fluorescent reporter proteins (dissertation Thomas Vogl, Synthetic biology to improve protein expression in *Pichia pastoris*, Graz University of Technology, 2015). In total I made nine different constructs bearing three different gRNAs in combination with either *Sp*Cas9, *Pp*Cas9 or *Hs*Cas9. Figure 15 shows a scheme of the CRISPR-Cas9 ribozyme constructs.

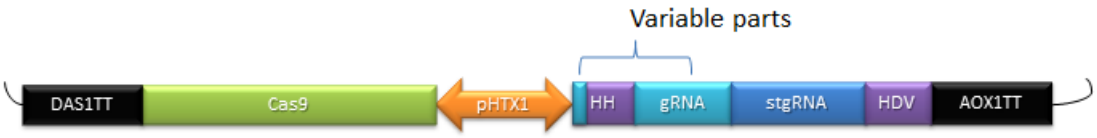


Figure 15: Design of CRISPR-Cas9 constructs using the bidirectional P_{HTX1} promoter and ribozymes for the expression of gRNAs. The constructs are based on the on the pPpT4-HTX1-PARS1 plasmid (A168). The bidirectional P_{HTX1} promoter directs the expression of either *Sp*Cas9, *Pp*Cas9 or *Hs*Cas9 in antisense orientation and the gRNA flanked by the HH and HDV ribozyme in sense orientation. A 6 bp sequence of the HH ribozyme has to be changed according to the variable sequence of the gRNA. The vector parts were not drawn in scale.

I transformed 100 ng circular vector DNA in *P. pastoris* CBS 7435 WT. The transformation rates differed strongly depending on the Cas9 sequence expressed. Expressing *Sp*Cas9 more than 2000 transformants were obtained. Approximately 500 colonies were counted on plates, on which the cells expressing *Hs*Cas9 and a gRNA were grown, and about 50 *Pp*Cas9-expressing colonies per construct were obtained. *Pp*Cas9 appeared to have a toxic effect on the cells indicated by the low transformation rate (Appendix S 2). The transformants were identically cultivated as described for the strains, which had been transformed with CRISPR-Cas9 plasmids bearing a RNA polymerase III promoter and again stamped on minimal media plates containing either glucose or glycerol as carbon source. I was able to generate strains, which show growth deficiencies similar to the *GUT1*- control with *Sp*Cas9, *Pp*Cas9 and *Hs*Cas9 (Figure 16). However, only a very small

number of glycerol deficient transformats (max. 3.5%) was obtained with *Sp*Cas9 or *Pp*Cas9. Whereas *GUT1* targeting efficiencies of 83% to 94% were reached, when *Hs*Cas9 was used in combination with the ribozyme flanked gRNAs. The transformation rates and the targeting efficiencies of all constructs are summarized in Figure 17.

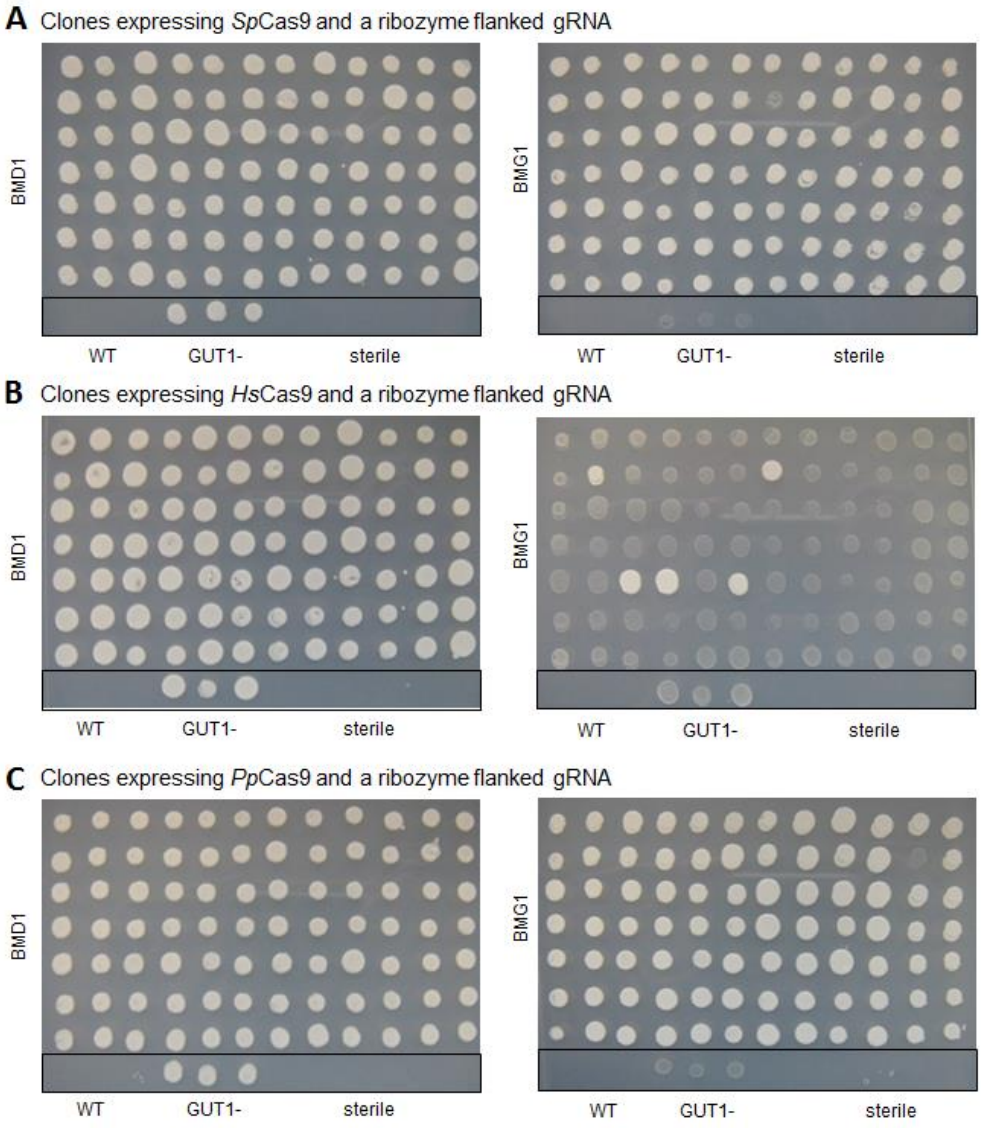


Figure 16: Targeting efficiencies using CRISPR-Cas9 constructs with various Cas9 sequences and ribozyme flanked gRNAs. The number of *GUT1* auxotrophic transformants depends on the Cas9 sequence used. 84 transformants were cultivated in 96-DWPs containing YDP-Zeo media for 48 h and then transferred with a metallic stamp on agar plates containing minimal media with either glucose (BMD1) or glycerol (BMG1) as C-source. CBS 7435 WT was inoculated, but the strain does not contain a Zeocin resistance cassette and should not be able to grow. The *GUT1*- control strain was generated during this thesis, it contains a Zeocin cassette, which was integrated in the *GUT1* locus. **A:** CRISPR-Cas9 vector with *Sp*Cas9 and gRNA3. Only one transformant shows a reduced growth behavior on glycerol. **B:** CRISPR-Cas9 vector with *Hs*Cas9 and gRNA3. 79 out 84 transformants show a reduced growth behavior on glycerol. **C:** CRISPR-Cas9 vector with *Pp*Cas9 and gRNA2 (42 transformants) and CRISPR-Cas9 vector with *Pp*Cas9 and gRNA3 (42 transformants). Only one transformant had retarded growth on glycerol.

Further I isolated genomic DNA of transformants with reduced and normal growth on glycerol, to prove that a CRISPR-Cas9 mediated genome modification had been performed. The *GUT1* locus was PCR amplified and the DNA was sent for sequencing. The sequencing results are summarized in Table 10. In contrast to *S. cerevisiae* where up to approximately 25 bp were removed and random bases were introduced (insertions) [202], one base upstream of the Cas9 cleavage position [176] was deleted in almost 100% of the sequenced clones. Most of the transformants, which were able to grow normally on glycerol, did not contain a mutation in the target sequence. A three basepair deletion was obtained in a few clones, when using gRNA4. The three basepair deletion did not cause a frameshift and also did not influence the protein functionality. One bp upstream of the Cas9 cleavage site was also deleted in the few glycerol deficient strains, which were obtained after the transformation with constructs containing either *Pp*Cas9 or *Sp*Cas9 and various gRNAs. As a control the plasmid containing only the *Hs*Cas9 without a gRNA to target *GUT1* was transformed in *P. pastoris*, but none of the transformants showed reduced growth on glycerol (Appendix S 3).

Table 10: Sequencing results of CRISPR-Cas9 transformants expressing *Hs*Cas9 and either gRNA2, gRNA3 or gRNA4 to target *GUT1*. A reduced growth behavior on glycerol is indicated by (-) and normal growth with a (+) symbol. The PAMs are written in bold letters, the gRNA sequences are underlined and the mutations are in bold and colored in red .

gRNA	Growth	Mutation	Sequenced transformants	Result
gRNA2	-	CGAGTACTCTACCTCT**-**CTCA**GG**ATGATATCAAAAG	8/8	One bp deleted
	+	CGAGTACTCTACCTCTGCTCA**GG**ATGATATCAAAAG	4/4	No mutation
gRNA3	-	TGCAATTTCCTCAGCC**-**GGCTG**GG**TTGAATGTC	8/8	One bp deleted
	+	TGCAATTTCCTCAGCCAGGCTG**GG**TTGAATGTC	4/4	No mutation
gRNA4	-	GTTGTTTGGTCCAAGA**-**GACA**GG**AAAGCCTCTTTA	7/8	One bp deleted
		GTTGTTTGGTCCAAG**--**GACA**GG**AAAGCCTCTTTA	1/8	Two bps deleted
	+	GTTGTTTGGTCCAAGAAGACA**GG**AAAGCCTCTTTA	2/4	No mutation
		GTTGTTTGGTCCAAGA**---**CA**GG**AAAGCCTCTTTACA	2/4	Three bps deleted

CRISPR-Cas9 genome editing in *P. pastoris* CBS 7435 *ku70*

In order to further characterize CRISPR-Cas9 mediated genome engineering in *P. pastoris* the CRISPR-Cas9 constructs, which contain the gRNAs flanked by ribozymes and various Cas9 sequences under control of P_{HTX1}, were transformed in a *KU70* deficient strain. Ku70p is involved in the NHEJ-repair mechanism and significantly increased HR-rates can be obtained using *P. pastoris* CBS 7435 *ku70* compared to the wild-type [199]. The transformation rates with the CRISPR-Cas9 plasmids were lower compared to the wildtype strain (Appendix S 2). Less than 15 transformants per plate were obtained with the constructs bearing either *Hs*Cas9 or

*Pp*Cas9 and one of the three gRNAs. None of the *Pp*Cas9 transformants showed reduced growth on glycerol, whereas up 73% of the clones bearing *Hs*Cas9 were unable to grow on glycerol. The number of clones obtained after the transformation with *Sp*Cas9 differed depending on the gRNA sequence. Transformations with *Sp*Cas9 and gRNA2 or gRNA4 resulted in more than 1000 transformants. None of the clones showed reduced growth on glycerol. Approximately 150 colonies were obtained, when a plasmid bearing *Sp*Cas9 and gRNA3 had been transformed. Seven transformants were cultivated in YDP-Zeo and transferred on minimal plates, of which all showed reduced growth on glycerol (Figure 17). I isolated genomic DNA of growth deficient and normally growing transformants of all constructs. Interesting I was not able to amplify the *GUT1* locus of auxotrophic mutants with the primers, which had been used for the amplification of genomic DNA from wildtype CRISPR-Cas9 transformants. *P. pastoris KU70*- transformants with wild type like growth did not contain a mutation in the target sequence.

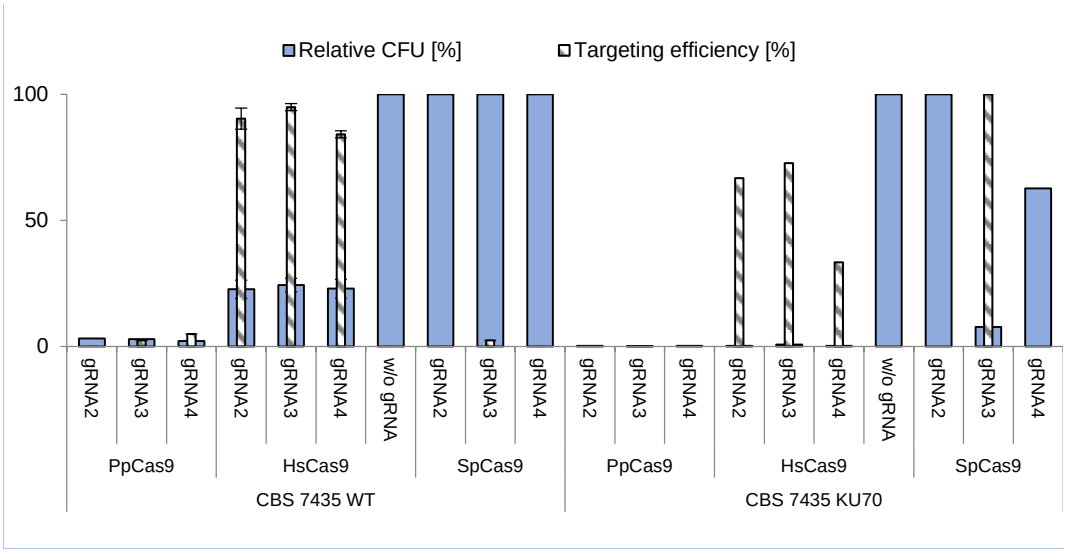


Figure 17: CRISPR-Cas9 induced NHEJ mutations in *P. pastoris* CBS 7435 WT and *ku70* strain. CRISPR-Cas9 plasmids containing either *Pp*Cas9, *Hs*Cas9 or *Sp*Cas9 and a gRNA to target *GUT1* were transformed in *P.pastoris* CBS 7435 wildtype and a *KU70* knockout strain (CBS7435 *ku70*) according to protocol described in the materials and methods section. The transformants were cultivated for 48 h transformants in 96-DWPs containing YDP-Zeocin media. The cells were transferred with a metallic stamp on minimal media agar plates with either glucose (BMD1) or glycerol (BMG1) as C-source. Functional Cas9 and gRNAs assemble and introduce a NHEJ-mediated mutation resulting in a reduced growth behavior on glycerol.

The NHEJ mechanims is used to repair ds breaks, which have been introduced with CRISPR-Cas9, when no homologous template is present in the cells. The CBS 7435 *ku70* knockout strain is lacking a protein, which is one of the key players in NHEJ-mediated DNA repair allowing the cells to seal strand breaks preferentially by HR. Introducing ds breaks with the CRISPR-Cas9 system is especially toxic for the *KU70* deficient strain and only a few transformants were

obtained, of which some showed reduced growth behavior on glycerol depending on the Cas9 sequence (HsCas9 in combination with gRNA2, gRNA3 or gRNA4, SpCas9 in combination with gRNA3) and the gRNA. It was not possible to amplify the *GUT1* locus from auxotrophic mutants with primers that have been used to amplify genomic DNA of wild type transformants. Most probably translocation events took place, which caused the rearrangement of non-homologous parts of the chromosomes [293].

CRISPR-Cas9 and donor DNA co-transformation in *P. pastoris* CBS 7435 WT and *P. pastoris* CBS 7435 *ku70*

The integration of a homologous DNA cassette at a genomic locus requires the HR-repair machinery of the cell and an endogenous or exogenous (transformed) homologous donor sequence. The specific integration rate can be increased by using site specific nucleases introducing a strand break at the desired integration locus [10]–[13]. Without a donor DNA, such double stranded breaks can only be repaired by the NHEJ mechanism, which may result in short indels in the ORF. In contrast, offering a donor DNA allows also to completely delete an ORF, to replace it with a different sequence or to insert sequences coding e.g. for a tag. In order to show beneficial effects of CRISPR-Cas9 introduced strand breaks on the HR-mediated integration rate of donor cassettes, co-transformation experiments using various donor cassettes were performed in *P. pastoris* CBS 7435 WT and CBS 7435 *ku70* (Figure 18, Figure 19).

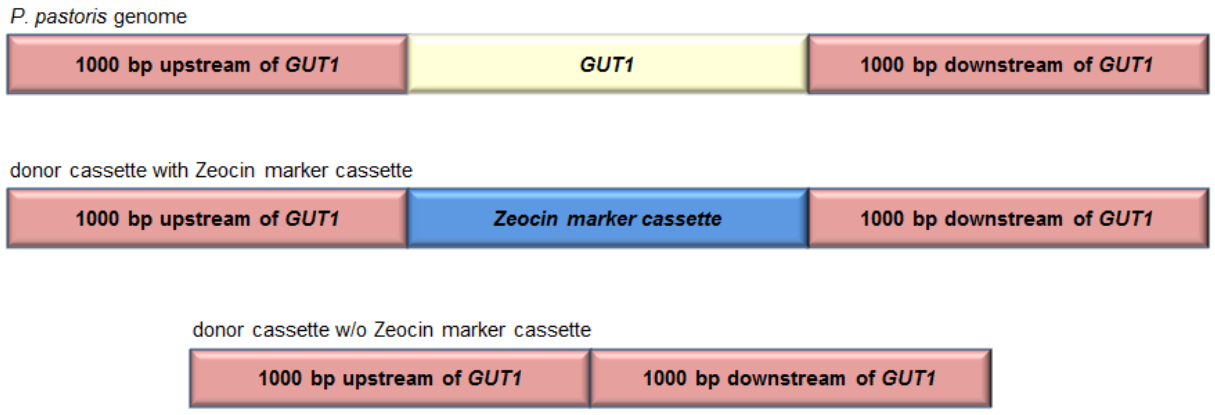


Figure 18: Overview of different *GUT1* knockout cassettes used for the co-transformantion with CRISPR-Cas9 vectors. The donor DNA with Zeocin marker cassette consisted of a Zeocin marker cassette flanked by homologous arms and the donor DNA w/o Zeocin marker cassette of a 1000 bp sequence upstream of the *GUT1* CDS directly fused a 1000 bp downstream sequence. The construction of the donor cassettes is described in the material and methods section. The elements are not drawn in scale.

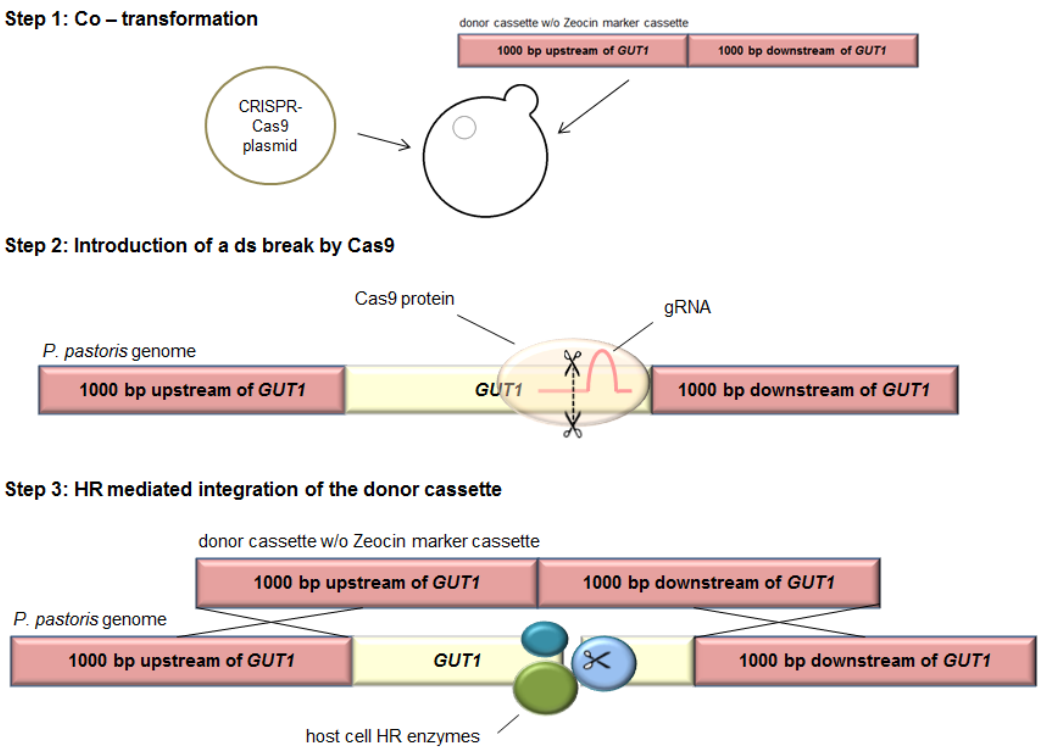


Figure 19: Principle of CRISPR-Cas9 mediated integration of donor cassettes to allow the complete deletion of an ORF without a resistance marker. As first step a CRISPR-Cas9 plasmid and a homologous donor cassette are co transformed into *P. pastoris* cells. The nuclease Cas9 introduces a double strand break at the target locus. Host cell HR enzymes mediate the integration of the homologous donor fragment. The elements are not drawn in scale.

I co-transformed a *GUT1* donor cassette (1 µg) and a CRISPR-Cas9 plasmid (100 ng) containing the *Hs*Cas9 and a gRNA, which targets the *GUT1* gene. The co-transformation experiments were performed with different donor cassettes: One consisted of a Zeocin marker cassette flanked by homologous arms (donor DNA with Zeocin marker cassette) and the other consisted of two homologous arms directly fused together (Figure 18). The cassette without a resistance marker would allow the complete removal of ORFs without leaving a resistance marker in the genome. The donor DNA containing the Zeocin marker cassette was used to determine the integration rates, when no CRISPR-Cas9 plasmid was co-transformed. The donor cassette containing the Zeocin cassette was additionally used for the co-transformation experiments, since the size of the donor DNA could have influenced the integration rates (2 kbp – donor DNA w/o Zeocin marker cassette, 3.8 kbp – donor DNA with Zeocin marker). However no differences in the integration efficiencies of the different cassettes could be observed.

By applying Zeocin selection, only cells, which contain a CRISPR-Cas9 plasmid survived after the transformation. It was not possible to discriminate between cells, which contain only a CRISPR-Cas9 plasmid and cells, which contain the CRISPR-Cas9 plasmid and a donor cassette. Ideally all living transformants contain a CRISPR-Cas9 plasmid and a donor cassette and the homologous recombination rate of the donor is increased up to 100% by the site specific strand break. In case this theory is correct, co-transforming CRISPR-Cas9 plasmids would allow the use of marker-free homologous recombination cassettes for the replacement of genomic DNA. This turned out to be not the case in the wildtype strain, where NHEJ remained the preferred pathway for repair. However, co-transforming a CRISPR-Cas9 plasmid in the *KU70*- strain further increased the homologous recombination rate (Figure 20).

The transformation rates were slightly increased in the *KU70*- strain (between 20 to 30 transformants per construct independently from the donor cassette used) compared to transformations, where only a CRISPR-Cas9 plasmid had been transformed (less than 15 viable clones per construct). The number of *P. pastoris* wildtype transformants (donor cassette + CRISPR-Cas9 plasmid) was similar to the transformations, where only a CRISPR-Cas9 plasmid was transformed (~500 viable clones). However these results have to be confirmed by repeating the transformations in triplicates. The transformants were cultivated in YPD-Zeocin media and stamped on minimal plates containing either glucose or glycerol as carbon source. I isolated the genomic DNA of a representative number of transformants (>10 per construct) and sequenced the *GUT1* locus. The majority of the *P. pastoris* wildtype transformants contained a NHEJ-repair mediated deletion mutation. This suggests that in the *P. pastoris* wild type strain NHEJ is the preferred repair pathway, even if a homologous DNA template is offered. In contrast 78-94% of the *KU70*- transformants integrated the donor cassettes (Figure 20). By co-transforming CRISPR-Cas9 plasmids it was possible to improve the HR-mediated integration rates in the *KU70*- strain up to 44% compared to the control (50% knockout frequency), where the solely the homologous donor cassette was transformed (no CRISPR-Cas9 plasmid).

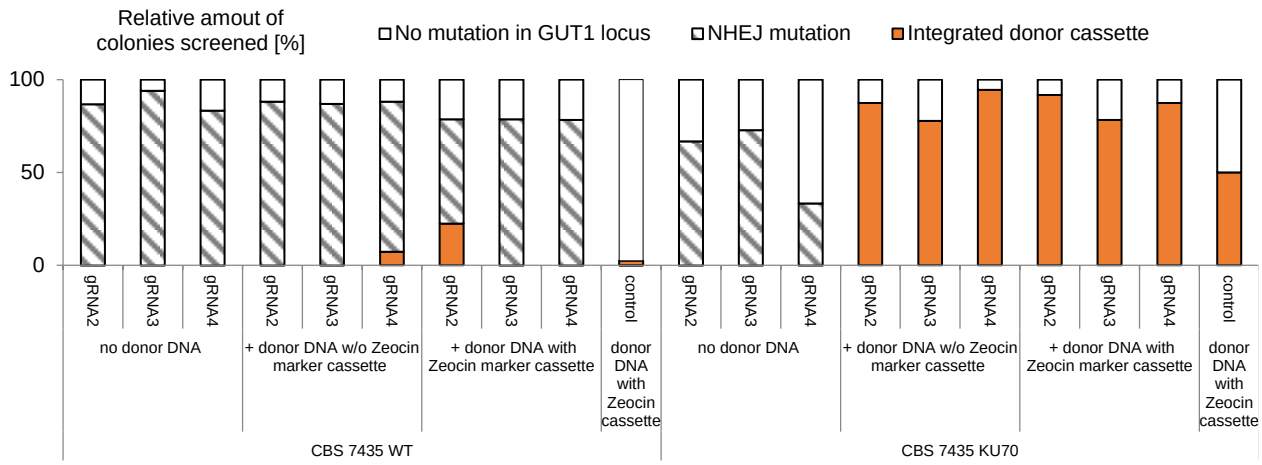


Figure 20: CRISPR-Cas9 mediated integration of homologous donor cassettes does not strongly affect the wildtype strain, but improves specific integration in the *KU70*- strain. 100 ng CRISPR-Cas9 circular plasmid DNA and 1 µg donor DNA were co-transformed in *P. pastoris* CBS7435 WT and CBS 7435 *ku70*. The construction of the donor DNA fragments is described in the materials and methods section. The transformants were cultivated for 48 h transformants in 96-DWPs containing YDP-Zeocin media. The cells were transferred with a metallic stamp on minimal media agar plates with either glucose (BMD1) or glycerol (BMG1) as C-source. >10 glycerol deficient transformants per construct were sequenced.

The CRISPR-Cas9 system was successfully used to integrate a marker-free knockout cassette in the *KU70*- strain. Typical knockout cassettes consist of a selection marker cassette flanked by two homologous arms, which are required for the site specific integration. Upon integration the selection marker is recycled for instance by the use of a site specific recombinase [199]. By transforming a marker-free knockout cassette in combination with a CRISPR-Cas9 plasmid to improve targeted integration, it was possible to completely remove the *GUT1* CDS in the *KU70*- strain at reasonable rates (approximately 8 out 10 transformants integrated the marker-free cassette). The CRISPR-Cas9 plasmid can be simply removed by growing the strains on non-selective media (see results and discussion: Plasmid maintenance and curing). The repair of harmful ds breaks is preferentially performed by the NHEJ mechanism in the *P. pastoris* wildtype strain and only in the *KU70*- strain, where integration of homologous cassette occurs at high rates independently from the CRISPR-Cas9 system [199], an improvement for the integration of homologous cassettes was observed. However the CRISPR-Cas9 system might be further improved for HR-mediated integration of donor fragments in *P. pastoris* wildtype strains by using the Cas9 nickase [17], [186]. The nickase introduces ss breaks, which are used as a starting point of HR-mediated DNA repair [177]. Thus, the frequency of harmful ds breaks and of (unwanted) NHEJ-repair mediated mutations could be reduced.

Plasmid maintenance and curing

In order to prove that the CRISPR-Cas9 plasmids can be removed (cured) from the mutant strains streak-out experiments were performed. Cell material of auxotrophic strains from the stamped minimal plates (no selection) was streaked out on YDP and YPD-Zeocin (100 µg/mL) agar plates. Although the cells were already grown on selection free media, they were still able to grow on YDP-Zeocin plates. After two days of incubation cell material from the YDP-plate was streaked out on a YDP-agar plate containing 100 µg/mL Zeocin and a YDP plate. The CRISPR-Cas9 mutant strains were able to grow on YPD-media, but did not grow on media supplemented with Zeocin (Figure 21). Thus, using CRISPR-Cas9 enables the generation of selection marker-free deletion strains in a short time. As a control the cells were also streaked out on BMG1 minimal media plates without antibiotics. The CRISPR-Cas9 *GUT1*- transformants show a reduced growth on glycerol similar to the *GUT1*- control, which contains a knockout cassette integrated in the *GUT1* locus.

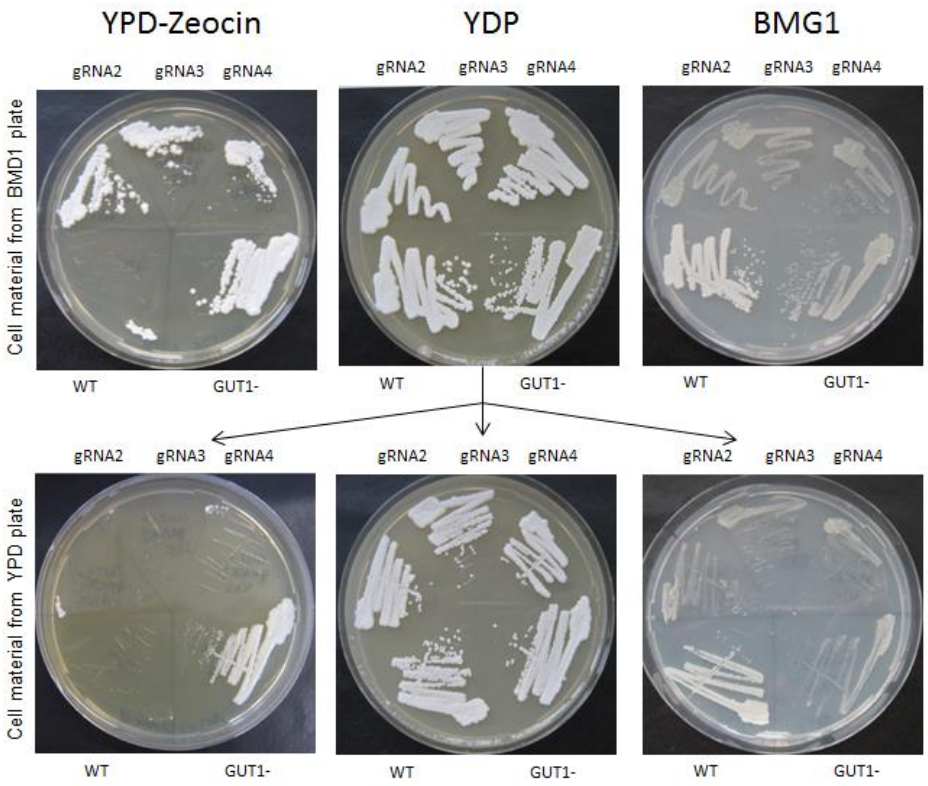


Figure 21: Plasmid maintenance of auxotrophic CRISPR-Cas9 transformants. The CRISPR-Cas9 transformants lost the ARS plasmid after growth on antibiotic free media. Cell material of CRISPR-Cas9 transformants bearing *Hs*Cas9 and either gRNA2, gRNA3 or gRNA4 on an ARS plasmid, which had been stamped on BMD1 plates, was streaked out on various agar plates. Cells from BMD1 plates can grow on YDP-Zeocin (100 µg/mL), YDP and BMG1 plates. Cell material from the YDP plate was restreaked on different agar plates. The cells can grow on YDP, but not on YDP-Zeocin plates (100 µg/mL), since they lost the ARS plasmid. The strains show a reduced growth behavior on glycerol compared to CBS 7435 WT on BMG1 (minimal media containing glycerol) plates. The *GUT1*- control strain contains a Zeocin cassette integrated in the *GUT1* locus and can grow on Zeocin containing media.

Characterization of the developed CRISPR-Cas9 system

Additional CRISPR-Cas9 target loci

In order to further characterize the developed CRISPR-Cas9 system I designed gRNAs to introduce NHEJ mutations in various *P. pastoris* genes, which are located on different chromosomes. *AOX1* encodes for the alcohol oxidase 1 and converts methanol to formaldehyde. *AOX1* knockout strains display a reduced phenotype, when grown on methanol as C-source (Mut^S) [201]. *MXR1*, *MPP1* and *PRM1* encode for transcriptional activators involved in the regulation of methanol regulated genes and activate the transcription of *AOX1*. The deletion of *MXR1* or *PRM1* also leads to a reduced growth phenotype on methanol [286]. For *MPP1*, so far no *P. pastoris* knockout strain had been reported, but in *P. angusta* (*Hansenula polymorpha*) knocking out the *MPP1* homologue led to a growth defect on methanol [294]. CRISPR-Cas9 plasmids expressing *Hs*Cas9 and different gRNAs flanked by ribozymes under the control of P_{HTX1} were transformed into *P. pastoris* wildtype cells. Random clones were picked, cultivated in 96-deep well plates and stamped on minimal plates containing either glucose or methanol as carbon source. In case a NHEJ mutation had been introduced in the target loci a reduced growth phenotype on methanol could be observed. The mutation frequencies of *PRM1, AOX1* and *MPP1* for the best gRNAs approach 100%. However, when some gRNAs were used, hardly any transformants with reduced growth on methanol were obtained (e. g. PRM1-gRNA2, MXR1-gRNA1, MXR1-gRNA3). The target sequences could be spliced out or inaccessible to the nuclease. Also the gRNAs might not be incorporated correctly into Cas9 due to sequence specific RNA-folding. Two of the gRNAs designed to target *MXR1* were hardly functional and an average targeting frequency of only 43% was obtained with MXR1-gRNA2, indicating that *MXR1* is difficult to target with CRISPR-Cas9. These results also suggest that it is beneficial to test more than one gRNA, when trying to introduce mutations in specific locus. Some gRNAs might not be functional at all and screening a larger number of transformants will not be increase the probability to identify a CRISPR-Cas9 mutant.

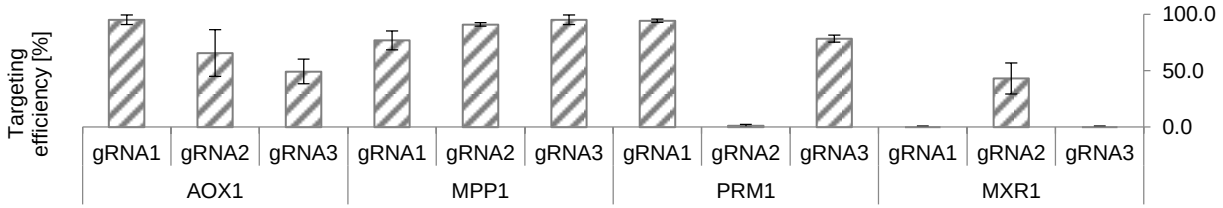


Figure 22: CRISPR-Cas9 targeting efficiencies of various target genes. Plasmids expressing *Hs*Cas9 and various gRNAs were transformed into *P. pastoris* CBS 7435 WT. The nuclease and the gRNA assemble and introduce a ds break in the target locus. Four different genes (*AOX1*, *MPP1*, *PRM1* and *MXR1*) were targeted using three different gRNAs. The mutation efficiencies vary depending on the gRNA sequence used. Mean and SD of three independent transformations are shown.

Multiplexing

CRISPR-Cas9 can be used to mutate several genomic loci by co-expressing different gRNAs [190], [191]. I also tried to simultaneously mutate two genomic loci, *GUT1* and *AOX1*, in *P. pastoris* by expressing two gRNAs on a CRISPR-Cas9 plasmid. The plasmids contained *Hs*Cas9, gRNA3 to target *GUT1* and a gRNA to target the *AOX1* locus. Upon transformation of the CRISPR-Cas9 plasmids Cas9 introduces a ds break in the *GUT1* as well as in the *AOX1* CDS. The strand breaks are then repaired by the NHEJ machinery. The plasmids were transformed into a *P. pastoris* wildtype strain and random clones were picked and cultivated in 96-deep well plates. Then the transformants were transferred with a metallic stamp on BMG1, BMM1 and BMD1 plates. In case multiplexing was successful the transformants displayed a reduced growth on methanol and glycerol plates.

Multiplexing had a negative effect on the transformation efficiency. A 10-fold reduction of CFUs was observed compared to transformations using CRISPR-Cas9 plasmids and a single gRNA. All transformants showed normal growth on BMD1 plates containing glucose as C-source. Reduced growth on glucose might have indicated detrimental effects, caused by the co-expression of two gRNAs. Both target loci were successfully deleted in 70.4% of the transformants, when GUT1-gRNA3 and AOX1-gRNA2 were combined. Although targeting efficiencies for AOX1-gRNA1 and GUT1-gRNA3 were approximately 90%, when only one gRNA was expressed, the mutation efficiency decreased to 3.5 %, when AOX1-gRNA1 and GUT1-gRNA3 were simultaneously expressed. The multiplexing rates obviously differ depending on the gRNA combination. Hardly any transformant contained a mutation in only one of the target loci, indicating that CRISPR-Cas9 can be used to modify two different genomic loci in *P. pastoris* at the same time. However toxic off-targeting effects multiply due to the increased number of gRNAs. The results have to be confirmed by repeating the experiments in triplicates.

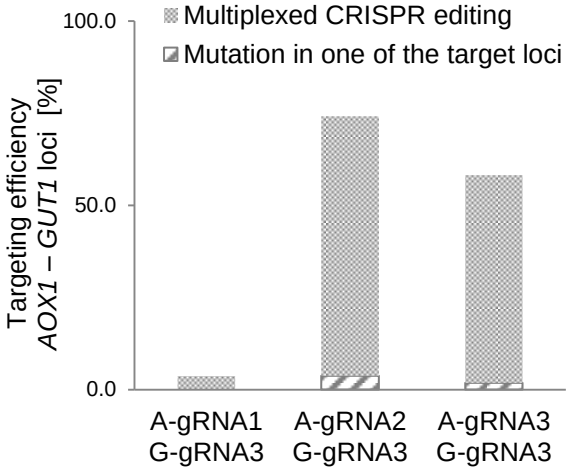


Figure 23: CRISPR-Cas9 multiplexing in *P. pastoris*. GUT1-gRNA3 and a gRNA to target *AOX1* (gRNA1, gRNA2 or gRNA3) were co-expressed on a CRISPR-Cas9 plasmid bearing *Hs*Cas9. The plasmids were transformed into *P. pastoris* CBS 7435 WT cells. Depending on the gRNA combination used multiplexing efficiencies up to 70.4% were obtained.

Synthetic TALENs to target *GUT1*

TALENs contain a DNA binding domain of single repeats that can be engineered to bind a desired DNA sequence, the catalytic domain of the *Fok*I nuclease and a NLS for the translocation to the eukaryotic nucleus. Two TALEN have to bind on adjacent DNA sequences and *Fok*I has to dimerize to introduce a strand break [18]. Strand breaks increase the rates of homologous recombination and introducing various errors (such as indels) by several orders of magnitude [10]–[13]. I designed TALEN, which cut similar to CRISPR-Cas9 in the *GUT1* CDS and planned to compare the targeting efficiency rates of the different genome engineering tools.

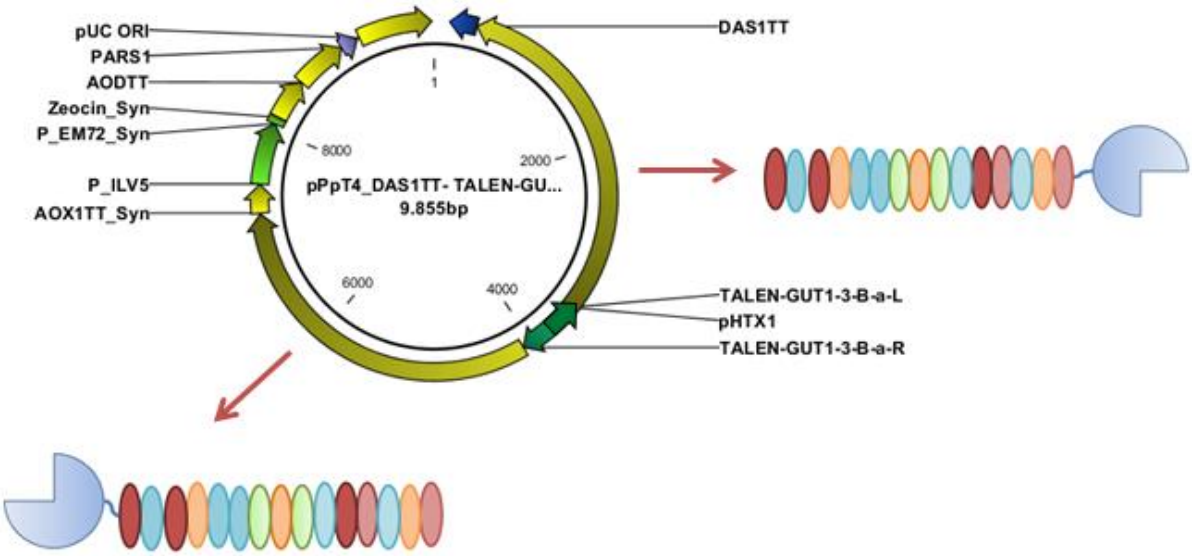


Figure 24: Design of vectors for the expression of TALENs, which target the *GUT1* locus in *P. pastoris*. The vector is based on the pPpT4-HTX1-PARS1 plasmid (A168). The CDSs of the monomeric TALENs, which bind in close proximity to enable the dimerization of *Fok*I and the introduction of a ds break, are cloned in the vector in a successive manner.

TALEN single repeats contain of 30-42 identical amino acids and the assembly of the DNA binding domains is time consuming and laborious. Due to the high repetitiveness the sequence of the DNA binding domain cannot be ordered as synthetic ds DNA fragments (gBlocks), but they have be assembled e.g. by a PCR based method or by hierarchical ligation as described in the introduction. Thus I ordered three synthetic TALENs on vectors of the company Genecopeia, which can be used for the expression in mammalian cells. The mammalian vectors contain the DNA binding domain to target the *GUT1* CDS, the SV40 NLS N-terminally and the *Fok*I monomer C-terminally fused to the DNA binding domain under the control of the *CMV* promoter

and a Hygromycin resistance cassette. Two TALEN vectors with different DNA binding domains (e. g. 3BaL and 3BaR) have to be co-transformed into the host. I amplified the TALEN CDS consisting of the NLS, the DNA binding domain and the *Fok*I sequence from the Genecopoeia vectors in order to clone them into a *P. pastoris* expression vector. I planned to express the two corresponding TALEN sequences under the control of the bidirectional histone promoter P_{HTX1} (Figure 24). However, PCRs with various primer pairs resulted in faint bands and a lot of background smear was obtained. I never succeeded to clone the insert fragments in the intended backbone vector. Thus I cloned the insert fragments in a pJET1.2 cloning vector and sent plasmid DNA for sequencing. Sequencing revealed that the insert was correct before and after the repetitive domain, but that large parts of the repeats were missing. I analyzed the sequence map of the TALEN monomers and discovered that the repeats are made of identical DNA sequences except from the RVD motif.

I also sequenced plasmid DNA, which was isolated from Genecopoeia strains, but here the repetitive DNA binding domain corresponded to the sequence map. In *E.coli* repetitive sequences are highly unstable and get lost easily. Most probably the company is using *E. coli* strains with a special genetic background (e.g. RecBCD-deficient strains [295]) with reduced recombination rates compared to our *E.coli* TOP10F' cloning strain. Self-binding of the repetitive domains might have also caused the low PCR yields and the large amount of unspecific bands.

The cloning and the transformation should be repeated with a suitable *E. coli* cloning strain, however it remains unclear if the repetitive sequences are stable in *P. pastoris*. TALEN vectors are expensive, since the DNA binding domain has to be purchased from external companies (~600 € for three TALEN pairs) and there are no *P. pastoris* TALEN vectors commercially available. Recloning of TALEN-CDS in *P. pastoris* vectors is troublesome due to the highly repetitive sequences. Despite all drawbacks TALEN are interesting for platform strain generation, since off-targeting sites do not occur in a mammalian sized genome. However RNA-guided *Fok*I nucleases (RFN), which are based on the CRISPR-Cas9 technology, combine the ease of reprogramming and the precise targeting similar to TALENs [197] and they represent a potent alternative to TALENs for applications, where off-targeting is not desired.

5. SUMMARY AND OUTLOOK

Implementing CRISPR-Cas9 in a given organism is reliant on a set of interdependent features. Efficient genome targeting can only be achieved, when all single components are correctly designed, expressed, produced and assembled in the host. The NLS for the import of Cas9 has to be recognized by the host nuclear import machinery. Cas9 has to be produced, folded and post translationally modified correctly. The nuclear import and the production of the 160 kDa protein should not have a toxic effect on the cells. The gRNA should remain in the nucleus and 5' or 3' RNA sequences, which might be added in the course of the transcription, must not have a negative influence on correct folding and functionality. Cas9 and the gRNA have to assemble in nucleus and the RNA-protein complex has to recognize and introduce a strand break at the targeting locus. Unspecified off-targeting and toxicity effects caused by overexpression of the single components must not have toxic or lethal effects on the cells.

I was able to implement a CRISPR-Cas9 system, which is functional in *P. pastoris*. The system currently relies on a human codon optimized Cas9 sequence (*Hs*Cas9) and a gRNA, which is flanked by two different ribozymes under control of the bidirectional histone promoter P_{HTX1}. The SV40 NLS, which is fused C-terminally to Cas9, is responsible for the translocation of the heterologous protein. The CRISPR-Cas9 system is expressed on an ARS-vector, which contains PARS1 for the autonomous replication and a Zeocin resistance marker cassette. The gRNA to target a desired genomic locus can be simply added to the vector by the aid of short gBlock fragment. The vector has to be opened with the RE-enyzme *Not*I, of which the recognition site is located between P_{HTX1} and the *AOX1* RNA polymerase II terminator. The gBlock, containing the gRNA flanked by the HH and the HDV ribozyme can be cloned into either with RE-enzymes or Gibson Assembly. Instructions for CRISPR-Cas9 vector design can be found in the Appendix.

Outlook

I characterized a set of heterologous and endogenous NLSs. However, only the C-terminal application of the SV40 sequence was tested for the import of Cas9. It remains unclear, if other NLSs have a positive effect on the CRISPR-Cas9 system. A strong NLS may be combined with a weak promoter for the expression of the heterologous protein (considering also homo- or heterologous transcription factors or interesting bacterial and bacteriophage proteins) to save cellular resources and improve the overall fitness.

Homologous donor templates were integrated at low levels in the *P. pastoris* CBS 7435 WT strain, when co-transformed with CRISPR-Cas9 plasmids. According to literature reports with other hosts the CRISPR-Cas9 system might be further improved for HR-mediated integration of donor fragments by using the Cas9 nickase [17], [186]. Cas9 consists of two catalytic domains and the Cas9 nickase can simply be generated by introducing a point mutation is in one of the domains. The Cas9 nickase introduces ss breaks, which are repaired by HR [177]. Thus, the frequency of harmful ds breaks and of (unwanted) NHEJ-repair mediated mutations is reduced.

A major general drawback of the CRISPR-Cas9 system is the high off-targeting rate. When applying CRISPR-Cas9 several gRNAs should be tested to introduce a mutation in a target locus to ensure that a certain phenotypic effect is caused by the deletion of the target gene and not due to an undesired off targeting mutation. The use of dimeric Cas9-*Fok*I fusion proteins [197] allows to cut only once in the *P. pastoris* genome. The fusion proteins consist of a catalytically inactive Cas9 fused to a *Fok*I monomer. Two RNA-guided *Fok*I nucleases (RFN) have bind in close proximity to enable the dimerization of *Fok*I. These dimeric RFNs can be used for the generation of platform strains and reduce time for double checking CRISPR-Cas9 mediated knockout and deletion effects.

At the moment the CRISPR-Cas9 plasmid is only available containing a Zeocin resistance marker cassette. However many interesting *P. pastoris* strains contain a Zeocin cassette integrated in the genome. In order to introduce mutations a CRISPR-Cas9 vector containing a different selection marker cassette might be useful. In addition ARS plasmid based expression can be up-regulated using Geneticin compared to Zeocin.

Currently, two ribozymes are required to guaranty to expression of a functional gRNA. A short sequence of the 5' cutting ribozyme always has to be changed depending on the variable sequence of the gRNA. In order to simplify the design and to shorten the time for vector construction, it would be beneficial if the 5' ribozyme can be omitted.

CRISPR-Cas9 is a powerful genome engineering method, which has been referred to as holy grail for genome editing [177]. It combines high targeting rates and enormous flexibility in terms reprogramming. The system, which was primarily described in *S. pyogenes*, was adapted in various eukaryotic species including whole animals, cell lines or various yeasts [55], [177], [296]. I developed a CRISPR-Cas9 system for biotechnological important yeast *P. pastoris*

during the course of this master thesis. The system was used to introduce deletions in various genomic loci, for the integration homologous cassettes in the *GUT1* locus and for multiplexing. Researchers can now apply CRISPR-Cas9 to generate *P. pastoris* knockout strains or to perform deletion studies for basic research as well as for industry related applications in a minimum of time.

6. BIBLIOGRAPHY

[1] P. A. Silver, J. C. Way, F. H. Arnold, and J. T. Meyerowitz, "Synthetic biology: Engineering explored.," *Nature*, vol. 509, no. 7499, pp. 166–7, May 2014.

[2] M. Heinemann and S. Panke, "Synthetic biology-putting engineering into biology.," *Bioinformatics*, vol. 22, no. 22, pp. 2790–9, Nov. 2006.

[3] K. M. Esvelt and H. H. Wang, "Genome-scale engineering for systems and synthetic biology.," *Mol. Syst. Biol.*, vol. 9, no. 641, Jan. 2013.

[4] T. M. Conrad, N. E. Lewis, and B. Ø. Palsson, "Microbial laboratory evolution in the era of genome-scale science.," *Mol. Syst. Biol.*, vol. 7, no. 509, Jan. 2011.

[5] S. F. Elena and R. E. Lenski, "Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation.," *Nat Rev Genet*, vol. 4, no. 6, pp. 457–469, Jun. 2003.

[6] M. R. Lieber, "The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway.," *Annu. Rev. Biochem.*, vol. 79, pp. 181–211, Jan. 2010.

[7] M. R. Capecchi, "Altering the genome by homologous recombination," *Sci.* , vol. 244 , no. 4910, pp. 1288–1292, Jun. 1989.

[8] A. Dudás and M. Chovanec, "DNA double-strand break repair by homologous recombination.," *Mutat. Res.*, vol. 566, no. 2, pp. 131–67, Mar. 2004.

[9] D. E. Barnes, "Non-homologous end joining as a mechanism of DNA repair.," *Curr. Biol.*, vol. 11, no. 12, pp. R455–7, Jun. 2001.

[10] P. Rouet, F. Smih, and M. Jasin, "Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease.," *Mol. Cell. Biol.*, vol. 14, no. 12, pp. 8096–106, Dec. 1994.

[11] F. Smih, P. Rouet, P. J. Romanienko, and M. Jasin, "Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells.," *Nucleic Acids Res.*, vol. 23 , no. 24, pp. 5012–5019, Jan. 1995.

[12] K. W. Caldecott, "Single-strand break repair and genetic disease," *Nat Rev Genet*, vol. 9, no. 8, pp. 619–631, Aug. 2008.

[13] F. Storici, C. L. Durham, D. A. Gordenin, and M. A. Resnick, "Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 25, pp. 14994–9, Dec. 2003.

[14] P. A. Carr and G. M. Church, "Genome engineering.," *Nat. Biotechnol.*, vol. 27, no. 12, pp. 1151–62, Dec. 2009.

[15] P. Perez-Pinera, D. G. Ousterout, and C. A. Gersbach, "Advances in targeted genome editing," *Curr. Opin. Chem. Biol.*, vol. 16, no. 3–4, pp. 268–277, Aug. 2012.

[16] A. Nern, B. D. Pfeiffer, K. Svoboda, and G. M. Rubin, “Multiple new site-specific recombinases for use in manipulating animal genomes.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 34, pp. 14198–203, Aug. 2011.

[17] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. a Doudna, and E. Charpentier, “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.,” *Science*, vol. 337, no. 6096, pp. 816–21, Aug. 2012.

[18] J. C. Miller, S. Tan, G. Qiao, K. a Barlow, J. Wang, D. F. Xia, X. Meng, D. E. Paschon, E. Leung, S. J. Hinkley, G. P. Dulay, K. L. Hua, I. Ankoudinova, G. J. Cost, F. D. Urnov, H. S. Zhang, M. C. Holmes, L. Zhang, P. D. Gregory, and E. J. Rebar, “A TALE nuclease architecture for efficient genome editing.,” *Nat. Biotechnol.*, vol. 29, no. 2, pp. 143–8, Feb. 2011.

[19] N. Sternberg, D. Hamilton, and R. Hoess, “Bacteriophage P1 site-specific recombination: II. Recombination between loxP and the bacterial chromosome.,” *J. Mol. Biol.*, vol. 150, no. 4, pp. 487–507, Aug. 1981.

[20] D. A. Sorrell and A. F. Kolb, “Targeted modification of mammalian genomes.,” *Biotechnol. Adv.*, vol. 23, no. 7–8, pp. 431–69, Nov. 2005.

[21] D. L. Hamilton and K. Abremski, “Site-specific recombination by the bacteriophage P1 lox-Cre system: Cre-mediated synapsis of two lox sites.,” *J. Mol. Biol.*, vol. 178, no. 2, pp. 481–486, Sep. 1984.

[22] A. Loonstra, M. Vooijs, H. B. Beverloo, B. a Allak, E. van Drunen, R. Kanaar, a Berns, and J. Jonkers, “Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 16, pp. 9209–14, Jul. 2001.

[23] D. P. Silver and D. M. Livingston, “Self-Excising Retroviral Vectors Encoding the Cre Recombinase Overcome Cre-Mediated Cellular Toxicity.,” *Mol. Cell*, vol. 8, no. 1, pp. 233–243, Nov. 2014.

[24] G. Proteau, D. Sidenberg, and P. Sadowski, “The minimal duplex DNA sequence required for site-specific recombination promoted by the FLP protein of yeast in vitro.,” *Nucleic Acids Res.* , vol. 14, no. 12, pp. 4787–4802, Jun. 1986.

[25] B. Sauer, “Functional expression of the cre-lox site-specific recombination system in the yeast Saccharomyces cerevisiae.,” *Mol. Cell. Biol.*, vol. 7, no. 6, pp. 2087–2096, 1987.

[26] N. L. Craig, “The Mechanism of Conservative Site-Specific Recombination.,” *Annu. Rev. Genet.*, vol. 22, no. 1, pp. 77–105, Dec. 1988.

[27] J.P. Cabaniols, L. Mathis, and C. Delenda, “Targeted gene modifications in drug discovery and development.,” *Curr. Opin. Pharmacol.*, vol. 9, no. 5, pp. 657–63, Oct. 2009.

[28] P. J. Enyeart, S. M. Chirieleison, M. N. Dao, J. Perutka, E. M. Quandt, J. Yao, J. T. Whitt, A. T. Keatinge-Clay, A. M. Lambowitz, and A. D. Ellington, “Generalized bacterial genome editing using mobile group II introns and Cre-lox.,” *Mol. Syst. Biol.*, vol. 9, no. 685, Jan. 2013.

[29] T. Schlake and J. Bode, “Use of Mutated FLP Recognition Target (FRT) Sites for the Exchange of Expression Cassettes at Defined Chromosomal Loci.,” *Biochemistry*, vol. 33, no. 43, pp. 12746–12751, Nov. 1994.

[30] R. Torres, A. García, M. Payá, and J. C. Ramirez, “Non-integrative lentivirus drives high-frequency cre-mediated cassette exchange in human cells.,” *PLoS One*, vol. 6, no. 5, p. e19794, Jan. 2011.

[31] S. Turan, M. Galla, E. Ernst, J. Qiao, C. Voelkel, B. Schiedlmeier, C. Zehe, and J. Bode, “Recombinase-mediated cassette exchange (RMCE): traditional concepts and current challenges.,” *J. Mol. Biol.*, vol. 407, no. 2, pp. 193–221, Mar. 2011.

[32] A. Baer and J. Bode, “Coping with kinetic and thermodynamic barriers: RMCE, an efficient strategy for the targeted integration of transgenes.,” *Curr. Opin. Biotechnol.*, vol. 12, no. 5, pp. 473–480, Oct. 2001.

[33] B. Sauer, “Recycling selectable markers in yeast,” *Biotechniques*, vol. 16, no. 6, pp. 1086–1088, Jun. 1994.

[34] T. Vogl, F. S. Hartner, and A. Glieder, “New opportunities by synthetic biology for biopharmaceutical production in Pichia pastoris.,” *Curr. Opin. Biotechnol.*, vol. 24, no. 6, pp. 1094–101, Dec. 2013.

[35] Z. Ivics and Z. Izsvák, “The expanding universe of transposon technologies for gene and cell engineering.,” *Mob. DNA*, vol. 1, no. 1, p. 25, Jan. 2010.

[36] T. Wicker, F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel, and A. H. Schulman, “A unified classification system for eukaryotic transposable elements.,” *Nat Rev Genet*, vol. 8, no. 12, pp. 973–982, Dec. 2007.

[37] I. Grabundzija, M. Irgang, L. Mátés, E. Belay, J. Matrai, A. Gogol-Döring, K. Kawakami, W. Chen, P. Ruiz, M. K. L. Chuah, T. VandenDriessche, Z. Izsvák, and Z. Ivics, “Comparative analysis of transposable element vector systems in human cells.,” *Mol. Ther.*, vol. 18, no. 6, pp. 1200–9, Jun. 2010.

[38] Y. Le Breton, N. P. Mohapatra, and W. G. Haldenwang, “In vivo random mutagenesis of Bacillus subtilis by use of TnYLB-1, a mariner-based transposon.,” *Appl. Environ. Microbiol.*, vol. 72, no. 1, pp. 327–33, Jan. 2006.

[39] C. Dumais Pope, L. Dhand, and N. P. Cianciotto, “Random mutagenesis of Legionella pneumophila with mini-Tn10.,” *FEMS Microbiol. Lett.*, vol. 124, no. 1, pp. 107–111, Nov. 1994.

[40] P. J. Fernandes, J. A. Powell, and J. A. Archer, “Construction of Rhodococcus random mutagenesis libraries using Tn5 transposition complexes.,” *Microbiology*, vol. 147, no. Pt 9, pp. 2529–36, Sep. 2001.

[41] C. F. Weil and R. Kunze, “Transposition of maize Ac/Ds transposable elements in the yeast Saccharomyces cerevisiae,” *Nat Genet*, vol. 26, no. 2, pp. 187–190, Oct. 2000.

[42] S. Parinov, I. Kondrichin, V. Korzh, and A. Emelyanov, “Tol2 transposon-mediated enhancer trap to identify developmentally regulated zebrafish genes in vivo.,” *Dev. Dyn.*, vol. 231, no. 2, pp. 449–59, Oct. 2004.

[43] K. A. Skipper, P. R. Andersen, N. Sharma, and J. G. Mikkelsen, “DNA transposon-based gene vehicles - scenes from an evolutionary drive.,” *J. Biomed. Sci.*, vol. 20, p. 92, Jan. 2013.

[44] Z. Ivics, P. B. Hackett, R. H. Plasterk, and Z. Izsvák, “Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells.,” *Cell*, vol. 91, no. 4, pp. 501–10, Nov. 1997.

[45] B. McClintock, “The origin and behavior of mutable loci in maize.,” *Proc. Natl. Acad. Sci.*, vol. 36 , no. 6 , pp. 344–355, Jun. 1950.

[46] C. Miskey, “The Frog Prince: a reconstructed transposon from Rana pipiens with high transpositional activity in vertebrate cells.,” *Nucleic Acids Res.*, vol. 31, no. 23, pp. 6873–6881, Dec. 2003.

[47] L. Cary, M. Goebel, B. Corsaro, and H. Wang, “Transposon mutagenesis of baculoviruses: Analysis of Trichoplusia ni transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses.,” *Virology*, vol. 172, no. 1, pp. 156–69, Sep. 1989.

[48] S. T. Thibault, M. a Singer, W. Y. Miyazaki, B. Milash, N. a Dompe, C. M. Singh, R. Buchholz, M. Demsky, R. Fawcett, H. L. Francis-Lang, L. Ryner, L. M. Cheung, A. Chong, C. Erickson, W. W. Fisher, K. Greer, S. R. Hartouni, E. Howie, L. Jakkula, D. Joo, K. Killpack, A. Laufer, J. Mazzotta, R. D. Smith, L. M. Stevens, C. Stuber, L. R. Tan, R. Ventura, A. Woo, I. Zakrajsek, L. Zhao, F. Chen, C. Swimmer, C. Kopczynski, G. Duyk, M. L. Winberg, and J. Margolis, “A complementary transposon tool kit for Drosophila melanogaster using P and piggyBac.,” *Nat. Genet.*, vol. 36, no. 3, pp. 283–7, Mar. 2004.

[49] X. Feng, A. L. Bednarz, and S. D. Colloms, “Precise targeted integration by a chimaeric transposase zinc-finger fusion protein.,” *Nucleic Acids Res.*, vol. 38, no. 4, pp. 1204–16, Mar. 2010.

[50] S. R. Yant, Y. Huang, B. Akache, and M. a Kay, “Site-directed transposon integration in human cells.,” *Nucleic Acids Res.*, vol. 35, no. 7, p. e50, Jan. 2007.

[51] J. B. Owens, D. Mauro, I. Stoytchev, M. S. Bhakta, M.-S. Kim, D. J. Segal, and S. Moisyadi, “Transcription activator like effector (TALE)-directed piggyBac transposition in human cells.,” *Nucleic Acids Res.*, vol. 41, no. 19, pp. 9197–207, Oct. 2013.

[52] A. M. Rushforth, B. Saari, and P. Anderson, “Site-selected insertion of the transposon Tc1 into a Caenorhabditis elegans myosin light chain gene.,” *Mol. Cell. Biol.*, vol. 13, no. 2, pp. 902–910, Feb. 1993.

[53] J. L. Bessereau, A. Wright, D. C. Williams, K. Schuske, M. W. Davis, and E. M. Jorgensen, “Mobilization of a Drosophila transposon in the Caenorhabditis elegans germ line.,” *Nature*, vol. 413, no. 6851, pp. 70–4, Sep. 2001.

[54] L. Cooley, R. Kelley, and A. Spradling “Insertional mutagenesis of the Drosophila genome with single P elements.,” *Science*, vol. 239, no. 4844, pp. 1121–8, Mar. 1988.

[55] T. Gaj, C. A. Gersbach, and C. F. Barbas, “ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering.,” *Trends Biotechnol.*, vol. 31, no. 7, pp. 397–405, Jul. 2013.

[56] J. G. Doench, C. P. Petersen, and P. A. Sharp, “siRNAs can function as miRNAs.,” *Genes Dev.*, vol. 17, no. 4, pp. 438–42, Feb. 2003.

[57] Y. Zeng, R. Yi, and B. R. Cullen, “MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 17, pp. 9779–84, Aug. 2003.

[58] G. Hutvágner and P. D. Zamore, “A microRNA in a multiple-turnover RNAi enzyme complex.,” *Science*, vol. 297, no. 5589, pp. 2056–60, Sep. 2002.

[59] M. Karbiener, C. Glantschnig, and M. Scheideler, “Hunting the Needle in the Haystack: A Guide to Obtain Biologically Meaningful MicroRNA Targets.,” *Int. J. Mol. Sci.*, vol. 15, pp. 20266–20289, Oct. 2014.

[60] F. Ratcliff, S. MacFarlane, and D. Baulcombe, “Gene silencing without DNA. RNA-mediated cross-protection between viruses.,” *Plant Cell*, vol. 11, no. 7, pp. 1207–16, Jul. 1999.

[61] L. He and G. J. Hannon, “MicroRNAs: small RNAs with a big role in gene regulation.,” *Nat Rev Genet*, vol. 5, no. 7, pp. 522–531, Jul. 2004.

[62] Y. Lee, C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Rådmark, S. Kim, and V. N. Kim, “The nuclear RNase III Drosha initiates microRNA processing.,” *Nature*, vol. 425, no. 6956, pp. 415–9, Oct. 2003.

[63] E. Bernstein, A. Caudy, S. M. Hammond, and G. J. Hannon, “Role for a bidentate ribonuclease in the initiation step of RNA interference.,” *Nature*, vol. 409, no. 6818, pp. 363–6, Jan. 2001.

[64] S. W. Knight and B. L. Bass, “A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans.,” *Science*, vol. 293, no. 5538, pp. 2269–71, Sep. 2001.

[65] C. Matranga, Y. Tomari, C. Shin, D. P. Bartel, and P. D. Zamore, “Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes.,” *Cell*, vol. 123, no. 4, pp. 607–20, Nov. 2005.

[66] Y. Tomari, C. Matranga, B. Haley, N. Martinez, and P. D. Zamore, “A protein sensor for siRNA asymmetry.,” *Science*, vol. 306, no. 5700, pp. 1377–80, Nov. 2004.

[67] S. M. Elbashir, “RNA interference is mediated by 21- and 22-nucleotide RNAs.,” *Genes Dev.*, vol. 15, no. 2, pp. 188–200, Jan. 2001.

[68] D. P. Bartel, “MicroRNAs: genomics, biogenesis, mechanism, and function.,” *Cell*, vol. 116, no. 2, pp. 281–97, Jan. 2004.

[69] M. Lagos-Quintana, R. Rauhut, W. Lendeckel, and T. Tuschl, “Identification of novel genes coding for small expressed RNAs.,” *Science*, vol. 294, no. 5543, pp. 853–8, Oct. 2001.

[70] N. C. Lau, L. P. Lim, E. G. Weinstein, and D. P. Bartel, “An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans.,” *Science*, vol. 294, no. 5543, pp. 858–62, Oct. 2001.

[71] R. C. Lee and V. Ambros, “An extensive class of small RNAs in Caenorhabditis elegans.,” *Science*, vol. 294, no. 5543, pp. 862–4, Oct. 2001.

[72] Z. Mourelatos, J. Dostie, S. Paushkin, A. Sharma, B. Charroux, L. Abel, J. Rappsilber, M. Mann, and G. Dreyfuss, “miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs.,” *Genes Dev.*, vol. 16, no. 6, pp. 720–8, Mar. 2002.

[73] L. P. Lim, N. C. Lau, P. Garrett-Engele, A. Grimson, J. M. Schelter, J. Castle, D. P. Bartel, P. S. Linsley, and J. M. Johnson, “Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs.,” *Nature*, vol. 433, no. 7027, pp. 769–73, Mar. 2005.

[74] B. P. Lewis, I. Shih, M. W. Jones-Rhoades, D. P. Bartel, and C. B. Burge, “Prediction of Mammalian MicroRNA Targets.,” *Cell*, vol. 115, no. 7, pp. 787–798, Dec. 2003.

[75] J. R. Lytle, T. a Yario, and J. a Steitz, “Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5’ UTR as in the 3' UTR.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 23, pp. 9667–72, Jun. 2007.

[76] L. Misquitta and B. M. Paterson, “Targeted disruption of gene function in Drosophila by RNA interference (RNA-i): a role for nautilus in embryonic somatic muscle formation.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 4, pp. 1451–6, Mar. 1999.

[77] S. M. Elbashir, J. Harborth, W. Lendeckel, a Yalcin, K. Weber, and T. Tuschl, “Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells.,” *Nature*, vol. 411, no. 6836, pp. 494–8, May 2001.

[78] P. J. Paddison, A. a Caudy, E. Bernstein, G. J. Hannon, and D. S. Conklin, “Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells.,” *Genes Dev.*, vol. 16, no. 8, pp. 948–58, Apr. 2002.

[79] T. R. Brummelkamp, R. Bernards, and R. Agami, “A system for stable expression of short interfering RNAs in mammalian cells.,” *Science*, vol. 296, no. 5567, pp. 550–3, Apr. 2002.

[80] F. V Rivas, N. H. Tolia, J. J. Song, J. P. Aragon, J. Liu, G. J. Hannon, and L. J.Tor, “Purified Argonaute2 and an siRNA form recombinant human RISC.,” *Nat. Struct. Mol. Biol.*, vol. 12, no. 4, pp. 340–9, Apr. 2005.

[81] O. Donzé and D. Picard, “RNA interference in mammalian cells using siRNAs synthesized with T7 RNA polymerase.,” *Nucleic Acids Res.*, vol. 30, no. 10, p. e46, Mar. 2002.

[82] D. Nikolova and D. Toncheva, “RNA interference - Regulations and application in oncology.,” *Journal of Cancer Molecules*, vol. 4, pp. 67–77, May 2008.

[83] R. Niwa and F. J. Slack, “The evolution of animal microRNA function.,” *Curr. Opin. Genet. Dev.*, vol. 17, no. 2, pp. 145–50, Apr. 2007.

[84] C. A. Sledz, M. Holko, M. J. de Veer, R. H. Silverman, and B. R. G. Williams, “Activation of the interferon system by short-interfering RNAs.,” *Nat. Cell Biol.*, vol. 5, no. 9, pp. 834–839, Sep. 2003.

[85] C. A. Sledz and B. R. G. Williams, “RNA interference in biology and disease.,” *Blood*, vol. 106, no. 3, pp. 787–94, Aug. 2005.

[86] A. L. Jackson, J. Burchard, J. Schelter, B. N. Chau, M. Cleary, L. Lim, and P. S. Linsley, “Widespread siRNA ‘off-target’ transcript silencing mediated by seed region sequence complementarity.,” *RNA*, vol. 12, no. 7, pp. 1179–87, Jul. 2006.

[87] F. Michel and J. Feral, “Structure and Activities of Group II Introns.,” *Annu. Rev. Biochem.*, vol. 64, no. 1, pp. 435–461, Jun. 1995.

[88] D. A. Mills, D. A. Manias, L. L. McKay, and G. M. Dunny, “Homing of a group II intron from Lactococcus lactis subsp. lactis ML3.,” *J. Bacteriol.*, vol. 179, no. 19, pp. 6107–11, Oct. 1997.

[89] A. Lambowitz and M. Caprara, “Group I and group II ribozymes as RNPs: clues to the past and guides to the future.,” *CSHL Press*, The RNA world ed. 2, pp. 451–486, 1999.

[90] N. N. Singh and a M. Lambowitz, “Interaction of a group II intron ribonucleoprotein endonuclease with its DNA target site investigated by DNA footprinting and modification interference.,” *J. Mol. Biol.*, vol. 309, no. 2, pp. 361–86, Jun. 2001.

[91] H. Guo, M. Karberg, M. Long, J. P. Jones, B. Sullenger, and A. M. Lambowitz, “Group II Introns Designed to Insert into Therapeutically Relevant DNA Target Sites in Human Cells.,” *Sci.* , vol. 289 , no. 5478 , pp. 452–457, Jul. 2000.

[92] R. Eskes, J. Yang, A. M. Lambowitz, and P. S. Perlman, “Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing.,” *Cell*, vol. 88, no. 6, pp. 865–74, Mar. 1997.

[93] B. Cousineau, D. Smith, S. Lawrence-Cavanagh, J. E. Mueller, J. Yang, D. Mills, D. Manias, G. Dunny, a M. Lambowitz, and M. Belfort, “Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination.,” *Cell*, vol. 94, no. 4, pp. 451–62, Aug. 1998.

[94] D. Smith, J. Zhong, M. Matsuura, A. M. Lambowitz, and M. Belfort, “Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming,” *Genes Dev.*, vol. 19, pp. 2477–2487, Oct. 2005.

[95] M. Karberg, H. Guo, J. Zhong, R. Coon, J. Perutka, and a M. Lambowitz, “Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria.,” *Nat. Biotechnol.*, vol. 19, no. 12, pp. 1162–7, Dec. 2001.

[96] C. L. Frazier, J. S. Filippo, A. M. Lambowitz, and D. A. Mills, “Genetic Manipulation of Lactococcus lactis by Using Targeted Group II Introns : Generation of Stable Insertions without Selection.,” *Appl. Environ. Microbiol.*, vol. 69, no. 2, pp. 1121–1128, Nov. 2003.

[97] H. Rawsthorne, K. N. Turner, and D. A. Mills, “Multicopy integration of heterologous genes, using the lactococcal group II intron targeted to bacterial insertion sequences.,” *Appl. Environ. Microbiol.*, vol. 72, no. 9, pp. 6088–93, Sep. 2006.

[98] J. Perutka, W. Wang, D. Goerlitz, and A. M. Lambowitz, “Use of Computer-designed Group II Introns to Disrupt Escherichia coli DExH/D-box Protein and DNA Helicase Genes.,” *J. Mol. Biol.*, vol. 336, no. 2, pp. 421–439, Feb. 2004.

[99] M. Mastroianni, K. Watanabe, T. B. White, F. Zhuang, J. Vernon, M. Matsuura, J. Wallingford, and A. M. Lambowitz, “Group II intron-based gene targeting reactions in eukaryotes.,” *PLoS One*, vol. 3, no. 9, p. e3121, Jan. 2008.

[100] A. M. Lambowitz and S. Zimmerly, “Group II introns: mobile ribozymes that invade DNA.,” *Cold Spring Harb. Perspect. Biol.*, vol. 3, no. 8, p. a003616, Aug. 2011.

[101] S. K. Sharan, L. C. Thomason, S. G. Kuznetsov, and D. L. Court, “Recombineering: a homologous recombination-based method of genetic engineering.,” *Nat. Protoc.*, vol. 4, no. 2, pp. 206–223, Feb. 2009.

[102] K. C. Murphy, “Use of Bacteriophage λ Recombination Functions To Promote Gene Replacement in Escherichia coli.,” *J. Bacteriol.*, vol. 180, no. 8, pp. 2063–2071, Apr. 1998.

[103] J. Muyrers, “Rapid modification of bacterial artificial chromosomes by ET- recombination.,” *Nucleic Acids Res.*, vol. 27, no. 6, pp. 1555–1557, Mar. 1999.

[104] H. M. Ellis, D. Yu, T. DiTizio, and D. L. Court, “High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 12, pp. 6742–6, Jun. 2001.

[105] J. A. Mosberg, M. J. Lajoie, and G. M. Church, “Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate.,” *Genetics*, vol. 186, no. 3, pp. 791–9, Nov. 2010.

[106] N. Eckert-Boulet, R. Rothstein, and M. Lisby, “Cell biology of homologous recombination in yeast.,” *Methods Mol. Biol.*, vol. 745, pp. 523–36, Jan. 2011.

[107] D. Yu, H. M. Ellis, E. C. Lee, N. A. Jenkins, N. G. Copeland, and D. L. Court, “An efficient recombination system for chromosome engineering in Escherichia coli.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 11, pp. 5978–83, May 2000.

[108] J. A. Mosberg, C. J. Gregg, M. J. Lajoie, H. H. Wang, and G. M. Church, “Improving lambda red genome engineering in Escherichia coli via rational removal of endogenous nucleases.,” *PLoS One*, vol. 7, no. 9, p. e44638, Jan. 2012.

[109] J. P. Muyrers, Y. Zhang, F. Buchholz, and A. F. Stewart, “RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners.,” *Genes Dev.*, vol. 14, no. 15, pp. 1971–82, Aug. 2000.

[110] M. Nakayama and O. Ohara, “Improvement of recombination efficiency by mutation of Red proteins,” *Biotechniques*, vol. 38, no. 6, pp. 917–924, Jun. 2005.

[111] H. H. Wang, F. J. Isaacs, P. a Carr, Z. Z. Sun, G. Xu, C. R. Forest, and G. M. Church, "Programming cells by multiplex genome engineering and accelerated evolution.," *Nature*, vol. 460, no. 7257, pp. 894–8, Aug. 2009.

[112] H. H. Wang, P.-Y. Huang, G. Xu, W. Haas, A. Marblestone, J. Li, S. P. Gygi, A. C. Forster, M. C. Jewett, and G. M. Church, "Multiplexed in vivo His-tagging of enzyme pathways for in vitro single-pot multienzyme catalysis.," *ACS Synth. Biol.*, vol. 1, no. 2, pp. 43–52, Feb. 2012.

[113] H. H. Wang, H. Kim, L. Cong, J. Jeong, D. Bang, and G. M. Church, "Genome-scale promoter engineering by coselection MAGE.," *Nat. Methods*, vol. 9, no. 6, pp. 591–3, Jun. 2012.

[114] F. J. Isaacs, P. A. Carr, H. H. Wang, M. J. Lajoie, B. Sterling, L. Kraal, A. C. Tolonen, T. A. Gianoulis, D. B. Goodman, N. B. Reppas, C. J. Emig, D. Bang, S. J. Hwang, M. C. Jewett, J. M. Jacobson, and G. M. Church, "Precise Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement.," *Sci.*, vol. 333, no. 6040, pp. 348–353, Jul. 2011.

[115] A. Thierry and B. Dujon, "Nested chromosomal fragmentation in yeast using the meganuclease I-Sce I: a new method for physical mapping of eukaryotic genomes.," *Nucleic Acids Res.*, vol. 20, no. 21, pp. 5625–5631, Nov. 1992.

[116] M. Belfort and R. J. Roberts, "Homing endonucleases: keeping the house in order.," *Nucleic Acids Res.*, vol. 25, no. 17, pp. 3379–88, Sep. 1997.

[117] B. Dujon, M. Belfort, R. A. Butow, C. Jacq, C. Lemieux, P. S. Perlman, and V. M. Vogt, "Mobile introns: definition of terms and recommended nomenclature.," *Gene*, vol. 82, no. 1, pp. 115–8, Oct. 1989.

[118] F. B. Perler, E. O. Davis, G. E. Dean, F. S. Gimble, W. E. Jack, N. Neff, C. J. Noren, J. Thorner, and M. Belfort, "Protein splicing elements: inteins and exteins-a definition of terms and recommended nomenclature.," *Nucleic Acids Res.*, vol. 22, no. 7, pp. 1125–7, Apr. 1994.

[119] D. R. Edgell, "Selfish DNA: Homing Endonucleases Find a Home.," *Current Biology*, vol. 19. pp. 115–117, Feb. 2009.

[120] B. S. Chevalier and B. L. Stoddard, "Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility.," *Nucleic Acids Res.*, vol. 29, no. 18, pp. 3757–3774, Apr. 2001.

[121] A. M. Lambowitz and M. Belfort, "Introns as Mobile Genetic Elements," *Annu. Rev. Biochem.*, vol. 62, no. 1, pp. 587–622, Jun. 1993.

[122] M. Jasin, "Genetic manipulation of genomes with rare-cutting endonucleases.," *Trends Genet.*, vol. 12, no. 6, pp. 224–8, Jun. 1996.

[123] F. S. Gimble and J. Wang, "Substrate recognition and induced DNA distortion by the PI-SceI endonuclease, an enzyme generated by protein splicing.," *J. Mol. Biol.*, vol. 263, no. 2, pp. 163–80, Oct. 1996.

[124] G. M. Argast, K. M. Stephens, M. J. Emond, and R. J. Monnat, "I-PpoI and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential in vitro enrichment.," *J. Mol. Biol.*, vol. 280, no. 3, pp. 345–53, Jul. 1998.

[125] S. Arnould, P. Chames, C. Perez, E. Lacroix, A. Duclert, J. C. Epinat, F. Stricher, A. S. Petit, A. Patin, S. Guillier, S. Rolland, J. Prieto, F. J. Blanco, J. Bravo, G. Montoya, L. Serrano, P. Duchateau, and F. Pâques, "Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets.," *J. Mol. Biol.*, vol. 355, no. 3, pp. 443–58, Jan. 2006.

[126] L. M. Seligman, K. M. Chisholm, B. S. Chevalier, M. S. Chadsey, S. T. Edwards, J. H. Savage, and A. L. Veillet, "Mutations altering the cleavage specificity of a homing endonuclease.," *Nucleic Acids Res.*, vol. 30, no. 17, pp. 3870–9, Sep. 2002.

[127] J. B. Doyon, V. Pattanayak, C. B. Meyer, and D. R. Liu, "Directed Evolution and Substrate Specificity Profile of Homing Endonuclease I-SceI.," *J. Am. Chem. Soc.*, vol. 128, no. 7, pp. 2477–2484, Jan. 2006.

[128] S. Grizot, A. Duclert, S. Thomas, P. Duchateau, and F. Pâques, "Context dependence between subdomains in the DNA binding interface of the I-CreI homing endonuclease.," *Nucleic Acids Res.*, vol. 39, no. 14, pp. 6124–36, Aug. 2011.

[129] R. J. Roberts, "A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes.," *Nucleic Acids Res.*, vol. 31, no. 7, pp. 1805–1812, Apr. 2003.

[130] B. S. Chevalier, T. Kortemme, M. S. Chadsey, D. Baker, R. J. Monnat, and B. L. Stoddard, "Design, activity, and structure of a highly specific artificial endonuclease.," *Mol. Cell*, vol. 10, no. 4, pp. 895–905, Oct. 2002.

[131] F. D. Urnov, E. J. Rebar, M. C. Holmes, H. S. Zhang, and P. D. Gregory, "Genome editing with engineered zinc finger nucleases.," *Nat Rev Genet*, vol. 11, no. 9, pp. 636–646, Sep. 2010.

[132] J. Miller, A. D. McLachlan, and A. Klug, "Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes.," *EMBO J.*, vol. 4, no. 6, pp. 1609–14, Jun. 1985.

[133] N. P. Pavletich and C. O. Pabo, "Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A.," *Sci.*, vol. 252, no. 5007, pp. 809–817, May 1991.

[134] J. H. Laity, B. M. Lee, and P. E. Wright, "Zinc finger proteins: new insights into structural and functional diversity.," *Curr. Opin. Struct. Biol.*, vol. 11, no. 1, pp. 39–46, Feb. 2001.

[135] C. O. Pabo, E. Peisach, and R. A. Grant, "Design and selection of novel Cys2His2 zinc finger proteins.," *Annu. Rev. Biochem.*, vol. 70, no. 1, pp. 313–340, Jun. 2001.

[136] Q. Liu, D. J. Segal, J. B. Ghiara, and C. F. Barbas, "Design of polydactyl zinc-finger proteins for unique addressing within complex genomes.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 94, no. 11, pp. 5525–30, May 1997.

[137] M. Isalan, A. Klug, and Y. Choo, "Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers.," *Biochemistry*, vol. 37, no. 35, pp. 12026–33, Sep. 1998.

[138] D. Carroll, "Genome engineering with zinc-finger nucleases.," *Genetics*, vol. 188, no. 4, pp. 773–82, Aug. 2011.

[139] R. R. Beerli and C. F. Barbas, “Engineering polydactyl zinc-finger transcription factors.,” *Nat. Biotechnol.*, vol. 20, no. 2, pp. 135–41, Feb. 2002.

[140] M. S. Bhakta, I. M. Henry, D. G. Ousterout, K. T. Das, S. H. Lockwood, J. F. Meckler, M. C. Wallen, A. Zykovich, Y. Yu, H. Leo, L. Xu, C. A. Gersbach, and D. J. Segal, “Highly active zinc-finger nucleases by extended modular assembly.,” *Genome Res.*, vol. 23, no. 3, pp. 530–8, Mar. 2013.

[141] B. Gonzalez, L. J. Schwimmer, R. P. Fuller, Y. Ye, L. Asawapornmongkol, and C. F. Barbas, “Modular system for the construction of zinc-finger libraries and proteins.,” *Nat. Protoc.*, vol. 5, no. 4, pp. 791–810, Apr. 2010.

[142] C. L. Ramirez, M. T. Certo, C. Mussolino, M. J. Goodwin, T. J. Cradick, A. P. McCaffrey, T. Cathomen, A. M. Scharenberg, and J. K. Joung, “Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects.,” *Nucleic Acids Res.*, vol. 40, no. 12, pp. 5560–8, Jul. 2012.

[143] J. C. Miller, M. C. Holmes, J. Wang, D. Y. Guschin, Y.-L. Lee, I. Rupniewski, C. M. Beausejour, A. J. Waite, N. S. Wang, K. a Kim, P. D. Gregory, C. O. Pabo, and E. J. Rebar, “An improved zinc-finger nuclease architecture for highly specific genome editing.,” *Nat. Biotechnol.*, vol. 25, no. 7, pp. 778–85, Jul. 2007.

[144] R. M. Gordley, C. A. Gersbach, and C. F. Barbas, “Synthesis of programmable integrases.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 13, pp. 5053–5058, Dec. 2008.

[145] J. Guo, T. Gaj, and C. F. Barbas, “Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases.,” *J. Mol. Biol.*, vol. 400, no. 1, pp. 96–107, Jul. 2010.

[146] N. Sun, Z. Abil, and H. Zhao, “Recent advances in targeted genome engineering in mammalian systems.,” *Biotechnol. J.*, vol. 7, no. 9, pp. 1074–87, Sep. 2012.

[147] W. Yan, C. Smith, and L. Cheng, “Expanded activity of dimer nucleases by combining ZFN and TALEN for genome editing.,” *Sci. Rep.*, vol. 3, p. 2376, Jan. 2013.

[148] C. Mussolino and T. Cathomen, “TALE nucleases: tailored genome engineering made easy.,” *Curr. Opin. Biotechnol.*, vol. 23, no. 5, pp. 644–50, Oct. 2012.

[149] S. Kay, S. Hahn, E. Marois, G. Hause, and U. Bonas, “A bacterial effector acts as a plant transcription factor and induces a cell size regulator.,” *Science*, vol. 318, no. 5850, pp. 648–51, Oct. 2007.

[150] P. Römer, S. Hahn, T. Jordan, T. Strauss, U. Bonas, and T. Lahaye, “Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene.,” *Science*, vol. 318, no. 5850, pp. 645–8, Oct. 2007.

[151] J. Boch and U. Bonas, “Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and Function.,” *Annu. Rev. Phytopathol.*, vol. 48, no. 1, pp. 419–436, Jul. 2010.

[152] J. Boch, “TALEs of genome targeting.,” *Nat. Biotechnol.*, vol. 29, no. 2, pp. 135–6, Mar. 2011.

[153] G. van den Ackerveken, E. Marois, and U. Bonas, “Recognition of the bacterial avirulence protein AvrBs3 occurs inside the host plant cell.,” *Cell*, vol. 87, no. 7, pp. 1307–16, Dec. 1996.

[154] A. J. Bogdanove, S. Schornack, and T. Lahaye, “TAL effectors: finding plant genes for disease and defense.,” *Curr. Opin. Plant Biol.*, vol. 13, no. 4, pp. 394–401, Aug. 2010.

[155] J. Boch, H. Scholze, S. Schornack, A. Landgraf, S. Hahn, S. Kay, T. Lahaye, A. Nickstadt, and U. Bonas, “Breaking the code of DNA binding specificity of TAL-type III effectors.,” *Science*, vol. 326, no. 5959, pp. 1509–1512, Dec. 2009.

[156] M. J. Moscou and A. J. Bogdanove, “A Simple Cipher Governs DNA Recognition by TAL Effectors.,” *Sci.*, vol. 326 , no. 5959 , p. 1501, Dec. 2009.

[157] J. Streubel, C. Blücher, A. Landgraf, and J. Boch, “TAL effector RVD specificities and efficiencies.,” *Nat. Biotechnol.*, vol. 30, no. 7, pp. 593–5, Jul. 2012.

[158] H. Scholze and J. Boch, “TAL effector-DNA specificity.,” *Virulence.*, vol. 1, no. 5, pp. 428–432, Sep. 2010.

[159] M. Zhang, F. Wang, S. Li, Y. Wang, Y. Bai, and X. Xu, “TALE: a tale of genome editing.,” *Prog. Biophys. Mol. Biol.*, vol. 114, no. 1, pp. 25–32, Jan. 2014.

[160] R. Morbitzer, P. Römer, J. Boch, and T. Lahaye, “Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 50, pp. 21617–22, Dec. 2010.

[161] F. Zhang, L. Cong, S. Lodato, S. Kosuri, G. M. Church, and P. Arlotta, “Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription.,” *Nat. Biotechnol.*, vol. 29, no. 2, pp. 149–53, Feb. 2011.

[162] J. P. Tremblay, P. Chapdelaine, Z. Coulombe, and J. Rousseau, “Transcription Activator-Like Effector Proteins Induce the Expression of the Frataxin Gene.,” *Hum. Gene Ther.*, vol. 23, no. 8, pp. 883–890, May 2012.

[163] M. Christian, T. Cermak, E. L. Doyle, C. Schmidt, F. Zhang, A. Hummel, A. J. Bogdanove, and D. F. Voytas, “Targeting DNA double-strand breaks with TAL effector nucleases.,” *Genetics*, vol. 186, no. 2, pp. 757–61, Oct. 2010.

[164] T. Li, S. Huang, X. Zhao, D. A. Wright, S. Carpenter, M. H. Spalding, D. P. Weeks, and B. Yang, “Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes.,” *Nucleic Acids Res.*, vol. 39, no. 14, pp. 6315–25, Aug. 2011.

[165] T. Cermak, E. L. Doyle, M. Christian, L. Wang, Y. Zhang, C. Schmidt, J. a Baller, N. V Somia, A. J. Bogdanove, and D. F. Voytas, “Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting.,” *Nucleic Acids Res.*, vol. 39, no. 12, p. e82, Jul. 2011.

[166] C. Mussolino, R. Morbitzer, F. Lütge, N. Dannemann, T. Lahaye, and T. Cathomen, “A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity.,” *Nucleic Acids Res.*, vol. 39, no. 21, pp. 9283–93, Dec. 2011.

[167] C. Mussolino, J. Alzubi, E. J. Fine, R. Morbitzer, T. J. Cradick, T. Lahaye, G. Bao, and T. Cathomen, “TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity.,” *Nucleic Acids Res.*, vol. 42, no. 10, pp. 6762–73, Jan. 2014.

[168] E. Weber, R. Gruetzner, S. Werner, C. Engler, and S. Marillonnet, “Assembly of designer TAL effectors by Golden Gate cloning.,” *PLoS One*, vol. 6, no. 5, p. e19722, Jan. 2011.

[169] N. E. Sanjana, L. Cong, Y. Zhou, M. M. Cunniff, G. Feng, and F. Zhang, “A transcription activator-like effector toolbox for genome engineering.,” *Nat. Protoc.*, vol. 7, no. 1, pp. 171–192, Jan. 2012.

[170] C. Uhde-Stone, N. Gor, T. Chin, J. Huang, and B. Lu, “A do-it-yourself protocol for simple transcription activator-like effector assembly.,” *Biol. Proced. Online*, vol. 15, no. 1, p. 3, Jan. 2013.

[171] J. L. Schmid-Burgk, T. Schmidt, V. Kaiser, K. Höning, and V. Hornung, “A ligation-independent cloning technique for high-throughput assembly of transcription activator–like effector genes.,” *Nat. Biotechnol.*, vol. 31, no. 1, pp. 76–81, Jan. 2013.

[172] D. Reyon, S. Q. Tsai, C. Khayter, J. A. Foden, J. D. Sander, and J. K. Joung, “FLASH assembly of TALENs for high-throughput genome editing.,” *Nat. Biotechnol.*, vol. 30, no. 5, pp. 460–5, May 2012.

[173] A. J. Bogdanove and D. F. Voytas, “TAL effectors: customizable proteins for DNA targeting.,” *Science*, vol. 333, no. 6051, pp. 1843–6, Sep. 2011.

[174] R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, and P. Horvath, “CRISPR provides acquired resistance against viruses in prokaryotes.,” *Science*, vol. 315, no. 5819, pp. 1709–12, Mar. 2007.

[175] L. A. Marraffini and E. J. Sontheimer, “CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA.,” *Science*, vol. 322, no. 5909, pp. 1843–5, Dec. 2008.

[176] G. Gasiunas, R. Barrangou, P. Horvath, and V. Siksnys, “Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 39, pp. E2579–86, Sep. 2012.

[177] G. Gasiunas and V. Siksnys, “RNA-dependent DNA endonuclease Cas9 of the CRISPR system: Holy Grail of genome editing?,” *Trends Microbiol.*, vol. 21, no. 11, pp. 562–7, Nov. 2013.

[178] F. J. M. Mojica, C. Díez-Villaseñor, E. Soria, and G. Juez, “Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria.,” *Mol. Microbiol.*, vol. 36, no. 1, pp. 244–246, Apr. 2000.

[179] R. Jansen, J. D. A. Van Embden, W. Gaastra, and L. M. Schouls, “Identification of genes that are associated with DNA repeats in prokaryotes.,” *Mol. Microbiol.*, vol. 43, no. 6, pp. 1565–75, Mar. 2002.

[180] I. Grissa, G. Vergnaud, and C. Pourcel, “The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats.,” *BMC Bioinformatics*, vol. 8, p. 172, Jan. 2007.

[181] F. J. M. Mojica, C. Díez-Villaseñor, J. García-Martínez, and E. Soria, "Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements.," *J. Mol. Evol.*, vol. 60, no. 2, pp. 174–82, Feb. 2005.

[182] R. Barrangou and P. Horvath, "The CRISPR System Protects Microbes against Phages, Plasmids.," *Microbe*, vol. 4, no. 5, pp. 224–230, May 2009.

[183] S. J. J. Brouns, M. M. Jore, M. Lundgren, E. R. Westra, R. J. H. Slijkhuis, A. P. L. Snijders, M. J. Dickman, K. S. Makarova, E. V Koonin, and J. van der Oost, "Small CRISPR RNAs guide antiviral defense in prokaryotes.," *Science*, vol. 321, no. 5891, pp. 960–4, Aug. 2008.

[184] E. R. Westra, D. C. Swarts, R. H. J. Staals, M. M. Jore, S. J. J. Brouns, and J. van der Oost, "The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity.," *Annu. Rev. Genet.*, vol. 46, pp. 311–39, Jan. 2012.

[185] K. S. Makarova, D. H. Haft, R. Barrangou, S. J. J. Brouns, E. Charpentier, P. Horvath, S. Moineau, F. J. M. Mojica, Y. I. Wolf, A. F. Yakunin, J. van der Oost, and E. V Koonin, "Evolution and classification of the CRISPR–Cas systems.," *Nat Rev Micro*, vol. 9, no. 6, pp. 467–477, Jun. 2011.

[186] B. Shen, W. Zhang, J. Zhang, J. Zhou, J. Wang, L. Chen, L. Wang, A. Hodgkins, V. Iyer, X. Huang, and W. C. Skarnes, "Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects.," *Nat. Meth.*, vol. 11, no. 4, pp. 399–402, Apr. 2014.

[187] P. D. Hsu, D. a Scott, J. a Weinstein, F. A. Ran, S. Konermann, V. Agarwala, Y. Li, E. J. Fine, X. Wu, O. Shalem, T. J. Cradick, L. A. Marraffini, G. Bao, and F. Zhang, "DNA targeting specificity of RNA-guided Cas9 nucleases.," *Nat. Biotechnol.*, vol. 31, no. 9, pp. 827–32, Sep. 2013.

[188] D. H. Haft, J. Selengut, E. F. Mongodin, and K. E. Nelson, "A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes.," *PLoS Comput. Biol.*, vol. 1, no. 6, p. e60, Nov. 2005.

[189] S. A. Shah, S. Erdmann, F. J. M. Mojica, and R. A. Garrett, "Protospacer recognition motifs: mixed identities and functional diversity.," *RNA Biol.*, vol. 10, no. 5, pp. 891–9, May 2013.

[190] L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, and F. Zhang, "Multiplex genome engineering using CRISPR/Cas systems.," *Science*, vol. 339, no. 6121, pp. 819–23, Feb. 2013.

[191] T. Sakuma, A. Nishikawa, S. Kume, K. Chayama, and T. Yamamoto, "Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system.," *Sci. Rep.*, vol. 4, p. 5400, Jan. 2014.

[192] L. A. Gilbert, M. H. Larson, L. Morsut, Z. Liu, G. a Brar, S. E. Torres, N. Stern-Ginossar, O. Brandman, E. H. Whitehead, J. A. Doudna, W. A. Lim, J. S. Weissman, and L. S. Qi, "CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.," *Cell*, vol. 154, no. 2, pp. 442–51, Jul. 2013.

[193] Y. Fu, J. A. Foden, C. Khayter, M. L. Maeder, D. Reyon, J. K. Joung, and J. D. Sander, "High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells.," *Nat. Biotechnol.*, vol. 31, no. 9, pp. 822–6, Sep. 2013.

[194] T. J. Cradick, E. J. Fine, C. J. Antico, and G. Bao, “CRISPR/Cas9 systems targeting β-globin and CCR5 genes have substantial off-target activity.,” *Nucleic Acids Res.*, vol. 41, no. 20, pp. 9584–92, Nov. 2013.

[195] F. A. Ran, P. D. Hsu, C. Lin, J. S. Gootenberg, A. Trevino, D. A. Scott, A. Inoue, and S. Matoba, “Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity.,”*Cell*, vol. 154, no. 6, pp. 1380–9, Sep. 2013.

[196] Y. Fu, J. D. Sander, D. Reyon, V. M. Cascio, and J. K. Joung, “Improving CRISPR-Cas nuclease specificity using truncated guide RNAs.,” *Nat. Biotechnol.*, vol. 32, no.3, pp. 279–84, Mar. 2014.

[197] S. Q. Tsai, N. Wyvekens, C. Khayter, J. A. Foden, V. Thapar, D. Reyon, M. J. Goodwin, M. J. Aryee, and J. K. Joung, “Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing.,” *Nat. Biotechnol.*, vol. 32, no. 6, Apr. 2014.

[198] B. K. Choi, P. Bobrowicz, R. C. Davidson, S. R. Hamilton, D. H. Kung, H. Li, R. G. Miele, J. H. Nett, S. Wildt, and T. U. Gerngross, “Use of combinatorial genetic libraries to humanize N-linked glycosylation in the yeast Pichia pastoris.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, pp. 5022–5027, Apr. 2003.

[199] L. Näätsaari, B. Mistlberger, C. Ruth, T. Hajek, F. S. Hartner, and A. Glieder, “Deletion of the Pichia pastoris KU70 homologue facilitates platform strain generation for gene expression and synthetic biology.,” *PLoS One*, vol. 7, no. 6, p. e39720, Jan. 2012.

[200] P. Li, A. Anumanthan, X. G. Gao, K. Ilangovan, V. V Suzara, N. Düzgüneş, and V. Renugopalakrishnan, “Expression of recombinant proteins in Pichia pastoris.,” *Appl. Biochem. Biotechnol.*, vol. 142, pp. 105–124, May 2007.

[201] D. R. Higgins and J. M. Cregg, “Pichia Protocols.,” *Humana Press*, 1998.

[202] J. E. DiCarlo, J. E. Norville, P. Mali, X. Rios, J. Aach, and G. M. Church, “Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems.,” *Nucleic Acids Res.*, vol. 41, no. 7, pp. 4336–43, Apr. 2013.

[203] B. Fahrenkrog and U. Aebi, “The nuclear pore complex: nucleocytoplasmic transport and beyond.,” *Nat Rev Mol Cell Biol*, vol. 4, no. 10, pp. 757–766, Oct. 2003.

[204] R. Y. H. Lim, U. Aebi, and B. Fahrenkrog, “Towards reconciling structure and function in the nuclear pore complex.,” *Histochem. Cell Biol.*, vol. 129, no. 2, pp. 105–16, Feb. 2008.

[205] P. L. Paine, L. C. Moore, and S. B. Horowitz, “Nuclear envelope permeability.,” *Nature*, vol. 254, pp. 109–114, Mar. 1975.

[206] A. Lange, R. E. Mills, C. J. Lange, M. Stewart, S. E. Devine, and A. H. Corbett, “Classical nuclear localization signals: definition, function, and interaction with importin alpha.,” *J. Biol. Chem.*, vol. 282, no. 8, pp. 5101–5, Feb. 2007.

[207] U. Kutay, F. R. Bischoff, S. Kostka, R. Kraft, and D. Görlich, “Export of importin alpha from the nucleus is mediated by a specific nuclear transport factor.,” *Cell*, vol. 90, no. 6, pp. 1061–71, Sep. 1997.

[208] E. Izaurralde, U. Kutay, C. von Kobbe, I. W. Mattaj, and D. Görlich, “The asymmetric distribution of the constituents of the Ran system is essential for transport into and out of the nucleus.,” *EMBO J.*, vol. 16, no. 21, pp. 6535–47, Nov. 1997.

[209] K. Stade, C. S. Ford, C. Guthrie, and K. Weis, “Exportin 1 (Crm1p) is an essential nuclear export factor.,” *Cell*, vol. 90, no. 6, pp. 1041–50, Sep. 1997.

[210] C. Dingwall and R. A. Laskey, “Nuclear targeting sequences-a consensus?,” *Trends Biochem. Sci.*, vol. 16, no. 12, pp. 478–481, Dec. 1991.

[211] J. Robbins, S. M. Dilwortht, R. A. Laskey, and C. Dingwall, “Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: Identification of a class of bipartite nuclear targeting sequence.,” *Cell*, vol. 64, no. 3, pp. 615–623, Jun. 2014.

[212] R. E. Lanford and J. S. Butel, “Construction and characterization of an SV40 mutant defective in nuclear transport of T antigen.,” *Cell*, vol. 37, no. 3, pp. 801–813, Jul. 1984.

[213] M. R. Hodel, A. H. Corbett, and A. E. Hodel, “Dissection of a nuclear localization signal.,” *J. Biol. Chem.*, vol. 276, no. 2, pp. 1317–25, Jan. 2001.

[214] E. Conti and J. Kuriyan, “Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha.,” *Structure*, vol. 8, no. 3, pp. 329–38, Mar. 2000.

[215] C. Dingwall, J. Robbins, S. M. Dilworth, B. Roberts, and W. D. Richardson, “The nucleoplasmin nuclear location sequence is larger and more complex than that of SV-40 large T antigen.,” *J. Cell Biol.*, vol. 107, no. 3, pp. 841–9, Sep. 1988.

[216] B. Hobl, B. Hock, S. Schneck, R. Fischer, and M. Mack, “Bacteriophage T7 RNA polymerase-based expression in Pichia pastoris.,” *Protein Expr. Purif.*, vol. 92, no. 1, pp. 100–4, Nov. 2013.

[217] G. Yang, H. Zhou, Y. Lu, Y. Lin, and S. Zhou, “Comparing expression of different forms of human DNA topoisomerase I in Pichia pastoris.,” *Enzyme Microb. Technol.*, vol. 34, no. 2, pp. 139–146, Feb. 2004.

[218] A. E. Gradoboeva and M. V Padkina, “Bovine gamma-interferon nuclear localization sequence provides translocation of recombinant protein to yeast Pichia pastoris cell nucleus,” *Cell and Tissue Biol.*, vol. 4, no. 6, pp. 566–71; Dec. 2010.

[219] M. Nelson and P. Silver, “Context affects nuclear protein localization in Saccharomyces cerevisiae.,” *Mol. Cell. Biol.*, vol. 9, no. 2, pp. 384–389, Feb. 1989.

[220] U. Fleig, S. S. Salus, I. Karig, and S. Sazer, “The Fission Yeast Ran Gtpase Is Required for Microtubule Integrity.,” *J. Cell Biol.*, vol. 151, no. 5, pp. 1101–1112, Nov. 2000.

[221] C. V Dang and W. M. Lee, “Identification of the human c-myc protein nuclear translocation signal.,” *Mol. Cell. Biol.*, vol. 8, no. 10, pp. 4048–4054, Oct. 1988.

[222] J. P. S. Makkerh, C. Dingwall, and R. A. Laskey, “Comparative mutagenesis of nuclear localization signals reveals the importance of neutral and acidic amino acids.,” *Curr. Biol.*, vol. 6, no. 8, pp. 1025–1027, Aug. 1996.

[223] A. K. Azad, T. Tani, N. Shiki, S. Tsuneyoshi, S. Urushiyama, and Y. Ohshima, “Isolation and molecular characterization of mRNA transport mutants in Schizosaccharomyces pombe.,” *Mol. Biol. Cell* , vol. 8, no. 5, pp. 825–841, May 1997.

[224] M. N. Hall, L. Hereford, and I. Herskowitz, “Targeting of E. coli β-galactosidase to the nucleus in yeast.,” *Cell*, vol. 36, no. 4, pp. 1057–1065, Jan. 2015.

[225] D. A. Jans, “The regulation of protein transport to the nucleus by phosphorylation.,” *Biochem. J.*, vol. 311, pp. 705–716, Nov. 1995.

[226] K. De Schutter, Y. C. Lin, P. Tiels, A. Van Hecke, S. Glinka, J. Weber-Lehmann, P. Rouzé, Y. Van de Peer, and N. Callewaert, “Genome sequence of the recombinant protein production host Pichia pastoris.,” *Nat. Biotechnol.*, vol. 27, no. 6, pp. 561–6, Jun. 2009.

[227] J. M. Cherry, E. L. Hong, C. Amundsen, R. Balakrishnan, G. Binkley, E. T. Chan, K. R. Christie, M. C. Costanzo, S. S. Dwight, S. R. Engel, D. G. Fisk, J. E. Hirschman, B. C. Hitz, K. Karra, C. J. Krieger, S. R. Miyasato, R. S. Nash, J. Park, M. S. Skrzypek, M. Simison, S. Weng, and E. D. Wong, “Saccharomyces Genome Database: the genomics resource of budding yeast.,” *Nucleic Acids Res.*, vol. 40, no. D1, pp. D700–5, Jan. 2012.

[228] M. C. Yuh and G. Blobel, “Karyopherins and nuclear import.,” *Current Opinion in Structural Biology*, vol. 11. pp. 703–715, Dec. 2001.

[229] C. S. Newlon and J. F. Theis, “The structure and function of yeast ARS elements.,” *Curr. Opin. Genet. Dev.*, vol. 3, no. 5, pp. 752–8, Oct. 1993.

[230] J. M. Cregg, K. J. Barringer, A. Y. Hessler, and K. R. Madden, “Pichia pastoris as a host system for transformations.,” *Mol. Cell. Biol.*, vol. 5, no. 12, pp. 3376–3385, Dec. 1985.

[231] I. Liachko, R. a Youngblood, K. Tsui, K. L. Bubb, C. Queitsch, M. K. Raghuraman, C. Nislow, B. J. Brewer, and M. J. Dunham, “GC-Rich DNA Elements Enable Replication Origin Activity in the Methylotrophic Yeast Pichia pastoris.,” *PLoS Genet.*, vol. 10, no. 3, p. e1004169, Mar. 2014.

[232] I. Liachko and M. J. Dunham, “An autonomously replicating sequence for use in a wide range of budding yeasts.,” *FEMS Yeast Res.*, vol. 14, no. 2, pp. 364–7, Mar. 2014.

[233] M. Méchali, “Eukaryotic DNA replication origins: many choices for appropriate answers.,” *Nat. Rev. Mol. Cell. Biol.*, vol. 11, no. 10, pp. 728–738, Oct. 2010.

[234] O. M. Aparicio, D. M. Weinstein, and S. P. Bell, “Components and dynamics of DNA replication complexes in S. cerevisiae: redistribution of MCM proteins and Cdc45p during S phase.,” *Cell*, vol. 91, no. 1, pp. 59–69, Oct. 1997.

[235] D. T. Stinchcomb, K. Struhl and R.W. Davis, “Isolation and characterisation of a yeast chromosomal replicator.,” *Nature*, vol. 282, no. 5734, pp. 39–43, Nov. 1979.

[236] I. Liachko, A. Bhaskar, C. Lee, S. C. C. Chung, B.-K. Tye, and U. Keich, “A comprehensive genome-wide map of autonomously replicating sequences in a naive genome.,” *PLoS Genet.*, vol. 6, no. 5, p. e1000946, May 2010.

[237] U. Keich, H. Gao, J. S. Garretson, A. Bhaskar, I. Liachko, J. Donato, and B. K. Tye, “Computational detection of significant variation in binding affinity across two sets of sequences with application to the analysis of replication origins in yeast.,” *BMC Bioinformatics*, vol. 9, p. 372, Jan. 2008.

[238] S. E. Celniker, K. Sweder, F. Srienc, J. E. Bailey, and J. L. Campbell, “Deletion mutations affecting autonomously replicating sequence ARS1 of Saccharomyces cerevisiae.,” *Mol. Cell. Biol.*, vol. 4, no. 11, pp. 2455–2466, Nov. 1984.

[239] Y. Marahrens and B. Stillman, “A yeast chromosomal origin of DNA replication defined by multiple functional elements.,” *Sci.*, vol. 255, no. 5046, pp. 817–23, Feb. 1992.

[240] K. Struhl, D. T. Stinchcomb, S. Scherer, and R. W. Davis, “High-frequency transformation of yeast: autonomous replication of hybrid DNA molecules.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 76, no. 3, pp. 1035–9, Mar. 1979.

[241] C. S. Chan and B. K. Tye, “Autonomously replicating sequences in Saccharomyces cerevisiae.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 77, no. 11, pp. 6329–6333, Nov. 1980.

[242] K. M. Esvelt, P. Mali, J. L. Braff, M. Moosburner, S. J. Yaung, and G. M. Church, “Orthogonal Cas9 proteins for RNA-guided gene regulation and editing.,” *Nat. Meth.*, vol. 10, no. 11, pp. 1116–1121, Nov. 2013.

[243] P. Mali, L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville, and G. M. Church, “RNA-Guided Human Genome Engineering via Cas9.,” *Sci.*, vol. 339, no. 6121, pp. 823–826, Feb. 2013.

[244] B. Shen, J. Zhang, H. Wu, J. Wang, K. Ma, Z. Li, X. Zhang, P. Zhang, and X. Huang, “Generation of gene-modified mice via Cas9/RNA-mediated gene targeting.,” *Cell Res.*, vol. 23, no. 5, pp. 720–3, May 2013.

[245] D. Liu, Z. Wang, A. Xiao, Y. Zhang, W. Li, Y. Zu, S. Yao, S. Lin, and B. Zhang, “Efficient gene targeting in zebrafish mediated by a zebrafish-codon-optimized cas9 and evaluation of off-targeting effect.,” *J. Genet. Genomics*, vol. 41, no. 1, pp. 43–6, Jan. 2014.

[246] C. Gustafsson, S. Govindarajan, and J. Minshull, “Codon bias and heterologous protein expression.,” *Trends Biotechnol.*, vol. 22, no. 7, pp. 346–53, Jul. 2004.

[247] A. Mellitzer, C. Ruth, C. Gustafsson, M. Welch, R. Birner-grünberger, R. Weis, T. Purkarthofer, and A. Glieder, “Synergistic modular promoter and gene optimization to push cellulase secretion by Pichia pastoris beyond existing benchmarks.,” *J. Biotechnol.*, vol. 191, pp. 187–195, Dec. 2014.

[248] K. Freeman, M. Gwadz, and D. Shore, “Molecular and genetic analysis of the toxic effect of RAP1 overexpression in yeast.,” *Genetics*, vol. 141, no. 4, pp. 1253–1262, Dec. 1995.

[249] G. Gill and M. Ptashne, “Negative effect of the transcriptional activator GAL4.,” *Nature*, vol. 334, no. 6184, pp. 721–724, Aug. 1988.

[250] C. J. Wilusz, M. Wormington, and S. W. Peltz, “The cap-to-tail guide to mRNA turnover.,” *Nat. Rev. Mol. Cell Biol.*, vol. 2, no. 4, pp. 237–246, Apr. 2001.

[251] J. D. Lewis and E. Izaurralde, “The role of the cap structure in RNA processing and nuclear export.,” *Eur. J. Biochem.*, vol. 247, pp. 461–469, Apr.1997.

[252] G. Dieci, G. Fiorino, M. Castelnuovo, M. Teichmann, and A. Pagano, “The expanding RNA polymerase III transcriptome.,” *Trends Genet.*, vol. 23, no. 12, pp. 614–22, Dec. 2007.

[253] P. Komarnitsky, E. J. Cho, and S. Buratowski, “Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription.,” *Genes Dev.*, vol. 14, no. 617, pp. 2452–2460, Oct. 2000.

[254] D. D. Licatalosi, G. Geiger, M. Minet, S. Schroeder, K. Cilli, J. B. McNeil, and D. L. Bentley, “Functional interaction of yeast pre-mRNA 3’ end processing factors with RNA polymerase II.,” *Mol. Cell*, vol. 9, pp. 1101–1111, May 2002.

[255] K. M. Hannan, R. D. Hannan and L. I. Rothblum, “Transcription by RNA polymerase I.,” *Front. Biosci.*, vol. 3, pp. 376–98, Mar. 1998.

[256] S. Nielsen, Y. Yuzenkova, and N. Zenkin, “Mechanism of Eukaryotic RNA Polymerase III Transcription Termination.,” *Sci.*, vol. 340, no. 6140, pp. 1577–1580, Jun. 2013.

[257] Q. Wang and L. Wang, “New methods enabling efficient incorporation of unnatural amino acids in yeast.,” *J. Am. Chem. Soc.*, vol. 130, no. 19, pp. 6066–7, May 2008.

[258] O. W. Ryan, J. M. Skerker, M. J. Maurer, X. Li, J. C. Tsai, S. Poddar, M. E. Lee, W. Deloache, J. E. Dueber, and P. Adam, “Selection of chromosomal DNA libraries using a multiplex CRISPR system in Saccharomyces cerevisiae.,” *Elife*, p. 03703, Aug. 2014.

[259] F. Farzadfard, S. D. Perli, and T. K. Lu, “Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas.,” *ACS Synth. Biol.*, vol. 2, no. 10, pp. 604–13, Oct. 2013.

[260] J. Y. Lee, C. F. Evans, and D. R. Engelke, “Expression of RNase P RNA in Saccharomyces cerevisiae is controlled by an unusual RNA polymerase III promoter.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 88, no. 16, pp. 6986–90, Aug. 1991.

[261] P. D. Good and D. R. Engelke, “Yeast expression vectors using RNA polymerase III promoters.,” *Gene*, vol. 151, no. 1–2, pp. 209–14, Dec. 1994.

[262] C. Marck, R. Kachouri-Lafond, I. Lafontaine, E. Westhof, B. Dujon, and H. Grosjean, “The RNA polymerase III-dependent family of genes in hemiascomycetes: comparative RNomics, decoding strategies, transcription and evolutionary implications.,” *Nucleic Acids Res.*, vol. 34, no. 6, pp. 1816–35, Jan. 2006.

[263] Y. Gao and Y. Zhao, “Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing.,” *J. Integr. Plant Biol.*, vol. 56, no. 4, pp. 343–9, Apr. 2014.

[264] M. J. Fedor and E. Westhof, “Ribozymes: The first 20 years.,” *Molecular Cell*, vol. 10, no. 4, pp. 703–704, Nov. 2002.

[265] M. J. Fedor and J. R. Williamson, “The catalytic diversity of RNAs.,” *Nat Rev Mol Cell Biol*, vol. 6, no. 5, pp. 399–412, May 2005.

[266] H. W. Pley, K. M. Flaherty, and D. B. McKay, “Three-dimensional structure of a hammerhead ribozyme.,” *Nature*, vol. 372, no. 6501, pp. 68–74, Nov. 1994.

[267] A. R. Ferre-D’Amare, K. Zhou, and J. A. Doudna, “Crystal structure of a hepatitis delta virus ribozyme.,” *Nature*, vol. 395, no. 6702, pp. 567–574, Oct. 1998.

[268] J. A. Doudna and T. R. Cech, “The chemical repertoire of natural ribozymes.,” *Nature*, vol. 418, no. 6894, pp. 222–228, Jul. 2002.

[269] L. Nissim, S. D. Perli, A. Fridkin, P. Perez-Pinera, and T. K. Lu, “Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells.,” *Mol. Cell*, vol. 54, no. 4, pp. 698–710, May 2014.

[270] P. D. Hsu, E. S. Lander, and F. Zhang, “Development and Applications of CRISPR-Cas9 for Genome Engineering.,” *Cell*, vol. 157, no. 6, pp. 1262–1278, Jun. 2014.

[271] J. H. Houtgraaf, J. Versmissen, and W. J. van der Giessen, “A concise review of DNA damage checkpoints and repair in mammalian cells.,” *Cardiovasc. Revascularization Med.*, vol. 7, pp. 165–172, Jul. 2006.

[272] A. Stotz and P. Linder, “The ADE2 gene from Saccharomyces cerevisiae: sequence and new vectors.,” *Gene*, vol. 95, pp. 91–98, Oct. 1990.

[273] P. J. Flynn and R. J. Reece, “Activation of transcription by metabolic intermediates of the pyrimidine biosynthetic pathway.,” *Mol. Cell. Biol.*, vol. 19, no. 1, pp. 882–888, Jan. 1999.

[274] J. Boeke, F. La Croute, and G. Fink, “A positive selection for mutants lacking orotidine-5′-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance.,” *Mol. Gen. Genet.*, vol. 197, no. 2, pp. 345–346, Nov. 1984.

[275] H. Sychrova and M. R. Chevallier, “Cloning and sequencing of the Saccharomyces cerevisiae gene LYP1 coding for a lysine-specific permease.,” *Yeast*, vol. 9, pp. 771–782, Jan. 1993.

[276] M. Ahmad, M. Hirz, H. Pichler, and H. Schwab, “Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production.,” *Appl. Microbiol. Biotechnol.*, vol. 98, no. 12, pp. 5301–5317, Jun. 2014.

[277] J. M. Cregg, K. R. Madden, K. J. Barringer, G. P. Thill, and C. A. Stillman, “Functional Characterization of the Two Alcohol Oxidase Genes from the Yeast Pichia pastoris.,” *Mol. Cell. Biol.*, vol. 9, no. 3, pp. 1316–23, Mar. 1989.

[278] F. W. Krainer, C. Dietzsch, T. Hajek, C. Herwig, O. Spadiut, and A. Glieder, “Recombinant protein expression in Pichia pastoris strains with an engineered methanol utilization pathway.,” *Microb. Cell Fact.*, vol. 11, p. 22, Jan. 2012.

[279] M. Du, M. B. Battles, and J. H. Nett, “A color-based stable multi-copy integrant selection system for Pichia pastoris using the attenuated ADE1 and ADE2 genes as auxotrophic markers.,” *Bioeng. Bugs*, vol. 3, no. 1, pp. 32–7, Jan. 2012.

[280] S. Harju, H. Fedosyuk, and K. R. Peterson, “Rapid isolation of yeast genomic DNA: Bust n’ Grab.,” *BMC Biotechnol.*, vol. 4, p. 8, Apr. 2004.

[281] D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, and H. O. Smith, “Enzymatic assembly of DNA molecules up to several hundred kilobases.,” *Nat. Meth.*, vol. 6, no. 5, pp. 343–345, May 2009.

[282] C. E. Seidman, K. Struhl, J. Sheen, and T. Jessen, “Introduction of Plasmid DNA into Cells.,” *Curr. Protoc. Mol. Biol.*, chapter 1, unit 1.8, May 2001.

[283] J. Lin-Cereghino, W. W. Wong, S. Xiong, W. Giang, L. T. Luong, J. Vu, S. D. Johnson, and G. P. Lin-Cereghino, “Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris.,” *Biotechniques*, vol. 38, no. 1, pp. 44–48, Jan. 2005.

[284] R. Weis, R. Luiten, W. Skranc, H. Schwab, M. Wubbolts, and A. Glieder, “Reliable high-throughput screening with Pichia pastoris by limiting yeast cell death phenomena.,” *FEMS Yeast Res.*, vol. 5, no. 2, pp. 179–89, Nov. 2004.

[285] J. Bai, D. J. Swartz, I. I. Protasevich, C. G. Brouillette, P. M. Harrell, E. Hildebrandt, B. Gasser, D. Mattanovich, A. Ward, G. Chang, and I. L. Urbatsch, “A gene optimization strategy that enhances production of fully functional P-glycoprotein in Pichia pastoris.,” *PLoS One*, vol. 6, no. 8, p. e22577, Jan. 2011.

[286] T. Vogl and A. Glieder, “Regulation of Pichia pastoris promoters and its consequences for protein production.,” *N. Biotechnol.*, vol. 30, no. 4, pp. 385–404, May 2013.

[287] D. Chelsky, R. Ralph, and G. Jonak, “Sequence requirements for synthetic peptide-mediated translocation to the nucleus.,” *Mol. Cell. Biol.*, vol. 9, no. 6, pp. 2487–2492, Jun. 1989.

[288] J. A. DeGrasse, K. N. DuBois, D. Devos, T. N. Siegel, A. Sali, M. C. Field, M. P. Rout, and B. T. Chait, “Evidence for a shared nuclear pore complex architecture that is conserved from the last common eukaryotic ancestor.,” *Mol. Cell. Proteomics*, vol. 8, no. 9, pp. 2119–30, Sep. 2009.

[289] Q. Vicens and E. Westhof, “Crystal structure of geneticin bound to a bacterial 16 S ribosomal RNA A site oligonucleotide.,” *J. Mol. Biol.*, vol. 326, no. 4, pp. 1175–1188, Feb. 2003.

[290] G. M. Ehrenfeld, J. B. Shipley, D. C. Heimbrook, H. Sugiyama, E. C. Long, J. H. Van Boom, G. A. Van der Marel, N. J. Oppenheimer, and S. M. Hecht, “Copper-dependent cleavage of DNA by bleomycin.,” *Biochemistry*, vol. 26, no. 3, pp. 931–942, Feb. 1987.

[291] P. R. Eriksson, D. Ganguli, V. Nagarajavel, and D. J. Clark, “Regulation of histone gene expression in budding yeast.,” *Genetics*, vol. 191, no. 1, pp. 7–20, May 2012.

[292] H. R. Waterham, M. E. Digan, P. J. Koutz, S. V Lair, and J. M. Cregg, “Isolation of the Pichia pastoris glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter.,” *Gene*, vol. 186, no. 1, pp. 37–44, Feb. 1997.

[293] E. Brunet, D. Simsek, M. Tomishima, R. DeKelver, V. M. Choi, P. Gregory, F. Urnov, D. M. Weinstock, and M. Jasin, “Chromosomal translocations induced at specified loci in human stem cells.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 26, pp. 10620–5, Jul. 2009.

[294] A. N. Leão-Helder, A. M. Krikken, I. J. Van der Klei, J. a K. W. Kiel, and M. Veenhuis, “Transcriptional Down-regulation of Peroxisome Numbers Affects Selective Peroxisome Degradation in Hansenula polymorpha,” *J. Biol. Chem.*, vol. 278, no. 42, pp. 40749–40756, 2003.

[295] N. Casali, “Escherichia coli host strains.,” *Methods Mol. Biol.*, vol. 235, pp. 27–48, 2003.

[296] J. K. Cheng and H. S. Alper, “The genome editing toolbox: a spectrum of approaches for targeted modification.,” *Curr. Opin. Biotechnol.*, vol. 30, pp. 87–94, Dec. 2014.

7. ABBREVIATIONS

ACS	ARS consensus sequence
ADE1	N-succinyl-5-aminoimidazole-4-carboxamide ribotide synthetase
ADE2	phosphoribosylaminoimidazole carboxylase
Ago2	argonaute RISC catalytic component 2
AOX1	alcohol oxidase 1
ARS	autonomously replicating sequences
ATP	adenosintriphosphat
Beta	ssDNA binding protein (Recombineering)
BMD1	buffered minimal glucose medium
BMG1	buffered minimal glycerol medium
bp	base pair
CAGE	Conjugative Assembly Genome Engineering
Cas	CRISPR associated (Cas) genes
Cas9	CRISPR associated protein 9
CBS	Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands
cDNA	complementary DNA
CDS	coding sequence
Cre	Causes recombination, site specific recombinase
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRi	CRISPR-Cas9 interference
CTD	C-terminal repeat domain
ds	double stranded
dsb	double stand break
eGFP	Enhanced green fluorescence protein
Exo	5’ -> 3’ exonuclease (Recombineering)
FLP	flippase, site specific recombinase
FRT	flippase recombination site
Gam	binds to the RecB subunit of RecBCD (Recombineering)
GAP	Glyceraldehydes-3-phosphate dehydrogenase

Gen	Geneticn
GOI	gene of interest
gRNA	guide RNA,
GUT1	Glycerol kinase
HDV	hepatitis delta virus ribozyme
HH	Hammerhead ribozyme
HR	homologous recombination
IEP	intron encoded protein
Indels	insertions and deletions
Kan	Kanamycin
loxP	Cre recombination site
MAGE	Multiplex automated genome engineering
miRNA	microRNA
NHEJ	non-homologous end joining
NLS	nuclear localization signal
OD	optical densitiy
ORC	Origin Recognition Complex
PAM	protospacer adjacent motif
PCR	*Polymerase Chain Reaction*
polyA	long chain of adenine nucleotides
polyT	long chain of thymin nucleotides
pUC ORI	*E. coli* origin of replication
RecBCD	inhibits the endogenous RecBCD nuclease
RFN	RNA-guided FokI nucleases
RISC	RNA-induced silencing complex
RMCE	recombinase-mediated cassette exchange
RNAi	RNA interference
RVDs	repeat variable diresidues
shRNA	short haipin RNA
siRNA	small interfering RNA

ss	single stranded
stgRNA	structural component of gRNA
SV40	Simian Virus 40
TALE	transcription activator-like *effector*
TALEN	transcription activator-like *effector nuclease*
TAL-TF	TAL–transcription factors
TIR	terminal inverted repeats
UTR	untranslated region
YPD	yeast extract peptone dextrose medium
Zeo	Zeocin
ZFN	zinc finger nucleases

8. LIST OF FIGURES

Figure 1: Genome engineering technologies……7
Figure 2: Essential parts of the CRISPR-Cas9 system……..….16
Figure 3: gRNA flanked by the 5' cleaving HH and the 3' cleaving HDV ribozyme……...28
Figure 4: Design of CRISPR-Cas9 vectors bearing a bidirectional RNA Polymerase II promoter and ribozymes…….43
Figure 5: Plasmid map of NLS vectors……..53
Figure 6: eGFP fluorescence obtained with different cultivation protocols……...54
Figure 7: Characterization of heterologous and endogenous NLSs in *P. pastoris*…….....……56
Figure 8: The nuclear import of *Pp*Cas9……...…….……57
Figure 9: Plasmid map of ARS-vectors……..58
Figure 10: sTomato fluorescence levels of ARS-strains……59
Figure 11: Streak outs of strains expressing the sTomato gene on ARS-vectors……...61
Figure 12: General design of cassettes used to knock out various auxotrophy genes in *P. pastoris* CBS 7435 WT and CBS 7435 *ku70*……...62
Figure 13: Plasmid map of a CRISPR-Cas9 vector with RNA polymerase III promoter for the expression of gRNAs……..65
Figure 14: *P. pastoris* CBS 7435 transformants bearing CRISPR-Cas9 plasmids with RNA polymerase III promoters……67
Figure 15: Design of CRISPR-Cas9 constructs using the bidirectional P_{HTX1} promoter and ribozymes for the expression of gRNAs…….69
Figure 16: Targeting efficiencies using CRISPR-Cas9 constructs with various Cas9 sequences and ribozyme flanked gRNAs……70
Figure 17: CRISPR-Cas9 induced NHEJ mutations in *P. pastoris* CBS 7435 WT and *ku70*…...72
Figure 18: Overview of different *GUT1* knockout cassettes used for the co-transformantion with CRISPR-Cas9 vectors……73
Figure 19: Principle of CRISPR-Cas9 mediated integration of donor cassettes to allow the complete deletion of an ORF without a resistance marker…….74
Figure 20: CRISPR-Cas9 mediated integration of homologous donor cassettes does not strongly affect the wildtype strain, but improves specific integration in the *ku70* strain…….76
Figure 21: Plasmid maintenance of auxotrophic CRISPR-Cas9 transformants…….77
Figure 22: CRISPR-Cas9 targeting efficiencies of various target genes…..……...……...78
Figure 23: CRISPR-Cas9 multiplexing in *P. pastoris*……79
Figure 24: Design of vectors for the expression of TALENs, which target the *GUT1* locus in *P. pastoris*……...……80

9. LIST OF TABLES

Table 1: Detailed comparison of genome engineering methods……17

Table 2: Heterologous NLSs used in this thesis…….22

Table 3: P. pastoris endogenous NLSs……...23

Table 4: gRNAs to target the *GUT1* locus in *P. pastoris* CBS 7435……..……..38

Table 5: RNA Polymerase III promoters used in this master thesis…….……..41

Table 6: Mammalian TALEN expression vectors containing a binding domain to target *GUT1* in *P. pastoris*……..…45

Table 7: *E. coli* strains generated during the master thesis……46

Table 8: *P. pastoris* strains generated during this master thesis…….……51

Table 9: HR frequencies in *P. pastoris* CBS 7435 WT and CBS 7435 *ku70* obtained after the transformation with various homologous cassettes……64

Table 10: Sequencing results of CRISPR-Cas9 tranformants expressing *Hs*Cas9 and either gRNA2, gRNA3 or gRNA4 to target *GUT1* ……71

10. APPENDIX

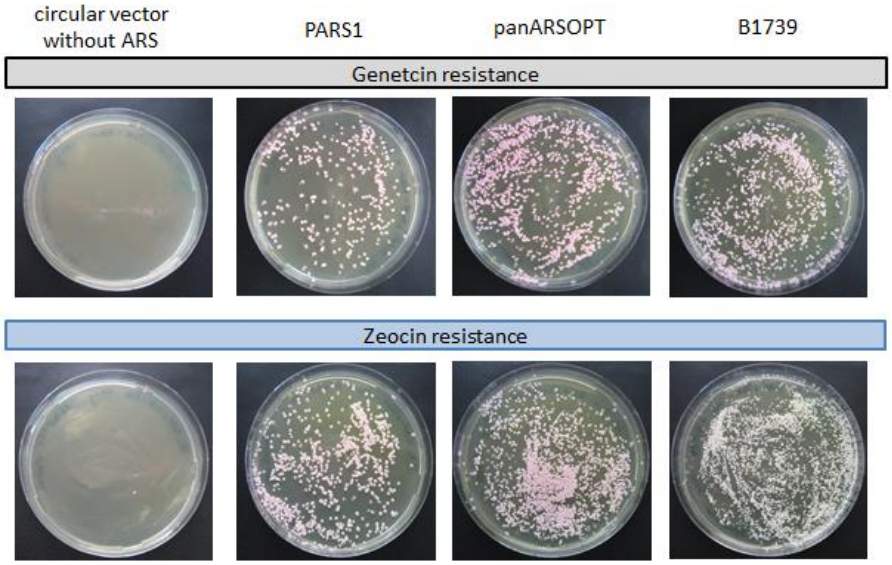
GUT1 DNA sequence

The gRNA sequences are highlighted

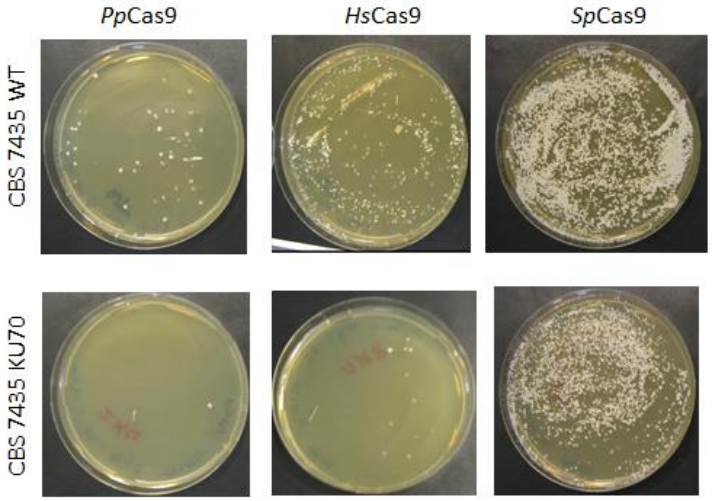
atgggaaaagactatacaccactagttgctaccatcgatattggtactacctccaccagagctattctttttgactaccacggtcaggaagtggc
caagcaccagatcgagtactctacctctgctcaggatgatatcaaaagaaagcgttctcagatcatctcttccgaaggtatttccctgacagttt
ctgacgacttggaagttgagtccgttgacaataaggctggtccaactttgcaatttcctcagccaggctgggttgaatgtcgtccaagtcacat
cttggccaacgccgttcagtgtcttgctgcttgtttagtcaccatggagaacaagaacttggaccgagatgaaaaaaataaatacaagctcata
tctatcggtgttgccaacatgagagagaccacggttgtttggtccaagaagacaggaaagcctctttacaacggtattgtgtggaacgatacc
agaaacaacgatattgttgacgagtacaccgccaagtactctgagaaggagagagaggaaatgcgaaccttgtgtggttgtccaatctccac
ctacttttctgcaaccaagttcagatggttactgaagcatgttcctgaggtcaaacaggcctatgacaatgctgatggggatctaatgtttggtac
tattgactcttggttgatttaccacttgactaacgaaaaatcccacgtcactgatgttaccaatgcctccagaaccaacttcatgaacattgaaac
caacaaatatgacgacagacttttgaaattctgggacgtcgatacttccaaagtcatccttccagaaatcagatcttccgcagaagtctacgga
cacttcaaggtcccacacttggagtctattggatatgttgagtcttaccttactgatgacgcccttgctcttctggaaactattgaaggtgctccttt
ggctggatgtctgggtgaccaatctgcctctttggttggacagttggctgtcagaaagggtgatgccaaatgtacatatggtaccggtgctttct
tgctgtacaacactggtgatcagactttgatttctgagcacggtgctttgaccactgtgggatattggttcccaggtttggatgagtccgaagat
ggcaaacactcttctaagccacagtatgctttggagggatcgattgctgtcgctggatctgtcgtgcaatggttaagagataaccttcgtttgatt
tccaaggctcaggacgtcggaccattggcttctcaagttgacaactctggaggtgtggtatttgttccagcattttcaggattgtttgccccttact
gggattccaactccagaggaaccattttcggtctgacccaatacacctcagcttctcatattgctagagctgctttggaaggtgtctgtttccaaa
ctagagccattttgaaggccatgatcagcgatgcaggagcttctgctgacttttggaggaatcatccaaggccactggccacaaccctctgt
cagttcttgccgtggacggaggtatgtccaaatcagacgagatgatgcagatccaagctgatattttgggtccatgtgtcactgttagacgttcc
atcaaccctgaatgtactgcactgggagctgccattgctgccggttttggtgtccctaaggaagatagaatttggggttccttgaaggaatgta
ccgaggccattcttgagggtaacaagatgtacttggctgcagggaacacttctttggacttcaaggccacattgagcgacgaggtcagaaga
aaggaatggagattgtgggaaaatgccattgcaaaggcaaagggctggcttaaggacactgcttaa

Supplementary data

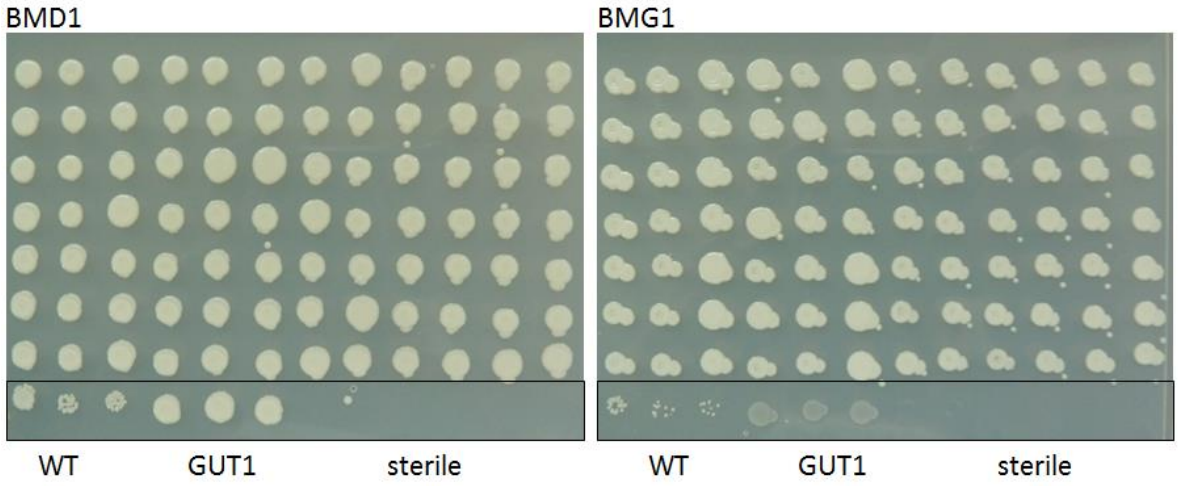
S 1: *P. pastoris* CBS 7435 WT transformants bearing various ARS plasmids. The vectors are based on the pPpT4_GAP plasmid (IMBT 6072). Different ARSs (PARS1, panARSOPT, B1739) were added upstream to the pUC ORI. The sTomato reporter gene is expressed under the control of P_{GAP}. The plasmids contain either a Zeocin or a Geneticin marker cassette. Ten ng circular plasmid DNA were used for the transformation. The transformants were plated on selective agar plates. As controls pPpT4_GAP vectors expressing sTomato, which did not contain one of the ARSs were transformed. No viable transformats were obtained, indicating that no additional sequence on the vector is recognized as ARS.



S 2: *P. pastoris* CBS 7435 WT and CBS 7435 *kU70* transformants bearing CRISPR-Cas9 plasmids with ribozyme-flanked gRNAs. The vectors are based on the pPpT4_HTX1-PARS1 plasmid (A168). The bidirectional P_{HTX1} promoter directs the expression of either *Sp*Cas9, *Pp*Cas9 or *Hs*Cas9 antisensewise and the gRNA2-GUT1 flanked by the HH and HDV ribozyme sensewise. The colony count differs between the transformations with the different Cas9 sequences. *Pp*Cas9 and *Hs*Cas9 appear to have a toxic effect in the WT as well as in the *KU70*- strain. Approximately 50 WT transformants were obtained with plasmids expressing *Pp*Cas9 and 500 WT transformants expressing *H*sCas9. Only very few viable CBS 7435 *ku70* transformants were obtained with *Hs*Cas9 or *Pp*Cas9. In contrast *Sp*Cas9 does not have a toxic effect neither in the WT nor in *KU70*- variant.



S 3: *GUT1* targeting efficiencies using a CRISPR-Cas9 construct expressing only *Hs*Cas9 without a ribozyme flanked gRNA. The gRNA is essential for CRISPR-Cas9 targeting and no obvious ds breaks have been introduced in *GUT1*. 84 transformants were inoculated in 96-DWPs containing YDP-Zeo media. The cells were transferred with a metallic stamp on agar plates containing minimal media with either glucose (BMD1) or glycerol (BMG1) as C-source. CBS 7435 WT was inoculated, but the strain does not contain a Zeocin resistance cassette and should not be able to grow. The *GUT1*- control strain was generated during this thesis. It contains a Zeocin cassette integrated in the *GUT1* locus.



CRISPR-Cas9 vector design

A *P. pastoris* CRISPR-Cas9 vector cloning vector is deposited in the strain collection of the institute (A171, IMBT 7414). The vector bears the *Hs*Cas9 under the control of the bidirectional P_{HTX1}. The gRNA to target your sequence of interest can be ordered on a synthetic ds DNA (gBlock). The vector has to be linearized with *Not*I resulting in a 7.7 kbp band and the gBlock (260 bp) can to be cloned in the vector by Gibson assembly. For RE-cloning the linearized vector has to be dephosphorylated (single RE-site cloning!) and the design of the gBlock has to be altered (no overhangs to the vector, addition of RE-sites). It is also possible to order different gRNAs on primers for site directed mutagenesis, however not only the 20 bp variable sequence of the gRNA, but also 6 bp sequence of the HH ribozyme has to be changed depending on the gRNA sequence. Thus the total length of the primer is at least 110 bp. Primers of this size are more expensive than gBlocks and additionally a PCR has to be performed.

The gBlock for Gibson cloning consists of:

Overhang to the P_{HTX1} promoter - **Variable part of the Hammerhead ribozyme** - Hammerhead ribozyme - **Variable part of gRNA** - Structural part of gRNA - HDV ribozyme - Overhang to the terminator

CCAGTTCAAGTTACCTAAACAAATCAAA**NNNNNN**CTGATGAGTCCGTGAGGACGAAACGAG
TAAGCTCGTC**NNNNNN**NNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATA
AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTGGCCGGCATGGTC
CCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGACTCAAGAG
GATGTCAGAATGCC

Vector maps of the IMBT 7414 (A171) plasmid as well as the plasmids to target *GUT1* (A178, A179, A180) are provided on a USB stick accompanying this master thesis. The variable part of the Hammerhead ribozyme has to pair with the first six bases of the variable part of the gRNA as described in the thesis. Thus the reverse complement sequence has to be entered at this position (http://www.bioinformatics.org/sms/rev_comp.html, 15/04/22). Useful instruction for ribozyme design and CRISPR-Cas9 can be found here: http://labs.biology.ucsd.edu/zhao/CRISPR_web/RGR_design_home_frame_set.html (15/04/22).

Here is an example for the design of the gBlock to target *GUT1*:

CCAGTTCAAGTTACCTAAACAAATCAAA**AACAAC**CTGATGAGTCCGTGAGGACGAAACGAG
TAAGCTCGTC**GTTGTT**TGGTCCAAGAAGACGTTTTAGAGCTAGAAATAGCAAGTTAAAATA
AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTGGCCGGCATGGTC
CCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGACTCAAGAG
GATGTCAGAATGCC

New CRISPR-Cas9 vectors for simplified construction will be provided soon (not part of the thesis, contact: aweninger@student.tugraz.at).

Selection and design of gRNAs

The gRNA consists of an 80 bp structural part (stg RNA) and a variable 20 bp part. The 20 bp variable part has to be complementary to the DNA sequence, which is targeted by Cas9. **Attention!** Cas9 binding is not possible at any position, but requires the presence of the PAM (protospacer adjacent motif,

sequence: N**GG**), due to the binding mechanism of Cas9: The Cas9 nuclease scans the DNA and starts unwinding of the DNA strands at the PAM. In case of base pairing between the gRNA and the complementary DNA strand, Cas9 is able introduce a ds break [17]. In theory, any DNA sequence, which habors NGG can be used as gRNA sequence. Following guidelines will help you to successfully design a gRNA.

- ➔ Retrieve the DNA sequence of the targeting region (CDS of the GOI) and check for introns. You should not select gRNAs from intron regions (assuming you want to introduce a frameshift mutation in the CDS of your GOI).
- ➔ Use to DNA2.0 gRNA design tool and insert your DNA sequence https://www.dna20.com/eCommerce/cas9/input
  The program is going to list gRNA-recommendations. Select the appropriate gRNA accordingly:
  - Avoid gRNAs with homodimers, hairpin motifs
  - The gRNA should bind in the middle of the CDS sequence, mutations at the rather end may not have an influence on the functionality of the protein
- ➔ The sequence, which is obtained from the DNA2.0 gRNA design tool, is the variable part of the gRNA (20bp) + NGG (the PAM sequence). Only insert the 20bp gRNA sequence in the vector map. You must not insert the PAM motif in the vector.
- ➔ Finally Blast your gRNA + PAM (!!) sequence (CRISPR cuts only, when PAM is available) for identical sequences. Check any combination for the PAM (AGG, TGG, GGG, CGG). If you find more than one identical sequence, do not use the gRNA (100% off – target region). If you obtain a sequence, which is almost 100% identical on the 3' end also think of using a different gRNA. Do not be too precise. There will always be off-target sites.

Hands-on CRISPR-Cas9:

Can this 20 bp sequence be a possible gRNA?

CGAGGCCAGAGCCTTGGAAA

- ➔ We don't know

Can this sequence be a possible gRNA?

CGAGGCCAGAGCCTTGGAAA**TGG**

- ➔ yes

You perform a Blast search and find out the following sequence occurs two times in the genome:

(A) CGAGGCCAGAGCCTTGGAAA
➔ no problem, off-targeting will only occur, if there is a PAM motif

(B) CGAGGCCAGAGCCTTGGAAA**TGG**
➔ 100% off-target region, do not use the sequence for targeting

(C) Is "CGAGGCCAGAGCCTTGGAAA**TGG**" the only off-targeting site?
➔ No, 100% off-targeting also occurs at CGAGGCCAGAGCCTTGGAAA**AGG,** CGAGGCCAGAGCCTTGGAAA**CGG** and CGAGGCCAGAGCCTTGGAAA**GGG** sites

List of materials

This list of materials is based on the Master theses of Florian Krainer and Lukas Sturmberger.

Chemicals

Acetic acid (100%)	Carl Roth GmbH, Karlsruhe, Germany
α-D(+)-glucose monohydrate	Carl Roth GmbH, Karlsruhe, Germany
Ammonia	Carl Roth GmbH, Karlsruhe, Germany
Ammonium acetate	Carl Roth GmbH, Karlsruhe, Germany
Ampicillin	Sigma-Aldrich GmbH, Vienna, Austria
Aqua bidest. "Fresenius"	Fresenius Kabi Austria GmbH, Graz, Austria
Bacto™ peptone	Becton Dickinson and Company, Sparks, MD, USA
Bacto™ yeast extract	Becton Dickinson and Company, Sparks, MD, USA
Bicine	Fluka Chemia AG, Basel, Switzerland
D-biotin	Fluka Chemia AG, Basel, Switzerland
Difco™ yeast nitrogen base w/o amino acids	Becton Dickinson and Company, Sparks, MD, USA
Dimethyl sulfoxide	Carl Roth GmbH, Karlsruhe, Germany
Dithiothreitol (DTT)	Carl Roth GmbH, Karlsruhe, Germany
D-sorbitol	Carl Roth GmbH, Karlsruhe, Germany
EDTA (disodium ethylenediaminetetraacetic acid)	Carl Roth GmbH, Karlsruhe, Germany
Ethanol abs.	Merck KGaA, Darmstadt, Germany
Ethidium bromide (≥ 98%)	Carl Roth GmbH, Karlsruhe, Germany
Ethylene glycol	Sigma-Aldrich GmbH, Vienna, Austria
Geneticin disulfate (G418 sulfate)	Carl Roth GmbH, Karlsruhe, Germany
Glycerol (≥ 98%)	Carl Roth GmbH, Karlsruhe, Germany
HCl (37%)	Merck KGaA, Darmstadt, Germany
K_2HPO_4	Carl Roth GmbH, Karlsruhe, Germany
K_2SO_4	Carl Roth GmbH, Karlsruhe, Germany
Kanamycin	Carl Roth GmbH, Karlsruhe, Germany
KCl	Merck KGaA, Darmstadt, Germany
KOH	Carl Roth GmbH, Karlsruhe, Germany
LB agar	Carl Roth GmbH, Karlsruhe, Germany
Methanol (≥ 99.8%)	Carl Roth GmbH, Karlsruhe, Germany
NaCl	Carl Roth GmbH, Karlsruhe, Germany
NaOH	Carl Roth GmbH, Karlsruhe, Germany
o-phosphoric acid (≥ 85%)	Carl Roth GmbH, Karlsruhe, Germany
Phenol:chloroform:isoamylalcohol (25.24:1)	Carl Roth GmbH, Karlsruhe, Germany
Tris	Carl Roth GmbH, Karlsruhe, Germany
Triton X-100	Carl Roth GmbH, Karlsruhe, Germany

Zeocin™	InvivoGen-Eubio, Vienna, Austria
DAPI (4′,6-Diamidino-2-phenylindole dihydrochloride)	Sigma-Aldrich GmbH, Vienna, Austria
Hoechst 33258	Sigma-Aldrich GmbH, Vienna, Austria

Instruments and devices

The laboratory equipment and all technical devices used in this thesis.

Centrifuges

Eppendorf Centrifuge 5810 R	Eppendorf AG, Hamburg, Germany
Eppendorf Centrifuge 5415 R	Eppendorf AG, Hamburg, Germany
Eppendorf Centrifuge 5415 D	Eppendorf AG, Hamburg, Germany

Shakers and incubators

Eppendorf Thermomixer comfort	Eppendorf AG, Hamburg, Germany
HT Infors Multitron shaker	Infors AG, Bottmingen, Switzerland
HT Infors Orbitron shaker	Infors AG, Bottmingen, Switzerland
HT Infors RS306 shaker	Infors AG, Bottmingen, Switzerland
Binder drying oven	Binder GmbH, Tuttlingen, Germany

PCR cycler

Applied Biosystems 2720 Thermal Cylcer Applied Biosystems	Inc., Foster City, CA, Unites States

Photometers and associated materials

Eppendorf BioPhotometer plus	Eppendorf AG, Hamburg, Germany
Semi-micro cuvette 10x4x45 mm, Polystyrene	Sarstedt Aktiengesellschaft & Co., Nümbrecht, Germany
NanoDrop 2000c Spectrophotometer	peqlab Biotechnologie GmbH, Polling, Austria

Platereader

SynergyMx Plate Reader	Biotek Inc., Winooski, United States

Gel electrophoresis and associated materials

BioRad PowerPac Basic Power Supply	Bio-Rad Laboratories Inc., Hercules, CA, United

	States
Biozym LE Agarose	Biozym Biotech Trading GmbH, Vienna, Austria
GelDoc-It™ Imaging System	UVP®, Cambridge, UK
Sub-cell GT	Bio-Rad Laboratories Inc., Hercules, CA, United States
Transilluminator Chroma 43	Vetter GmbH, Germany
Digital printer UP-D897	Sony, Vienna, Austria
6x DNA Loading Dye	Fisher Scientific - Austria GmbH, Vienna, Austria
GeneRuler™ 1kb DNA-Ladder	Fisher Scientific - Austria GmbH, Vienna, Austria

Pipettes

Eppendorf Research® Series 2100 Adjustable Volume, Single Channel Pipette, 0.1-2.5 µL	Eppendorf AG, Hamburg, Germany
Eppendorf Research® Series 2100 Adjustable Volume, Single Channel Pipette, 0.5 -10 µL	Eppendorf AG, Hamburg, Germany
Eppendorf Research® Series 2100 Adjustable Volume, Single Channel Pipette, 2 - 20 µL	Eppendorf AG, Hamburg, Germany
Eppendorf Research® Series 2100 Adjustable Volume, Single Channel Pipette, 20 - 200 µL	Eppendorf AG, Hamburg, Germany
Eppendorf Research® Series 2100 Adjustable Volume, Single Channel Pipette, 100 - 1000 µL	Eppendorf AG, Hamburg, Germany
Biohit Proline® multichannel electronic pipettor, 8 channels, 5-100 µL	Biohit Oyj, Helsinki, Finnland
Biohit Proline® multichannel electronic pipettor, 8 channels, 50-1200 µL	Biohit Oyj, Helsinki, Finnland
Acura® manual 855 multichannel micropipette, 5-50µl	Socorex Isba S.A., Ecublens, Swiss

Microtiterplates

96 well PS Microplater sterile	Greiner Bio-One GmbH, Frickenhausen, Germany
Nunc™ MicroWell™ 96-Well Optical-Bottom Plates with Polymer Base	Thermo Fisher Scientific Inc., Rochester, NY, United States
Bel-Art 96-Well Deep Well Plates	Bel-Art Products, Wayne, NJ, United States
Cover for deep well plate	Bel-Art Products, Wayne, NJ, United States
MicroAmp® Optical 96-Well Reaction Plate	Applied Biosystems, Foster City, CA, US

MicroAmp® Optical Adhesive Covers	Applied Biosystems, Foster City, CA, USA

Electroporation materials

Bio-Rad BioRad Gene Pulser 1652076	Bio-Rad Laboratories Inc., Hercules, CA, United States
Capacitance Extender 1652087	Bio-Rad Laboratories Inc., Hercules, CA, United States
Pulse Controller P/N 1652098	Bio-Rad Laboratories Inc., Hercules, CA, United States
Electroporation cuvettes EP-102	Cell Projects Ltd, Kent, United Kingdom

Other devices

MT PG12001-S DeltaRange Balance	Mettler Toledo Inc., Greifensee, Switzerland
Heidolph MR 2002 Mangentic Stirrer	Heidolph Instruments, Schwabach, Germany
Heidolph MR 3000 Magnetic Stirrer	Heidolph Instruments, Schwabach, Germany
inoLab pH 720 pH-Meter	WTW GmbH, Weilheim, Germany
Hamilton® Polyplast lab pH electrode	Sigma-Aldrich GmbH, Vienna, Austria
Vortex Genie 2	Scientific Industries Inc., Bohemia, NY, United States
Certoclav LVEL 12L	CertoClav GmbH, Traun, Austria
arium® basic ultrapure water system	Sartorius Stedim Biotech GmbH, Göttingen, Germany
Sartorius Analytical BL120S	Sartorius Stedim Biotech GmbH, Göttingen, Germany

Reaction tubes

Microcentrifuge tubes, 1.5 mL with lid	Greiner Bio GmbH, Frickenhausen, Germany
PP-Tube, sterile, 15 mL	Greiner Bio GmbH, Frickenhausen, Germany
PP-Tube, sterile, with/without support skirt, 50 mL	Greiner Bio GmbH, Frickenhausen, Germany

Pipette tips

Pipette tips, micro P10	Greiner Bio-One GmbH, Frickenhausen, Germany
Pipette tips 200	Greiner Bio-One GmbH, Frickenhausen, Germany

Pipette tips 1000	Greiner Bio-One GmbH, Frickenhausen, Germany
Biohit Tips 300 µL Single Tray	Biohit Oyj, Helsinki, Finnland
Biohit Tips 1200 µL Bulk	Biohit Oyj, Helsinki, Finnland

Microscopy equipment

Leica DM LB microscope	Leica Mikrosysteme GmbH, Austria
Microscope slides, Nr.0656	Carl Roth GmbH, Karlsruhe, Germany
Immersion oil 50 ml, Nr. X899.1	Carl Roth GmbH, Karlsruhe, Germany
Coverslips, Nr. 0657	Carl Roth GmbH, Karlsruhe, Germany

Enzymes

All enzymes used in this thesis.

Enzyme	**Company**
FastDigest *Swa*I	Fisher Scientific GmbH, Vienna, Austria
FastDigest *Pci*I	Fisher Scientific GmbH, Vienna, Austria
FastDigest *EcoR*I	Fisher Scientific GmbH, Vienna, Austria
FastDigest *BamH*I	Fisher Scientific GmbH, Vienna, Austria
FastDigest *Not*I	Fisher Scientific GmbH, Vienna, Austria
FastDigest *Bgl*II	Fisher Scientific GmbH, Vienna, Austria
FastDigest *Xho*I	Fisher Scientific GmbH, Vienna, Austria
FastDigest *Xba*I	Fisher Scientific GmbH, Vienna, Austria
*Nhe*I	Fisher Scientific GmbH, Vienna, Austria
*Sal*I	Fisher Scientific GmbH, Vienna, Austria
*Sfi*I	New EngIand Biolabs, Ipswich, MA, Unites States
Taq DNA Polymerase	Promega GmbH, Mannheim, Germany
Phusion High Fidelity DNA Polymerase	Fisher Scientific GmbH, Vienna, Austria

T4 DNA Ligase	Fisher Scientific GmbH, Vienna, Austria
FastAP Thermosensitive Alkaline Phosphatase	Fisher Scientific GmbH, Vienna, Austria
T5 Exonuclease	New EngIand Biolabs, Ipswich, MA, Unites States
Taq DNA Ligase	New EngIand Biolabs, Ipswich, MA, Unites States
GoTaq® Polymerase	Promega GmbH, Mannheim, Germany

Software and webtools

(Date: 2/6/2015)

CRISPR gRNA Design tool	https://www.dna20.com/eCommerce/cas9/input
EMBOSS 6.3.1: freak	http://mobyle.pasteur.fr/cgi-bin/portal.py#forms::freak
RNA secondary structure prediction	http://www.genebee.msu.su/services/rna2_reduced.html
Expasy Translate Tool	http://web.expasy.org/translate/
Basic Local Alignment Search Tool	http://blast.ncbi.nlm.nih.gov/Blast.cgi
Gene Designer 1.1.4.1	DNA2.0 Inc., Menlo Park, CA, USA
CLC Main Workbench	QIAGEN N.V., Spoorstraat, Netherlands
GNU Image Manipulation Program (Gimp) 2.8	The GIMP Development Team, www.gimp.org

Primer and gBlocks

S 4 lists all primers and gBlocks generated in this master thesis. The synthetic DNA fragments were ordered from IDT, Integrated DNA Technologies, Inc., Coralville, Iowa. The primers and gBlocks were stored in the internal primer collection at -20°C.

S 4: Primer and gBlocks used during this thesis

Primer name	Sequence 5' --> 3'	Internal collection number
Nuclear localization signals		
pAOX1-nNLS SV40-eGFP-Gib	GACAActtgagaagatcaaaaaacaactaattattcgaaacgATGCCAAAGAAGAAAAGAAAAGTTGCTAGCAAAGGAGAAGAACTTTTCACTG	P14190
AOX1TT-cNLS SV40-eGFP-Gib	GCAAATGGCATTCTGACATCCTCTTGATTAAACTTTTCTTTTCTTCTTTGGCTTGTACAATTCATCCATGCCATGTGT	P14205
pAOX1-nNLS c-myc-eGFP-Gib	GACAActtgagaagatcaaaaaacaactaattattcgaaacgATGCCAGCTGCTAAGAGAGTTAAGTTGGATGCTAGCAAAGGAGAAGAACTTTTCACTG	P14191
AOX1TT-cNLS c-myc-eGFP-Gib	GCAAATGGCATTCTGACATCCTCTTGATTAATCCAACTTAACTCTCTTAGCAGCTGGCTTGTACAATTCATCCATGCCATGTGT	P14206
pAOX1-nNLS nucleoplasmin-eGFP-Gib	GACAActtgagaagatcaaaaaacaactaattattcgaaacgATGAAGAGACCTGCTGCTGCCACTAAGAAAGCAGGGCAAGCTAAGAAGAAGAAGGCTAGCAAAGGAGAAGAACTTTTCACTG	P14192
AOX1TT-cNLS nucleoplasmin-eGFP-Gib	GCAAATGGCATTCTGACATCCTCTTGATTACTTCTTCTTCTTAGCTTGCCCTGCTTTCTTAGTGGCAGCAGCAGGTCTCTTCTTGTACAATTCATCCATGCCATGTGT	P14193
pAOX1-nNLS tr Mata2-eGFP-Gib	GACAActtgagaagatcaaaaaacaactaattattcgaaacgATGAAGATTCCAATTAAGGCTAGCAAAGGAGAAGAACTTTTCACTG	P14207
AOX1TT-cNLS tr Mata2-eGFP-Gib	GCAAATGGCATTCTGACATCCTCTTGATTACTTAATTGGAATCTTCTTGTACAATTCATCCATGCCATGTGT	P14208

pAOX1-eGFP-Gib	GACAActtgagaagatcaaaaaacaactaattattcgaaacgATGGCTAGC AAAGGAGAAGAACTTTTCACTG	P14203
AOX1TT-eGFP-Gib	GCAAATGGCATTCTGACATCCTCTTGATTACTTGTAC AATTCATCCATGCCATGTGT	P14204
gBlock-cNLS-pichia1	CCCTGTCCTTTTACCAGACAACCATTACCTGTCGACA CAATCTGCCCTTTCGAAAGATCCCAACGAAAAGCGT GACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTG GGATTACACATGGCATGGATGAATTGTACAAGAAAG GAAGGCGGGCTAATGCCTCAAAGAAGAAGAAGTAAT CAAGAGGATGTCAGAATGCCATTTGCCTGAGAGATG CAGGCTTCATTTTTGATACTTTTTTATTTGTAACCTAT ATAGTATAGGATTTTTTTTGTCATTTTGTTTCTTCTCG TACGAGCTTGCTCCTGATCAGCCTATCTCGCAGCAGA TGAATATCTTGTGGTAGGGGTTTGGGAAAATCATTCG AGTTTGATGTTTTTCTTGGTATTTCCCACTCCTCTTCA GAGTACAGAAGATTAAGTGAGACCTTCGTTTGTGCG GATCCTTCAGTAATGTCTTGTTTCTTTTGT	P14279
gBlock-cNLS-pichia2	CCCTGTCCTTTTACCAGACAACCATTACCTGTCGACA CAATCTGCCCTTTCGAAAGATCCCAACGAAAAGCGT GACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTG GGATTACACATGGCATGGATGAATTGTACAAGAAGC AAAAGGTTCTACGAGCTCATATCGATAAGCAAAAAA AGAAGGGTCATTAATCAAGAGGATGTCAGAATGCCA TTTGCCTGAGAGATGCAGGCTTCATTTTTGATACTTTT TTATTTGTAACCTATATAGTATAGGATTTTTTTTGTCA TTTTGTTTCTTCTCGTACGAGCTTGCTCCTGATCAGCC TATCTCGCAGCAGATGAATATCTTGTGGTAGGGGTTT GGGAAAATCATTCGAGTTTGATGTTTTTCTTGGTATTT CCCACTCCTCTTCAGAGTACAGAAGATTAAGTGAGAC CTTCGTTTGTGCGGATCCTTCAGTAATGTCTTGTTTCT TTTGT	P14280

gBlock-cNLS-pichia3	CCCTGTCCTTTTACCAGACAACCATTACCTGTCGACA CAATCTGCCCTTTCGAAAGATCCCAACGAAAAGCGT GACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTG GGATTACACATGGCATGGATGAATTGTACAAGAAGA GAAAACTTGAAGAAGAGGAAGGGTCAAAGAGAAAC AAACGGATAAAAGGTTAATCAAGAGGATGTCAGAAT GCCATTTGCCTGAGAGATGCAGGCTTCATTTTTGATA CTTTTTTATTTGTAACCTATATAGTATAGGATTTTTTT TGTCATTTTGTTTCTTCTCGTACGAGCTTGCTCCTGAT CAGCCTATCTCGCAGCAGATGAATATCTTGTGGTAGG GGTTTGGGAAAATCATTCGAGTTTGATGTTTTTCTTG GTATTTCCCACTCCTCTTCAGAGTACAGAAGATTAAG TGAGACCTTCGTTTGTGCGGATCCTTCAGTAATGTCT TGTTTCTTTTGT	P14281
gBlock-cNLS-pichia4	CCCTGTCCTTTTACCAGACAACCATTACCTGTCGACA CAATCTGCCCTTTCGAAAGATCCCAACGAAAAGCGT GACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTG GGATTACACATGGCATGGATGAATTGTACAAGAAGC GACCTCTGGAAATAGAGCAGGAAGAAACATATTCGA AAAGAAAGAAGAGTACTATATAATCAAGAGGATGTC AGAATGCCATTTGCCTGAGAGATGCAGGCTTCATTTT TGATACTTTTTTATTTGTAACCTATATAGTATAGGATT TTTTTTGTCATTTTGTTTCTTCTCGTACGAGCTTGCTC CTGATCAGCCTATCTCGCAGCAGATGAATATCTTGTG GTAGGGGTTTGGGAAAATCATTCGAGTTTGATGTTTT TCTTGGTATTTCCCACTCCTCTTCAGAGTACAGAAGA TTAAGTGAGACCTTCGTTTGTGCGGATCCTTCAGTAA TGTCTTGTTTCTTTTGT	P14282

gBlock-cNLS-pichia5(SWI5)	CCCTGTCCTTTTACCAGACAACCATTACCTGTCGACA CAATCTGCCCTTTCGAAAGATCCCAACGAAAAGCGT GACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTG GGATTACACATGGCATGGATGAATTGTACAAGAAAA AGTTCGTCAGAAATCATGATCTTCGAAGGCATAAAA AGAAATAATCAAGAGGATGTCAGAATGCCATTTGCC TGAGAGATGCAGGCTTCATTTTTGATACTTTTTTATTT GTAACCTATATAGTATAGGATTTTTTTTGTCATTTTGT TTCTTCTCGTACGAGCTTGCTCCTGATCAGCCTATCTC GCAGCAGATGAATATCTTGTGGTAGGGGTTTGGGAA AATCATTCGAGTTTGATGTTTTTCTTGGTATTTCCCAC TCCTCTTCAGAGTACAGAAGATTAAGTGAGACCTTCG TTTGTGCGGATCCTTCAGTAATGTCTTGTTTCTTTTGT	P14283
gBlock-cNLS-SWI5	CCCTGTCCTTTTACCAGACAACCATTACCTGTCGACA CAATCTGCCCTTTCGAAAGATCCCAACGAAAAGCGT GACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTG GGATTACACATGGCATGGATGAATTGTACAAGAAGA AGTACGAAAACGTTGTTATCAAGAGATCCCCAAGAA AGAGAGGUAGACCAAGAAAATAATCAAGAGGATGT CAGAATGCCATTTGCCTGAGAGATGCAGGCTTCATTT TTGATACTTTTTTATTTGTAACCTATATAGTATAGGAT TTTTTTTGTCATTTTGTTTCTTCTCGTACGAGCTTGCTC CTGATCAGCCTATCTCGCAGCAGATGAATATCTTGTG GTAGGGGTTTGGGAAAATCATTCGAGTTTGATGTTTT TCTTGGTATTTCCCACTCCTCTTCAGAGTACAGAAGA TTAAGTGAGACCTTCGTTTGTGCGGATCCTTCAGTAA TGTCTTGTTTCTTTTGT	P14284
gBlock-nSWI5	GAGAAGATCAAAAAACAACTAATTATTCGAAACGAT GAAGAAGTACGAAAACGTTGTTATCAAGAGATCCCC AAGAAAGAGAGGUAGACCAAGAAAAGCTAGCAAAG GAGAAGAACTTTTCACTG	P14842
Autonomously replicating sequences		
pGAP-sTOM-fw-Gib	CCCTATTTCAATCAATTGAACAACTATCAAAACACAA TGGTTTCTAAGGGTGAGGAAGTTATCAAG	P14196
AOXTT-sTom-rv-Gib	CAAATGGCATTCTGACATCCTCTTGATTACTTATAAA GCTCGTCCATACCGTACAAGAA	P14197
sTom-fw-Gib	gcatacgttaagcatccagcagacattcctgacta	P14198
sTom-rv-Gib	Tagtcaggaatgtctgctggatgcttaacgtatgc	P14199

AODTT-PARS1-fw-Gib	ACTTGGATCTGATTACCTTAGCTGCATCGAGATAAGC TGGGGGAACATTC	P14200
pUCORI-PARS1-rv-Gib	GATCTTTTCTACGGGGTCTGACGCTCAGTGGTACCTC GACAATTAATATTTACTTATTTTGGTCAACCCC	P14201
pGAP-rv-Gib	AGGACACCAAGACATTTCTACAAAAAATTTAAATGA ATAATAACTGTGTATTTTTCAGTGTTCCCG	P14202
AODTT-B1739-fw-Gib	GAAACTTGGATCTGATTACCTTAGCTGCATAATGACA ACTGTGGGGTTCGAACCCACGCCTCCGGAGAGACCA GAACCTTAATCTGGCGCCTTAGACCAACTCGGCCAAA TTGTCTCTACGTAAGTATACACGAAAGATGTCATATA TAAGGGAGGTACCACTGAGCGTCAGAC	P14210
G-Block-panARSOPT	TTGGATCTGATTACCTTAGCTGCATCAACATCTTTGG ATAATATCAGAATGAGAAAGAACAGATACGCAGTAC GTTTTTTGGTGAGCTCTTTGCACTTCTTTAGTTCTTTC CATCAATATCAGTTttTTAaaCttTTAgGACTAAaAgTGAT GTTTAACTTCAAaATgTTTAAAaTTTTGTTCTTCCCGAC GTTCATTAAGAATACTAATACACTTTAATAATTAtTTT AATATTTtgTTCTAaATAATGACtTTTAATTAAAAAAGA TAAAATATAAAAACATCATAATAACTCACCAGAGGT TAAGAACAAAAAAACAAATTAGATATCTGCTAATCC AATATAGTTAAATCAATCTTTCCTTGGTATAATGGGT ATATTACATATATTTCAAGGACCGACACTCCTACCAA ATATCTAAAATTTACCATATTAACATAACATGTATAT AAACGTCAAATCATAATCAGCACTAGGTACCACTGA GCGTCAGACCCC	P14259

Control strains

3UTR-GUT1-F-Gibson	CCTGGCCTTTTGCTGGCCTTTTGCTCagagcagctgtaattatatt atcatgttaggtca	P14359
3UTRGUTR	GTGTTTGCTGTAGGATGACCTAGATTTAAATATAAGA GGAAACAACGTTCGTATCGTGA	P14252
5UTRGUTF	CACGATACGAACGTTGTTTCCTCTTATATTTAAATCT AGGTCATCCTACAGCAAACACC	P14253
5UTR-GUT1-R-Gibson	CTAAGATACGTTCCGTTCCtatagtagatatatctgtggtatagtgtgaa aaagtagaag	P14360
3UTR-AOX1-F-Gibson	CCTGGCCTTTTGCTGGCCTTTTGCTCTCAAGAGGATG TCAGAATGCCATTTG	P14357

3UTRAOXR	GTAGGCGCTGGGATTTAAATTTCTCTAAAACAAGATAGTGCCCCTCAAG	P14248
5UTRAOXF	GGCACTATCTTGTTTTAGAGAAATTTAAATCCCAGCGCCTACAATGATG	P14249
5UTR-AOX1-R-Gibson	GCTAAGATACGTTCCGTTCCCGTTTCGAATAATTAGTTGTTTTTTGATCTTCTCAAGTTG	P14358
3UTR-ADE2-F-Gibson	CTTTTGCTGGCCTTTTGCTCATTTAGTATTGTTTTTTAATAGATGTATATATAATAGTACACGTAACTTATCTATTC	P14361
3UTRADER	GTGGGGTGAGGCAGATGATTTAAATTATGCTCCAGCTGTTTCATTCTTTTCC	P14256
5UTRADEF	GGAAAAGAATGAAACAGCTGGAGCATAATTTAAATCATCTGCCTCACCCCAC	P14257
5UTR-ADE2-R-Gibson	GCTAAGATACGTTCCGTTCCGAGAGTTTGGAAGTTTTTTAATCGTTTCTACGATTC	P14362
3UTR-F-ADE1	CTTTTGCTGGCCTTTTGCTCATGATTAGTACCCTCCTCGCCTTTTTC	P14562
5UTR-R-ADE1	GCTAAGATACGTTCCGTTCCACTTGGAAGGATACAGCAAAGTGTG	P14563
5UTR-F-ADE1	GCTTCAATGTGTATATCAAGGAGGTTAATTTAAATAGCTCTCTGCGAAAATGTCAAGAAT	P14564
3UTR-R-ADE1	ATTCTTGACATTTTCGCAGAGAGCTATTTAAATTAACCTCCTTGATATACACATTGAAGC	P14565
Arg4TT-KpnI-rv	GGGGTCTGACGCTCAGTGGTACCAATGCGAGGATGCTG	P14245
pArg4-BglII-fv	CCTATACTTTCTAGAGAATAGGAACTTCAGATCTGGAACGGAACGTATCTTAGCATGG	P14246
Parg_rv	TGAATCTGTCGCACAACCATGC	P14356
5ADE1_fw	CTCTCTGCGAAAATGTCAAGAATCTTGTTG	P14618
5ADE1_rv	GGCGAGGAGGGTACTAATCATACTTGGAAGGATACAGCAAAGTGTGTTG	P14619
3ADE1_fw	CAACACACTTTGCTGTATCCTTCCAAGTATGATTAGTACCCTCCTCGCC	P14620

3ADE1_rv	CTCCTTGATATACACATTGAAGCAACTTTCCT	P14621
5AOX1_fw	AAATCCCAGCGCCTACAATGATG	P14622
5AOX1_rv	CAAATGGCATTCTGACATCCTCTTGACAATTGGAACC AGTCGCAATTATGAAAGTAAG	P14623
3AOX1_fw	CTTACTTTCATAATTGCGACTGGTTCCAATTGTCAAG AGGATGTCAGAATGCCATTTG	P14624
3AOX1_rv	GATAGTGCCCCTCAAGTTGATGAG	P14625
5GUT1_fw	AAATCTAGGTCATCCTACAGCAAACACC	P14626
5GUT1_rv	GACCTAACATGATAATATAATTACAGCTGCTCGTAGA AGAAGAGTCTTTTTCAGTCCTTG	P14627
3GUT1_fw	CAAGGACTGAAAAAGACTCTTCTTCTACGAGCAGCT GTAATTATATTATCATGTTAGGTC	P14628
3GUT1_rv	AAATATAAGAGGAAACAACGTTCGTATCGTGATC	P14629

CRISPR-Cas9 constructs with RNA Polymerase III promoters

RNA Polymerase II promoters

pucori-SNR52-fw-Gib	GCCTTTTGCTGGCCTTTTGCTCATCTTTGAAAAGATA ATGTATGATTATGCTTTCACTC	P14267
stgRNA-Xho-SNR52-rv-Gib	CTTGCTATTTCTAGCTCTAAAACTCGAGGATCATTTA TCTTTCACTGCGGAGAAGTTTC	P14268
pucori-RPR1-fw-Gib	GCCTTTTGCTGGCCTTTTGCTCATCTGCCAATTGAACA TAACATGGTAGTTACATATAC	P14269
stgRNA-Xho-RPR1-rv-Gib	CTTGCTATTTCTAGCTCTAAAACTCGAGCTGCCAATC GCAGCTCCCAGAGTTTCGTTC	P14270
pucori-pLYS-fw-Gib	GCCTTTTGCTGGCCTTTTGCTCACTAGTGGGAAGGAC ACGAATCAAATGTAAAATC	P14271
stgRNA-Xho-pLYS-rv-Gib	CTTGCTATTTCTAGCTCTAAAACTCGAGGCTCCTCAT CAGGGGCTCGAACCCTGGAC	P14272
pucori-pMET-fw-Gib	GCCTTTTGCTGGCCTTTTGCTCAGTATCTAAGTAGCTG TATGCTGCTTGAAGTC	P14273
stgRNA-Xho-pMET-rv-Gib	CTTGCTATTTCTAGCTCTAAAACTCGAGTGCTCCACG GGAGGTTCGAACTCTCGACC	P14274

pucori-pSER-fw-Gib	TTTGCTGGCCTTTTGCTCAAAGTATCACAGTTGGATT AATTAGATATGAAAACTAATTG	P14275
stgRNA-Xho-pSER-rv-Gib	CTTGCTATTTCTAGCTCTAAAACTCGAGGCGTCACAG ACAGGATTCGAACCTGCGCAG	P14276
pucori-SUP4-fw-Gib	CTTTTGCTGGCCTTTTGCTCAAGTATACTCTTTCTTCA ACAATTAAATACTCTCGGTAG	P14277
stgRNA-Xho-SUP4-rv-Gib	CTTGCTATTTCTAGCTCTAAAACTCGAGTCTCCCGGG GGCGAGTCGAACGCCCGATC	P14278
pUCori-SCR1-fw-S-cer	CTGGCCTTTTGCTGGCCTTTTGCTCATGATCAACTTAG CCAGGACATCCATAG	P14296
stgRNA-SCR1-rv-S-cer	TTTAACTTGCTATTTCTAGCTCTAAAACTCGAGTATG GTTCAGGACACACTCCATCCCC	P14297
gBlock-pSNR52-K-lactis	CTGGCCTTTTGCTGGCCTTTTGCTCAttggtgaacccaatgggaa atgtgggtagactactttctttaatttatcattgatctcatggaaatctcttttgaaagtggaca cgatactttcatctaacgttcctctagcagaagtggataaagtggaaccgtaaccgttacc agtctcagtttcaataaaatctggttcaatggtgtccataattggccttactcgctgggatat ggtaattaattgtaaaaaaagaataaaatagaggtttttacttataagtcaattcagaagga aggcaacatgtttttaatcctttgatacaagcgttatttgtatcaatcagagattgttcacgc atagttattcatcatttgagattcgaactgccgagaaagtaacCTCGAGTTTTA GAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14319
gBlock-pRPR1-S-pombe	CTGGCCTTTTGCTGGCCTTTTGCTCAaagcttgaataggtgttgta aagtgttgatttatgtgacgctgagggtgcgcatgaaaggaatgttgggtcacgattatta aacagtttgctagcttggacacttgagtattggaagttgttgaattctaaaaaactttcagtt gatttgaatagttgctgttgccaaaaaacataacctgtaccgaagaaccatgctggacgt acgggcgaacgccgcacttcctcaaattcaaacgcgttgaaaagcgcacagctcgttg agggggtaaggtcggagaaacatcttcgttgcgtgctcgtgaggagcgaagaacgaa cgttctgccgaatgtaccagaaattcaatcagtatggcctcgtttgtcgtacctgaCTC GAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG CTAGTC	P14320
gBlock-pASN-P-angusta	CTGGCCTTTTGCTGGCCTTTTGCTCAacggagcgtagttgatcat ctgctgggcctgcttgtttctctgtatcttgtgcttctggaggcccagtctagcctgttcaga ctggtacttcgccgtcaacgcgatagtgttgtttatatcgtctagctgatacgaaaaaggc acagagtccaggaccattttgttcgtgctggaaagttttcggttcacgaaaattatattggc caaaaaatcttggggaggctgtttggtgggttatatcagagtacgttaagcaagatgacg aataagacttcgtggccaagctggtttaaggcatgcgactgttaatcgcaagatcgtgag ttcgatcctcaccgaggtcgttcgtaaCTCGAGTTTTAGAGCTAGAA ATAGCAAGTTAAAATAAGGCTAGTC	P14321
gRNAs		
SNR52-GUT1-gRNA1-stgRNA-fw	GCAGTGAAAGATAAATGATCCTAGTTGCTACCATCG ATATTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG CTAGTC	P14454

SNR52-GUT1-gRNA2-stgRNA-fw	GCAGTGAAAGATAAATGATCCGAGTACTCTACCTCTG CTCTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC TAGTC	P14455
SNR52-GUT1-gRNA3-stgRNA-fw	GCAGTGAAAGATAAATGATCTGCAATTTCCTCAGCCA GGCTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC TAGTC	P14456
SNR52-GUT1-gRNA4-stgRNA-fw	GCAGTGAAAGATAAATGATCGTTGTTTGGTCCAAGA AGACTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG CTAGTC	P14457
SNR52-GUT1-gRNA5-stgRNA-fw	GCAGTGAAAGATAAATGATCGTCCCACACTTGGAGT CTATTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG CTAGTC	P14458
RP2-RPR1-GUT1-gRNA2-stgRNA-fw	GGAGCTGCGATTGGCAGCGAGTACTCTACCTCTGCTC TTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTA GTC	P14566
RP2-RPR1-GUT1-gRNA3-stgRNA-fw	GGAGCTGCGATTGGCAGTGCAATTTCCTCAGCCAGGC TTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTA GTC	P14567
RP2-RPR1-GUT1-gRNA4-stgRNA-fw	GGAGCTGCGATTGGCAGGTTGTTTGGTCCAAGAAGA CTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCT AGTC	P14568
RP3-SUP4-GUT1-gRNA2-stgRNA-fw	CGCCCCCGGGAGACGAGTACTCTACCTCTGCTCTTTT AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14569
RP3-SUP4-GUT1-gRNA3-stgRNA-fw	CGCCCCCGGGAGATGCAATTTCCTCAGCCAGGCTTTT AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14570
RP3-SUP4-GUT1-gRNA4-stgRNA-fw	CGCCCCCGGGAGAGTTGTTTGGTCCAAGAAGACTTTT AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14571
RP4-LYS-GUT1-gRNA2-stgRNA-fw	gcccctgatgaggagcCGAGTACTCTACCTCTGCTCTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14572
RP4-LYS-GUT1-gRNA3-stgRNA-fw	gcccctgatgaggagcTGCAATTTCCTCAGCCAGGCTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14573
RP4-LYS-GUT1-gRNA4-stgRNA-fw	gcccctgatgaggagcGTTGTTTGGTCCAAGAAGACTTTTAG AGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14574
RP5-MET-GUT1-gRNA2-stgRNA-fw	gaacctcccgtggagcaCGAGTACTCTACCTCTGCTCTTTTAG AGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14575
RP5-MET-GUT1-gRNA3-stgRNA-fw	gaacctcccgtggagcaTGCAATTTCCTCAGCCAGGCTTTTAG AGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14576
RP5-MET-GUT1-gRNA4-stgRNA-fw	gaacctcccgtggagcaGTTGTTTGGTCCAAGAAGACTTTTAG AGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14577
RP6-SER-GUT1-gRNA2-stgRNA-fw	cgaatcctgtctgtgacgcCGAGTACTCTACCTCTGCTCTTTTAG AGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14578

RP6-SER-GUT1-gRNA3-stgRNA-fw	cgaatcctgtctgtgacgcTGCAATTTCCTCAGCCAGGCTTTTAG AGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14579
RP6-SER-GUT1-gRNA4-stgRNA-fw	cgaatcctgtctgtgacgcGTTGTTTGGTCCAAGAAGACTTTTA GAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14580
RP7-SNR52-GUT1-gRNA2-stgRNA-fw	cgaactgccgagaaagtaacCGAGTACTCTACCTCTGCTCTTTT AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14581
RP7-SNR52-GUT1-gRNA3-stgRNA-fw	cgaactgccgagaaagtaacTGCAATTTCCTCAGCCAGGCTTTT AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14582
RP7-SNR52-GUT1-gRNA4-stgRNA-fw	cgaactgccgagaaagtaacGTTGTTTGGTCCAAGAAGACTTTT AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14583
RP8-RPR1-GUT1-gRNA2-stgRNA-fw	gcctcgtttgtcgtacctgaCGAGTACTCTACCTCTGCTCTTTTAG AGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14584
RP8-RPR1-GUT1-gRNA3-stgRNA-fw	gcctcgtttgtcgtacctgaTGCAATTTCCTCAGCCAGGCTTTTA GAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14585
RP8-RPR1-GUT1-gRNA4-stgRNA-fw	gcctcgtttgtcgtacctgaGTTGTTTGGTCCAAGAAGACTTTTA GAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14586
RP9-ASN-GUT1-gRNA2-stgRNA-fw	ctcaccgaggtcgttcgtaaCGAGTACTCTACCTCTGCTCTTTTA GAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14587
RP9-ASN-GUT1-gRNA3-stgRNA-fw	ctcaccgaggtcgttcgtaaTGCAATTTCCTCAGCCAGGCTTTTA GAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14588
RP9-ASN-GUT1-gRNA4-stgRNA-fw	ctcaccgaggtcgttcgtaaGTTGTTTGGTCCAAGAAGACTTTTA GAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	P14589
RP10-SCR1-GUT1-gRNA2-stgRNA-fw	GGAGTGTGTCCTGAACCATACGAGTACTCTACCTCTG CTCTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG CTAGTC	P14590
RP10-SCR1-GUT1-gRNA3-stgRNA-fw	GGAGTGTGTCCTGAACCATATGCAATTTCCTCAGCCA GGCTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG CTAGTC	P14591
RP10-SCR1-GUT1-gRNA4-stgRNA-fw	GGAGTGTGTCCTGAACCATAGTTGTTTGGTCCAAGAA GACTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG CTAGTC	P14592

stgRNA-Stuffer	TAAATGATCCTCGAGTTTTAGAGCTAGAAATAGCAA GTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAA GTGGCACCGAGTCGGTGGTGCTTTTTTTGTTTTTTATG TCTggcgcgccGACCTCTGTTGCCTCTTTGTTGGACGAAC CATTcACCGGTGTCTTaTACTTAAAGGGCAGTGGTATC ACTGAAGACTTCCAGTCCCTAAAGGGTAAGAAGATC GGTTACGTTGGTGATTCGGTAAGATCCAAATCGATGA ATTGACCAAGCACTACGGTATGAAGCCAGAAGACTA CACtGCtGTCAGATGTGGTATGAATGTCGCCAAGTACA TCATCGAAGaTAAGATTGATGCtGGTATTGGTATCGAA TGTATGCAACAAGTCGAATTGGAAGAGTACTTGGCC AAGCAAGGCAGACCAGCTTCTGATGCTAAAATGTTG AGAATTGACAAGTTGGCTTGCTTGGGTTGCTGTTGtTT CTGTACCGTTCTTTACATCTGCAACGATGAATTTTTG AAGAAaAACCCTGAAAAGGTCAGAAAGTTCTTGAAA GCCATCAAGAAGGCAACCGACTACGTTCTAGCCGAC CCTGTGAAGGCTTGGAAAGAATACATCGACTTCAAG CCTCAATTGAACAACGATCTATCATACAAGCAAACC AAAGATGTTACGCTTACTTCTCTTCATCTTTGTACAAT GTTgagctcCTCGAGTTTTTTGTAGAAATGTCTTGGTGTC CTCGTC	P14354
stgRNA-Stuffer_fw	CTCGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATA A	P14482
stgRNA-stuffer_rv	CCAAGACATTTCTACAAAAAACTCGAGgag	P14483
***P. pastoris* codon optimized Cas9**		
Cas9-1-1	CCCTATTTCAATCAATTGAACAACTATCAAAACACAG AATTCCGAAACGATGGACAAGAAGTATTCTATCGGA CTTGACATTGGAACTAACTCAGTGGGTTGGGCTGTGA TCACCGATGACTACAAGGTTCCATCTAAGAAGTTGAA GGTCTTGGGAAACACCGACAGACACGGAATCAAGAA GAACCTGATTGGTGCTTTGCTTTTCGATTCTGGAGAA ACCGCAGAGGCTACTCGTTTGAAAAGAACTGCCAGA AGAAGATACACTCGTCGTAAGAACAGAATCTGTTAC TTACAAGAAATCTTCTCTAATGAAATGGCCAAGGTTG ACGATTCTTTCTTCCACAGATTGGAGGAATCCTTCTT GGTTGAGGAAGATAAGAAGCACGAGAGACACCCAAT CTTCGGTAACATCGTGGACGAGGTTGCCTACCACGAG AAGTACCCAACTATCTACCACTTGAGAAAGAAACTT GCTGATTCTACTGATAAGGTCGATTTGAGATTGATCT ACTTGGCCTTGGCTCACATGATCAAGTTCCGTGGTCA CTTCCTGATTGAAGGTGACCTTAACCCTGACAACTCT GATGTTGACAAGTTGTTCATTCAATTGGTCCAAACTT ACAACCAGCTTTTCGAAGAAAACCCTATCAACGCATC TCGTGTGGACGCTAAGGCAATCCTTTCCGCTAGATTG TCCAAGTCCCGTCGTTTGGAAAACTTGATCGCTCAGC TTCCAGGTGAGAAGAAGAACGGTTTGTTTGGTAACTT GATTGCCCTGTCCTTGGGTCTTACTCCAAACTTCAAA TC	P14452

Cas9-1-2	CCTTGGGTCTTACTCCAAACTTCAAATCCAACTTCGA CCTTGCCGAAGACGCCAAGTTGCAGTTGTCTAAGGAT ACCTACGATGATGACCTGGACAACTTGCTTGCCCAAA TCGGTGACCAATACGCTGACCTTTTCCTTGCTGCTAA GAACTTGTCCGACGCCATTCTGTTGTCCGACATTCTT AGAGTGAACTCTGAGATTACTAAGGCTCCATTGTCAG CTTCAATGATCAAGAGATACGATGAACACCACCAAG ACTTGACTTTGTTGAAAGCATTAGTCAGACAGCAACT TCCAGAAAAGTACAAAGAAATCTTCTTCGATCAATCA AAGAACGGTTATGCTGGTTACATCGACGGTGGTGCCT CTCAAGAGGAGTTCTACAAGTTCATCAAGCCAATCTT AGAAAAGATGGACGGTACTGAGGAGCTGCTGGCCAA GTTGAATAGAGAAGACTTGTTGCGTAAGCAGAGAAC TTTCGACAACGGTTCTATTCCTTACCAGATCCACCTT GGAGAGTTGCATGCTATCCTTAGAAGACAAGAAGAC TTCTACCCATTCCTGAAGGACAACAGAGAGAAGATC GAGAAGATTCTTACCTTTAGAATCCCATACTACGTCG GTCCATTAGCAAGAGGAAACTCTCGTTTCGCTTGGAT GACCAGAAAGTCAGAGGAGACTATCACTCCTTGGAA CTTCGAGGAGG	P14453
gBlock Cas9-2	ACTCCTTGGAACTTCGAGGAGGTCGTGGACAAGGGT GCTTCCGCACAATCTTTCATCGAGAGAATGACTAACT TTGACAAGAACCTTCCAAACGAAAAGGTGCTTCCAA AGCACTCCTTGTTGTATGAATACTTCACCGTTTACAA TGAGTTGACCAAAGTTAAGTACGTTACCGAGGGTAT GAGAAAGCCTGCATTTCTGTCAGGAGAACAAAAGAA AGCCATTGTTGATTTGTTGTTCAAGACCAATAGAAAG GTCACTGTTAAGCAGCTTAAGGAGGATTACTTCAAGA AGATTGAGTGCTTCGATTCTGTCGAAATCTCTGGTGT TGAGGATAGATTCAACACCTCTTTGGGTACTTATCAC GACTTGTTGAAGATCATCAAGGACAAGGACTTCTTGG ATAACGAAGAGAACGAAGACATCTTGGAGGACATCG TTCTGACTTTGACTTTGTTTGAAGACAAAGAAATGAT TGAGGAGCGTTTGAAGAAGTATGCAAATCTTTTCGAC GACAAGGTTATGAAGCAGTTGAAGCGTCGTCACTAC ACCGGTTGGGGTAGATTGTCTCGTAAGTTGATCAATG GAATCCGTGACAAGCAGTCTGGTAAGACTATCTTAG ACTTCTTGAAGTCTGACGGTTTCGCTAACAGAAACTT CATGCAATTGATCAACGACGATTCCCTGACCTTCAAA GAAGCAATCCAGAAGGCTCAAGTTTCTGGACAAGGT CACTCTTTGCACGAACAAATCGCTAACTTGGCTGGTT CTCCAGCAATCAAGAAGGGTATTCTTCAAACCGTTAA GATCGTTGACGAGCTTGTTAAGGTTATGGGTCATAAG CCTGAGAACATTGTTATCGAAATGGCTAGAGAAAAC CAAACTACTCAGAAAGGTCAGAAGAACTCCAGAGAG AGAATGAAGAGAATCGAGGAGGGAATCAAAGAATT GGGTTCACAAATCTTGAAGGAGCATCCAGTCGAAAA CAC!CCAATTGCA	P14352

gBlock Cas9-3	GGAGCATCCAGTCGAAAACACCCAATTGCAAAACGA GAAGTTGTACCTTTACTACTTGCAAAACGGTAGAGAC ATGTACGTTGACCAAGAATTGGACATCAACAGACTG TCCGATTACGATGTTGATCACATTGTCCCACAATCTT TCTTGAAGGACGATTCCATCGACAACAAGGTCCTGAC TAGATCAGACAAGAACCGTGGTAAGTCTGATAACGT CCCTTCTGAGGAAGTTGTTAAGAAGATGAAGAACTA CTGGAGACAATTGTTGAACGCAAAACTTATCACCCA AAGAAAGTTTGATAACTTGACCAAGGCTGAGAGAGG TGGATTGTCTGAGTTAGACAAGGCTGGTTTCATCAAG CGTCAGTTGGTCGAGACTAGACAAATCACTAAACAC GTGGCTCAGATCCTTGATTCCAGAATGAACACCAAGT ACGATGAAAACGATAAGTTGATCCGTGAAGTTAAGG TGATCACCCTTAAGTCAAAACTTGTGTCAGACTTCAG AAAGGATTTCCAGTTCTACAAGGTTAGAGAGATCAA CAACTACCATCATGCCCACGATGCCTACCTTAACGCT GTCGTTGGTACTGCTCTTATCAAGAAGTACCCTAAGT TAGAATCTGAGTTTGTCTACGGTGACTACAAGGTTTA CGACGTTAGAAAGATGATCGCCAAGTCCGAACAAGA AATCGGAAAGGCCACTGCTAAGTACTTCTTCTATTCC AACATCATGAACTTCTTCAAGACTGAGATTACTTTGG CCAACGGAGAGATCAGAAAACGTCCTTTGATTGAAA CTAACGGTGAAACCGGTGAGATTGTCTGGGACAAGG GTAGAGACTTCGCCACCGTTCGTAAGGTCTTGTCTAT GCCACAAGTCAACATTGTCAAGAAGACCGAAGTTCA GACCGGTGGTTTCTCCAAAGAATCAATCCTGCCTAAG CGTAACTCTGACAAGTTGATTGCTCGTAAGAAGGATT GGG!ACCCAAAGA	P14353
gBlock Cas9-4	ATTGCTCGTAAGAAGGATTGGGACCCAAAGAAGTAC GGTGGATTTGATTCACCTACCGTTGCTTACTCTGTGTT AGTCGTTGCTAAGGTGGAGAAGGGTAAGTCTAAGAA GTTGAAGTCTGTGAAGGAATTGTTGGGTATCACTATC ATGGAAAGATCCTCTTTCGAAAAGAACCCAATTGACT TCCTTGAGGCCAAGGGATACAAGGAAGTCCGTAAAG ATTTGATCATCAAGTTGCCAAAGTATTCTTTGTTTGA GTTGGAAAACGGTAGAAAGAGAATGTTAGCCTCTGC TGGTGAATTGCAAAAGGGTAATGAGCTGGCCTTGCC ATCCAAGTACGTCAACTTTCTGTACTTGGCTTCCCAC TACGAGAAGCTGAAGGGTTCCCCAGAGGACAACGAA CAAAAGCAATTGTTCGTCGAGCAGCATAAGCATTACT TGGATGAAATCATTGAACAGATTTCTGAGTTCTCCAA GCGTGTCATTCTTGCTGACGCTAACCTTGATAAGGTC TTATCTGCCTACAACAAGCACAGAGACAAACCTATC AGAGAACAGGCTAAGAACATCATCCACTTGTTCACC CTTACCAACTTGGGTGCACCAGCTGCTTTCAAGTACT TTGACACCACTATTGAGAGAAACAGATACAAATCTA TCAAAGAAGTTCTGGACGCCACTCTTATCCACCAATC CATTACTGGTTTGTACGAAACTCGTATTGACTTGTCT CAACTTGGTGGAGATCCAAAGAAGAAAAGAAAAGTT TAAGCGGCCGCTCAAGAGGATGTCAGAATGCCATTT GCCTG	P14355
Cas9.1-fw	CCCTATTTCAATCAATTGAACAACTATCAAAACACAG	P14377

Cas9.2-rv	TGCAATTGGGTGTTTTCGACTGG	P14378
Cas9.3-fw	GGAGCATCCAGTCGAAAACACC	P14379
Cas9.4-rv	CAGGCAAATGGCATTCTGACATCC	P14380
Cas9-1fw-new	CAATTGAACAACTATCAAAACACAGAATT	P14400
Cas9-1rv	GTCCACGACCTCCTCGAAG	P14401
Cas9-2fw	ACTCCTTGGAACTTCGAGGAGG	P14402
Cas9-1-1-rv	GATTTGAAGTTTGGAGTAAGACCCAAGG	P14459
Cas9-1-2-fw	CCTTGGGTCTTACTCCAAACTTCAAATC	P14460

CRISPR-Cas9 constructs with RNA Polymerase II promoter and ribozymes

pHTX1-DAS1TT-Gibson	CTCAAACTATATTAAAACTACAACAGAATTCACGGG AAGTCTTTACAGTTTTAGTTAGG	P14784
DAS1TT-pHXT1-fw	AATTCTGTTGTAGTTTTAATATAGTTTGAGTATGAGA TGG	P14785
AOXTT-HTX1-rv-Gibson	CTGACATCCTCTTGAGCGGCCGCTTTGATTTGTTTAG GTAACTTGAACTGGATGTATTAG	P14786
DAS1TT-Cas9-Gibson	CTAAAACTGTAAAGACTTCCCGTCTCGAGTTAAACTT TTCTTTTCTTCTTTGGATCTCCA	P14787
pHTX1-Cas9-Gibson	CTCAAACTATATTAAAACTACAACAGAATTCCGAAA CGATGGACAAGAAGTATTCTATCG	P14788
DAS1TT-spCas9-Gibson	TAAAACTGTAAAGACTTCCCGTCTCGAGTTAAACTTT TCTTTTCTTCTTTGGGTCACCTC	P14831
HTX1-spCas9-Gibson	CTATATTAAAACTACAACAGAATTCCGAAACGATGG ATAAGAAATACTCAATAGGCTTAG	P14832
DAS1TT-hsCas9-Gibson	TAAAACTGTAAAGACTTCCCGTCTCGAGTTAAACTTT TCTTTTCTTCTTTGGGTCTCCAC	P14833
HTX1-hsCas9-Gibson	CTCAAACTATATTAAAACTACAACAGAATTCCGAAA CGATGGACAAGAAGTACTCC	P14834
RZ-GUT1-gRNA2-RZ	CCAGTTCAAGTTACCTAAACAAATCAAATACTCGCTG ATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCCG AGTACTCTACCTCTGCTCGTTTTAGAGCTAGAAATAG CAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGCACCGAGTCGGTGCTTTTGGCCGGCATGGTC CCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTT CGGCATGGCGAATGGGACTCAAGAGGATGTCAGAAT GCC	P14835

RZ-GUT1-gRNA3-RZ	CCAGTTCAAGTTACCTAAACAAATCAAAATTGCACTG ATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCTG CAATTTCCTCAGCCAGGCGTTTTAGAGCTAGAAATAG CAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGCACCGAGTCGGTGCTTTTGGCCGGCATGGTC CCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTT CGGCATGGCGAATGGGACTCAAGAGGATGTCAGAAT GCC	P14836
RZ-GUT1-gRNA4-RZ	CCAGTTCAAGTTACCTAAACAAATCAAAAACAACCT GATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCG TTGTTTGGTCCAAGAAGACGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAA AAAGTGGCACCGAGTCGGTGCTTTTGGCCGGCATGGT CCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCT TCGGCATGGCGAATGGGACTCAAGAGGATGTCAGAA TGCC	P14837
Tagged Cas9		
GAP-GFP-fw-Gibson	ATCAATTGAACAACTATCAAAACACAATGGCTAGCA AAGGAGAAGAACTTTTC	P14789
Cas9-GFP-rv-Gibson2	CGATAGAATACTTCTTGTCCATCGTTTCCTTGTACAAT TCATCCATGCCATGTGTAATC	P14838
His-Cas9-fw	GAACAACTATCAAAACACAGAATTCCGAAACGATGC ATCATCATCATCATCATGACAAGAAGTATTCTATCGG ACTTGAC	P14830
***Sp*Cas9 and *Hs*Cas9**		
spCas9_fw	GAACAACTATCAAAACACAGAATTCCGAAACGATGG ATAAGAAATACTCAATAGGCTTAG	P14683
spCas9-SV40_rv	TCCTCTTGAGCGGCCGCTTAAACTTTTCTTTTCTTCTT TGGGTCACCTCCTAGCTGACTC	P14684
hsCas9_fw	GAACAACTATCAAAACACAGAATTCCGAAACGATGG ACAAGAAGTACTCCATTGGG	P14685
hsCas9-SV40_rv	TCCTCTTGAGCGGCCGCTTAAACTTTTCTTTTCTTCTT TGGGTCTCCACCGAGCTGAGAG	P14686
***GUT1* PCR and sequencing**		
seq-pGUT1-332..308-fwd	tgggttcaatggcgtttgagttag	P14332
3UTRGUTR	GTGTTTGCTGTAGGATGACCTAGATTTAAATATAAGA GGAAACAACGTTCGTATCGTGA	P14252
PGUTseq2	GGTACTTTGCCGACTCCTC	P09523
GUTout3prR1	ATGAAGTTAGTAAGGTTCTTGATGAAGC	P10040

pGUT1fwd	GAAAAGGTTTACTATCCCGATTTAGGCGAAAAGAG	P12485
TALEN		
5GUT1-TALEN-fw-new	ctaactagagaacccactgcttactgg	P14903
3GUT1-TALEN-fw-new	Acacattccacagaattaattcgcg	P14904
GUT1-TALEN-fw-new	Atggccccaaagaagaagcg	P14901
GUT1-TALEN-rv-new	Ttatgagcggaaattgatctcgcc	P14902
TALENseq1	gcttcgcgatcttcagcaactgg	P14867
TALENseq2	gaggacatcacatcgagtggca	P14868
Additional CRISPR-Cas9 targets		
RZ-AOX1-gRNA1-RZ	CCAGTTCAAGTTACCTAAACAAATCAAAATCAAACT GATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCT TTGATATCCTAGTTCTAGGGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAA AAAGTGGCACCGAGTCGGTGCTTTTGGCCGGCATGGT CCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCT TCGGCATGGCGAATGGGACTCAAGAGGATGTCAGAA TGCC	P15076
RZ-AOX1-gRNA2-RZ	CCAGTTCAAGTTACCTAAACAAATCAAAAGTCATCTG ATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCAT GACTTCCAAGCCGAGGGCGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAA AAAGTGGCACCGAGTCGGTGCTTTTGGCCGGCATGGT CCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCT TCGGCATGGCGAATGGGACTCAAGAGGATGTCAGAA TGCC	P15077
RZ-AOX1-gRNA3-RZ	CCAGTTCAAGTTACCTAAACAAATCAAAGTGTAGCTG ATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCCT ACACCTACCCAGTTTGCCGTTTTAGAGCTAGAAATAG CAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGCACCGAGTCGGTGCTTTTGGCCGGCATGGTC CCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTT CGGCATGGCGAATGGGACTCAAGAGGATGTCAGAAT GCC	P15078
RZ-MXR1-gRNA1-RZ	CCAGTTCAAGTTACCTAAACAAATCAAAATTGAACTG ATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCTT CAATGGGACCACACCCTCGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAA AAAGTGGCACCGAGTCGGTGCTTTTGGCCGGCATGGT CCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCT TCGGCATGGCGAATGGGACTCAAGAGGATGTCAGAA TGCC	P15079

RZ-MXR1-gRNA2-RZ	CCAGTTCAAGTTACCTAAACAAATCAAACTGTGCCTG ATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCGC ACAGCAACTGCTCTGATGGTTTTAGAGCTAGAAATAG CAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGCACCGAGTCGGTGCTTTTGGCCGGCATGGTC CCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTT CGGCATGGCGAATGGGACTCAAGAGGATGTCAGAAT GCC	P15080
RZ-MXR1-gRNA3-RZ	CCAGTTCAAGTTACCTAAACAAATCAAAATTCTACTG ATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCTA GAATCCCTATTTTCTGTCGTTTTAGAGCTAGAAATAG CAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGCACCGAGTCGGTGCTTTTGGCCGGCATGGTC CCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTT CGGCATGGCGAATGGGACTCAAGAGGATGTCAGAAT GCC	P15081
RZ-MPP1-gRNA1-RZ	CCAGTTCAAGTTACCTAAACAAATCAAAAAATGGCT GATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCC CATTTGACCCTAATGAACAGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAA AAAGTGGCACCGAGTCGGTGCTTTTGGCCGGCATGGT CCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCT TCGGCATGGCGAATGGGACTCAAGAGGATGTCAGAA TGCC	P15082
RZ-MPP1-gRNA1-RZ	CCAGTTCAAGTTACCTAAACAAATCAAAGAAAATCT GATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCA TTTTCAGTAGACGCTGACGGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAA AAAGTGGCACCGAGTCGGTGCTTTTGGCCGGCATGGT CCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCT TCGGCATGGCGAATGGGACTCAAGAGGATGTCAGAA TGCC	P15083
RZ-MPP1-gRNA3-RZ	CCAGTTCAAGTTACCTAAACAAATCAAATATGAACTG ATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCTT CATACCCATCTGGGCCGGGTTTTAGAGCTAGAAATAG CAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGCACCGAGTCGGTGCTTTTGGCCGGCATGGTC CCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTT CGGCATGGCGAATGGGACTCAAGAGGATGTCAGAAT GCC	P15084
RZ-PRM1-gRNA1-RZ	CCAGTTCAAGTTACCTAAACAAATCAAAGAAGGTCT GATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCA CCTTCCAGTAAGAAGAGACGTTTTAGAGCTAGAAAT AGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGA AAAAGTGGCACCGAGTCGGTGCTTTTGGCCGGCATG GTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATG CTTCGGCATGGCGAATGGGACTCAAGAGGATGTCAG AATGCC	P15085

RZ-PRM1-gRNA2-RZ	CCAGTTCAAGTTACCTAAACAAATCAAACTATGGCTG ATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCCC ATAGAATCTAATCTAGGGGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAA AAAGTGGCACCGAGTCGGTGCTTTTGGCCGGCATGGT CCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCT TCGGCATGGCGAATGGGACTCAAGAGGATGTCAGAA TGCC	P15086
RZ-PRM1-gRNA3-RZ	CCAGTTCAAGTTACCTAAACAAATCAAAATATACCTG ATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCGT ATATTCGGTGATGGCTTGGTTTTAGAGCTAGAAATAG CAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGCACCGAGTCGGTGCTTTTGGCCGGCATGGTC CCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTT CGGCATGGCGAATGGGACTCAAGAGGATGTCAGAAT GCC	P15087
Multiplexing		
pAOXsyn-pHHF2-fw-Gibson	GGAACACTGAAAAATACACAGTTATTATTCCGATTTG GCACTTTTTGCCATCAG	P15005
HH-AOX1-gRNA1-pHHF2-rv-Gibson	TCACGGACTCATCAGTTTGATATTTATTGATTATTTGT TTATGGGTGAGTCTAGAAAAGG	P15006
HH-AOX1-gRNA2-pHHF2-rv-Gibson	TCACGGACTCATCAGATGACTATTTATTGATTATTTG TTTATGGGTGAGTCTAGAAAAGG	P15007
HH-AOX1-gRNA3-pHHF2-rv-Gibson	TCACGGACTCATCAGCTACACATTTATTGATTATTTG TTTATGGGTGAGTCTAGAAAAGG	P15008
pHHF2-Rz-AOX1-gRNA1-Rz-DAS1TT	ATCAAACTGATGAGTCCGTGAGGACGAAACGAGTAA GCTCGTCTTTGATATCCTAGTTCTAGGGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATC AACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTGGCC GGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGC AACATGCTTCGGCATGGCGAATGGGACGTAGATTTG GCCACTAACGGGTTAGTAGTTGTGTAAGTCTATTAAA TTTGATTTTTGTTTATGGATGATCATCGTAGTGGCTAT CTGTTTACCTGTAGGACATCCTAGGGTGGGATGGTGA TGTACACCCCCTCAATCTTCAGATGCAACACTATGTG GTAGGTCATTGACATAAGGTTTAGGAAAGACCTGTTT TTTGACCAATAAATGGAACAGGAAGGAAAGGAGGAA CCAGTTTACGAACCCCGTCGACCCTTGTGACTGACAC TTTG	P15009

pHHF2-Rz-AOX1-gRNA2-Rz-DAS1TT	AGTCATCTGATGAGTCCGTGAGGACGAAACGAGTAA GCTCGTCATGACTTCCAAGCCGAGGGCGTTTTAGAGC TAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTAT CAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTGGC CGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGG CAACATGCTTCGGCATGGCGAATGGGACGTAGATTT GGCCACTAACGGGTTAGTAGTTGTGTAAGTCTATTAA ATTTGATTTTTGTTTATGGATGATCATCGTAGTGGCTA TCTGTTTACCTGTAGGACATCCTAGGGTGGGATGGTG ATGTACACCCCCTCAATCTTCAGATGCAACACTATGT GGTAGGTCATTGACATAAGGTTTAGGAAAGACCTGTT TTTTGACCAATAAATGGAACAGGAAGGAAAGGAGGA ACCAGTTTACGAACCCCGTCGACCCTTGTGACTGACA CTTTG	P15010
pHHF2-Rz-AOX1-gRNA3-Rz-DAS1TT	GTGTAGCTGATGAGTCCGTGAGGACGAAACGAGTAA GCTCGTCCTACACCTACCCAGTTTGCCGTTTTAGAGC TAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTAT CAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTGGC CGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGG CAACATGCTTCGGCATGGCGAATGGGACGTAGATTT GGCCACTAACGGGTTAGTAGTTGTGTAAGTCTATTAA ATTTGATTTTTGTTTATGGATGATCATCGTAGTGGCTA TCTGTTTACCTGTAGGACATCCTAGGGTGGGATGGTG ATGTACACCCCCTCAATCTTCAGATGCAACACTATGT GGTAGGTCATTGACATAAGGTTTAGGAAAGACCTGTT TTTTGACCAATAAATGGAACAGGAAGGAAAGGAGGA ACCAGTTTACGAACCCCGTCGACCCTTGTGACTGACA CTTTG	P15011