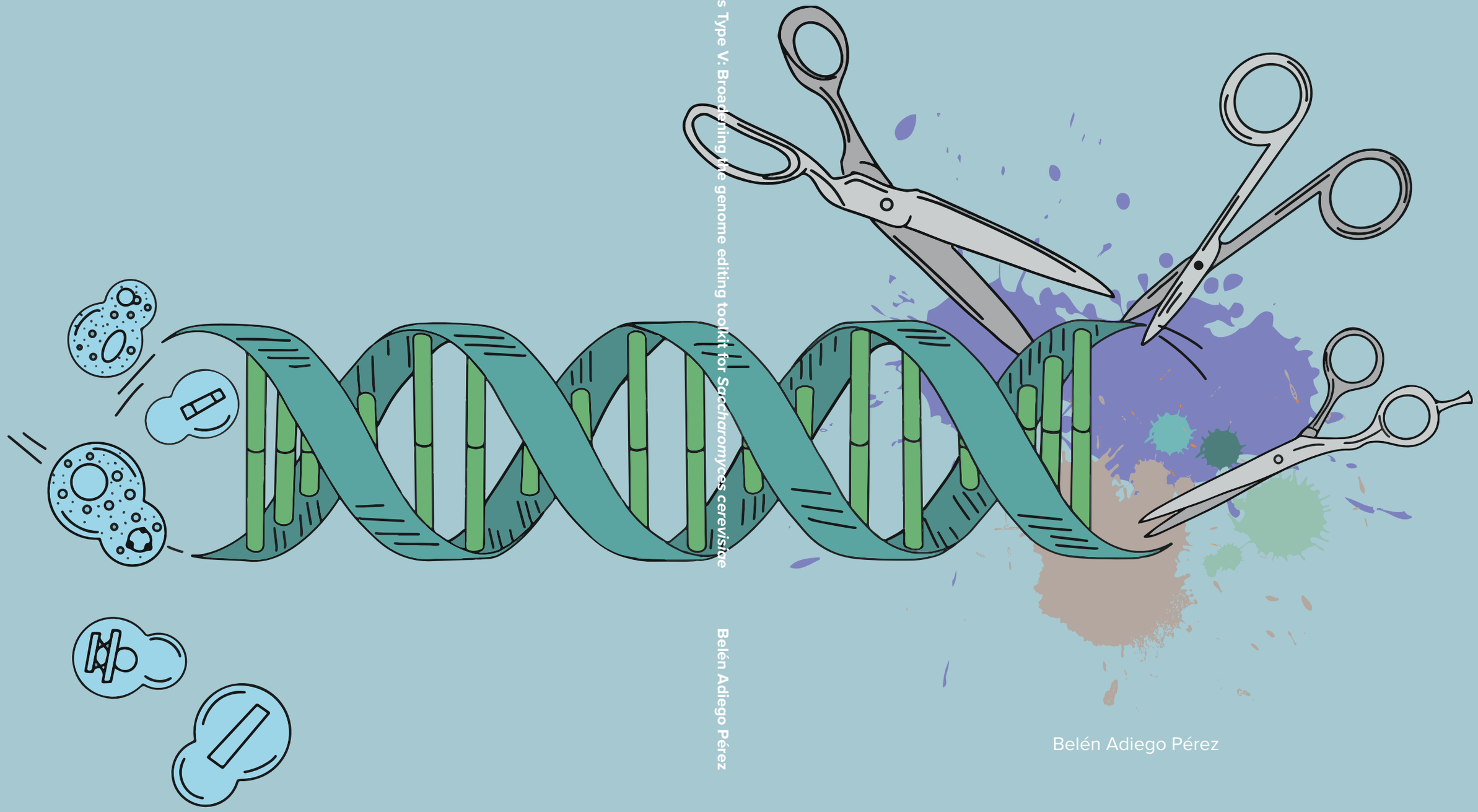
The colors of CRISPR-Cas Type V:

Broadening the genome editing toolkit for *Saccharomyces cerevisiae*



Belén Adiego Pérez

The colors of CRISPR-Cas Type V: Broadening the genome editing toolkit for *Saccharomyces cerevisiae*

Belén Adiego Pérez

Propositions

1. Diversity of type V CRISPR-Cas systems emphasizes the requirement of a single and systematic nomenclature system for designation of effectors. (This thesis)

2. *Saccharomyces cerevisiae* should not be considered a model yeast for development of CRISPR-Cas tools. (This thesis)

3. The H-index system does not stimulate the peer-reviewed publication of negative results.

4. Governmental obligation of open access publication increases the power that publishing entities have over the publication process.

5. The terms “health” and “mental health” are prejudicially analysed separately.

6. Unnecessary flying will not be avoided until train travelling becomes affordable.

Propositions belonging to the thesis entitled:

The colors of CRISPR-Cas Type V:

Broadening the genome editing toolkit for *Saccharomyces cerevisiae*

Belén Adiego Pérez

Wageningen, 13th September 2022

The colors of CRISPR-Cas Type V:

Broadening the genome editing toolkit for *Saccharomyces cerevisiae*

Belén Adiego Pérez

The colors of CRISPR-Cas Type V:

Broadening the genome editing toolkit for *Saccharomyces cerevisiae*

Belén Adiego Pérez

Thesis Committee

Promotor
Prof. Dr J. van der Oost
Wageningen University & Research

Co-promotor
Dr S. D'Adamo
Wageningen University & Research

Other members
Dr D.C. Swarts, Wageningen University & Research
Dr J.G. Schaart, Wageningen University & Research
Dr B. Vonk, DSM
Prof. Dr M. Kleerebezem, Wageningen University & Research

This research was conducted under the auspices of the Graduate School VLAG (Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences)

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr. A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 13 September 2022
at 1:30 p.m. in the Omnia Auditorium.

TABLE OF CONTENTS

Chapter 1	Introduction and thesis outline	9
Chapter 2	Multiplex Genome Editing of Microorganisms Using CRISPR-Cas	25
Chapter 3	The miniature CRISPR-Cas12m effector binds DNA to block transcription	55
Chapter 4	Exploring the potential of *Mm*Cas12m as gene silencing tool in *Saccharomyces cerevisiae*	103
Chapter 5	Exploring *Mm*Cas12m potential as a base editing tool in *Saccharomyces cerevisiae*	127
Chapter 6	Type V-K CRISPR-associated transposon system – towards its application in eukaryotic organisms	149
Chapter 7	Adaptation of SIBR-Cas for its use in yeast	171
Chapter 8	Thesis summary and general discussion	191
	Thesis summary	192
	General discussion	194
APPENDICES		213

Belén Adiego Pérez
The colors of CRISPR-Cas Type V: Broadening the genome editing toolkit for *Saccharomyces cerevisiae*,
250 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2022)
With references, with summary in English

ISBN: 978-94-6447-324-7
DOI: https://doi.org/10.18174/574280

ABBREVIATIONS

AAV	Adeno-associated virus
CDS	Coding sequence
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
dDNA	Donor DNA
DSB	Double Strand Break
gRNA	guide RNA
HDR	Homology-directed repair
HR	Homologous recombination
MGE	Mobile genetic elements
NLS	Nuclear localization signal
NMD	Non-sense mediated decay
NT	Non-targeting
ORF	Open reading frame
PAM	Protospacer adjacent motif
PCR	Polymerase chain reaction
pre-crRNA	pre-CRISPR RNA
PTC	Premature termination codon
RBS	Ribosome binding site
RNP	Ribonucleoprotein
sgRNA	single-guide RNA
SIBR	Self-splicing Intron-Based Riboswitch
T	Targeting
TALE	Transcription activator-like effectors
TALEN	Transcription activator-like effector nuclease
ZF	Zinc Finger
ZFN	Zinc Finger nuclease

CHAPTER 1

Introduction and thesis outline

1.1 A historical approach to biotechnology

Human beings have unknowingly benefited from the use of biotechnology since prehistorical times. The use of microorganisms, and specifically yeast, to produce fermented products like beer or wine can be traced back to different historical moments and has been linked to the settling down of nomadic civilizations (*1*). For instance, there is archaeological evidence of cereal-based beer production in stone mortars from the Middle East (11,750 – 9,750 BC) (*2*), of rice beer production in East Asia (7,050 – 6,750 BC) (*3*), and of wine in the South Caucasus region (6,000–5,800 BC) (*4*). Neolithic agricultural civilizations considered fermentations to be spontaneous processes and their products, to be gifted by Gods (*2*, *5*).

In 1775, the term "yeast" was used for the first time to describe the "existing bodies in the liquid during fermentation, which float on the surface and later sink to the bottom" by Krünitz (*6*). Several authors contributed to describe the morphology of these bodies and their replicative and vegetative nature during the 18th and 19th centuries (*7*). However, it was not until 1857 when Louis Pasteur demonstrated the direct participation of microorganisms in fermentation processes and associated each fermentation to a specific organism. Improvement of fermentation processes was then possible by means of isolating the microorganisms in charge of each type of fermentation, by implementing sterile procedures and avoiding contamination. Isolation of yeast was not possible until 1883, when Emil Christian Hansen isolated the yeast responsible for beer production, *Saccharomyces carlsbergensis* (*8*). The publication of the method he used allowed for the establishment of an isolation method for other yeasts. This eventually resulted in the achievement of a pure culture of *Saccharomyces cerevisiae* (also known as baker's yeast), undoubtedly the most widely used yeast for both fundamental studies and industrial applications (*9*). Throughout the years, an extended amount of knowledge was accumulated on the biology of *S. cerevisiae*, leading to considering this organism as a model yeast.

In 1917, the term biotechnology was introduced for the first time by the Hungarian engineer Karl Erkey (*10*). Biotechnology can be defined as "the use of living things, especially cells and bacteria, in industrial processes" or "the use of living organisms or their products to modify or improve human health and human environment" (*11*). Subsequent discoveries during the 20th century, such as penicillin by Alexander Fleming (1928), broadened the knowledge on the ability of microorganisms to produce useful products for human beings. Based on these discoveries, we nowadays benefit from natural producers such as *Aspergillus niger* for citric acid production (*12*), *S. cerevisiae* for bioethanol production (*13*, *14*) or *Corynebacterium glutamicum* for amino acid production (*15*).

1.2 The impact of genetics in industrial biotechnology

Natural producers of compounds of interest are widely used in the biotechnological industry because of their tolerance to the product of interest and their innate ability to produce it. However, production levels of these compounds are often limited by the nature of the producers or the ability to cultivate them in industrial settings. Advances in the genetics field during the 19th and 20th centuries have certainly boosted progress of biotechnological processes.

Contemporarily to the discoveries of Pasteur on the role of microorganisms in fermentation processes, in 1865, Gregory Mendel (1822-1884) established the existence of "factors of heredity" related to observable traits. Mendel also described how these factors, which we currently refer to as "genes", were passed from generation to generation (*16*). Sadly, his observations were shadowed until 1900, when three scientists independently re-discovered Mendel's laws and prepared the fundaments for the Mendelian community. Two years later, also based on several studies on the morphology of chromosomes conducted in the end of the 19th century, it was proposed that chromosomes could bear these Mendelian factors. However, the morphological nature of the described chromosomes and genes did not occur until the 1940s (*17*), when experimental observations by Oswald Avery and colleagues connected purified nucleic acids to the genes (*17*). In 1953, James Watson and Francis Crick, thanks to the essential work of Rosalind Franklin (*18*), finally resolved the structure of DNA (*19*). This remarkable event unleashed a series of discoveries in the field of molecular biology during the second half of the 20th century. Among them, the central dogma of molecular biology was described (*20*), the genetic code was deciphered (*21*) and the technology of recombinant DNA was born in the 1970s (*22*). All these initial discoveries were then substantially improved with the invention of PCR and other molecular biology techniques (*23*). Nowadays, automated DNA sequencing and improved computational tools allow the deciphering of complete genomes, helping the characterization of the metabolic networks required to produce interesting biotechnological compounds.

1.3 From natural producers to engineered cell factories

Undoubtedly, advances in genetics throughout the 19th and 20th centuries have allowed us to boost biotechnological production of industrial compounds. During this time, natural producers of biotechnological compounds have also been characterized to reveal the metabolic routes that support their production processes. For instance, *S. cerevisiae* was characterized as natural producer of ethanol through fermentation. Natural producers have also been subjected to iterative mutagenesis approaches (*24*) or adaptive laboratory evolution to increase production titers or tolerance to high product levels (*25*). In the case of *S. cerevisiae*, this has led to obtaining improved bioethanol producers which are also more tolerant to high temperatures or stress conditions (*26*). However, simultaneous improvement of multiple traits is a complex task, as improvement of certain traits often affects other interesting features.

Thanks to the acquired knowledge in the fields of metabolism, genetics and molecular biology, genetic modification of natural producers of biotechnological compounds was possible. The use of genetic engineering tools for the rational modification of the metabolism of organisms is defined as metabolic engineering. The main objective of this discipline is to produce specific metabolites with high yields (*27*). Implementation of iterative rational modifications, strain building, and phenotype characterization and assessment cycles resemble classical Design-Build-Test-Learn (DBTL) cycles. These DBTL cycles are frequently used in engineering disciplines, and now also in synthetic biology (*28*). In fact, metabolic engineering approaches are focused on the elimination of by-products, the broadening of substrate ranges (*29*) and

the engineering of stress tolerance strategies (*29–31*). At the same time, they also allow the incorporation of complete metabolic networks in heterologous hosts to expand the product range (*32*, *33*). In this way, besides its classical use in the production of alcoholic beverages, food or glycerol (*33*), *S. cerevisiae* is now being used for the production of bioethanol, chemical building blocks and pharmaceuticals (*34*).

Both for genetic improvement of natural producers and the use of non-native hosts for heterologous production of biotechnological compounds, incorporation of non-native DNA or modification of native DNA is required. The heterologous expression of non-native pathways could be sustained by using extra-chromosomal DNA (i.e., centromeric or episomal plasmids). However, chromosomal integration by homologous recombination (HR) is generally preferred in an industrial context because of the inherent instability of plasmid vectors, and to avoid the need for continuous selective pressure to maintain the plasmids (*33*). Efficient genome editing tools are thus required to modify the genome of biotechnological producers. These tools are essential to decrease the time in the building step of DBTL cycles. Moreover, they should allow the use of non-model organisms with interesting characteristics as cell factories (*35*).

1.4 Classical targeted genome modification in yeast

The modifications required for metabolic engineering approaches include: (i) the addition of new DNA fragments (knock-in), (ii) the modification of already existing genomic DNA sequences (mutations), or (iii) the deletion of complete genomic sequences (knockout). These approaches are used to re-direct metabolic fluxes, engineer co-factor dependency, avoid the formation of by-products, or implement heterologous metabolic pathways, enzymatic or transport activities (*36*, *37*).

Genome modification of *S. cerevisiae* and other yeasts is based on the activity of DNA repair pathways. Upon the event of a double strand break (DSB), two main different pathways can be activated to repair the damaged DNA: the homology direct repair (HDR) pathway and the non-homologous end-joining (NHEJ) pathway. The activity of the first pathway is exploited by the supply of synthetic repair fragments (homologous regions that flank the desired genetic edit) that will recombine and integrate the provided heterologous DNA, while NHEJ is an error-prone repair system that can lead to the incorporation of deletions or insertions at the repaired site (*38*, *39*). Ideally, integration constructs are directed to target locations in the genome based on the homologous recombination (HR) activity of the HDR pathway. In *S. cerevisiae*, the HDR pathway is highly active and DNA fragments can be recombined into targeted genomic locations which only carry 30-50 base pair (bp) homology regions. However, this high activity of the HDR pathway is not representative for other yeasts such as *Yarrowia lipolytica* or *Kluyveromyces marxianus* (*40*). In these organisms, low transformation efficiencies are the first limiting step in the achievement of genome-edited strains. Secondly, the high activity of the NHEJ pathway limits the HR recombination events by the HDR pathway (*41*, *42*). Therefore, targeted integration of DNA fragments in non-model yeasts requires the use of longer homologous regions, whose length can vary from several hundreds to several thousands of base-pairs (*43*). Finally, the high activity of the NHEJ pathway may result in

random integration of supplied DNA fragments. Therefore, these series of obstacles altogether hinder the screening process to obtain mutants with the desired integration products in several industrially-relevant yeasts (*44*). In some instances, random integration approaches of heterologous DNA (using NHEJ or transposons) have been used in metabolic engineering strategies (*40*). However, characterization of the integration event(s) and evaluation of the phenotype of the obtained strains are required to exclude any potential disruption of important genes, making the process laborious and therefore not convenient in strain building programs.

All the above-mentioned homologous recombination events require the use of selection markers in the integration constructs. The use of amino acid biosynthetic enzymes as markers is widely employed among auxotrophic laboratory strains, while it is not commonly used in industrial set-ups (*45*). In this context, dominant drug-resistance markers are used due to the prototroph nature of industrial strains. However, a limited number of drug-resistance markers are available for each species, resulting in the need of recycling markers between consecutive transformation rounds (*46*). Moreover, the use of drug-resistance markers and incorporation of heterologous DNA qualifies the obtained strains as genetically-modified organisms (GMOs), further limiting their use in specific industrial set-ups. For these reasons, scientists have developed alternative methods to generate clean or almost clean genome modified organisms and help in the elimination of selection markers from genomic locations (*44*). Some of the methods developed for *S. cerevisiae* include the use of the Cre-*lox* system, which leaves a small scar on the genome after use (*47*) or the use of the *delitto perfetto* technique (*48*), that allows to obtain clean knockouts (Figure 1.1). Still, these systems require multiple transformation and marker selection rounds or leave scars in the genome that can contribute to genome instability (*44*).

A way to increase targeted modification efficiencies is by introducing double strand breaks (DSBs) in the DNA in a targeted manner. DNA sequence-specific binding proteins, such as Zinc fingers (ZFs) or transcription activator-like effectors (TALEs), have been fused to the FokI endonuclease to be used as Zinc fingers nucleases (ZFNs) (*49*) or transcription activator-like effector nucleases (TALENs), (*50*, *51*). The FokI endonuclease works as a dimer and, therefore, a pair of ZFNs or TALENs are required to target opposite DNA strands of the same target site to introduce a DSB. ZFs recognize trinucleotide sequences in a specific and context-dependent manner, while TALENs can recognize single nucleotides (*52*, *53*). Therefore, by independent combination of multiple ZF or TALE domains fused to a FokI endonuclease, long DNA sequences can be targeted. This approach has been used to enhance homologous recombination in yeast upon the introduction of double strand breaks (DSBs) in a targeted manner (*54*) and can be used to generate clean knockouts. However, programming these tools to cleave new targets requires protein design and subsequent laborious genetic engineering to adjust the specificity of the DNA binding proteins, which limits the wide applicability of these tools. In the last decade, a new group of tools have been developed based on CRISPR-Cas systems, which has revolutionized genome editing.

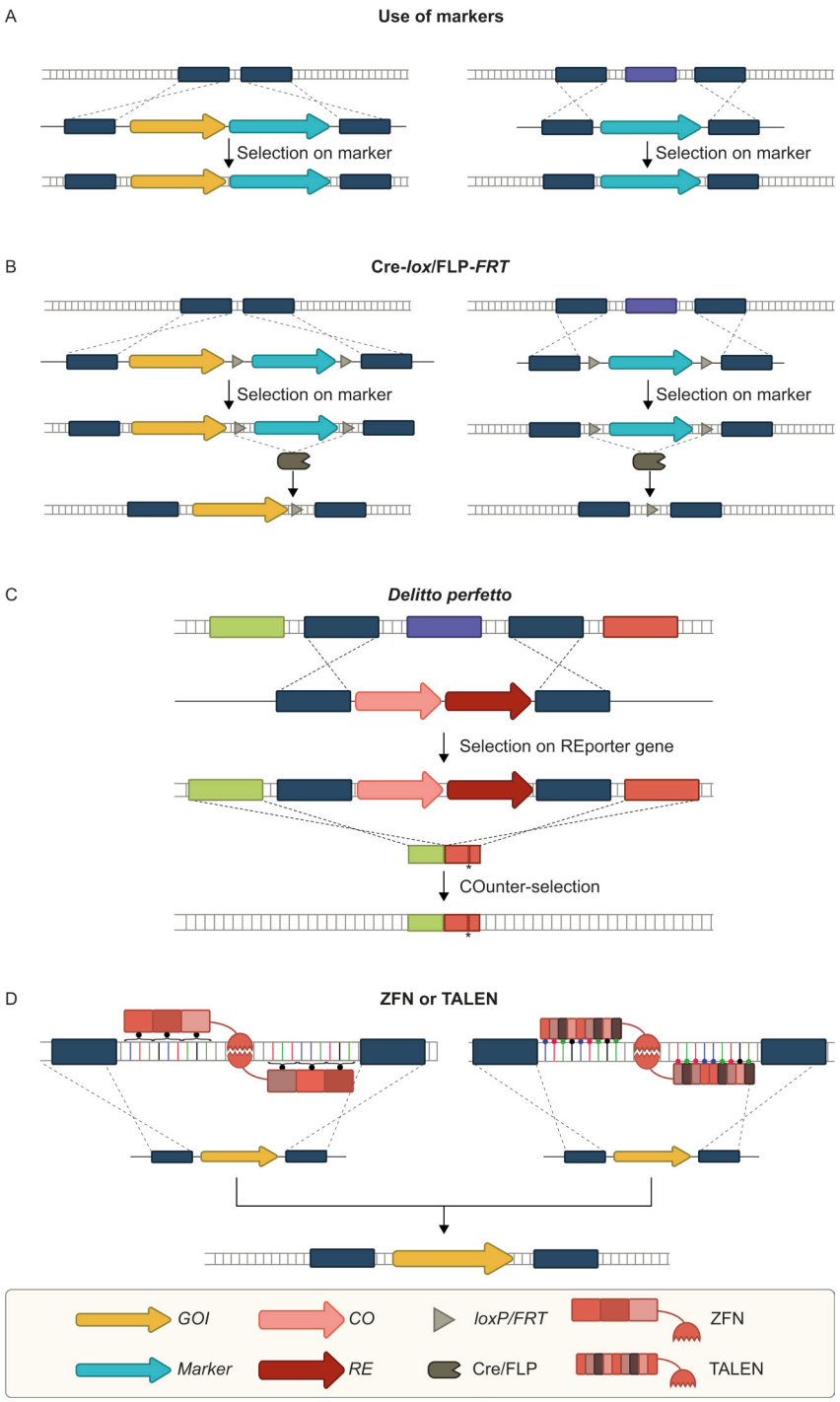


Figure 1.1. Targeted genome editing strategies used in *Saccharomyces cerevisiae* and other non-model yeasts for deletion or integration of DNA fragments. (A) Use of markers. This strategy is used for both integration of foreign DNA fragments (left) or deletion of genome sections (right). Heterologous fragments containing the gene of interest (GOI) are provided with homologous regions to the DNA. The size of these homologous regions varies depending on the organism, from 50 bp in *S. cerevisiae* to several hundreds or thousands of base pairs in some non-model yeasts. Heterologous DNA contains a marker for which selection is applied. **(B)** Cre-*lox* or FLP/*FRT*. This strategy can be used for both integration of foreign DNA fragments (left) or deletion of genome regions (right). Heterologous DNA is integrated by homologous recombination and selected through cultivation on selective conditions. Expression of a recombinase (Cre or Flippase (FLP)) allows recombination between short recognition sites (*loxP* or *FRT*, respectively). After recombination, a small scar remains in the modification site. **(C)** *Delitto perfetto* technique. This method can be used for the deletion of genomic sequences and introduction of point mutations. DNA is first incorporated by homologous recombination into the target site. This DNA harbors a reporter (*RE*) and a counter selection gene (*CO*). First, recombination individuals are grown under selective conditions for expression of the reporter (*RE*) gene. Then, deletion of the targeted genomic site by recombination of small DNA fragments is promoted by cultivation in counter-selection conditions. Recombination fragments can contain single point mutations (symbolized with " * "). **(D)** Use of zinc finger nucleases (ZFNs) or transcription activator-like effector nucleases (TALENs). ZFs recognize nucleotide triplets and fusion of multiple ZFs allows for recognition of longer DNA sequences. On the other hand, TALEs recognize single nucleotides and fusion of multiple TALEs allows for recognition of longer DNA sequences. Addition of a FokI nuclease domain to ZFs or TALEs forms ZFNs and TALENs, respectively. Dimerization of two FokI domains is required for introduction of double strand breaks (DSBs) in the DNA. Therefore, a pair of ZFNs or TALENs are required to target complementary DNA strands for the introduction of DSBs. DSBs are then repaired by classical DNA repair pathways: homology directed repair (HDR) or non-homologous end joining (NHEJ). The activity of the HDR pathway is required for integration of homologous DNA fragments following this strategy.<<

1.5 The CRISPR-Cas era

CRISPR-Cas systems are adaptive immune systems that protect archaea and bacteria from invading viruses, plasmids or other types of mobile genetic elements (MGEs) (*55*, *56*). These systems consist of clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins. These two elements are often located adjacent to each other.

The first evidences pointing to the use of CRISPR-systems for adaptive immunity against foreign genetic elements were inferred from the identification of viral sequences in CRISPR spacers by bioinformatic analyses (*57–60*). Later, through phage-infection experiments, it was proven that acquisition of fragments of the invading DNA into the CRISPR array confers resistance to the invading phage (*55*). CRISPR-Cas systems are able to store information about previous infection events in their CRISPR arrays, by means of the incorporation of short DNA sequences originating from the invader elements (called "spacers") between repeated and often palindromic sequences (the "repeats") (*55*, *61*). Thus, CRISPR-Cas systems function as adaptive immune systems over the course of three stages: adaptation, expression, and interference (*62–65*).

In the adaptation stage, Cas proteins identify the invading DNA and incorporate new invader-derived sequences into the CRISPR array. Cas1 and Cas2 are the main representative proteins

for this function in CRISPR-Cas systems and are present in almost all CRISPR-Cas systems (*66*). CRISPR arrays are preceded by leader sequences that contain motifs for the adaptation stage. Integration of new spacers into the CRISPR array derives from a duplication event of the leader-proximal repeat, which confers a directionality to the array (*66*). Moreover, these leader sequences also contain promoters that drive transcription of the CRISPR array into a long precursor CRISPR RNA (pre-crRNA) during the expression stage (*67*). This pre-crRNA is then processed into mature short crRNAs with the aid of other host or dedicated Cas proteins with RNase activity (*68–70*). In some cases, transactivating CRISPR RNAs (tracrRNA) (*71*, *72*) or short-complementarity untranslated RNAs (scoutRNA) (*73*) are also required in the maturation step. Each mature crRNA bears a spacer sequence that guides the sequence-recognition antiviral response (*68*) and a part of the CRISPR repeat.

In the interference stage, the crRNA (bearing the spacer sequence) combines with one or multiple Cas proteins to form active Cas-crRNA complexes. Sometimes, tracrRNA or scoutRNA sequences are also required for the targeting activity. These ribonucleoprotein (RNP) complexes then scan the cell in search of protospacer/target sequences that match the spacer sequences contained in the crRNAs. For these systems to be efficient defense systems, it is imperative that they are able to differentiate between foreign and native DNA. To avoid self-targeting of the spacer sequences contained in the CRISPR arrays, the recognition of a CRISPR-subtype characteristic protospacer adjacent motifs (PAMs, present in invader DNA, but not in CRISPR array on host genome) is first required. Upon identification of the required PAM, a conformational change in the Cas/crRNA complex is promoted, which allows unwinding of the target DNA and the initiation of an R-loop formation by hybridization of the crRNA to the protospacer (*74*). Upon successful base-pairing of the crRNA with its cognate target, the Cas protein(s) of the RNP complex generally proceed(s) with degradation of the targeted nucleic acid (*64*, *71*, *75*).

1.6 Classification of CRISPR-Cas systems

The continuous arms race between prokaryotes and their specific viruses or invading MGE promotes plasticity of CRISPR-Cas systems over evolution. Mining of genomic and metagenomic sequences has uncovered the existence of a wide variety of CRISPR-Cas systems (*76*, *77*). At present, these systems are classified into 2 classes, each with 3 types that include many subtypes. The classification is based on relevant structural and mechanistic features of the CRISPR-Cas systems (*78*). Class 1 systems include multi-subunit protein (Cascade-like) effector complexes that coordinate their activity to target nucleic acids. On the other hand, Class 2 systems are represented by single multi-domain effector proteins. Class 1 and class 2 are further subdivided into types I, III and IV and types II, V and VI, respectively (Figure 1.2). Type I, II and V target DNA (with some exceptions), whereas type III and VI target RNA (*79*, *80*). In addition, many cases of non-specific collateral cleavage have been described for types III, V, VI and eventually type I (*81*), the biological function of which is to induce dormancy or even death of the infected host cell, as a final attempt to avoid further spread of the pathogenic MGE throughout the population.

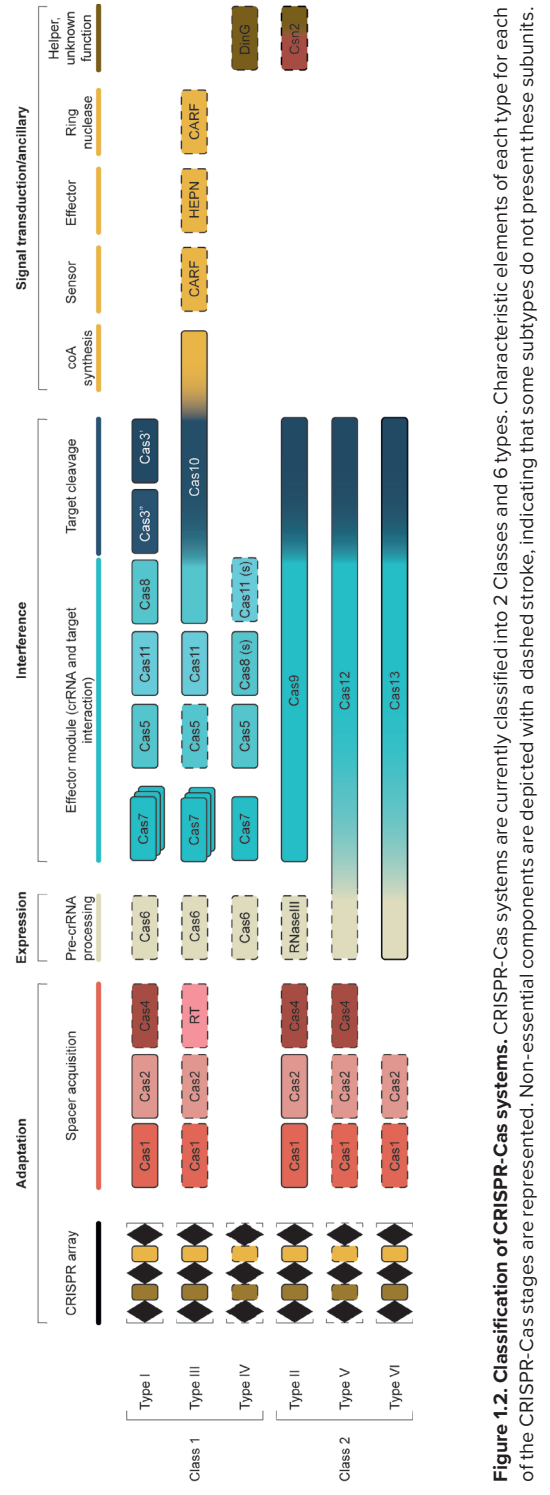


Figure 1.2. Classification of CRISPR-Cas systems. CRISPR-Cas systems are currently classified into 2 Classes and 6 types. Characteristic elements of each type for each of the CRISPR-Cas stages are represented. Non-essential components are depicted with a dashed stroke, indicating that some subtypes do not present these subunits.

1.7 Use of CRISPR-Cas systems as genome editing tools in yeast

In the last decade, class 2 CRISPR-Cas systems have been repurposed as genome editing tools in a wide variety of organisms due to their effector module simplicity. By re-programming the sequence of the spacer part of the crRNA, class 2 effectors can be directed to specifically introduce DBSs in target genomic locations (*71*, *82*). This in turn activates DNA repair pathways and promotes the targeted integration of DNA fragments into specific locations. Among class 2 effectors, the type II-A effector Cas9 has been widely adopted as genome editing tool and used in many organisms (*71*, *82*, *83*). Cas9 requires the action of host RNase III and a tracrRNA to process its pre-crRNA (*70*). For the interference step, the crRNA and the tracrRNA can be synthetically combined into a sgRNA, which simplifies the design (*82*). Upon combination of the sgRNA and the Cas9 protein, the RNP complex scans the DNA in search of 5'-NGG-3' PAM situated at the 3' end of the protospacer. Recognition of the correct PAM allows crRNA pairing to DNA and R-loop formation. Subsequentially, Cas9 introduces blunt DSB via its HNH and RuvC nuclease domains.

One of the organisms in which Cas9 was first tested and evaluated was the model yeast *S. cerevisiae* (*84*). Since then, the use of CRISPR-Cas9 in *S. cerevisiae* has been extensively adopted (*85*, *86*). Introduction of DSBs in this organism activates the highly efficient HDR pathway (*87–89*). Therefore, upon supply of an adequate repair template with homologous regions, various targeted genetic manipulations can be achieved, such as DNA integration, gene deletion, gene disruption or single nucleotide modifications. With the current system, close to 100% modification efficiencies can be achieved in single targeting events with short donor DNA fragments (< 1 kb) (*84*, *90*). However, editing efficiencies of Cas9 are still strictly coupled to the activity of the HDR repair pathway, which limits the use of Cas9 for multiplexed genome editing approaches, for the integration of long DNA fragments or to specific cell phases.

Current strain building programs require the simultaneous modification of multiple loci for the integration of several DNA constructs or the disruption of different native expression pathways. The programmability of Cas9 has allowed to target multiple sites at the same time, which decreases strain engineering programs span. However, introducing multiple DSBs in *S. cerevisiae* has a cytotoxic effect due to the fragmentation of the chromosomes and, therefore, the number of loci that can be simultaneously modified is limited by the number of individual cells that survive the multiplexed genome editing event. Temporal control over the CRISPR-Cas elements used for genome editing could theoretically improve the outcome of multiplexed genome editing events, by increasing the number of individual cells that survive the multiplexing process. Although some approaches have already focused on the use of inducible regulatory sequences at the DNA (*91*), RNA (*92*, *93*) or protein level (*94*), these have mainly been developed for *S. cerevisiae* or require the use of markers to achieve high homologous recombination efficiencies.

Modification of production strains for metabolic engineering purposes often requires the introduction of long biosynthetic pathways. However, even when using Cas9, integration efficiencies rapidly decline with increasing sizes of the integration fragments and recombination remains active only during S or G2 phases of the cell cycle (*94–96*). Still, large DNA sequences up to 24 kb can be integrated in several copy numbers by targeting repetitive sequences such as the Ty transposon sequences across the chromosomes, although it requires extended Cas9 and sgRNA expression for several days (*97*).

Therefore, genome editing tools should be improved to allow the integration of long DNA constructs by establishing HDR-independent processes or by modifying the HDR activity. Some attempts of using HDR-independent methods have been focused on the use of transposases linked to deactivated Cas proteins, yet their use has been mainly explored for animal cell genome editing (*98–101*). Moreover, the efficiencies and specificities of these systems require improvement.

Multiple genomic loci can be targeted and modified simultaneously by expressing multiple guide RNAs. The main limitation of this approach is the need of multiple sgRNA expression cassettes (*84*, *90*) or of engineering complex pre-crRNA processing systems based on the co-expression of endoribonucleases (*93*), self-cleaving tRNAs (*102*, *103*) or ribozymes (*83*).

The use of CRISPR-Cas tools in *S. cerevisiae* has not been limited to the incorporation of permanent large DNA modifications or deletions. CRISPR-Cas systems have also been used for silencing and activation of gene expression by combining deactivated Cas proteins to activation or repression transcriptional domains (*104*, *105*). Modification of Cas9 and combinations of nickase or deactivated variants to base editors or prime editors has also allowed the use of CRISPR-Cas proteins for the incorporation of targeted point mutations in yeasts (*106–109*).

Recently, the type V effector protein Cas12a (described in more detailed below) has also been used for genome editing (*110*, *111*) and gene regulation (*112*) in the model yeast. This has allowed straightforward multiplexed editing events due to the pre-crRNA processing activity of the type V nuclease (*110*, *111*). Moreover, it has allowed editing of AT-rich regions, less accessible for Cas9 (*111*, *112*). To date, no other Cas protein has been used for genome editing in the yeast *S. cerevisiae*. Characterization of novel nucleases may allow for targeting of previously inaccessible sites, increasing editing efficiencies or decreasing off-target activities. Moreover, novel functionalities might be described for novel effectors which might be interesting for genome editing purposes.

The use of CRISPR-Cas tools has also revolutionized genome editing in multiple species of non-model yeasts. However, its application is limited due to the absence of replicative plasmids (for some species), the absence of characterized RNA polymerase III promoters for sgRNA expression or, most importantly, the low homologous recombination efficiencies of these

yeasts, where NHEJ is the dominant repair pathway of DSBs (*113*, *114*). The latter issue makes the screening for correct targeted integrations a laborious task. Therefore, improvements in the currently established CRISPR-Cas technologies (largely based on the use of Cas9) or implementation of novel CRISPR-Cas based tools could advance the genome editing programs in these yeasts. Several strategies such as the use of long homologous recombination arms (*115–117*), NHEJ knockout strains (*35*, *118–120*) or cell cycle synchronization (*96*, *121*) have been used to enhance homologous recombination efficiencies. However, most of these strategies are organism-dependent and require the use of markers to boost integration of the repair template.

1.8 Type V CRISPR-Cas systems

Type V systems are represented by the signature protein and effector Cas12. Among all Cas12 orthologs, Cas12a (initially called Cpf1) is the most extensively characterized (*122*, *123*). Cas12a is an RNA-guided multi-domain protein able to induce targeted dsDNA cleavage (*122*, *124*). Cas12a presents a bilobed structure with a REC and a NUC lobe (*124*, *125*). Several domains are embedded in the NUC lobe: (i) a Wedge (WED) domain, (ii) a PAM interacting (PI) domain, (iii) a single RuvC domain responsible for dsDNA cleavage,(iv) a Bridge Helix (BH) domain, and (v) a Nuclease (Nuc) domain that remains inactive (*124*, *126*). Cas12a can process its own pre-crRNA without the aid of neither RNaseIII nor a tracrRNA (*122*, *123*). Like many other Cas12 subtypes, also Cas12a proteins requires a T-rich PAM upstream of the target sequence. Upon PAM recognition, downstream sequence interrogation starts by extension of the crRNA-DNA heteroduplex and is followed by the R-loop formation. Correct crRNA base-pairing and R-loop formation trigger cleavage of the displaced strand (non-target strand) by the RuvC domain. Subsequent rearrangement of the target strand to position it close to the active site of the RuvC domain for cleavage seems to be prompted by the Nuc domain. This results in the generation of 5' staggered ends upon target cleavage of the DNA strands (*126*). Mutations in the Nuc domain only prevent cleavage of the target strand (*124*), while deactivation of the RuvC domain results in a completely deactivated nuclease (*122*).

Mining of genomic and metagenomic databases has allowed the identification of many novel CRISPR-Cas Type V systems (*127*, *128*). Some of them lack characteristic elements of CRISPR *loci* such as the adaptation module (*cas1* and *cas2*) or an adjacent CRISPR array. Others present a RuvC-deactivated nuclease domain or have been shown to have two different nuclease activities (*76*). Moreover, collateral nucleic acid targeting activities have also been related to some recently characterized type V nucleases (*129*). Interestingly, significantly smaller nucleases of this CRISPR-Cas type have been discovered (*130–133*). Experimental characterization of some of these *in silico* predicted systems is still required. Exploration of their activity could widen the genome editing toolkit for eukaryotic organisms, improve genome editing programs and allow editing of difficult-to-access genomic regions.

Therefore, in this thesis we explore the diversity of newly predicted type V CRISPR-Cas systems and assess their use in the model yeast *S. cerevisiae*. Moreover, we propose the improvement

of already existing CRISPR-Cas tools to overcome current genome editing limitations in non-model yeasts. We believe that novel CRISPR-Cas systems could eventually yield improved editing efficiencies, require different PAMs or offer new functionalities for genome editing. By expanding the current genome editing toolbox, we aim to reduce the time and labor needed in genome editing programs into DBTL cycles and simplify the generation of genetically modified organisms for biotechnological purposes. Moreover, functional characterization of these systems in the model yeast *S. cerevisiae* can be used as a starting point for implementation of novel genome editing tools in non-model yeasts or complex eukaryotic systems.

1.9 Thesis outline

Chapter 1 provides a short historical introduction on the use of yeast in biotechnological processes and the importance of genome editing tools in metabolic engineering approaches. Initially, commonly used approaches for genome editing in yeast are described. Next, CRISPR-Cas systems are introduced and described as adaptive immune systems in bacteria and archaea. Detail is provided on the re-purposing of these systems as genome editing tools and their application in yeast. Finally, we focus on the diversity of type V effectors, which have been recently characterized and could eventually be harnessed for genome editing in yeast.

Chapter 2 discusses the use of CRISPR-Cas tools for multiplex genome editing and regulation of microorganisms. The chapter details different approaches for expression of multiple guides, analyses different donor DNA delivery strategies and explains different methods to increase the activity of the homology direct repair pathway over the non-homologous end joining or alternative NHEJ repair pathways. Finally, the challenges and outlook of multiplex genome editing are analyzed.

Chapter 3 presents the characterization, both *in vitro* and *in vivo* (in *Escherichia coli*) of a novel subtype V-M, and its effector Cas12m. The selected *Mm*Cas12m protein originates from *Mycolicibacterium mucogenicum*. In this chapter we show that Cas12m is an RNA-guided DNA binding protein without an active RuvC domain. We prove that the characterized binding activity can be used in *E. coli* for both gene silencing and plasmid interference. Via fusion of Cas12m to a deaminase protein, we show that Cas12m base editors have a characteristic double editing window.

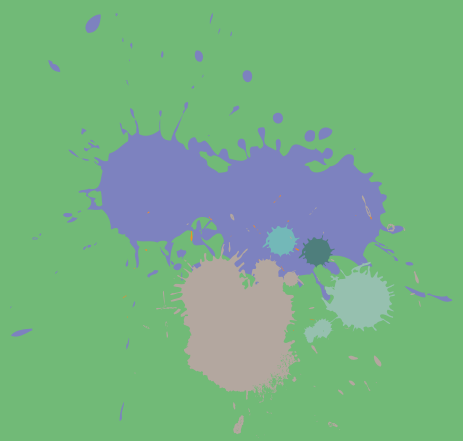
Chapter 4 assesses the expression and activity of Cas12m in the yeast *Saccharomyces cerevisiae*. We successfully localize Cas12m in the nucleus of *S. cerevisiae* and detect expression of the protein with the expected protein size. We use eGFP expressing strains to determine whether Cas12m can be used as a genome silencing tool as observed in *E. coli*. We compare its activity to the activity of previously used CRISPR-Cas-based silencing tools (dCas12a-Mxi1 and dCas9-Mxi1), and we conclude that the gene repression strength of Cas12m is lower.

Chapter 5 assesses the activity of Cas12m base editors in the yeast *S. cerevisiae*. Two different cytidine deaminase base editors are tested to edit the characterized editing window obtained in *E. coli*. Targeted base editing is observed in low population frequencies (< 2%) even when the editing period is extended over time. Unspecific activity of the base editors is observed at similar frequencies as target editing, indicating that future use of these base editors would need further optimization.

Chapter 6 focuses on the adaptation of the type V-K CRISPR-associated transposon system for *S. cerevisiae*. First, the transposon system is tested in *E. coli* at temperatures that were lower than described previously to validate the system *in vitro*. Next, nuclear localization signals are fused to the N- or C-terminus of each of the complex components to assess possible disruption of the complex activity due to the fusion of the small peptides. These nuclear localization signals (NLS) will be required for translocation of the complex to the nucleus of *S. cerevisiae*. Placement of the NLS peptides at the C-terminus of two of the proteins of the transposition complex is shown to disrupt the activity of the transposon complex, while NLS peptides at the N-termini generally worked well. Finally, the CRISPR-associated transposon system is expressed in *S. cerevisiae* following two different strategies for delivery of the donor DNA, unfortunately not resulting in the desired transposition under any of the tested conditions.

Chapter 7 reports the adaptation of a recently developed CRISPR-Cas tool (SIBR-Cas) for its use in *S. cerevisiae*. Originally, this tool is developed for time control of the expression of Cas12a in prokaryotic organisms and allows engineering of wild-type bacteria. We show that the original design of the tool is not suitable for its use in eukaryotes. Modification of the design of the tool allows for controlled induction of the activity of Cas12a in *S. cerevisiae* and opens the door for the use of this tool in non-model yeasts.

Chapter 8 summarizes the results of this thesis on the expansion of CRISPR-Cas based tools in *S. cerevisiae*. A discussion on the status and outlook of CRISPR-Cas tools in yeast is presented, focusing on the main improvement points of the field. The diversity and applicability of type V systems is discussed, together with the transferability strategies of these novel discovered effectors to eukaryotic systems in general and to yeast in particular.



CHAPTER 2

Multiplex Genome Editing of Microorganisms Using CRISPR-Cas

Belén Adiego-Pérez[1]
Paola Randazzo[2]
Jean Marc Daran[2]
René Verwaal[3]
Johannes. A. Roubos[3]
Pascale Daran-Lapujade[2]
John van der Oost[1*]

[1] Laboratory of Microbiology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands

[2] Department of Biotechnology, Delft University of Technology, Van der Maasweg 9, 2629 HZ Delft, The Netherlands

[3] DSM Biotechnology Center, Alexander Fleminglaan 1, 2613 AX Delft, The Netherlands

* corresponding author

CHAPTER ADAPTED FROM PUBLICATION:

FEMS MICROBIOLOGY LETTERS, **VOLUME 366, ISSUE 8, APRIL 2019, FNZ086, HTTPS://DOI.ORG/10.1093/FEMSLE/FNZ086**

2.1 Abstract

Microbial production of chemical compounds often requires highly engineered microbial cell factories. During the last years, CRISPR-Cas nucleases have been repurposed as powerful tools for genome editing. Here, we briefly review the most frequently used CRISPR-Cas tools and describe some of their applications. We describe the progress made with respect to CRISPR-based multiplex genome editing of industrial bacteria and eukaryotic microorganisms. We also review the state of the art in terms of gene expression regulation using CRISPRi and CRISPRa. Finally, we summarize the pillars for efficient multiplexed genome editing and present our view on future developments and applications of CRISPR-Cas tools for multiplex genome editing.

2.2 Introduction

Industrial microbiology plays a key role in the transition towards a more sustainable industry to produce food and feed ingredients, bio-based materials, biofuels and direct synthesis of cosmetic and pharmaceutical compounds (*134*). Oil-based production processes are gradually being substituted by bio-based processes, in which genetically engineered microorganisms are generally crucial to achieve cost-effective productivities and yields (*135*, *136*). The implementation of CRISPR-Cas tools has revolutionized genome editing and mitigated the investment in the metabolic engineering programs required to generate highly engineered microbial cell factories (*137*, *138*).

CRISPR-Cas systems (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated proteins) are bacterial and archaeal adaptive immune defence systems, which can be repurposed as versatile genetic editing or regulation tools in a broad range of organisms. The effector endonucleases of these systems are guided by short RNA molecules encoded by CRISPR arrays. Native CRISPR arrays consist of a succession of spacers originating from invader organisms separated by direct repeats (*55*, *58*). Transcription of the CRISPR array results in a long precursor-crRNA transcript (pre-crRNA), that is subsequently being processed to short functional CRISPR RNA (crRNA) guides (*68*). To date, six different types of CRISPR-Cas systems (I-VI) have been described that are divided into two major classes (Class 1 and Class 2) (*77*, *127*, *139–141*). This review focuses on the application of DNA-targeting class 2 CRISPR systems (included in types II and V), that all consist of large multi-domain effector proteins able to use crRNA guides to target complementary DNA. Recent reviews have covered applications of CRISPR-Cas editing in bacteria (*142*), in *Streptomyces* (*143*), in filamentous fungi (*144*), in yeast (*85*, *113*), in microalgae and cyanobacteria (*145*), and in general industrial microorganisms (*146*).

After an introduction of single target genome editing tools, we focus on the spectacular development of multiplexed genome editing by Cas9 (type II) and Cas12a (type V) in industrial microorganisms. Both bacterial and eukaryotic examples are described, although more attention is given to yeast and filamentous fungi, since the diversity of strategies using Cas endonucleases for genome editing applications is more extensive in this group of organisms.

2.3 Single target genome editing and regulation

Since its establishment as a genome editing tool (*71*, *147*), Cas9 from *Streptococcus pyogenes* (*Sp*Cas9) has become the most widely used RNA-guided endonuclease for genome editing and transcription regulation purposes (Table 2.1). Expression of this type II Cas nuclease together with a guide RNA (gRNA) is sufficient for generating targeted blunt double-stranded breaks (DSBs). The gRNA bound by *Sp*Cas9 consists of two small RNA molecules: a CRISPR RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA). To simplify gRNA expression, a synthetic chimeric construct named single guide RNA (sgRNA) can be synthesized by fusing the tracrRNA and the crRNA (*71*). Targeting of complementary DNA sequences (protospacers) by the Cas9:gRNA complex requires a protospacer adjacent motif (PAM), in case of Cas9 positioned downstream of the target sequence (*148*). Correct PAM identification and base-pairing will trigger cleavage of the non-target and target DNA strands by the RuvC and HNH nuclease domains, respectively (*71*, *147*) (Figure 2.1).

Table 2.1. Characteristics of the most commonly used Cas orthologues for genome editing.

	Cas9		Cas12a	
Ortholog	*Sp*Cas9	*Fn*Cas12a	*As*Cas12a	*Lb*Cas12a
Subtype	II-A		V-A	
Organism of origin	*Streptococcus pyogenes*	*Francisella novicida*	*Acidaminococcus sp.*	*Lachnospiraceae bacterium*
Nuclease domain	HNH, RuvC		RuvC	
tracrRNA	Yes		No	
PAM (5'-3')	NGG	TTTV	TTTV	TTTV
Size (amino acids)	1368	1302	1307	1228
RNA processing	No/RNaseIII	Yes/WED III	Yes/ WED III	Yes/ WED III
Minimum guide length (mature)	~100 nt		~44 nt	
Reference(s)	(*70*)		(*149*)	

The more recently characterized endonuclease Cas12a (formerly called Cpf1) (type V) can cleave dsDNA directed by a crRNA, hence without the requirement of a tracrRNA (*122*) (Table 2.1). Cas12a does not possess an HNH domain, and its RuvC domain has been demonstrated to cleave both the non-target and the target DNA strands (*126*, *150*). Moreover, Cas12a is able to process its crRNA guide autonomously (*126*, *149*, *151*), while Cas9 relies on the activity of an additional non-Cas, dsRNA (crRNA/tracrRNA) targeting ribonuclease (RNaseIII) (*70*). Both Cas9 and Cas12a can use multiple crRNA guides for creating simultaneous DSBs at different target loci in the genome (Figure 2.1). Recently, two distinct Cas12 subtypes (Cas12b, CasX/Cas12e) were shown to also edit genomes of bacteria and mammalian cells (*152*, *153*).

Genomic DSBs can be repaired by homology-directed repair (HDR), non-homologous end joining repair (NHEJ) or alternative non-homologous end joining systems such as microhomology-mediated end joining (MMEJ) (*154–156*). The error-prone NHEJ repair system is often most prevalent in eukaryotes (*39*) (Figure 2.1), whereas it has been predicted to be encoded by only ~26% of publicly available prokaryotic genomes (*157–159*). The less studied alternative MMEJ repair system has been reported to also be present in bacteria and fungi (*154*). This repair system has been proven to be active together with other repair mechanisms in the fungi *Aspergillus niger* and *Yarrowia lipolytica* (*160*) or in NHEJ-free bacteria (*156*). HDR can be used in a targeted way: (i) to insert DNA fragments in targeted genomic locations; (ii) to delete small and large DNA fragments; or (iii) to introduce point mutations. HDR requires the introduction of a single or double-stranded DNA repair fragment into the cell, called donor DNA (dDNA), encoding the desired novel property or designed nucleotide change. To avoid targeting after the designed change, the recombinant sequence generally contains one or more silent mutations in the protospacer or PAM recognition sequence or partial deletion thereof. On the other hand, in organisms with highly active NHEJ or MMEJ repair systems, the introduction of non-specific insertions (only in case of NHEJ) and/or deletions (indels) in a certain target sequences can lead to gene disruption (*82*).

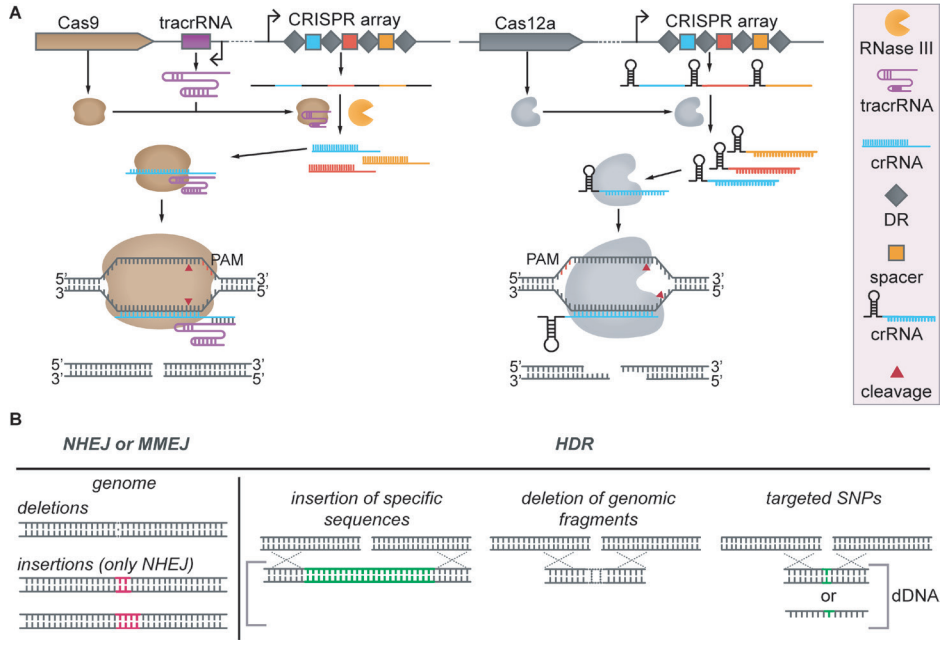


Figure 2.1. General functioning of CRISPR-Cas systems types II-A and V-A, Cas9 and Cas12a, respectively and general DNA repair mechanisms. (A) Cas9 and Cas12a expression and cleavage schemes. Left panel: Cas9 requires tracrRNA transcription and RNase III expression for CRISPR array transcript processing. Cas9 forms a complex with crRNA and tracrRNA and cleaves target DNA generating blunt ends. Right panel: Cas12a processes its own CRISPR array transcript to obtain individual crRNAs without the requirement of any tracrRNA or RNAse III co-expression. Cas12a stays in complex with crRNA and cleaves target DNA generating staggered ends. **(B)** Double strand break (DSB) repair mechanisms. DSBs can be repaired via non-homologous end joining (NHEJ), alternative non-homologous end joining repair pathways such as microhomology-mediated end joining (MMEJ), or via homologous direct recombination. NHEJ and MMEJ repair pathways can lead to the incorporation of deletions or insertions (only in case of NHEJ) in the targeted region. HDR is combined with the supplementation of donor DNA (dDNA), which can be double stranded or single stranded. dDNA can be used for insertion of long DNA sequences, deletion of genomic fragments or introduction of single point mutations (SNPs).

Inactive or deactivated versions of both Cas9 and Cas12a (named dCas9 and dCas12a) have been designed by substituting one or more of the catalytic amino acids in the nuclease domains (*71*, *122*, *147*, *150*). These variants have been used to regulate gene expression in many organisms since they retain the target-binding ability (*161*). By directing dCas9 or dCas12a to the promoter or coding sequence of a target gene, transcriptional repression (silencing) can be achieved by steric hindrance of the RNA polymerase and/or of transcription factors required for transcription of the target gene. This CRISPR interference (CRISPRi) technique has initially been established in *Escherichia coli*, resulting in significant transcriptional repression when targeting either the promoter or the non-template DNA strand of an open reading frame (*162*, *163*). In eukaryotic microorganisms, gene repression is normally achieved by fusing repressor domains such as the mammalian transcriptional repressor domain Mxi1 or the Krüppel-associated box (KRAB domain) to the C-terminus of dCas9 or dCas12a (*104*, *164*). Recently, native repression

domains have been characterized in the yeast *Saccharomyces cerevisiae* with multiple Cas9 orthologs (*165*). This practice is more common in eukaryotic organisms since the use of only a dCas9:gRNA-complex seems not to be sufficient to significantly block transcription (*166*). Moreover, fusions of these deactivated variants to transcription activation domains are used to achieve gene activation (CRISPRa). In eukaryotes, VP16, VP64, Gal4^{AD} or the synthetic VPR activator domains have been used successfully (*104*, *167*, *168*), while the omega (ω) subunit of the RNA polymerase has been used in bacteria (*163*, *169*).

2.4 Multiplex genome editing

Editing of multiple loci is often required to introduce multiple heterologous genes and to fine-tune metabolic networks of microbial cell factories. In the pre-CRISPR era, iterative rounds of genome editing making use of selection markers were necessary to build strains expressing multiple-gene expression pathways. Establishment of marker-free CRISPR-Cas tools brought powerful nuclease-mediated multiplex genome engineering capabilities, considerably saving time and resources in strain construction programs. The multiplexing capabilities of CRISPR-Cas systems as genome editing tools have been widely exploited with Cas9 and more recently with Cas12a. One of the first microbial applications demonstrating the multiplexing capabilities of Cas9 was performed with the native Cas9 system of *Streptococcus pneumoniae* and two spacers expressed from a synthetic array integrated into the genome (*170*). After this proof of principle, multiple studies explored the multiplexing capabilities of the endonuclease Cas9 in mammalian cells (*82*, *171*), as well as in industrially relevant prokaryotic and eukaryotic microorganisms (Table 2.2).

The crRNA processing activity of the recently characterized Cas12a increases the simplicity of multiplexing. By expressing a single CRISPR array containing multiple spacers under the transcriptional regulation of a single promoter and terminator, multiple loci can be targeted simultaneously (*149*). Therefore, there is no need of supplying multiple targeting expression constructs. Recent studies have demonstrated the multiplex editing potential of the Cas12a endonuclease in a wide range of microorganisms (Table 2.2).

Aneuploidy and polyploidy are common conditions among eukaryotic industrial microorganisms. The requirement for simultaneous targeting of multiple alleles in non-haploid strain results in a decrease of the CRISPR editing efficiencies (*172*). The term "*cis*-multiplexing" is used when targeting a single genomic locus found multiple times across the genome of a non-haploid organism. "*Trans*-multiplexing" refers to the simultaneous introduction of modifications in multiple genes which occur in more than one copy in the genome (*173*). Several examples of both *cis*- and *trans*- multiplexing have recently been described (Table 2.2).

Table 2.2. Multiplexed genome editing events in industrial microorganisms using CRISPR-Cas systems.

Specie **[strain(ploidy)]**	**Cas nuclease tool (expression), Plasmid (replication origin)/ genome integrated**	**strategy for multiplexed gRNA expression/delivery (expression), plasmid (replication origin)/ genome integrated**	**Type of donor DNA: HFs; amount/concentration**	**Type of modification: N° target, Editing efficiency**	**Reference**
PROKARYOTES					
E. coli [MG1655]	*Sp*Cas9 (inducible), plasmid expression (repA101 ori)	several sgRNA expression cassettes (constitutive), plasmid expression	circular dsDNA (pMB1 ori):~300 bp	Knockouts: 2; 100% 3; 88.3% 4; >30%	(*174*)
E. coli [MG1655]	*Sp*Cas9 (constitutive), plasmid expression (repA101 (Ts) ori)	several sgRNA expression cassettes (constitutive), plasmid expression (pMB1 ori)	circular dsDNA (pMB1 ori): 250 - 550 bp	Knockouts: 2; 97%±4%; 3; 47%±8%	(*175*)
E. coli [MG1655]	*Sp*Cas9 (inducible), plasmid expression (ColE1 ori)	several sgRNA expression cassettes (inducible), plasmid expression (pMB1 ori)	linear ssDNA: 70 bp; 5 pmol	Short insertions: 2; ~70%	(*176*)
E. coli [MG1655]	*Sp*Cas9 (constitutive), plasmid expression (p15A ori)	several sgRNA expression cassettes (constitutive), plasmid expression (ColE1 ori)	linear ssDNA: ~89 bp; 50 pmol	Point mutations: 2; 83% 3; 23%	(*177*)
Streptococcus pneumoniae [crR6c]	*Sp*Cas9 (constitutive), genome integrated	native-like CRISPR array (constitutive), genome integrated	linear dsDNA: not mentioned;0.7 ng/µl to 2.5 µg/µl	Deletions: 2; 75%	(*170*)
Streptomyces lividans	*Sp*Cas9 (constitutive), plasmid expression (pSG5rep)	several sgRNA expression cassettes (constitutive), plasmid expression (pSG5rep)	circular dsDNA (oriT): 1 kB	Short deletions (20-34 bp): 2; 100% (4/4)	(*178*)

Table 2.2. Multiplexed genome editing events in industrial microorganisms using CRISPR-Cas systems. *(continued)*

Specie **[strain(ploidy)]**	**Cas nuclease tool (expression), Plasmid (replication origin)/ genome integrated**	**strategy for multiplexed gRNA expression/delivery (expression), plasmid (replication origin)/ genome integrated**	**Type of donor DNA: HFs; amount/concentration**	**Type of modification: N° target, Editing efficiency**	**Reference**
Streptomyces coelicolor [M145]	*Sp*Cas9 (constitutive), plasmid expression (pSG5rep)	several sgRNA expression cassettes (constitutive), plasmid expression (pSG5rep)	circular dsDNA (pSG5): ~1 kB	Deletions (768-1053 bp): 2; 29-54%	(*179*)
E. coli [MG1655]	*Fn*Cas12a (inducible), plasmid expression (repA 101)	native-like CRISPR array (constitutive), plasmid expression (pSC101 ori)	circular dsDNA (oriE): 500 bp	Gene insertions: 3; ~20%	(*180*)
Streptomyces coelicolor [M145]	*Fn*Cas12a (constitutive), plasmid expression (pSG5rep)	native-like CRISPR array (constitutive), plasmid expression (pSG5)	circular dsDNA (pSG5): ~1 kB	Knockouts: 2; 75%	(*181*)
EUKARYOTES					
Saccharomyces cerevisiae [CEN.PK113-7D (n), CEN.PK2-1c (n), CEN.PK122 (2n)]	*Sp*Cas9 (constitutive), genome integrated	several sgRNA expression cassettes (RNA pol III promoter, constitutive), plasmid expression (multicopy)	Linear dsDNA: 60 bp; 12pmols	Knockouts: 2, 100%; 6, 65%;	(*90*)

Table 2.2. Multiplexed genome editing events in industrial microorganisms using CRISPR-Cas systems. *(continued)*

Specie **[strain(ploidy)]**	**Cas nuclease tool (expression), Plasmid (replication origin)/ genome integrated**	**strategy for multiplexed gRNA expression/delivery (expression), plasmid (replication origin)/ genome integrated**	**Type of donor DNA: HFs; amount/concentration**	**Type of modification: N° target, Editing efficiency**	**Reference**
Saccharomyces cerevisiae [CEN.PK2–1c (n)]	*Sp*Cas9 (constitutive), genome integrated	several sgRNA expression cassettes with homology flanks to a linearized plasmid backbone plasmid expression (multicopy)	Linear dsDNA: 500 bp; 0.6-1.54pmols	Knockouts: 3, 64%;	(*88*)
Saccharomyces cerevisiae [CEN.PK113-5D (n)]	*Fn*Cpf1 (constitutive), genome integrated	native-like CRISPR array (RNA pol III promoter, constitutive) plasmid expression (multicopy)	Linear dsDNA: 60 bp; 12pmols	Knockouts: 2, 100%; 4, 85%	(*111*)
Saccharomyces cerevisiae [CEN.PK113-7D (n), Ethanol Red (2n)]	*Sp*Cas9 (constitutive), plasmid expression (multicopy)	several sgRNA expression cassettes (RNA pol III promoter, constitutive), plasmid expression (multicopy)	Linear ssDNA: 40 bp; 300pmols	Knockouts: 2, 91-98%	(*182*)
Saccharomyces cerevisiae [BY4741 (n)]	*Fn*Cpf1 (constitutive), centromeric plasmid	native-like CRISPR array (RNA pol III promoter, constitutive) plasmid expression (multicopy)	Linear dsDNA: 50 bp	Multi-gene integrations: 2, 52%; 3, 43%	(*183*)

Table 2.2. Multiplexed genome editing events in industrial microorganisms using CRISPR-Cas systems. *(continued)*

Specie **[strain(ploidy)]**	**Cas nuclease tool (expression), Plasmid (replication origin)/ genome integrated**	**strategy for multiplexed gRNA expression/delivery (expression), plasmid (replication origin)/ genome integrated**	**Type of donor DNA: HFs; amount/concentration**	**Type of modification: N° target, Editing efficiency**	**Reference**
S. cerevisiae [204508; ATCC (mated) (2n)]	*Sp*Cas9 (constitutive) plasmid expression (multicopy)	synthetic array of ribozyme-flanked sgRNA (RNA pol III promoter, constitutive) plasmid expression (multicopy)	Linear dsDNA: 50 bp; 44.94 pmol	Deletions: 2, 43%; 3, 19%	(*173*)
Saccharomyces cerevisiae [CEN.PK113-7D (n)]	*Lb*Cpf1, *As*Cpf1 or *Fn*Cpf1 (constitutive) centromeric plasmid	native-like crRNA-array with homology flanks to a linearized plasmid backbone (RNA pol III promoter, constitutive) plasmid expression (multicopy)	Linear dsDNA: 50 bp; 90-120fmols	Multi-gene integrations: 3, 91%	(*110*)
Saccharomyces cerevisiae [BY4741 (n), CEN. PK2-1c (n)]	i*Sp*Cas9 (constitutive), plasmid expression (multicopy)	native-like crRNA-array (RNA pol III promoter, constitutive), separate expression of tracrRNA plasmid expression (multicopy)	circular dsDNA, at 5' of each spacer sequence (multicopy): 50bp; 142fmols	Knockouts: 3, 27-87%	(*87*)

Table 2.2. Multiplexed genome editing events in industrial microorganisms using CRISPR-Cas systems. *(continued)*

Specie **[strain(ploidy)]**	**Cas nuclease tool (expression), Plasmid (replication origin)/ genome integrated**	**strategy for multiplexed gRNA expression/delivery (expression), plasmid (replication origin)/ genome integrated**	**Type of donor DNA: HFs; amount/concentration**	**Type of modification: N° target, Editing efficiency**	**Reference**
Saccharomyces cerevisiae [CEN.PK113-7D (n)]	*Sp*Cas9 (constitutive), genome integrated	synthetic crRNA-array (RNA pol III promoter), separate expression of *Pa*Csy4 plasmid expression (multicopy)	Linear dsDNA: 60 bp; 12pmols	Knockouts: 2, 100%; 4, 96%	(*93*)
Saccharomyces cerevisiae [BY4741 (n)]	*Sp*Cas9 (constitutive), centromeric plasmid	Several HDV ribozyme-sgRNA expression cassettes (RNA pol III promoter, constitutive) plasmid expression (centromeric)	Linear dsDNA, barcoded: 60bp; 55pmols	Knockouts: 2, 65-87.5%; 3, 57.5-75%; 4, 27.5-15%	(*184*)
Saccharomyces cerevisiae [CEN.PK2-1C (n)]	*Sp*Cas9 (constitutive), centromeric plasmid	several sgRNA expression cassettes (RNA pol III promoter, constitutive) plasmid expression (multicopy)	Linear dsDNA: 500 bp; 700fmols	Multi-gene integrations: 3, 84%	(*185*)
Saccharomyces cerevisiae [CEN.PK111-27B (n)]	*Sp*Cas9 (constitutive), centromeric plasmid	several sgRNA expression cassettes (RNA pol III promoter, constitutive) plasmid expression (multicopy)	Linear dsDNA: 50bp; 4pmols	Multi-gene integrations: 2, 58%; 3, 30.6%	(*89*)

Table 2.2. Multiplexed genome editing events in industrial microorganisms using CRISPR-Cas systems. *(continued)*

Specie **[strain(ploidy)]**	**Cas nuclease tool (expression), Plasmid (replication origin)/ genome integrated**	**strategy for multiplexed gRNA expression/delivery (expression), plasmid (replication origin)/ genome integrated**	**Type of donor DNA: HFs; amount/concentration**	**Type of modification: N° target, Editing efficiency**	**Reference**
Saccharomyces cerevisiae [CEN.PK2-1C]	*Sp*Cas9 (constitutive), genome integrated	several sgRNA expression cassettes (RNA pol III, constitutive), some target more than one site; plasmid expression (multicopy)	Linear dsDNA: 60 bp; 26.96 pmol	Deletions: 9, 50% (only 2 transformants on plate)	(*186*)
Saccharomyces cerevisiae [CEN.PK 113-5D]	*Sp*Cas9 (constitutive), multicopy plasmid	synthetic crRNA-array (with one RNA pol III promoter for the expression of four gRNAs), gRNAs between $tRNA^{gly}$ sequences; plasmid expression (multicopy)	Linear dsDNA: 50 bp; 266.9 pmol	Deletions (8 bp): 8, 86.7%	(*102*)
Ogataea parapolymorpha [CBS 11895 (n)]	*Sp*Cas9 (constitutive), centromeric plasmid	synthetic array of ribozyme-flanked sgRNA (RNA pol II promoter, constitutive); plasmid expression (centromeric)	Linear dsDNA: 480bp, 1.6pmols	Knockouts: 2, 2-5%	(*120*)
Ogataea polymorpha [CGMCC7.89 (n)]	i*Sp*Cas9 (constitutive), genome integrated	several sgRNA expression cassettes expressed (RNA pol III promoter, constitutive); genome integrated	Linear dsDNA: 1kb, 1.7pmols	Multi-gene integrations: 3, 30.56 ± 2.40%	(*187*)

Table 2.2. Multiplexed genome editing events in industrial microorganisms using CRISPR-Cas systems. *(continued)*

Specie **[strain(ploidy)]**	**Cas nuclease tool (expression), Plasmid (replication origin)/ genome integrated**	**strategy for multiplexed gRNA expression/delivery (expression), plasmid (replication origin)/ genome integrated**	**Type of donor DNA: HFs; amount/concentration**	**Type of modification: N° target, Editing efficiency**	**Reference**
Yarrowia lipolytica [ATCC 201249 (n), ATCC MYA-2613 (n)]	*Sp*Cas9 (constitutive), plasmid expression (centromeric)	synthetic array of ribozyme-flanked sgRNAs (RNA pol II promoter, constitutive); plasmid expression (centromeric)	On multicopy plasmid: ~450bp	Knockouts: 2, 36.7 ± 8.5%; 3, 19.3 ± 9.2 %	(*119*)
Penicillium chrysogenum [DS68530]	*Sp*Cas9 (transient) delivered as a RNP	*in vitro* synthetized sgRNA in RNP, protoplast-mediated transformation; transient expression	Linear dsDNA: ≥ 1kb, 1-11 µg	Cassette integration: 2; 50%	(*188*)
Aspergillus nidulans [IBT27263 (n)]	*Sp*Cas9 (constitutive) plasmid expression (centromeric)	synthetic array of tRNA-flanked sgRNAs (RNA pol III promoter, constitutive); plasmid expression (centromeric)	Linear ssDNA: 45bp, 1µmol	Multi-purpose: 3, 90%;	(*189*)
Trichoderma reesei [ATCC 13631, (n); ATCC 56765 (n)]	*Sp*Cas9 (inducible) genome integrated	*in vitro* synthetized sgRNA delivery by protoplasts transformation; transient expression	Linear dsDNA: 200bp, 296pmols	Knockouts: 2, 16-45%; 3, 4.2%	(*190*)

Table 2.2. Multiplexed genome editing events in industrial microorganisms using CRISPR-Cas systems. *(continued)*

Specie **[strain(ploidy)]**	**Cas nuclease tool (expression), Plasmid (replication origin)/ genome integrated**	**strategy for multiplexed gRNA expression/delivery (expression), plasmid (replication origin)/ genome integrated**	**Type of donor DNA: HFs; amount/concentration**	**Type of modification: N° target, Editing efficiency**	**Reference**
Scheffersomyces stipitis [UC7, (n)]	*Sp*Cas9 (constitutive) plasmid expression (centromeric)	several sgRNA expression cassettes (RNA pol III promoter, constitutive) plasmid expression (centromeric)	Linear dsDNA: 500bp	Knockouts: 2, 40%;	(*191*)
Kluyveromyces lactis [ATCC8585 (n)]	*Sp*Cas9 (constitutive), genome integrated	several sgRNA expression cassettes with homology flanks to a linearized plasmid backbone plasmid expression (multicopy)	Linear dsDNA: 1kb, 0.6-1.54pmols	Multi-gene integration: 3, 2.1%	(*88*)
Myceliophthora thermophila [ATCC 42464, (n)]	*Sp*Cas9 (constitutive), plasmid expression (centromeric)	several sgRNA expression cassettes (RNA pol III promoter, constitutive), transient expression	Linear dsDNA: 600bp; ~12pmols	Knockouts: 2, 61–70%; 3, 30% ; 4, 22%	(*192*)
Saccharomyces pastorianus [CBS1483, (n#)]	*Sp*Cas9 (constitutive), plasmid expression (centromeric)	synthetic array of ribozyme-flanked sgRNA (RNA pol II promoter, constitutive) plasmid expression (centromeric)	Linear dsDNA: 60bp; 12pmols	Knockouts: 2, 100%	(*193*)

Table 2.2. Multiplexed genome editing events in industrial microorganisms using CRISPR-Cas systems. *(continued)*

Specie **[strain(ploidy)]**	**Cas nuclease tool (expression), Plasmid (replication origin)/ genome integrated**	**strategy for multiplexed gRNA expression/delivery (expression), plasmid (replication origin)/ genome integrated**	**Type of donor DNA: HFs; amount/concentration**	**Type of modification: N° target, Editing efficiency**	**Reference**
Komagataella phaffi [CBS7435 (n)]	*Sp*Cas9 (constitutive), plasmid expression (centromeric)	several sgRNA (ribozyme-flanked) expression cassettes (RNA pol II promoter, constitutive) plasmid expression (centromeric)	Linear dsDNA: 1kb; ~400-770fmols	Knockouts: 2, 69±13%	(*194*)
Phaeodactylum tricornutum (2n)	*Sp*Cas9 (transient), delivered as a RNP	*in vitro* synthetized sgRNA in RNP, biolystic delivery transient expression	----------	Knockouts: 2, 65-100%; 3, 15,4%	(*195*)
Magnaporthe oryzae (2n)	*Sp*Cas9 (transient), delivered as a RNP	*in vitro* synthetized sgRNA in RNP, protoplast-mediated transformation transient expression	Linear dsDNA: ~40 bp	Knockouts by SNP: 2, 3.4-12.3%	(*196*)
Fusarium fujikuroi NJtech 02, CCTCC M2015614 # anaeuploid	*Sp*Cas9 (constitutive), plasmid expression (centromeric)	several sgRNA expression cassettes (RNA pol III promoter, constitutive), plasmid expression (centromeric)	----------	Knockouts by disruption: 2, 20.8%; 3, 4.2%	(*197*)

2.5 gRNA expression systems for efficient multiplex genome editing

In most reported cases of genome editing of industrial microorganisms, and in all bacterial examples, gRNAs are generally expressed from multicopy plasmids that, after successful editing, can be removed through counter selection (*87*, *198*) or growth on non-selective medium (*90*, *186*). In case of multiplex genome editing, two strategies have been described for the delivery of the different gRNAs: (i) multiple gRNA expression cassettes transcribing a sgRNA molecule from one or more plasmids (Figure 2.2); (ii) polycistronic expression of gRNAs, either inspired by native CRISPR systems or by synthetically designed ones (Figure 2.3). Individual gRNAs can also be *in vitro* transcribed and supplied directly to filamentous fungi or microalgae (Table 2.2).

2.5.1 Multiplexing using multiple single gRNA expression cassettes.

The initial attempts of multiplex genome editing using *Sp*Cas9 relied on combined expression of several individual gRNA expression cassettes. In bacteria, different strength natural or synthetic promoters (constitutive or inducible) are routinely used for gRNA expression (Table 2.2). In eukaryotes, efficient gRNA expression might be controlled by an RNA polymerase-III-dependent promoter. In a cell that expresses a nuclear-localized *Sp*Cas9, this gRNA can direct the nuclease to its target. In addition, fusions of gRNAs and tRNA auto-splicing sequences (used as promoters) have also been demonstrated to yield multiple functional gRNAs in eukaryotes (*199*, *200*). Alternatively, gRNAs flanked with self-cleaving ribozymes on both ends have been expressed from RNA polymerase II-dependent promoters to provide transcripts with modified ends and increase transcript stability in the nucleus (*194*, *201–203*). Moreover, expression of a bacterial T7 RNA polymerase in the yeasts *Y. lipolytica* and *Kluyveromyces lactis* has been used for guide expression (*204*).

In bacteria, the maximal number of reported simultaneous editing events using *Sp*Cas9 varies from organism to organism (Table 2.2). Expression of multiple gRNAs from multiple expression cassettes is done from plasmid-borne or genome-integrated constructs. Most reported examples of multiplex genome editing using gRNAs expressed from multiple expression cassettes have been performed in *E. coli* (*175–177*), with a maximum of four genes targeted simultaneously and an editing efficiency of ~30% (*174*). The authors of the study developed a CRISPR multiplex genome editing technique which uncouples transformation and editing. This separation is achieved by inducing Cas12a expression only after transformation and seems to be key, together with recombineering, for increased editing efficiencies (*174*, *205*).

In eukaryotes, the combination of several gRNA expression modules in a single amplicon or spread over several co-transformed plasmids resulted in successful editing in several organisms such as *S. cerevisiae* (*88–90*, *103*, *182*, *184*, *185*), *Komagataella phaffi* (*194*), the xylose-utilizing yeast *Scheffersomyces stipitis* (*191*), the enzyme producer *Myceliophthora thermophila* (*192*), *K. lactis* (*88*), *Fusarium fujikuroi* (*197*), the methylotrophic yeast *Ogataea polymorpha* (*187*) and the oleaginous yeast *Y. lipolytica* (*119*) (Table 2.2). In many organisms, high editing efficiencies allow for straightforward screening and selection of the desired mutant

without introducing selectable markers in their genome. However, the reported efficiencies for targeting two or more sites simultaneously in eukaryotes vary significantly from one organism to another (i.e., from 2 to 100%). Using *Sp*Cas9, nine editing events in *S. cerevisiae* is the highest number of simultaneous modifications reported in microbes to date (Table 2.2) (*186*).

In some filamentous fungi, the gRNA molecules can be synthetized *in vitro* and co-transformed with Cas9-encoding plasmids. In microalgae, *in vitro* synthetized gRNAs can be delivered together with *in vitro* produced Cas proteins as ribonucleoprotein complexes (RNPs) as well (*144*, *145*, *188*, *190*). Simultaneous double editing has been achieved by following this strategy in *Penicillium chrysogenum* (*188*, *206*), in *Phaeodactylum tricornutum* (*195*) and in the rice blast fungus *Magnaporthe oryzae* (*196*). Although transformation of *in vitro* synthetized gRNA simplifies gRNA cloning work and avoids the requirement of identifying effective RNA polymerase III promoters in eukaryotes, efficient selection of transformants may still require chromosomal integration of selectable markers located in the dDNA molecules (*188*, *190*). These markers can be removed from the genome by using counter-selection markers or Cre-recombinase-based approaches.

Finally, although *Sp*Cas9 has been successfully implemented in microalgae, we found few examples for the use thereof for multiplexed genome editing (*145*, *207*). Recently, a triple knockout strain of the microalgae *Phaeodactylum tricornutum* was generated in a single step transformation using six different Cas9-based RNP complexes (*195*).

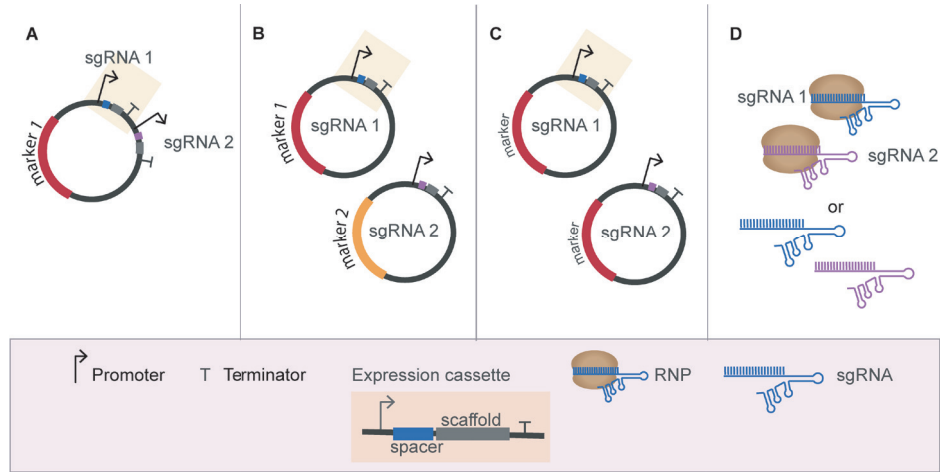


Figure 2.2. Multiplexing using single gRNA expression cassettes. (A) Expression of several sgRNA cassettes from a single expression vector. **(B)** Expression of several sgRNA cassettes from multiple expression vectors (each harboring a different marker). **(C)** Expression of several sgRNA cassettes from multiple expression vectors (all harboring the same marker). **(D)** Transient supplementation with in vitro assembled RNPs or in vitro transcribed gRNAs.

2.5.2 Multiplexing using gRNA polycistronic cassette

Cas9-dependent multiplex genome editing efforts reported to date are generally based on the combination of individual expression cassettes required for single targeting editing experiments. Hence, they commonly suffer from (i) repeated usage of equal sets of promoters and terminators for gRNA expression, (ii) requirement for multiple selection markers, and more importantly, (iii) labour-intensive expression cassettes and plasmid construction. By expressing multiple gRNAs under the control of a single promoter, polycistronic expression cassettes overcome these hurdles. To date, two polycistronic cassette-based approaches have been shown that enable multiplex genome editing (Figure 2.3).

The first approach uses tandem arrays of chimeric sgRNAs that require a dedicated maturation mechanism to release the individual sgRNAs after transcription (Figure 2.3A and B). The Csy4 ribonuclease cleaves synthetic precursor sgRNA arrays when interspaced by 28-nt sequences recognized by Csy4 (*208*). As such, efficient quadruplex editing by *Sp*Cas9 has been accomplished in *S. cerevisiae* (*93*). Alternatively, cleavage of the crRNA array can be performed by self-cleaving RNA sequences. As such, efficient duplex editing was achieved in the non-conventional yeasts *Ogataea parapolymorpha* and *Saccharomyces pastorianus* using sgRNAs flanked with Hammerhead and Hepatitis Delta Virus ribozymes (*120*, *193*). Triplex editing was achieved in a diploid *S. cerevisiae* strain with a synthetic array of sgRNAs flanked by ribozymes (*173*). In addition, tandem fusion of multiple sgRNAs-tRNAs enabling sgRNAs processing by the native tRNA-maturation system was shown to promote *Sp*Cas9-assisted multiplexed genome editing in filamentous fungi (*189*).

The second approach for polycistronic gRNA expression resembles native CRISPR arrays (Figure 2.3C and D). The expression of Cas9 is required together with the transcription of a CRISPR array with multiple crRNAs and a tracrRNA. With this approach, an array of two guides was expressed and processed in a self-targeting system in *E. coli* for plasmid removal. Processing of the CRISPR array transcript occurred via unknown native endoribonucleases (*176*). Again, resembling native CRISPR systems, a three-spacers array and a tracrRNA were co-expressed in *S. cerevisiae* and up to three genes were deleted with efficiencies ranging from 27-87% (depending on the sequence of the targeting guides) (*87*). More recently, several orthologues from the class V endonuclease Cas12a (*As*Cas12a, *Lb*Cas12a and *Fn*Cas12a) (*122*), were shown to deliver efficient multiplex genome editing of *E. coli* (*180*), *Streptomyces coelicolor* (*181*) and *S. cerevisiae* (*110*, *111*, *183*).

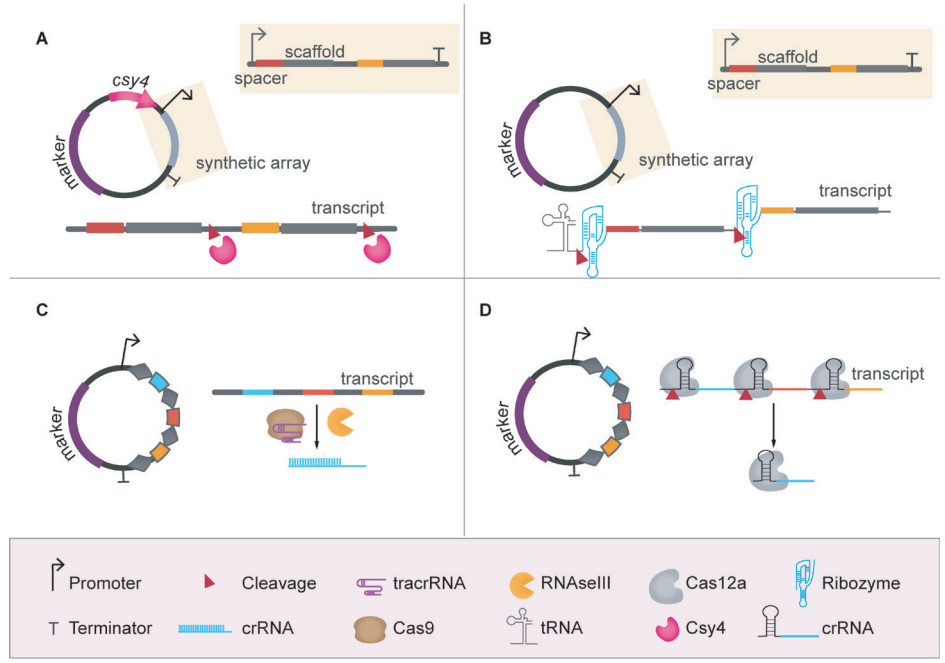


Figure 2.3. Multiplexing using gRNA polycistronic cassettes. (A) Expression of gRNAs from synthetic array dependent on Csy4 processing. In this case, Csy4 has to be co-expressed. **(B)** Expression of gRNAs from synthetic array dependent on endoribonuclease splicing. In most of the reviewed examples, these synthetic arrays are expressed using tRNAs as RNA pol III promoters. **(C)** Expression of gRNAs from native-like CRISPR array dependent on Cas9, tracrRNA and RNAse III processing. **(D)** Expression of gRNAs from native-like CRISPR array dependent on Cas12a processing.

2.6 Efficient delivery of donor DNA for repair of CRISPR-mediated DNA breaks

Current multiplexing strategies are based on two different procedures for the simultaneous delivery of dDNA fragments. On the one hand, transient delivery of these homology repair templates is accomplished by co-transforming multiple linear dDNAs with a corresponding set of gRNAs. This approach was implemented for efficient introduction of multiple barcoded repairs fragments (short DNA fragments that contain a unique sequence tag) (*83*), and editing of heterologous metabolic pathways in both eukaryotic and prokaryotic genomes (*89*, *90*, *110*, *170*, *177*, *185*, *187*). While dDNA is often delivered in double-stranded configuration, it should be noted that successful repair using single-stranded DNA (ssDNA) has also been achieved in *S. cerevisiae* (*84*, *182*), *Aspergilli* (*189*), in *E. coli* (*176*, *177*), and in mammalian cells (*209*, *210*). Alternatively, a stable source of dDNA sequences is provided when these are cloned in one multi-copy plasmid (*119*, *174*, *175*, *178–181*). In a more elegant way, these sequences can be cloned in tandem to the corresponding gRNAs under control of a single promoter. These hybrids are long transcripts which include multiple repair template-crRNA sequences between

CRISPR repeats to be further processed into shorter gRNAs. These gRNAs will still include the transcribed sequence of each dDNA template (*87*, *211*, *212*).

Repair efficiencies of the double-stranded DNA break at a targeted locus depend on the size of the homology flanks (HFs) of the dDNA. Commonly used dDNA sequences may vary from short-sized HFs (~50 bp for *S. cerevisiae* or ~50- 100 bp for *E. coli*), to longer, PCR-based HFs for many bacteria, non-conventional yeasts and fungi (200-1000 bp) (Table 2.2). Long HFs have been shown to increase the efficiency of HDR for both single target and multiplex genome editing in bacteria and eukaryotic microorganisms (*180*, *185*, *187*). In *S. cerevisiae*, multiplex gene deletions can be obtained simultaneously by using oligo-sized dDNAs with short HFs (~50 bp), either as single-stranded (*182*) or as annealed double-stranded dDNAs (*90*, *93*, *111*) (Table 2.2). The amount of dDNA has previously been considered a key factor for enhancing CRISPR-Cas-mediated editing (*88*). Cells are generally transformed with a relatively high concentration of dDNAs, in the order of picomoles (*88*, *90*). This concentration is higher than the one usually used for *in vitro* plasmid assembly procedures in *S. cerevisiae* (*213*).

2.7 Methods for increasing HDR frequencies versus NHEJ and alternative NHEJ repair

The potential of CRISPR-Cas9/Cas12a for implementing precise HDR-based genome editing is often hindered by (i) the presence of NHEJ or alternative-NHEJ repair systems, (ii) the presence of inefficient HDR systems, or (iii) an unfavourable balance between NHEJ and HDR repair mechanisms.

In bacteria, Cas9 and Cas12a are mainly used as counter-selection tools: the endonucleases create DSBs causing cell death due to the absence or poor efficiency of NHEJ repair systems. Only those cells which successfully obtained and integrated appropriate repair templates into their genome avoid the occurrence of persistent chromosomal breaks (*38*). The expression of recombineering systems based on the λ-Red recombinase is extensively used in organisms such as *E. coli*, *Lactobacillus* (*214*) or *Pseudomonas putida* (*215*) to boost the efficiency of HDR and increase the number of recombinants (*216*).

In the case of a multiplex engineering approach in yeasts, the desired genome editing (and the cell viability) relies on the stochastic allocation of dDNAs into the nucleus. In *S. cerevisiae*, the relatively high activity of the HDR machinery facilitates precise editing at multiple loci simultaneously (*90*, *93*, *111*). In contrast, the prevalent NHEJ system in non-conventional yeasts (or other difficult-to-engineer organisms, such as microalgae) might need a more extensive screening due to the high heterogeneity of transformants (*187*, *195*). To improve the efficiency of HDR in some yeasts such as *Y. lipolytica*, the *KU70* gene responsible for DSB repair in the NHEJ pathway was disrupted (*121*). A similar approach is often applied in non-conventional yeasts, e.g. in *Naumovozyma castellii*, where simultaneous deletion of the orthologues of *KU70* and *KU80* completely abolished NHEJ repair during CRISPR-based editing (*217*). Alternatively,

researchers have inhibited certain endogenous DNA repair components to favor HDR in CRISPR experiments.

The balance between double-stranded break repair pathways is further influenced by the cell cycle phase in which a cell is. By making use of hydroxyurea-mediated cell cycle arrest (S-phase), the frequency of targeted integration is significantly increased in multiple fungi (*96*) and demonstrated specifically for CRISPR-based genome editing of *Y. lipolytica* (*121*). Finally, an improvement of HDR efficiency in single target genome editing has recently been achieved by active recruitment of the dDNA to the DSB making use of the ability of some proteins to bind DSBs in *S. cerevisiae* (*212*). Similar methods could be considered for obtaining an increased HDR frequency in a multiplexing set-up for a broader variety of microbes.

2.8 Multiplex gene repression and activation

Multiplexing can also be exploited for CRISPRi/CRISPRa approaches (Table 2.3). Again, both dCas9 and dCas12a have been used for controlling gene expression. In many microorganisms, CRISPRi and CRISPRa have mainly been used to re-direct the carbon flux towards the production of the desired product. This usually requires fine-tuning the expression of multiple genes, which can be achieved by multiplexed gRNA expression for polygenic targeting. Multiplexing using CRISPRi has been explored in several prokaryotic industrial organisms such as *E. coli* (*218*, *219*), *S. coelicolor* (*181*, *220*), *P. putida* (*221*), *Bacillus subtilis* (*222*), *Corynebacterium glutamicum* (*15*) and the cyanobacterium *Synechocystis* (*223*, *224*). Furthermore, different approaches have been explored in the yeasts *S. cerevisiae* (*104*, *165*), *Y. lipolytica* (*225*, *226*) and *Kluyveromyces marxianus* (*227*). Independent studies on the implementation of Cas12a for CRISPRi purposes reported changes in repression strength after altering the order of the spacers in the CRISPR array (*228*). Other studies did not observe the same effect (*149*, *219*, *229*). This suggests that the phenomenon could be gRNA-sequence-dependent and therefore most likely related to transcript secondary structure formation.

Fewer examples can be found of the use of CRISPRa because of the requirement of functional transcription activators. Frequently, CRISPRa applications are used in combination with CRISPRi approaches (*165*). Combination of CRISPRi and CRISPRa using RNA scaffolds was accomplished using similar designs in *E. coli* (*230*) and *S. cerevisiae* (*165*, *231*). In both cases, protein-binding RNA sequences were fused to the 3' end of sgRNAs of Cas9. These protein-recruitment RNA sequences have a high affinity for certain proteins, such as the MCP (used both in *E. coli* and *S. cerevisiae*) or the PCP and Com proteins (tested only in *S. cerevisiae*). By fusing the transcriptional activation domain SoxS (in *E. coli*) or VP64 (in *S. cerevisiae*) to these RNA binding proteins, expression of a gene downstream a target promoter can be enhanced, and pathway fluxes sequentially directed (*230*, *231*).

Table 2.3. Multiplexed genome regulation events in industrial microorganisms using CRISPRi and CRISPRa.

Specie **[strain(poidy)]**	**Cas nuclease tool (CRISPRi/CRISPRa), expression, promoter** **Plasmid (replication origin)/genome integrated**	**Strategy for multiplexed gRNA expression/delivery (expression)** **Plasmid (replication origin)/genome integrated**	**Goal**	**Nº targets**	**Ref**
PROKARYOTES					
E. coli	d*Sp*Cas9 (CRISPRi) - constitutive expression, trc promoter	several sgRNA expression cassettes (constitutive)	increase malate titer	3 targets	(*218*)
[B0013]	plasmid expression	plasmid expression (CloDF13ori)			
E. coli	dd*As*Cas12a (CRISPRi) - constitutive expression, j23100 promoter	several sgRNA expression cassettes (j23119-SpeI constitutive promoter)	proof of principle	3 targets	(*219*)
[MG1655]	plasmid expression (p15A ori)	plasmid expression (ColE1 ori)			
E. coli	d*Sp*Cas9 (CRISPRI/CRISPRa) - constitutive expression, endogenous S. pyogenes promoter	several sgRNA expression cassettes (j23119 constitutive promoter) fused to protein binding sequences (scaffold RNAs)	proof of principle. Activation of ethanol biosynthesis	2 targets	(*230*)
[MG1655]	plasmid expression	plasmid expression			
Bacillus subtilis	d*Sp*Cas9 (CRISPRi) - - inducible expression, *xylA* promoter	several sgRNA expression cassettes (P_{veg} constitutive promoter)	increase N-acetylglucosamine titer	3 targets	(*222*)
[BNY]	genome integrated	genome integrated			
Streptomyces coelicolor	d*Sp*Cas9 (CRISPRi) - constitutive expression, ermE*p promoter	several sgRNA expression cassettes (j23119p constitutive promoter)	proof of principle (knock-out 4 pigmented antibiotic synthesis)	3 and 4 targets	(*220*)
	genome integrated	genome integrated			
Pseudomonas putida	d*Spa*Cas9 (CRISPRi) - inducible expression, LacI-P*tac* promoter	several sgRNA expression cassettes (P*tet* promoter)	proof of principle	2 targets	(*221*)
	genome integrated	plasmid expression (oriV, Rep)			

Table 2.3. Multiplexed genome regulation events in industrial microorganisms using CRISPRi and CRISPRa. *(continued)*

Specie **[strain(poidy)]**	**Cas nuclease tool (CRISPRi/CRISPRa), expression, promoter** **Plasmid (replication origin)/genome integrated**	**Strategy for multiplexed gRNA expression/delivery (expression)** **Plasmid (replication origin)/genome integrated**	**Goal**	**Nº targets**	**Ref**
Corynebacterium glutamicum	d*Sp*Cas9 (CRISPRi) - inducible expression, P*tac* promoter	several sgRNA expression cassettes (P*tac* inducible promoter)	increase aminoacid production	3 targets	(*15*)
	plasmid expression	plasmid expression			
Bacillus subtilis	d*Sp*Cas9 (CRISPRi) - inducible expression, PxylA promoter	several sgRNA expression cassettes (P*veg* promoter)	increase GlcNAc titer	3 targets	(*222*)
[BNY]	genome integrated	genome integrated			
Synechocystis	d*Sp*Cas9 (CRISPRi) - constitutive expression, PpsbA2 promoter	several sgRNA expression cassettes (P_{L31} constitutive promoter)	reduction of PHB and glycogen accumulation during nitrogen starvation	4 targets	(*223*)
	genome integrated	genome integrated			
Synechocystis	d*Sp*Cas9 (CRISPRi) - inducible expression, P_{L22} promoter	several sgRNA expression cassettes (P_{L22} constitutive promoter)	carbon flux re-direction for production of fatty alcohols	6 targets	(*224*)
	genome integrated	genome integrated			
Streptomyces coelicolor	dd*Fn*Cas12a (CRISPRi) - constitutive expression, *ermEp** promoter	native-like CRISPR array (*kasOp** constitutive promoter)	proof of principle (knock-out 3 pigmented antibiotic synthesis)	3 targets	(*181*)
[M145]	genome integrated	genome integrated			
EUKARYOTES					
Saccharomyces cerevisiae	dCas9-VPR (CRISPRi and CRISPRa)	TEF1p-tRNA-sgRNA-tRNA (constitutive)	Increase 2,3-butanediol titer	5 targets (4 interference, 1 activation)	(*103*)
[CEN.PK2-a and Sigma 10560-4A]	plasmid expression (centromeric)	plasmid expression (centromeric)			

Table 2.3. Multiplexed genome regulation events in industrial microorganisms using CRISPRi and CRISPRa. *(continued)*

Specie **[strain(poidy)]**	**Cas nuclease tool (CRISPRi/CRISPRa), expression, promoter** **Plasmid (replication origin)/genome integrated**	**Strategy for multiplexed gRNA expression/delivery (expression)** **Plasmid (replication origin)/genome integrated**	**Goal**	**Nº targets**	**Ref**
Saccharomyces cerevisiae	d*Sp*Cas9 (CRISPRi/CRISPRa) - inducible expression, pGal10 promoter	several sgRNA (with RNA scaffolds) expression cassettes (SNR52p constitutive RNA pol III promoter)	proof of principle, obtention of different violacein biosynthetic pathway products	3 targets	(*231*)
	genome integrated	plasmid expression (centromeric)			
Saccharomyces cerevisiae	d*Lb*Cas12a-VP (CRISPRa), d*Sp*Cas9-RD1152 (CRISPRi), *Sa*Cas9 (CRISPRa) - constitutive expression, pTDH3 promoter	several sgRNA and gRNA in synthetic array between Csy4 recognition sites (TEF1p RNA pol II constitutive promoter)	Increase β-carotene production	3 targets (1 activation, 1 interference and 1 deletion)	(*165*)
	genome integrated	plasmid expression			
Kluyveromyces marxianus	d*Sp*Cas9 (CRISPRi) - constitutive expression, *TEF1p* promoter	several sgRNA expression cassettes (RPR1-tRNA^{gly} ,RNA pol III constitutive promoter)	Increase ethyl acetate production	6 targets (4 genes)	(*227*)
	plasmid expression (centromeric)	plasmid expression (centromeric)			
Yarrowia lipolytica	d*Sp*Cas9-Mxi1 (CRISPRi) - constitutive expression, UAS1B8-TEF	several sgRNA expression cassettes (SCR1-tRNA^{gly} RNA pol III constitutive promoter)	Increase HR by repression of the NHEJ machinery	3 targets (2 genes)	(*225*)
	plasmid expression	plasmid expression			
Saccharomyces cerevisiae	d*Sp*Cas9-VPR (CIRSPRi and CRISPRa)	synthetic array of ribozyme-flanked sgRNAs (Gal1p, RNA pol II inducible promoter)	Proof of principle	2 targets	(*232*)
[BY4741 (n)]	plasmid expression (centromeric)	plasmid expression (centromeric)			

Table 2.3. Multiplexed genome regulation events in industrial microorganisms using CRISPRi and CRISPRa. *(continued)*

Specie **[strain(poidy)]**	**Cas nuclease tool (CRISPRi/CRISPRa), expression, promoter** **Plasmid (replication origin)/genome integrated**	**Strategy for multiplexed gRNA expression/delivery (expression)** **Plasmid (replication origin)/genome integrated**	**Goal**	**Nº targets**	**Ref**
Saccharomyces cerevisiae	d*Sp*Cas9-VPR (CIRSPRi and CRISPRa)	synthetic array of ribozyme-flanked sgRNAs (TEF1p, RNA pol II constitutive promoter)	Proof of principle	4 targets	(*232*)
[BY4741 (n)]	plasmid expression (centromeric)	plasmid expression (centromeric)			
Yarrowia lipolytica	d*Sp*Cas9 or d*Fn*Cas12a (CRISPRi) - constitutive expression	several sgRNA expression cassettes (SCR1-tRNA^{gly} RNA pol III constitutive promoter)	proof of principle, decrease protodeoxy-violaceinic acid	3 targets	(*226*)
[ATCC 201249]	plasmid expression (centromeric)	plasmid expression (centromeric)			

2.9 Challenges and outlook of multiplexing

In recent years, remarkable progress has been achieved in the field of multiplexed genome editing. Besides broadening the spectrum of microorganisms that can be engineered using CRISPR-Cas endonucleases, novel approaches have recently been implemented in terms of gRNA and dDNA delivery for increasing multiplexing efficiency. Below, a summary is provided of the major challenges related to the development of efficient multiplexing CRISPR-Cas systems in microorganisms.

1. *Guide design*: availability of highly efficient guides (sgRNA, crRNA) obtained either through software prediction based on generalized well defined guide-design principles, preferential PAM domains and secondary structure prediction (*228*, *233–237*), or through pre-characterization of the functionality of individual guide performance in single target genome editing experiments.
2. *dDNA design and delivery*: dDNA can be part of a vector, provided as a linear fragment, single or double-stranded, a single fragment of multiple fused repair templates or provided as multiple DNA fragments. Proper characterization of such elements (e. g. by determination of optimal size of homology sequences) for the specific host organism to be edited is required. Additionally, dDNA might be stabilized by chemical modifications (*238*).
3. *Controllable expression of CRISPR elements*: tight expression control of guides and nucleases, via tuning of expression either by systematic variation of constitutive promoters or by using inducible systems. Optimized promoter and terminator sequences are required to fine-tune the expression of the different CRISPR elements (*174*).
4. *Innovative plasmid assembly methods*: smart DNA construction schemes to develop single or multi-plasmid systems containing elements of the CRISPR-Cas system (gRNA, Cas, dDNA, selective marker), including techniques to incorporate repetitive DNA elements such as the repeat sequences of CRISPR arrays (*103*, *228*, *239*). Recently, a cloning-free approach was developed in *S. cerevisiae* for the obtainment of up to 6 simultaneous deletions using *Sp*Cas9 with a efficiency of 23.3%. The strategy combined multiple successful strategies presented in this manuscript. Three transcripts (each of them containing two gRNAs flanked by $tRNA^{gly}$ sequences were expressed from a single plasmid encoding for *Sp*Cas9. The assembly of the plasmid via Golden Gate reaction was performed in the yeast (*102*).
5. *Organism-specific CRISPR tools*: smart choice of the CRISPR-Cas expression approach depending on final application of the production strain or the target organism of choice. For instance, CRISPR-Cas tools can be combined with the introduction of dDNA containing selective markers, which makes screening and selection of positive clones more efficient. This approach can be used in proof of principle studies, whereas in other cases, marker-free strains are important. In a similar way, guides and Cas nucleases can be co-expressed from plasmid-borne expression cassettes or expressed sequentially in strains pre-expressing the Cas nuclease from a second plasmid or from a genome integrated copy.

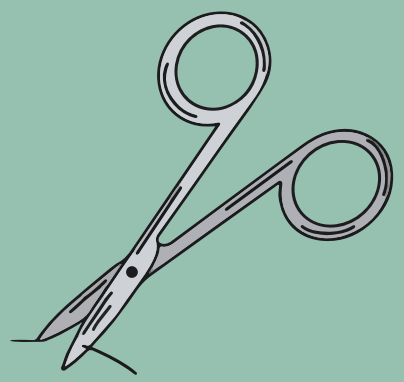
6. *Editing conditions*: optimization of organism-specific CRISPR-Cas delivery systems and recovery protocols. Cell synchronization protocols in combination with CRISPR-Cas systems have been used in human cells to enhance HDR versus NHEJ repair (*240*). In some yeasts, the highest rate of HDR over NHEJ is shown during S-phase (*96*). Therefore, this strategy has been already proposed for its use in industrial microorganisms (*120*).
7. *Novel or improved endonucleases*: these endonucleases should have alternative or less -stringent PAM recognition selection. In addition, nucleases are preferred that are smaller, more specific and more active, as reviewed elsewhere (*241*). Cas9 variants with distinct PAM-recognizing features have been obtained by laboratory evolution (*242*), as well as by structure-guided protein engineering of Cas12a (*241*). The discovery and characterization of novel Cas endonucleases (*153*)- can also extend the PAM compatibility (*130*, *152*, *243*). At the same time, multiple endonucleases can be used in orthogonal designs in order to edit the genome and regulate gene expression simultaneously (*165*, *244*). Alternatively, orthogonal designs can incorporate modified gRNAs, as described for Cas12a (*245*).

In multiplexed genome editing experiments, a negative correlation is experienced between the number of targets and the amount of obtained colonies after transformation in most microorganisms. The introduction of DSBs dramatically reduces the cell survival rate, and this causes limited numbers of simultaneous modifications as a delicate balance between DNA cleavage and repair needs to be established. Alternatively, multiplexed single-base editing does not depend on DSB generation nor dDNA supply and can be used to introduce nucleic acid base changes at a targeted window of DNA (*246*, *247*). In this approach, deactivated or nickase Cas nuclease variants coupled to base editors (cytidine or adenine deaminases) are directed to a target site by the gRNA (*248*). Nucleotide changes are introduced in a targeted DNA window rather than in a precise DNA position, which makes this technique less accurate for the introduction of single point mutations. Multiplexed single-base editing has been recently implemented in prokaryotic microorganisms (*249*, *250*). Useful applications of this technique remain limited to phenotype modifications linked to single nucleotide polymorphisms (SNPs) and to the possibility of introducing stop-codons for gene inactivation in coordination with the presence of a PAM sequence in a determined window (*251*). Although very promising, further development and control of this base editing approach are required in order to broaden the type of modifications and to avoid unwanted mutations (*106*).

Advances in multiplexed genome editing of microorganisms can significantly accelerate future strain construction programs of cell factories with unprecedented efficiencies. Therefore, the CRISPR revolution continues: new tools and workflows are being developed to broaden the range of functionalities of currently used CRISPR-Cas systems as well as knowledge about the mechanism of these systems. As identified in this review, dedicated optimization of each of the elements involved in CRISPR-Cas genome editing is crucial for efficient multiplex genome editing and for stretching the number of simultaneous editing events.

2.10 Acknowledgements:

P.D-L, J.A.R and J.v.d.O received funding from the research programme Building Blocks of Life by the Netherlands Organisation for Scientific Research (NWO, project 737.016.005), and J.A.R and R.V. also received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 764591 (ITN-SynCrop).



CHAPTER 3

The miniature CRISPR-Cas12m effector binds DNA to block transcription

Wen Y. Wu[1*†]
Prarthana Mohanraju[1*]
Chunyu Liao[2]
Belén Adiego-Pérez[1]
Sjoerd C. A. Creutzburg[1]
Kira S. Makarova[4]
Karlijn Keessen[1]
Timon A. Lindeboom[1]
Tahseen S. Khan[1]
Stijn Prinsen[1]
Rob Joosten[1]
Winston X. Yan[3]
Anzhela Migur[2]
Charlie Laffeber[5]
David A. Scott[3]
Joyce H.G. Lebbink[5,6]
Eugene V. Koonin[4]
Chase L. Beisel[2,7]
John van der Oost[1†]

[1]Laboratory of Microbiology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands.

[2]Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz-Centre for Infection Research (HZI), 97080 Würzburg, Germany.

[3]Arbor Biotechnologies, Cambridge, MA 02139, USA.

[4]National Centre for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.

[5] Department of Molecular Genetics, Oncode Institute, Erasmus MC Cancer Institute, Erasmus University Medical Center, 3000 CA, Rotterdam, The Netherlands.

[6] Department of Radiation Oncology, Erasmus University Medical Center, 3000 CA Rotterdam, The Netherlands.

[7]Medical Faculty, University of Würzburg, 97080 Würzburg, Germany.

*** THESE AUTHORS CONTRIBUTED EQUALLY TO THIS WORK.**

[†]TO WHOM CORRESPONDENCE SHOULD BE ADDRESSED: W.Y.W. (WEN.WU@WUR.NL), J.V.D.O. (JOHN.VANDEROOST@WUR.NL)

3.1 Abstract

CRISPR-Cas are prokaryotic adaptive immune systems. Cas nucleases generally use RNA guides to specifically bind and cleave DNA or RNA targets. We here describe the experimental characterization of a bacterial CRISPR effector protein Cas12m representing a novel subtype V-M. Despite being less than half the size of Cas12a, Cas12m processes a pre-crRNA to generate mature crRNA guides. Cas12m recognizes a 5'-TTN' protospacer-adjacent motif (PAM) and stably binds double-stranded DNA (dsDNA). Cas12m lacks a typical RuvC nuclease catalytic site and accordingly fails to cleave nucleic acids either specifically or collaterally. Despite lacking target cleavage activity, the high binding affinity of Cas12m to dsDNA targets allows for interference as demonstrated by its ability to protect bacteria against invading plasmids by down-regulating invader replication. Based on these molecular characteristics, we repurposed Cas12m by fusing it to a cytidine deaminase that resulted in base editing within a unique window.

3.2 Introduction

Most of the bacterial and archaeal CRISPR-Cas systems provide adaptive immunity against invading mobile genetic elements (*252*). The systems are highly diverse structurally and functionally, and have been divided into two classes, each with three types (*139*). Class 1 systems possess multi-protein, Cascade-like effector complexes that bind crRNA guides to target either DNA or RNA. Class 2 systems encompass single-protein effectors (Type II Cas9, Type V Cas12, Type VI Cas13), that also use their crRNA guides to target DNA or RNA containing a sequence complementary to the spacer portion of the crRNA. Upon binding of a nucleic-acid target, interacting nuclease proteins (Class 1) or fused nuclease domains (Class 2) are generally activated, resulting in different target interference mechanisms (*247*, *253*, *254*). A recent genomic survey of CRISPR-Cas systems resulted in a major expansion of type V, with many new sub-types discovered (*76*, *255*).

After the original discovery of Cas12a (originally called Cpf1 (*140*, *256*)), subsequent experimental characterization of this nuclease (*122*, *151*) and its distinct homologs from other type V subtypes (Cas12b and Cas12c (*141*)) revealed crRNA-guided DNA interference suggesting a role in defense that is typical of CRISPR-Cas (Figure 3.1A). By screening the rapidly growing genomic and metagenomic databases, five small Cas12 variants were initially discovered (V-U1 to V-U5; (*76*)). Two of these small variants have recently been characterized: Cas12f (V-U3, also referred to as Cas14a/b/c (*130*, *257*)) and the T7-like transposon-associated, catalytically inactive Cas12k (V-U5; (*258*)). In addition, using similar approaches, six new subtypes have been subsequently characterized: Cas12d/Cas12e (originally named CasY/CasX (*243*)), Cas12g/Cas12h/Cas12i (*129*) and the phage-derived Cas12j (CasΦ) (*259*) (Figure 3.1A).

Characterization of type V subtypes has revealed a highly conserved PAM specificity (T-rich motif at 5' end of protospacer) and two highly conserved domains, a RuvC-like nuclease and Bridge helix (Figure 3.1B). Apart from these domains, however, remarkable structural and mechanistic diversity of Cas12 has been observed. The first surprise was the discovery of two structurally and functionally distinct types of guides: either a short CRISPR-derived crRNA that is processed by Cas12 itself (*122*, *126*, *151*), or a crRNA that base pairs with a second, auxiliary RNA (tracrRNA/scoutRNA) that is processed by an RNase III-type ribonuclease (*130*, *141*) (Figure 3.1A). Other variable features are the nature of the target nucleic acid (DNA or RNA) and the role of the RuvC domain. In different Cas12 variants, RuvC was found to catalyze guide specific, on-target cleavage or nicking of dsDNA, cleavage of RNA (*129*), or simply binding of the target DNA (*258*). Besides on-target nucleic acid cleavage by RuvC, most Cas12 variants also have collateral cleavage activity, where the target-activated RuvC domain ambiguously cleaves ssDNA or ssRNA (*129*, *260*). Apart from defense of bacterial and archaeal hosts against mobile genetic elements (MGEs), some Cas12 effectors are encoded by MGE genomes. One of these, Cas12k, has been shown to mediate RNA-guided transposition of Tn7-like transposons (*258*, *259*). Cas12 effectors also substantially differ in size, ranging from less than 500 to more than 1300 amino acids (*76*).

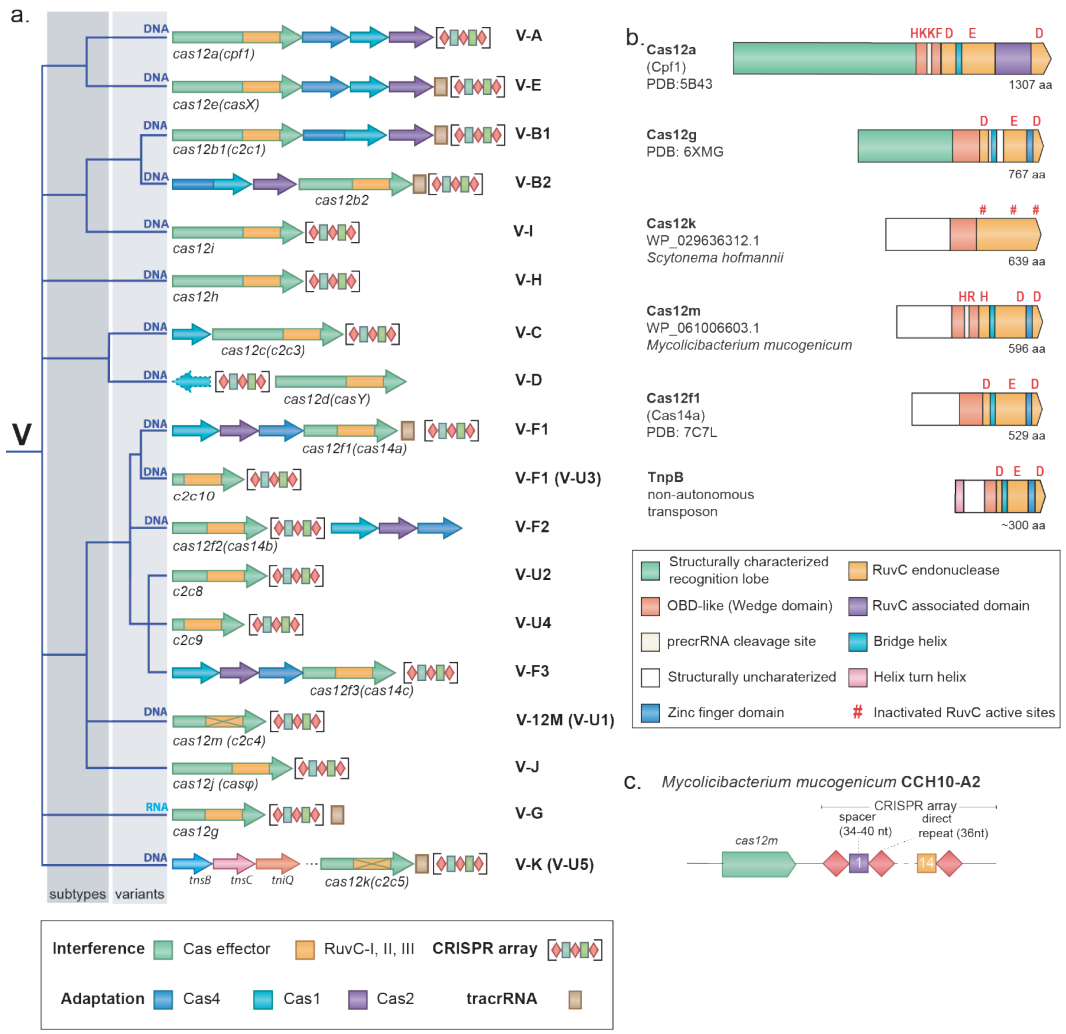


Figure 3.1 Subtype V-M CRISPR-Cas systems. (A) A dendrogram showing the likely evolutionary relationships between the types and subtypes of type V CRISPR-Cas systems. **(B)** Domain architectures of Cas12a, Cas12g, Cas12k, Cas12m and Cas12f1 proteins compared with transposon-encoded TnpB. Proteins are drawn roughly to scale and the number of amino acids is indicated for each protein. **(c)** Organization of the V-M CRISPR-Cas locus in the genome of *Mycolicibacterium mucogenicum* CCH10-A2.

Cas12 proteins are homologous to RNA-guided TnpB nucleases encoded by IS605 family transposons (*76*, *261*, *262*). The sequences of large Cas12 proteins, such as Cas12a, Cas12b and Cas12c, are weakly similar to TnpB, but the smaller Cas12-U proteins show greater similarity and are thought to be evolutionary intermediates between TnpB and larger Cas12 effectors. Phylogenetic analyses suggest that different Cas12-U proteins independently evolved from

different groups of TnpB (*76*, *261*, *262*). Among the V-U systems that remain to be characterized, U2 and U4 cluster within the Cas12f clade in the Cas12 phylogeny (Figure 3.1A, Figure S3.1). Here we describe the experimental analysis of the Cas12-U1, a deeper branching variant (Figure 3.1A, Figure S3.1). Loci of subtype U-1 (hereafter called V-M; Figure 3.1B) were identified in the genomes of diverse bacteria (*76*). We present *in silico* and experimental analyses that reveal relevant features of *Mm*Cas12m from *Mycolicibacterium mucogenicum* CCH10-A2 (hereafter called Cas12m for brevity) (Figure 3.1C), and compare these to other Cas12 subtypes. The results suggest that subtype V-M is a unique CRISPR defense system that functions through silencing the expression of genes of invading MGE. Based on these molecular features, we repurposed Cas12m as a base editor and showing a distinct editing window.

3.3 Results

3.3.1 In silico analysis indicates Cas12m targeting MGEs in the absence of a tracrRNA

The current set of subtype V-M CRISPR-Cas loci consists of 104 non-redundant proteins, encoded in 129 contigs. All V-M loci consist of a single gene encoding the Cas12m effector protein and a CRISPR array (Figure 3.1C), with the implication that V-M systems depend on other CRISPR systems, at least for adaptation. Remarkably, of the 104 V-M loci, 41 were located in a genomic context suggestive of a mobile genetic element origin, 5 within a putative (pro)phage genome and one inside a Tn7-like transposon (data not shown). Comparison of the spacers from the V-M arrays with the NCBI nucleotide sequence database detected 26 protospacer matches, at least 10 of which correspond to plasmid targets. Specifically, the CRISPR array of the V-M system from *Mycolicibacterium mucogenicum* appears to target genes encoding hypothetical proteins on plasmids of *Mycolicibacterium* sp. TY66 (data not shown). At the typical location of the Cas12m PAM, 5' of the protospacer, short T-rich motifs were identified, 5'-(G/C)-T-T-(G/C) (Figure S3.2).

Some Cas12 variants require a second RNA species, either a tracrRNA or an alternatively called scoutRNA, for crRNA processing and interference (*263*). Therefore, we searched the V-M loci for the presence of tracrRNA-like sequences using a previously described prediction approach (*264*). We did not identify candidate tracrRNA in any of the V-M loci, suggesting that the pre-crRNA is processed by Cas12m itself as previously reported for Cas12a (*149*, *151*) and Cas12c/h/i/j (*129*, *149*, *265*). The apparent absence of a tracrRNA in V-M systems is in line with the partial palindromic arrangement of the repeat sequences of the CRISPR arrays that most likely results form a pseudoknot structure (Figure S3.3), a feature that correlates with tracrRNA-independent pre-crRNA processing by Cas12 nucleases (*124*, *149*, *151*). Because no candidate tracrRNAs were identified in the V-M loci, several conserved residues in the predicted OBD domain could contribute to the catalytic site for pre-crRNA processing (Figure S3.4A), as discussed below.

Like all Cas12 subtypes, and particularly the smaller protein variants, Cas12m showed substantial similarity to TnpB (*255*). The similarity between Cas12 and TnpB involves mainly the RuvC domain, which contains three conserved motifs contributing to the nuclease catalytic site (RuvC-I, RuvC-II, and RuvC-III), with the arginine-rich Bridge helix domain separating RuvC-I and RuvC-II (Figure 3.1B and Figure S3.4). A typical feature of the catalytically active RuvC nuclease domain is the presence of three or four negatively charged amino acids that coordinate two divalent cations (Mg^{2+}/Mn^{2+}). In TnpB and in all but two Cas12 subtypes, three active site residues are highly conserved (D-E-D) (Figure 3.1B and Figure S3.5). The two exceptions are the catalytically inactive Cas12k that completely lost its active site residues and Cas12m in which up to two of the metal-coordinating residues are replaced (H-D/E-D) (Figure 3.1B and Figure S3.6) (*76*, *258*). Another domain of TnpB, a Zn Finger (HxxC---CxxC), is conserved in a subset of Cas12 variants including Cas12f and Cas12m (Figure 3.1B).

3.3.2 Cas12m recognizes a 5'-TTN PAM

The Cas12m system of *Mycolicibacterium mucogenicum* CCH10-A2 was selected for experimental study based on the presence of a long CRISPR array, with degenerate repeats at one end of the array indicating the array polarity (*266*). Because this bacterial strain has not been previously cultivated (*267*), its V-M CRISPR locus was heterologously expressed in *Escherichia coli*. The *E. coli* Rosetta (DE3) cells were transformed with two plasmids, one containing an *E. coli* codon-harmonized *cas12m* gene and the other containing a minimal CRISPR array (repeat-spacer-repeat) harboring a spacer targeting the *lac* promoter (P_{lac}). To determine if Cas12m was able to target dsDNA, the PAM was first determined. The PAM plays a central role in the target selection by dsDNA cleaving CRISPR systems. In the absence of a PAM, Cas nucleases generally cannot stably bind the target, even in the presence of a protospacer that is perfectly complementary to the spacer (*268*). As mentioned above, among the spacers in the V-M CRISPR locus of *M. mucogenicum*, two of them partially matched sequences of plasmids from this bacterium (*269*), with a putative 5'-T rich PAM (5'-CTTC and 5'-GTTG) (Figure S3.2).

To determine whether Cas12m recognizes a 5'-T rich PAM, we adapted the previously developed PAM-SCANR assay (*270*), a high-throughput *E. coli*-based positive and tunable screen for PAM determination (Figure 3.2A). This assay is based on a catalytically inactive crRNA-guided Cas nuclease (dCas) blocking the −35 element within the promoter upstream of *lacI*. In the absence of binding (due to a non-PAM) by the dCas effector, LacI is produced, binds to the *lac* operator upstream the green fluorescent protein (GFP) gene, thereby blocking its expression. If, in the presence of a PAM, dCas binds to the -35 region, *lacI* is repressed, resulting in expression of GFP (*270*). Based on the assumption that Cas12m would be able to cleave dsDNA, we generated an effector plasmid (pCas-dCas12m) encoding a "dead" *cas12m* gene by mutating one of the potential RuvC active site residues (D485A). We decided to name the D485A mutant "dCas12m", after the convention used for catalytically inactive derivatives of Cas12a. In addition, a CRISPR array plasmid (pCRISPR-PS) was constructed with a spacer targeting a matching protospacer (PS) upstream of the −35 element of the *lacI* promoter in the target PAM-SCANR plasmid (pTarget-PS). The four nucleotides immediately upstream of this target sequence were randomized to generate the PAM library. A CRISPR array plasmid with a non-targeting spacer (CRISPR-NT) was used as a negative control. *E. coli* cells were transformed with the pCas-dCas12m, combined with either pCRISPR-PS or pCRISPR-NT, as well as the pTarget-PS plasmid. After cultivation, fluorescent cells were isolated by fluorescence-activated cell sorting (FACS) (Figure 3.2B). Comprehensive screening by next-generation sequencing of the pre-sorted and post-sorted PAM libraries followed by analyses of the target-flanking sequences showed that binding of target dsDNA by dCas12m depended on a 5'-NTTM-3' consensus PAM (Figure 3.2C, D). Weakly recognized PAMs were also detected by titrating the Isopropyl β-D-1-thiogalactopyranoside (IPTG) levels to down-regulate the strength of LacI repression (Figure S3.7A). The presence of T nucleotides at the -2 and -3 positions of the 5'-PAM appeared to be the most crucial for PAM recognition (Figure S3.7B). Thus, the PAM recognized by Cas12m is similar to that of the other characterized type V effector proteins (*122*, *129*, *271*, *272*). To validate the PAM and to clarify the ambiguity at the -1 and -4 PAM positions, we generated a set of 16 different plasmids (pTarget-GFP) containing a protospacer adjacent to a 5'-NTTN-3' PAM sequence on the promoter upstream of the *gfp* target gene (Figure S3.7C). *E. coli* cells harboring pCas-dCas12m and the CRISPR array plasmid with a spacer targeting the promoter (pCRISPR-promoter) were transformed with the pTarget-GFP plasmids and assayed for silencing of GFP fluorescence resulting from efficient dsDNA binding (Figure S3.7C). As a control, we analyzed the catalytically inactive type V-A effector (dCas12a) of *Francisella tularensis* subsp. *novicida* U112 (pCas-dCas12a), with its corresponding crRNA guide targeting the same protospacer (Figure S3.7D). Efficient GFP repression was observed for all the tested PAM variants, confirming the PAM sequence 5'-(N)TTN-3' for dCas12m and 5'-TTTV-3' for dCas12a. In addition, this analysis revealed robust *in vivo* crRNA-guided dsDNA binding by both dCas12m and dCas12a (Figure S3.7D).

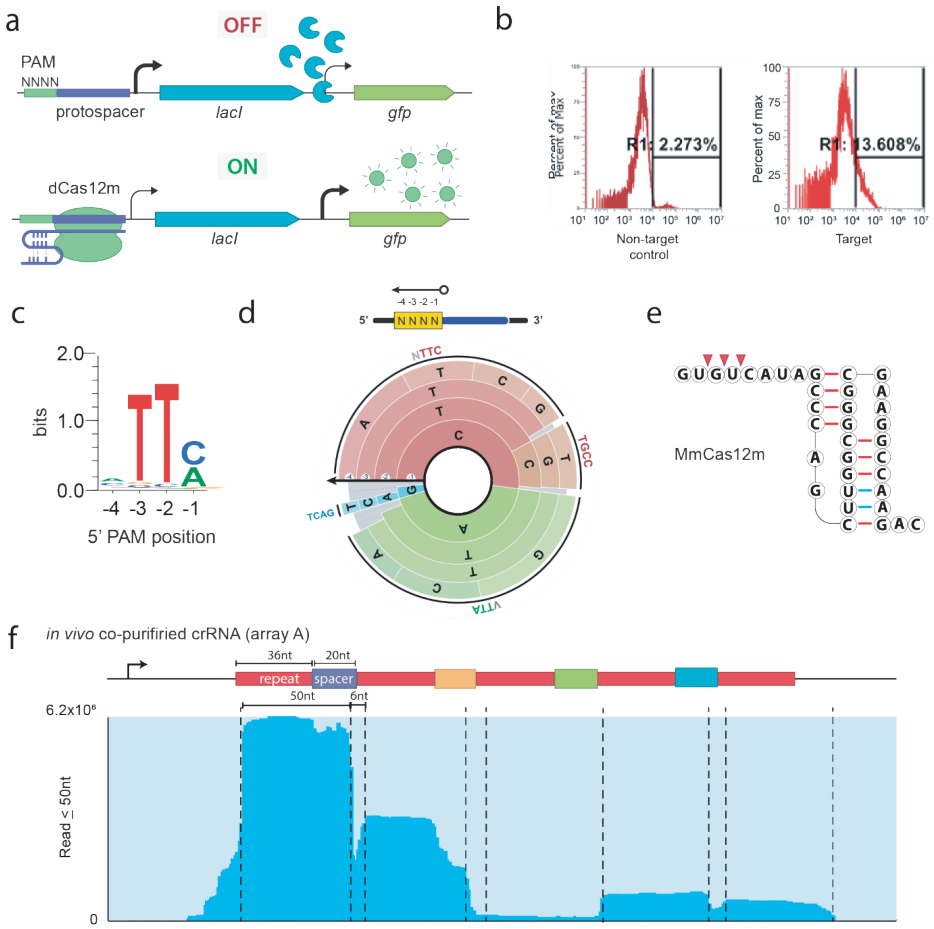


Figure 3.2. Cas12m recognizes dsDNA targets flanked by a 5′-TTM PAM and does not require a tracrRNA. (A) Schematic illustrating the in vivo PAM screen achieved by PAM-SCANR consisting of a library of randomized 5' PAM sequences (4N) cloned upstream of the lacI promoter. Immediately downstream of *lacI* there is a LacI-dependent *lacZ* promoter controlling expression of GFP. A catalytically dead Cas12m (dCas12m) protein is directed to a protospacer within the lacI promoter, resulting in GFP fluorescence only in the presence of a functional PAM. **(B)** Cells harboring a targeting or non-targeting spacer against the pTarget-PS plasmid that led to a GFP fluorescence were isolated by fluorescence-activated cell sorting (FACS). The Y-axis represents the percentage of 10.000 cells and the X-axis represents GFP fluorescence. **(C)** Plasmids from the FACS-sorted cells were extracted and sequenced to determine functional PAM sequences. A weblogo for the Cas12m PAM as determined by NGS sequencing of plasmids from sorted fluorescent cells grown in 0 μM IPTG. **(D)** PAM-wheel representation of enriched PAMs in 0 μM IPTG. Arrow indicates sequences moving from inner to outer wheel. Colors indicate relative frequency of the innermost nucleotide (-1 position). The area of the sector within the PAM wheel is directly proportional to the relative PAM enrichment in the library. **(E)** RNA-seq analysis of co-purified crRNA from cells expressing Cas12m and a 4-spacer CRISPR-array (array a). Reads ≤50 nt were mapped on the expression construct (pCRISPR). **(F)** Predicted pseudoknot of the crRNA of Cas12m (see also Figure S3.3) and indicated cleavage sites for pre-crRNA auto-processing by Cas12m.

3.3.3 Cas12m is responsible for pre-crRNA processing

To further characterize Cas12m and to validate the lack of tracrRNA requirement, Cas12m and a minimal CRISPR-array were co-expressed in *E. coli* (Figure S3.8A). Subsequent analysis demonstrated the presence of co-purified RNAs in the size range of the mature crRNAs (Figure S3.8B). In addition, we performed an *in vitro* processing assay using purified recombinant Cas12m and a two-spacer array (repeat-spacer-repeat-spacer-repeat), and observed the intact pre-crRNA, intermediate products and RNAs that, according to their size, could correspond to mature guides (Figure S3.8C). RNA-seq analysis of these putative crRNAs showed that cleavage predominantly occurred after the 2nd nucleotide of the repeats (Figure 3.2E, F). Similar crRNA processing patterns were observed when Cas12m and the CRISPR-array were co-expressed in a cell-free transcription-translation system (TXTL) (Figure S3.9) (*273*). Co-expression of Cas12m and a four-spacer CRISPR-array in TXTL followed by Northern blot analysis corroborated Cas12m-catalyzed pre-crRNA processing (Figure S3.10) (*228*).

Examination of the multiple alignment of Cas12m proteins suggested five conserved charged residues located in the oligonucleotide binding domain (OBD, also referred to as Wedge domain) that could be implicated in the pre-crRNA processing (Figure S3.4). Each of these residues was mutated individually (R241A, R249A, H269A, R270A and R287A) and a double mutant of two adjacent residues was constructed as well (H269A and R270A). These Cas12m mutants were assayed for silencing of *gfp* expression *in vivo* using four-spacer and single-spacer CRISPR-arrays (Figure S3.11). Single mutations appear to have little to no effect on GFP silencing, whereas the double mutation (H269A, R270A) does suppress the silencing activity of dCas12m, suggesting that at least these two residues are involved in crRNA binding and/or processing.

To test whether target site binding by Cas12m results in crRNA-guided dsDNA cleavage, a target plasmid containing a 5'-CTTA PAM adjacent to the previously used PAM-SCANR protospacer (pTarget-CTTA) was generated. This plasmid transformed into *E. coli* cells harboring the effector plasmid encoding the wild-type Cas12m protein (pCas-Cas12m) with either the pCRISPR-PS or the control pCRISPR-NT plasmid. Notably, upon transformation with the pTarget-CTTA plasmid, no depletion in the number of transformants was observed for the cells harboring pCas-Cas12m and pCRISPR-PS, as compared with the strain harboring pCas-Cas12m and the control pCRISPR-NT plasmids (Figure 3.3A). In contrast, the dsDNA-targeting Cas12a control resulted in the expected substantial depletion of transformants (Figure 3.3A). This result indicates that, at least under the tested conditions, CRISPR-Cas12m did not cleave dsDNA in the heterologous *E. coli* host (Figure 3.3A). To confirm the inability of Cas12m to cleave dsDNA, we repeated the same experiment, but with a plasmid (pCRISPR-GFP) containing a different spacer targeting the 3' region of the *gfp* gene. Again, we did not observe any drop in the number of transformants as compared to the non-target control, indicating that Cas12m lacks DNase activity. Still, because of the commonly used nomenclature of RuvC-disrupted dCas effectors, we decided to stick to the name dCas12m. However, we did observe a substantial decrease of the GFP fluorescence signal in cells that harbored the pCas-Cas12m, pCRISPR-GFP and pTarget-GFP plasmid (Figure 3.3B). This *in vivo* GFP

repression was much lower or undetectable in case of pCas-dCas12m (Figure 3.3B). Thus, the observed silencing of gene expression by Cas12m was, at least to some extent, RuvC-dependent (Figure 3.3B).

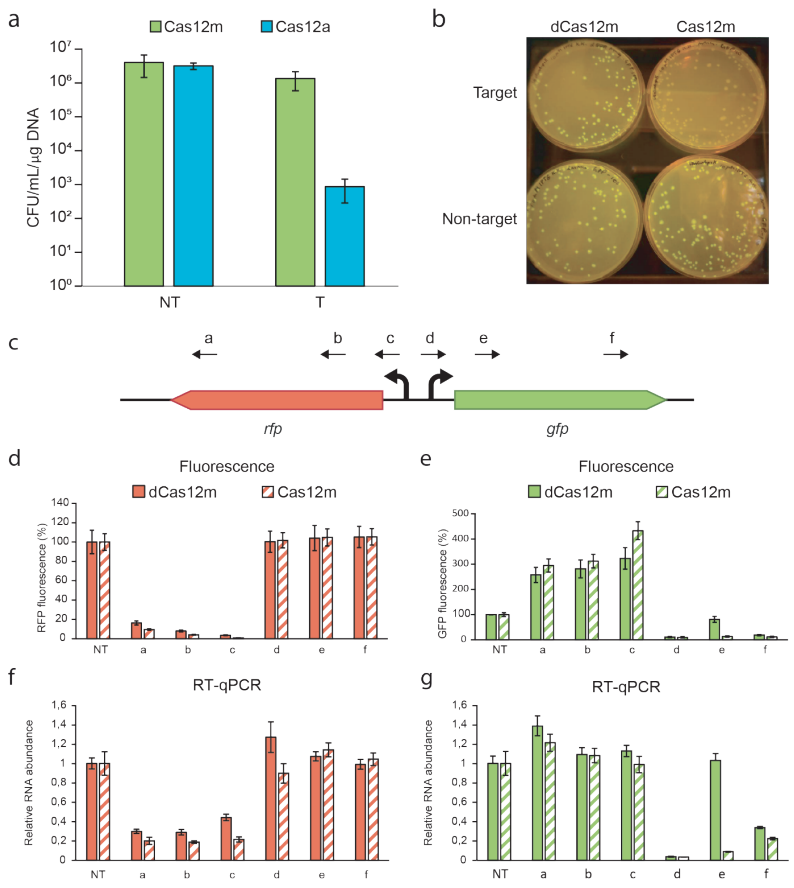


Figure 3.3. Cas12m does not cleave dsDNA but reduces mRNA levels. (A) Results of the *in vivo* dsDNA targeting experiment showing OD_{600} measurements from cultures of *E. coli* harboring the pTarget-PS plasmid transformed with pCas-Cas12m and pCRISPR-PS compared to cells transformed with pCas-*Fn*Cas12a and pCRISPR-Cas12a-PS plasmid. **(B)** Qualitative comparison of GFP fluorescence in the cells harboring pTarget-GFP transformed with pCas-dCas12m with either pCRISPR-GFP (Target) or pCRISPR-NT (Non-target) versus the cells harboring pTarget-GFP transformed with pCas-Cas12m with either pCRISPR-GFP (Target) or pCRISPR-NT (Non-target). **(C)** Schematic of the pTarget-divergent, including the *rfp* and *gfp* genes under the transcriptional control of two different constitutive promoters, P_{taq} and P_{laclq}. The arrows indicate crRNAs used for targeting by Cas12m and dCas12m proteins using the respective spacers (a to f). The PAM is located at the beginning of the arrow. **(D)** RFP fluorescence detected in the cells upon dCas12m and Cas12m targeting using the individual spacers (n = 3; error bars represent mean ± SEM). NT refers to a non-targeting spacer. **(E)** GFP fluorescence detected in the cells upon dCas12m and Cas12m targeting using the individual spacers (n = 3; error bars represent mean ± SEM). NT refers to a non-targeting spacer; No PAM refers to a spacer targeting a protospacer containing a 5'-GGGC PAM (non-functional PAM). **(F)** Relative *rfp* mRNA in the cells upon dCas12m and Cas12m targeting using the individual spacers (n = 3; error bars represent mean ± SEM) by RT-qPCR. NT refers to a non-targeting spacer. **(G)** Relative *gfp* mRNA in the cells upon dCas12m and Cas12m targeting using the individual spacers (n = 3; error bars represent mean ± SEM). NT refers to a non-targeting spacer.

To determine whether the silencing activity of Cas12m was due to target-activated, collateral RNA degradation, as in the case of the type VI effector Cas13a, we generated a target plasmid (pTarget-divergent) with two fluorescence reporter genes, *rfp and gfp*, under the transcriptional control of two divergent constitutive promoters, P_{taq} and P_{laclq}, respectively (Figure 3.3C). *E. coli* cells harboring either the pCas-dCas12m or pCas-Cas12m and the pTarget-divergent plasmids were transformed with a set of CRISPR array plasmids (pCRISPR-(A-F)) containing spacers targeting different locations on the promoters and on the coding strand of either *rfp* or *gfp,* since pCRISPR-GFP targeted the coding strand (Figure 3.3C). Exclusive repression of the targeted reporter gene was observed, indicating that interference by Cas12m occurred locally and was unlikely to originate from collateral nuclease activity. Again, the wild-type Cas12m generally was more efficient than dCas12m in silencing the expression of the reporter gene (Figure 3.3 D, E). In addition to fluorescence measurements, the amounts of transcripts were also measured by quantitative reverse transcription PCR (RT-qPCR), showing similar trends to those observed with the fluorescent signal (Figure 3.3F and G). The observed increase in GFP fluorescence upon repression of the *rfp* gene is most likely due to the relief of the burden on the transcription and translation machinery to produce both GFP and RFP as *gfp* transcripts also increase when targeting *rfp* (Figure 3.3F and G). However, this effect was not observed for RFP when repressing *gfp,* which might be due to the stronger P_{taq} promoter occupying most of the transcriptional machineries and driving high *rfp* expression.

When guided by spacer-e, Cas12m efficiently repressed *gfp*, whereas dCas12m inadequately repressed *gfp*. (Figure 3.3F, G, crRNA E). The poor efficiency of spacer-e suggests sequence- and context-dependent loss of silencing activity, most likely due to hindering RNA secondary structures similar to the previous observations on Cas12a (*228*, *274*). However, this spacer still allowed for silencing by Cas12m, suggesting stronger binding. Collectively, these observations suggest a distinct Cas12 mechanism whereby crRNA-guided binding by Cas12m to a transcriptionally active dsDNA inhibits mRNA transcription (Figure 3.3D-G) (*162*, *275*).

After demonstrating local transcription repression by Cas12m, the next step was to silence both fluorescent proteins simultaneously. To test whether *in vivo* multiplex gene silencing was feasible, a CRISPR-array containing two spacers was used to silencing RFP and GFP, respectively (Figure S3.12). Indeed, both RFP and GFP silencing was achieved, demonstrating efficient *in vivo* multiplex silencing by both Cas12m and dCas12m.

To test the specificity of Cas12m, we examined the mismatch tolerance between the spacer and the protospacer. Single mismatches, tiled 2-nucleotide mismatches and tiled 4-nucleotide mismatches, were introduced across the protospacer in the target *gfp* gene (Figure S3.13A). Cas12m appeared to be relatively tolerant to most single mismatches (Figure S3.13B), except for the mismatch at the PAM-proximal position-8. This is in contrast to Cas12a which, according to an *in vivo* analysis (*276*), is highly sensitive to single or double substitution in most positions between 1 and 18. Double and quadruple mismatches at PAM-proximal positions 1 to 11 severely impaired the silencing activity of Cas12m (Figure S3.13C and D), suggesting the existence

of a seed-like sequence (*277*). Notably, although some mismatches impaired the activity of dCas12m in GFP repression, the effect of the mismatches on GFP silencing by the wild-type Cas12m was much less pronounced (Figure S3.13B-D).

To rule out that Cas12m has target-activated ssRNA cleavage activity, we incubated a purified Cas12m protein first with a crRNA guide, then with a complementary dsDNA plasmid target, and then, with either a target or a non-target RNA. However, under these *in vitro* conditions, Cas12m appeared to be incapable of cleaving either of the RNAs (Figure S3.14). Likewise, a similar *in vitro* approach with linear target DNA also did not reveal ant transcript cleavage (not shown). Hence, the observed interference of Cas12m, can best be explained by Cas12m forming a transcriptional roadblock.

3.3.4 Cas12m inhibits transcription

To further investigate whether the dsDNA target-dependent silencing activity by Cas12m extends further than one gene, we generated a target plasmid (pTarget-Operon) containing a bi-cistronic operon with two fluorescence reporter genes, *rfp and gfp* (Figure 3.4A). *E. coli* cells harboring either the pCas-dCas12m or pCas-Cas12m and the pTarget-operon were transformed with different CRISPR array plasmids (pCRISPR) containing spacers targeting either the non-coding (A1-F1) or the coding (A2-F2) strand at different locations across the entire operon (Figure 3.4A). As expected, crRNA guides that target dsDNA sequences in the proximity of the promoter region resulted in greatly reduced GFP and RFP fluorescent signals, indicating efficient transcriptional silencing of both genes (Figure 3.4B and C, crRNAs A1/A2). Strikingly, although the transcriptional silencing of the fluorescent reporter genes by the dCas12m protein was weak for crRNA guides that targeted dsDNA towards the end of the operon (crRNAs D2/E1/E2), relatively strong repression of both the red and the green fluorescence signals was observed in cells expressing wild-type Cas12m (Figure 3.4B and C, crRNAs D2/E1). The loss of red as well as green fluorescence upon binding to the downstream *gfp* gene indicates that transcription, translation or stability of the entire mRNA was affected by Cas12m (Figure 3.4B and C, crRNAs D1/D2). Moreover, the crRNA guides that targeted dsDNA sequences downstream of the terminator (crRNAs E2/F1/F2) did not cause detectable loss of fluorescence, suggesting that Cas12m mediated RNA-guided transcriptional silencing. Silencing appeared to be confined to the transcripts of the target DNA, in contrast to the collateral cleavage activity that has been reported for type V and type VI effectors (*129*, *265*).

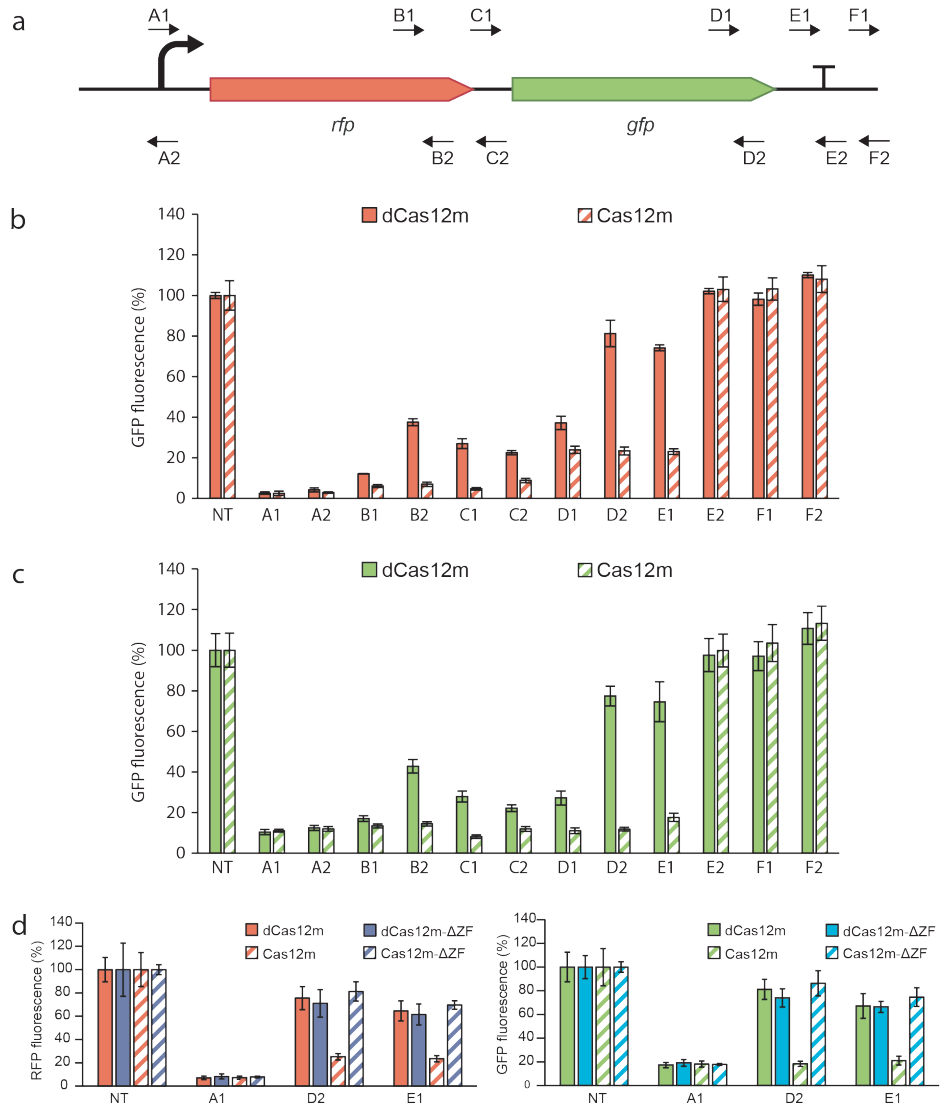


Figure 3.4. RuvC and ZF domain contribute to silencing by Cas12m. (A) Schematic of the pTarget-operon, including the bi-cistronic operon encoding the *rfp* and *gfp* genes. The arrows indicate the crRNAs used for targeting by Cas12m and dCas12m proteins (A1 to F2) with the PAM located at the beginning of the arrow. **(B)** RFP fluorescence detected in the cells upon dCas12m and Cas12m targeting using the individual spacers (n = 3; error bars represent mean). NT refers to a non-targeting spacer. **(C)** GFP fluorescence detected in the cells upon dCas12m and Cas12m targeting using the individual spacers (n = 3; error bars represent mean ± SEM). NT refers to a non-targeting spacer. (D) RFP (left) and GFP (right) fluorescence detected in the cells upon dCas12m, Cas12m, dCas12m-ΔZF and Cas12m-ΔZF targeting using the individual spacers (n = 3; error bars represent mean ± SEM). NT refers to a non-targeting spacer.

Besides the RuvC domain, a Zn Finger (ZF) is conserved in a subset of Cas12 variants including Cas12m. Therefore, the ZF domain was also mutated in both Cas12m and dCas12m to generate double mutants (H549A & C552A) Cas12m-ΔZF and dCas12m-ΔZF, respectively. Cas12m-ΔZF and dCas12m-ΔZF variants appeared to silence pTarget-operon using crRNA guides A1, D2 and E1. All four Cas12m variants (Cas12m, dCas12m, Cas12m-ΔZF and dCas12m-ΔZF) silenced RFP and GFP with equal efficiency with guides targeting the promoter, indicating similar dsDNA binding properties (Figure 3.4B and C, crRNA A1). Further downstream in the coding region, however, where the silencing efficiency of Cas12m still works relatively well, interference by all three mutants (dCas12m, Cas12m-ΔZF and dCas12m-ΔZF) is severely affected (Figure 3.4D).

To validate the observations of the silencing activity of Cas12m in different *in vivo* experiments, we used the TXTL system to evaluate GFP silencing by Cas12m. The Cas12m and its CRISPR array were initially pre-expressed in a TXTL reaction generating RNPs targeting *degfp*, a *gfp* variant exhibiting higher expression in TXTL (Figure S3.15A). This reaction was subsequently used in a new TXTL reaction containing a deGFP plasmid (pdeGFP) The GFP fluorescence was measured over time to assess deGFP repression (Figure S3.15B). deGFP repression was achieved in TXTL using Cas12m and dCas12m (Figure S3.15C). dCas12a was used as a control using a non-targeting spacer (NT) and a spacer targeting the promoter (crRNA 1). Similar to our operon repression experiment (Figure 3.4A), Cas12m yielded higher silencing activity compared to dCas12m when targeting the transcribed region of *gfp* (Figure S3.15C, crRNA 2). The same spacers were then tested *in vivo*, with similar results (Figure S3.15D). These findings corroborated targeted transcriptional silencing by Cas12m.

3.3.5 Cas12m defends against MGE through silencing

Given that Cas12m is unable to cleave dsDNA, the question arises: could type V-M systems provide defense against MGE? To test this, pTarget-GFP plasmids were transformed into *E. coli* cells harboring the effector plasmid encoding Cas12m or dCas12m and a pCRISPR plasmid expressing a minimal CRISPR-array (repeat-spacer-repeat). Spacers targeted essential regions on the plasmid involved in plasmid replication, including the origin of replication (ori) and DnaA binding boxes (*278*) (Figure 3.5A). Binding to these essential regions can disrupt plasmid replication and thereby inhibit further plasmid propagation. In addition, *gfp* was also targeted as a non-essential gene using spacers D2 (GFP1) and E1 (GFP2) from the previous operon silencing experiment (Figure 3.5A). The potential inhibition of plasmid propagation by Cas12m and dCas12m was compared to interference by dCas12a and Cas12a, controls that interfere by binding and cleaving dsDNA, respectively.

In line with its anticipated role in anti-MGE defense, Cas12m efficiently silenced *gfp* (Figure 3.3B). In addition, Cas12m substantially lowers the transformation efficiency of pTarget-GFP when targeting regions essential for replication of the plasmid (Ori1, Ori2 and DnaA) (Figure 3.5B). Again, the interference efficiency of dCas12m is lost by the RuvC mutation. No difference in transformation efficiency was detected when non-essential regions were targeted with Cas12m, dCas12m (Figure 3.5B, N.E.S (non-essential sequence), GFP1 and GFP2).

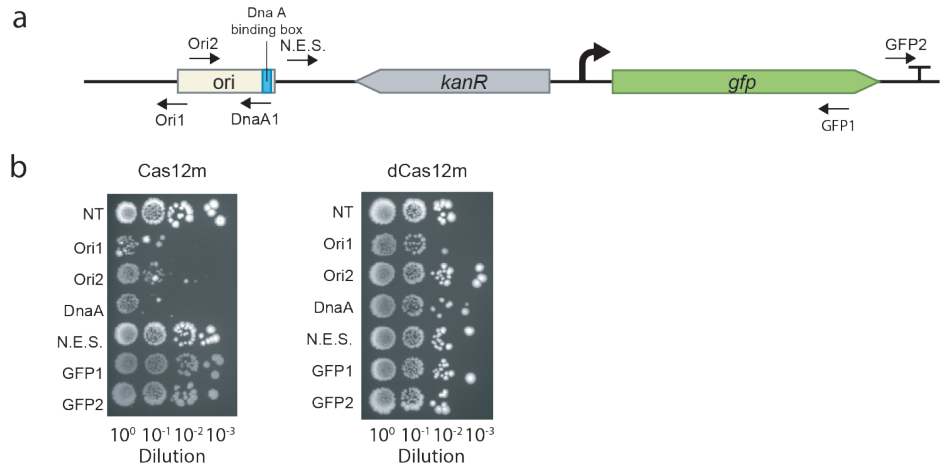


Figure 3.5. Binding by Cas12m or dCas12m can inhibit plasmid propagation. (A) Spacers for targeting the pTarget plasmid (linear representation) in a transformation assay. Arrows indicate and location and orientation of the targeted region with the PAM located at the beginning of the arrow. **(B)** Drop plating of a transformation assay using Cas12m, dCas12m. Rows represent spacers used for targeting and columns represent sample dilutions. N.E.S. = non-essential sequence. Transformation assay was performed using biological duplicates (Figure S3.16).

In addition, Cas12a and dCas12a were included as controls (Figure S3.16B). dCas12a inhibited plasmid propagation similarly to dCas12m, except for spacers Ori1 and DnaA. For most of the tested spacers, the dsDNA-cleaving nuclease Cas12a caused a more pronounced inhibition of transformation than any of the Cas proteins that only bound DNA. Thus, cleavage of dsDNA is more efficient in plasmid clearance than dsDNA binding that presumably blocks replication initiation (*279*). Moreover, as could be expected, in the case of Cas12a, plasmid propagation can be inhibited even by targeting non-essential regions (Figure 3.4A, Figure S3.16B). Cas12a failed to affect plasmid propagation with spacer GFP2 and had only a weak inhibitory effect with spacer DnaA, which most likely reflects an inefficient spacer in general (*228*, *274*).

3.3.6 Cas12m binds stronger to dsDNA than dCas12m in vitro

Most of the DNA-targeting class 2 effectors possess the ability to recognize, bind and cleave dsDNA substrates (*71*, *122*, *260*, *280*, *281*). We first compared the binding of Cas12a, Cas12m, dCas12m and dCas12m-ΔZF (all in complex with the aforementioned (PS) crRNA guide (Figure 3.2A)) to complementary dsDNA fragments immobilized on a chip surface using surface plasmon resonance (SPR) (Figure 3.6, Figure S3.17). All proteins form complexes with the dsDNA during the 1-minute association phase, with Cas12a binding more rapidly than Cas12m and its variants at the same protein concentrations. A major fraction of all these complexes remains stably bound and no difference in dissociation rate is apparent between Cas12a and Cas12m variants if dissociation is monitored for 1 minute (Figure 3.6, Figure S3.17). However, if we monitor dissociation for 5 hours, clear differences can be observed, with the majority of

wild type Cas12m remaining bound for the full duration of the experiment, dCas12m completely dissociating, and the Zinc finger mutant and Cas12a displaying an intermediate behavior. The dCas12m-ΔZF and dCas12m mutants have considerably lower affinities for their target DNA than Cas12m. This indicates that the observed increased silencing *in vivo* by Cas12m is most likely due to stronger binding to dsDNA.

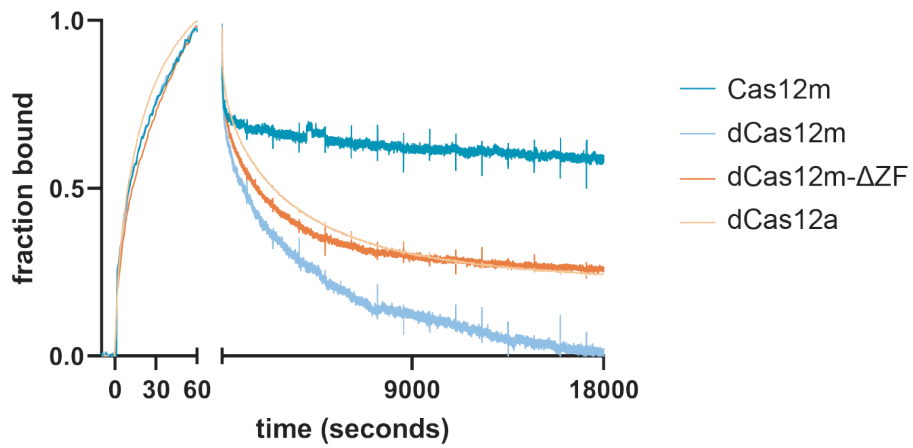


Figure 3.6. Relative binding and dissociation of crRNA guided Cas12m, dCas12m, dCas12m-ΔZF and Cas12a to double stranded DNA analyzed using Surface Plasmon Resonance. Proteins (63 nM) and buffer were added at time 0 and 60 seconds, respectively. Spikes during the extended dissociation phase are caused by consecutive buffer intakes of the microfluidic system.

3.3.7 Cas12m cytidine base editor has unique base editing window

The Cas12m protein characterized in this study enables robust gene silencing *in vivo* in *E. coli* while not cleaving dsDNA. Taking advantage of these properties, we repurposed Cas12m for genome editing. To this end, Cas12m cytidine base editors (CBE) were constructed, which allows for guided single nucleotide C-to-T substitution (*106*, *282*). Although no nucleic acid cleavage by Cas12m was observed in our experiments (see above), we used dCas12m for constructing BEs; a control experiment with homologous Cas12m base editors resulted in similar quantitative/qualitative editing (Figure S3.18). dCas12m-CBE1 was constructed by fusing dCas12m to a cytidine deaminase (CDA), a uracil glycosylase inhibitor (UGI), and an LVA degradation tag (AANDENYALVA) which is included to reduce the half-life of the protein (Figure 3.7A).

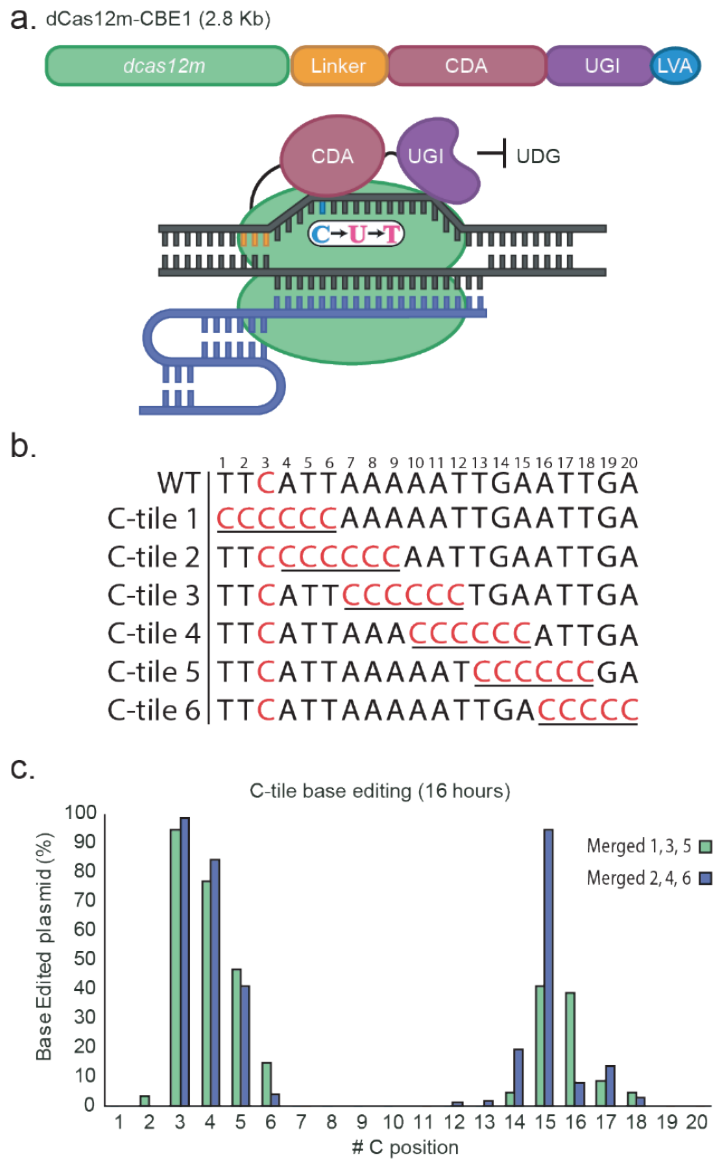


Figure 3.7. A dCas12m base editor achieves C-to-T editing in an atypical window. (A) Schematic of dCas12m-CBE1. dCas12m-CBE1 recognizes a 5' TTN PAM (orange) and binds to its target. Once an R-loop is formed, CDA (bordeaux) deaminases a C to a U, which is converted to a T after replication. UGI (purple) blocks UDG repair. **(B)** Overview of the C-tile targets used to characterize the editing window of dCas12m-CBE1. The wildtype sequence contains a C at position 3 and serves as an internal standard for base editing. C-tile 1 to C-tile 6 plasmids contain six consecutive C's in the sequence and shifts three position toward 3' end until position 20 is reached. **(C)** Deep sequencing results of dCas12m-CBE1 targeting the C-tile plasmids after 16 hours. Data from plasmids of uneven and even numbers were fused for easier data overview corresponding to 'Merged 1, 3, 5' and 'Merged 2, 4, 6', respectively. Y-axis represents base edited plasmids in % of the whole plasmid population and X-axis represents the C position within the protospacer. Unmerged data can be found in Table S3.1.

This architecture is analogous to that of the previously developed Cas9-based BE (*250*). To determine its base editing window, dCas12m-CBE1 was used to edit various pTarget plasmids. We generated six variants of pTarget, each containing six consecutive cytosine bases at different positions of the protospacers (C-tile plasmids; Figure 3.7B). These boxes of six C's shift 3 positions towards the 3' end up to the 20th position (Figure 3.7B), ensuring a complete C coverage on the protospacer. In addition to the C-tiles, a fixed at position 3 (C3) was used as an internal standard for base editing. Once transformed, the *E. coli* cells were grown in M9TG medium for 1-2-3 days. Deep sequencing of the targeted sequence was performed to assess base editing of the whole population. Sequence analysis showed that base editing occurred in each of the six C-tile plasmids, with efficient C3 base editing (>90% editing) in all plasmids (Table S3.1).

Next, the results of the uneven C-tile plasmids (1, 3 and 5) and of the even C-tile plasmids (2, 4 and 6) were merged, revealing a unique base editing window (Figure 3.7C). Interestingly, it was found that dCas12m-CBE1 catalyzed base editing in two different regions within the base editing window instead of the single region for previously described Cas9 or Cas12a-based Bes (*282*, *283*). C-to-T editing occurred both in a PAM-proximal (positions 2-5) and in a PAM-distal (positions 13-19) regions. Positions 3 and 4 were found to have the highest base editing efficiency of >75%. Base editing efficiency for C15 varied between plasmids, 41% for C-tile 5 and 94% for C-tile 6. Such differences were most pronounced in position C15 but were also observed in other positions, such as C6, C16 and C17, and were likely caused by sequence-specific base editing biases, i.e., context dependent base editing.

To improve the focus of the base editing window of dCas12m-CBE1, a series of dCas12m-CBE1 variants were constructed containing different linker length (dCas12-CBE1.A-D) (Figure S3.19A), which targeted C motif plasmids. C motif plasmids contain a tiled C motif (CxxCxxCxxCxxCxxCxxC), starting at every first (C1 motif), second (C2 motif) or third (C3 motif) nucleotide of the protospacer (Figure S3.19B). Reducing the linker length of dCas12m-BE1 showed no substantial decrease in the base editing window (Figure S3.19C). However, substituting the coding region of either CDA or UGI by using a *Homo sapiens* codon optimized sequence led to a decrease in overall base editing efficiency and thereby a narrower base editing window, which is most likely due to reduced protein production of dCas12m-CBE1 (Figure S3.19B and C). In short, we've characterized a Cas12m protein from type V-M that processes its own pre-crRNA, recognizes a 5'TTN- PAM and binds dsDNA. In addition, Cas12m was utilized in constructing Cas12m-based cytidine base editors, which allows for guided single nucleotide C-to-T substitution in a unique base editing window.

3.4 Discussion

Here we describe the characterization of a miniature type V CRISPR-Cas system, subtype V-M (formerly V-U1). The key player of this system is the relatively small Cas12m effector protein (596 amino acids) that has distinct functional features compared to other Cas12 subtypes. Of the six CRISPR-Cas types, Type V is the most diverse CRISPR-Cas type that comprises 13 distinct, well-defined subtypes (A to M) (Figure 3.1, Figure S3.1). Phylogenetic analysis indicates that the effectors of different Cas12 subtypes evolved independently from different groups of transposon-encoded TnpB nucleases that are thought to use an RNA-guided transposition mechanism (*76*, *261*, *262*). The CRISPR-Cas variants that initially were called subtype V-U (*76*) have been reclassified as genuine subtypes after their experimental characterization. They appear to represent intermediate stages of the evolution of type V systems from TnpB. Most of these proteins are relatively small, being about the same size as typical TnpB (400-500 amino acids). In addition, they show high levels of sequence similarity to TnpB, in contrast to the much larger effector proteins of other subtypes, such as Cas12a-d. The latter, fully developed and highly efficient CRISPR effectors are thought to have evolved from TnpB by accretion of various, poorly conserved protein domains (*76*). Different type V variants appear to correspond to different stages on this evolutionary path, which is reflected in the protein sizes. From this perspective, Cas12m seems to represent an early evolutionary stage (Figure 3.1A, Figure S3.1, Figure S3.20).

Structural analyses of Cas12 effectors from different subtypes (*255*) show that the conserved RuvC residues (Figure 3.1B, DED) constitute the catalytic site for target nucleic acid cleavage. Similarly to other nucleases that share the RNase H fold, including bacterial RuvC, Argonaute (PIWI domain) and Cas9 (RuvC-like domain), these negatively charged residues form an essential part of the binding site for a pair of divalent cations (Mg^{2+}/Mn^{2+}) that position a water molecule appropriately for a nucleophilic attack of a phosphorus atom in the backbone of a target nucleic acid, resulting in target cleavage through hydrolysis of the scissile phosphodiester bond (*126*, *284*, *285*). In case of Cas12m, the metal-binding site of the RuvC domain is disrupted by substitutions of one or more of the cation-binding amino acids (Figure 3.1B), and the identity of the substitutions is rather variable among members of the Cas12m subtype (Figure S3.6). In particular, the Cas12m from *M. mucogenicum* contains the triad HDD, instead of the DED catalytic triad that is conserved in all catalytically active TnpB and Cas12 variants. In agreement with this lack of conservation of the catalytic residues, we never detected any target cleavage activity of Cas12m. In all Cas12m systems, however, the strong conservation of closely-spaced two aspartic acid residues at a position that correspond to the location of the third metal binding ligand of the canonical Cas12 RuvC domain (Figure S3.6) may suggest that at least this residue still plays some structural and/or functional role.

Rather than cleavage of the target nucleic acid, we demonstrate both *in vivo* and *in vitro* that Cas12m causes interference through strong crRNA-guided binding of dsDNA targets. Initial comparison of binding kinetics of Cas12m with a catalytically inactive variant of the well-established Cas12a system (dCas12a) did reveal similar dissociation rates (Figure S3.17).

However, when monitoring for longer periods of time, the stability of the Cas12m RNP complex with its dsDNA target appears to be substantially stronger than that of dCas12a (Figure 3.6). Comparison of the performance of the wild type Cas12m (RuvC HDD) with mutant dCas12m (RuvC HAD) reveals interesting differences. When targeting non-transcribed promoter regions, both variants can lead to robust silencing most likely by blocking docking of the RNA polymerase complex. However, when targeting sites in the transcribed coding region, a trend is observed that targeting sites that are more distant from the promoter, results in steadily decreasing silencing efficiency (as previously reported for silencing by other Cas effectors, e.g. dCas9 (*163*). This is observed to some extent for the wild type Cas12m, but much more pronounced so for the mutant dCas12m. In addition, Cas12m variants in which the ZF motif is disrupted behave similarly as the RuvC mutant. These differences in silencing potential appear to correlate with a drop in the target binding stability, as observed in the extended monitoring of dissociation of the Cas12m variants (Figure 3.6). This reveal that despite the lost capacity to cleave target dsDNA, the catalytic inactive RuvC domain of Cas12m plays a role in the crRNA binding of dsDNA. This role of Cas12m-RuvC in target binding could also explain the observation that dCas12m is more sensitive for mismatches than the wild type (Figure S3.13). As previously described for other Cas effectors (e.g. Cas9 (*286*)) derived variants in which target affinity was decreased resulted in increased specificity). The enhanced specificity of dCas12m may be beneficial for certain applications, including targeted base editing (see below).

We show here that Cas12m employs an interference mechanism that is distinct from all CRISPR-Cas systems characterized to date. Rather than target cleavage, Cas12m functions through RNA-guided binding to complementary dsDNA targets of MGEs, that impedes either MGE replication or transcription of essential MGE genes. Next to Cas12k, Cas12m is the second type V effector that lacks target cleavage activity. The two inactivated Cas12 variants are similar mechanistically but distinct biologically: despite the loss of the nuclease activity, Cas12m is likely to maintain a defense function, whereas Cas12k has been repurposed by a transposon to locate a specific site for RNA-guided transposition (*258*, *287*). Several examples have been reported of CRISPR-Cas systems that have been adopted by MGEs: transposons, but also viruses and plasmids (*81*). The detection of anti-plasmid spacers in subtype V-M arrays (Figure S3.2) suggests that control of plasmid replication could be the main function of the V-M systems. In this context, downregulating an MGE by binding to the origin of replication or silencing the expression of an essential gene could be sufficient to control its propagation rate while mitigating the cost of CRISPR in terms of inhibition of horizontal gene transfer and autoimmunity. In addition, the silencing strategy would allow to control the gene expression of MGEs that are integrated in the host genome. Whereas DNA-cleaving CRISPR systems control gene expression by using partial matching crRNA guides (*81*), the DNA-binding Cas12m system can use perfectly matching guides for the same purpose. Given that almost half of the V-M systems are located on plasmids, their additional role could involve inter-plasmid competition (*81*). Interestingly, very recently similar conclusions have been reported on the targeting mechanism of a Cas12c system, that also functions through target binding but not cleaving (*288*).

Some recently discovered Cas12 effectors are not even half the size of the smallest Cas9 or Cas12a proteins. This potentially makes these miniature Cas12 variants appealing candidates for packaging in FDA-approved Adeno-Associated Viruses (AAVs), as delivery tool for *in vivo* genome engineering therapies (*289*, *290*). Indeed, AAV-based delivery to human cells has recently been demonstrated for the V-F systems (*132*). Thus, investigation of the mechanisms of these compact CRISPR-Cas systems is not only contributing to understanding the evolution of anti-MGE defense in prokaryotes, but also holds potential for new developments in biotechnology.

The PAM–dependent DNA-binding ability of Cas12m can be utilized to recruit transcriptional activators or repressors (*166*) as well as base editing enzymes (*282*). As a proof-of-concept, we repurposed Cas12m towards cytidine base editors (CBEs) tailored to the C-to-T substitutions. Compared to Cas9- or Cas12a-based CBEs, dCas12m-CBE1 has a unique base editing window that consists of two regions located at the PAM-proximal (positions 2-5) and PAM-distal (positions 13-19) end of the protospacer. This unique feature of dCas12m-CBE1 can be due the small size of dCas12m, where dCas12m is unable to protect the single-stranded displaced strand and thereby allowing it to be exposed for deamination. A recently developed dCas12f1 adenine base editor has also been reported to base edit both at the PAM-proximal and -distal end, although editing on the PAM-distal was found to have significantly lower editing efficiency (*291*). Moreover, the base editing window of dCas12m-CBE1 (positions 2-5 & 13-19) is distinct from that of the dCas12a-based CBE (position 8-13), and as such is a potentially useful, complementary addition to the base editing toolbox for targeting 5' T-rich PAMs. Depending on the desired genome adjustment in a certain cell, the choice should be made on whether using BEs that include the wild type Cas12m (stronger binding, longer editing) or rather the mutant dCas12m (more specific, less off targeting).

To summarize, we characterized a novel Cas12 subtype with several unique features. Rather than crRNA-guided target cleavage, Cas12m effector proteins appear to control MGE propagation by crRNA-guided target binding, most likely by blocking the invader's replication and/or expression. Apart from revealing fundamental mechanistic features, the small Cas12m protein from type V-M were repurposed to develop CBEs. These findings underscore the incredible diversity of CRISPR-Cas systems and their potential for a wide range of applications.

3.5 Method Details

3.5.1 Sequence analysis

The *Mm*Cas12m sequence (WP_061006603.1) was used as a query for web version of the PSIBLAST program (*292*), which was run against NCBI non-redundant protein sequence database with E-value=1e-4. Phylogenetic analysis was used to verify the V-M family assignments as follows: all sequences identified PSIBLAST search were combined with TnpB-like proteins and previously identified V-U and V-F1 proteins (*255*). Multiple alignment of this sequence set was constructed by MUSCLE v5 (option super5) and a phylogenetic tree was

constructed using approximate likelihood method implemented in the FastTree program. The FastTree program was run with WAG evolutionary model and gamma distributed site rates. All proteins that belong to a well-supported branch containing both previously identified V-U1 sequences and new sequences were considered members of V-M family (Figure S3.19). Extended loci (30 gene up- and downstream of V-M effector gene) were annotated using PSI-BLAST program with collection of position specific scoring matrices (PSSMs) from CDD database (*293*). Multiple alignment of V-M proteins was built using the MUSCLE v3 program (*294*). CRISPR arrays were predicted using web version of the CRISPRfinder (*295*). Spacer matches were identified using the BLASTN (megablast option; E-value=1e-4) program which was run against viral database (*296*), Genbank non-redundant nucleotide database and NCBI whole-genome shotgun contig (WGS) database. Searches were performed using either stand alone or web version of the BLASTN program (*292*). The WGS search was restricted by contigs of the same taxonomic lineage affiliation as the genome from which respective spacer was originated (data not shown).

3.5.2 Bacterial strains and growth conditions

Bacterial strains used for the cloning and propagation of plasmids in the current study are *E. coli* DH5α and DH10β. For protein expression, the *E. coli* Rosetta (DE3) (EMD Millipore) was used. For transformation assays or fluorescence loss assays, *E. coli* BW25113 strain lacking the *lacI, lacZ* genes and the type I-E CRISPR-Cas system was used since it was previously developed for the PAM-SCNR assays (*270*). The *E. coli* strains were routinely cultured at 37 °C and 220 rpm, unless specified, in either Luria Bertani medium (LB) [10 g L^{-1} peptone (Oxoid), 5 g L^{-1} yeast extract (BD), 10 g L^{-1} NaCl (Acros)] or M9TG minimal medium [1xM9 salts (Sigma), 10 g L^{-1} tryptone (Oxoid), 5g L^{-1} glycerol (Acros)]. Plasmids were maintained with ampicillin (100 mg mL^{-1}), chloramphenicol (35 mg mL^{-1}), and/or kanamycin (50 mg mL^{-1}) as needed. Liquid media was supplemented with IPTG as specified.

3.5.3 Plasmid construction

All plasmids and spacers used in this study can be found in Table S3.2 and Table S3.3, respectively. The plasmids constructed and the oligonucleotides (IDT) used in cloning and sequencing are listed in Table S3.4. *E. coli* codon-harmonized *Cas12m* gene was inserted into the plasmid pML-1B backbone (obtained from the UC Berkeley MacroLab, Addgene #29653) by ligation-independent cloning using oligonucleotides to generate a protein expression construct encoding the Cas12m polypeptide sequence (residues 1–596) fused with an N-terminal tag comprising a hexahistidine sequence and a Tobacco Etch Virus (TEV) protease cleavage site.

The three plasmids used for the PAM-SCANR screening platform were based on the previously published protocol (*270*). The *cas12m* and *dcas12m* genes were inserted into the pBAD33 vector backbone under the control of the constitutive J23108 promoter to generate the pCas-Cas12m and pCas-dCas12m plasmids, respectively. The pCRISPR guideRNA plasmid series were generated by inserting a CRISPR array downstream the constitutive J23119 promoter in

pBAD18 backbone. The pTarget-PS plasmid is comprised of the PAM-SCANR NOT gate-based circuit in a pAU66 plasmid backbone.

The pCas-dCas12m and pCas-Cas12m plasmids were constructed using NEBuilder® HiFi DNA Assembly (NEB). The fragments for assembling the plasmids were amplified by PCR using Q5® High-Fidelity 2X Master Mix (NEB). The dCas12m gene fragment was created by site-directed mutagenesis of the aspartic acid in the RuvC domain to an alanine (D485A). Zinc finger mutants pCas-Cas12m-ΔZF and pCas-dCas12m-ΔZF were constructed by NEBuilder® HiFi DNA Assembly (NEB). Backbone pCas-Cas12m was digested with restriction enzymes AvrII and HindIII and Cas12m fragments were amplified by PCR.

More in-depth cloning details of various pCas plasmids can be found in Table S3.5. The pCas-(d)Cas12m-BE1 was constructed using NEBuilder® HiFi DNA Assembly (NEB). DNA fragments used in the assembly were amplified by PCR using Q5® High-Fidelity 2X Master Mix (NEB). The pCas-(d)Cas12m-BE1was then used to construct pCas-(d)Cas12m-BE1.A-D by Golden Gate cloning. The vector and linker were PCR amplified to introduce flanking SapI restriction sites. To enable more straightforward cloning of the other Cas12m-BEs, pCas-RFP-UGI-Entry was constructed. pCas-RFP-UGI-Entry which contains a *rfp* and an UGI gene. The *rfp* gene is flanked BbsI restriction sites, which can be used for Golden Gate cloning and visualization of correctly assembled plasmid by the absence of RFP fluorescence. pCas-RFP-UGI-Entry was used to construct pCas-(d)Cas12m-BE2 and pCas-(d)Cas12m-BE3.

The pTarget plasmids used for base editing (C-tile and C-motif plasmids) were constructed using a fragment of pTarget-divergent digested with the restriction enzymes, AatII and KpnI and subsequent ligating it to a short protospacer sequence. Protospacer sequences containing complementary overhangs were created by annealing two oligonucleotides (Table S3.4).

The pCRISPR plasmids for Cas12m were constructed by restriction-digestion and ligation. By PCR amplification, a BbsI restriction site and a CRISPR repeat was added as an overhang to the vector fragment. The amplified fragment was digested (using KpnI and BbsI enzymes) and ligated to a spacer-repeat sequence generated by annealing two oligonucleotides containing complementary overhangs. Using the same method, a pCRISPR_NT plasmid was created, containing a spacer flanked by BbsI sites. Other CRISPR plasmids containing the different targeting spacers were created using pCRISPR-NT by digestion and ligation. Longer CRISPR-arrays such as the four-spacer CRISPR-array were created by annealing two oligonucleotides to create spacer-repeat fragments. Fragments were design to contain compatible overhangs to other spacer-repeat fragments. Spacer-repeat fragments are ligated together and PCR amplified to yield spacer-repeat-spacer-repeat-spacer flanked by BbsI restriction sites. The amplified linear fragment is then cloned into pCRISPR by digestion and ligation.

The pTarget-GFP plasmid was constructed using BamHI restriction and ligation of a linear P_{laclq} and GFP gene fragment amplified from the pTarget-PS plasmid. pTarget-GFP containing different PAMs were constructed by site directed mutagenesis. The pTarget-operon plasmid

was constructed by digesting the pTarget-GFP plasmid with BamHI enzyme to generate a linear vector which was assembled with an mRFP fragment containing compatible overhangs using the NEBuilder® HiFi DNA Assembly. The pTarget-divergent plasmid was constructed using a fragment of pTarget-GFP digested with the restriction enzymes, AatII and BamHI and subsequent ligated with a mRFP fragment under the control of a Taq promoter.

For testing the mismatch tolerance, targets were ordered as an oligonucleotide pair, which was phosphorylated with T4 PNK and annealed. The backbone pTarget-MM-BsmBI-entry was linearized with BsmBI and ligated to the target adaptors to create series pTarget-MM-[x], where x is the position from 1 to 20 on the protospacer where the mismatch is introduced. A frameshift (pTarget-MM-[FS] was made in the *gfp* by digesting pTarget-MM-[WT] with BstBI, filling in the overhang with Klenow fragment and re-circularizing the plasmid. The CRISPR-array plasmids pCRISPR-MM-[WT] were created using the same method described above.

The pCRISPR plasmids for Cas12a were constructed by restriction digestion of pCas-Cas12a-pCRISPR-RFP with restriction enzyme BbsI. BbsI digestion removes a *rfp* gene flanked by two Cas12a repeats. Spacers are created by annealing two oligonucleotides containing complementary overhangs and subsequently ligated to pCRISPR-Cas12a.

3.5.4 Cas12m protein expression and purification

The purification protocol was adapted from established Cas12a purification methods previously (*297*). Briefly, the *cas12m* gene was heterologous expressed in *E. coli* Rosetta (DE3) (EMD Millipore) and purified using a combination of Ni^{2+} affinity, cation exchange and gel filtration chromatography steps. Three liters of LB growth medium with 100 µg mL^{-1} ampicillin was inoculated with 30 mL overnight culture of *E. coli* cells containing the expression construct. Cultures were grown to an OD600nm of 0.5 - 0.6; expression was induced by the addition of IPTG to a final concentration of 0.2 mM and incubation was continued at 18 °C overnight. Cells were harvested by centrifugation and the cell pellet was resuspended in 50 mL lysis buffer (20 mM Tris-HCl pH 8, 500 mM NaCl, 5mM imidazole, supplemented with protease inhibitors (Roche). Cells were lysed by sonication and the lysates were centrifuged for 45 min at 4 °C at 30,000x g to remove insoluble material. The clarified lysate was applied to a 5 mL HisTrap HP column (GE Healthcare). The column was washed with 10 column volumes of wash buffer (20 mM Tris/HCl pH 8, 250 mM NaCl, 20 mM Imidazole) and bound protein was eluted in elution buffer (20 mM Tris/HCl pH 8, 250 mM NaCl, 250 mM Imidazole). Fractions containing pure proteins were pooled and TEV protease was added in a 1:100 (w/w) ratio. The sample was dialyzed against Dialysis buffer (20 mM HEPES-KOH pH 7.5, 250 mM KCl) at 4 °C overnight. For further purification the protein was diluted 1:1 with 10 mM HEPES KOH (pH 7.5) and loaded on a HisTrap Heparin HP column (GE Healthcare). The column was washed with IEX Buffer A (20 mM HEPES-KOH pH 7.5, 150 mM KCl) and eluted with IEX Buffer B (20 mM HEPES-KOH pH 7.5, 2 M KCl) by applying a gradient from 0% to 50% over a total volume of 60 ml. Peak fractions were analyzed by SDS-PAGE and fractions containing the Cas12m protein were combined, and DTT (Sigma-Aldrich) was added to a final concentration of 1 mM.

The protein was fractionated on a HiLoad 16/600 Superdex 200 gel filtration column (GE Healthcare) and eluted with SEC buffer (20mM HEPES-KOH pH 7.5, 500mM KCl, 1mM DTT). Peak fractions were combined, concentrated to 10 mg mL^{-1}, flash frozen in liquid nitrogen and either used directly for biochemical assays or frozen at -80°C for storage.

3.5.5 Pre-crRNA processing

The pre-crRNA processing assay was conducted with ~varying amounts of Cas12m nuclease and ~100nM pre-crRNA. The assay was conducted in Cas9 Nuclease Reaction Buffer (NEB), in a total volume of 15 µL, at 37ºC for an hour and quenched with 2 µL proteinase K (NEB) at 30 ºC for 30 minutes. Subsequently, the samples were analyzed on a 10% urea-PAGE gel stained with SYBR™ Gold Nucleic Acid Stain (Invitrogen).

3.5.6 Transformation assay

E. coli cells harboring the pCas and pCRISPR plasmids were made chemically competent and were transformed with 2 ng of the pTarget plasmid. After recovery, the transformation mix was diluted 2 µL:200 µL M9TG medium in a 96 well 2 mL master block (Greiner) and sealed using a gas-permeable membrane (Sigma, AeraSeal™) and grown overnight at 37 °C at 900 rpm overnight selecting only for pCas and pCRISPR. After 16h, cells were diluted in PBS and drop plated (3 µL) on LB agar selecting for all three plasmids. Plates are grown overnight at 37 °C and colonies were analyzed the following day.

3.5.7 PAM-SCANR assay

A day prior to sorting, *E. coli* cells harboring the pCas and pCRISPR plasmids were made chemically competent and were transformed with the pTarget-PS plasmid containing the randomized 4N PAM library. After recovery, the transformation mix was used to inoculate 10 mL LB medium (1:100) and grown overnight. The next day, the culture was used to inoculate 10 mL LB medium (1:100) and supplemented with different concentrations (0, 10, 1000 µM) of IPTG and cultured to an OD_{600} of ~0.5. Subsequently, the cultures were diluted 1:100 in phosphate buffer saline (PBS) and GFP-positive cells were sorted using a Sony SH800S Cell Sorter. GFP was excited using a blue laser (485 nm) and detected using a 525/50 filter. Pure cultures of either GFP expressing fluorescent or non-fluorescent cells were used as controls to set the gating and the sensitivity for the forward scatter, side scatter and photomultiplier tubes (PMT). A minimum of 100,000 single cell events were sorted and collected in 5 mL LB medium and grown overnight at 37 ˚C. The following day, the culture was used to inoculate (1:100) 10 mL fresh LB medium and grown for 3 hours. The cultures were diluted 1:100 in PBS and sorted for GFP positive cells. 500,000 single cell events were collected in 1 mL PBS, which was then immediately re-sorted to collect 50,000 single cell events in 5 mL LB medium and grown overnight. The next day, the culture was used to inoculate (1:100) 10 mL LB medium and grown overnight. The next day, plasmids were extracted and sent for deep sequencing.

3.5.8 Fluorescence repression assays

For the silencing assays, *E. coli* cells harboring either the pCas-dCas12m or the pCas-Cas12m and the corresponding target plasmids were made chemically competent and transformed with the pCRISPR library. For the 5'-NTTN PAM determination assays, cells harboring either the pCas-dCas12m or the pCas-Cas12m and the pCRISPR plasmid were made competent and then transformed with the target plasmid. For the mismatch tolerance assays, chemically competent *E. coli* BW225 cells harboring either targeting plasmid pCRISPR-MM-[WT] or non-targeting plasmid pCRISPR-BbsI, and either pCas-Cas12m or pCas-dCas12m were transformed with pTarget-MM-[x]. After recovery, the transformation mix was diluted 2 µL:200 µL M9TG medium in a 96 well 2 mL master block (Greiner) and sealed using a gas-permeable membrane (Sigma, AeraSeal™) and grown overnight at 37 °C at 900 rpm overnight. The next day, the cells were diluted 1:10000 in triplicate in fresh M9TG medium in a 96-wells masterblock and grown overnight at 37°C. Overnight cultures were then used for fluorescence measurements.

3.5.9 Plate reader measurements

Overnight cultures were diluted 1:10 in 200 µL PBS for the mismatch tolerance assays and measured on a Biotek Synergy MX microplate reader a Synergy MX microplate reader. GFP and RFP fluorescence were measured with an excitation of 485 nm and 555 nm, respectively and an emission at 585 nm. GFP and RFP were measured with gain of 75 and 100, respectively.

Fluorescence was calculated as $$\frac{average\left(\frac{Fl_{x_{targeting}}-Fl_{Blank}}{OD600_{x_{targeting}}-OD600_{Blank}}\right)-average\left(\frac{Fl_{FS}-Fl_{Blank}}{OD600_{FS}-OD600_{Blank}}\right)}{average\left(\frac{Fl_{x_{non-targeting}}-Fl_{Blank}}{OD600_{x_{non-targeting}}-OD600_{Blank}}\right)-average\left(\frac{Fl_{FS}-Fl_{Blank}}{OD600_{FS}-OD600_{Blank}}\right)}$$

3.5.10 RT-qPCR analysis

10 mL LB with 50 mg mL^{-1} kanamycin, 34 mg mL^{-1} chloramphenicol and 100 mg mL^{-1} ampicillin was inoculated 1:1000 from a preculture. Cells were grown to an OD600 of 0.6 and cooled down on ice-water. Cells were pelleted and resuspended in 250 µL of 50 mM Tris-HCl pH8, 10 mM EDTA and 10 mM DTT. Cells were then lysed with 250 µL of [0.2 M NaOH and 1% SDS]. Protein, genomic DNA and SDS were precipitated by adding 250 µL [1.8 M potassium acetate and 1.2 M acetic acid]. Debris was pelleted in a microcentrifuge tube and 650 µL was transferred to a new Eppendorf tube. RNA was precipitated by adding 650 µL isopropanol and centrifuging for 5 minutes at maximum speed. RNA pellets were washed with 500 µL of [10 mM Tris-HCl pH8 and 70% ethanol] and dried in a laminar flow cabinet. Pellets were dissolved in 100 µL DNAseI buffer (NEB) with 0.25 µL DNAse I (NEB) and incubated at 37 °C for 30 minutes. First, 300 µL of DNAseI buffer was added and then 200 µL of Roti aqua phenol (Roth). The phases were separated by centrifugation and 300 µL of the aqueous phase was transferred to a new Eppendorf tube. 300 µL of isopropanol was added to the aqueous phase and the mixture was loaded on a silica column (Thermo K0702). The RNA was washed twice with 400 µL [10 mM Tris-HCl pH8, 70% ethanol and 100 mM NaCl]. Finally, the RNA was eluted into 50 µL of [1 mM Tris-HCl pH8, 0.1 mM EDTA]. The RNA was diluted to 1 g/L in water and cDNA was generated with the Maxima H minus (Thermo) reverse transcriptase. RT-qPCR was performed

with the SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad) using cDNA derived from 10 ng of total RNA in a 10 µL reaction.

3.5.11 In vitro TXTL assay

TXTL experiments were conducted in the laboratory of Chase Beisel at the Helmholtz Institute for RNA-based Infection Research in Würzburg, Germany. The TXTL reaction is consisted of myTXTL® master mix, pCas, pCRISPR and p70a-deGFP. The myTXTL® Sigma 70 Master Mix and p70a-deGFP was purchased from Arbor Biosciences. pCas and pCRISPR were plasmids used for in vivo silencing and were prepared by midiprep using the ZymoPURE™ II Plasmid Midiprep Kit (Zymo Research), followed by PCR purification using the DNA Clean & Concentrator-5 (Zymo Research). TXTL reactions were prepared according to (*298*). pCas (4nM) and pCRISPR (4nM) were first pre-incubated in a TXTL reaction for 16 hours in 29^{0}C. 1 ul of the pre-incubated mix was added together with pTarget-eGFP (1mM) in a new TXTL mix with an end volume of 12µL. The final reaction was pipetted into a 96-well plate using a Labcyte Echo 525 acoustic liquid dispensing system. Each well contained a 3µL reaction with four replicates per sample. The 96-well plate was then incubated for 16 hours at 29^{0}C in a Synergy Neo2(Biotek) plate reader. deGFP fluorescence was measured every 3 min with an excitation and emission of 485 and 528, respectively. Also, bandwidth and again were set to 13nm and 60, respectively. For pre-crRNA processing, the pCas and pCRISPR mixture was incubated at 29 °C for five hours in a thermocycler, and total RNA was extracted using Direct-zol RNA MiniPrep kit following the manufacturer's instructions (Zymo Research).

3.5.12 Northern blot

For Northern blotting analysis, 5 µg of each RNA sample obtained from TXTL was put on an 8% polyacrylamide gel (7 M urea) at 300 V for 140 min. RNA was transferred onto Hybond-XL membranes (Amersham Hybond-XL, GE Healthcare) using an Electroblotter using 50 V for 1 h at 4 °C (Tank-Elektroblotter Web M, PerfectBlue) and crosslinked with UV-light for a total of 0.12 Joules (UV-lamp T8C; 254 nm, 8 W). Hybridization occured overnight in 17 ml Roti-Hybri-Quick buffer with 5 µL γ-32P-ATP end-labeled oligodeoxyribonucleotides at 42 °C. The membrane was visualized using a Phosphorimager (Typhoon FLA 7000, GE Healthcare).

3.5.13 Surface plasmon resonance

A 50-nucleotide biotinylated oligo (Table S3.4) containing a DNA target site and its complementary strand were obtained from IDT (Leuven, Belgium) and solubilized in 25 mM HEPES, 150 mM KCl pH 7.5. To create dsDNA, the oligonucleotides were mixed at equal ratio, heated to 95°C and slowly cooled to room temperature. Surface plasmon resonance (SPR) spectroscopy was carried out at 25°C on a Biacore T100 (Cytiva, Marlborough USA). A series S CM5 sensor chip surface was modified with 2500 response units (RU) of streptavidin (Invitrogen) using the amine coupling kit (Cytiva). Flow cell 1 was used as reference. On flow cell 3, 21 RU of dsDNA was immobilized. Ribonucleotide protein complexes were formed by mixing protein with 1.4-fold molar excess of Cas12m and Cas12a PAM-SCANR RNA (Table S3.4) to a final concentration of 500 nM protein in SPR buffer (20 mM Hepes, 150mM KCl,

10 mM $MgCl_2$ and 0.05% Tween 20, pH 7.8). Dilutions of this complex in SPR buffer (500-3.9 nM) were injected over the chip surface for 60 seconds at 50 µl/min and dissociation was monitored for a further 60 seconds (for the complete set) or 5 hours (for an additional 63 nM injection). Flow cells were regenerated using 4 consecutive injections of 4 M $MgCl_2$. Because this not only removed the protein but also the unwound complementary DNA strand, dsDNA was reformed by injection of the complementary oligonucleotide (1 µM) for 100 seconds at 5 µl/min overflow cell 3 before every new cycle. After double reference correction a two-state model was fit to the data using BioEvaluation Software (Cytiva) to obtain kinetic and affinity constants for binding to target DNA. Plots were created using GraphPad Prism version 8.3.1.

3.5.14 Base editing assay

E. coli BW25113 strain lacking the *lacI, lacZ* genes and the type I-E CRISPR-Cas system was used for base editing assay. Cells harboring pCRISPR-C-tile or pCRISPR-C motif plasmids and their corresponding pTarget plasmids were made chemically competent and transformed with the different *Mm*Cas12m base editor (pCas) plasmids. After recovery, the transformation mix was diluted 2 µL:200 µL M9TG medium in a 96 well 2 mL master block (Greiner). Master block was then sealed using a gas-permeable membrane (Sigma, AeraSeal™) and grown overnight at 37 °C at 900 rpm overnight. The following day, the cells were diluted 1:10000 in fresh M9TG medium in a 96-wells master block and grown overnight at 37°C. 20 µL *E. coli* cultures were taken every at time point 16, 24 and 48 hours for C-tile base editing, whereas samples were only taken at 40 hours for C-motif base editing. Base edited region was PCR amplified by using 2 µL cultures in a 50 µL PCR reaction using Q5® High-Fidelity 2X Master Mix (NEB). Amplified fragments were purified using DNA Clean & Concentrator™-5 (Zymo Research) and sent for sequencing.

3.6 Acknowledgements

We would like to thank Sanne Klompe, Jasper Groen, Yuxin Zhang, Patrick Barendse for their technical assistance and Christian Sudfeld for his assistance in the cell sorting experiments. J.v.d.O. is supported by the Dutch Research Council (NWO) through a NWO/TOP grant (714.015.001) and a NWO/BBOL grant (737.016.005). K.S.M. and E.V.K. are supported by the intramural program of the U.S. Department of Health and Human Services (to the National Library of Medicine). J.H.G.L. and C.L. are supported by the gravitation program CancerGenomiCs.nl from the Netherlands Organisation for Scientific Research (NWO), part of the Oncode Institute, which is partly financed by the Dutch Cancer Society. C.L.B. is supported by a Consolidator grant from the European Research Council (865973).3

3.7 Competing interests

A patent application has been filed related to this work, with W.Y.W and J.v.d.O. as inventors. W.X.Y and D.A.S. are employees and shareholders of Arbor Biotechnologies, Inc.

3.8 Corresponding author

Correspondence and requests for materials should be addressed to J.v.d.O.

3.9 Supplementary materials

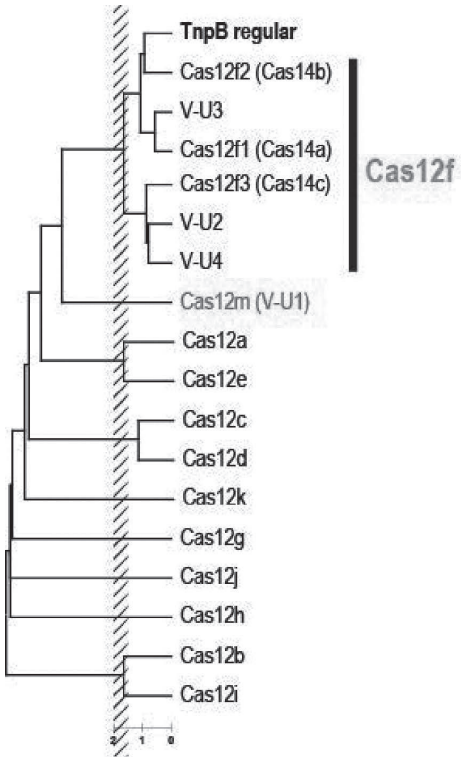


Figure S3.1. UPGMA dendrogram showing similarity between different families of Type V effectors. The dendrogram was built using the UPGMA (unweighted pair group method with arithmetic mean) method and is based on matrix of HHalign (*299*) scores calculated for all against all pairwise alignments with the length coverage 33% or more. The alignments for the respective families were taken from (*255*), except for the Cas12m family for which an updated alignment (104 proteins) was used. The dashed rectangle corresponds to the tree depth D between 1.5 and 2 (D = 2 roughly corresponds to the pairwise HHsearch similarity score of exp(2D) ≈ 0.02 relative to the self-score) and reflect the tree depth where the subtype assignment is uncertain and subject for additional consideration.

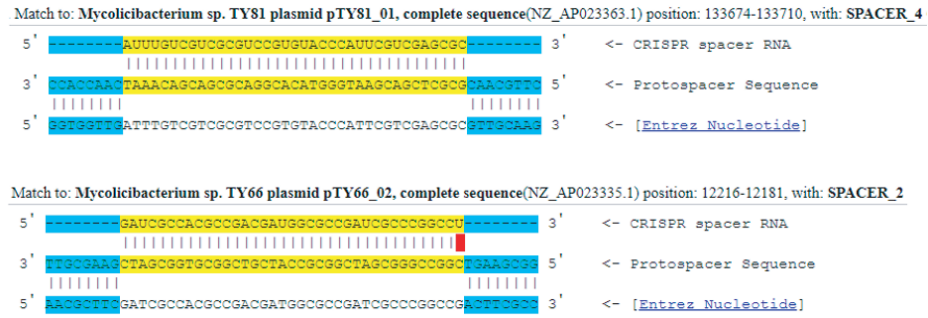


Figure S3.2. Native spacers matching protospacers using CRISPRtarget. Native spacers were analyzed with CRISPRtarget (http://crispr.otago.ac.nz/CRISPRTarget/crispr_analysis.html) using default settings. Red squares represent mismatches.

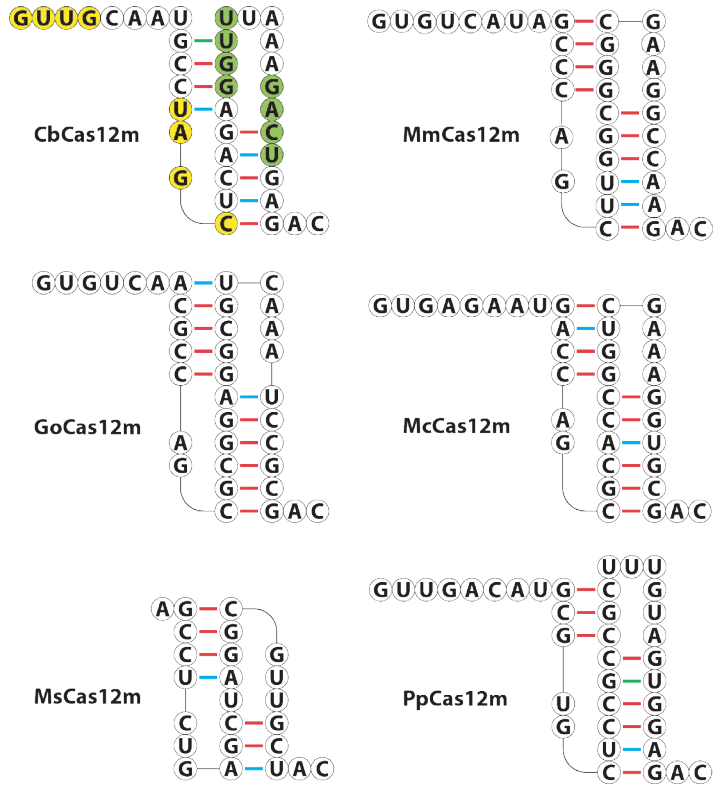


Figure S3.3. Type V-M CRISPR repeats from different bacteria. *Fn*Cas12a and *As*Cas12a crRNA structures are based on Xtal structures. Putative Cas12m pseudoknot structures in the CRISPR RNA repeat regions as predicted by Vsfold (*300*) except in case of *Cba*Cas12m the predicted base pairing deviates from the structure shown here. *Fn*Cas12a: *Francisella tularensis* subsp. *novicida* U112 Cas12a; *As*Cas12a: *Acidaminococcus* sp. BV3L6 Cas12a; *Cb*Cas12m: *Clostridiales bacterium* DRI 13 Cas12m; *Mm*Cas12m: *Mycolicibacterium mucogenicum* CCH10-A2 Cas12m: *Mc*cas12m: *Mycobacterium conceptionense* MLE Cas12m; *Ms*Cas12m: *Meiothermus silvanus* DSM 9946 Cas12m.

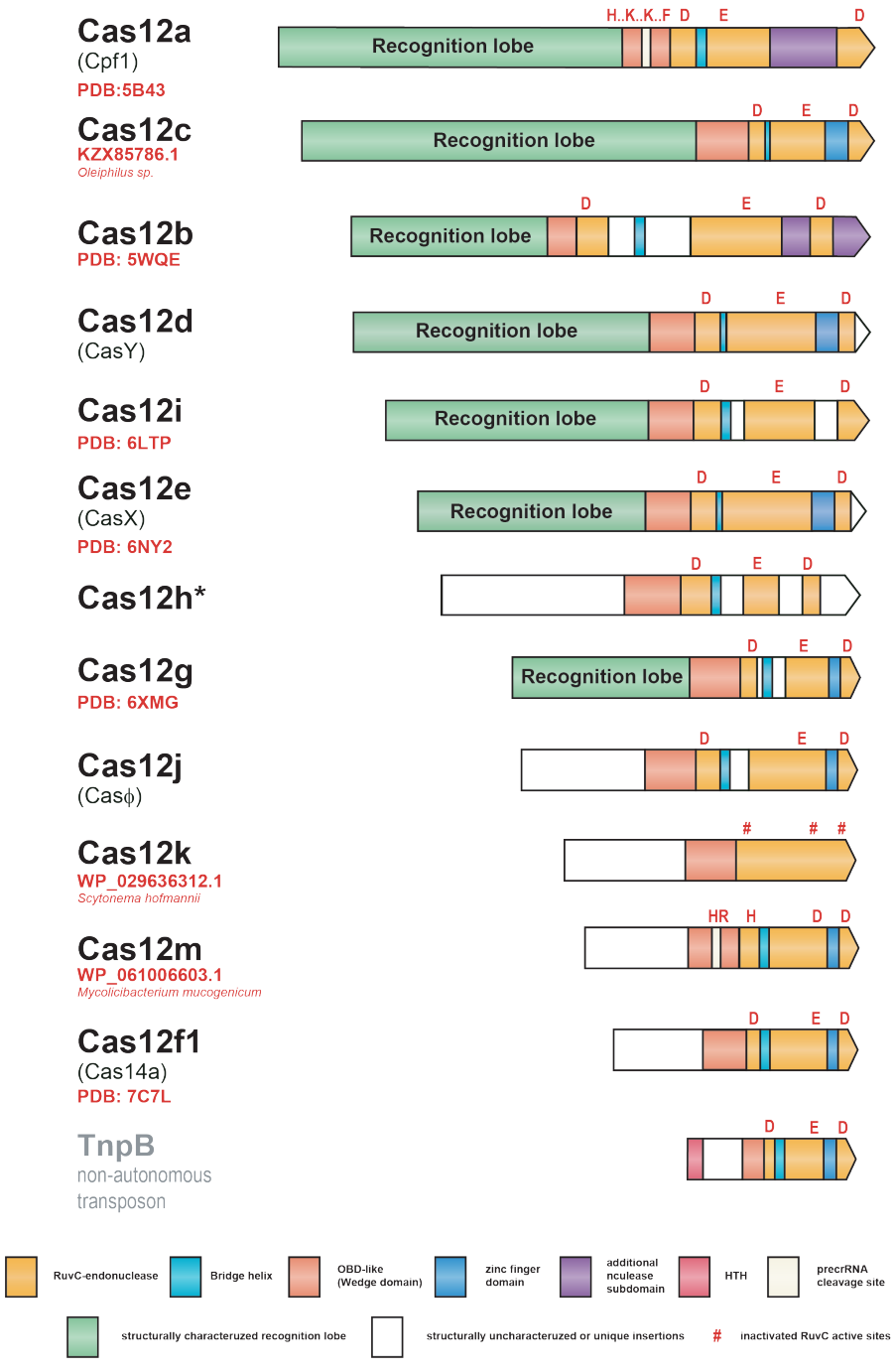


Figure S3.5. Domain architecture of type V effector proteins.

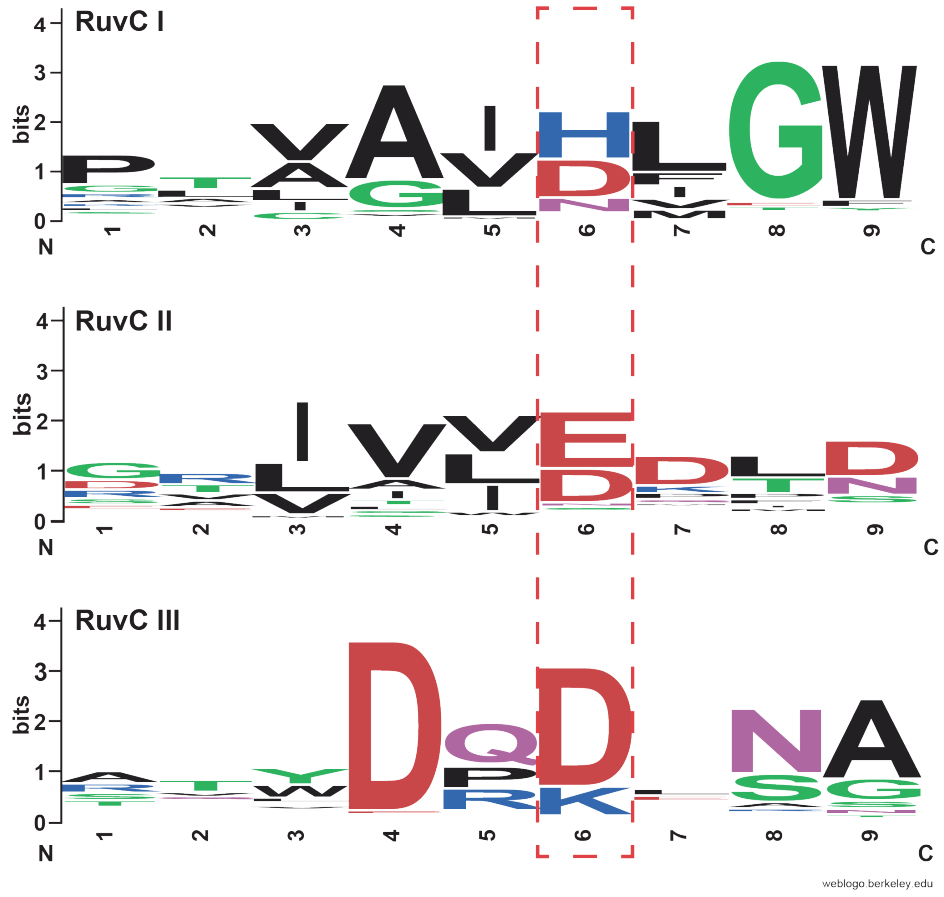


Figure S3.6. Web logo of conserved amino acid residues found in RuvC I, II and III using 104 Cas12m sequences. The logos was built using Web Logo site (https://weblogo.berkeley.edu/logo.cgi) for alignment of 104 Cas12m proteins with default parameters.

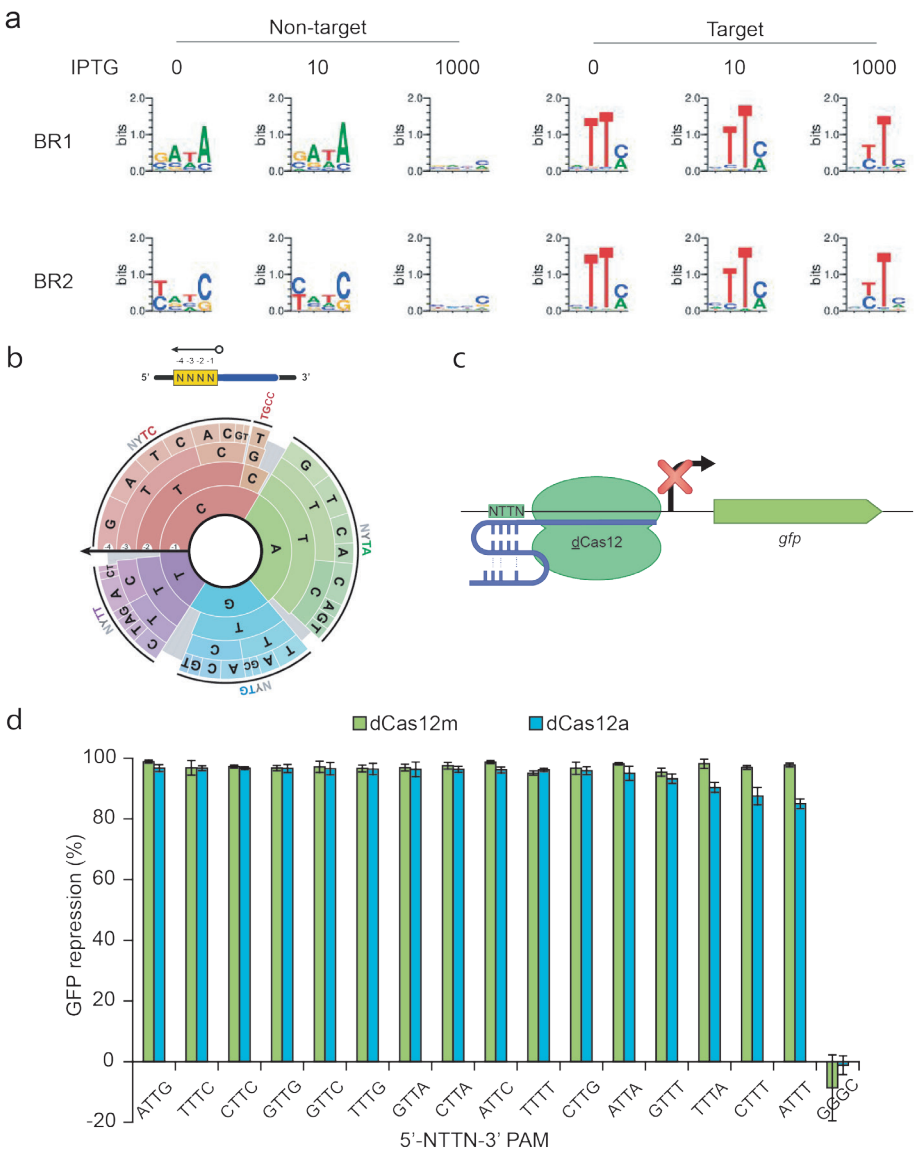


Figure S3.7. Cas12m PAM determination. (A) Deep sequencing analysis of PAM-SCANR after FACS sorting. Plasmids from the FACS-sorted cells were extracted and sequenced to determine functional PAM sequences. Sequence logo for the Cas12m PAM at different IPTG concentrations (0, 10, 1000 µM) as determined by NGS sequencing. NT: non-targeting, T: targeting, BR1 and BR2 are two independent biological replicates. Letter height at positions -4, -3, -2, -1 (left to right) at the 5' end of the protospacer reflects the degree of conservation at each position. **(B)** PAM-wheel representation of enriched PAMs in 1000 µM IPTG. Arrow indicates sequences moving from inner to outer wheel. Colors indicate relative frequency of the innermost nucleotide (-1 position). The area of the sector within the PAM wheel is directly proportional to the relative PAM enrichment in the library. **(C)** Schematic of the pTarget-GFP encoding the *gfp* gene. The protospacer flanked by 5-NTTN-3' PAM upstream of the promoter is targeted by the dCas12m and dCas12a proteins using the respective crRNAs. **(D)** GFP repression detected in

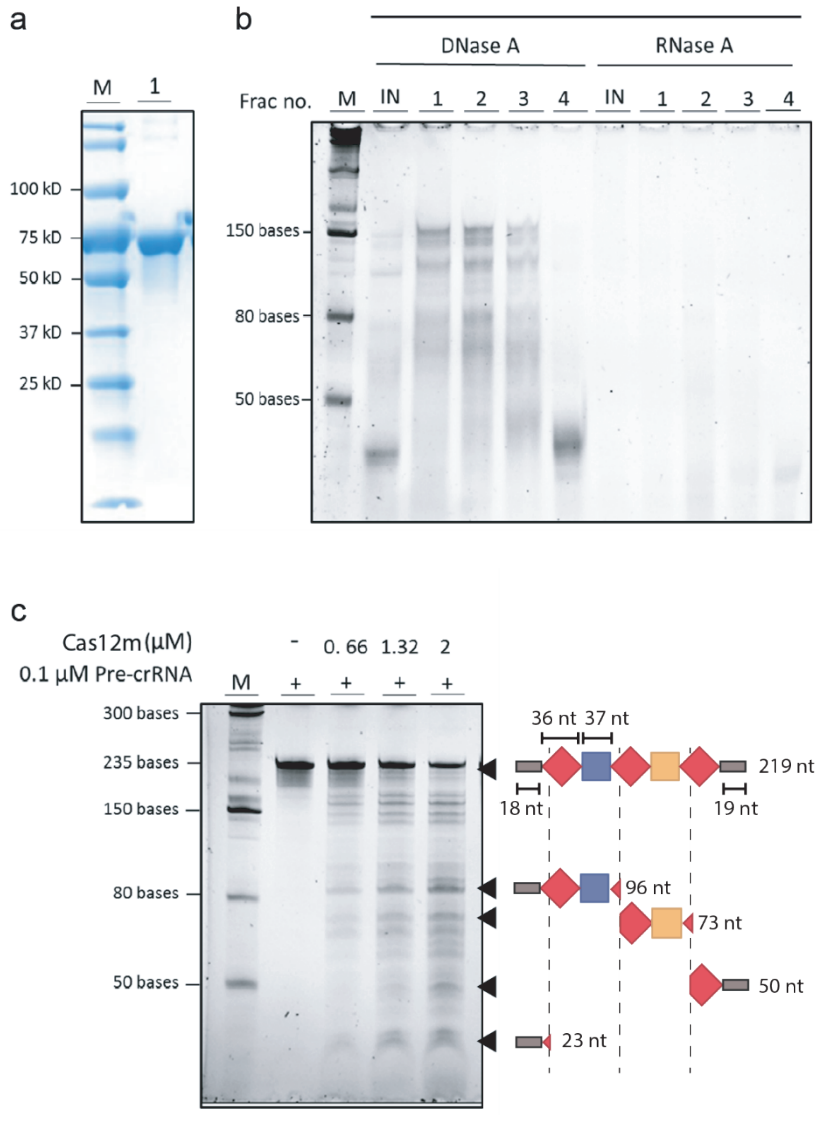


Figure S3.8. Co-purified nucleic acids and pre-crRNA processing by Cas12m. (A) Coomassie blue stained SDS-PAGE gel in which the purified Cas12m protein (66.2 kD) is visualized**. (B)** Co-purified nucleic acids from Cas12m treated with enzymes as indicated. M: low range ssRNA ladder (NEB), IN: input fraction for Size Exclusion Chromatography (SEC) 1-4: different fractions from the SEC purification. **(C)** 10% Urea-PAGE gel on which the processed pre-crRNA transcripts were resolved. RNA was visualized after staining with SYBR-gold. M: low range ssRNA ladder (NEB).

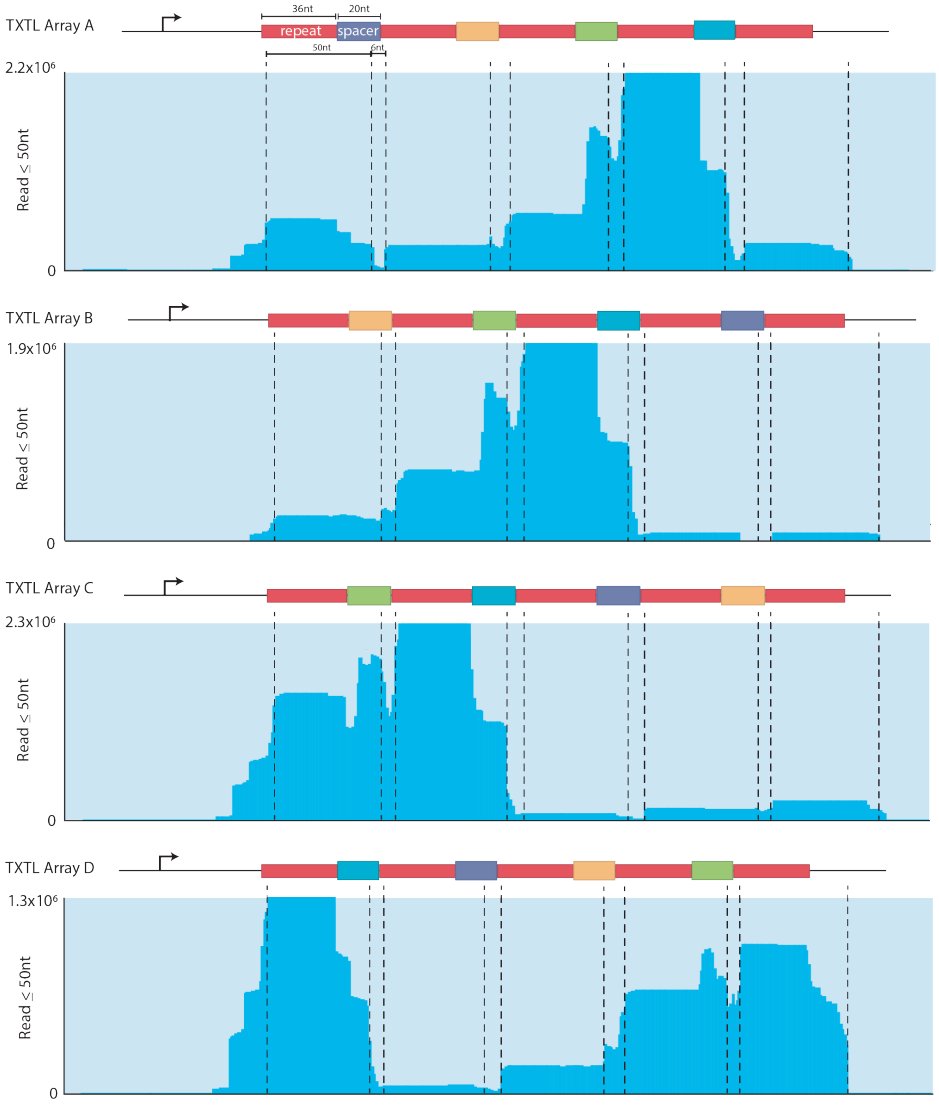


Figure S3.9. Co-purified nucleic acids and pre-crRNA processing by Cas12m from a TXTL reaction. Array (A-D) and *Mm*Cas12m were co-expressed in TXTL, followed by small-RNA isolation and RNA-seq analysis. Reads ≤50 nt were mapped on the expression construct (pCRISPR).

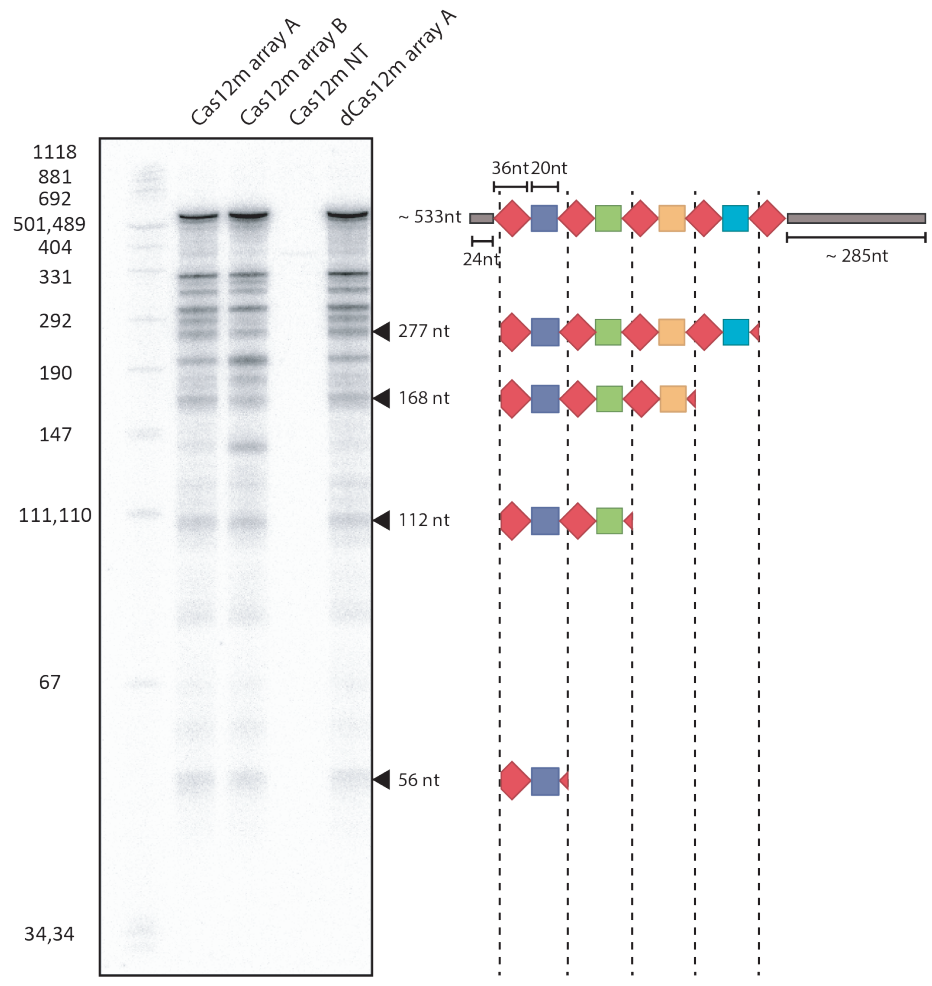


Figure S3.10. Processing by Cas12m and dCas12m by northern blot analysis in TXTL. A plasmid expressing the Cas12m CRISPR-array containing four spacers were incubated with a plasmid expressing Cas12m or dCas12m in TXTL. RNA was visualized by northern blot, using a probe that binds to the first spacer of array A (purple). Array B is similar to that of array A, only the order of spacers a shifted (yellow-green-blue-purple).

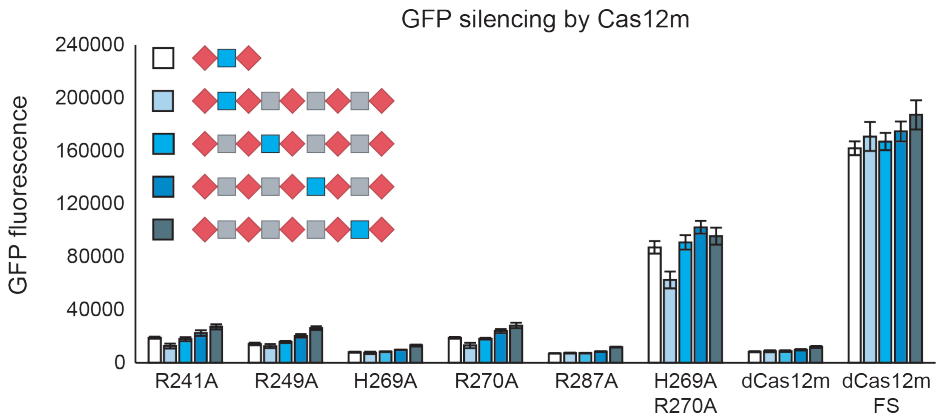


Figure S3.11. GFP silencing assay by single-spacer and four-spacer CRISPR-array using Cas12m processing mutants. Blue spacer indicates the spacers that targets the promoter region of a constitutively expressing GFP (Figure S3.7B). dCas12m and a dCas12 frameshift (FS) was used as positive and negative controls, respectively.

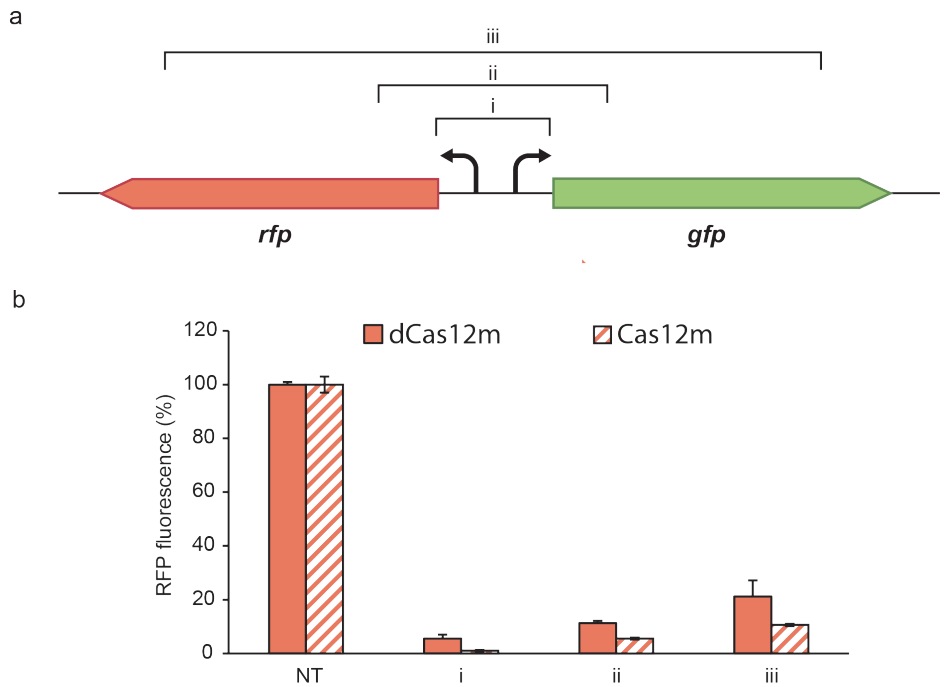


Figure S3.12. Cas12m can be used for multiplex transcriptional silencing. (A) Schematic of the pTarget-divergent including the *rfp* and *gfp* genes under the transcriptional control of two different constitutive promoters, P_{taq} and P_{laclq}. i to iii indicate the crRNA spacer pairs used in the pCRISPR array plasmid to target the *gfp* and *rfp* using the Cas12m and DCas12m proteins. **(B)** RFP fluorescence detected in the cells upon DCas12m or Cas12m targeting using the crRNA spacer pairs is shown on the Y-axis and the different mismatches are shown on the X-axis (n = 3; error bars represent mean ± SEM). NT refers to a non-targeting spacer **(C)** GFP fluorescence detected in the cells upon DCas12m or Cas12m targeting using the crRNA spacer pairs is shown on the Y-axis and the different mismatches are shown on the X-axis (n = 3; error bars represent mean ± SEM). NT refers to a non-targeting spacer.

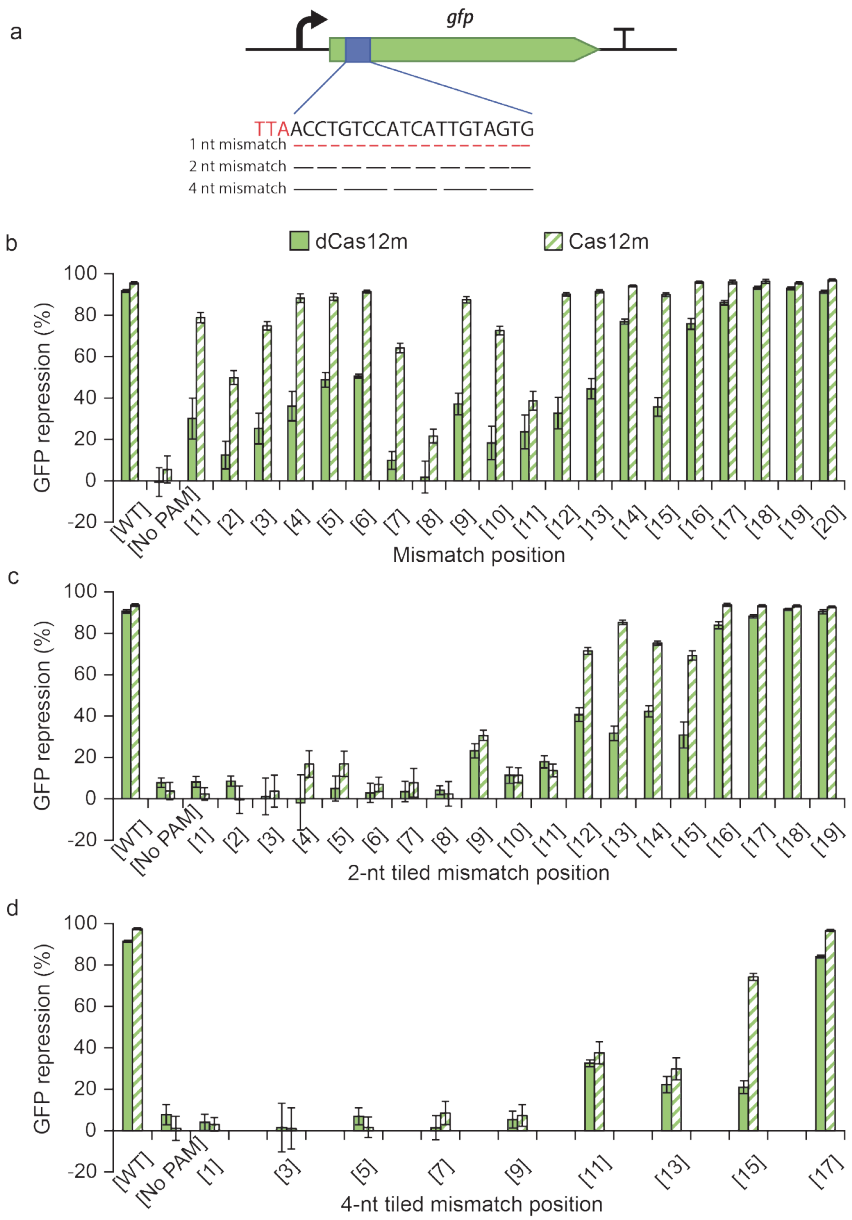


Figure S3.13. Tolerance of Cas12m to mismatched crRNAs. (A) Schematic gfp silencing to assess mismatch tolerance. Mismatches are tiled through the protospacer in one, two or four nucleotides. **(B)** Comparison of the mismatch tolerance of Cas12m with dCas12m for single mismatches across the protospacer sequence. **(C)** Comparison of the mismatch tolerance of Cas12m with dCas12m for 2-nucleotide mismatches tiled across in the target sequence. **(D)** Comparison of the mismatch tolerance of Cas12m with dCas12m for 4-nucleotide mismatches tiled across in the target sequence. GFP repression detected in the cells upon dCas12m or Cas12m targeting is shown on the Y-axis and the different mismatches are shown on the X-axis (n = 3; error bars represent mean ± SEM). No PAM refers to a spacer targeting protospacer next to GGGC motif (non-functional PAM).

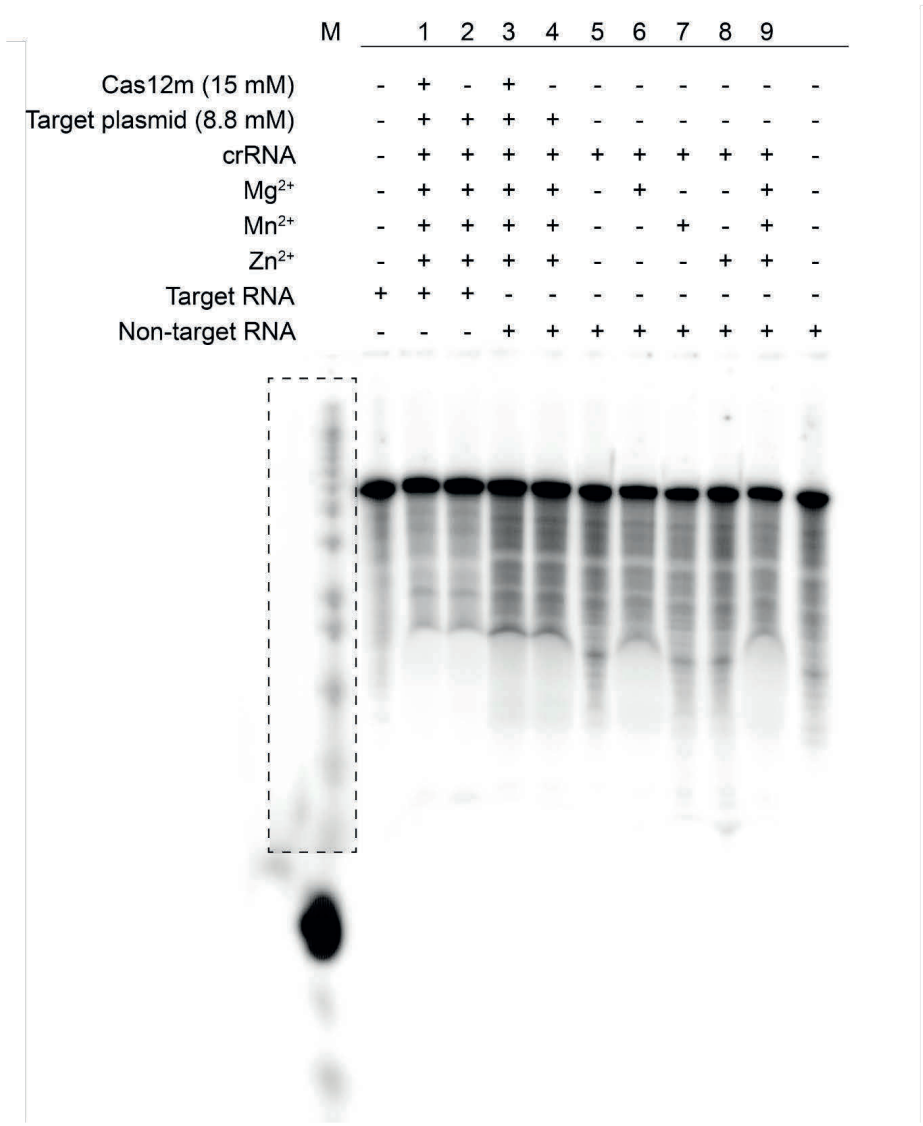


Figure S3.14. Test for in vitro dsDNA activated RNA cleavage by Cas12m. Urea-PAGE assessing the ability of Cas12m protein incubated with a crRNA and an activator target DNA to cleave a [γ-^{32}P] ATP labelled target or a non-target substrate RNA. Dashed rectangle line indicates different exposure.

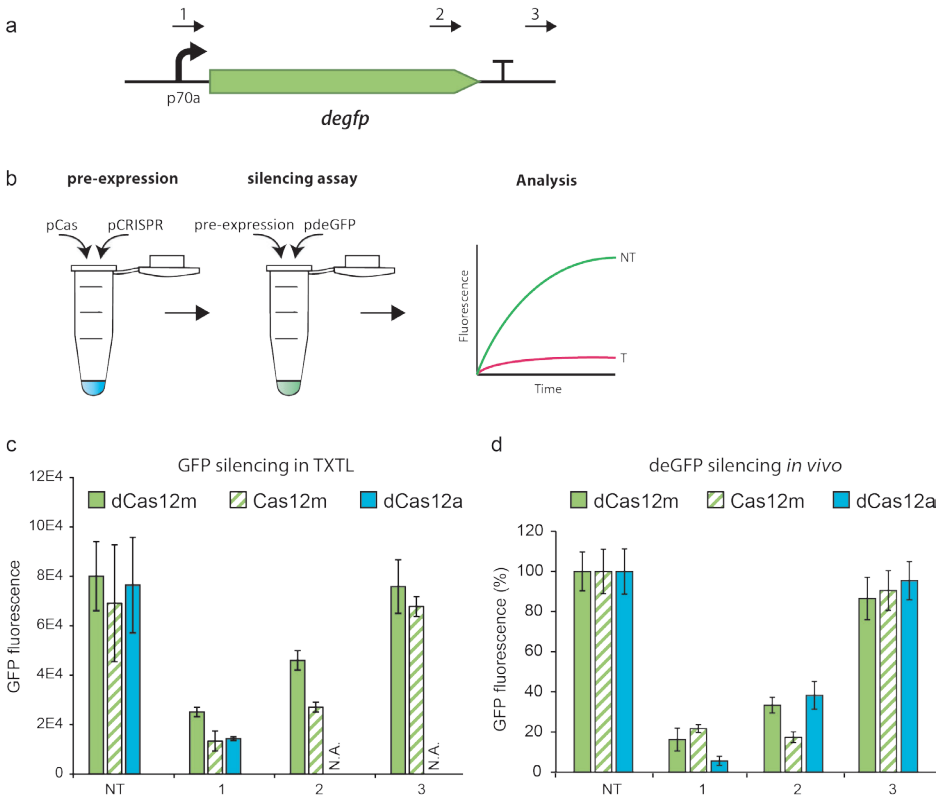


Figure S3.15. Cas12m targeting deGFP in a TXTL system and in vivo. (A) Workflow of deGFP silencing in TXTL. pCas and pCRISPR are pre-expressed in a TXTL reaction, which is then used in a subsequent TXTL reaction containing pdeGFP. The final TXTL reaction is incubated overnight and GFP fluorescence is measured over time to assess GFP silencing. **(B)** Schematic of the pdeGFP regulated by P70a promoter. The arrows indicate the crRNAs used for targeting by Cas12m and dCas12m proteins (crRNA 1, 2, and 3). **(C)** GFP fluorescence detected in TXTL upon dCas12m and Cas12m targeting using the individual spacers (n = 4; error bars represent mean ± SEM). NT refers to a non-targeting spacer. FndCas12a was used as a positive control with crRNA NT and crRNA 1. N.A.= not available. **(D)** GFP fluorescence detected in in cells upon dCas12m, Cas12m and dCas12a targeting deGFP using spacers 1,2 and 3 (n = 3; error bars represent mean ± SEM). NT refers to a non-targeting spacer.

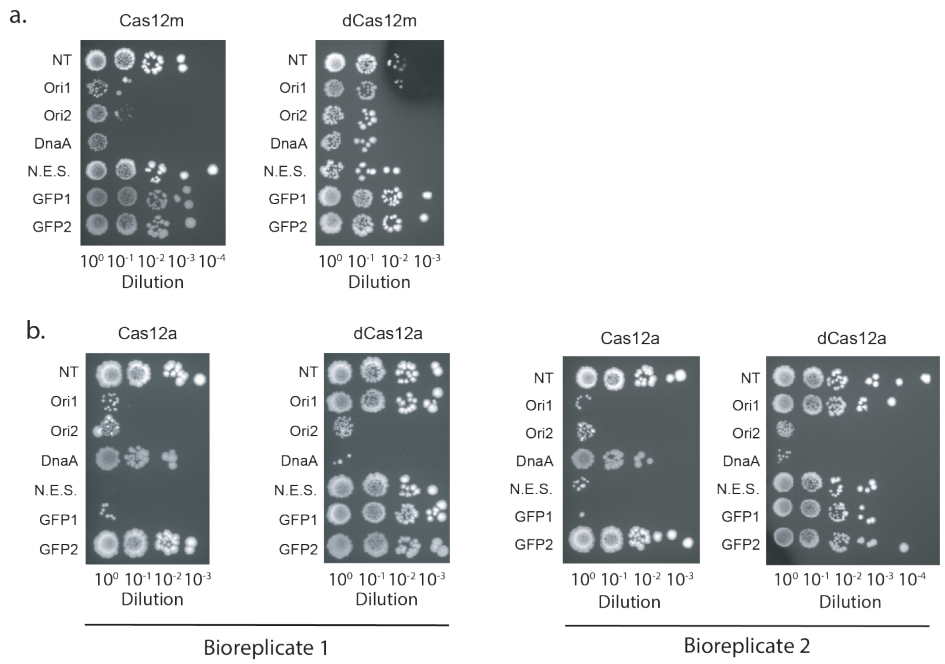


Figure S3.16. Drop plating of a transformation assay using Cas12m, dCas12m, Cas12a and dCas12a. Rows represent spacers used for targeting (Figure 3.6A) and columns represent sample dilutions. N.E.S. = non-essential sequence. **(A)** Transformation assay of Cas12m and dCas12m. Transformation assay was performed using biological duplicates (Figure 3.6). **(B)** Transformation assay of Cas12a and dCas12a Transformation assay was performed using biological duplicates.

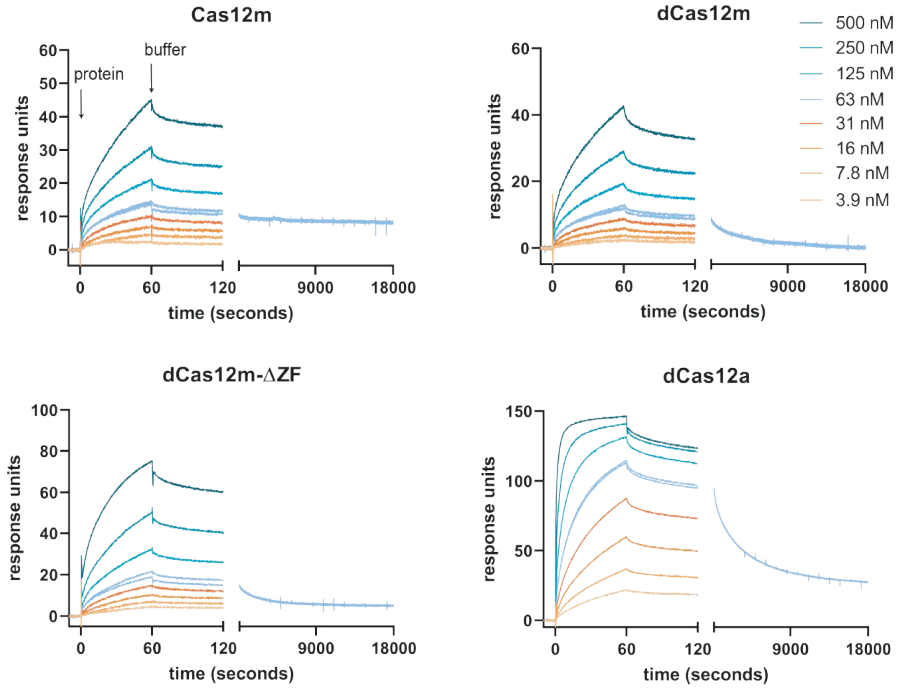


Figure S3.17. Binding of PAM-SCANR-bound Cas12m, dCas12m, dCas12m-ΔZF and dCas12a to dsDNA monitored using Surface Plasmon Resonance. Increasing concentrations of protein (colored curves) were injected and association was monitored for 1 minute, followed by 1 minute dissociation for the complete concentration series, or 5 hours for the 63 nM condition.

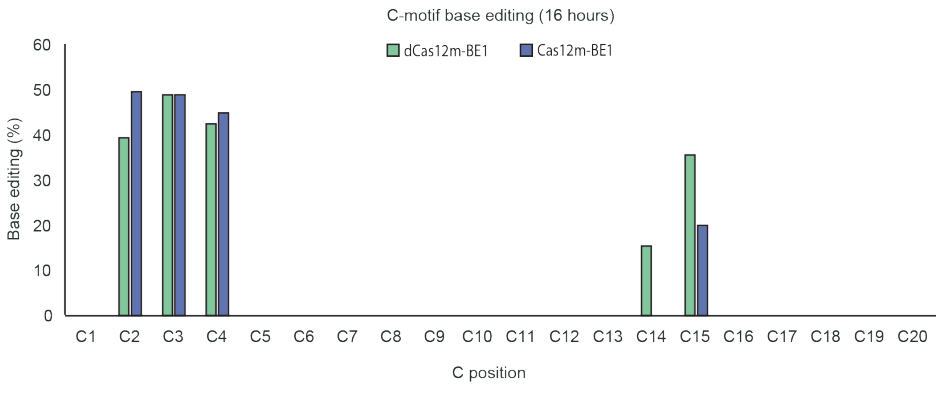
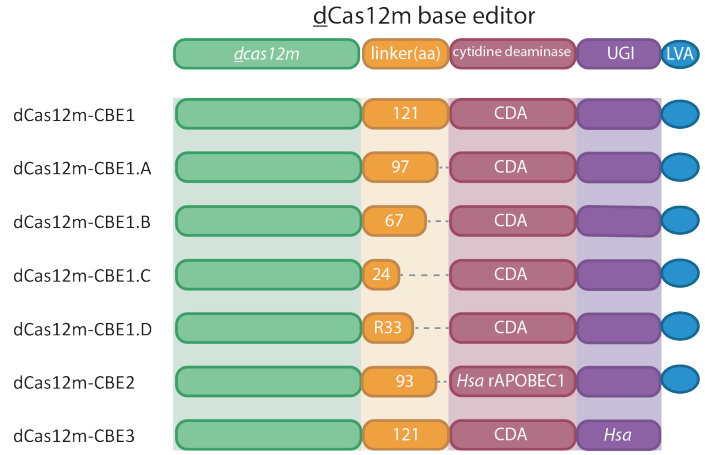


Figure S3.18. Base editing by dCas12m-CBE1 and Cas12m-CBE1. Data was obtained by merging three data sets obtained from base editing C-motif plasmids (Figure S3.18B). Base editing % was determined by sanger sequencing C-motif plasmids after sixteen hours and subsequent analysis using EditR (*301*).

a.



b.

c.

Construct	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
MmuBE_E1	0	85	96	100	37	0	0	0	0	0	0	0	28	55	99	99	17	1	48	0
MmuBE_E1.A	0	78	38	100	38	0	0	0	0	0	0	0	24	41	82	83	0	0	13	0
MmuBE_E1.B	0	0	100	100	0	0	0	0	0	0	0	0	19	34	81	45	0	24	1	0
MmuBE_E1.C	0	67	72	100	38	0	0	0	0	0	0	0	27	38	84	83	0	2	47	0
MmuBE_E1.D	0	99	65	100	42	0	0	0	0	0	0	0	3	52	82	67	0	1	10	0
MmuBE_E2	0	0	81	100	0	77	49	0	0	0	0	18	23	0	0	0	0	28	1	0
MmuBE_E3	0	100	22	19	75	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0

Figure S3.19. Characterizing various dCas12m-CBEs. (A) Schematic of different dCas12m-CBEs. All dCas12m-CBEs consist of a dCas12m (green), linker (orange), cytidine deaminase (bordeaux), UGI (purple). *Hsa* indicates *Homo sapiens* codon optimized. **(B)** Base editing targets consisting of a C on every first, second and third position of each trinucleotide. These plasmids were named C1, C2 and C3 motif, respectively. **(C)** Heat map representing % of base edited C's using different variants of dCas12m-BEs. Data was obtained by merging C1, C2 and C3 motif data. Base editing % was determined by sanger sequencing C-motif plasmids after sixteen hours and subsequent analysis using EditR (*301*).

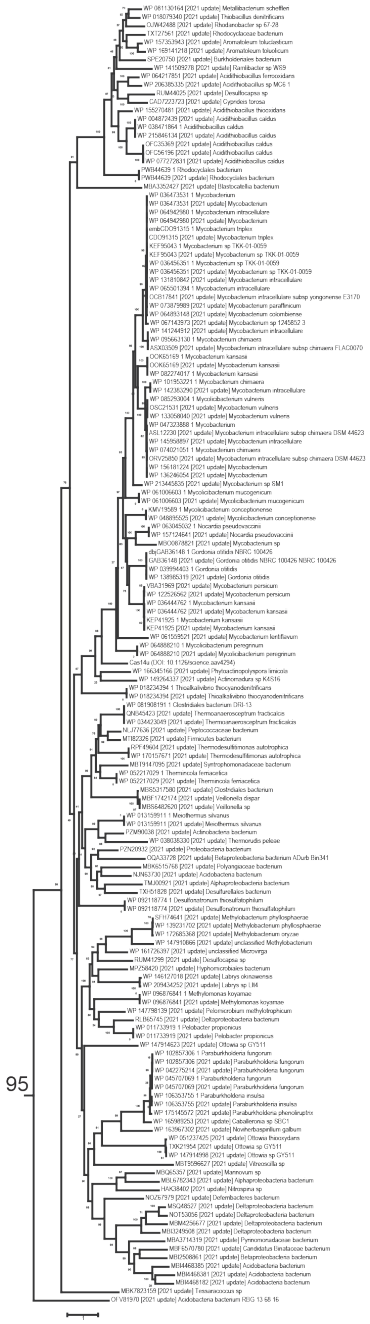
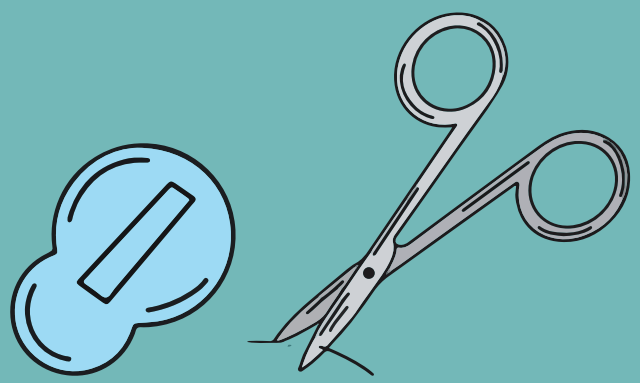


Figure S3.20. A subtree corresponding to V-M protein family. The subtree was exported from a large phylogenetic tree containing 9613 TnpB-like sequences and previously identified V-U and V-F proteins (see Materials and Methods for details). Sequences retrieved in this work designated by words “2021 update”. Support values calculated using FastTree program are shown for each branch.

Additional supplementary materials, tables and figures of this chapter can be accessed via

https://figshare.com/s/e9dcf65a833a8890af83



CHAPTER 4

Exploring the potential of *Mm*Cas12m as gene silencing tool in *Saccharomyces cerevisiae*

Belén Adiego-Pérez[1]
Wen Y. Wu[1]
Kirsten Dekkers[1]
Guus Verver[1]
Paola Randazzo[2]
Pascale Daran-Lapujade[2]
Sarah D'Adamo[3]
John van der Oost[1]

[1] Laboratory of Microbiology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands

[2] Department of Biotechnology, Delft University of Technology, Van der Maasweg 9, 2629 HZ Delft, The Netherlands

[3] Bioprocess Engineering & AlgaePARC, Wageningen University and Research, P.O. Box 16, 6700 AA Wageningen, The Netherlands

4.1 Abstract

CRISPR interference has previously been described for gene regulation purposes in the yeast *Saccharomyces cerevisiae* by using dCas9 or dCas12a. Recently several CRISPR-associated proteins have been discovered that potentially could be used as novel gene silencing tools. In this study, we assessed the potential of the recently characterized bacterial *Mm*Cas12m DNA-binding protein as a CRISPR interference tool in the yeast *S. cerevisiae*. We also compared its ability as gene silencing tool with *Sp*dCas9 and *Fn*dCas12a. We observed gene silencing activity when using the native *Mm*Cas12m in *S. cerevisiae*, albeit less efficient than the well-established gene silencing systems. Further optimization is required to develop *Mm*Cas12m into an efficient CRISPR interference tool.

4.2 Introduction

Gene regulation is a powerful tool to study the function of genes in a given organism (*162*, *166*, *231*, *302*, *303*). Moreover, modulation of gene expression is a key aspect in metabolic engineering (*304*, *305*). A wide range of molecular tools has recently become available to efficiently achieve this task (*166*, *169*). In this work we set out to develop novel CRISPR-Cas tools for gene regulation purposes.

In the last decade, gene regulation tools based on CRISPR-Cas systems have been developed for either gene activation (CRISPRa) or gene interference (CRISPRi) purposes (*164*, *306*, *307*). In prokaryotes, these CRISPRa/CRISPRi systems are based on the use of a deactivated Cas protein (dCas) and a guide RNA targeting a specific location. By directing dCas9 or dCas12a to the promoter region of the coding sequence of the gene-of-interest, transcriptional repression can be achieved by steric hindrance, in which the Cas protein blocks binding of the RNA polymerase and/or of transcription activators (*162*, *163*). On the other hand, to increase the expression of a target gene, the fusion of the deactivated Cas proteins and the omega (ω) subunit of the RNA polymerase complex has been proven sufficient in bacteria (*163*, *169*). In eukaryotes such as the yeast *Saccharomyces cerevisiae*, directing dCas9 or dCas12a to the promoter region has frequently been demonstrated not to be sufficient for efficient transcriptional silencing (*104*, *162*). However, fusing transcription repressor domains to the deactivated versions of the Cas endonucleases appears to be a successful strategy (*166*). Fusing either the mammalian transcriptional repressor Mxi1 domain or the Krüppel-associated box (KRAB) domain to the C-terminus of dCas9 or dCas12a are the most commonly applied strategies for CRISPR-based silencing in eukaryotes (*104*, *164*). Similarly, fusion products between deactivated Cas proteins and the VPR domain have allowed for CRISPRa in eukaryotic organisms (*308–310*).

To date, *Sp*dCas9 has undoubtedly been the most extensively used Cas protein for gene regulation purposes in a wide range of organisms (*311*, *312*). That is also true for the model organism *S. cerevisiae* (*313*, *314*). Recently, a dCas12a tool generated through the fusion of this Cas protein to the Mxi1 repressor domain has been developed (*112*). Although gene repression levels were not compared to dCas9-Mxi1 repression levels, the authors generated a novel gene repression tool by making use of some inherent advantages of the Cas12a protein: the preference for a T-rich PAM and its pre-crRNA processing capacity allowing for multiplexed gene regulation. At the same time, the authors presented the application of a non-dCas9-based gene regulation tool and opened the door to explore for novel Cas candidates.

In this chapter, we evaluated the capabilities of a novel Cas protein from the recently characterized CRISPR-Cas type V-M system (previously called type V-U1) from *Mycolicibacterium mucogenicum* CCH10, consisting of the *Mm*Cas12m effector protein (Figure 4.1A). *Mm*Cas12m is a 596 amino acid protein with an inactive RuvC domain. *In silico* analysis based on detection of sequences with a strong complementarity to the direct repeat sequence yielded no predicted tracrRNA-like sequences in the adjacent DNA sequences to the CRISPR

locus (*315*) (Chapter 3). This fact, together with small RNAs being co-purified when expressing the protein together with a small CRISPR array in *E. coli* suggested that *Mm*Cas12m is able to process its own pre-crRNA into mature crRNAs, similar to what has been reported or suggested for several other type V-CRISPR systems: Cas12a/h/i/j (*122*, *129*, *259*). This observation was confirmed in an *in vitro* set-up in which purified recombinant *Mm*Cas12m protein and a minimal CRISPR array (repeat-spacer-repeat-spacer-repeat) yielded processed intermediates and seemingly mature crRNAs (*316*). Through an adapted PAM-SCANR assay (*270*), a 5'-NTTN-3' PAM similar to that of other characterized type V effectors could be identified (*129*, *272*, *315*). However, despite numerous attempts, no dsDNA cleavage could be detected neither in the heterologous *E. coli* host nor under *in vitro* conditions. Further experiments showed that *Mm*Cas12m is able to target and bind dsDNA both *in vivo* (in *E. coli*) and *in vitro*. This dsDNA binding activity allowed *Mm*Cas12m to protect *E. coli* against invading plasmids by impairing plasmid replication. Experimental set-ups in *E. coli* indicated singleplex and multiplex crRNA-guided *Mm*Cas12m-mediated transcriptional silencing through dsDNA binding (*317*) (Figure 4.1B).

In this study, we postulated that *Mm*Cas12m would be able to silence gene expression in a eukaryotic system, such as the yeast *Saccharomyces cerevisiae*. The inherently deactivated RuvC domain in *Mm*Cas12m and its inability to cleave dsDNA allows for the use of the native *Mm*Cas12m sequence to test this hypothesis. *Mm*Cas12m presents the same advantages as dCas12a regarding its capacity to process its own pre-crRNA and the requirement of an AT-rich PAM. This last characteristic is a useful feature for silencing promoters as they generally are AT-rich.

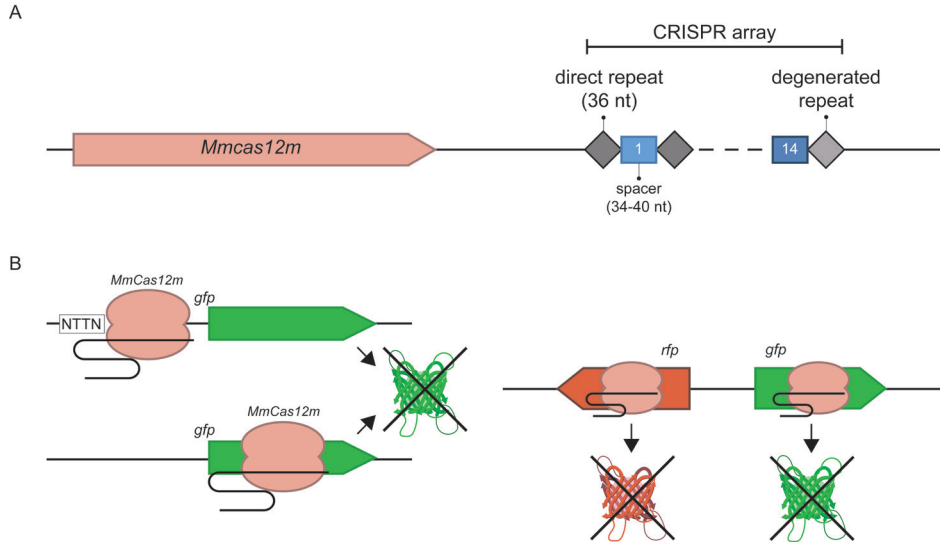


Figure 4.1. *Mycolicibacterium mucogenicum* CCH10-A2 type V-U1 CRISPR-Cas system. (A) Organization of the CRISPR-Cas locus of *Mycolicibacterium mucogenicum* CCH10-A2. No adaptation module is associated to this CRISPR-Cas locus. The CRISPR array contains 14 spacers (colored rectangles) between direct repeats (grey diamonds). Last repeat (light grey diamond) shows a slightly degenerated sequence which generally indicates the directionality of the transcribed pre-crRNA. **(B)** Demonstrated activities of *Mm*Cas12m when heterologously expressed in *E. coli*. Gene silencing targeting the promoter or the coding region (left) and multiplex gene silencing.

4.3 Results

4.3.1 Expression of *Mm*Cas12m in the eukaryotic host *S. cerevisiae*

In order to use *Mm*Cas12m as a gene expression regulator in *S. cerevisiae*, the Cas protein needed to be properly translocated to the nucleus. For this purpose, we fused the nuclear localization tag Nucleoplasmin sequence "KRPAATKKAGQAKKKK" (NLS) to the C-terminus of *Mm*Cas12m, as has previously been described in case of Cas12a (*110*, *111*). Fusion products to the N-terminus of *Mm*Cas12m seemed to diminish the activity of the protein in *E. coli* (*317*). The construct was expressed from a multicopy plasmid for this purpose. In order to determine the cellular localization of *Mm*Cas12m, mRuby2 was fused at the C-terminus of the NLS sequence.

The location of *Mm*Cas12m-mRuby2 in living cells of *S. cerevisiae* was visualized using contrast microscopy (Figure 4.2B). The genetic material of the cells was stained using DAPI (Figure 4.2A). The overlay of the two microscopy images shows that the signals of the *mRuby2* and the nuclear stain mostly overlap, which confirms localization of *Mm*Cas12m in the nucleus of *S. cerevisiae* (Figure 4.2C). An observation during the obtention of the images was that the intensity of the signal of *Mm*Cas12m-mRuby2 was weaker compared to the signal obtained from the nuclear stain DAPI or the mRuby2 control strains. This may be caused by unproper folding

of the fusion protein *Mm*Cas12m-mRuby2 or by degradation of the protein due to metabolic burden by overexpression of the protein from a multicopy plasmid. Moreover, expression of *Mm*Cas12m under these conditions hampered the cultivation of the strain, rendering it unviable in multiple occasions (data not shown). Also, under the microscope some cells showed apparently low or almost no mRuby2 associated fluorescence (Figure 4.2B), which was confirmed by subsequent analysis of the population by flow cytometry (Figure S4.1).
Flow cytometry data showed that the *Mm*Cas12m-mRuby2 population did not express fluorescence in a uniform manner. Controls with mRuby2 and mRuby2-NLS showed that addition of an NLS to mRuby2 resulted in loss of fluorescence in a part of the population (Figure S4.1). This effect increased when expressing *Mm*Cas12m-mRuby2 with an NLS from a multicopy plasmid, with only ~17% of the population being fluorescent.

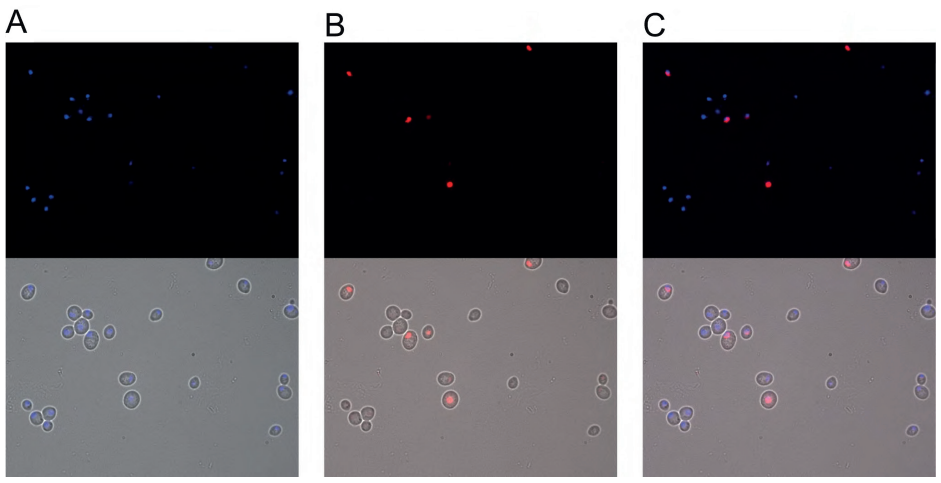


Figure 4.2. Visualization of *Mm*Cas12m in the nucleus of *S. cerevisiae* strain YSTB117. (A) Nucleic acid stained by DAPI, used to delimitate nuclear position. Top picture shows fluorescence signal and bottom signal shows the overlay of fluorescence with the phase contrast image. **(B)** Visualization of *Mm*Cas12m-mRuby2. Top picture shows fluorescence signal and bottom signal shows the overlay of fluorescence with the phase contrast image. **(C)** Overlay of the two fluorescent signals. Images shown are representative of the examined population. Top picture shows the overlay of the fluorescence signals from the DAPI stained and the *Mm*Cas12m-mRuby2.

Western blot analysis of *Mm*Cas12m expression strains also showed that expression of the protein at high levels might cause problems for *S. cerevisiae* (Figure S4.2). Detection of *Mm*Cas12m expressed from a multicopy plasmid using Western blot was not possible either because of multiple failed attempts of cultivating the strain, or because of apparent loss of the construct when the strain grew (which resulted in the absence of signal on Western blot). In a similar way, Western blot detection of *Mm*Cas12m fused to mRuby2 (YSTB117) was inconsistent. Only when the strain showed a slow growth rate (not quantified) we could detect *Mm*Cas12m-mRuby2 expression.

In contrast to expression from a multicopy plasmid, however, *Mm*Cas12m production was consistently detected when expressed from a genome-integrated copy (YSTB211) (Figure S4.2). To proceed with further experiments, the *Mm*Cas12m impact on yeast growth was assessed. YSTB211 resulted in a growth rate of 0.37 ± 0.008 h^{-1}, while YSTB164 (isogenic background) had a growth rate of 0.39 ± 0.004 h^{-1}. This result shows that expression of *Mm*Cas12m from a genome-integrated copy has a minimal effect on *S. cerevisiae* growth, although statistically significant (Figure S4.3).

4.3.2 Fluorescence repression assays of *Mm*Cas12m in *S. cerevisiae*

Various GFP fluorescent strains expressing *Mm*Cas12m (with or without HA tag for protein detection) or *Mm*Cas12m-Mxi1 from a genome integrated copy were designed to test the ability of *Mm*Cas12m as genome silencing tool in *S. cerevisiae*. Expression of *Mm*Cas12m from a genome integrated copy was preferred since expression of *Mm*Cas12m from a multicopy plasmid caused expression and growth problems. Strains expressing *Fn*dCas12a-Mxi1 and *Sp*dCas9-Mxi1 were used as positive controls for repression activity, as their use in *S. cerevisiae* had previously been reported (*104*, *112*, *165*, *166*). To be able to quantify the eventual silencing activity of *Mm*Cas12m, we designed several spacers targeting either the coding or the non-coding strand of the promoter regulating the expression of *egfp*, or the coding sequence of *egfp* (Figure 4.3A). These guides were expressed from multicopy plasmids maintained with an antibiotic marker. In every test, a strain expressing a non-targeting (NT) spacer was added as negative control for silencing. Moreover, a *Sp*Cas9-Mxi1 strain expressing a targeting (T) spacer was always included as positive control and internal control between experiments.

Initially, all spacers were tested in the *egfp* expressing strain expressing *Mm*Cas12m (Figure 4.3). Only spacer p2 (targeting the non-coding strand in the promoter region) showed some gene repression activity when comparing the levels of fluorescence to the strain expressing a NT spacer for both variants of *Mm*Cas12m (with and without HA tag). However, repression was only statistically significant for the variant with the HA tag (t-test, p-value=0.0459). In this case, fluorescence was diminished by 21.1 ± 1.8 % (1.3 ± 0.1-fold decrease). Later, the transcriptional repressor Mxi1 was fused to *Mm*Cas12m to eventually increase the repression levels observed with *Mm*Cas12m alone. Mxi1 has been shown to interact with the yeast Sin3-histone deacetylase homolog complex (*318*) and therefore, can promote gene silencing in a targeted way when fused to a dCas protein. However, the repression levels achieved with *Mm*Cas12m fused to Mxi1 with the p2 spacer (achieving a similar 20.9% ± 1.7%) were considered not statistically significant (t-test, p-value=0.0693). Likewise, none of the other guides targeting the promoter region or the CDS of *egfp*, decreased the fluorescence levels in the testing strain in a significant way (Table S4.5).

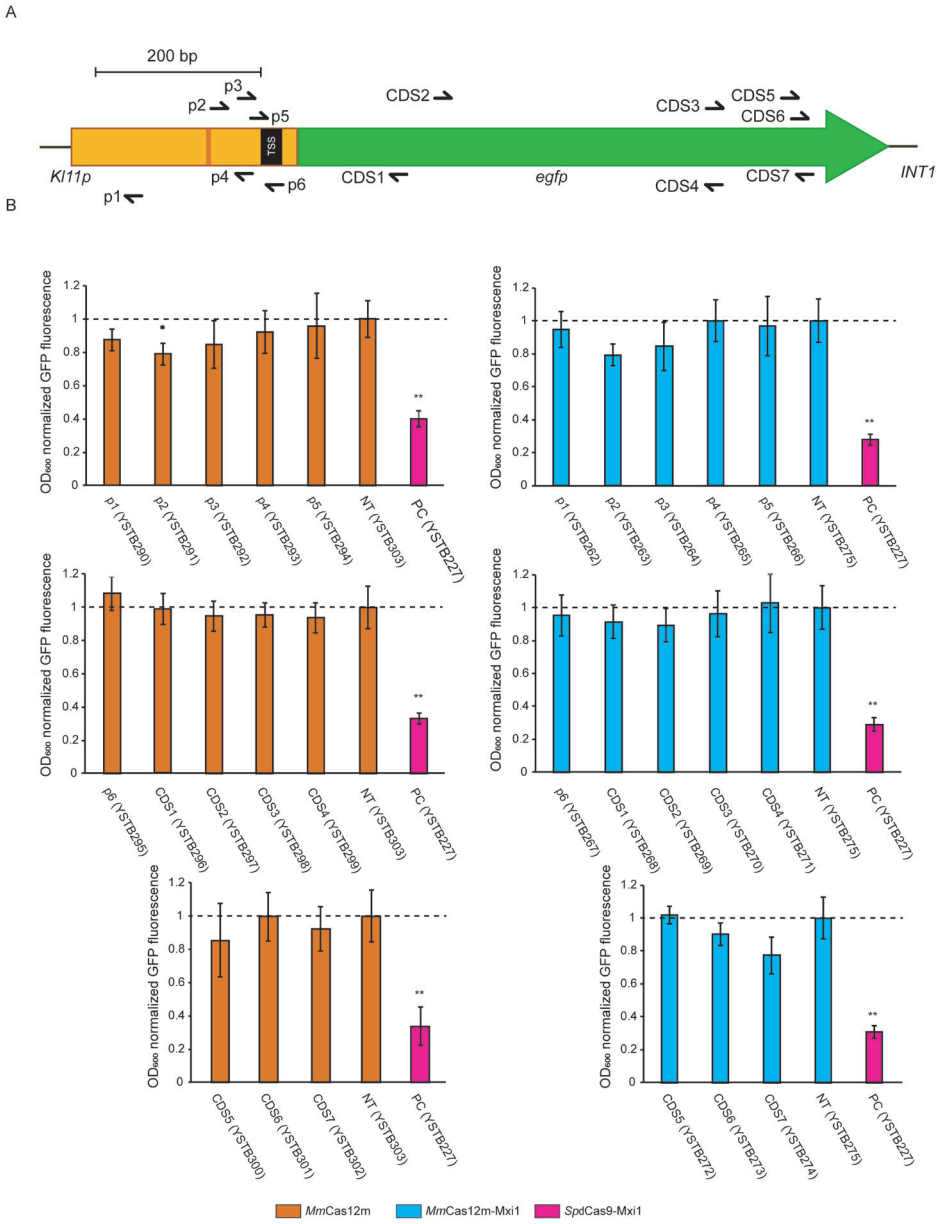


Figure 4.3. *egfp* silencing using *Mm*Cas12m and *Mm*Cas12m-Mxi1. (A) Scheme of *Mm*Cas12m targeting *KI11p* promoter and *egfp* coding sequence. Spacers targeting the coding strand are marked with a black arrow pointing left and spacers targeting the non-coding strand are marked with a black arrow pointing to the right. Promoter region is marked in yellow while *egfp* coding sequence is depicted in green. Black square indicates TSS and red line indicates TATA box. **(B)** OD_{600} normalized-fluorescence measurements of different strains expressing GFP and *Mm*Cas12m (red), *Mm*Cas12m-Mxi1 (blue) or *Sp*dCas9 (pink). Statistical significance is indicated with symbols (*, 0.05 > p-value > 0.01; **, p-value < 0.01). Statistical significance was calculated with a t-test comparing each strain to its congenic strain expressing a non-targeting (NT) spacer. A positive control with *Sp*dCas9 and a *egfp* targeting spacer is included in each testing set.

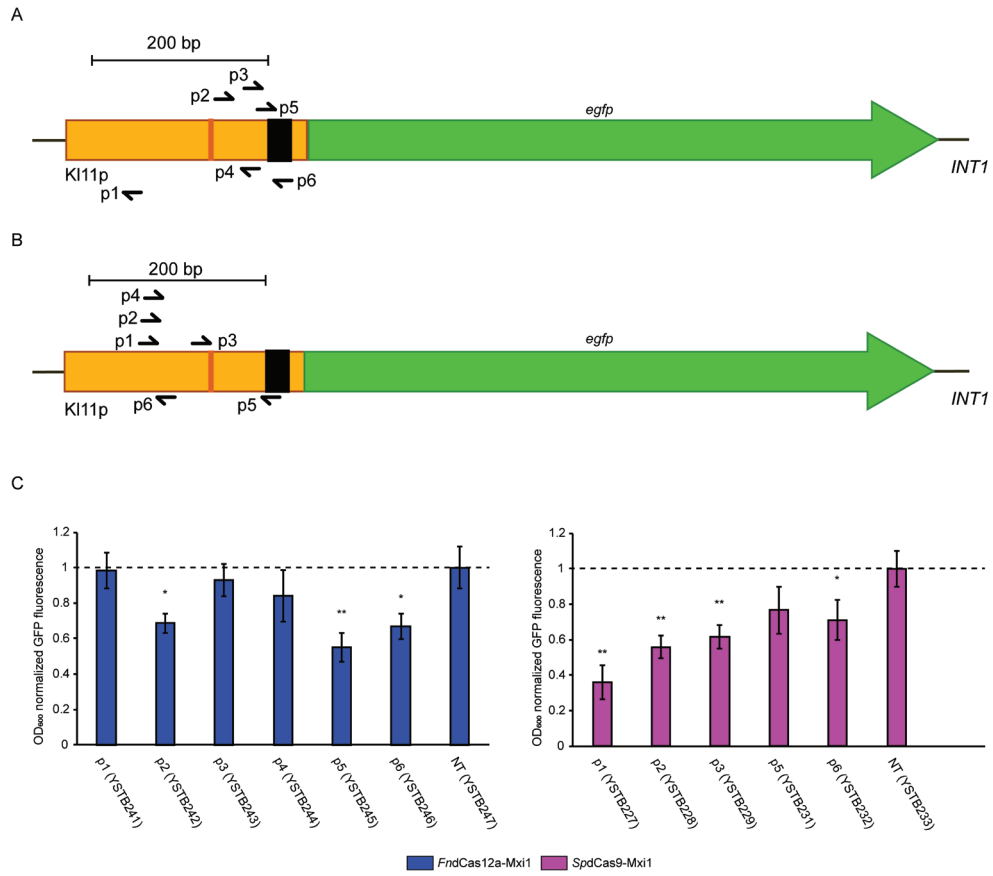


Figure 4.4. *egfp* silencing using *Fn*dCas12a-Mxi1 and *Sp*dCas9-Mxi1. (A) Scheme of *Fn*dCas12a-Mxi1 targeting *KI11p* promoter upstream of *egfp* coding sequence. Spacers targeting the coding strand are marked with a black arrow pointing left and spacers targeting the non-coding strand are marked with a black arrow pointing right. **(B)** Scheme of *Sp*dCas9 targeting *KI11p* promoter upstream of *egfp* coding sequence. Spacers targeting the coding strand are marked with a black arrow pointing left and spacers targeting the non-coding strand are marked with a black arrow pointing right. **(C)** OD_{600} normalized-fluorescence measurements of different strains expressing eGFP and *Fn*dCas12a-Mxi1 (blue - left) or *Sp*dCas9-Mxi1 (purple – right). Statistical significance is indicated with symbols (*, 0.05 > p-value > 0.01; **, p-value < 0.01). Statistical significance was calculated with a t-test comparing each strain to its congenic strain expressing a non-targeting (NT) spacer.

In parallel, fluorescent strains expressing *Fn*dCas12a-Mxi1 and several spacers targeting the *egfp* promoter region were also tested to assess the accessibility of the nuclease to the designed target sites (Figure 4.4A). Since *Mm*Cas12m and *Fn*Cas12a share a common PAM (5'-TTV-3'), we designed all the spacers according to this PAM and we could test all the spacers in the promoter region for both *Mm*Cas12m and *Fn*dCas12a. When assessing the silencing activity of *Fn*dCas12a-Mxi, spacer p2 (t-test, p-value=0.0139), spacer p5 (t-test, p-value=0.0055) and spacer p6 (t-test, p-value=0.0143) showed statistically significant silencing activities. These

spacers targeted the non-coding strand (p2 and p5) or the coding strand (p6) of the DNA at the promoter region. Repression levels of 45.2 ± 6.7% (1.8±0.3 fold) were achieved with the best crRNA (p5). At the same time, when using *Sp*dCas9-Mxi1 (Figure 4.4B), all designed spacers targeting the promoter region of *egfp* (except for p5) showed significant gene repression, achieving repression levels of 63.8±17.0 (2.8±0.7 repression-fold).

4.4 Discussion

In this work, we systematically tested *Mm*Cas12m for CRISPRi in *S. cerevisiae* and we compared it to previously used CRISPRi systems based on *Fn*dCas12a-Mxi1 and *Sp*dCas9-Mxi1. In the eukaryotic organism, we first confirmed nuclear localization of *Mm*Cas12m. Then, we showed that expression of the protein from a multicopy plasmid resulted either in non-optimal growth, or in non-proper expression of *Mm*Cas12m. Expression of the protein from a chromosome-integrated gene did not have a pronounced effect on yeast growth, and hence allowed for testing of the silencing activity on an eGFP-expressing strain. Silencing using *Mm*Cas12m was only detected with one of the tested spacers, allowing for 20% reduction of fluorescence. Fusion of the repressor domain Mxi1 to *Mm*Cas12m did not increase silencing levels at any specific target location. Both *Fn*dCas12a and *Sp*dCas9 allowed for higher repression levels with multiple tested spacers (up to 45% and 64%, respectively), validating the presented silencing assay.

Expression of CRISPR-Cas proteins can lead to cell burden or even toxicity, which might prevent the use of certain Cas proteins for genome editing or regulation (*319*). Although we observed nuclear localization of *Mm*Cas12m in *S. cerevisia*e, the fluorescence signal was not homogeneous throughout the population. This might indicate some expression burden associated with high level expression of *Mm*Cas12m. In addition, strains expressing *Mm*Cas12m from a high copy number plasmid exhibited impaired growth on multiple occasions and the protein could not be detected by Western blot. These observations could indicate that high expression levels of *Mm*Cas12m might be toxic for *S. cerevisiae* cells and therefore, high copy number plasmid expression might not be stable. Attempts to obtain stable expression of *Mm*Cas12m in human cells also resulted in cellular toxicity, hampering the cultivation of the transformed cell lines (data not shown). Similar detrimental effects have been reported for other Cas proteins such as Cas9 (*173*, *182*) and Cas12a in case of *S. cerevisiae* (*110*, *111*). Toxicity of Cas proteins has also been reported in other organisms, potentially representing a limitation for the development of novel tools (*145*, *320*, *321*). However, as reported here, toxicity problems can be solved by means of tuning protein expression levels or using inducible expression systems. Still, most of the times, unsuccessful efforts of using novel Cas proteins in any organism remain unpublished, which hampers the assessment of the importance of this issue.

Distinct deactivated Cas proteins are able to achieve different silencing levels in the same organism (*165*). Although *Mm*Cas12m was successfully translocated to the nucleus of *S. cerevisiae*, significant eGFP fluorescence repression of approximately 20% could only be observed when

using 1 out of the 7 tested spacers (p2), targeting the non-coding strand (anti-sense strand) of the *egfp* promoter. The mentioned spacer targeted just upstream (overlapping on 1 bp) of the promoter TATA box, which makes it different from the other two spacers targeting the same strand (p3 and p5). These spacers targeted the region between the TATA box and the TSS. For dCas9 silencing, positioning of the spacers downstream or in close proximity to the TATA box results in gene repression, effect that we can partially and minimally reproduce using *Mm*Cas12m, only when directing it to the TATA box sequence. Strand bias effect (not reported for dCas9 or dCas12a (*112*, *322*)) could not be assessed for *Mm*Cas12m in yeast since only one spacer lead to gene repression.

In order to increase this minimal repression observed, we followed the common strategy of fusing Cas proteins to a repressor domain such as the KRAB, Mxi1 or MIG1 repressor domains (*104*, *112*, *164*). Unfortunately, fusion of *Mm*Cas12m to Mxi1 did not increase the silencing activity of *Mm*Cas12m, at least not with the current fusion-protein design. Several explanations could explain the inability of *Mm*Cas12m to silence gene expression. First, the choice of regulatory sequences (promoter and terminator) and integration site for the expression of *egfp*. For each promoter, there seems to be a more effective region to target, but this region is not well-defined and varies from promoter to promoter (*112*). It is possible that the combination of promoter-*egfp* chosen for this study limits the repression levels that theoretically could be obtained. However, by means of using *Fn*dCas12a and *Sp*dCas9 under the same set-up, we demonstrated that repression levels of up to 64% (2.8-fold repression) could be obtained under the same conditions in a stable and reproducible way. Secondly, *Mm*Cas12m transient DNA binding or low DNA affinity to the DNA in yeast could explain its low repression levels, since addition of the Mxi1 repression domain did not improve silencing activity as it occurs with *Sp*dCas9 or *Fn*dCas12a (*104*, *164*, *323*). Normally, in many eukaryotic organisms, steric hindrance of the RNA polymerase by targeting the promoter region is not efficient enough to achieve high silencing levels. Although a single study reported an 18-fold repression in yeast when targeting the *TEF1* promoter with *Sp*dCas9 alone (*166*), deactivated Cas proteins are often fused to repressor domains such as the Mxi1, KRAB, TUP1 or MIG1 for efficient gene silencing (*309*, *324*). Although there seems to be limitations in the transferability of design rules when combining Cas proteins (dCas12a, dCas9) with repressor domains (Mxi1, KRAB, TUP1 or MIG1), Mxi1 improves silencing in yeast with both *Sp*dCas9 and *Fn*dCas12a (*112*, *323*). Therefore, a low residence time of *Mm*Cas12m onto its target could explain the absence of improvement in silencing activity when *Mm*Cas12m was fused to Mxi1. However, this hypothesis would contradict the Surface Plasmon Resonance results analyzing the binding of *Mm*Cas12m to dsDNA (Chapter 3) (*316*). Further experimental assessment is required to confirm the poor or transient binding of *Mm*Cas12m to yeast DNA *in vivo*. Single molecule tracking microscopy could help to elucidate the movements of *Mm*Cas12m inside the nucleus of yeast and estimate its binding time to DNA (*325*, *326*), which could help to explain why *Mm*Cas12m is not yet ready to be used for silencing in yeast.

In our experiment, we chose integration site 1 (*INT1*) to integrate *KI11p:egfp:CYC1* for testing purposes. Accessibility to the target integration site has been previously demonstrated (*110*).

Moreover, we also tested the accessibility of each of the specific designed promoter targets for *Mm*Cas12m by using *Fn*dCas12a, which showed higher repression activity than *Mm*Cas12m targeting the same specific sequences. Targeting guides for *Sp*dCas9-Mxi1, *Fn*dCas12a-Mxi1 and *Mm*Cas12m were designed to target a 200 bp window upstream of the transcription start site (TSS) as described before (*322*). The absolute fold-repression values obtained in this study with previously tested nucleases (*Sp*dCas9 and *Fn*dCas12a) were lower than values reported in other CRISPRi studies performed in *S. cerevisiae* (*112*, *166*, *232*). Clearly, different set-ups and measuring protocols impair direct comparisons of silencing assays. Expression levels of the reporter, of the guides or of the Cas protein and design of target sequences have an effect on the reported silencing effect. Although *Sp*dCas9-Mxi1 and *Fn*dCas12a-Mxi1 cannot be easily compared in the absence of a target that shares both PAMs, in our study we achieved higher repression levels when using *Sp*dCas9-Mxi1 than *Fn*dCas12a-Mxi1. To our knowledge, only one other study has compared the repression activity of both *Fnd*Cas12a-Mxi1 and *Sp*dCas9-Mxi1 on the same promoter region, and they report similar outcomes (*165*).

In our work, we expressed for the first time *Mm*Cas12m in a eukaryotic organism. Although the protein was localized in the nucleus of *S. cerevisiae*, expression problems allowed only for a low-copy expression system. Moreover, when the protein was tested for silencing activity in *S. cerevisiae*, only limited silencing efficiencies were detected under the tested conditions. These results may suggest that the implementation of *Mm*Cas12m as a genome regulation tool is still limited to prokaryotic organisms, and further research is needed to better understand its behavior in eukaryotic cells, and to optimize its performance.

4.5 Materials and methods

4.5.1 *E. coli* strains and growth conditions

E. coli DH5α and DH10B were used for plasmid construction and maintenance. Cultures were routinely grown in Luria Bertani (LB) medium (10 g·L^{-1} Peptone (Oxoid), 5 g·L^{-1} Yeast Extract and 10 g·L^{-1} NaCl). Ampicillin (100 µg/mL) was supplemented when required.

4.5.2 *S. cerevisiae* strains and growth conditions

S. cerevisiae strains built in this study belong to the *S. cerevisiae* CEN.PK113-5D background. Used and generated strains for this study can be found in Table S4.4. *S. cerevisiae* was grown on YPD medium (10 g·L^{-1} Peptone, 10 g·L^{-1} Yeast Extract, 2% glucose) or Synthetic Medium (SMG) medium (3 g·L^{-1} KH_2PO_4, 0.5 g·L^{-1} $MgSO_4 \cdot 7H_2O$, 5 g·L^{-1} $(NH_4)_2SO_4$, 1 mL·L^{-1} of a trace element solution, and 1 mL·L^{-1} of a vitamin solution as previously described (*327*). When required, the media was supplemented with 200 µg/mL G418-sulfate (Geneticin). When selection with G418 was required on SMG media, $(NH_4)_2SO_4$ was replaced with 3 g·L^{-1} K_2SO_4 and 2.3 g·L^{-1} urea to avoid pH drop (*328*).

For growth rate determination, synthetic complete medium (SC) [6.7 g·L^{-1} yeast nitrogen base without amino acids and with ammonium sulfate (SIGMA), 2.0 g·L^{-1} Yeast Synthetic Drop-out Medium Supplements (Sigma Aldrich) and 150 mg·L^{-1} uracil] was used. 2% Glucose was used in all cases as a sole carbon source.

4.5.3 Plasmid construction

The plasmids used in this study and the oligonucleotides (IDT and SIGMA) used for cloning and sequencing can be found in Figure S4.1 and Table S4.2, respectively. *Mm*Cas12m sequence was codon optimized for *S. cerevisiae* and synthetized by IDT as two overlapping gBlocks Gene Fragments (IDT). Sequence can be found in Supplementary material 4.1.

Plasmids used for the *Mm*Cas12m localization experiment (PL-156, PL-157 and PL-164) were built using pUDE731 as backbone and introducing the mRuby2 coding sequence in combination with a nuclear localization signal or the sequence coding for *Mm*Cas12m.

PL-074 was constructed to correct the SUP4 terminator sequence to its original length by PCR amplification of pUD628 and subsequently re-circularization by blunt-end ligation (pUD628 was a gift from Jean-Marc Daran). PL-074 was subsequentially used as backbone for expression of guides for the different effectors tested in this study.

PL-098 was constructed by incorporation of a spacer targeting the *INT1* locus as an overhang in the forward primer used for linearization of PL-074 by PCR amplification.

To incorporate *Mm*Cas12m spacers and repeats into PL-074 backbone, two different strategies were followed. Initially, pCRISPR-*Mm*-NT (BbsI) (Chapter 3) was digested with BbsI-HF® and different pairs of pairing oligo nucleotides were used to ligate the plasmid. Plasmids PL-115 to PL-118, PL-130 to PL-132, PL-138 and PL-162 were built in this way. Later primers BG16700 and BG14156 were used to obtain different DNA fragments containing the *Mm*Cas12m repeat-spacer-repeat sequences (A0140 to A0143, A0150 to A0152 and A0154). These DNA fragments were digested in a two-step protocol with restriction enzymes KpnI and BtgZI, blunted with T4 DNA polymerase (NEB) and cloned into a linearized PL-074 backbone (fragment A0128). Plasmids PL-120 to PL-123, PL-134 to PL-136, PL-139 and PL-163 were built in this way.

Plasmid PL-163 (containing a BsaXI restriction site) was subsequentially digested with BsaXI and used for building targeting plasmids (PL-181 to PL-186) by ligation of different pairs of annealed oligonucleotides.

To incorporate *Sp*Cas9 repeats into the PL-074 backbone, PL-175 was first built based on a NEBuilder® HiFi DNA Assembly of a linearized product of PL-074 (fragment A0193) and a DNA fragment containing the structural region of Cas9 gRNA obtained from p426-SNR52p-gRNA.CAN1.Y-SUP4t (fragment A0192). A BsaXI restriction site sequence was incorporated by amplification of PL-175 with a forward primer containing a BsaXI restriction site, and subsequentially ligated to build PL-176. PL-176 was subsequentially digested with BsaXI and ligated with previously annealed pairing oligonucleotides to form all targeting plasmids for

*Sp*Cas9 (PL-187 to PL-193). In a similar way, PL-179 was built in order to incorporate a BsaXI restriction sites between the *Fn*Cas12a repeats in PL-074.

Easiness of cloning was increased by incorporation of an RFP expression cassette between BsaXI restriction sites in all the above mentioned BsaXI plasmids (PL-176, PL-179 and PL-163). Plasmids PL-196, PL-198 and PL-200 were built which contain an RFP expression cassette between BsaXI restriction sites and the effector repeats. The backbones for each plasmid were amplified from the previously constructed BsaXI containing plasmids. Plasmids were obtained by NEBuilder® HiFi DNA Assembly. Plasmid PL-200 was subsequentially digested with BsaXI and used for building *Fn*Cas12a targeting plasmids (PL-201 to PL-207) by ligation of different pairs of annealed oligonucleotides.

4.5.4 Generation of strains for microscopy and flow cytometry

S. cerevisiae CEN.PK113-5D was transformed with PL-156, PL-157, PL-164 and PL-140 multicopy plasmids to obtain YSTB111, YSTB112, YSTB117 and YSTB115, respectively. Transformants were selected on SMG (not supplemented) agar plates. Obtention of transformants was confirmed by PCR of the transformed plasmids and strains were stored in SMG medium with 15% glycerol at -80 °C.

4.5.5 Generation of *S. cerevisiae egfp* chromosomally integrated expression strain

A codon optimized *egfp* expression cassette was integrated into integration site 1 (*INT1*) as followed in previous studies (*110*). *S. cerevisiae* strain YSTB013 (CEN.PK113-5D background strain expressing *Fn*Cas12a from a multicopy plasmid) was transformed with four DNA fragments and a *INT1* plasmid targeting guide to obtain YSTB163. Integration of the *egfp* cassette was based on the *in vivo* recombination of four DNA fragments (a 5' and 3' homology fragments to *INT1*, a fragment containing *Kl11p* (*Kluyveromyces lactis* KLLA0F20031g promoter) and a fragment encoding for eGFP). YSTB164 (*egfp* expression strain) was obtained after curing YSTB163 (obtained as indicated in Table S4.4).

4.5.6 Design of *egfp* targeting spacers

Spacers to target the promoter region or the coding sequence of *egfp* were designed using the ChopChop online tool (*329*). Spacers were designed to target either the target or the non-target strand of the promoter and the coding sequence of *egfp*. All spacers were checked for off-targeting in *S. cerevisiae* genome by the ChopChop online algorithm (*329*). The sequence of the tested spacers can be found in Table S4.3. All spacers were tested for self-complementarity or complementarity to the DR (in case of Cas12a spacers) using RNAfold web server.
In our design, the *Kluyveromyces lactis* promoter *KLLA0F20031g* (*Kl11p*) was used to express *egfp*. The transcription start site of the *Kl11p* promoter was determined using the YeasTSS database (*330*) and predicted to be 44 to 18 bases upstream of the *egfp* open reading frame. *Kl11p* targeting sequences were designed to target the region 200 bp upstream of the TSS as this region had been described as a good target for gene repression using CRISPRi in *S. cerevisiae* in previous works (*104*, *164*). Spacers were expressed under the regulation of the *SNR52* promoter and the *SUP4* terminator from a multicopy plasmid.

4.5.7 Generation of test strains for silencing with different effectors

Genomic constructs encoding for Cas effectors tested for silencing or used as controls (*Mm*Cas12m, *Mm*Cas12m-Mxi1, *Fn*dCas12a-Mxi1 and *Sp*dCas9-Mxi1) were genome integrated at the *INT2* site of strain YSTB164. Strains YSTB203, YSTB205, YSTB209 and YSTB2011were built following this strategy. Strains were generated by *in vivo* homologous recombination of four DNA fragments; a 5' and 3' *INT2* homologous flanks, the cassette encoding for the effector of interest and a *URA3* expression cassette. Recombination among fragments was facilitated by connector sequences described before (*110*). Strain YSTB212 was built as control which did not express any effector construct but harbored the same *URA3* expression cassette as previously mentioned strains.

CRISPR guides for each effector were expressed from multicopy plasmids under the control of the *SNR52* promoter and the *SUP4t* terminator on the PL-074 backbone. Each guide was flanked by the required repeats depending on the effector protein they were co-expressed with. Strains were obtained by transformation of the effector expressing strains with guides expressing plasmids. Strains YSTB227 to YSTB233 were obtained for testing *Sp*dCas9-Mxi1; strains YSTB241 to YSTB247 were obtained for testing *Fn*dCas12a-Mxi1; strains YSTB262 to YSTB275 were obtained for testing *Mm*Cas12m-Mxi1 and strains YSTB290 to YSTB303 were obtained for testing *Mm*Cas12m. Strain YSTB304 was built as a control with an empty backbone.

4.5.8 Growth rate determination

Strains YSTB211 and YSTB164 were grown in SC medium (supplemented with uracil) in 500 mL shake flasks. Shake flasks were filled with 100 mL of medium and incubated at 30°C and 200 rpm. Growth was monitored by measuring optical density (600 nm) using UV-1800 Spectrohpotometer (Shimadzu). Maximum specific growth rate for each strain was calculated from at least 6 time points, 3 biological replicates and 2 technical measurements.

4.5.9 Phase contrast microscopy

An Olympus BX41 system equipped with a 100x/1.30 numeric aperture oil-immersion objective was used for phase contrast microscopy. Excitation of mRuby2 was performed with X-Cite® 120Q equipped with a 530-550 nm excitation filter. Excitation of DAPI dyed cells was performed using the same laser with a 405 nm filter. Images were acquired using the Infinity Analyse software interface (Lumenera Corporation).

4.5.10 Flow cytometry

Strains were inoculated in SC (-ura) medium from single colonies on plates. Cultures were grown overnight at 30°C and 200 rpm, washed twice in 1 X PBS and diluted to OD 0.5 for an optimal sorting rate of 2000 cells/s. Fluorescence of diluted cultures was measured using

the Attune NxT flow cytometer (ThermoScientific). 30,000 single cell events were used to obtain average mRuby2 fluorescence distribution among the population (excitation 561 nm, emission filter 585/16).

4.5.11 SDS and Western blot

Cell extracts were isolated in 200 µL of Lysis buffer (50 mM Tris-HCl pH 7.6, 150mM NaCl and 1 mM EDTA). To each tube, 30 µL of 0.5-0.75 mm glass beads were added. Cells were disrupted using a FastPrep-24 ™ 5G Instrument (MP Biomedicals) in 4 rounds of bead beating at 6.5 ms^{-1} for 25 s. Tubes were kept on ice for 1 min between cycles to avoid temperature rise. Equal amounts of supernatant were used for SDS and Western blot analysis.

Proteins were separated in 10% Mini-PROTEAN® TGX™ Precast Protein Gels (#4561033) and transferred to Nitrocellulose membranes (#IB301032) using the iBlot® 2 Dry Blotting System (# IB21001). Membranes were blocked in blocking buffer (TBS-T, 5% skimmed milk), incubated with primary antibody Thermo Scientific™ Pierce™ HA Tag Mouse anti-Tag, Clone: 2-2.2.14 (#11553060) and secondary antibody Goat Anti-Mouse IgG Antibody (H&L) [HRP], pAb (#A00160). Visualization of signal was done using SuperSignal(TM) West Pico PLUS Chemiluminescent Substrate (#34577) and signal was captured using G:BOX Chemi XRQ Imager (Syngene).

4.5.12 Growth conditions for silencing experiments

S. cerevisiae fluorescent strains harboring Cas effector proteins genes integrated in the genome and the corresponding target plasmids were used for silencing assays. For wake-up cultures, three individual colonies were picked from fresh YPD+G418 agar plates. Biological replicates were grown in 2.5-3 mL YPD+G418 in 24-deep-well plates with Breath-Easy® membranes (Diversified Biotec). Plates were grown overnight at 30°C and 250 rpm. Inoculums were prepared by transferring the required volume to fresh 24-deep-well plates filled with 3 mL YPD+G418, which were grown under the same conditions for approximately 10 hours and until exponential phase was achieved. Exponentially growing cultures were used to inoculate the measurement cultures by transferring the required volume for an initial OD_{600} of 0.05. Plates were incubated for 12-13 hours until measurements were performed.

4.5.13 Plate reader measurement of silencing experiments

500 µL of each well were transferred to a deep conical 96 well plate in duplicate. Plates were spun down at 3500 x g for 10 min. Supernatant was discarded by inversion and cells were resuspended in 500 µL of sterile PBS. Plates were spun down a second time at 3500 x g for 10 min and cells were resuspended in 500 µL PBS for measurements. 100 µL of each resuspension was transferred to a black 96 well black/clear bottom plate. Volume was corrected to 200 µL with PBS. A fluorescent end-point measurement was performed using the Synergy Neo 2 Biotek plate reader. Fluorescence was measured at 30°C after a 30 s double orbital shaking step. Top measurements were taken using a gain of 125, excitation wavelength of 485/20 and emission wavelength of 515/5.

Fluorescence ratios were calculated considering the following formula:

$$\text{Ratio} = \frac{\left[\frac{F_{x,targeting}-F_{Blank}}{OD_{600,x,targeting}-OD_{600,blank}}\right]_{av} - \left[\frac{F_{non-fluorescent}-F_{Blank}}{OD_{non-fluorescent}-OD_{600,blank}}\right]_{av}}{\left[\frac{F_{non-targeting}-F_{Blank}}{OD_{non-targeting}-OD_{600,blank}}\right]_{av} - \left[\frac{F_{non-fluorescent}-F_{Blank}}{OD_{non-fluorescent}-OD_{600,blank}}\right]_{av}}$$

Fluorescence ratios were used to run t-Tests over pairs of variables (comparing the value of test spacers to the one of the NT spacer). GraphPad t Test Calculator was used for this purpose, website: https://www.graphpad.com/quickcalcs/ttest1/ (accessed March 2021).

4.6 Acknowledgements

This study has been financially supported by the research program Building Blocks of Life by the Netherlands Organization for Scientific Research (NWO, project 737.016.005).

4.7 Supplementary materials

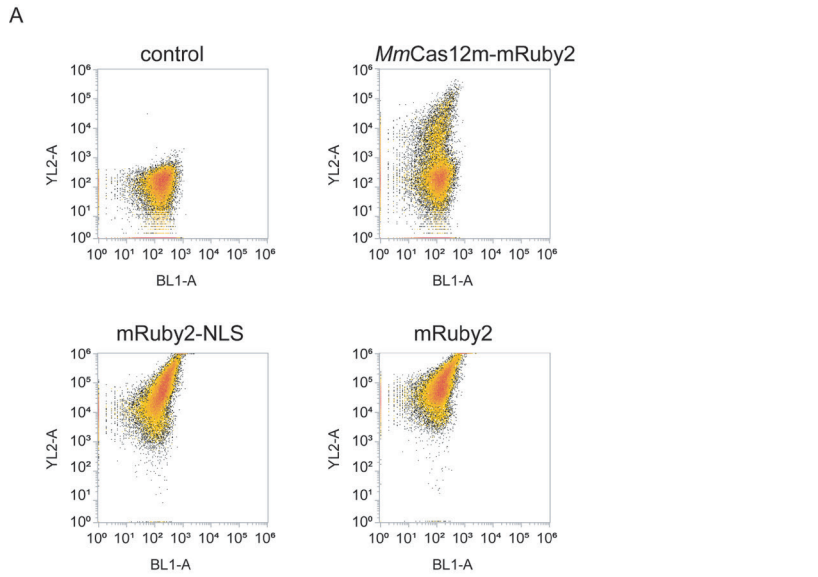


Figure S4.1. *Saccharomyces cerevisiae* CEN.PK113-5D expressing *Mm*Cas12m-mRuby2, mRuby2-NLS or mRuby2. (A) YL2-A vs BL1-A density plot. Each individual dot represents an individual particle passed through the laser and classified as a singlet. For each population, 30.000 single events are plotted. Control cells (not expressing *mRuby2*) are plotted on the low left quadrant while mRuby2-NLS and mRuby2 expressing cells are plotted on the top left quadrant. *Mm*Cas12m-mRuby2 expressing cells present a heterogeneous population with some fluorescent cells and a majority of not fluorescent cells. **(B)** Histograms of each population: *Mm*Cas12m-mRuby2 (salmon), mRuby2-NLS (pink) and mRuby2 (brown). Approximately 17% of single cells from the *Mm*Cas12m-mRuby2 population present fluorescence while >93% of single cells from the mRuby2-NLS or the mRuby2 populations present fluorescence. The mRuby2-NLS presents a low level of fluorescence compared to the mRuby2 population.

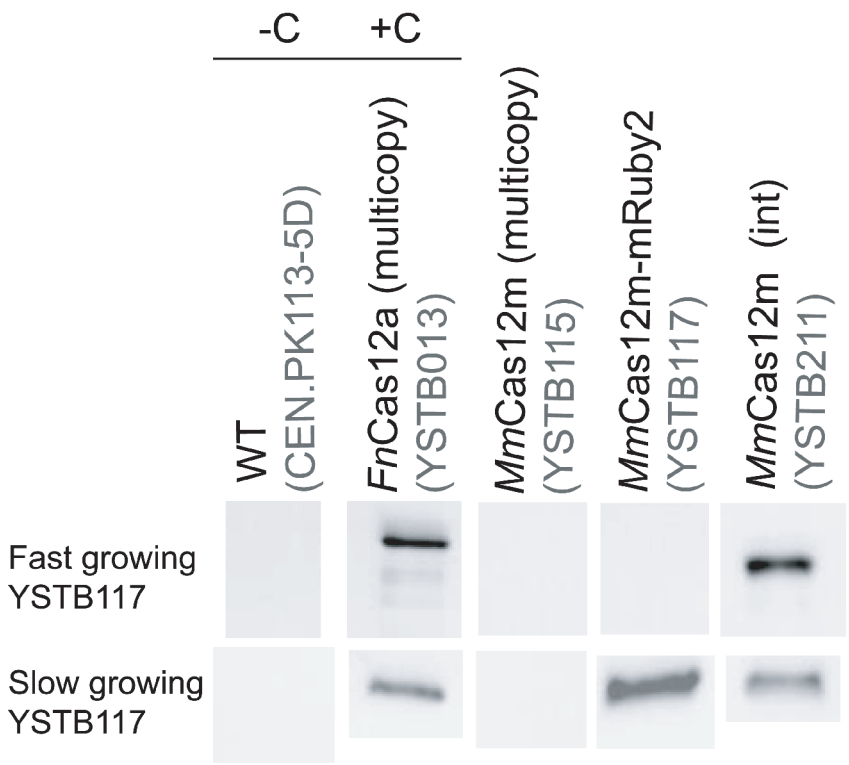
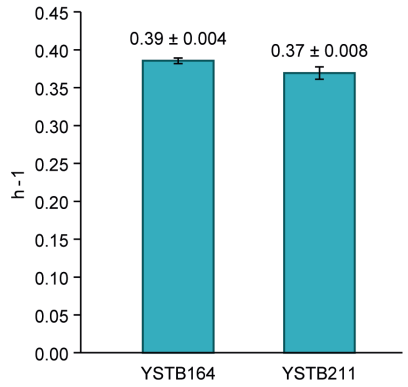


Figure S4.2. Western blot detection of HA tag fused to *Mm*Cas12m or *Fn*Cas12a (as positive control). No protein was detected on WT strain and positive control showed always signal. *Mm*Cas12m could not be detected when expressed from a multicopy plasmid (YSTB115). For strain YSTB117, *Mm*Cas12m could be detected only when the strain grew at a slow growth rate. *Mm*Cas12m was always detected when expressed from a genome integrated copy (YSTB211).

A



B

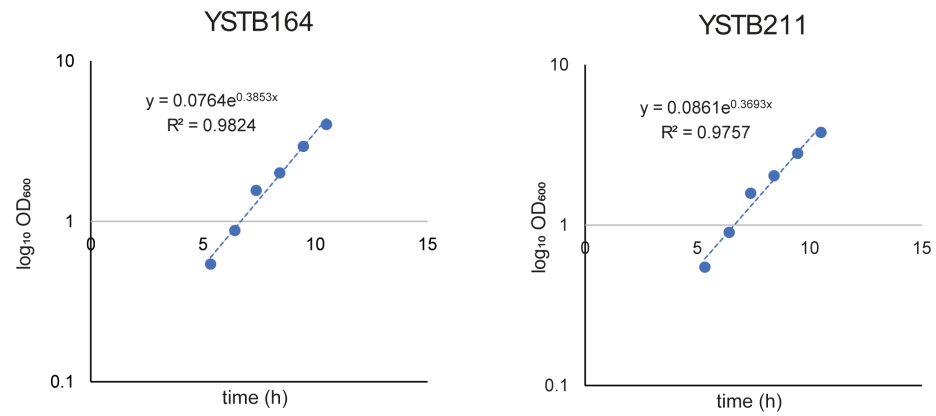
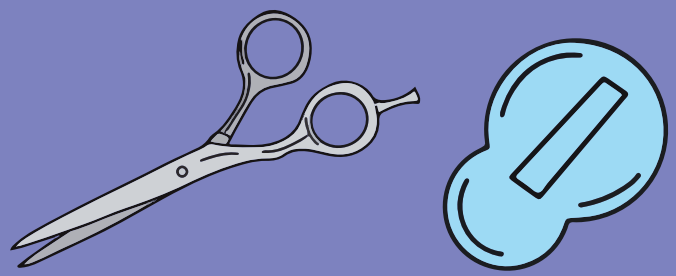


Figure S4.3. Specific growth rate of strain constitutively expressing *Mm*Cas12m from a genome integrated construct at *INT2* (YSTB211) and its congenic controls YSTB164 (expressing eGFP). (A) Difference in growth rates is considered statistically significant. Statistical significance was calculated with a t-test (p-value 0.0339). **(B)** Growth rate profiles for strains YSTB164 and YSTB211.

Additional supplementary materials, tables and figures of this chapter can be accessed via

https://figshare.com/s/2387b15aea490da1d82d



CHAPTER 5

Exploring *Mm*Cas12m potential as a base editing tool in *Saccharomyces cerevisiae*

Belén Adiego-Pérez[1]
Wen Wu[1]
Sarah D'Adamo[2]
John van der Oost[1]

[1] Laboratory of Microbiology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands

[2] Bioprocess Engineering & AlgaePARC, Wageningen University and Research, P.O. Box 16, 6700 AA Wageningen, The Netherlands

5.1 Abstract

Base editing has previously been achieved in the yeast *Saccharomyces cerevisiae* by using nCas9 or dCas9 in combination with the *Pm*CDA1-UGI or APOBEC1 deaminases. Until now, no other Cas protein has been reported for base editing purposes in this organism. In this study, we assessed the activity of the recently characterized *Mm*Cas12m DNA-binding protein fused to the synthetic deaminase parts of the previously described Cas9 base editors, i.e. *Pm*CDA1-UGI and APOBEC1-YEE-UGI. Deaminase-dependent, specific base editing has been observed at the target site, the reporter gene *ADE2*. Unfortunately, the editing frequency appeared very low, possibly due to non-optimal binding of *Mm*Cas12m to the yeast DNA. Moreover, the inability of this protein to introduce DNA nicks may limit its potential. Nevertheless, these findings indicate that *Mm*Cas12m base editors are expressed in *S. cerevisiae* and able to target specific locations in the genome of the yeast. Apart from optimization of *Mm*Cas12m, analyzing the performance of other Cas12m homologs eventually might allow the use of this novel Cas12 subtype in eukaryotic organisms.

5.2 Introduction

With the recent advances in genetics, scientists have been working to develop precise and efficient genome editing tools to modify the genomes of living cells (*331*, *332*). In the last decade, CRISPR-Cas systems have revolutionized the genome editing field as programmable tools that can generate RNA-guided DNA double-strand breaks (DSBs) (*82*, *333*). Generation of DSBs awakens cellular repair mechanisms, such as non-homologous end joining (NHEJ) or microhomology-mediated end joining (MMEJ) (*334*, *335*). These processes repair the break in an imprecise, error-prone manner, typically resulting in the introduction of random insertions or deletions (indels) (*334*, *335*). Alternatively, homology-directed repair (HDR) mechanisms allow for precision repair, through the incorporation of specific DNA sequences at the site of the DSB, derived either from a natural genetic copy of that DNA sequence, or from a synthetic repair template consisting of a desired edit flanked by homology arms (*155*). Although targeted and defined DNA modifications can be introduced using CRISPR systems and HDR, this last cellular process only operates in actively dividing cells. Moreover, in some organisms the HDR activity is low and thus out-competed by non-homologous repair (*336–338*). Recently, CRISPR-associated base editing has been presented as a promising technology to generate precise DNA modifications in an HDR and repair template-independent manner, opening new doors for its therapeutic use in human cells (*108*, *282*, *339*).

Base editors are built by fusing components of CRISPR systems and base modification enzymes. This allows for guide-dependent targeting, and for the deamination of nucleotides within a nucleotide window at (or close to) a CRISPR target sequence (*108*). These deaminase domains allow C-to-T (in case of cytosine base editors, CBE) (*282*, *340*) or A-to-G (*339*) (in case of adenine base editors, ABE) conversions. In the process of C-to-T deamination, cytosine (C) bases are initially converted into uracil (U). In order to prevent the recognition of the deaminated base by a Uracil DNA glycosylase and its further removal into an a-basic site (*341*), a Uracil DNA glycosylase inhibitor (UGI) is fused to the deaminase (*282*). This prevents unwanted C-to-A or C-to-G substitutions and increases base editing efficiency (*342*). Commonly, catalytically deactivated Cas9 (dCas9) or nickase Cas9 (nCas9) are fused to CBEs or ABEs to introduce single nucleotide changes in the selected target sequence (*282*, *339*). While deactivated variants of dCas9 can be generated by the simultaneous substitution of active site residues of both the HNH and the RuvC domains, two different nCas9 can be generated by individual deactivation of each of the catalytic domains. In the context of base editing, the use of nCas9 (D10A) has been proven useful to increase base editing efficiencies since introduction of a cut in the target strand promotes mismatch repair using the edited non-target strand as template (*106*, *339*). The extended use of Cas9 base editors is limited to the encounter of an 5'-NGG-3' protospacer adjacent motif (PAM) and a defined window determined by the combination of base editor and the specific Cas protein (*343*). In order to explore alternative PAMs and editing windows, other modified nucleases such as the type V-A dCas12a (*283*, *344*) or the Type V-F Cas12f (Cas14) (*132*, *291*) have been proposed as complementary DNA base editing systems. However, only dCas9 and nCas9 have been used in the yeast *Saccharomyces cerevisiae* for base editing purposes to date.

Recent bioinformatic analyses have resulted in the identification of novel putative nucleases that have been functionally characterized under *in vivo* and *in vitro* conditions (*76*, *130*, *133*, *140*, *257*, *345*). Among these new candidates, *Mm*Cas12m is a particularly appealing type V-U1 predicted endonuclease due to its small size. Previous research performed *in vitro* and in *Escherichia coli* showed that, like other Type V endonucleases, *Mm*Cas12m requires a 5'-TTN-3' PAM and it can process its own pre-crRNA. In contrast to other Type V endonucleases (*122*, *346*), *Mm*Cas12m can bind DNA without introducing DSB, which allowed for the development of gene silencing and base editing applications in *E. coli*. The characterization of different *Mm*Cas12m based C-to-T base editors in this model bacterium defined a base editing window consisting of two regions: a PAM-proximal region at positions 2-5 and a PAM-distal region at positions 13-19. This dual editing region has not been reported by any other previously developed base editor (*108*), although a recently developed A-to-G Cas12f1 base editor showed some base editing activity at position 3 and positions 8, 10 and 14 (*291*). Confirmation of the dual window activity in other organisms would establish *Mm*Cas12m base editors as promising base editor tools to use in eukaryotic systems, such as human cells. Moreover, the reduced size of the *Mm*Cas12m base editor constructs makes them interesting candidates for their eventual therapeutic use in size-limited delivery strategies such as adeno-associated virus (AAV) particles.

Previous research attempted to use *Mm*Cas12m in a eukaryotic organism by showing that *Mm*Cas12m could be expressed and translocated to the nucleus of *Saccharomyces cerevisiae* (Chapter 4). At best, however, very low activity was detected when this tool was used for genome silencing approaches, possibly due to the transient binding of *Mm*Cas12m to its target. Still, the possibility of using *Mm*Cas12m for base editing in yeast remains open, since base editing requires only transient binding of the Cas protein and generates a permanent change on the DNA. However, *Mm*Cas12m-mediated base editing activity has not yet been reported in any eukaryotic organism.

The objective of this research was to determine whether *Mm*Cas12m-based base editors are suitable for base editing in the eukaryotic organism *S. cerevisiae*. For this purpose, an *in vivo* assay was designed to introduce stop codons in the coding sequence of *ADE2* by C-to-T base editing. Confirmation of base editing activity in the model organism *S. cerevisiae* would therefore open the possibility for future applications of *Mm*Cas12m base editors in higher eukaryotic organisms, including human cells.

5.3 Results

5.3.1 Effect on growth of *Mm*Cas12m-*Pm*CDA1-UGI expression

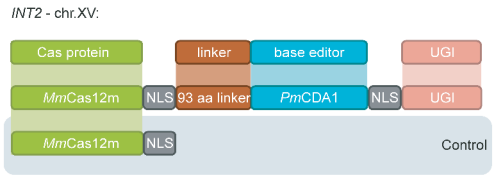
Initially, we built a yeast strain carrying a single copy of a *S. cerevisiae* codon-optimized *Mm*Cas12m-*Pm*CDA1-UGI gene integrated into its genome (strain YSTB305). To allow for translocation to the nucleus, a nuclear localization signal (NLS) was fused to the C-terminus of *Mm*Cas12m, followed by a 93 amino acids linker to which the base editing module was attached. The genes coding for the *Pm*CDA1 and UGI were codon optimized for expression in humans and separated by a second NLS (Figure 5.1A). Through standard homologous recombination, the *Mm*Cas12m-*Pm*CDA1-UGI construct was integrated into the *INT2* integration site (*110*) in chromosome XV together with the *URA3* gene from *Kluyveromyces lactis* for selection purposes. We confirmed the correct integration of the construct by PCR and sequencing of the genomic locus.

To assess the impact on growth of *Mm*Cas12m-*Pm*CDA1-UGI expression, we compared the growth rate of the generated strain (YSTB305) to the isogenic control strains YSTB164, which harbors a genome integrated *eGFP* expression construct, and YSTB211, which expresses both *eGFP* and *Mm*Cas12m from genome integrated copies. Growth rates of 0.31 ± 0.011 h^{-1}, 0.39 ± 0.004 h^{-1} and 0.37 ± 0.008 h^{-1} were determined for each strain, respectively, showing that the expression of *Mm*Cas12m fused to a base editor from a chromosomal-integrated copy has a significant impact on *S. cerevisiae* physiology, compared to the expression of *Mm*Cas12m alone (Figure S5.1) (t-test: N = 3, p-value = 0.0003 (for YSTB164) and p-value = 0.002 (for YSTB211)).

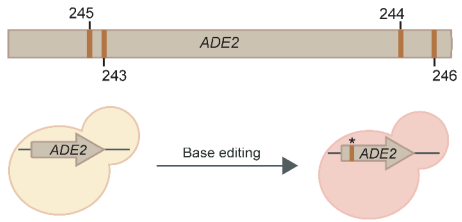
5.3.2 Base editing from a genome integrated expression cassette for *Mm*Cas12m-*Pm*CDA1-UGI

Strain YSTB305 was transformed with multicopy plasmids expressing *Mm*Cas12m RNA guides targeting the *ADE2* gene. This gene can be used as a reporter in *S. cerevisiae* because *ADE2* knockout strains accumulate P-ribosylamino imidazole, an intermediate of the adenine biosynthetic pathway that is oxidized into a red pigment in aerobic conditions (*347*). This process allows for easy discrimination of knockout and wild-type colonies on plates (Figure 5.1B). Based on the presence of a 5'-TTN PAM, we designed four targeting guides to generate C-to-T mutations in specific positions of the *ADE2* coding sequence, creating a nonsense mutation that prematurely stopped protein translation (Figure 5.1B and C). Base editing was assessed on plates after 2 hours of recovery in YPD, after 24 hours of recovery in YPD and after 24 hours of liquid growth in YPD and antibiotic G418. Recovery and growth in liquid medium were conducted to assess editing after an increased time of exposure to the base editor.

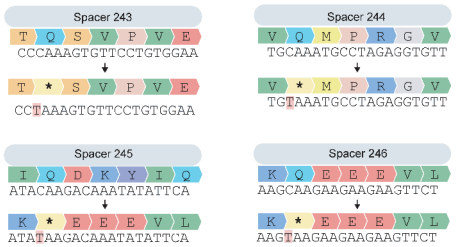
A



B



C



D

	Spacer 243 (YSTB321)	Spacer 244 (YSTB322)	Spacer 245 (YSTB323)	Spacer 246 (YSTB324)	Non-targeting (YSTB325)	Control *Mm*Cas12m (YSTB320)
Number of observed red colonies	11	14	2	14	4	0
Number of sequenced colonies	2	6	1	3	3	0
Genotype of sequenced strains	0 / 2	1 / 5	0 / 1	0 / 3	3 / 0	----------

on-target outside target

Figure 5.1. Set-up for *Mm*Cas12m genome-integrated base editing in *S. cerevisiae*. (A) Scheme of the *Mm*Cas12m-*Pm*CDA1-UGI and *Mm*Cas12m genome integrated constructs at *INT2* in chromosome XV. **(B)** Targeting positions of the 4 designed spacers (243-246) in the *ADE2* gene and phenotypical assessment of *ADE2* knockout strains. **(C)** Sequence of the designed spacers and introduction of stop codons. **(D)** Distribution of mutations in sequenced red strains.

Red colonies were observed only on plates with more than 1000 CFU and in the presence of the base editor. In all cases, the estimated percentage of red colonies per plate was <1% (Table S5.1). However, we were not able to assess the total number of CFU per plate because of the high colony density. Moreover, red colonies were often found in close contact or mixed with white colonies. Surprisingly, red colonies were also obtained on the transformation plates with the non-targeting guide (NT) (Figure 5.1D), but not in the strain expressing *Mm*Cas12m without the base editor (YSTB211) when transformed with the NT guide. Among the three conditions used for plating, plating after 24 hours of recovery in YPD yielded the lower number of red colonies, suggesting a possible loss of plasmid. However, cultivating the transformed strains in liquid medium under antibiotic selection for 24 hours did not show an increase in the number of red colonies per plate, compared to the plates obtained after recovering the strains for 2 hours in YPD in the absence of antibiotic (Table S5.1). This indicates that 24 hours seems not to be enough time to significantly increase the activity of base editing over the population.

Isolated red colonies on the transformation plates were re-streaked on fresh YPD + G418 plates to check for *ADE2* WT contamination. Sequencing of the *ADE2* locus of some of the isolated red colonies showed that 3 out of the 4 designed guides yielded targeted base editing, albeit at a low frequency (Figure 5.1D). Moreover, desired target modifications were observed only in the presence of their corresponding guide. In some of the sequenced colonies, other C-to-T mutations were also detected in the *ADE2* gene (at positions 1415 and 1517, data not shown), as well as an unclear genotype with no apparent mutation in the *ADE2* coding sequence (Figure 5.1D). Red colonies from the transformation plates with non-targeting guides did not show any of the desired C-to-T conversions.

5.3.3 Base editing from a plasmid-borne expression cassette

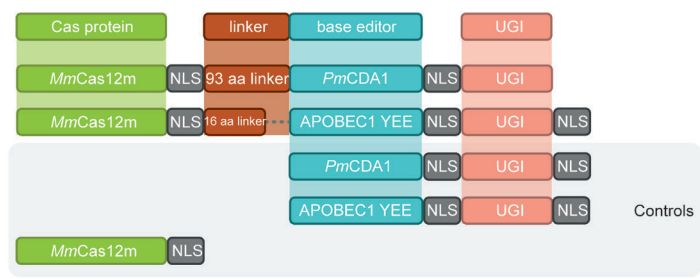
We explored other base editor combinations by building a new set of centromeric plasmids to assess *Mm*Cas12m base editing activity. All combinations were tested in biological triplicates. *Mm*Cas12m-*Pm*CDA1-UGI and *Mm*Cas12m-APOBEC1-YEE-UGI base editing activity was assessed after 2 hours of recovery on YPD medium and after 5 days of growth in liquid medium with G418 selection (Figure 5.2A).

Red colonies were observed on plates in all transformations when base editor domains were being transformed, either after 2 hours of recovery, or after 5 days of growth in liquid medium with antibiotic pressure. No plate showed more than 1.5% of red colonies (Table S5.2 and Table S5.3). As expected, the negative control expressing *Mm*Cas12m without base editors did not yield any red colonies. Controls with transformation of base editor domains alone (*Pm*CDA1-UGI or APOBEC1-YEE-UGI) yielded red colonies (Figure 5.2, Table S5.2 and Table S5.3), although mutations were not localized at any of the designed target sites. Among all sequenced red colonies from transformation plates obtained after 2 hours of recovery in YPD, targeted base editing with *Mm*Cas12m-*Pm*CDA1-UGI was shown with an approximate overall maximum frequency of 0.2%, showing that not all red colonies presented targeted editing. Using that recovery protocol, no targeted base editing could be detected with *Mm*Cas12m-

APOBEC1-YEE-UGI. However, we did observe targeted base editing for both *Mm*Cas12m-*Pm*CDA1-UGI and *Mm*Cas12m-APOBEC1-YEE-UGI at low frequencies (up to 1 every ~79 (1.3%) and ~76 (1.3%) colonies, respectively) on plates after growing the transformants for 5 days on liquid YPD with antibiotic G418. Therefore, both the number of red colonies obtained per plate and the frequency of targeting products relatively increased after extended incubation (5 days) in liquid medium, compared to the value obtained after 2 hours of recovery upon transformation.

Among all red colonies, only a fraction presented the desired genotype. For instance, we only detected targeted base editing by using *Mm*Cas12m-*Pm*CDA1-UGI expressed from the centromeric plasmid with spacer 246 (Figure 5.2B, Table S5.2 and Table S5.3). When this spacer was expressed, 9/10 genotyped red colonies showed the desired C-to-T modification after 5 days of growth in liquid medium with G418. Although red colonies were present on all other transformation plates, sequencing showed that C-to-T mutations were only present at non-targeted positions, or that no mutations were present in the *ADE2* coding sequence at all. In the case of *Mm*Cas12m-APOBEC1-YEE-UGI, we observed targeted base editing with spacers 244, 245 and 246 only after 5 days of growth in liquid medium with G418. Spacer 245 showed the highest targeted editing frequency with 36/39 red colonies presenting the desired genotype. Both *Mm*Cas12m-*Pm*CDA1-UGI and *Mm*Cas12m-APOBEC1-YEE-UGI showed some background base editing activity when expressed with a non-targeting guide or without guide.

A



B

*Mm*Cas12m-*Pm*CDA1 -UGI (YSTB347)	Spacer 243 (YSTB356) 0 / 3	Spacer 244 (YSTB357) 0 / 4	Spacer 245 (YSTB358) 0 / 4	Spacer 246 (YSTB359) 1 / 9	Non-targeting spacer (YSTB360) 0 / 3
*Mm*Cas12m-APOBEC1 -YEE-UGI (YSTB353)	Spacer 243 (YSTB376) 0 / 1	Spacer 244 (YSTB377) 0 / 1	Spacer 245 (YSTB378) 3 / 36	Spacer 246 (YSTB379) 0 / 1	Non-targeting spacer (YSTB380) 0 red colonies
*Pm*CDA1-UGI (YSTB354)	No guide control (YSTB385) 0 red colonies				
APOBEC1 -YEE-UGI (YSTB351)	No guide control (YSTB370) 0 / 4				
*Mm*Cas12m (YSTB355)	Non-targeting spacer (YSTB390) 0 red colonies				

Controls — on-target — outside target

Figure 5.2. Set-up for *Mm*Cas12m plasmid-borne base editing in *S. cerevisiae*. (A) Plasmid constructs used for expression of the *Mm*Cas12m-*Pm*CDA1-UGI and *Mm*Cas12m-APOBEC1-YEE-UGI constructs. Controls are indicated in the shaded area. **(B)** Distribution of mutations in sequenced red colonies obtained after plating cultures grown for 5 days in YPD+G418 medium. In case transformations did not show red colonies on plate, no graph is included.

Overall, this data indicates occurrence of the designed targeting activity of *Mm*Cas12m base editors, but at a relatively low frequency. We observed the desired C-to-T modifications only when the corresponding targeting spacers were expressed. At the same time, red colonies were also obtained that were at least partly the result of off-target base editing.

5.4 Discussion

In this study we tested for the first time *Mm*Cas12m C-to-T base editors in *S. cerevisiae*. We used a genome integrated base editor (*Pm*CDA1-UGI) and two centromeric plasmid-expressed base editors (*Pm*CDA1-UGI and APOBEC1-YEE-UGI). We successfully introduced targeted mutations using *Mm*Cas12m fused to either *Pm*CDA1-UGI or APOBEC1-YEE-UGI, although at rather low frequencies, even after 5 days of growth in liquid medium (<1.5%). These low base editing frequencies were observed both by plasmid-borne and chromosomal expression of the *Mm*Cas12m base editing tested variants. Moreover, we identified some spacers that yielded higher editing frequencies than others, which most likely reflects binding efficiencies being spacer sequence-dependent (*274*). Under the tested conditions, we also observed unspecific activity under control conditions when the base editors were expressed alone or fused to *Mm*Cas12m along with a NT guide. This unspecific activity yielded colonies with the same phenotype as the correctly edited ones but carried no mutation at the target site. This unspecific base editing activity was in the same order of magnitude as the targeted base editing which could indicate low affinity of the *Mm*Cas12m-base editors for the yeast DNA.

The study performed was based on the phenotypic assessment of introduced mutations through base editing. This set-up is based on the accumulation of a red pigment in colonies on plate upon successful editing, and it has been validated in previous Cas9 and Cas12a characterization assays in yeast (*84*, *111*, *348*, *349*). In the case of Cas9 and Cas12a, the high activity of the nucleases allowed for a precise quantification of the base editor efficiencies (*84*, *102*, *111*). However, when testing *Mm*Cas12m base editors, even upon plating of highly dense cultures, we observed only limited base editing events. The low number of observations in relation to the total amount of colonies prevented us from presenting our data in terms of exact mutation frequencies with standard deviations even when biological replicates were used (Table S5.1**,** Table S5.2 and Table S5.3). Other approaches have been used in literature to assess base editing, such as targeting a fluorescence protein-encoding gene and analyzing a population through flow cytometry. However, we believe these approaches would have missed the low number of observations in our population (*350*). Future directions could include deep sequencing of population samples to help quantify the base editing efficiencies (*106*). Yet, it is important to consider that the sequencing depth of each sample should be high enough to detect a significant amount of base editing events in each population.

Base editing on the designed target loci was observed only when the specific targeting spacer was expressed and never under the control conditions. These results indicate that the *Mm*Cas12m base editors were correctly expressed, translocated to the nucleus and able to target specific locations in the genome of *S. cerevisiae*. However, the low number of targeted

base editing events detected in our assays was comparable to the amount of unspecific activity we observed under control conditions. This might indicate poor binding of *Mm*Cas12m-base editors to yeast DNA. The low observed editing frequencies are not in line with reports of nCas9-*Pm*CDA1 or dCas9-*Pm*CDA1 base editors, which targeted activity is 1000-fold higher than the unspecific activity (*106*). Our data could indicate that *Mm*Cas12m is not able to stably bind the DNA of yeast, unlike dCas9. Other even smaller Cas proteins (such as Cas12f1), were also reported to have no detectable activity in mammalian cells, most likely due to inefficient binding (*291*), which is also discussed in Chapter 4.

In our study, we observed that the base editing frequencies obtained were not comparable to the ones obtained using nCas9(D10A) base editors in *S. cerevisiae* (*106*, *351*, *352*). The use of nCas9(D10A)-*Pm*CDA1-UGI in this model yeast allowed to achieve mutation frequencies varying from 16 to 47% in a similar assay to the one performed in this work, based on the disruption of *ADE1* (*106*). Similarly, the use of several nCas9-APOBEC-BE3 variants allowed for mutation frequencies of *CAN1* up to ~60% of the population (*352*). Although these systems use the same deaminases used in our work, the main difference probably lies in the ability of nCas9 (D10A) to generate a nick on the non-edited (target strand) (*71*). The introduction of this nick allows the cell repair mechanisms to use the base-edited, non-target strand as a repair template for the target strand (*339*), increasing the mutation frequencies obtained in the population (*106*). Beyond the suggested and discussed inefficient binding presented by *Mm*Cas12m to DNA in yeast, the inability of *Mm*Cas12m to introduce nicks, or to eventually generate a nickase, prevented us from achieving higher base editing efficiencies. A similar case was described for the yeast *Yarrowia lipolytica*, where dCas9-*Pm*CDA1 activity could not be observed on a phenotypic selection assay (*353*) although dCas9 can be used in this yeast for silencing purposes (*226*). Instead, base editing activity was reported for nCas9-*Pm*CDA1 after 24 hours of expression of the base editor (*353*) which shows, once more, that the introduction of the nick on the non-edited strand might be crucial to increase base editing efficiencies also in yeast. Although lower activities compared to nCas9 base editors were expected, the higher targeted vs. unspecific activity ratio of dCas9-*Pm*CDA1-UGI indicates that this construct would achieve higher base editing frequencies than *Mm*Cas12m-*Pm*CDA1-UGI. However, no similar assay based on the frequency of phenotype-based mutants in a population has been developed for dCas9-*Pm*CDA1-UGI (*106*).

Binding base editors (not nickase) cannot induce repair mechanisms on the non-base edited strand because they do not cleave it and therefore, a higher chance of reversion of the introduced modification on the base-edited strand is implied in their use (*282*).The modification of *Mm*Cas12m towards a nickase variant is not straight forward, since an earlier attempt to restore its inactivated RuvC catalytic residues (HHD > DED) did not yield a DNA-cleaving protein (*317*). Thus, the difficulty in generating a *Mm*Cas12m nickase might represent the major limitation for the use of *Mm*Cas12m base editors in eukaryotic organisms. Nevertheless, it should be noted that some Cas12m clades do possess the canonical RuvC catalytic residues (DED), suggesting they did not lose target nuclease activity. Assessment of the nuclease

activity of these homologs could facilitate the development of more efficient Cas12m base editors.

This work represents one of the first attempts to establish non-Cas9 base editors in the yeast *S. cerevisiae*. Other similar reports consisted in the use of APOBEC1 (a non-optimized version of APOBEC1-YEE) fused to RNA-targeting protein dCas13a (*350*). In a similar way to our study, Cas13a activity in *S. cerevisiae* had not been previously reported and no base editing activity could be detected when fusing dCas13a to APOBEC1. To date, the only successful base editing report in yeast not using dCas9 or nCas9 was performed in the fission yeast *Schizosaccharomyces pombe,* where a low 6% editing was observed on RNA when using dCas13a fused to the ADAR base editor (*354*). Similarly, dCas12a base editors have only been developed for bacteria and mammalian cells, but no application has yet been reported for its use in yeast (*355*).

With our constructs we were able to base edit positions +3, +4 and +15 of the protospacer in the target gene, similar to previously described results in *E. coli* (*317*). However, the base editing frequencies that we observed in our population were lower than the ones described in the model bacterium, where efficiencies close to 100% were reached (*317*). Different DNA topology, chromatin configuration and differences in the mismatch repair mechanisms between bacteria and yeast might interfere with the obtainment of the desired mutations (*343*).

It is important to consider that the constitutive expression of *Mm*Cas12m base editors might result in increased mutation rates and/or decreased growth rates. All tested base editors were constitutively expressed either from genome-integrated expression constructs or centromeric plasmids, similarly to what has been reported before for *Mm*Cas12m base editors in *E. coli*. Constitutive expression of base editors is known to result in an increased off-target mutation rate across the genome (*106*). In our study, we observed a reduction in growth rate caused by the expression of the genome-integrated *Mm*Cas12m-*Pm*CDA1 base editor. This may suggest a toxic effect from the constitutive expression of the base editor due to unspecific activity across the genome. In future studies, this can be circumvented by induction of the base editor expression as previously reported (*106*, *351*, *356*). However, the chosen approach did not hamper any measurement in our study due to the overall low activity of the studied protein (*106*, *351*, *356*).

In our work we showed that successful characterization of genome editing tools in *E. coli* is not a guarantee for their functioning in the yeast *S. cerevisiae*. Particularly, type V Cas proteins are extremely diverse and present a wide variety of functions, many of which still need to be characterized (*345*, *346*). A better understanding of the protein mechanism of action in their native hosts or under heterologous expression conditions will help to develop novel tools in eukaryotic organisms. Optimization of the studied protein (or related Cas12m proteins) and more careful characterization of the protein binding mechanism *in vivo* could help to develop a reduced size tool for its use in eukaryotic systems.

5.5 Materials and methods

5.5.1 Strains and growth conditions

E. coli DH5α and DH10B were used for cloning and propagation of plasmids. *E. coli* was routinely cultured at 37 °C and 200 rpm in Luria Bertani medium (LB) [10 g L^{-1} tryptone (Oxoid), 5 g L^{-1} yeast extract (BD), 10 g L^{-1} NaCl (Acros)]. Ampicillin was added for plasmid maintenance (100 μg·mL^{-1}).

S. cerevisiae CEN.PK113-5D or CEN.PK113-7D were used for all base editing characterization experiments as background strains. A complete list of strains can be found in Table S5.4. *S. cerevisiae* was routinely cultured in YPD medium [10 g L^{-1} yeast extract (BD), 20 g L^{-1} peptone (Oxoid) and 20 g L^{-1} glucose (Fisher Scientific B.V.)] or synthetic medium (SMG) [3 g L^{-1} KH_2PO_4 (VWR International B.V.), 0.5 g L^{-1} $MgSO_4 \cdot 7H_2O$, 5 g L^{-1} $(NH_4)_2SO_4$, 1 ml L^{-1} of a trace element solution, and 1 ml L^{-1} of a vitamin solution as previously described (*327*)]. Synthetic complete medium (SC) [6.7 g L^{-1} yeast nitrogen base without amino acids and with ammonium sulfate (SIGMA), 2.0 g L^{-1} Yeast Synthetic Drop-out Medium Supplements (Sigma Aldrich) and 150 mg L^{-1} uracil] was used for growth rate determination. 2% Glucose was used in all cases as a sole carbon source. All yeast cultures were grown at 30 °C, with shaking at 200 rpm in case of liquid cultures. When selection was required, media was supplemented with 200 mg mL^{-1} G418 (Geneticin) and/or 100 mg mL^{-1} Nourseothricin (ClonNAT). When selection with G418 was required on SMG media, $(NH_4)_2SO_4$ was replaced with 3 g L^{-1} K_2SO_4 (Sigma Aldrich) and 2.3 g L^{-1} filter-sterilized urea (Acros Organics) to avoid pH drop (*328*). Solid media were obtained by addition of 20 g L^{-1} agar (Oxoid).

5.5.2 Molecular biology techniques

PCR reactions for diagnosis purposes were performed using OneTaq® DNA Polymerase (New England Biolab, NEB) following manufacturer's instructions. Q5® High-Fidelity 2X Master Mix (NEB) was used for high fidelity amplifications according to supplier's instructions. Oligonucleotides used for cloning and sequencing were purchased from either IDT or Sigma Aldrich. When required, DNA fragments were purified from agarose gels using Zymoclean™ Gel DNA Recovery Kit (Zymogen). Bacterial plasmids were isolated using GeneJET Plasmid Miniprep Kit (Thermo Scientific). Gene insertions were confirmed by diagnostic PCR and Sanger sequencing (Macrogen, Germany).

5.5.3 Growth rate determination

To evaluate potential growth defects on strains constitutively expressing *Mm*Cas12m fused to *Pm*CDA1 base editor, strains YSTB305 and YSTB164 were grown in SC medium (supplemented with uracil) in 500 mL shake flasks. Shake flasks were filled with 100 mL of medium and incubated at 30°C and 200 rpm. Growth was monitored by measuring optical density (600 nm) using UV-1800 Spectrohpotometer (Shimadzu). Maximum specific growth rated for each strain were calculated from 3 biological replicates and 2 technical measurements.

5.5.4 Selection of target sites and design of crRNA

Protospacers allowing for screening of the C-to-T base editing activity were selected to generate premature stop codons in the open reading frame of the *ADE2* gene of *S. cerevisiae*. To this end, 4 protospacers were identified carrying a cytosine in position 3, 4 or 15. Plasmids PL-243 to PL-246 were constructed to contain the corresponding spacer sequences (Table S5.5).

5.5.5 Plasmid construction

The plasmids constructed in this study are listed in Table S5.5, as well as the assembly methods employed. Oligonucleotides used for cloning and sequencing are listed in Table S5.6. Briefly, *ADE2* targeting plasmids carrying the CRISPR repeats and crRNAs for *Mm*Cas12m (PL-243 to PL-246) were built by digesting PL-196 with BsaXI (NEB) and ligating previously annealed pairing oligonucleotides with required overhangs.

5.5.6 Construction of genome-integrated base editor strains

An initial small-scale experiment was designed with strains expressing the *Mm*Cas12m-*Pm*CDA1 base editor from a genome integrated copy. YSTB164 (Chapter 1) was used as the parental strain for transformation to obtain a *S. cerevisiae* strain expressing the *S. cerevisiae* codon-optimized *Mm*Cas12m fused to a *Pm*CDA1-UGI fusion construct. YSTB164 was transformed by using the lithium acetate method (*357*) with five linear fragments to obtain YSTB305: A0246 and A0247 were obtained by genomic amplification from *S. cerevisiae* CEN.PK113-5D by using primer pairs BG20385-BG20386 and BG20387-BG20388, respectively; A0295 was obtained by amplification of the *Mm*Cas12m gene from PL-140 with primers BG20383 and BG21049; A0296 was obtained by amplification of the human optimized *Pm*CDA1-UGI fusion construct from pSI-Target-AID-NG with primers BG21050 and BG21051; and A0297 was obtained by amplification of the *Kl*URA3 cassette from pUDE731. A control strain without the base-editor fusion construct was built by using YSTB164 as the parental strain and then transforming it with linear fragments A0246, A0247, A0248 and A0245 to obtain YSTB211. A0245 was obtained by amplifying the *Mm*Cas12m construct from PL-140 with primers BG20383 and BG20384, and A0248 was obtained by amplifying the *KlURA3* cassette from pUDE731 with primers BG20389 and BG20390. Transformants were plated on SMG media and incubated at 30°C until colonies appeared. Genomic integrations were confirmed by diagnostic PCR and sequencing the integration locus.

5.5.7 Construction of plasmid-borne base editor strains

S. cerevisiae CEN.PK113-7D was transformed with centromeric plasmids for expression of *Mm*Cas12m-*Pm*CDA1 (PL-277) or *Mm*Cas12m-APOBEC1-YEE (PL-296) to obtain strains YSTB347 and YSTB353, respectively. Control strains were built by transforming centromeric plasmids for expression of *Mm*Cas12m (PL-302), *Pm*CDA1 (PL-301) or APOBEC1-YEE (PL-293) to obtain strains YSTB355, YSTB354 and YSTB351, respectively. Transformants were plated on YPD agar plates containing Nourseothricin and incubated at 30 °C until colonies appeared. Diagnostic PCR was used to check for plasmid acquisition.

5.5.8 Assessment of base editing activity with genome-integrated *Mm*Cas12m-*Pm*CDA1 base editor

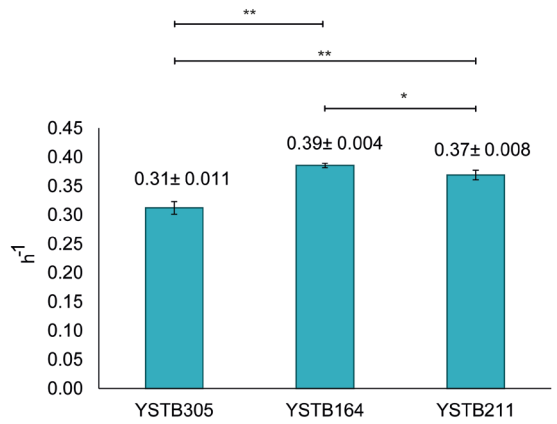
Subsequently, strains YSTB305 and YSTB211 were transformed with plasmids PL-243 to PL-246 and PL-139 (NT) to assess base editing activity. Transformants were plated on YPD selective media containing 200 mg mL^{-1} G418 (Geneticin) after 2 hours of recovery. At the same time, transformants were also inoculated in YPD medium with and without selection for overnight growth. Cultures were subsequently plated on YPD + G418 plates. Colonies were obtained on plates after 3 days of incubation at 30 °C. To favor red pigment accumulation in case of the *ADE2* knockouts, plates were incubated for an extra day at 4 °C. Pure or mixed red colonies were individually picked and re-streaked in YPD + G418 medium until single red colonies were isolated. High-fidelity genomic DNA amplification using primers BG11655 and BG11656 and Q5® High-Fidelity DNA Polymerase (NEB) was performed on isolated red colonies. PCR products were purified and analyzed with Sanger sequencing (Macrogen) with primers BG11656, BG11655, BG16485 and BG12982.

5.5.9 Assessment of base editing activity with plasmid-borne *Mm*Cas12m-base editors

Obtained strains were transformed in triplicate with plasmids PL-243, PL-244, PL-245, PL-246 and PL-139 to assess base editing activity. Transformations were recovered for 2 hours at 30 °C. After recovery, 100 µL of each transformation was used to inoculate 3 mL YPD + G418 in 24 deep well plates. Plates were covered with Breathe-Easy® sealing membranes (Diversified Biotech) and a 24-well plate sandwich cover. Plates were incubated at 30 °C and 250 rpm. After 48 hours, cultures were diluted 1:30 in fresh 3 mL YPD + G418 + Nourseothricin and incubated at 30 °C and 250 rpm. Every 24 hours, a new 1:30 dilution was performed in a fresh 24-deep well plate. After 5 days, 90 µL of a 10^{-4} dilution and 90 µL of a 10^{-5} dilution were plated on YPD agar + G418 + Nourseothricin plates. Plates were incubated for 3 days at 30 °C, after which phenotype and genotype were assessed as previously described.

5.6 Supplementary materials

A



B

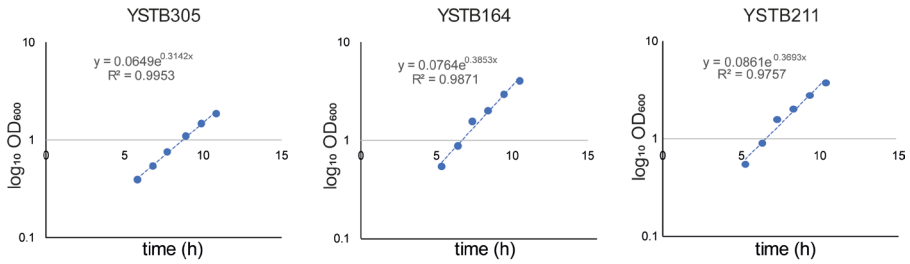


Figure S5.1. Specific growth rate of strain constitutively expressing *Mm*Cas12m-*Pm*CDA1-UGI base editor from a genome integrated construct at *INT2* and *egfp* at *INT1* (YSTB305) and its congenic control strains YSTB164 (expressing *egpf* at *INT1* and *Mm*Cas12m at *INT2*). Statistical significance was calculated with a t-test. (*) for p-value < 0.05 and (**) for p-value < 0.01.

Table S5.1. Population base editing frequencies with *Mm*Cas12m-*Pm*CDA1-UGI base editor expressed from a genome integrated copy (YSTB305). Strain expressing *Mm*Cas12m (YSTB211) was included as control. Plates presented more than 1000 colonies (exact number not assessed). Among the number of sequenced colonies, the number of colonies presenting targeted base editing is indicated in red.

		Plated after 2 h of recovery		Plated after 24 h incubation in YPD		Plated after 24 h incubation in YPD + G418	
		Number of red colonies per plate	Colonies sent for sequencing	Number of red colonies per plate	Colonies sent for sequencing	Number of red colonies per plate	Colonies sent for sequencing
YSTB211	Spacer 243 (YSTB316)	0		0		0	
	Spacer 244 (YSTB317)	0		0		0	
	Spacer 245 (YSTB318)	0		0		0	
	Spacer 246 (YSTB319)	0		0		0	
	Spacer 139 (YSTB320)	0		0		0	
YSTB305	Spacer 243 (YSTB321)	8		1		2	2
	Spacer 244 (YSTB322)	5		0		9	6 (1)
	Spacer 245 (YSTB323)	1		0		1	1
	Spacer 246 (YSTB324)	3		2	1	9	2
	Spacer 139 (YSTB325)	2	2	1		1	1

Table S5.2. Population base editing frequencies with *Mm*Cas12m base editors expressed from centromeric plasmids and plated after transformation. Base editing was tested using both *Pm*CDA1-UGI and APOBEC1-YEE-UGI deaminases although targeted base editing was only observed with *Mm*Cas12m- *Pm*CDA1-UGI base editor (frequency 0.2%).

		Red colonies	CFU/plate	Approximate % of red colonies	Targeted mutations	Targeted mutations (%)	Other mutations
Mm Cas12m-Pm CDA1 (YSTB347)	spacer 243	0	>1000	0.0%		0.0%	0
		0	>1000	0.0%		0.0%	0
		0	>1000	0.0%		0.0%	0
	spacer 244	0	990	0.0%		0.0%	0
		0	>1000	0.0%		0.0%	0
		0	>1000	0.0%		0.0%	0
	spacer 245	0	>1000	0.0%		0.0%	0
		0	>1000	0.0%		0.0%	0
		1	590	0.2%	1	0.2%	0
	spacer 246	0	>1000	0.0%		0.0%	0
		0	669	0.0%		0.0%	0
	non-targeting spacer	0	>1000	0.0%		0.0%	0
		0	>1000	0.0%		0.0%	0
		0	>1000	0.0%		0.0%	0
Mm Cas12m-APOBEC1-YEE (YSTB353)	spacer 243	0	>1000	0.0%		0.0%	0
		0	>1000	0.0%		0.0%	0
		0	>1000	0.0%		0.0%	0
	spacer 244	1	>1000	0.1%	0	0.0%	1
		1	>1000	0.1%	0	0.0%	1
		1	>1000	0.1%	0	0.0%	1
	spacer 245	0	>1000	0.0%		0.0%	0
		1	>1000	0.1%	0	0.0%	1
		1	>1000	0.1%	0	0.0%	1
	spacer 246	0	>1000	0.0%		0.0%	0
		1	>1000	0.1%	0	0.0%	1
	non-targeting spacer	1	>1000	0.1%	0	0.0%	1
		0	>1000	0.0%		0.0%	0
		0	>1000	0.0%		0.0%	0
APOBEC1-YEE control (YSTB351)	no guide	0	510	0.0%		0.0%	0
		0	560	0.0%		0.0%	0
		0	362	0.0%		0.0%	0
Pm CDA1 control (YSTB354)	no guide	1	202	0.1%	0	0.0%	1
		0	259	0.0%		0.0%	0
		0	192	0.0%		0.0%	0
Mm Cas12m control (YSTB355)	non-targeting spacer	0	95	0.0%		0.0%	0
		0	89	0.0%		0.0%	0
		0	91	0.0%		0.0%	0

Table S5.3. Population base editing frequencies with *Mm*Cas12m base editors expressed from centromeric plasmids and plated after 5 days of cultivation in liquid selective medium. Base editing was observed with both *Pm*CDA1-UGI and APOBEC1-YEE-UGI deaminases with maximal targeted mutation frequency per plate of 1.3% for both base editors. Both base editors showed background activity (*Mm*Cas12m-*Pm*CDA1-UGI when expressed with a non-targeting guide and APOBEC1-YEE-UGI when expressed without a guide). CFU/plate was estimated from the dilution plates when numbers are in italics.

		Red colonies	CFU/plate	Approximate % of red colonies	Targeted	Targeted mutations (%)	Other mutations	Total on-target	Total outside target	Total red colonies
Mm Cas12m-Pm CDA1 (YSTB347)	spacer 243	1	1094	0.1%	0	0.0%	1			
		0	170	0.0%	0	0.0%	0			
		1	1113	0.1%	0	0.0%	1			
		1	138	0.7%	0	0.0%	1			
		---	---	---	---	---	---			
		---	---	---	---	---	---	0	3	3
	spacer 244	0	845	0.0%	0	0.0%	0			
		0	113	0.0%	0	0.0%	0			
		1	993	0.1%	0	0.0%	1			
		1	129	0.8%	0	0.0%	1			
		1	1002	0.1%	0	0.0%	1			
		1	143	0.7%	0	0.0%	1	0	4	4
	spacer 245	0	948	0.0%	0	0.0%	0			
		0	123	0.0%	0	0.0%	0			
		1	962	0.1%	0	0.0%	1			
		1	121	0.8%	0	0.0%	1			
		2	1167	0.2%	0	0.0%	2			
		0	179	0.0%	0	0.0%	0	0	4	4
	spacer 246	2	958	0.2%	2	0.2%	0			
		2	159	1.3%	2	1.3%	0			
		1	940	0.1%	0	0.0%	1			
		2	158	1.3%	2	1.3%	0			
		3	1102	0.3%	3	0.3%	0			
		0	164	0.0%	0	0.0%	0	9	1	10
	non-targeting spacer	1	1970	0.1%	0	0.0%	1			
		1	197	0.5%	0	0.0%	1			
		0	2150	0.0%	0	0.0%	0			
		1	215	0.5%	0	0.0%	1			
		0	1520	0.0%	0	0.0%	0			
		0	152	0.0%	0	0.0%	0	0	3	3
Mm Cas12m-APOBEC1-YEE (YSTB353)	spacer 243	0	3840	0.0%	0	0.0%	0			
		0	384	0.0%	0	0.0%	0			
		0	2950	0.0%	0	0.0%	0			
		0	295	0.0%	0	0.0%	0			
		0	3000	0.0%	0	0.0%	0			
		1	300	0.3%	0	0.0%	1	0	1	1
	spacer 244	0	1750	0.0%	0	0.0%	0			
		0	175	0.0%	0	0.0%	0			
		1	2610	0.0%	1	0.04%	0			
		0	261	0.0%	0	0.0%	0			
		0	3080	0.0%	0	0.0%	0			
		0	308	0.0%	0	0.0%	0	1	0	1
	spacer 245	14	3230	0.4%	14	0.4%	0			
		5	323	1.5%	4	1.2%	1			
		4	2810	0.1%	4	0.1%	0			
		2	281	0.7%	2	0.7%	0			
		10	2290	0.4%	9	0.4%	2			
		3	229	1.3%	3	1.3%	0	36	3	39
	spacer 246	0	2930	0.0%	0	0.0%	0			
		0	293	0.0%	0	0.0%	0			
		1	2080	0.05%	1	0.05%	0			
		0	208	0.0%	0	0.0%	0			
		0	2180	0.0%	0	0.0%	0			
		0	218	0.0%	0	0.0%	0	1	0	1
	non-targeting spacer	0	1960	0.0%	0	0.0%	0			
		0	196	0.0%	0	0.0%	0			
		0	2540	0.0%	0	0.0%	0			
		0	254	0.0%	0	0.0%	0			
		0	2500	0.0%	0	0.0%	0			
		0	250	0.0%	0	0.0%	0	0	0	0
APOBEC1-YEE control (YSTB351)	no guide	0	2090	0.0%	0	0.0%	0			
		1	209	0.5%	0	0.0%	1			
		1	2210	0.0%	0	0.0%	1			
		0	221	0.0%	0	0.0%	0			
		1	2100	0.0%	0	0.0%	1			
		1	210	0.5%	0	0.0%	1	0	4	4
Pm CDA1 control (YSTB354)	no guide	0	1690	0.0%	0	0.0%	0			
		0	169	0.0%	0	0.0%	0			
		0	2470	0.0%	0	0.0%	0			
		0	247	0.0%	0	0.0%	0			
		0	1910	0.0%	0	0.0%	0			
		0	191	0.0%	0	0.0%	0	0	0	0
Mm Cas12m control (YSTB355)	non-targeting spacer	0	2070	0.0%	0	0.0%	0			
		0	207	0.0%	0	0.0%	0			
		0	2250	0.0%	0	0.0%	0			
		0	225	0.0%	0	0.0%	0			
		0	1930	0.0%	0	0.0%	0			
		0	193	0.0%	0	0.0%	0	0	0	0

Additional supplementary materials, tables and figures of this chapter can be accessed via:

https://figshare.com/s/90aa26769fa38e8adc30

CHAPTER 6

Type V-K CRISPR-associated transposon system – towards its application in eukaryotic organisms

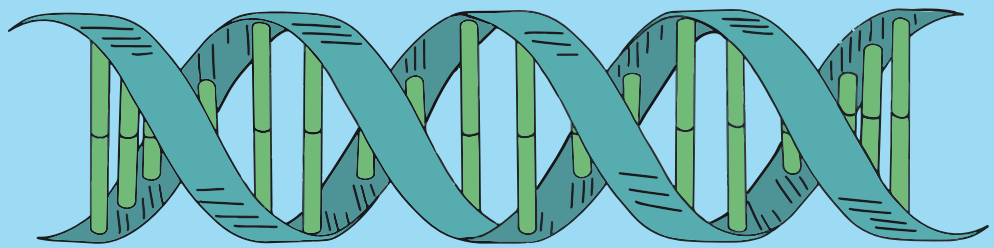
Belén Adiego-Pérez[1]
Guus Ververt[1]
Klaudia Ciurkot[2,3]
Hans Roubos[2]
Sarah D'Adamo[4]
John van der Oost[1]

[1] Laboratory of Microbiology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands

[2] DSM Biotechnology Center, Delft 2613 AX, The Netherlands

[3] Department of Chemistry, University of Hamburg, Hamburg 20146, Germany

[4] Bioprocess Engineering & AlgaePARC, Wageningen University and Research, P.O. Box 16, 6700 AA Wageningen, The Netherlands



6.1 Abstract

CRISPR-based tools have been harnessed in the last years for targeted incorporation of DNA sequences in prokaryotic and eukaryotic genomes. However, incorporation of large DNA fragments is a process dependent on the efficiency of the homology directed repair (HDR) mechanism. This process is usually either cell-stage dependent or not active at all in many organisms. CRISPR-associated transposon systems have recently been discovered and relevant mechanistic details have been revealed. These systems appear to be the result of the genetic incorporation of the gene(s) encoding a CRISPR effector into a gene cluster encoding a Tn7-like transposon. This chimeric system has been demonstrated to allow for RNA-guided DNA transposition in an HDR-independent manner. In this work, we set out to use of the ShCAST type V-K CRISPR-associated transposon system for genome editing in eukaryotic organisms. First, we investigate the effect of adding nuclear localization signals to each of the proteins of the complex on transposition. Although we observe that two configurations hamper transposition, six others do not affect the complex functionality and allow for transposition in *Escherichia coli*. Then, we show transposition occurring at lower temperatures than the ones previously reported in *E. coli* under *in vivo* conditions. Finally, when transplanting the system to *Saccharomyces cerevisiae,* we did not observe transposition. Considering the most recently published work, we speculate on different research directions towards the application of RNA-guided DNA transposition in eukaryotic organisms, and discuss recently reported alternatives.

6.2 Introduction

In the last years, clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR- associated Cas proteins have been harnessed for genome editing purposes in wide range of prokaryotic and eukaryotic organisms. This has contributed to substantially improving the genome editing efficiencies in various yeasts and fungi, for which a multitude of CRISPR-based genetic tools have been developed over the last decade (*84*, *94*, *110*, *111*, *113*, *120*, *187*). Some of these systems allow for the introduction of double strand breaks (DSBs) in the DNA of the target organism, which are lethal if not repaired. Some cells have an active homology directed repair (HDR) mechanism to fix these DSBs with the utilization of DNA fragments containing homologous sequences. Apart from natural recombination substrates, these DNA fragments can also be supplied externally and can contain any desired heterologous DNA sequence. This feature of HDR has been utilized to engineer new metabolic pathways into chromosomal locations and has allowed for control over the modifications introduced in the DNA sequence, as opposed to other DSBs repair pathways. However, HDR efficiencies decrease with the increasing size of the DNA repair templates, which limits the size of the DNA that can be introduced in a single event (*116*, *359*, *360*). Moreover, the HDR mechanism takes place in actively dividing cells and the efficiency of this process over other repair processes such as non-homologous end joining (NHEJ) is highly cell type and organism dependent (*96*, *361*). This means that, even though CRISPR-Cas systems have been established in multiple organisms, generated DSBs are often repaired by NHEJ, which limits the efficiency of HDR-mediated DSB repair. This is the case, for instance, in wild type (WT) strains of industrially relevant non-model yeasts such as *Kluyveromyces marxianus*. In these species, the favored DNA repair mechanism is NHEJ, and HDR typically occurs in less than 30% of the events upon DSB occurrence (*94*, *120*). Mammalian cells and plant cells also show low HDR efficiencies, which limits direct introduction of DNA sequences for therapeutic and production purposes (*362*, *363*).

Different strategies have been adopted throughout the years to increase the HDR efficiency over other repair pathways. Cell cycle synchronization and limitation of Cas9 expression to the $S/G_2/M$ cell cycle stages have been proven as valid strategies to increase the rates of HDR in eukaryotic cells (*94*, *96*, *121*, *364*). Expression of heterologous recombinases such as the l-red recombinase (*365–368*) in *E.coli* and related bacteria has also been shown to be an efficient strategy to increase homologous recombination when coupled to CRISPR-Cas9 as counter-selection systems. Temporal control of the action of the Cas nucleases has also been demonstrated to increase the efficiency of homologous recombination (*94*, *369*). Alternatively, gene silencing (*118*), protein inhibition (*370*) or knockout of some of the competing NHEJ pathway elements (LigD/Ku subunits) (*24–26*) has been shown to increase HDR activity in human cells and some non-conventional yeasts. Finally, the possibility of fusing characterized transposases to deactivated Cas proteins such as deactivated Cas9 (dCas9) has also been proposed and validated. Fusion of dCas9 to different transposases (the piggyBac transposase (*374*), the Sleeping Beauty (*98*), or the TnpA transposase (*258*)) have been used to influence the target site selection by directing the transposases to specific target sites determined by

the sgRNA. Although showing limited success or being restricted to a ssDNA donor, these systems do not rely on native DSB repair mechanisms, but rather on some external integration machinery and serve as proofs of concept for the development of novel tools. Moreover, they potentially could be used to integrate long DNA fragments (*374*). Laboratory evolution of some of these systems yielded integration efficiencies over 20%, with low or non-detectable off-target transposition events (*100*).

A very well studied bacterial transposon system is the Tn7 transposon. The classical Tn7 transposon is composed by proteins TnsA, TnsB, TnsC and TnsD and/or TnsE (*375*, *376*). In the canonical Tn7 transposon, TnsA and TnsB act together as an heteromeric transposase and catalyze the DNA strand breakage and joining reactions required for transposon mobilization from donor to target DNA (*377*, *378*). TnsB cleaves and joins the 3' ends of the transposon, while TnsA catalyzes the cleavage of the 5' end (*378*). Thus, they allow for cut-and-paste transposition by excising the transposon from the donor DNA (*379*). TnsC is an AAA+ ATPase protein that requires ATP to bind to DNA (*380*, *381*). This protein acts as a regulator protein and interacts both with the transposition core proteins TnsA and TnsB, as well as with TnsD or TnsE (*382*, *383*). These last two proteins determine the transposition fate: the TnsABC+TnsD complex directs transposition to *attTn7* sites in bacterial chromosomes while TnsABC+TnsE promotes transposition to non-*attTn7* sites found in mobile genetic elements (MGE) such as conjugative plasmids, although at lower efficiencies (*383*) (Figure 6.1A). Interestingly, natural CRISPR-associated Tn7-like transposon systems were recently described by bioinformatic analyses in sequenced bacterial genomes of Cyanobacterial, Streptomycetaceae and Vibrionaceae species (*384*, *385*). These analyses identified several Tn7-like transposon systems harboring incomplete CRISPR systems as cargo DNA between the characteristic transposon left and right terminal sequences, also called left and right elements (RE or LE).

These Tn7-like CRISPR-associated transposases (CASTs) were identified together with incomplete CRISPR systems of type I-B, I-F and V-U5, hereafter called V-K (*384*). All these systems share the following characteristics: (i) the absence of the adaptation module (*cas1*, *cas2* or *cas4*) in the *cas* operon; (ii) the absence either of a nuclease subunit (Cas3 in case of Class 1 complexes), or of a functional nuclease domain (RuvC, in case of Class 2 effectors) and; (iii) the absence of *TnsE* and substitution of this transposon subunit by the incomplete CRISPR system (*385*). Therefore, it has been proposed that these CRISPR-associated transposases utilize the recruited CRISPR systems to target transposition to MGEs and therefore facilitate horizontal transfer of the transposable element (*384*) (Figure 6.1B).

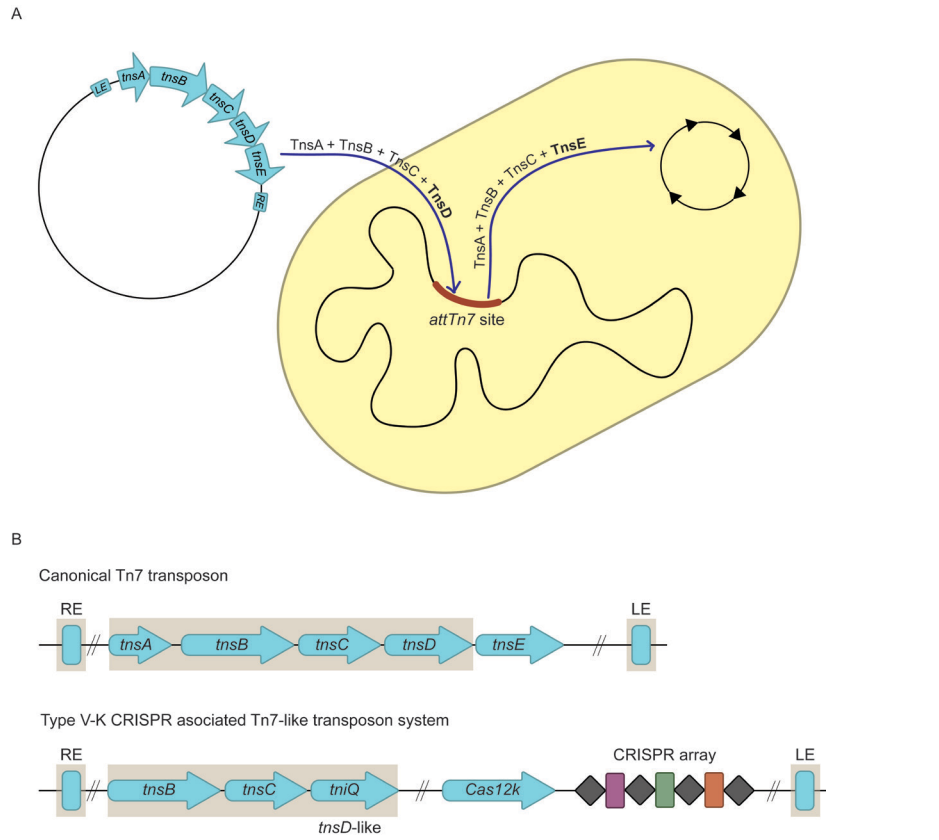


Figure 6.1. Canonical Tn7 transposon and association to type V-K CRISPR systems. (A) Canonical Tn7 transposition mechanism. Transposition to *attTn7* chromosomal sites is guided by TnsABC+D while transposition to mobile genetic elements (MGEs) is guided by TnsABC+TnsE. **(B)** Gene composition of canonical Tn7 transposons and type V-K CRISPR-associated Tn7-like transposons. Common components are shaded in grey.

Among all experimentally characterized systems, the type V-K system is the most compact one (*384*). This system is composed of *tnsB*, *tnsC* and *tniQ* genes together with a type V-K nuclease (Cas12k) and a CRISPR array (*258*) (Figure 6.1B) and has been used for RNA-guided DNA transposition *in vivo* and *in vitro* (*258*, *386*, *387*). Unlike Type-I CRISPR-associated transposon systems, this system allows for unidirectional transposition products. Similar to the transposon family Tn5053, the V-K system is characterized by the absence of the transposase subunit TnsA (*386*, *388–390*). This has been proven to lead to a high number of cointegration products with insertion of the whole donor plasmid and a duplication of the cargo sequence by a copy-and-paste mechanism (*391*). Many Tn5053-like transposons include a *tniR* resolvase that prevents the formation of the cointegrates (*390*, *391*). Because the type V-K CRISPR-associated transposases are not associated to any *tniR* resolvase, cointegrates are being generated, which is a non-desirable feature of the system (*386*, *388*). Supply of linear or 5' nicked donor DNA could prevent

the formation of co-integrants (*388*). To date, this system has been shown to allow RNA-guided DNA transposition in *Escherichia coli* although at lower specificity and efficiency than a type I-F CRISPR-associated transposon and in some other bacteria (*392*, *393*). The use of any CRISPR-associated transposon system has not been reported yet in any eukaryotic system (*391*).

Because of its simplicity, we here describe attempts to use a type V-K CRISPR-associated transposon system from *Scytonema hofmani* (ShCAST) to promote targeted DNA integration in eukaryotic systems by first using the model yeast *Saccharomyces cerevisiae*. Initially, we tested whether the system could function at lower temperatures than the ones tested to date (*258*). Aiming to use this system in organisms that grow at lower temperatures (e.g., *S. cerevisiae*), we investigated whether RNA-guided DNA transposition could take place *in vivo* in *E. coli* at lower temperatures than 37 °C. Moreover, activity of the RNA-guided transposon system in the eukaryotic organism can only be achieved when all protein components are successfully translocated to the nucleus. It has been shown that all components are necessary for transposition to take place (*258*). Normally, positively charged nuclear localization signal (NLS) peptides are fused to the N- and/or C-terminus of the proteins of interest for that purpose (*84*, *111*). However, addition of nuclear localization signals can disturb the protein structure or functionality (*317*), potentially generating inactive complexes. For these reasons, we studied the effect of addition of NLS peptides to each of the proteins of the complex at both the C and the N terminus of the proteins.

Finally, we used the gained insights to test transposition in the yeast *S. cerevisiae*. Ideally, this CRISPR-associated transposon systems could be used for integration of DNA in a homologous recombination-independent manner, which could be a useful tool for genome editing of eukaryotic cells with low HDR efficiencies.

6.3 Results

6.3.1 Temperature dependence of *in vivo* transposition activity of ShCAST

Previous work of the type V-K CRISPR-associated transposon system (ShCAST) was carried out in *E. coli* at 37 °C *in vivo* and in a range of temperature between 25-50 °C, *in vitro* (*258*). In the *in vitro* tests, it was observed that transposition did occur at 37 °C but not at 25 °C (*258*). With the aim to use the system in organisms growing at lower temperatures (e.g. *S. cerevisiae*), we tested the system *in vivo* in *E. coli* at different temperatures using the previously described ShCAST 2-plasmid system (*258*). We transformed the pHelper plasmid expressing the different ShCAST components (TnsB, TnsC, TniQ and Cas12k) and a sgRNA targeting PSP49 (*258*) into an *E. coli* strain harboring the pDonor plasmid. PSP49 targeted an *E. coli* genomic location proven to be highly accessible by the ShCAST transposition machinery. We grew the transformed populations overnight in liquid medium at different temperatures (25 °C, 30 °C and 37 °C). A sample from each liquid culture was used for PCR amplification of the genomic target site (Figure 6.2A). Transposition was observed *in vivo* at either 25, 30 or 37 °C (Figure 6.2B) and confirmed by sequencing results. No quantification of transposition efficiencies was performed for any temperature.

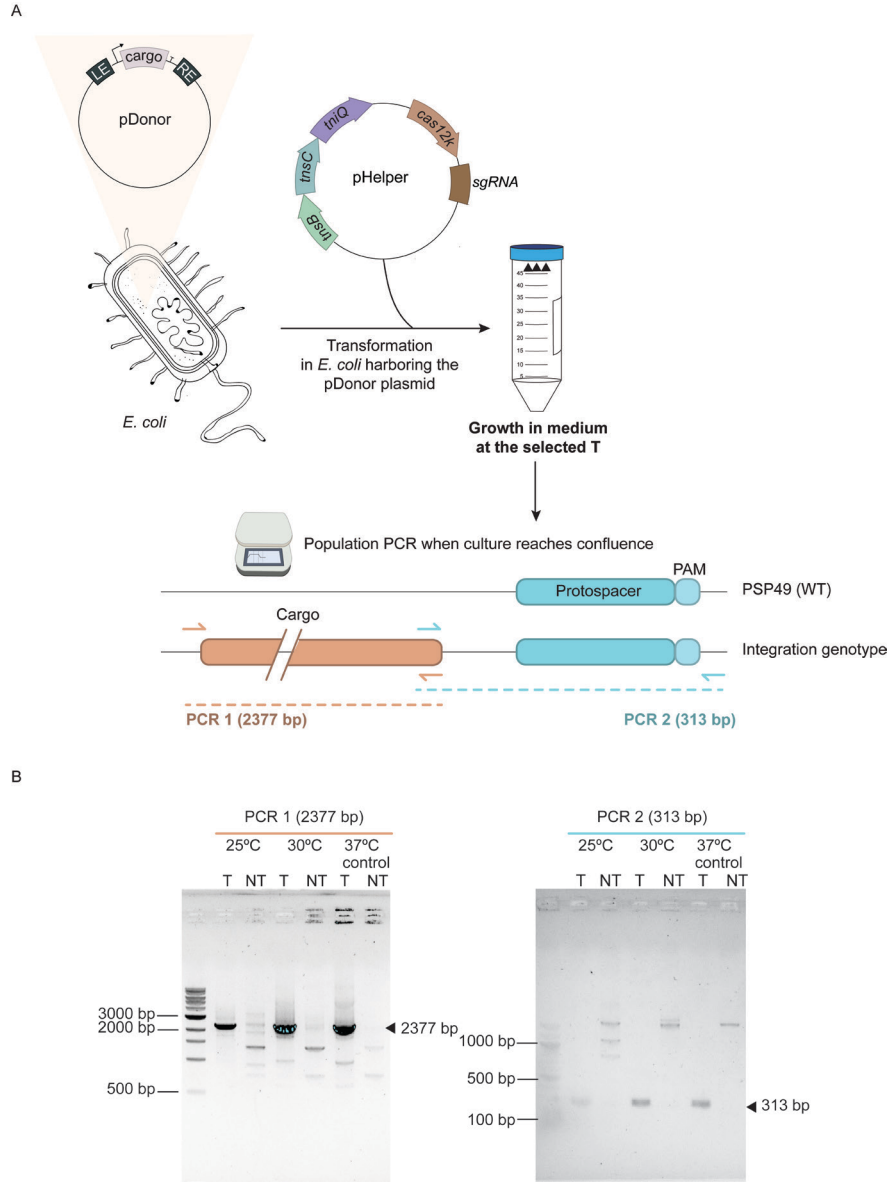
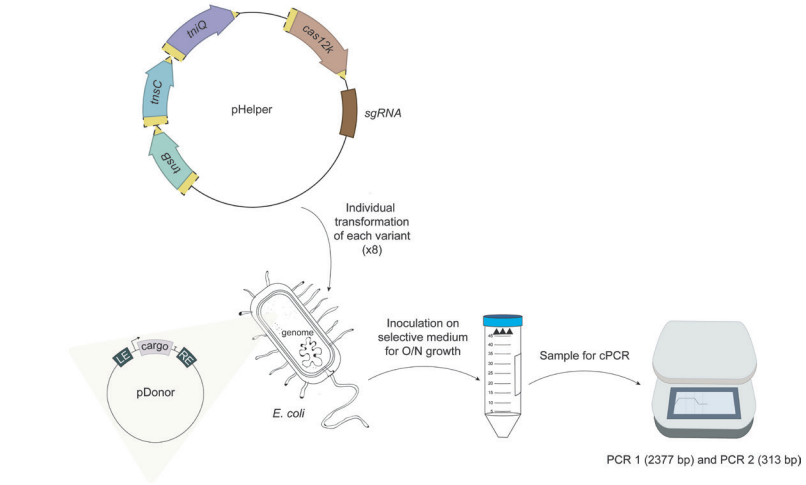


Figure 6.2. Diagnostic PCR of *E. coli* populations after ShCAST transposition. (A) Scheme of the followed protocol for qualitative assessment of ShCAST transposition in *E. coli*. A pHelper plasmid was transformed in *E. coli* harboring a pDonor plasmid. After heat shock and recovery at the selected temperatures (25 °C, 30 °C or 37 °C), transformants were inoculated in LB liquid medium and grown with antibiotic selection at test temperatures. Cultures were cultivated overnight until culture reached confluence. Population samples were used for PCR. Two different PCR reactions (PCR1 and PCR2) at the site of integration were used to diagnose target transposition in a qualitative way at a population level. **(B)** 1% agarose gel with obtained PCR1 and PCR2 products to demonstrate the transposition, and to investigate the effect of temperature (25ºC, 30ºC and 37ºC) on the wild-type transposition ShCAST system in *E. coli*. PCR reactions on populations transformed with a NT guide show a non-specific band pattern.

6.3.2 Effect of nuclear localization signal peptide fusion to individual elements of the type V-K CRISPR associated system

To assess protein and complex stability upon NLS addition, we used the previously described ShCAST 2-plasmid system as a starting point (*258*). We fused the Nucleoplasmin sequence "KRPAATKKAGQAKKKK" (*394*, *395*) to TnsB, TnsC, TniQ and Cas12k, independently to both their N and C-terminus. Effect of the addition of the sequence was evaluated for each variant, for the total of 8 different constructs. Transposition was assessed qualitatively at a population level in *E. coli* following the previously described transposition protocol (Figure 6.2A) (*258*). The 8 plasmids built for this assay were independently transformed into a *E. coli* strain harboring the pDonor plasmid. This plasmid contained the cargo sequence between the transposon's right and left elements (RE and LE). On-target transposition was confirmed by PCR amplification and Sanger sequencing of the target locus (Supplementary material 6.1) from a population sample. Sequencing results also showed mixed peaks, which could indicate the occurrence of the co-integration events recently described in literature for the ShCAST system (*386*, *388*, *391*). The results obtained from this assay showed that addition of the positively charged peptide sequence in 6 cases did not interfere with RNA-guided DNA transposition, except for the two constructs with the NLS at the C-terminus of TnsB, and at the C-terminus of TnsC (Figure 6.3A). Since all Tn7-like genes were expressed from an operon construct, addition of the short peptide encoding sequence at the 3' end of *tnsB* and *tnsC* could be affecting expression of the downstream genes *tnsC* and *tniQ*, respectively, by disrupting the respective ribosome binding sites (RBSs). However, even after addition of a second RBSs upstream of the coding sequences of *tnsC* and *tniQ*, no transposition could be observed (Figure 6.3A), which suggests that the positioning of the NLS in these proteins may interfere with the protein folding and/or activity, or with functional complex formation.



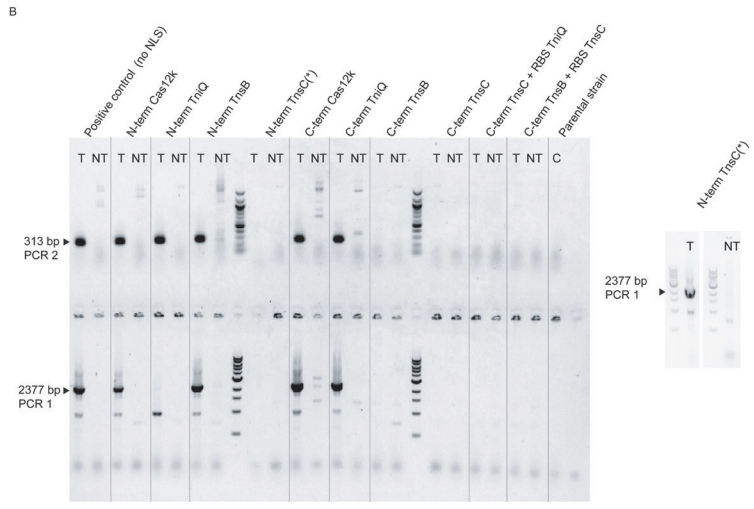


Figure 6.3. Scheme of the ShCAST test in *E. coli* and effect of NLS addition to individual ShCAST elements. (A) Scheme of the followed protocol for qualitative assessment of ShCAST transposition in *E. coli*. Each of the pHelper plasmid versions encoding for the different NLS-protein fusion variants of the ShCAST system (initially, a total of 8 variants) were individually transformed in *E. coli* harboring a pDonor plasmid. Two extra variants were designed after the first results which included an extra RBS upstream *tnsC* and *tniQ*. After heat shock and recovery, transformants were inoculated in LB liquid medium and grown with antibiotic selection until culture reached confluence. Two different PCR reactions (PCR1 and PCR2) at the site of integration were used to diagnose target transposition in a qualitative way at a population level. **(B)** 1% agarose gel with PCR1 and PCR2 results over the *E. coli* populations obtained after transforming the different NLS-protein fusion variants of the ShCAST system. Placement of the NLS sequence (N-term or C-term) is indicated on top of the agarose gel. Initially, 8 variants were built. Two more variants ("C-term TnsC + RBS TniQ" and "C-term TnsB + RBS TnsC") were added to the set-up. A positive control for transposition (a non-modified ShCAST system, without NLS addition) was included (first two samples on the left) as a control for transposition and absence of it. Each system was transformed with a targeting (T) and a Non-Targeting (NT) sgRNA guide. The parental strain was included as negative control for PCR. The effect of the N-terminal NLS on TnsC(*) had to be repeated because of PCR technical issues and only PCR 1 was used to diagnose transposition.

6.3.3 ShCAST activity in the yeast *S. cerevisiae*

To test the ShCAST activity in *S. cerevisiae*, we designed a 2-plasmid system for expression of the ShCAST system and aimed to integrate a donor DNA sequence into a previously described accessible integration site (*INT59*). The first plasmid of the system was the pHelper plasmid, a centromeric plasmid which encodes the CRISPR-associated Tn7-like transposon system proteins (TnsB, TnsC, TniQ and Cas12k). For proper expression of the proteins in yeast, all gene sequences were codon optimized. After determining that the NLS signals should not be placed at the C-termini of TnsB and TnsC because of hampered transposition, we avoided these positions for NLS placement. We fused all nuclear signals to the N-terminus of the Tn7-like proteins (TnsB, TnsC and TniQ) and to the C-terminus of Cas12k. From the results obtained in *E. coli,* placement of the NLS sequences to the N- or the C- terminus of Cas12k did not avoid transposition. For each of the components of the ShCAST system, we selected pairs of promoter-terminator sequences leading to similar protein expression (*396*). Details of the selected promoter and terminator sequences can be found in Supplementary sequence 6.1-4.

Two different strategies were designed and tested for supply of the transposable sequence carrying the cargo DNA in *S. cerevisiae*. Both strategies shared the sequence of the cargo, an expression cassette encoding a red fluorescent protein (*mRuby2*) for easy detection (Figure 6.4). For both strategies, first a *S. cerevisiae* strain harboring the pHelper plasmid was obtained.

The first strategy aimed to replicate *E. coli* tests in which supply of the donor DNA was performed from a circular plasmid (pDonor). Execution of this approach allows for RNA-guided DNA transposition in *E. coli*, although it has been linked to co-integration reports of the plasmid backbone and replication of the cargo DNA due to the absence of TnsA in ShCAST. Therefore, three linear DNA fragments with 50 bp shared homologous sequences were transformed into *S. cerevisiae* harboring the pHelper plasmid. The three linear fragments corresponded to: (i) the transposable element containing the cargo flanked by right and left transposon ends; (ii) the single guide RNA expression cassette and; (iii) a plasmid backbone. Upon transformation, these elements should recombine *in vivo* and generate a circular plasmid (pDonor) harboring the cargo sequence to match *E. coli* transposition conditions.

After transformation of *S. cerevisiae* harboring the pHelper plasmid following the first delivery strategy, cultures were recovered in YPD medium, overnight at 4 different temperatures (20 °C, 30 °C, 34 °C or 37 °C) before plating and incubating at 30 °C. Red-fluorescent colonies were found on all agar plates, independently of the recovery temperature (supplementary Figure S6.1A). However, colony PCR over the integration locus yielded WT genotypes for all tested red-fluorescent colonies, indicating the absence of transposition. Moreover, fluorescence of individual clones was progressively lost after cultivation of transformants over 92 hours without selection for the pDonor plasmid, and no fluorescent cells could be recovered (validated at flow cytometry, data not shown). This indicated that initial florescence observed on all agar plates was due to plasmid-borne expression rather than to expression after transposon-mediated chromosomal integration.

Since the type V-K CRISPR-associated transposon system does not encode for a TnsA subunit, it was tested in parallel whether supply of the donor DNA as a linear fragment could allow RNA-guided DNA transposition and decrease the number of co-integration events reported in *E. coli* (*386*, *388*). Therefore, in this second strategy we supplied two linear DNA fragments that should recombine into a plasmid upon incorporation in *S. cerevisiae* (plasmid backbone and sgRNA expression cassette) and a third linear DNA fragment that contained the sequence of the cargo DNA. This last DNA fragment did not contain any homologous sequence to any of the supplied linear DNA fragments and should thereafter be used as a linear donor DNA. After transformation of *S. cerevisiae* harboring the pHelper plasmid following this second strategy, cultures were recovered in YPD medium, overnight at 4 different temperatures (20 °C, 30 °C, 34 °C and 37 °C) before plating and incubating agar plates at 30 °C. Agar plates obtained after following this second strategy showed a lower number of fluorescent colonies or no fluorescent colonies at all (for the transformation recovered at 20 °C). Fluorescent colonies were only observed in populations recovered at 30, 34 and 37 °C (supplementary Figure S6.1B). However, isolation of the colonies from any of the plates was not possible, which indicates that fluorescence did not originate from chromosomal integration of *mRuby2*. Controls for random integration or random plasmid assembly were carried out at 20 °C and no fluorescent colonies were observed on plates (supplementary Figure S6.1C). All together, these results indicate that transposition did not occur under the tested conditions in *S. cerevisiae*.

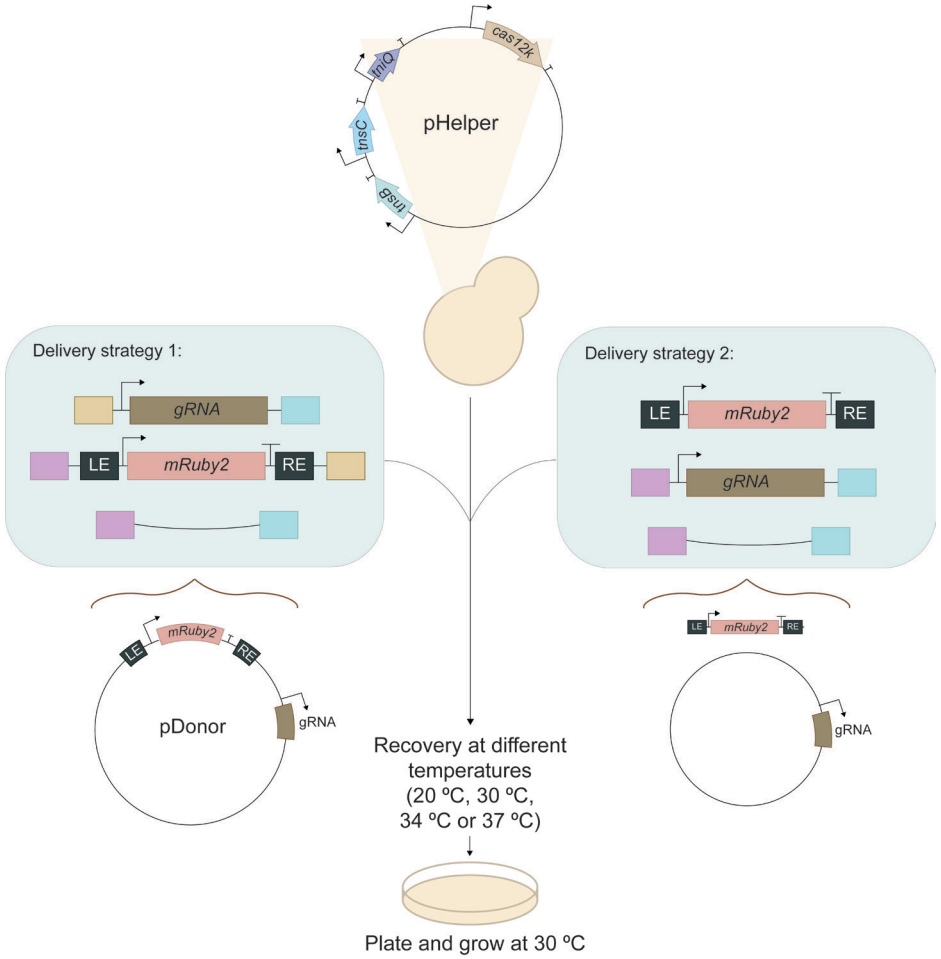


Figure 6.4. Scheme of the ShCAST test in *S. cerevisiae*. Each of the ShCAST components was expressed with a unique combination of promoter and terminator and introduced in the yeast *S. cerevisiae*. A second transformation was used to incorporate the donor sequence following two different strategies. In strategy 1, three linear fragments with homologous regions were used to transform the yeast, and they carried: (1) a gRNA cassette, (2) a *mRuby2* cassette between right and left ends and (3) pNR1120 backbone, respectively. In strategy 2, the donor DNA was introduced in the yeast as a linear fragment without homologous regions, along with a gRNA cassette and a pNR1120 backbone with homologous regions. Transformed yeasts were screened for fluorescence.

6.4 Discussion

In this work, we aimed to adapt the ShCAST system that has recently been established in different bacterial systems to *S. cerevisiae* as a proof of principle for other eukaryotic systems. Adoption of the CRISPR-associated transposon systems into different organisms requires study of its activity at different temperatures. Although the ShCAST system originates from the mesophilic cyanobacteria *Scytonema hofmanni*, *in vitro* characterization studies of the transposition complex

led to the conclusion that the system is active from 37 °C to 50 °C, and that no transposition was detected at 25 °C (*258*). As shown for different Cas12 variants, temperature is a key parameter that can impair Cas proteins activity (*397*, *398*). However, transposition dynamics *in vivo* and *in vitro* are difficult to compare because of the different nature of both systems. In our study, we showed that in *E. coli* transposition can still take place *in vivo* at 25 and 30 °C , suggesting that the system can be adapted to work in plants or yeasts, which are mostly cultured at temperatures below 37 °C. Efficiency of the transposition at these temperatures was not assessed.

Translocation of proteins into the nucleus of eukaryotic cells requires the addition of nuclear localization signal (NLS), a sequence encoding a short stretch of positively charged amino acids that enable protein import into the nucleus. Although translocation to the nucleus has been shown for some Cas proteins without the addition of an NLS (*112*, *399*), it is a common procedure to include these synthetic localization peptides (*84*, *111*, *123*). In the case of protein complexes, translocation to the nucleus can take place upon pre-assembly of the protein complexes in the cytoplasm (*400*). Therefore, addition of a single NLS should trigger the translocation of the whole complex. However, when expressing heterologous complexes in eukaryotic cells, generally all protein subunits are labelled for their translocation to the nucleus (*401*, *402*). Often, the addition of non-native small peptides or small proteins to the proteins of interest can disturb their folding and/or activity (*291*, *317*, *403*), therefore a proper assessment needed to be done also for the ShCAST. In this work, we assessed and showed that addition of NLS to single subunits did not interfere with *in vivo* transposition, when the NLS is placed at the N terminus of any of the ShCAST proteins or the C-terminus of Cas12k or TniQ. However, we observed that when the NLS was placed to the C-terminus of TnsB or TnsC, *in vivo* transposition did not occur. This indicates that TnsB and C are likely more sensitive to the position of NLS, possibly affecting relevant interactions in ShCAST protein complex. It should be noted, that in the *E. coli* trials only a single NLS per complex was present, whereas in the yeast experiments each of the 4 subunits had an NLS. Therefore, we cannot rule out that presence of an NLS on each of the ShCAST subunits might impair complex formation. A future test in *E. coli* should mimic the yeast strategy and test the transposition activity in case of multiple NLS tags on 2-4 subunits of the complex. Moreover, it still remains unclear whether nuclear labelling of a single ShCAST complex protein could allow translocation of a cytoplasm pre-assembled complex.

The ShCAST system has recently been reported to undergo a series of coordinated structural rearrangements that eventually lead to RNA-guided DNA transposition, most likely requiring subunit interactions that are not fully characterized to date (*404*, *405*). For instance, TnsC is an AAA^+ ATPase protein that polymerizes to form helical hexamers around the target DNA. This allows for the formation of filaments that remodel the DNA duplex (*405*). At the same time, it interacts directly with Cas12k and TnsB. The latter transposase protein is in charge of cleaving and joining the 3′ ends of the transposon DNA (*391*, *406*). It has been shown that the C-terminus of TnsB interacts with TnsC to unlock its ATPase activity and trigger filament disassembly of TnsC. Addition of the small positively charged peptide to the C-terminus of these proteins can destabilize protein folding, or any of the above-mentioned enzymatic activities or protein

interactions. Pull down assays with recombinant proteins containing each of the NLS could help elucidate which interactions are being disrupted by addition of the positively charged small peptides. Alternatively, lack of production of the individual proteins could indicate instability or degradation of the specific protein.

During the preparation of this chapter, another manuscript was released which indicates that placement of fusion proteins at the C-terminus of either TnsB or TnsC proteins abolished ShCAST transposition in *E. coli* (*407*). These results corroborate our results and show that, especially in the context of a multi-subunit complex, some proteins are more sensitive to protein fusions than others. On the other hand, the same study shows that certain protein fusions between the components of the ShCAST system are allowed (*407*). The use of ShCAST elements fusion products could potentially stabilize the complex and reduce the amount of NLS tags that should be added to the system. For instance, the interaction of TnsC with Cas12k has been described in the well-characterized ShCAST system. At the same time, the C-terminal fusion of TnsC to Cas12k did not avoid transposition in *E. coli* (*407*). Protein fusion products could minimize the number of protein interactions required to undergo transposition, stabilize complex formation, and facilitate the transposition process in other organisms.

After the characterization work in *E. coli*, we assessed the activity of the ShCAST system in *S. cerevisiae*. However, our assays carried out at different temperatures *in vivo* in *S. cerevisiae* did not show any RNA-guided transposition events for ShCAST systems, indicating that the system will need further optimization. To date, no CRISPR-associated Tn7-like transposon system (type I-F nor type V-K) has been described to work in any eukaryotic organism. Although in our study we did not investigate the reasons why RNA-guided DNA transposition in *S. cerevisiae* is not occurring, we propose several future research approaches that could help the establishment of this system in eukaryotes, based on the recently published characterization works (*404*, *405*, *407*, *408*),

In the first place, the ShCAST system contains a type Cas12k (type V-U5) protein. Other small size Cas12 variants such as Cas12m (type V-U1) or Cas12f1 (type V-U3) (*397*) initially showed low or no activity in eukaryotic cells. This has been attributed to poor protein binding to eukaryotic dsDNA *in vivo*. In the case of Cas12f1, wild type protein sequences showed no detectable activity in mammal cells (*291*, *397*). Modifications in the Cas12k protein sequence by addition of positively charged amino acids in the nucleic acids interacting domains could increase DNA and gRNA binding affinity to Cas12k. In the case of Cas12f1, this strategy allowed for the detection of dsDNA cleavage in human cells (*291*). Moreover, modifications in the gRNA structure (or individual crRNA and tracrRNA) associated to Cas12k might increase its stability *in vivo* and render detectable editing efficiencies. Again, this strategy has been shown effective for Cas12f1 in human cells (*132*, *291*). Moreover, single particle studies allowing to track the trajectory of proteins inside a cell (*325*) could help understand Cas12k behavior inside the nucleus. Second, the process of DNA transposition requires the orchestration of TnsC, TnsB and TniQ, to act and interact together with Cas12k for DNA transposition to take place. As shown in recently published work, TnsC polymerizes and forms filaments around the DNA

(*404*, *405*). However, this has only been shown to happen under *in vitro* conditions or in the bacterial DNA context. DNA structure in eukaryotic cells shows higher organization levels, with histone proteins packing the DNA into nucleosomes and a higher level of DNA packaging with the formation of chromatin fibers. Although the DNA in bacteria also interacts with proteins (*409*), the packaging level of eukaryotic DNA may hamper the formation of TnsC filaments around the DNA helix and therefore avoid further transposition events. As mentioned above, protein engineering and expression of fusion protein products of the ShCAST system could allow to promote or stabilize protein interactions between the proteins of the complex and to DNA (*407*). On the other hand, successful genome editing by different types of natural and synthetic CRISPR-associated effector proteins (nucleases, base editors, prime editors) has been described, showing that it should be possible to also use CRISPR-associated transposons for engineering eukaryotic DNA (*392*, *410*).

Further experimental characterization is required for an eventual use of the ShCAST transposon system or other Tn7-like transposon systems in eukaryotic organisms. Following the results of the present study, the correct expression and localization of each of the ShCAST components in *S. cerevisiae* should be corroborated. Fusion of small activable fluorescent tags could be an option to detect proper protein expression and localization with minimal decrease of protein expression. Besides, further protein structural work needs to clarify the directionality of the TnsC growing filaments since two different hypotheses have been proposed to date that need to be validated (*404*, *405*). Moreover, the role of TniQ into capping TnsC filaments needs to be put into context in the *in vivo* transposition set-up. Finally, the interaction of TnsB to TnsC or to any other subunit of the complex should be further characterized, as well as the rearrangements during the different steps of the transposition process. Structural information could clarify the order of the events that lead to RNA-guided DNA transposition and therefore be used to propose eventual protein fusion products between ShCAST components that could help to stabilize the complex and allow for RNA-guided DNA transposition in eukaryotic cells.

6.5 Materials and methods

6.5.1 Strains and growth conditions

E. coli DH5α and DH10B were used for cloning and propagation of plasmids. *E. coli* Pir1 were used for maintenance of pDonor plasmid and *E. coli* transposition assessment experiments. *E. coli* was routinely cultured at 37°C and 200 rpm in Luria Bertani medium (LB) [10 g L^{-1} tryptone (Oxoid), 5 g L^{-1} yeast extract (BD), 10 g L^{-1} NaCl (Acros)]. Ampicillin and kanamycin were added for plasmid maintenance at final concentrations of 100 µg·mL^{-1} and 50 µg·mL^{-1}, respectively.

S. cerevisiae CEN.PK113-7D was used as a recipient strain for assessment of the ShCAST in *S. cerevisiae*. *S. cerevisiae* was routinely cultured at 30 °C and 200 rpm in YPD medium (20 g L^{-1} peptone, 10 g L^{-1} yeast extract, 20 g L^{-1} glucose), unless specified otherwise. Agar was added to a final concentration of 2% for cultivation on plates. G418 was added for plasmid maintenance at a final concentration of 200 µg·mL^{-1}, when required.

6.5.2 Molecular biology techniques

PCR reactions for diagnosis purposes were performed using OneTaq® DNA Polymerase (New England Biolab, NEB) or Phusion polymerase (NEB) following manufacturer's instructions. Q5® High-Fidelity 2X Master Mix (NEB) was used for high fidelity amplifications according to supplier's instructions. Oligonucleotides used for cloning and sequencing were purchased from either IDT or Sigma Aldrich and can be found in Table S6.1 and Table S6.2. When required, DNA fragments were purified from agarose gels using Zymoclean™ Gel DNA Recovery Kit (Zymogen) or Wizard SV Gel and PCR Clean-Up System (Promega). HiFi assembly (NEB) was used for construction of plasmids. All other required enzymes were acquired at NEB. Bacterial plasmids were isolated using GeneJET Plasmid Miniprep Kit (Thermo Scientific). Gene insertions were confirmed by diagnostic PCR and Sanger sequencing (Macrogen, Germany).

6.5.3 Plasmid construction for *E. coli* ShCAST expression plasmids

In order to assess the effect of adding small positively charged peptides used as nuclear localization signals to each of the proteins of the ShCAST system, we based our assay on the 2 plasmids developed by Strecker *et al.*,(*258*). pHelper_ShCAST was a gift from Feng Zhang (Addgene plasmid # 127922; http://n2t.net/addgene:127922; RRID:Addgene_127922) and pDonor_ShCAST_kanR was a gift from Feng Zhang (Addgene plasmid # 127924 ; http://n2t.net/addgene:127924 ; RRID:Addgene_127924). Plasmids used for this assessment were built as indicated in Table S6.3. Intermediate cloning plasmids for cloning of spacers were built which expressed *rfp* between SapI restriction sites (PL-215 to PL-223 and PL-285). The *rfp* expressing plasmids were subsequently digested with SapI (NEB) and used for cloning PSP49 targeting (*258*) and non-targeting spacers by oligonucleotide annealing and ligation.

6.5.4 Transposition assessment in *E. coli*

E. coli Pir1 cells harboring the pDonor_ShCAST_kanR plasmid were made chemically competent (*411*) and used for transposition assessment. 50 µL aliquots were transformed with 5 ng of pHelper plasmid variants with a Targeting guide (T) or a Non-targeting guide (NT) (PL-224 to PL-241 and PL-297 to PL-300). Cells were recovered in 1 mL LB for 20 minutes and inoculated in 5 mL LB supplemented with ampicillin and kanamycin. When assessing the effect of temperature on transposition, recovery was done at 25 or 30°C, respectively. 1 µL of overnight confluent cultures was used for PCR using primers BG21035 and BG22585 (PCR 2) and BG21036 and BG22587 (PCR 2). Amplicons were also obtained with Q5 polymerase and sequenced by Sanger sequencing.

6.5.5 Plasmid construction for *S. cerevisiae* ShCAST expression plasmids

A pHelper plasmid was designed for expression of the ShCAST components in *S. cerevisiae*. Sequences of each of the ShCAST genes and details on the chosen promoter and terminator sequences for gene expression can be found in Supplementary materials (Supplementary sequence 6.1-6.4). All genes were codon optimized for *S. cerevisiae* and ordered from Twist Bioscience. The designed plasmid contained a CEN/ARS replication sequence for maintenance in *S. cerevisiae*. DNA linear fragments were introduced into *S. cerevisiae* for *in vivo* plasmid assembly. Plasmids were recovered from transformed yeast cells and transformed in *E. coli* for plasmid amplification and maintenance. Plasmid sequence was confirmed by Sanger sequencing using the BigDye™ Terminator v3.1 Cycle Sequencing Kit (ThermoFisher Scientific) and NucleoSEQ columns for dye terminator removal (Macherey-Nagel). The analysis was performed with Genetic Analyzer 3500XL (Applied Biosystems).

Different pDonor plasmids were designed for expression of gRNA. Plasmids used for strategy 1 also contained the cargo sequence expressing *mRuby2* between ShCAST right and left ends (Supplementary sequence 6.5). Plasmids were assembled by *in vivo* assembly in *S. cerevisiae*. Spacer sequences were designed to target the intergenic non-coding region (*INT59*), previously characterized as highly accessible site for genome editing using Cas12a and Cas9. Spacer sequences and gRNA cassettes used for transformation can be found in Supplementary sequences 6.6-7. All sequences contain the 50 bp linkers used for homologous recombination of fragments required for *in vivo* assembly.

6.5.6 Transposition assessment in *S. cerevisiae*

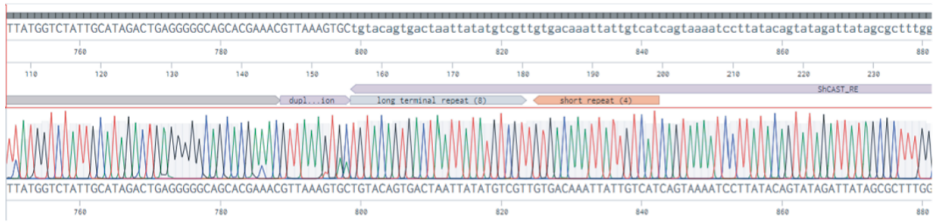
A *S. cerevisiae* strain harboring the pHelper plasmid was transformed following two different strategies to assess RNA-guided DNA transposition in *S. cerevisiae*. Following strategy 1, three linear fragments were transformed into the yeast strain: (1) pNR1120 backbone with NatMX cassette, (2) gRNA expression cassette with SNR52 promoter and SUP4 terminator, (3) cargo sequence containing the ShCAST right and left ends and expressing the *mRuby2* sequence between *S. cerevisiae* compatible promoter and terminator. All these linear fragments were flanked by 50 bp sequences that allowed *in vivo* assembly of all transformed fragments. Following strategy 2, two linear fragments were transformed into the yeast strain: (1) pNR1120 backbone with NatMX cassette and (2) gRNA expression cassette with SNR52 promoter and SUP4 terminator. In this case, the cargo sequence did not contain 50 bp flanking sequences for *in vivo* assembly. Linear fragments were introduced in the yeast using the LiAc/PEG transformation protocol (*412*) and recovered at different temperatures overnight (20°C, 30°C, 34°C and 37°C). Reagents required for yeast transformation were obtained from Sigma Aldrich (Lithium acetate dihydrate (LiAc) and deoxyribonucleic acid sodium salt from salmon testes (ssDNA)) and Merck (polyethylene glycol 4000 (PEG)).

Cells were plated in dilution series on selection plates. Fluorescence of transformants was visualized with a Qpix 450 colony picker using FITC wavelength (Molecular Devices). Colony PCR to assess targeted transposition was performed using Phusion Polymerase. In case of strategy 1, cells were resuspended in non-selective liquid medium and cultured for plasmid loss. Cultures were analyzed using flow cytometry. Flow cytometry was conducted using BD FACSAria Fusion (Becton–Dickinson). Detection of events was set such that 20,000 events were measured for single cells and double cells were excluded from the analysis. The signal of fluorescent proteins was detected with a bandpass filter at 610/20 nm for RFP. The data was recorded using BD FACSDiva 8.0.2 software. Measurements were performed after 24, 48, 72 and 92 hours of cultivation without selective pressure for the pDonor plasmid. In case of strategy 2, fluorescent cells spotted on plates were re-streaked on non-selective plates for isolation.

6.6 Supplementary materials

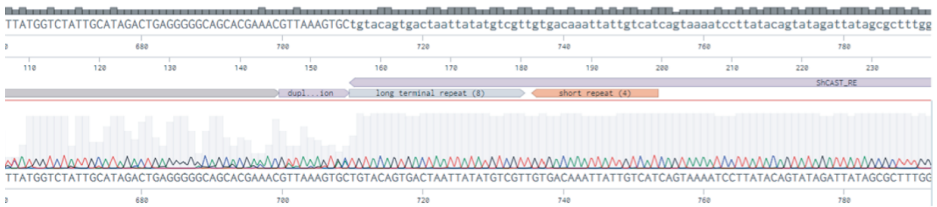
Supplementary material 6.1. Sanger sequencing results of amplified chromosomal location (PSP49). Sequencing reactions show that transposition was achieved at the target locations. However, mixed peaks were observed for almost all locations which could indicate mixed amplifications of simple and co-integration transpositions, as previously described for the ShCAST system.

No NLS (Wild type)



N-term Cas12k

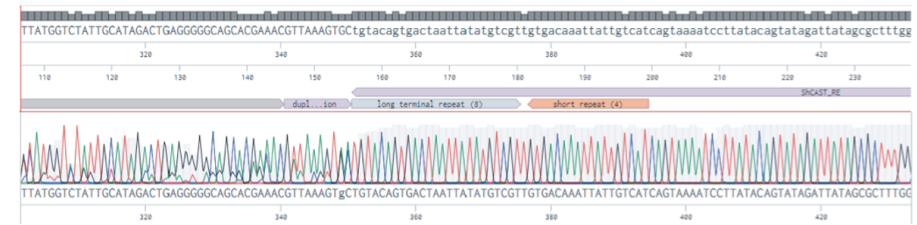
N-term TniQ



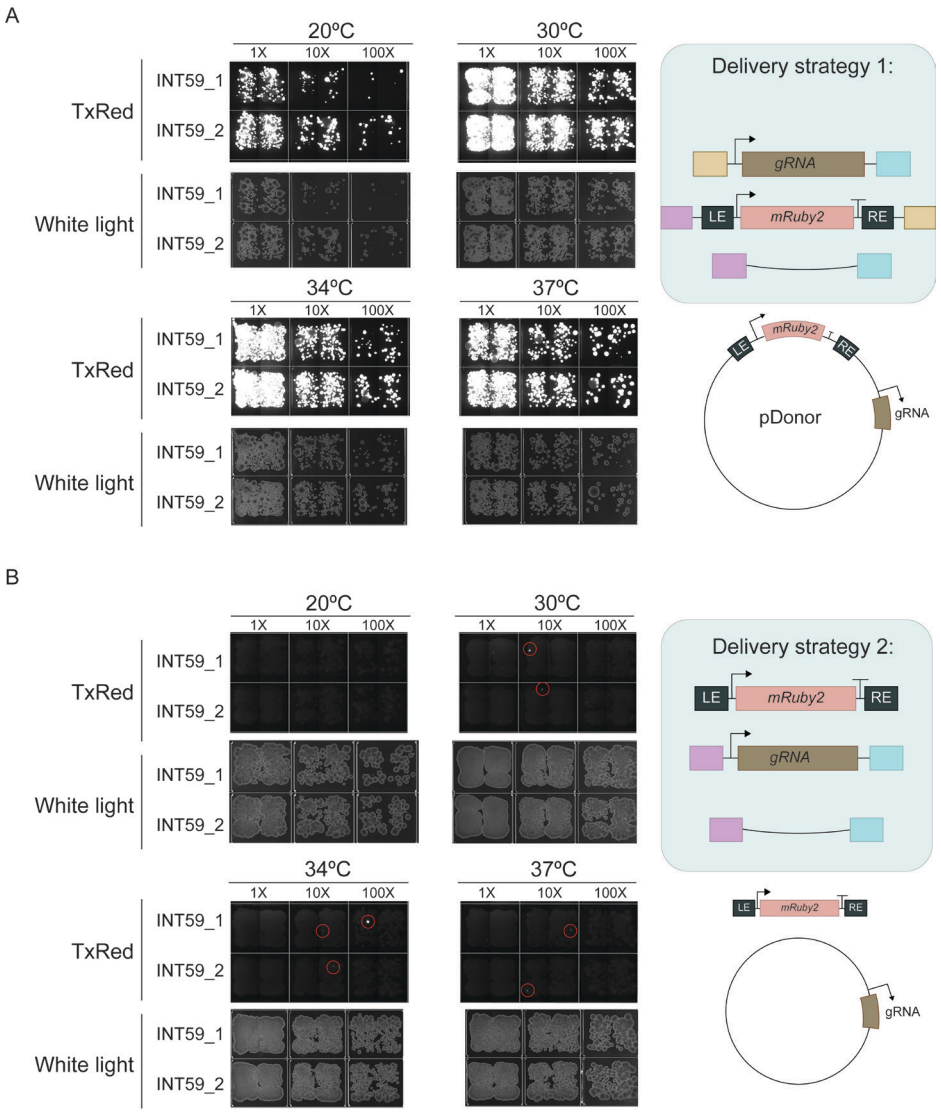
N-term TnsB

N-term TnsC

C-term Cas12k



C-term TniQ



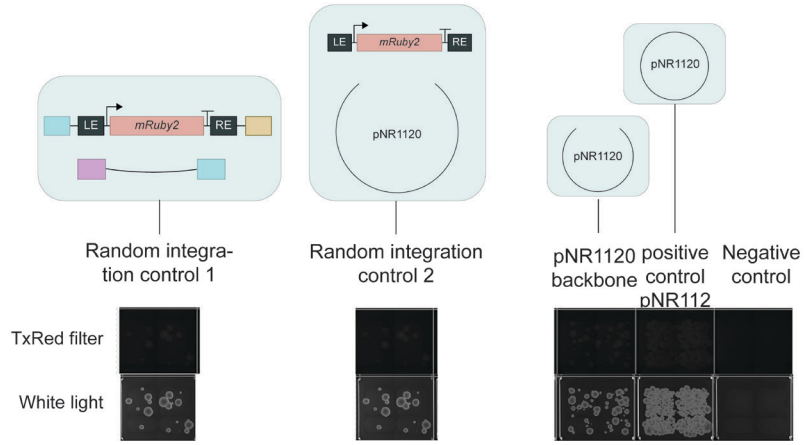
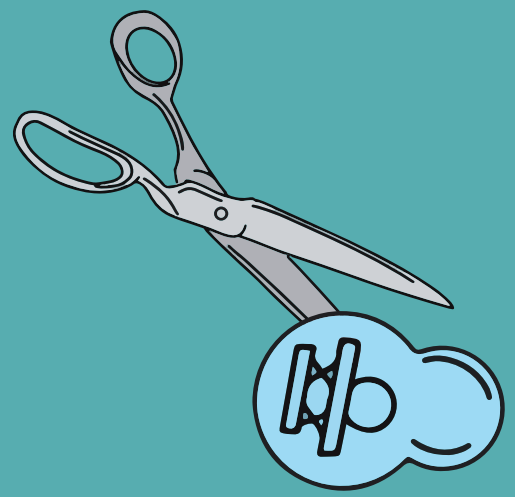


Figure S6.1. Transformation results in *Saccharomyces cerevisiae* using different delivery strategies. **(A)** Transformation results in *Saccharomyces cerevisiae* using pDonor delivery strategy 1. Three linear fragments were used for *in vivo* plasmid assembly into a single plasmid containing the *mRuby2* expression cassette, the gRNA expression cassette and the pNR1120 backbone. After transformation, cells were recovered at 20, 30, 34 or 37°C overnight and plated on selective plates. Colonies were screened for fluorescence under the TxRed fluorescence filter. **(B)** Three linear fragments were used for *in vivo* plasmid assembly into a single plasmid containing the gRNA expression cassette and the pNR1120 backbone. The cargo sequence expressing the *mRuby2* expression cassette between ShCAST right and left ends was introduced as a linear fragment without homology sequences. After transformation, cells were recovered at 20, 30, 34 or 37°C overnight and plated on selective plates. Colonies were screened for fluorescence under the TxRed fluorescence filter. **(C)** Control reactions used to check for random integration or random plasmid assembly. Controls were performed at 20°C.

Additional supplementary materials, tables and figures of this chapter can be accessed via

https://figshare.com/s/a08a21101f28ee9d4a7b



CHAPTER 7

Adaptation of SIBR-Cas for its use in yeast

Belén Adiego-Pérez*[1]
Constantinos Patinios*[1]
Sjoerd Creutzburg[1]
Adini Q Arifah[1]
Sarah D'Adamo[2]
Raymond Staals[1#]
John van der Oost[1#]

[1] Laboratory of Microbiology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands

[2] Bioprocess Engineering & AlgaePARC, Wageningen University and Research, P.O. Box 16, 6700 AA Wageningen, The Netherlands

* both authors contributed equally

[#] correspondence: raymond.staals@wur.nl, john.vanderoost@wur.nl

7.1 Abstract

Non-model yeasts bring added value to industrial processes because of their ability to grow at a broad range of conditions using various substrates. Genomic modification of these organisms is often required to implement novel production pathways. However, integration of heterologous DNA at target locations is limited by low homologous recombination efficiencies. The implementation of CRISPR-Cas tools has substantially improved the efficiency of the homology-directed repair (HDR) pathway. Yet, because non-homologous end joining is often the predominant repair pathway in these yeasts, screening for the desired HDR-dependent modifications is a laborious process. In this work, we accomplished the first steps towards the implementation of SIBR-Cas in non-model yeasts by establishing its use in the model yeast *S. cerevisiae*. SIBR-Cas is a CRISPR-based tool that we recently developed to increase homology-directed repair efficiencies in bacteria by delaying cleavage by Cas12a. The temporal control of the activity of the Cas protein allows for more recombination events to take place before counter-selection is executed. We optimized SIBR-Cas for its application in eukaryotic cells and showed induced activity in the yeast *S. cerevisiae*. We discuss further steps for transferring the acquired knowledge to non-model yeasts. Eventually, implementation of SIBR-Cas in these organisms may increase homologous recombination efficiencies and as such contribute to speeding up strain engineering.

7.2 Introduction

Yeast cell factories are being used to produce a wide variety of chemicals (*413*). The yeast *Saccharomyces cerevisiae* is conveniently chosen to produce organic acids, proteins, isoprenoids or alcohols (*414*). Robustness while culturing, availability of genome sequences and of genome editing tools make this yeast an ideal candidate as a chassis for production of industrially relevant biotechnological products. However, in the recent years, interest for the popularly called “non-conventional” yeasts has been increasing (*413*). Their ability to grow on a broader range of substrates under different conditions makes them attractive cell factory alternatives.

One example of an increasingly used yeast is the Crabtree negative and thermotolerant *Kluyveromyces marxianus* (*415*). This yeast has been used to produce flavors, bioethanol, biomass or native enzymes at an industrial level (*416*). A characteristic feature of this yeast is its ability to grow on lactose, which makes it an interesting organism to use on the dairy industry waste streams (*415*, *416*). In order to explore further uses of this yeast and exploit its metabolism, several efforts have been made towards the facilitation of genome modifications in this generally regarded as safe (GRAS) organism (*372*, *417*). Transformation protocols adapted from *S. cerevisiae* procedures allow for the uptake of linear DNA and incorporation into the chromosomal DNA. Integration of single DNA fragments into chromosomal targeted positions requires the activity of the homology-directed repair (HDR) pathway. Yet, the HDR activity is relatively low compared to the highly efficient non-homologous end joining (NHEJ) pathway (*42*). Wild type strains integrate linear DNA randomly via the dominant NHEJ repair mechanism. Alternatively, linear DNA fragments can be incorporated into targeted locations via the less frequent HDR pathway. For this to occur, the DNA fragments need to be flanked by homologous DNA sequences and include a marker for their selection. As opposite to *S. cerevisiae*, where HDR efficiencies are close to 100% (*418*), targeted integration efficiencies in *K. marxianus* vary from 0-14% (*371*, *419*). Therefore, construction of mutants with target modifications require the screening of a large number of individuals until the correct mutant is found (*417*).

Several strategies have been adopted which aim to increase HDR over NHEJ in non-model yeasts. Among the most popular ones, disruption of the NHEJ pathway by deletion of the *KU80* gene has resulted in increased HDR efficiencies (*41*, *119*, *371*). However, utilization of these deletion strains is not widely accepted throughout the scientific community because of the instability of the mutated strains against environmental stress (*420*). Recently, several CRISPR-Cas toolkits have been developed which streamline genome editing in *K. marxianus* by means of introducing targeted DNA double strand breaks (DSBs)(*41*, *120*, *421*). The use of CRISPR-Cas9 in WT *K. marxianus* strains allows to incorporate specific DNA sequences into target sites by HDR. Following this approach, no selection markers are needed and higher target editing efficiencies are achieved compared to traditional linear fragment transformation and marker-based selection (*41*, *120*, *372*, *421*). Yet, only 24-28% of the obtained colonies incorporate the DNA at the target location by HDR (*41*, *120*). In the event of a double strand break, NHEJ is favored over HDR, which limits HDR efficiencies when using CRISPR-Cas

tools. The action of this mechanism hampers the introduction of specific DNA sequences in a targeted manner and requires the investigation of alternative methods.

Multiple efforts have also been conducted to increase specificity and homologous recombination (HR) efficiencies of CRISPR-Cas tools via the regulation of the expression and activity of its modules. Most of the following strategies allow for spatiotemporal control of the CRISPR-Cas module. Some of the efforts focus on the control of the Cas protein expression, and others on the control of the guide RNA (gRNA). The ones controlling the Cas nuclease include the use of small molecules to activate Cas9 at a protein level (*422*), the control of protein localization by fusion of the Cas protein to an estrogen receptor (*423*), the use of split Cas proteins delivered via different recombinant adeno-associated viruses (rAAVs) (*424*), the use of transcription regulators (*84*), the fusion of degradation tags to Cas proteins (*94*), or the use of split Cas proteins which fusion can be induced with light (*425*) or by addition of some chemical ligand (*426*). Among the strategies focused on the control of the gRNA, we can distinguish photoactivation (*427*, *428*), photocaging (*429*), ligand activation (*430*) or the use of inhibitory aptamers interacting with the gRNA (*431*).

A Self-splicing Intron-Based Riboswitch (SIBR) associated to *Fn*Cas12a (SIBR-Cas) has recently been developed for its application in prokaryotic organisms. The system is based on the restoration of the *Fn*Cas12a reading frame by a theophylline-inducible Self-splicing Intron-Based Riboswitch (SIBR) (Figure 7.1). Therefore, induction of the system takes place at the mRNA level. The self-splicing intron is placed downstream of the start codon of *Fn*Cas12a, to avoid alterations in the coding sequence (CDS) by introduction of the exonic flanking regions. These exonic flanking regions are required for splicing of the intron but can be modified to drive different splicing efficiencies. Placement of the intron at this position allows the transferability of this technology to virtually any other gene of interest (GOI) in bacteria. The sequence of the intron is based on the bacteriophage T4 Group I *td* self-splicing intron (*432*, *433*). Modification of the intron by addition of a specific theophylline-binding sequence (RNA aptamer) to the P6 stem loop allowed to induce the splicing of the intron. Moreover, theophylline has been used to induce the splicing both *in vivo* and *in vitro* (*432*). As other naturally occurring introns, the T4 Group I *td* intron contains in-frame stop codons when retained in the pre-mRNA. Therefore, when no splicing occurs (in the absence of theophylline), a short peptide is formed due to the ribosomal encounter with premature termination codons in the pre-mRNA transcript. Only upon induction of the splicing system by theophylline, the intron contained in the pre-mRNA is spliced and a mature mRNA is formed. In the case of SIBR-Cas, this mRNA is translated into *Fn*Cas12a, which cleaves its target DNA in combination with the corresponding gRNA. The SIBR-Cas technology allows to delay the activity of the Cas protein by tight induction of the system. An overall higher editing efficiency is then achieved in bacteria by postponing Cas12a cleavage, achieving editing efficiencies of 38 ± 6.41 % in *Escherichia coli* MG1655, 70% ± 9.85% in *Pseudomonas putida* KT2440 and 100% ± 0% in *Flavobacterium* IR1 (*369*). Apparently, by delaying the induction of intron self-splicing, more events of homologous recombination can take place before Cas12a cleavage and counter-selection occurs (*369*).

The same principle could be used in non-model yeasts to promote targeted chromosomal integration by homologous recombination.

In this work, we describe the first steps towards the implementation of SIBR-Cas in yeast to increase the efficiencies of HDR over NHEJ. For this purpose, we adjusted the previously developed SIBR-Cas technology for engineering in bacteria, to allow for using it in the model yeast *S. cerevisiae*. The successfully established SIBR-Cas in *S. cerevisiae* paves the way for its application in other yeasts, and potentially in other eukaryotes.

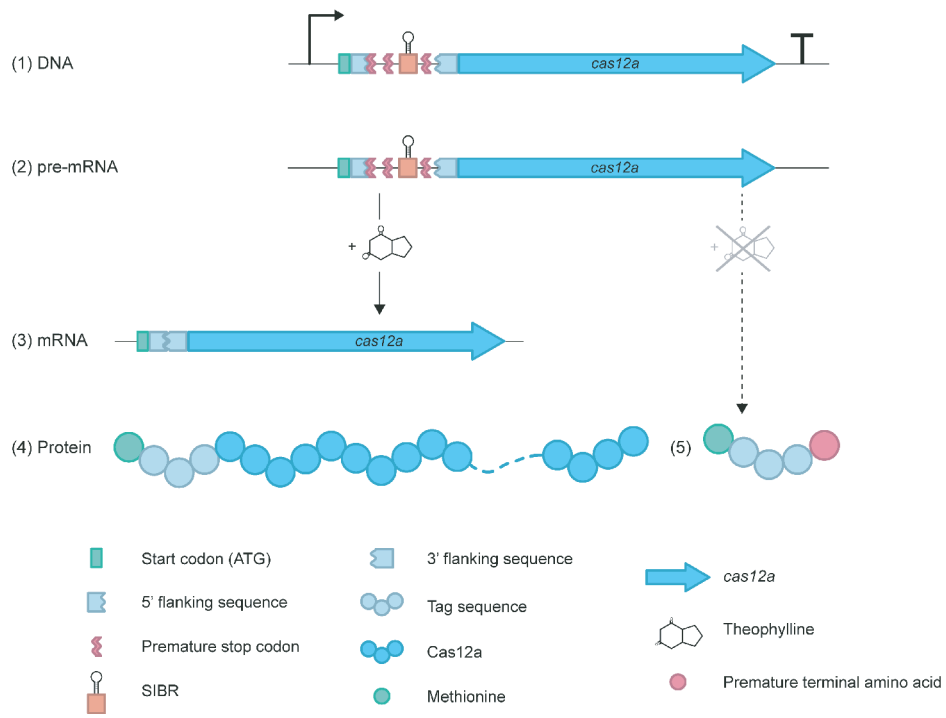


Figure 7.1. SIBR-Cas mechanism in bacteria. (1) The SIBR sequence is located after the start codon of the Cas12a open reading frame (ORF). The intron contains a modified version of the bacteriophage T4 Group I *td* self-splicing intron with a synthetic, theophylline-responsive aptamer. The intron is flanked by exonic flanking sequences (5' flanking sequence and 3' flanking sequence). Four different variants of the exonic flanks were previously developed for the SIBR-Cas system which, upon splicing, all leave different tags of 4-amino acids attached to the Cas12a protein. **(2)** Transcription leads to the formation of a pre-mRNA containing the SIBR sequence. In the presence of theophylline (left), **(3)** the SIBR is spliced, and the complete Cas12a mRNA transcript is formed and **(4)** translated. **(5)** In absence of theophylline, premature stop codons included in the SIBR sequence abort translation at an initial stage.

7.3 Results

7.3.1 High concentration of theophylline does not inhibit growth in *S. cerevisiae*

To investigate the functionality of SIBR in eukaryotic systems, we expressed SIBR-Cas tool (previously developed for bacteria; (*369*)) in the model yeast *S. cerevisiae*. By showing the functionality of SIBR-Cas in *S. cerevisiae*, we expect that the transferability of the tool to other yeasts should also be feasible.

SIBR-Cas is based on the use of theophylline as an inducer. Therefore, theophylline should be able to penetrate the *S. cerevisiae* cells and subsequently interact with the theophylline aptamer of the T4 *td* intron to cause its splicing. The use of theophylline has been previously reported as an inducer molecule for the activity of ribozymes and riboswitches in *S. cerevisiae* (*434*). The theophylline concentration used for induction in *S. cerevisiae* varied between different studies, being the 2 – 5 mM range the most common concentration. Nevertheless, we sought to define the limits of *S. cerevisiae* growth in presence of theophylline and for this reason, we conducted *S. cerevisiae* growth assays at different concentrations of theophylline (0-20 mM) in complex medium. In contrast to earlier reports (*435*), no significant growth defect was observed when cultivating *S. cerevisiae* in increasing concentrations of theophylline (0 - 20 mM) (Figure S7.1).

7.3.2 Direct translation of SIBR-Cas in *S. cerevisiae* does not lead to control over the activity of Cas12a

As an initial experimental set-up for SIBR-Cas in *S. cerevisiae*, we used the SIBR-Cas design that was previously used successfully in different wild-type bacteria. In more detail, we placed the 4 different SIBR variants (Int1 to Int4) downstream of the ATG start codon of the *Fncas12a* gene, which resulted in plasmids PL-286 to PL-289. Each intron variant presents a different splicing efficiency upon induction, with Int1 having the lowest and Int4 the highest (*369*, *436*).

To assess induction of SIBR-Cas and *Fn*Cas12a activity in *S. cerevisiae*, we used a 2-plasmid system based on previous research with *Fn*Cas12a in the model yeast (*111*). One plasmid was used to constitutively express either a WT *Fn*Cas12a, or the 4 different SIBR-*Fn*Cas12a variants. A second plasmid expressed either a guide targeting the *ADE2* reporter gene, or a non-targeting guide used as control. Because *S. cerevisiae* generally has very low NHEJ activity, functional expression of *Fn*Cas12a and its targeting guide should be lethal in the absence of a repair template, resulting in a major reduction in the number of colonies compared to a non-targeting guide (*84*).

We obtained strains YSTB391, YSTB392 and YSTB394 which harbored PL-286 (*Fn*Cas12a with SIBR Int1), PL-287 (*Fn*Cas12a with SIBR Int2) and PL-289 (*Fn*Cas12a with SIBR Int4), respectively. Strain YSTB013 was obtained which harbored plasmid pUDE731 (for constitutive expression of *Fn*Cas12a). Strains YSTB391, YSTB392, YSTB394 and YSTB013 were transformed with PL-074 (expressing an *ADE2* targeting spacer, used as targeting (T) plasmid) or PL-207 (expressing a non-targeting (NT) guide), independently. Strain CEN.PK113-5D was co-transformed with PL-288 (*Fn*Cas12a with SIBR Int3) and PL-074 (T) to confirm that a similar assay could be set-up by co-transformation of plasmids, which would allow us to simplify the approach in follow-up experiments. We plated all transformants on minimal medium with theophylline (5 mM). As expected, the combination of pUDE731 with the non-targeting PL-207 did not show colony reduction whereas the combination of pUDE731 with the targeting plasmid PL-074 yielded no colonies on plate. The combination of the SIBR-Cas variants (PL-286, PL-287 and PL-289) with the non-targeting PL-207 did not result in colony reduction. Surprisingly, when either of the four SIBR-Cas variants (Int1-Int4) was combined with the targeting PL-074, no reduction in the number of colonies could be observed (Figure S7.2).

In order to determine the cause hindering inducible splicing of the intron, we also tested the system with increasing concentrations of Mg^{+2} (Figure S7.3) which alleviates the self-splicing inhibition of certain components *in vitro* (*432*). However, when comparing results of transformations with a targeting guide to results of transformations with a non-targeting guide, no reduction in the number of colonies was observed upon induction with theophylline (5 mM). Controls with constitutively expressed *Fn*Cas12a did show a clear decrease in the number of colonies (Figure S7.3). Apparently, the SIBR-Cas design that does work well in different bacteria, did not lead to functional expression of Cas12a in yeast.

In order to pinpoint the cause hindering the use of SIBR-Cas in yeast, we set out to systematically analyze the different elements of the SIBR system. Starting with theophylline, we could rule out the inability of *S. cerevisiae* to take up theophylline since previously, theophylline has been reported to act successfully as an inducer for ribozymes and riboswitches in the model yeast (*434*). We also looked into different medium parameters (Figure S7.3) (*437*, *438*) and components such as antibiotics (*437*, *439–443*) or co-factors (*444–446*) that could be interfering with group I intron self-splicing. We used minimal medium for our assays and an antibiotic which was reported not to interfere with intron splicing *in vitro* (*440*). To eliminate the possibility that the T4 *td* intron could not splice in *S. cerevisiae*, we omitted the theophylline aptamer from the self-splicing intron. We placed the intron after the ATG start codon (PL-363) and after codon 67 (PL-364) of *Fn*Cas12a. Position 67 was chosen as the closest position to the ATG start codon in *Fncas12a* CDS which allowed for the integration of the intron respecting the exon flanking sequence requirements. The use of the T4 *td* self-splicing intron in *S. cerevisiae* had been previously reported (*447*). Therefore, we expected these two approaches to yield active *Fn*Cas12a proteins, and a reduction of transformants in the presence of a targeting guide (and in the absence of a repair fragment). Otherwise, presence of inhibitors in the medium should prevent self-splicing of the intron. We co-transformed the newly constructed SIBR plasmids (PL-363 and PL-364) with targeting (PL-074) and non-targeting (PL-207) guides expressing vectors (Figure 7.2). A reduction in the number of colonies was observed when comparing the outcome of both transformations with a T and a NT guide, indicating splicing of the T4 *td* intron and translation of a functional *Fn*Cas12a protein (Figure 7.2).

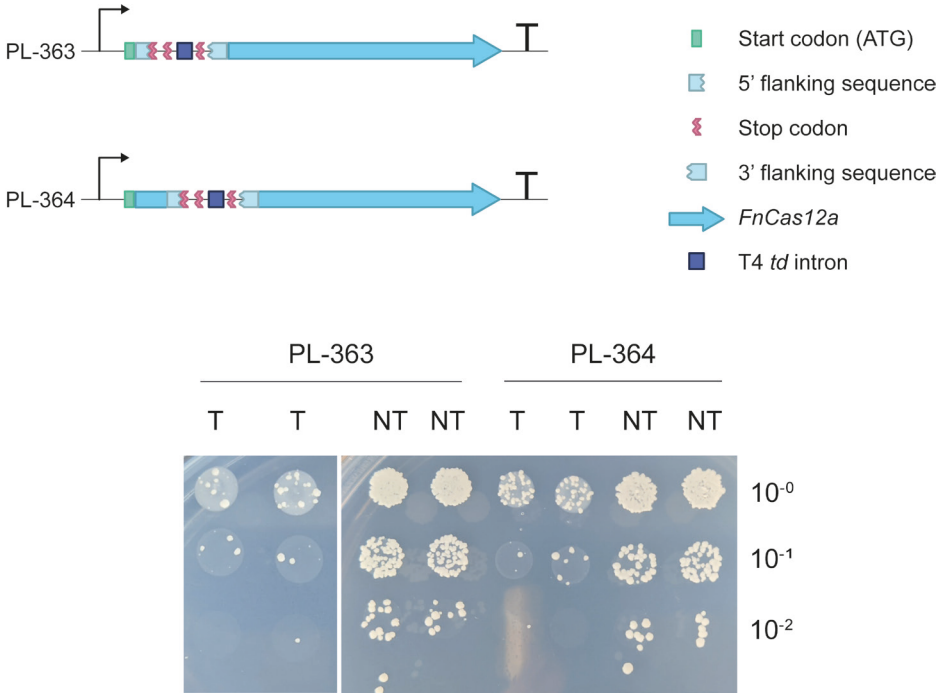


Figure 7.2. Bacteriophage T4 *td* intron test in *S. cerevisiae*. Placing of T4 *td* intron after the ATG start codon (PL-363) and at position 67 (PL-364) of *Fn*Cas12a CDS. CEN.PK113-5D was co-transformed with PL-363 and a targeting guide encoding plasmid show a reduction in the number of colonies compared to the transformation with the NT guide encoding plasmid. The same can be observed for the PL-364 plasmid, indicating that self-splicing of the T4 *td* intron is taking place and an active *Fn*Cas12 is able to cleave yeast DNA. Two biological replicates are shown for each transformation.

The main difference between the splicing requirements between the T4 *td* intron and the SIBR is the requirement of theophylline for the latter to induce self-splicing. After assumption of theophylline uptake and confirmation of T4 *td* intron splicing in the *Fn*Cas12a CDS, we hypothesized that the presence of premature stop codons in the SIBR-Cas intron triggers degradation of the pre-mRNA by non-sense mediated mRNA decay pathway before theophylline can induce its splicing (*448*, *449*). Such mechanism would be competing with theophylline for transcripts and avoiding induced splicing.

7.3.3 Adaptation of SIBR-Cas for eukaryotic use

Non-sense mediated decay (NMD) is a “quality control” regulatory pathway that degrades mRNA transcripts reaching the cytoplasm which contain premature stop codons (also referred to as premature termination codons (PTCs)), among other targets (*448–450*). This mechanism is present in eukaryotes and it is a translation-coupled mechanism that appears to vary substantially across different eukaryotic linages (*451*, *452*). Our hypothesis is that the placement of the SIBR sequence directly after the ATG start codon, as initially tested in

S. cerevisiae, could be promoting the action of the surveillance mechanism prior to induced splicing, resulting in decay of the transcript (NMD).

To prevent the possibility of NMD acting before the theophylline-induced splicing of the SIBR occurs, we re-designed the SIBR constructs and adapted them for their expression in eukaryotic cells. Although the triggers of NMD are not fully understood in , placement of PTCs close to the 3’ end of the transcript has been shown to reduce the activity of the NMD pathway (*453*). In this way, the adjusted design of SIBR-Cas may allow for its application in the yeast *S. cerevisiae* as eukaryotic SIBR-Cas in order to by-pass NMD (Figure 7.3A). Using an *in house* script (*454*) to respect exon flanking sequence requirements (*369*), we screened the *Fncas12a* CDS for a region which could be modified with synonymous codons respecting such requirements. We found position 859 in *Fncas12a* CDS to be suitable for placement of the SIBR. The 5’ exonic flanking sequence 5’-AAAGAGTCGCT-3’ and the 3’ exonic flanking sequence 5’-CTT-3’ do respect the amino acidic sequence of *Fn*Cas12 upon excision of the intron. At the same time, position 859 allowed to place the intron before the RuvC I, II and III domains. The RuvC domain is responsible for dsDNA cleavage by Cas12a (*126*) and by placing the intron at that position we ensured that translation of the RuvC domain would not occur in the absence of theophylline. To test NMD by-pass, we built two constructs: one containing the WT T4 *td* intron (without the theophylline aptamer) at position 859 and another one containing the SIBR (with the theophylline aptamer) also at position 859. We used pUDE731 as a control for the constitutive expression of *Fn*Cas12a.

Placement of the T4 *td* intron at position 859 allowed for intron self-splicing and consequently, for expression of an active *Fn*Cas12a. Co-expression of a targeting guide allowed cleavage from the Cas protein (Figure 7.3B), indicating that intron placement was not problematic at the tested position. Addition of the theophylline aptamer to the intron allowed for inducible control of splicing. Therefore, *Fn*Cas12a activity could be induced through addition of theophylline (Figure 7.3B). We tested induction of the system using theophylline concentrations from 5 - 20 mM. Unlike what we observed when testing the unmodified version of SIBR-Cas (i.e., SIBR inserted directly after the ATG start codon), a slight reduction in the number of colonies obtained under induction conditions could already be observed on plates containing 5 mM theophylline. Induction levels comparable to constitutive expression of *Fn*Cas12a were achieved at 20 mM theophylline with the eukaryotic-modified SIBR-Cas (eSIBR-Cas). Altogether, these results do demonstrate that SIBR-Cas could be adapted for its use in *S. cerevisiae*, most likely also providing possibilities for its use in other yeasts and eukaryotic systems.

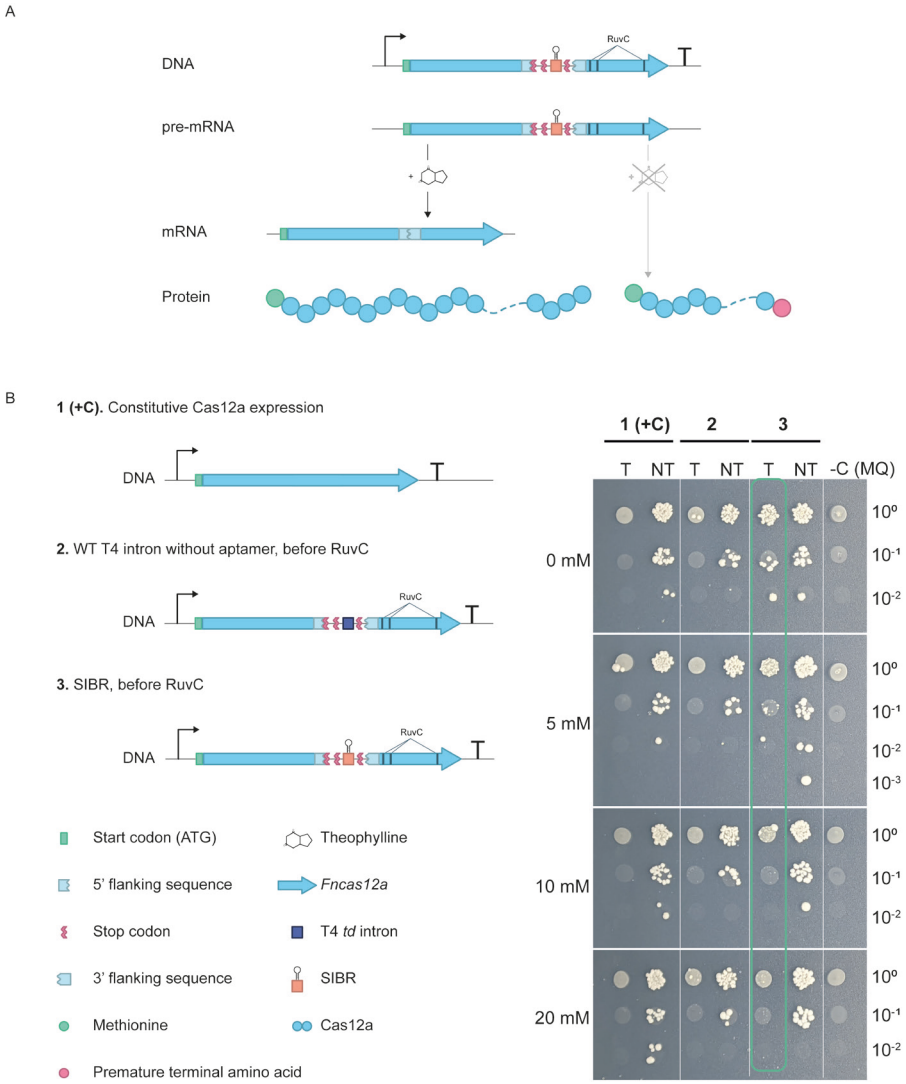


Figure 7.3. Adaptation of SIBR-Cas to *S. cerevisiae*. (A) The SIBR sequence was placed at position 859 of *Fn*Cas12a, allowing production of *Fn*Cas12a upon induction with theophylline. **(B)** Representative output of the test of SIBR-Cas in *S. cerevisiae*. Constitutive expression of *Fn*Cas12a in combination with a targeting guide (T) causes a decrease in the number of colonies obtained after transformation. Insertion of the T4 *td* intron at position 859 of *Fn*Cas12a allows for splicing of the self-splicing intron and cleavage of *Fn*Cas12a in presence of a targeting guide (T). Co-expression of *Fn*Cas12a with the SIBR sequence at position 859 and a targeting guide allows for cleavage only in presence of increasing concentrations of theophylline, allowing for low cleavage activities at 5 mM and high cleavage activities comparable to the control at 20 mM theophylline.

7.4 Discussion

In this study, we performed the first steps towards the application of SIBR-Cas in eukaryotic organisms. In bacteria, this tight control delays CRISPR-Cas counter-selection activity (i.e., lethal cleavage of wild type sequence), allowing more time for recombination events to occur. Following a similar rationale, we aimed to increase homologous recombination efficiencies in eukaryotic systems, by granting more time for the editing event to occur before CRISPR-Cas cleaves the DNA. We adopted the original design of SIBR-Cas, in which 4 different SIBR variants were positioned downstream of the ATG start codon, for its application in the model yeast *S. cerevisiae*. However, initially the implementation of SIBR-Cas constructs was not successful. Potentially, this could be the result of nonsense-mediated mRNA decay (NMD) (*448*, *449*, *452*), a eukaryotic surveillance mechanism that (amongst others) targets premature, intron-containing transcripts reaching the cytoplasm (*449*, *453*, *455*). This control mechanism might target the intron-containing transcript before theophylline-induced self-splicing occurred. Hence, we modified the original design of SIBR-Cas by placing the intron closer to the 3' end of the mRNA transcript of *Fn*Cas12. Indeed, this allowed us to successfully induce *Fn*Cas12a cleavage activity by theophylline addition, thereby establishing a eukaryotic-modified SIBR-Cas (eSIBR-Cas) system. Although NMD could have been responsible for disturbing the functionality of the original SIBR-Cas design in yeast, we cannot rule out other explanations at this stage, since a confirmation of the NMD activity over the SIBR transcripts is still required.

To date, we developed eSIBR-Cas for the control of Cas12a. Apart from Cas12a, the SIBR approach could ideally be transferred to Cas9 which has been more broadly utilized among non-model yeasts (*114*, *194*). Moreover, similarly to what it is proposed for bacteria (*369*), the SIBR could eventually be used to regulate expression of any gene when placed towards the 3'- end of transcripts, as long as the sequence encoded upstream of the PTCs encodes for a non-functional protein. Finally, the SIBR could also be used to allow the translation of C-terminal degradation tags upon splicing, which would allow for induced protein degradation when exposing to theophylline.

Other technologies have been developed in *S. cerevisiae* which use ribozymes or riboswitches induced by the theophylline RNA aptamer. Most studies use a maximum theophylline concentration of 10 mM. In our study, increasing theophylline concentrations from 5 – 20 mM correlated with an increased activity of *Fn*Cas12a in *S. cerevisiae*, which is likely related to increased splicing of SIBR. In the absence of a repair template, this resulted in increased cell death. However, 20 mM theophylline was required to achieve similar activity levels as the ones achieved with the constitutive expression of *Fn*Cas12a. The used concentrations for induction of SIBR-Cas in bacteria were lower, probably reflecting differences in theophylline uptake and sensitivity limits (*369*).

Certainly, implementation of eSIBR-Cas in *S. cerevisiae* opens the door for implementation of the tool in non-model yeasts. Ideally, eSIBR-Cas could be used as a tool in a broad range of organisms without any further modifications. Group I introns are present in multiple organisms,

including yeasts (*456*). In these organisms, some of these introns are autocatalytic and do not require the assistance of other proteins for their splicing. This represents an advantage compared to other inducible systems such as the use of inducible promoters, characterization of which is often limited in non-model organisms and transferability between species is not always possible (*457*, *458*).

Unfortunately, due to time limitations, transferring eSIBR-Cas into *K. marxianus* was not possible. Future studies should focus on the adoption of this tool for this organism, among other yeasts. Our results suggest that implementation of the tool in different organisms may require further tuning of the splicing conditions. Temperature, pH of the medium or the presence of T4 *td* intron inhibitors in the medium (e.g., some antibiotics) might limit the application of eSIBR-Cas in some organisms. Theophylline toxicity and uptake need to be investigated before application of the tool in a new organism. Although theophylline does not severely affect the growth of several yeasts (*459–461*), a working concentration range needs to be established for each organism beforehand. Toxicity of theophylline has only been reported for the *S. cerevisiae strain W303* (*MATa, his3-11, 15 trp1-1 leu2-3 ura3-1 ade2-1*) (*435*). However, adenine auxotrophy for this yeast strain might explain the observed theophylline toxicity. Uptake of adenine could be affected by the presence of theophylline in the medium, as it has been described for other yeasts like *S. pombe* for other methylxanthines such as the homolog caffeine (*462*, *463*). In a similar way, in other yeasts and higher eukaryotes, theophylline is described as an adenosine receptor antagonist (*459*, *460*), blocking the uptake of purines. These considerations should be taken into account at the time to test this tool in other organisms.

Overall, we showed that eSIBR-Cas is a promising tool for genome editing. The induction of *Fn*Cas12a activity at the post-transcriptional level in *S. cerevisiae* is a milestone in the transfer of this tool to non-model yeasts. The universality of the Group I introns should help in implementing this tool in different organisms, although host-specific optimization may be required. Ideally, eSIBR-Cas will help to increase homologous recombination efficiencies over NHEJ in non-model yeasts. This will facilitate the engineering process of non-model yeasts and repurpose them into more flexible platforms for industrial uses.

7.5 Materials and methods

7.5.1 Strains and growth conditions

E. coli DH5α and DH10B were used for the cloning and propagation of plasmids. *E. coli* was routinely cultured at 37 °C and 200 rpm in Luria Bertani medium (LB) [10 g L^{-1} peptone (Oxoid), 5 g L^{-1} yeast extract (BD), 10 g L^{-1} NaCl (Acros)]. Ampicillin was added for plasmid maintenance (100 μg mL^{-1}).

Saccharomyces cerevisiae CEN.PK113-5D (MATa; ura3-52) was used for testing SIBR-Cas in baker's yeast. *S. cerevisiae* was grown in minimal medium (yeast nitrogen base with ammonium sulfate as nitrogen base). Geneticin (G418) was used as selective antibiotic when required. In these occasions, glutamate was used as nitrogen base to maintain the pH stable. Solid media were obtained by addition of 20 g L^{-1} agar (Oxoid).

7.5.2 Molecular biology techniques

PCR reactions for diagnosis purposes were performed using OneTaq® DNA Polymerase (NEB) following manufacturer's instructions. Q5® High-Fidelity 2X Master Mix (NEB) was used for high fidelity amplifications according to supplier's instructions. Oligonucleotides used for cloning and sequencing were ordered from IDT or Sigma Aldrich (those longer than 60 bp). When required, DNA fragments were purified from agarose gels using Zymoclean Gel DNA Recovery Kit (Zymogen). Plasmids were generally assembled by HiFi® assembly (NEB), unless indicated otherwise. Repair template was cloned into plasmids using previously incorporated SapI restriction sites. Bacterial plasmids were isolated using GeneJET Plasmid Miniprep Kit (Thermo Scientific). Gene insertions were confirmed by diagnostic PCR and Sanger sequencing (Macrogen, Germany).

7.5.3 Plasmid construction

The plasmids constructed in this study are listed in Table S7.1. Oligonucleotides used for cloning are listed at Table S7.2. *S. cerevisiae* expression plasmids were based on the pUDE731 backbone. Intron sequences tested in this study were incorporated by HiFi Assembly. Non-targeting spacer sequence used for transformation controls was introduced by PCR or digestion-ligation.

7.5.4 Theophylline toxicity assay

A 40 mM theophylline stock solution was prepared by dissolving 0.286 g of theophylline anhydrous (Sigma-Aldrich) in 40 mL dH_2O. Theophylline was dissolved by vortexing, solution was filtered through 0.22 µm filters and added fresh to media.

4 X YP (40 g/L yeast extract, 80 g/L peptone) medium was prepared and diluted accordingly to accommodate tested theophylline concentrations with 50% glucose (2% final concentration), the theophylline stock solution and dH_2O. Two *S. cerevisiae* CEN.PK113-5D individual colonies were grown overnight in 5 mL YPD at 30 °C and 200 rpm in 50 mL Greiner tubes. After overnight incubation, cultures were inoculated in technical duplicates in 96 well plates. Individual wells were filled to 200 µL with YPD medium and inoculated with the corresponding volume for a starting OD_{600} of 0.05. Measurements were performed on a Synergy Neo2 plate reader (Agilent) connected to Gen5 software. Double orbital continuous shaking (3mm orbital, 282 cpm) with slow orbital speed were used as settings. OD_{600} was monitored for 24 hours every 10 minutes. Growth curves were generated for each condition (growth under different theophylline concentrations) and compared to the control condition (0 mM theophylline).

7.5.5 SIBR-Cas test in *S. cerevisiae*

To test SIBR-Cas in *S. cerevisiae*, CEN.PK113-5D or derivate strains (see table) were transformed with 500 ng of plasmid by the LiAc/SS carrier DNA/PEG method (*357*). When 2 plasmids were co-transformed, 500 ng of each plasmid were added to the transformation mixture. Recovery was performed in YPD or minimal medium for 2 hours at 30 ºC. Geneticin (G418) was used for selection of plasmids. Plates were incubated for 3-5 days at 30 °C until colonies appeared. All transformations were performed, at least, in duplicate.

To test SIBR-Cas in *S. cerevisiae*, PL-286, PL-287 and PL-289 were transformed in *S. cerevisiae* CEN.PK113-5D in order to obtain YSTB391, YSTB392 and YSTB394, respectively (Table S7.3). Transformants were plated on SC (-ura). Obtained strains were transformed with PL-096 (targeting 4 genomic sites) and PL-207 (NT guide). CEN.PK113-5D was co-transformed with PL-288 and PL-096. YSTB013 was transformed with PL-096 and PL-207 (NT guide) as control for cleavage and absence of cleavage, respectively. Transformations were plated in dilution series in minimal medium with glutamate as nitrogen source and G418 as selection marker. Theophylline was added to the medium at concentration 5 mM. The same procedure was followed to test the effect of increasing Mg^{+2} concentrations in the medium by adding $MgCl_2$ in the medium at concentrations 4 mM, 8 mM and 12 mM. This time, only strain expressing the *Fn*Cas12a intron 4 variant (YSTB394) was used, since higher splicing efficiencies had been reported for that variant.

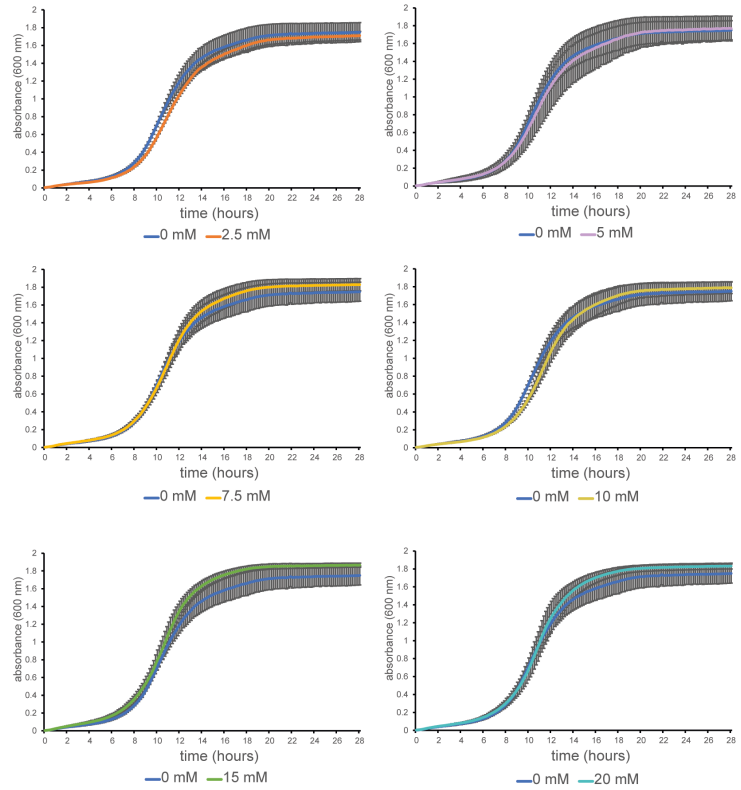
To test self-splicing of T4 *td* intron after the ATG start codon and at position 67, PL-363 and PL-364 were co-transformed with PL-074 (targeting *ADE2*) and PL-207 (Non-targeting) in *S. cerevisiae* CEN.PK113-5D. Transformations were plated on minimal medium with glutamate as nitrogen source and G418 as selection antibiotic in dilution series.

To test the introduction of the SIBR and the T4 *td* intron into the middle of the CDS of *Fn*Cas12a, PL-319 and PL-320 were co-transformed with PL-074 (targeting *ADE2*) and PL-207 (Non-targeting) in *S. cerevisiae* CEN.PK113-5D. *S. cerevisiae* CEN.PK113-5D was also co-transformed with pUDE731 and PL-074 or PL-207 as controls for cleavage and absence of cleavage. Theophylline was added to induction plates at a concentration of 5, 10 and 20 mM. Transformants were plated in dilution series on minimal medium with glutamate as nitrogen source and G418 as antibiotic selection agent.

To test the introduction of the SIBR into the middle of the CDS of *iCas9*, plasmids PL-339, PL-340, PL-341 and PL-342 were co-transformed with PL-074 (targeting *ADE2*) and PL-207 (NT). PL-338 was used as positive control for cleavage and absence of cleavage and co-transformed with PL-074 (targeting *ADE2*) and PL-207 (NT). Theophylline was added to induction plates at a concentration of 20 mM. Transformants were plated in dilution series on minimal medium with glutamate as nitrogen source and G418 as antibiotic selection agent.

7.6 Supplementary materials

A



B

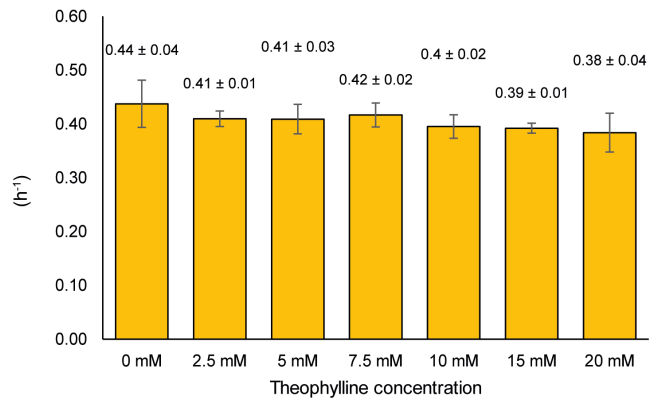


Figure S7.1. Theophylline toxicity on *S. cerevisiae* growing on complex YPD medium (0 mM – 20 mM). **(A)** Growth curves of *S. cerevisiae* CEN.PK113-5D in presence of theophylline (0 mM – 20 mM). Each graph represents the growth curve of one condition compared to the control condition (0 mM theophylline). **(B)** Estimated specific growth rates of each growth condition.

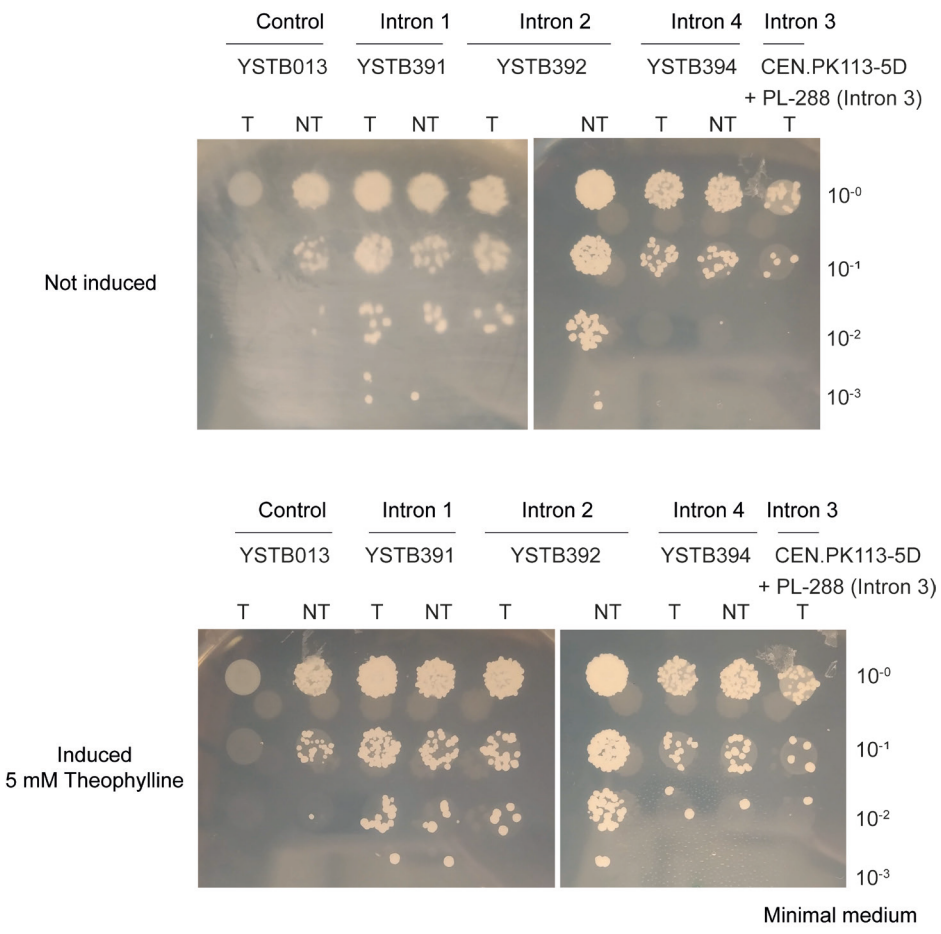


Figure S7.2. SIBR-Cas test in *S. cerevisiae* in minimal medium and concentrations of theophylline (5 mM). SIBR-Cas variants with intron 1 to intron 4 (strains YSTB391, YSTB392 and YSTB394) were tested with targeting (T) and non-targeting (NT) guides. Positive control was obtained with YSTB013 transformed with a targeting (T) and a non-targeting (NT) guide. A reduction in the number of colonies was only observed with the control strain and not with any of the SIBR-Cas strains.

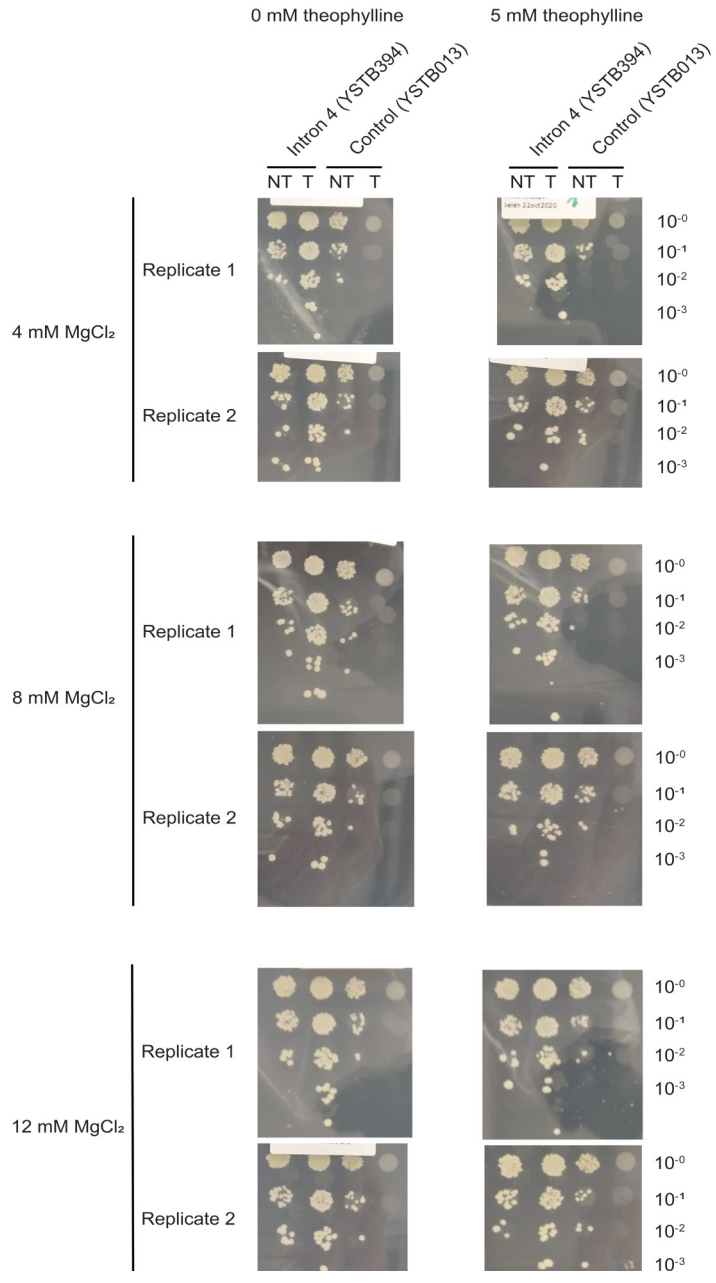


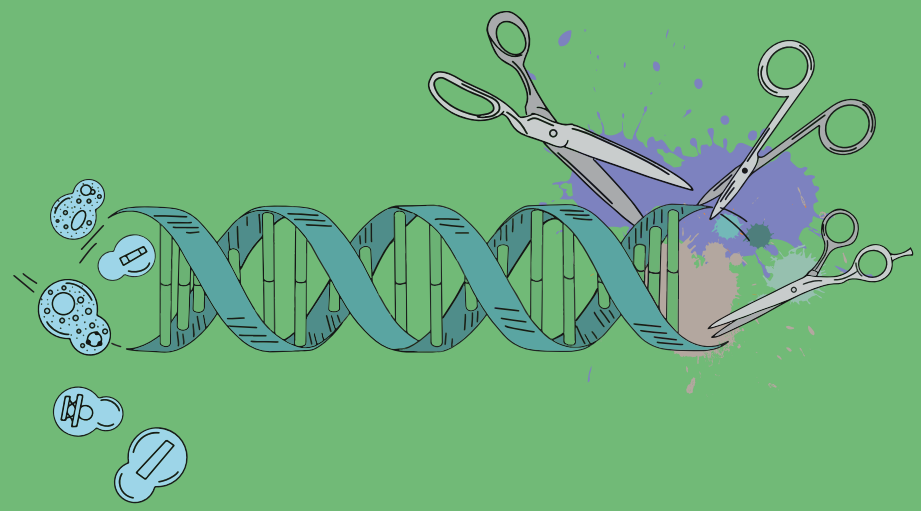
Figure S7.3. SIBR-Cas test in *S. cerevisiae* in minimal medium and increased concentrations of $MgCl_2$ with 5 mM theophylline. SIBR-Cas variant with intron 4 (strain YSTB394) was tested with targeting (T) and non-targeting (NT) guides. Positive control was obtained with YSTB013 transformed with a targeting (T) and a non-targeting (NT) guide. A reduction in the number of colonies was only observed with the control strain and not with any of the SIBR-Cas strains in any of the tested condition. Increased concentrations of $MgCl_2$ did not promote intron splicing.

Additional supplementary materials, tables and figures can be found here:

https://figshare.com/s/fb9aba5f1850122b0605

CHAPTER 8

Thesis summary and general discussion



THESIS SUMMARY

The use of microorganisms for biotechnological processes allows the natural and heterologous production of industrially relevant compounds. Genome editing tools are required to incorporate heterologous biosynthetic production pathways or to fine-tune already existing metabolic pathways. The application of CRISPR-Cas systems in prokaryotes and eukaryotes has revolutionized genome editing programs in the last years. Natural diversity of these CRISPR-Cas systems allows the development of novel tools with different characteristics, which can grant higher editing efficiencies, lower toxicity effects when heterologously expressed, integration of large DNA constructs or access to novel targets. Moreover, modification of characterized tools can facilitate genome editing of difficult-to-modify organisms, improving genome editing programs for the construction of industrially relevant strains. In this thesis, we explored the diversity of type V CRISPR-Cas effectors and performed the first steps towards their application in the yeast *Saccharomyces cerevisiae*.

In **Chapter 1**, we introduced the use of yeast in biotechnological applications throughout history and how these have benefited from the advances in the field of genetics and genome engineering. We introduced classical genome editing methods in *S. cerevisiae* and the revolution of CRISPR-Cas-based tools. Among all described systems, we focused on the type V for the further development of novel CRISPR tools for genome editing in *S. cerevisiae* and non-model yeasts.

In **Chapter 2**, we reviewed the multiplexing capabilities of CRISPR-Cas based tools and the different strategies adopted in the recent years for targeting of multiple loci, both in eukaryotic and prokaryotic microorganisms, using CRISPR-Cas tools. We assessed gRNA design strategies, donor DNA delivery strategies and the main bottlenecks of multiplex genome editing. We described the use of a type V effector, Cas12a, because of its capability of processing its pre-crRNA without the aid of additional proteins or RNA species.

In **Chapter 3** we focused on a novel candidate to type V effector, *Mm*Cas12m. We characterized this effector as a DNA binding protein not able to cleave DNA. We showed this protein is able to process its pre-crRNA and that it can be conveniently used for multiplex applications in *Escherichia coli*. We hypothesized the biological role of this protein in the native organism and showed that it can impair plasmid replication when directed to the origin of replication of an invading plasmid. Finally, we fused several base editor domains to the protein and repurposed *Mm*Cas12m as a small size base editor in *E. coli*. Characterization of the fusion products showed their activity on a dual base editing window in the protospacer.

Following with the promising results in *E. coli*, in **Chapter 4** we assessed whether the novel type V effector could be used in *S. cerevisiae* for genome silencing purposes. Initially, we showed nuclear localization of the protein and confirmed proper protein size. However, low-to-no silencing effect was observed when directing *Mm*Cas12m to the promoter region of a genome-integrated copy of *eGFP*. Therefore, in **Chapter 5**, we aimed to detect the activity of *Mm*Cas12m in yeast throughout a permanent effect assay. For this purpose, we designed several *Mm*Cas12m base editors targeting the *ADE2* reporter gene and aiming to convert specific cytosines into thymidines, for the introduction of stop codons that would knockout the target gene. In our assays, we detected targeted base editing activity albeit at low frequencies, not comparable with previously developed dCas9 or nCas9 base editors. The work presented in these two chapters show that *Mm*Cas12m is not ready to be used as genome editing or regulation tool in *S. cerevisiae*.

In **Chapter 6** we focused on another Type V CRISPR-Cas system, the type V-K. Type V-K presents Cas12k, another DNA binding effector not able to cleave DNA. In this case, Cas12k is associated to a Tn7-like transposon system with 3 other components (TnsB, TnsC and TniQ). We assessed the activity of this CRISPR-associated transposon system at different temperatures in *E. coli* and detected transposition at 25, 30 and 37 °C. Moreover, we assessed the effect of the addition of fusion peptides to the terminal ends of each of the proteins of the complex. We showed that 2 out of the 8 tested positions compromise RNA-guided DNA transposition in *E. coli*. We used this knowledge to add nuclear localization signals to each of the proteins of the complex and expressed them in *S. cerevisiae*. In this eukaryotic model organism, no RNA-guided DNA transposition was observed.

Finally, in **Chapter 7**, we repurposed SIBR-Cas for its use in *S. cerevisiae*. This tool, initially developed for its application in wild-type bacteria, is based on the control of the splicing of a group I intron placed at the 3'-terminus of a Cas12a transcript. Splicing is induced by the addition of theophylline, which allows translation of Cas12a. In bacteria, delay of the expression of Cas12a allows homologous recombination to take place before CRISPR-Cas counterselection. In order to adapt this tool for *S. cerevisiae,* we positioned the intron downstream in the mRNA transcript in order to avoid native control mechanisms. This new design allowed induction of splicing of the intron upon theophylline addition. Although this tool was only tested in the model yeast, we envision that time control over the expression of Cas12a should boost homologous recombination and improve genome editing programs in non-model yeasts.

GENERAL DISCUSSION

8.1 Impact of CRISPR-Cas tools on yeast research

The use of yeast for biotechnological applications has been present throughout the history of humankind (Chapter 1) (*5*, *16*). The more we know about yeast metabolism and genetics, the more successful we will be in modifying its activity to fit our interest (*37*, *464*). In this way, the well-known Baker's yeast is currently used as production organism for ethanol, pharmaceuticals, and enzymes. Moreover, in the last years, the interest in using yeast for industrial processes has been expanded to non-model yeasts, which may have typical features such as growth on complex substrates or at higher temperatures (*40*).

Through design, build, test and learn (DBTL) cycles, yeast strains can be improved to boost the production of relevant industrial products (*28*). Following the classification of a strain as genetically accessible, efficient genetic toolboxes are required for performing these iterative engineering cycles. In addition, knowledge on yeast metabolism, availability of sequenced genomes and access to metabolic models, is essential for designing strategies to modify the DNA in a targeted manner, to eventually obtain a yeast with the desired metabolism. In this way, production strains can be generated in which synthetic and/or heterologous production pathways are incorporated, or in which competing by-products are eliminated.

CRISPR-Cas tools have resulted in a revolution in the field of genome editing of many different organisms, and their application in yeast has exponentially increased over the last years. In the model Baker's yeast (*Saccharomyces cerevisiae*), these technologies not only increased homology directed repair efficiencies upon induction of double-strand DNA breaks (DSB), they also improved multiplex editing programs, by allowing the simultaneous target modification of multiple sites with high efficiencies. Moreover, CRISPR interference (CRISPRi) and activation (CRISPRa), as well as prime and base editor tools have been developed for fundamental research applications. Still, a limited number of Cas proteins are being used in yeast, with Cas9 being the most commonly employed one and Cas12a and Cas13 used in a smaller number of cases (Figure 8.1).

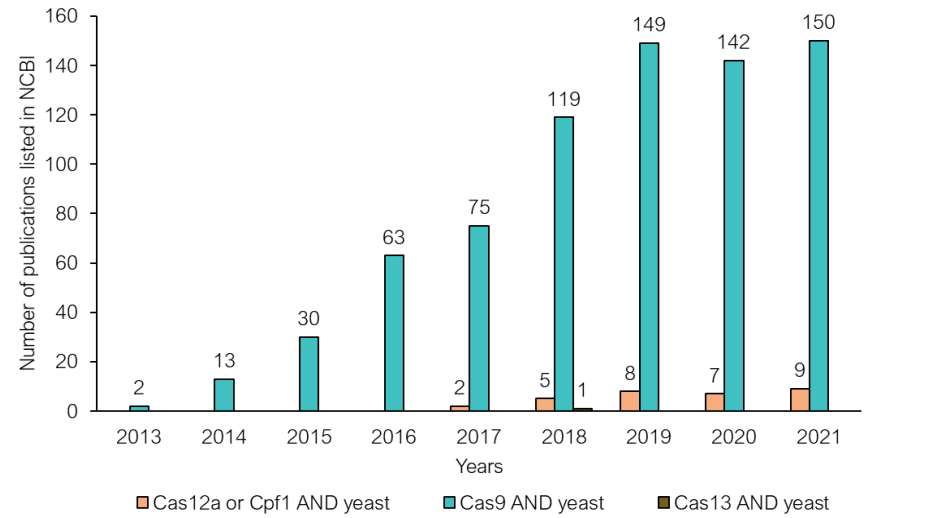


Figure 8.1. Application of CRISPR-Cas tools in yeast. Number of publications listed in NCBI with the combined search "Cas9 AND yeast" (blue), "Cas12a OR Cpf1 AND yeast" (orange) or "Cas13 AND yeast" (brown).

8.2 Bottlenecks in the field of genome editing in yeast

Since its first application in the yeast *S. cerevisiae* (*84*), Cas9 has been widely adopted for genome editing applications across both model and non-model yeasts (*83*, *193*, *349*). The availability of guide design tools for Cas9 and its proven activity in multiple organisms have prompted its use across both *S. cerevisiae* and other industrially relevant yeasts. To a more limited extent, Cas12a has also been applied for genome editing, exploiting its native capabilities for multiplexing (*110–112*). Advances in the application of CRISPR-Cas technologies in this model yeast are focused on several bottlenecks detailed in the next sections.

8.2.1 Multiplex editing with CRISPR-Cas systems

CRISPR-Cas technologies have been established and improved at different levels both in *S. cerevisiae* and in non-model yeasts. Since close-to-100% editing efficiencies can be obtained when using Cas9 or Cas12a in *S. cerevisiae*, recent advances have been focused on generating multiple targeted modifications simultaneously in this organism. One of the most impressive studies regarding multiplexing concerns the generation of a hexose transporter-free *S. cerevisiae* strain using Cas9 (*186*). In this work, throughout only three transformation rounds, 21 hexose transporter genes were deleted from the *S. cerevisiae* genome (*186*). High sequence homology among the target genes was used to target up to 3 genes with the same guide RNA (gRNA). Following this approach, up to 8 different *loci* could be targeted simultaneously with the expression of only 4 different gRNAs. There are several strategies to express multiple guides to drive multiplexed genome editing in yeast (Chapter 2).

However, since the publication of this chapter (*465*), a novel strategy has been developed based on the delivery of linear DNA fragments (*466*). Using T7 RNA polymerase-expressing yeast strains, multiple linear fragments with gRNA genes under the control of a T7-polymerase promoter can be transformed into yeast (encoding for the T7 RNA polymerase) and be transcribed into functional Cas9 guides. Although editing efficiencies are lower than obtained when using plasmid-based approaches, this strategy considerably accelerates editing programs by avoiding amplification and cloning steps. Still, the main challenges and opportunities (already detailed in chapter 2, (*465*)) remain: (i) guide design and predictability, (ii) donor DNA design and delivery, (iii) controllable expression of CRISPR elements, (iv) innovative plasmid assembly methods, (v) editing conditions, and (vi) discovery of novel and improved endonucleases.

The work presented in this thesis aimed to address some of these challenges, one of them regarding the control of the expression of CRISPR-Cas elements. To this extent, we implemented SIBR-Cas (*369*) in *S. cerevisiae* (Chapter 7). Originally, SIBR-Cas was developed as a prokaryotic genome editing tool and it employed a self-splicing intron located on the 5'-end of the Cas12a transcript (*369*). Generation of the intron-containing transcript prevents translation of the full Cas12a due the premature stop codons contained in the sequence of the intron. Upon addition of the inducer theophylline, the self-splicing intron is spliced from the transcript together with the stop codons it contains, allowing controllable expression of the Cas12a nuclease. This means that by delaying the counter selection activity of Cas12a, a longer time for homologous recombination is granted, which in turn yields a higher number of recombinants in wild type bacterial organisms. After adapting SIBR-Cas for its use in *S. cerevisiae*, we showed temporal control over the activity of Cas12a by the addition of theophylline. We envision that the time control over the activity of Cas12a provided by the SIBR system can be used to enhance multiplexed editing. Parallel to what was described in bacteria, we aim to provide a repair template in a circular form for its maintenance in the cell. Although DNA recombination from plasmid DNA occurs at low rates (*467*), we anticipate that by postponing Cas12a expression, a greater fraction of the population will recombine the provided repair templates before experiencing Cas12a counterselection activity.

Novel nucleases could be used in multiplexing genome editing programs. In this regard, we characterized a novel Cas12 variant, Cas12m (Chapter 3), and applied it to *S. cerevisiae* (Chapter 4 and 5). Characterization experiments in *E. coli* proved the ability of Cas12m to process its pre-crRNA and its multiplexing capabilities, by efficiently and simultaneously silencing the expression of *gfp* and *rfp* genes. Cas12m belongs to the wide type V, with multiple nuclease and collateral activities described among the characterized effectors. Among them, only Cas12a (*149*), Cas12c (*73*), Cas12i (*129*), Cas12j (*259*) and Cas12m have been described to display pre-crRNA processing activity, although some require the presence of additional RNA species. This RNase activity makes them interesting nucleases for multiplexed editing, since they do not require additional proteins to process their pre-crRNAs. However, as other recently characterized small size type V effectors, Cas12m displayed low or no activity in eukaryotic systems (Chapter 4). Therefore, its multiplexing capabilities were not evaluated in *S. cerevisiae*.

8.2.2 Beyond canonical use of CRISPR systems for the integration of DNA fragments

Targeted integration of DNA fragments in yeasts with limited homologous recombination efficiencies is a challenge. While induction of targeted DSBs increases editing efficiencies at targeted loci in these yeasts, NHEJ-repaired colonies are yet predominant among the edited populations. Integration of DNA fragments through DSBs-independent mechanisms could solve this issue since it would avoid recruitment of the DNA repair machinery. In this regard, we presented the Type V Tn7-like CRISPR-associated transposon system (Chapter 6). Expression of four proteins (TnsB, TnsC, TniQ and Cas12k) is sufficient for RNA-guided DNA transposition in an HDR-independent manner in bacteria (*258*). This system has been used for editing of *E. coli* and *Pseudomonas putida* among other bacterial species, but still, no activity has been reported in eukaryotic systems. We assessed its activity in *S. cerevisiae* and we did not detect RNA-guided DNA transposition (Chapter 6). The studied RNA-guided DNA transposition requires the sequential activity of multiple proteins from the transposition complex. In the context of a eukaryotic cell, Cas12k needs to stably interact with DNA. As seen for other type V effector proteins, such as Cas12f1, interaction of a nuclease with DNA in the eukaryotic context is not always strong (*132*). We postulated that inefficient DNA interaction could be one of the causes hampering the activity of the type V CRISPR-associated transposon system. DNA binding assays of Cas12k in eukaryotic systems might help elucidate whether low binding affinity of the protein is the bottleneck in the RNA-guided DNA transposition process. Next, TnsB needs to be active and cleave the transposon DNA from the donor DNA molecule. Moreover, TnsC should be able to polymerize around the yeast DNA and interaction of TnsB with TnsC should trigger ATP hydrolysis by TnsC and DNA decoupling (*405*). Assessment of each individual step inside the nucleus of *S. cerevisiae* would shed light on the feasibility of using this mechanism in yeast. By fusing a purification tag on each protein of the transposition complex, *S. cerevisiae* pull-down assays could be employed to determine which protein-protein interactions successfully take place in the yeast.

The use of non-native recombinases or transposases that recognize specific integration sequences has also been proposed as an alternative strategy to integrate heterologous DNA sequences without the generation of DSB. Fusion of serine-recombinases or transposases to deactivated Cas9 (dCas9) has been proven as a valid strategy to direct these enzymes to specific genomic sites and integrate DNA fragments in an HDR-independent manner (*468*, *469*). However, some of these approaches still require the previous characterization of naturally existing recognition sites or introduction of those in the target genome (*468*). Recently, a large set of serine recombinases has been evaluated for DNA integration in human cells (*470*). Efficiencies varying from 40-70% have been obtained without selection, although this technology is still limited to the previous incorporation of landing pads recognized by the recombinases (*470*). Undoubtedly, these systems are promising candidates for genome editing tools to integrate DNA fragments in yeasts with limited homologous recombination efficiencies, similar to the CRISPR-associated transposon systems.

8.2.3 Improvement of homologous recombination efficiencies

The most common strategy to enhance the activity of the HDR pathway over the NHEJ pathway is the use of knockout strains for the Ku70 or the Ku80 subunits (*43*, *373*). Such strains have been generated for industrially-relevant species of *Yarrowia lipolytica* and *Kluyveromyces marxianus*, among others (*43*, *353*, *373*). However, their use at industrial level is not ideal since they might present lower stability against environmental changes because of the deficiency in the DNA repair pathway. Therefore, several genome editing tools have been focused on boosting the activity of HDR over NHEJ.

In eukaryotic systems, homologous recombination activity is maximal in actively dividing cells (during S, G2 and M phases). Moreover, this activity is also promoted by inducing DSBs with, for instance, Cas effectors. Therefore, we proposed the use of SIBR-Cas as a way to induce the activity of Cas12a after allowing the population of cells to reach an actively dividing, and thus homologous recombination active, state. We hypothesize that this could allow for more recombination events to take place before the induction of the Cas activity, which in time will also enhance the recombination efficiency upon introduction of the DSB. Although such a system could also be implemented with inducible promoters, transferability of promoters between different yeast species is sometimes challenging. Instead, group I splicing introns are present in multiple domains of life (*471*, *472*) and are, therefore, are good candidates to be established successfully in multiple species. Although we did not manage to test our hypothesis because of time constraints, we showed that we can promote activity of Cas12a by inducing the splicing of the SIBR in *S. cerevisiae*. However, this system could virtually be used in any other yeast. A similar “easy-to-transfer-between-yeasts” system has been recently proposed, which uses linear fragments for expression of guides and Cas encoding genes and can simultaneously be used as DNA repair template (*94*). Moreover, this system is based on the degradation of Cas9 in the G1 phase and accumulation of the protein in the S, G2 and M phases, via its fusion to a tag (*364*). In this way, expression of Cas9 is limited to the cell cycle phases when the HDR pathway is active.

8.2.4 Use of CRISPR-Cas tools as silencing or single nucleotide editing tools

In the last years, multiple genome editing or regulation tools derived from CRISPR systems have been developed. Initially, dCas9 proteins were fused to repressor domains (Mxi1, KRAB) to mediate transcriptional repression in several yeasts (*104*, *323*). Most of the transcriptional repression tools are based on the use of dCas9, although recently dCas12a-Mxi1 has been showed as a powerful genome silencing tool in *S. cerevisiae*. This is due to the increased availability of Cas9 target design tools and the limited application of Cas12a in yeasts other than *S. cerevisiae*. In our work, we also tested and compared the silencing power of dCas12a-Mxi1 to dCas9-Mxi1, showing higher repression effects when using dCas9-Mxi1 (Chapter 4). Comparison of our results to previously reported data in literature is a difficult task, due to the lack of standardized setups, testing conditions or the lack of inclusion of several Cas proteins in the same study.

The use of CRISPR-Cas tools for genome regulation purposes is based on the deactivation of the catalytic residues in charge of DNA cleavage. In Cas12 variants, these catalytic residues are found in the RuvC domain. Therefore, Cas12 variants naturally harboring deactivated RuvC domains should also be able to mediate gene repression/activation *in vivo*. We showed that this is the case for *Mm*Cas12m in *E. coli* (Chapter 3). However, when we assessed the activity of *Mm*Cas12m in *S. cerevisiae* (Chapter 4), low-to-no repression activity by *Mm*Cas12m was observed in *S. cerevisiae*. Moreover, fusion of a repressor domain to *Mm*Cas12m did not improve the silencing effect. As discussed in the mentioned chapter, this could be due to unstable binding of the protein to the DNA in yeast, as suggested for other small Cas12 variants in human cells or plants (*130*). However, this is in contradiction with the surface plasmon resonance results obtained *in vitro* (Chapter 3), which showed that *Mm*Cas12m is able to stably bind to DNA. Following our hypothesis, and in order to improve the silencing activity of Cas12m in yeast, a similar laboratory evolution experiment to the one performed on Cas12f1 in human cells (*291*) could be performed. First, random mutagenesis could be performed on the *Mm*Cas12m sequence. The obtained library could then be transformed into an *eGFP*-expressing yeast and silencing levels could be determined through flow cytometer by directing *Mm*Cas12m to the *eGFP* promoter. Sorting of the strongest eGFP repressors could allow the identification of improved transcriptional repressors. Multiple rounds of this process should allow the improvement of *Mm*Cas12m as silencing tool in *S. cerevisiae* by modifying its sequence and, eventually, generating stronger DNA-binding *Mm*Cas12m variants.

Combination of novel Cas12 variants with DNA-modifying enzymes could eventually be used for base editing applications. This is the case for *Mm*Cas12m, that can be used as a base editor in *E. coli*. However, similarly to what we observed with its silencing activity, low base editing frequencies were observed at a population level with *Mm*Cas12m in *S. cerevisiae* (Chapter 5). Interestingly, in the cases where we observed targeted base editing, only the aimed nucleotide was modified. Most likely, this is due to a globally low activity of the *Mm*Cas12m base editor, which coincidentally minimizes the off-target activity of the fusion protein (*473*). At the same time, it can be related with the previously hypothesized unstable binding of *Mm*Cas12m to the DNA in yeast (Chapter 4). Finally, *Mm*Cas12m is a catalytically-deactivated nuclease and, to our knowledge, cannot be converted into a nickase by restoring the catalytic DED triad. This limits the use of this Cas12 variant in combination with base editors in eukaryotic organisms. For their use in eukaryotic systems, base editors should ideally nick the target strand to allow repair using the non-target and edited strand as template (*282*). Therefore, for *Mm*Cas12m to be implemented as an efficient base editing tool, modification of the protein sequence to turn it into an effectively DNA-binding nickase is required. Alternative, other Cas12m homologs with a predicted active RuvC domain could be used for this purpose.

8.2.5 Innovative CRISPR-Cas based tools

Besides DSB-inducing and transcription regulation tools, novel CRISPR tools such as prime editors, serine-recombinases or similar are not extensively used in yeast. High homologous recombination efficiencies in *S. cerevisiae* allow the incorporation of point mutations using DSB-generating CRISPR-Cas tools. Therefore, more complex approaches such as base editing (*106*, *107*, *356*) or prime editing (*109*) remain as proof-of-concept applications and only tested in *S. cerevisiae*. For instance, Cas9-derived base editors have only been implemented in *S. cerevisiae* (*106*, *107*) and *Yarrowia lipolytica* (*353*). To date, no Cas12a-derived base editors have been used, while their characterization has taken place for human applications in depth (*283*). Another example is the use of Cas13, which activity has only been shown in *Schizosaccharomyces pombe*, and not in *S. cerevisiae*.

Application of novel technologies such as twin prime editors (*410*), serine recombinases (*470*) or the combination of both (*410*) does not represent a substantial advantage over already existing technologies in *S. cerevisiae*. This is therefore not a priority for the expansion of the *S. cerevisiae* CRISPR-based toolbox. Still, lack of proof-of-principle experiments in the model yeast limits their application in non-model yeasts, where these tools might be more valuable.

8.3 Diversity of Cas12 variants as a source for novel genome editing tools

Recent access to genomic and metagenomic data has resulted in the identification of a wide range of novel putative CRISPR-Cas systems. Among newly identified candidates, especially a large number of Cas12 variants have been discovered and characterized in the recent years, revealing diverse functionalities and sizes (*77*) (Table 8.1). The rapidly expanding diversity of CRISPR-Cas systems emphasizes the importance of a systematic and unique nomenclature system for designation of newly characterized effectors (*140*, *255*, *474*), which would avoid confusion.

Table 8.1. Type V effector proteins and characteristics. Effectors names are provided with the Cas12-based nomenclature. Alternative names are provided as “old nomenclature”. PDB ID are provided for Binary complexes (BC) and ternary complexes (TC).

Effector	Old nomenclature	PAM	Subtype	Effector size (aa)	crRNA: effector	Additional RNA required	length of tracrRNA	pre-processing activity	Substrate	Collateral activity	PDB ID	Ref.
Cas12a	Cpf1	5’- (T) TTV-3’	V-A	1226-1307, 1100-1300	1:1	No		Yes	dsDNA and ssDNA	ssDNA	5NG6 (BC), 6I1L (TC)	(*122*)
Cas12b	C2c1	5’-TTN-3’	V-B	1108-1129	1:1	tracrRNA	~ 130 nt	No	dsDNA	ssDNA	5u33	(*141*)
Cas12c	C2c3	5’-TG-3’/5’-TN-3’	V-C	1209-1330	1:1	tracrRNA (scoutRNA)	~ 100 nt	Yes	dsDNA	ssDNA		(*73*, *141*)
Cas12c2	C2c3	5’-TN-3’	V-C	1218	1:1	tracrRNA	71 nt	Yes	dsDNA	ssDNA	7V93 (BC), 7V94 (TC)	(*475*)
Cas12d	CasY	5’-TR-3’	V-D	~ 1200	1:1	scoutRNA	50- to 100-nt	Yes	dsDNA	ssDNA		(*73*, *243*)
Cas12e	CasX	5’-TTCN-3’	V-E	< 1000	1:1	tracrRNA	~ 110 nt	?	dsDNA	ssDNA		(*153*, *476*),
Cas12f1	Cas14a	5’-TTR-3’	V-F1	400-700	1:2	tracrRNA	~170 nt	No	dsDNA	ssDNA	7C7L (TC), 7L48 (BC), 7L49 (TC)	(*130*, *133*, *257*, *477*)
Cas12f1	c2c10		V-F1 (V-U3)	400-700	?	tracrRNA		No	?	?		(*130*, *257*)
Cas12f2	Cas14b	5’-TTAT-3’	V-F2	400-700	?	tracrRNA		?	dsDNA	?		(*133*)
Cas12f3	Cas14c		V-F3	400-700	?	tracrRNA		?	?	?		(*257*)
Cas12g		No		720-830	1:1	tracrRNA	92 nt	No	ssRNA	ssDNA, ssRNA	6XMG (TC)	(*129*, *131*)
Cas12h		5’-RTR-3’	V-H	870-933	1:1	No		?	dsDNA and ssDNA	ssDNA		(*129*)

Table 8.1. Type V effector proteins and characteristics. Effectors names are provided with the Cas12-based nomenclature. Alternative names are provided as "old nomenclature". PDB ID are provided for Binary complexes (BC) and ternary complexes (TC). *(continued)*

Effector	Old nomenclature	PAM	Subtype	Effector size (aa)	crRNA: effector	Additional RNA required	length of tracrRNA	pre-processing activity	Substrate	Collateral activity	PDB ID	Ref.
Cas12i1		5'-TTN-3'	V-I	1033-1093	1:1	No		Yes	dsDNA (nickinig) and ssDNA	ssDNA	6W5C (TC), 6W62 (BC)	(*129*, *478*)
Cas12i2		5'-TTN-3'	V-I	1054	1:1	No		Yes	dsDNA	ssDNA	6LTR, 6LTU and 6LU0 (TC), 6LTP (BC)	(*479*)
Cas12j1, Cas12j2	CasΦ-1, CasΦ-2	5'-TTN-3'		700-800	1:1	No		Yes	dsDNA	ssDNA		(*480*)
Cas12j3				700-800	1:1	No		No			7ODF	
Cas12k	c2c5	5'-GTN-3'	V-K (V-U5)	639	1:1	tracrRNA	216 nt	?	dsDNA	No	7N3P (TC), 7N3O (BC)	(*404*, *408*)
Cas12m	C2c4	5'-NTTN-3'	V-M (V-U1)	596	1:1	No		Yes	dsDNA	No		(*316*)
c2c8	c2c8		V-U2									
c2c9	c2c9		V-U4									
TnpB		TAM: 5'-TTGAT-3'		400	omega RNA:1	No		?	dsDNA			(*262*)

8.3.1 Diverse mechanisms of Cas12 variants

Type V CRISPR-Cas systems are characterized by effector proteins that contain a single nuclease domain (RuvC) (*122*, *126*). These RuvC domains have been shaped by evolution to give rise to divergent sequences and domain architectures. The most likely ancestor of all Cas12 variants (*127*, *255*), is a TnpB transposon of the IS200/IS605 and IS607 transposon families, that recently has been demonstrated to function as an RNA-guided DNA endonuclease allowing for RNA guided transposition (*261*, *262*). Although it has not been found in association with CRISPR arrays, TnpB uses small RNA guides derived from the right-end element of the transposon, and can generate DSBs next to 5'-TTGAT transposon-associated motifs (*262*). TnpB includes the characteristic Cas12 RuvC motif with a well-conserved catalytic triad (DED). Characterization of this evolutionary ancestor and characterization of evolutionary-derived Cas12 variants has increased (and will increase) the number of available tools for genome editing.

Differences in domain architecture of Cas12 variants are the basis of the current type V classification (*255*). The Type V sub-types differ substantially with respect to pre-crRNA processing capacities (ability to self-process or RNaseIII/tracrRNA-dependent cleavage), PAM requirements, target specificity and collateral activity (Figure 8.2).

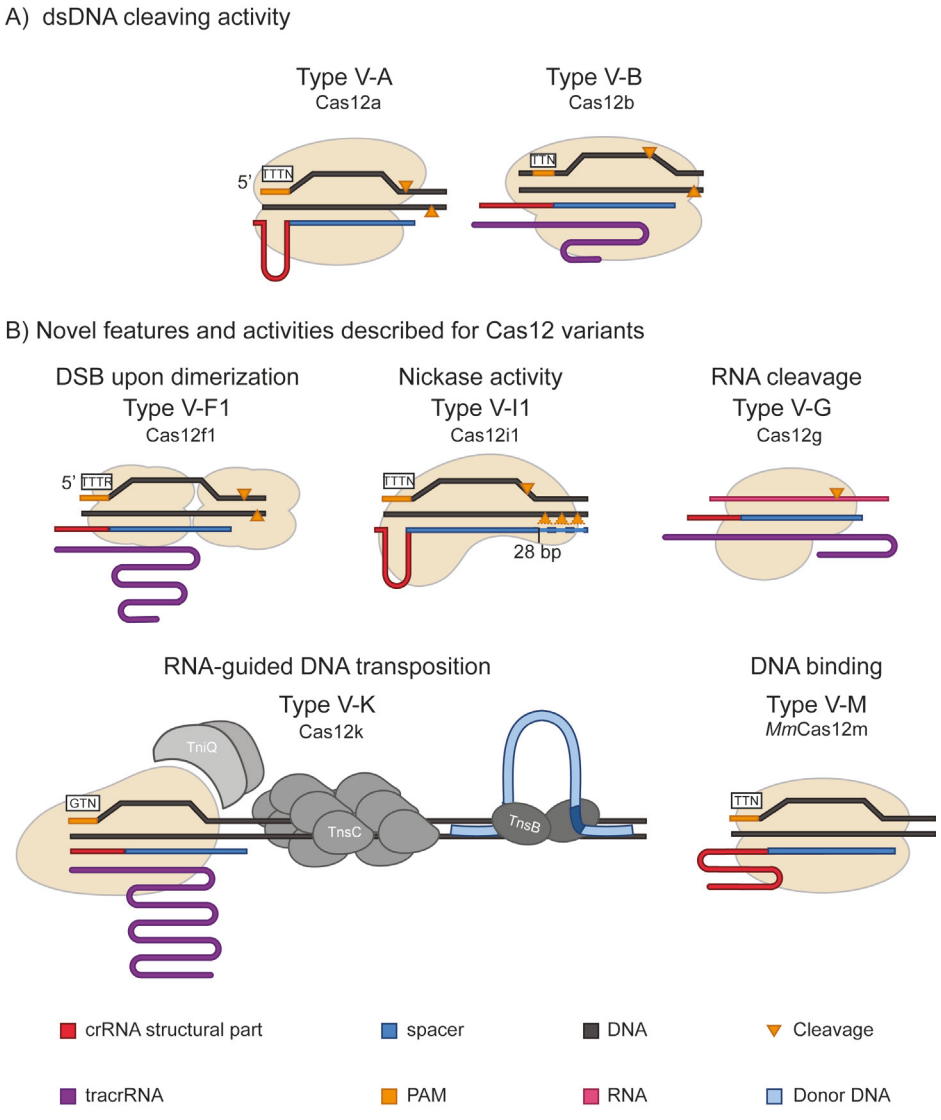


Figure 8.2. Beyond dsDNA cleavage activities described for Cas12 variants. (A) Double-strand DNA cleavage activity of Cas12 variants, described for Cas12a, Cas12b, Cas12c, Cas12d, Cas12e, Cas12h and Cas12j. **(B)** Novel activities described for several Cas12 variants: dsDNA cleavage upon dimerization has been described for Cas12f1, nickase activity has been described for Cas12i1 when the crRNA:DNA heteroduplex is shorter than 28 bp, RNA cleavage activity has been described for Cas12g, RNA-guided DNA transposition activity has been described for Cas12k in combination with TnsB, TniQ and TnsC and DNA binding activity has been described for *Mm*Cas12m.

Pre-crRNA processing activity in the absence of tracrRNA is one of the most attractive features of Cas12a, as it simplifies multiplexing applications (*151*). This feature has also been shown for other small Cas12 variants, such as Cas12i (*129*), Cas12j (*259*) and Cas12m (*316*) (Chapter 3). However, the guide maturation mechanisms used by these enzymes varies is diverse. Cas12a and Cas12i use a catalytic site that resides in the Wedge (WED) domain to cleave their crRNA upstream of the crRNA pseudoknot in a divalent cation-independent process (*126*, *478*). A similar mechanism has also been suggested for Cas12m (*316*), although details of the catalytic site remain elusive (Chapter 3). In contrast, Cas12j1 and Cas12j2 have been shown to use the RuvC domain for this task in a metal-dependent mechanism (*259*). On the other hand, and similar to type II effectors, Cas12b (*152*), Cas12e (*153*), Cas12f1 (*133*), Cas12g (*131*) and Cas12k (*258*) require the assistance of a tracrRNA for generating a long double-stranded RNA structure (crRNA repeat/tracrRNA anti-repeat) that most likely allows for RNaseIII pre-processing. The length and secondary structure of the identified tracrRNAs are diverse (Table 8.1). Finally, a new CRISPR-related non-coding RNA species has been described for subtypes V-C and V-D, the short-complementarity untranslated RNA (scoutRNA) (*73*). This tracrRNA analog is required by Cas12c and Cas12d effectors for crRNA maturation and DNA cleavage. It has been claimed that scoutRNAs differ from previously described tracrRNAs in their secondary structure and the presence of a five-nucleotide conserved sequence essential for activity (*73*). In the case of Cas12c, a metal-dependent mechanism involving the RuvC domain seems to be responsible for pre-crRNA processing, since dCas12c (D928A) is not able to process its pre-crRNA (*475*). Finally, pre-crRNA processing by some other recently described Cas12 variants has either not yet been studied (Cas12h) or not observed (Cas12g1). In the case of Cas12g1, crRNA pre-processing was not observed neither with nor without a tracrRNA, which suggests that additional host factors may be required for *in vivo* crRNA maturation (*129*).

A common feature for almost all characterized Cas12 variants to date is the requirement of a PAM upstream of the protospacer, although a wide diversity of sequence requirements has been described. As described for *Mm*Cas12m (Chapter 3), several Cas12 variants, (e.g., Cas12a, Cas12b, Cas12i orthologs), require T-rich PAMs. However, even between closely related Cas12a orthologs, some variants have been described which accommodated multiple G's in their PAM sequences (*128*, *258*, *481*). Interestingly, the singular RNA-targeting Cas12g does not require a PAM. Moreover, different domains are involved in PAM recognition among Cas12 variants. While most variants use the PAM interacting (PI) domain to interact with the different PAMs, Cas12b and Cas12e have been shown to use the REC and the WED domains for the same purpose (*153*, *272*).

The nuclease activity of Cas12 variants originates from their RuvC nuclease domain. Active RuvC domains in Cas12 variants share a conserved catalytic triad (DED). In other nucleases (e.g. Cas9), this triad recruits a divalent cation (Mg^{2+}) that coordinates a water molecule required for the hydrolysis of DNA (*482–484*). Substitutions of one or more of these residues diminish or completely abolish the nuclease activity of Cas12 variants, as has been reported for Cas12k and Cas12m (*122*, *131*). The nuclease activity of this domain is activated upon R-loop formation

and hybridization of the crRNA to the target strand (TS). This induces a conformational change in Cas12 variants that involves the displacement of a lid motif, which covers the RuvC domain in its inactive state (*475*). Opening of the RuvC domain allows access of the RuvC catalytic site to the non-target strand, resulting in its cleavage. Subsequentially, the TS is repositioned towards the active site and gets cleaved (*126*, *485*). This mechanism is shared among multiple Cas12 variants such as Cas12a (*126*), Cas12b (*272*), Cas12e (*153*), Cas12i and Cas12g (*131*, *478*). Interestingly, several phylogenetically-related Cas12b and Cas12i orthologs have been initially characterized as dsDNA nickases. Their nickase activity has been related to the lack of access of the TS to the RuvC active site upon NTS cleavage. In the case of some Cas12b orthologs, cleavage of the TS has been achieved upon incorporation of positively charged amino acids in the pocket between the crRNA:TS-DNA hybrid and the RuvC active site, suggesting that these residues might be pulling the TS for cleavage at the RuvC active sites. In the case of Cas12i, access of the TS to the RuvC active site is only granted after propagation of the crRNA:DNA heteroduplex to position 28, which promotes an inward movement of the REC1 domain towards the heteroduplex (*478*). This allows cleavage of the TS only when using spacers longer than 28 nucleotides (*478*). However, this characteristic is not shared with Cas12i2, which is able to cleave dsDNA using spacers of 14 nt (*479*). Better understanding of structural features of such Cas12 variants might allow rational engineering of these proteins, which may result in effector proteins with increased efficiencies or new functionalities.

Interestingly, some Cas12 variants with predicted RuvC domains do not present the conserved DED catalytic triad in charge of nucleic acid cleavage (*127*). These variants have been shown to bind dsDNA but not cleave it, as it is the case for two of the Cas12 effectors studied in this thesis, Cas12m and Cas12k. In the case of Cas12m, we showed that this nuclease is able to bind DNA and interfere with plasmid replication, potentially explaining its biological function (Chapter 3) (*316*). In the case of Cas12k, the protein is involved in directing a Tn7-like transposon to target loci, by interacting with TniQ, TnsC and TnsB proteins (*258*). Finally, a relevant, unique activity has been described for Cas12g, a protein that is able to target ssRNA. This protein shares the bilobed architecture described for Cas12a, Cas12b, Cas12e and Cas12i (*478*), and it has a canonical DED catalytic triad (*131*). However, differences in the REC1 and REC2 domains might be the reason of its unique RNA targeting activity (*131*). The small size of Cas12g (767 aa) together with its RNA-targeting activity, makes it an interesting candidate for *in vivo* transcriptome engineering applications (*131*).

Recently, other small Cas13 variants (type VI) have been characterized which are also able to target RNA in a similar way as Cas12g does (*486*). These variants belong to the subtypes VI-D (928 aa) and VI-BT (775–804 aa), act in a protospacer flanking sequence-independent manner *in vivo* and use two ribonuclease domains (HEPN) that together form an active site that catalyzes target and collateral RNA cleavage (*486–488*). The mentioned Cas12g and small Cas13d and Cas13bt variants could be interesting candidates to assay as RNA-guided RNA degradation tools, since activity of Cas13a has never been reported in *S. cerevisiae*.

Another remarkable characteristic of the type V effector RuvC domains is the activation of collateral cleavage upon target recognition (*260*). Besides its biological role in initiating dormancy or even suicide of the infected host cell (*81*), this activity has been harnessed for multiple biotechnological applications (*260*, *489–491*). Collateral activity towards ssDNA has been described for almost all Cas12 variants (except Cas12m and Cas12k) (*130*), with Cas12g alone also harnessing collateral ssRNA cleavage.

Finally, crystallization and cryo-EM structures of Cas12 variants allowed elucidation of important details regarding complex stoichiometries and protein/guide/target interactions. One of the most unique findings regarding complex stoichiometry is the dimerization requirement of subtype V-F1 effectors in presence of a single copy of a tracrRNA and a crRNA (*133*, *492*). To date, this feature has not been described for any other Cas12 variant, nor for the closest evolutionary neighbor, the TnpB transposon. A protein:reRNA (right end element RNA, specific for TnpB) molar ratio of 1:1 has been determined for the TnpB RNP complex through mass photometry (*262*). Other relevant findings are the insights on complex stoichiometry required for RNA-guided DNA transposition using the type V CRISPR-associated transposon system. Multiple TnsC monomers are required to polymerize and form filaments around the DNA, while two TniQ monomers interact with two TnsC monomers each (*404*). At the same time, TnsB is only able to interact with TnsC when the last is part of a filament. Therefore, for both type V-F and type V-K systems, protein polymerization is a requirement for proper functioning of the effector complex, which might hamper the application of these systems in eukaryotic organisms. Nucleosomal configuration of DNA in eukaryotic systems might limit these protein interactions, although the Cascade complex has successfully been expressed functionally in human cells resulting in genome editing activity (*401*).

8.3.2 Current application of Cas12 nucleases

Cas12 nucleases have been used as genome editing tools in a wide diversity of organisms. Although first characterized *in vitro* and in *E. coli*, most of them have been adapted for genome editing applications in eukaryotic organisms with different success levels (*398*, *493*). Undoubtedly, Cas12a is the most widely adopted type V effector for genome editing applications in eukaryotic organisms. In this thesis, we used it both for genome editing and genome regulation applications (Chapter 4 and Chapter 7). Other variants such as Cas12b (*152*), Cas12e (*153*, *476*) and Cas12f (*132*, *397*, *494*) have also been used in mammalian cells with different degree of success. However, the low activity of some of these thermophilic variants at moderate temperatures, currently limits their application for genome editing purposes (*397*). We also attempted to express Cas12f1 and Cms1 (a small Cas12a ortholog) (*495*) in *S. cerevisiae* but observed no activity of the mentioned proteins in the eukaryotic model organism (data not included), although protein expression was successful. In the case of Cms1, activity of this ortholog has only been reported in plants (*93,* personal communication Schaart *et al.,*). Structural comparison of *Sm*Cms1 to *Lb*Cas12a shows the absence of the Nuc domain, which might alter the conformation and/or flexibility of the active RuvC domain, affecting the cleaving activity. Regarding Cas12f1, we attempted to use the *As*Cas12f1 ortholog in *S. cerevisiae* for

both genome editing and base editing and observed no activity of the protein for any of the applications. Similar observations have been reported by other authors when using Cas12f1 in plants and human cells (*291*, *397*). This indicates that functional transplantation of some type V effectors in eukaryotic systems remains challenging for unknown reasons. For instance, some Cas12f variants required protein and/or tracrRNA engineering for detectable activity in mammalian cells (*132*).

Among Cas12 variants, the association of Cas12k with Tn7-like transposition systems presents a novel functionality for type V effectors. Cas12k presents a deactivated RuvC catalytic triad and thus, does not induce dsDNA breaks (*258*). Association with Tn7-like transposon systems allows RNA-guided DNA integration independently of the host's homology-directed recombination system. Although this type of coexistence and interaction between a CRISPR-Cas subtype and a transposition system is not unique for subtype V-K (*384*, *496*), this is the most compact system described to date allowing RNA-guided DNA transposition. Activity of this system has been proven *in vitro* and in multiple bacteria (*393*), although with high off-target rates. We confirmed activity of the system in *E. coli* and determined critical protein fusion products preventing transposition (Chapter 6). We showed that protein fusion products to the C-terminus of TnsB and TnsC prevent transposition, highlighting the importance of these positions (*404*). Although we confirmed that all proteins can be individually tagged with nuclear localization tags, we failed in establishing RNA-guided DNA transposition in *S. cerevisiae*. Already in *E. coli* it has been shown that the type V-K transposition system is prone to yield co-integration products, consisting of the integration of the transposon and the plasmid backbone. Attempts to prevent this by supplying a linear donor DNA fragment (*386*) did not result in the desired transposition. To date, activity of CRISPR-associated transposon systems has been limited to bacterial systems. Future studies should focus on the identification of the limiting step that prevents transposition in eukaryotes. For instance, localization studies of each of the type V CRISPR-associated transposon system should determine whether all complex proteins localize in the nucleus. Moreover, protein pull-down purification assays in eukaryotic organisms could be used to determine which interactions are currently taking place. Finally, fusion products between components of the Tn7-like transposition complex could facilitate transposition. A similar approach has already been used in *E. coli*, yielding a functional RNA-guided transposition complex (*407*).

The small size of some Cas12 variants could encourage the development of transposon engineered systems based on the fusion of a DNA-binding Cas12 variant or a deactivated Cas12 variant to a transposase. Similar approaches have already been developed for dCas9 and the sleeping beauty (*469*) or piggyBac transposases (*100*, *101*), as well as for prime editors in combination with serine recombinases (*410*).

8.4 Future prospects for transferability of CRISPR-Cas technologies to yeast

Current strategies to express CRISPR-Cas systems in yeast and other eukaryotic organisms involve common protein expression strategies used for expression of proteins (Figure 8.3). Commonly, the gene sequence of novel CRISPR-Cas effectors is codon optimized/harmonized for expression in the organism of choice. Optimization for human expression has successfully been used for expression of CRISPR-Cas editors in yeast (*84*, *111*). Choice of adequate promoter sequences for expression of the CRISPR effectors or the spacer sequences is also key in compartmentalized eukaryotic organisms. The use of RNA polymerase II promoters is already implemented for expression of the Cas effectors, although the expression level of the effectors often needs to be fine-tuned to prevent toxicity (*111*). For instance, *Mm*Cas12m could not be expressed in high levels in yeast due to expression issues and instability of the protein (Chapter 4). Another important step when expressing DNA-targeting Cas effectors in yeast or other eukaryotic systems is the addition of nuclear localization signals to the proteins which activity is required in the nucleus. During this thesis, we confirmed that *Fn*Cas12a can reach the nucleus of *S. cerevisiae* without a nuclear localization signal ((Chapter 7), data not shown). Most likely, presence of positively charged amino acids in the *Fn*Cas12a sequence allows for nuclear translocation in the absence of a specific nuclear localization signal. This was also observed when using Cas12a for genome silencing (*112*). Still, it is important to evaluate whether addition of nuclear localization signals to Cas effectors hampers the activity of the proteins or of the eventual complexes these proteins may be part of. For instance, we showed the potentially negative effect of adding nuclear localization signals on the type V Tn7-like CRISPR-associated transposon (Chapter 6). Finally, these interrogation studies can be used to assess whether N- or C- terminal positions can be used to fuse specific domains. For instance, addition of base editor domains to the N-terminal of Cas12m has been reported as detrimental for the binding activity of this protein (Chapter 3). Still, even after taking care of all these improvement items, expression of Cas effectors in a novel organism may not lead to detection of its activity.

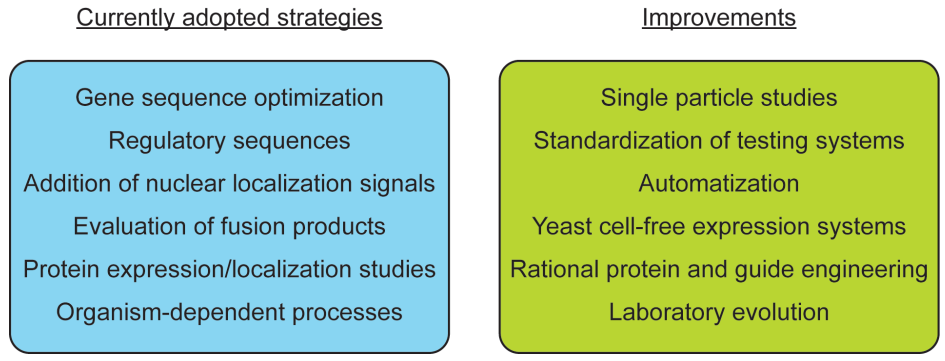


Figure 8.3. Current and eventual strategies to adopt for the implementation of CRISPR-Cas tools in yeast. Currently adopted strategies for the application of CRISPR-Cas effectors in yeast (left) and eventual strategies to be implemented for improvement of the process (right).

CRISPR-Cas systems originate from bacterial and archaeal systems. In their natural environments, these prokaryotic systems interact with DNA and RNA from the host or viral systems. The repurposing of these systems for genome editing applications in eukaryotic systems requires proper interaction of the Cas effectors with eukaryotic DNA. As shown in this thesis, expression of the proteins and proper translocation to the nucleus is not sufficient to detect equivalent activities as observed in bacteria. Although bacterial genomes also present DNA-associated, genome condensation proteins such as the factors of inversion stimulation (FIS) or the heat-unstable nucleoid proteins (HU), the topology of the DNA between prokaryotes and eukaryotes is vastly different (*497*). A clear example of this bottleneck is the fact that deactivated Cas9 or Cas12a proteins alone can be used to repress gene expression in bacteria in an effective way (*498*, *499*). On the other hand, fusion of repressor-recruiting domains is necessary to achieve an equivalent repression effect in yeast (*164*). From a fundamental point of view, single particle tracking experiments through super resolution microscopy could help elucidating the behavior of novel Cas effectors in live eukaryotic cells (*325*, *500*). Comparison of diffusion coefficients for different Cas effectors could then provide information of different DNA interaction patterns.

During this thesis, we also evaluated the transferability to *S. cerevisiae* of SIBR-Cas, a CRISPR-Cas based tool initially developed for bacteria. The tunability of this tool at RNA level represented a challenge for the use of this tool in the model yeast. SIBR-Cas is based on the control of the expression of *Fn*Cas12a via an inducible intron placed at the 5'- end of its transcript (*369*). The intron contains premature termination codons that prevent the translation of *Fn*Cas12a in the absence of splicing. Splicing of the intron can be induced by theophylline addition. In *S. cerevisiae*, a native mechanism prevented the direct transfer of the SIBR-Cas system (Chapter 7). Non-sense mediated mRNA decay is a surveillance mechanism in eukaryotic organisms that prevents the accumulation of premature termination codons-containing transcripts in eukaryotes (*448*, *449*). Therefore, SIBR-containing transcripts are natural targets for this pathway, among others. Induction of splicing in *S. cerevisiae* by addition of theophylline was achieved when the intron was placed further downstream into the *Fn*Cas12a transcript. This modification seemed to avoid the activity of non-sense mediated mRNA decay. In this case, transferability of SIBR-Cas to *S. cerevisiae* required the by-pass of a eukaryotic control mechanism not present in bacteria, highlighting the importance of considering organism-dependent processes and mechanisms for the implementation of CRISPR-based tools. Future steps will focus on the use of this tool in non-model yeasts, where the time control over the activity of Cas12a could be used in favor of allowing more homologous recombination events before the activity of Cas12a.

Assessing the activity of novel Cas effectors as genome editing or gene expression regulation tools requires powerful *in vivo* characterization systems. Differences in testing setups hinders direct comparison of editing efficiencies or expression change. For instance, the vast majority of studies assessing the activity of genome editing tools in yeast present gene editing efficiencies as number of edited colonies over number of obtained colonies per transformation (*90*, *111*, *120*), obviating the number of obtained colonies. Moreover, in multiple occasions, results from transformations incorporating selection markers are directly compared to transformations without selection (*94*). Standardized characterization systems or the inclusion of control proteins should be a staple in the assessment of the activity of novel proteins, as done when assessing the silencing activity of Cas12m in yeast (Chapter 4).

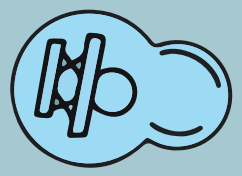
As shown in this thesis, characterization of novel Cas12 variants has yielded multiple variants with low or no activity in eukaryotic systems. The wide variety of systems and sequences available allows for big assessment programs in which more than a handful of proteins are simultaneously tested for desired activity in such systems. However, this requires a certain degree of automatization and standardization not available in all laboratories. Eukaryotic cell-free systems could help in the future to speed-up initial screening programs (*273*, *500*). Moreover, to date several Cas12 variants have been engineered to increase their activity in eukaryotic systems. For instance, Cas12b or Cas12f1 variants required rational protein engineering approaches or underwent laboratory evolution programs to show detectable activities in human cells (*132*, *152*). High-resolution cryo-EM structures or crystal structures of Cas12 variants can also provide information at the level of residue interaction with the guide sequences or the target DNA. Following this approach, guide RNAs as well as protein domains can be, to some extent, rationally engineered. For instance, the DNA interaction domains of several Cas12f1 variants have been engineered by incorporation of several positively charged amino acids to presumably increase protein:DNA and/or RNA affinities (*291*).

Increasing fundamental knowledge on Cas12 variants should enlarge the list of available genome editing and regulation tools in *S. cerevisiae* and other yeasts. Moreover, diversity of Cas systems and, specifically, of Cas12 systems ensures the future characterization of novel Cas effectors useful for genome editing or other applications. Knowledge sharing regarding positive or negative results on the application of novel genome editing tools would also considerably speed up the development of novel tools. These tools should then help shortening strain building programs focused on the industrial use of *S. cerevisiae* and non-model yeasts.

8.5 Conclusion

Evolution has provided a wide range of CRISPR-Cas effectors which can be eventually used as genome editing tools. Fundamental research on the functionality of these effectors is key for their application in different organisms. However, application of novel genome editing tools in different organisms can be challenging. Organisms-dependent conditions might alter the functionality of the implemented tools. In the case of CRISPR-Cas systems, this can lead to lack of activity of the implemented tools. Bigger screening programs with simultaneous testing of multiple Cas effectors might increase the chance to find active variants in the organism we want to modify. Rational protein engineering approaches or laboratory evolution strategies can improve Cas effector activities. Undoubtedly, a lot more CRISPR-Cas based applications, in many more organisms, are still to come.

APPENDICES



REFERENCES

1. N. P. Money, *The rise of yeast: how the sugar fungus shaped civilisation* (Oxford University Press, 2018).
2. L. Liu, J. Wang, D. Rosenberg, H. Zhao, G. Lengyel, D. Nadel, Fermented beverage and food storage in 13,000 y-old stone mortars at Raqefet Cave, Israel: Investigating Natufian ritual feasting. *J. Archaeol. Sci. Reports*. **21**, 783–793 (2018).
3. J. Wang, L. Jiang, H. Sun, Early evidence for beer drinking in a 9000-year-old platform mound in southern China. *PLoS One*. **16**, e0255833 (2021).
4. P. McGovern, M. Jalabadze, S. Batiuk, M. P. Callahan, K. E. Smith, G. R. Hall, E. Kvavadze, D. Maghradze, N. Rusishvili, L. Bouby, O. Failla, G. Cola, L. Mariani, E. Boaretto, R. Bacilieri, P. This, N. Wales, D. Lordkipanidze, Early Neolithic wine of Georgia in the South Caucasus. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E10309–E10318 (2017).
5. S. Bhatia, D. Goli, *Introduction to Pharmaceutical Biotechnology, Volume 2* (2018).
6. J. G. Krünitz, *Bier 1775: aus J.G. Krünitz Oeconomische Encyclopädie* (Olms Presse, 1976; https://books.google.nl/books?id=TBAzAQAAMAAJ).
7. C. N. Frey, History and Development of the Modern Yeast Industry. *Ind. Eng. Chem.* **23**, 340 (1931).
8. T. Riis, The Story of Emil Chr, Hansen. *J. Eur. Econ. Hist.* **41**, 159 (2012).
9. J. Nielsen, Yeast Systems Biology: Model Organism and Cell Factory. *Biotechnol. J.* **14** (2019), doi:10.1002/BIOT.201800421.
10. S. Wu-Pong, Y. Rojanasakul, *Biopharmaceutical Drug Design and Development* (1999).
11. W. L. Tang, H. Zhao, Industrial biotechnology: Tools and applications. *Biotechnol. J.* **4**, 1725–1739 (2009).
12. Z. Tong, X. Zheng, Y. Tong, Y. C. Shi, J. Sun, Systems metabolic engineering for citric acid production by *Aspergillus niger* in the post-genomic era. *Microb. Cell Fact.* **18**, 1–15 (2019).
13. A. Tesfaw, F. Assefa, Current Trends in Bioethanol Production by *Saccharomyces cerevisiae* : Substrate, Inhibitor Reduction, Growth Variables, Coculture, and Immobilization. *Int. Sch. Res. Not.* **2014**, 1–11 (2014).
14. S. H. Mohd Azhar, R. Abdulla, S. A. Jambo, H. Marbawi, J. A. Gansau, A. A. Mohd Faik, K. F. Rodrigues, Yeasts in sustainable bioethanol production: A review. *Biochem. Biophys. Reports*. **10**, 52–61 (2017).
15. S. Cleto, J. V. Jensen, V. F. Wendisch, T. K. Lu, *Corynebacterium glutamicum* Metabolic Engineering with CRISPR Interference (CRISPRi). *ACS Synth. Biol.* **5**, 375–385 (2016).
16. A. S. Verma, S. Agrahari, S. Rastogi, A. Singh, Biotechnology in the Realm of History. *J. Pharm. Bioallied Sci.* **3**, 321 (2011).
17. O. T. Avery, C. M. MacLeod, M. McCarty, Studies on the chemical nature of the substance inducing transformation of *Pneumococcal* types. *J. Exp. Med.* **79**, 137–158 (1944).
18. A. Klug, Rosalind Franklin and the discovery of the structure of DNA. *Nature*. **219**, 808–844 (1968).
19. F. Crick, J. Watson, Molecular structure of nucleic acids. *Nature*. **171** (1953).
20. F. Crick, Central Dogma of Molecular Biology. **227**, 561–563 (1970).
21. F. H. C. Crick, The origin of the genetic code. *J. Mol. Biol.* **38**, 367–379 (1968).
22. S. N. Cohen, A. C. Y. Chang, H. W. Boyert, R. B. Hellingt, Construction of Biologically Functional Bacterial Plasmids In Vitro (R factor/restriction enzyme/transformation/endonuclease/antibiotic resistance). *Proc. Natl. Acad. Sci.* **70**, 3240–3244 (1973).

23. R. K. Saiki, D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, H. A. Erlich, Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. *Science (80-.).* **239**, 487–491 (1988).
24. B. R. Glick, C. L. Patten, *Molecular biotechnology: principles and applications of recombinant DNA* (John Wiley & Sons, 2017), vol. 34.
25. L. Caspeta, T. Castillo, J. Nielsen, Modifying yeast tolerance to inhibitory conditions of ethanol production processes. *Front. Bioeng. Biotechnol.* **3**, 184 (2015).
26. S. Yi, X. Zhang, H. xin Li, X. xia Du, S. wei Liang, X. hua Zhao, Screening and Mutation of *Saccharomyces cerevisiae* UV-20 with a High Yield of Second Generation Bioethanol and High Tolerance of Temperature, Glucose and Ethanol. *Indian J. Microbiol.* **58**, 440 (2018).
27. D. M. Wuest, S. Hou, K. H. Lee, Metabolic Engineering. *Compr. Biotechnol. Second Ed.* **3**, 617–628 (2011).
28. R. Liu, M. C. Bassalo, R. I. Zeitoun, R. T. Gill, Genome scale engineering techniques for metabolic engineering. *Metab. Eng.* **32**, 143–154 (2015).
29. J. Becker, E. Boles, A modified *Saccharomyces cerevisiae* strain that consumes L-arabinose and produces ethanol. *Appl. Environ. Microbiol.* **69**, 4144–4150 (2003).
30. M. M. Zhang, Y. Wang, E. L. Ang, H. Zhao, Engineering microbial hosts for production of bacterial natural products. *Nat. Prod. Rep.* **33**, 963 (2016).
31. W. Ye, W. Zhang, T. Liu, G. Tan, H. Li, Z. Huang, Improvement of ethanol production in *Saccharomyces cerevisiae* by high-efficient disruption of the *ADH2* gene using a novel recombinant TALEN vector. *Front. Microbiol.* **7**, 1067 (2016).
32. D. E. Cameron, C. J. Bashor, J. J. Collins, A brief history of synthetic biology. *Nat. Rev. Microbiol. 2014 125*. **12**, 381–390 (2014).
33. S. Le Borgne, in *Recombinant Gene Expression* (Springer, 2012), pp. 451–465.
34. J. Hou, K. E. J. Tyo, Z. Liu, D. Petranovic, J. Nielsen, Metabolic engineering of recombinant protein secretion by *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **12**, 491–510 (2012).
35. V. Meyer, Genetic engineering of filamentous fungi - Progress, obstacles and future trends. *Biotechnol. Adv.* **26**, 177–185 (2008).
36. J. Lian, S. Mishra, H. Zhao, Recent advances in metabolic engineering of *Saccharomyces* cerevisiae: New tools and their applications. *Metab. Eng.* **50**, 85–108 (2018).
37. J. Nielsen, J. D. Keasling, Engineering Cellular Metabolism. *Cell.* **164**, 1185–1197 (2016).
38. I. Mougiakos, E. F. Bosma, W. M. de Vos, R. van Kranenburg, J. van der Oost, Next Generation Prokaryotic Engineering: The CRISPR-Cas Toolkit. *Trends Biotechnol.* **34**, 575–587 (2016).
39. K. S. Pawelczak, N. S. Gavande, P. S. VanderVere-Carozza, J. J. Turchi, Modulating DNA Repair Pathways to Improve Precision Genome Engineering. *ACS Chem. Biol.* **13**, 389–396 (2018).
40. A. K. Löbs, C. Schwartz, I. Wheeldon, Genome and metabolic engineering in non-conventional yeasts: Current advances and applications. *Synth. Syst. Biotechnol.* **2**, 198–207 (2017).
41. Y. Nambu-Nishida, K. Nishida, T. Hasunuma, A. Kondo, Development of a comprehensive set of tools for genome engineering in a cold-and thermo-tolerant *Kluyveromyces marxianus* yeast strain, doi:10.1038/s41598-017-08356-5.
42. B. M. A. Abdel-Banat, S. Nonklang, H. Hoshida, R. Akada, Random and targeted gene integrations through the control of non-homologous end joining in the yeast *Kluyveromyces marxianus*. *Yeast*. **27**, 29–39 (2010).
43. V. Meyer, M. Arentshorst, A. El-Ghezal, A. C. Drews, R. Kooistra, C. A. M. J. J. van den Hondel, A. F. J. Ram, Highly efficient gene targeting in the *Aspergillus niger kusA* mutant. *J. Biotechnol.* **128**, 770–775 (2007).
44. M. G. Fraczek, S. Naseeb, D. Delneri, History of genome editing in yeast. *Yeast.* **35**, 361–368 (2018).

45. S. Le Borgne, Genetic Engineering of Industrial Strains of *Saccharomyces cerevisiae*. *Methods Mol. Biol.* **824**, 451–465 (2012).
46. K. R. Kildegaard, J. A. Arnesen, B. Adiego-Pérez, D. Rago, M. Kristensen, A. K. Klitgaard, E. H. Hansen, J. Hansen, I. Borodina, Tailored biosynthesis of gibberellin plant hormones in yeast. *Metab. Eng.* **66**, 1–11 (2021).
47. B. Sauer, Functional expression of the cre-lox site-specific recombination system in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **7**, 2087 (1987).
48. F. Storici, M. A. Resnick, in *Genetic Engineering* (2003), pp. 189–207.
49. Y. Doyon, T. D. Vo, M. C. Mendel, S. G. Greenberg, J. Wang, D. F. Xia, J. C. Miller, F. D. Urnov, P. D. Gregory, M. C. Holmes, Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat. Methods 2010 81.* **8**, 74–79 (2010).
50. M. Christian, T. Cermak, E. L. Doyle, C. Schmidt, F. Zhang, A. Hummel, A. J. Bogdanove, D. F. Voytas, Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics*. **186**, 757–761 (2010).
51. T. Li, S. Huang, X. Zhao, D. A. Wright, S. Carpenter, M. H. Spalding, D. P. Weeks, B. Yang, Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res.* **39**, 6315–6325 (2011).
52. A. J. Bogdanove, D. F. Voytas, TAL effectors: customizable proteins for DNA targeting. *Science*. **333**, 1843–6 (2011).
53. R. R. Beerli, C. F. Barbas, Engineering polydactyl zinc-finger transcription factors. *Nat. Biotechnol. 2002 202*. **20**, 135–141 (2002).
54. T. Li, S. Huang, W. Z. Jiang, D. Wright, M. H. Spalding, D. P. Weeks, B. Yang, TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.* **39**, 359 (2011).
55. R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, P. Horvath, CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. **315**, 1709–12 (2007).
56. K. S. Makarova, Y. I. Wolf, E. V. Koonin, Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res.* **41**, 4360–4377 (2013).
57. A. Bolotin, B. Quinquis, A. Sorokin, S. Dusko Ehrlich, Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology.* **151**, 2551–2561 (2005).
58. F. J. M. Mojica, C. Díez-Villaseñor, J. García-Martínez, E. Soria, Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *J. Mol. Evol.* **60**, 174–182 (2005).
59. C. Pourcel, G. Salvignol, G. Vergnaud, CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology.* **151**, 653–663 (2005).
60. K. S. Makarova, N. V. Grishin, S. A. Shabalina, Y. I. Wolf, E. V. Koonin, A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct*. **1**, 1–26 (2006).
61. F. J. M. Mojica, C. Díez-Villaseñor, E. Soria, G. Juez, Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.* **36**, 244–246 (2000).
62. L. A. Marraffini, CRISPR-Cas immunity in prokaryotes. *Nature*. **526**, 55–61 (2015).
63. J. van der Oost, M. M. Jore, E. R. Westra, M. Lundgren, S. J. J. Brouns, CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci.* **34**, 401–407 (2009).

64. J. E. Garneau, M. È. Dupuis, M. Villion, D. A. Romero, R. Barrangou, P. Boyaval, C. Fremaux, P. Horvath, A. H. Magadán, S. Moineau, The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*. **468**, 67–71 (2010).
65. R. Barrangou, L. A. Marraffini, CRISPR-cas systems: Prokaryotes upgrade to adaptive immunity. *Mol. Cell*. **54**, 234–244 (2014).
66. I. Yosef, M. G. Goren, U. Qimron, Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* **40**, 5569 (2012).
67. Y. Wei, M. T. Chesne, R. M. Terns, M. P. Terns, Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in *Streptococcus thermophilus*. *Nucleic Acids Res.* **43**, 1749–1758 (2015).
68. 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, J. van der Oost, F. Zhang, Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science (80-.).* **321**, 960–964 (2008).
69. J. Carte, R. Wang, H. Li, R. M. Terns, M. P. Terns, Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **22**, 3489–3496 (2008).
70. E. Deltcheva, K. Chylinski, C. M. Sharma, K. Gonzales, Y. Chao, Z. A. Pirzada, M. R. Eckert, J. Vogel, E. Charpentier, CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*. **471**, 602–607 (2011).
71. M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. **337**, 816–21 (2012).
72. T. Karvelis, G. Gasiunas, A. Miksys, R. Barrangou, P. Horvath, V. Siksnys, crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*. *RNA Biol.* **10**, 841–851 (2013).
73. L. B. Harrington, E. Ma, J. S. Chen, I. P. Witte, D. Gertz, D. Paez-Espino, B. Al-Shayeb, N. C. Kyrpides, D. Burstein, J. F. Banfield, J. A. Doudna, A scoutRNA Is Required for Some Type V CRISPR-Cas Systems. *Mol. Cell*. **79**, 416-424.e5 (2020).
74. D. Gleditzsch, P. Pausch, H. Müller-Esparza, A. Özcan, X. Guo, G. Bange, L. Randau, PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA Biol.* **16**, 504–517 (2019).
75. E. R. Westra, P. B. G. van Erp, T. Künne, S. P. Wong, R. H. J. Staals, C. L. C. Seegers, S. Bollen, M. M. Jore, E. Semenova, K. Severinov, W. M. de Vos, R. T. Dame, R. de Vries, S. J. J. Brouns, J. van der Oost, CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. *Mol. Cell*. **46**, 595–605 (2012).
76. S. Shmakov, A. Smargon, D. Scott, D. Cox, N. Pyzocha, W. Yan, O. O. Abudayyeh, J. S. Gootenberg, K. S. Makarova, Y. I. Wolf, K. Severinov, F. Zhang, E. V Koonin, Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol.* **15**, 169–182 (2017).
77. E. V Koonin, K. S. Makarova, F. Zhang, Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* **37**, 67–78 (2017).
78. E. V. Koonin, K. S. Makarova, Evolutionary plasticity and functional versatility of CRISPR systems. *PLOS Biol.* **20**, e3001481 (2022).
79. F. Hille, H. Richter, S. P. Wong, M. Bratovič, S. Ressel, E. Charpentier, The Biology of CRISPR-Cas: Backward and Forward. *Cell*. **172**, 1239–1259 (2018).
80. O. O. Abudayyeh, J. S. Gootenberg, S. Konermann, J. Joung, I. M. Slaymaker, D. B. T. Cox, S. Shmakov, K. S. Makarova, E. Semenova, L. Minakhin, K. Severinov, A. Regev, E. S. Lander, E. V. Koonin, F. Zhang, C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science*. **353**, aaf5573 (2016).
81. P. Mohanraju, C. Saha, P. van Baarlen, R. Louwen, R. H. J. Staals, J. van der Oost, Alternative functions of CRISPR–Cas systems in the evolutionary arms race. *Nat. Rev. Microbiol.* **20**, 351–364 (2022).

82. L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, F. Zhang, Multiplex genome engineering using CRISPR/Cas systems. *Science*. **339**, 819–823 (2013).
83. O. W. Ryan, J. H. D. Cate, Multiplex engineering of industrial yeast genomes using CRISPRm. *Methods Enzym.* **546**, 473–489 (2014).
84. J. E. DiCarlo, J. E. Norville, P. Mali, X. Rios, J. Aach, G. M. Church, Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* **41**, 4336–4343 (2013).
85. V. Stovicek, C. Holkenbrink, I. Borodina, CRISPR/Cas system for yeast genome engineering: advances and applications. *FEMS Yeast Res.* **17** (2017), doi:10.1093/femsyr/fox030.
86. J. Lian, M. HamediRad, H. Zhao, Advancing Metabolic Engineering of *Saccharomyces cerevisiae* Using the CRISPR/Cas System. *Biotechnol. J.* **13**, 1–11 (2018).
87. Z. Bao, H. Xiao, J. Liang, L. Zhang, X. Xiong, N. Sun, T. Si, H. Zhao, Homology-Integrated CRISPR–Cas (HI-CRISPR) System for One-Step Multigene Disruption in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* **4**, 585–594 (2015).
88. A. A. Horwitz, J. M. Walter, M. G. Schubert, S. H. Kung, K. Hawkins, D. M. Platt, A. D. Hernday, T. Mahatdejkul-Meadows, W. Szeto, S. S. Chandran, J. D. Newman, Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. *Cell Syst.* **1**, 88–96 (2015).
89. T. Jakočiūnas, I. Bonde, M. Herrgård, S. J. Harrison, M. Kristensen, L. E. Pedersen, M. K. Jensen, J. D. Keasling, T. Jakociunas, I. Bonde, M. Herrgard, S. J. Harrison, M. Kristensen, L. E. Pedersen, M. K. Jensen, J. D. Keasling, Multiplex metabolic pathway engineering using CRISPR/Cas9 in *Saccharomyces cerevisiae*. *Metab. Eng.* **28**, 213–222 (2015).
90. R. Mans, H. M. van Rossum, M. Wijsman, A. Backx, N. G. A. Kuijpers, M. van den Broek, P. Daran-Lapujade, J. T. Pronk, A. J. A. van Maris, J.-M. M. G. M. Daran, CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res*. **15** (2015), doi:10.1093/femsyr/fov004.
91. N. S. Mccarty, A. E. Graham, L. Studená, R. Ledesma-amaro, Multiplexed CRISPR technologies for gene editing. *Nat. Commun.* **11**, 1–13 (2020).
92. B. Minkenberg, M. Wheatley, Y. Yang, CRISPR/Cas9-Enabled Multiplex Genome Editing and Its Application. *Prog. Mol. Biol. Transl. Sci.* **149**, 111–132 (2017).
93. R. Ferreira, C. Skrekas, J. Nielsen, F. David, Multiplexed CRISPR/Cas9 Genome Editing and Gene Regulation Using Csy4 in *Saccharomyces cerevisiae*. *ACS Synth Biol*. **7**, 10–15 (2018).
94. D. Ploessl, Y. Zhao, M. Cao, S. Ghosh, C. Lopez, M. Sayadi, S. Chudalayandi, A. Severin, L. Huang, M. Gustafson, Z. Shao, A repackaged CRISPR platform increases homology-directed repair for yeast engineering. *Nat. Chem. Biol. 2021.* **18**, 38–46 (2021).
95. M. Liu, S. Rehman, X. Tang, K. Gu, Q. Fan, D. Chen, W. Ma, Methodologies for improving HDR efficiency. *Front. Genet.* **9**, 691 (2019).
96. V. Tsakraklides, E. Brevnova, G. Stephanopoulos, A. J. Shaw, Improved Gene Targeting through Cell Cycle Synchronization. *PLoS One*. **10**, e0133434 (2015).
97. S. Shi, Y. Liang, M. M. Zhang, E. L. Ang, H. Zhao, A highly efficient single-step, markerless strategy for multi-copy chromosomal integration of large biochemical pathways in *Saccharomyces cerevisiae*. *Metab. Eng.* **33**, 19–27 (2016).
98. A. Kovač, C. Miskey, M. Menzel, E. Grueso, A. Gogol-Döring, Z. Z. Ivics, A. Kovac, C. Miskey, M. Menzel, E. Grueso, A. Gogol-Döring, Z. Z. Ivics, RNA-guided retargeting of *Sleeping Beauty* transposition in human cells. *Elife*. **9**, 1–19 (2020).
99. L. Ye, J. J. Park, M. B. Dong, Q. Yang, In vivo CRISPR screening in CD8 T cells with AAV–Sleeping Beauty hybrid vectors identifies membrane targets for improving immunotherapy for glioblastoma. *Nat. Biotechnol.* **37**, 1302–1313 (2019).

100. M. Pallarès-Masmitjà, D. Ivancic, J. Mir-Pedrol, J. Jaraba-Wallace, T. Tagliani, B. Oliva, A. Rahmeh, A. Sánchez-Mejías, M. Güell, M. Güell, D. Ivančić, J. Mir-Pedrol, J. Jaraba-Wallace, T. Tagliani, B. Oliva, A. Rahmeh, A. Sánchez-Mejías, M. Güell, Find and cut-and-transfer (FiCAT) mammalian genome engineering. *Nat. Commun.* **12**, 1–9 (2021).
101. D. Z. Hazelbaker, A. Beccard, G. Angelini, P. Mazzucato, A. Messana, D. Lam, K. Eggan, L. E. Barrett, A multiplexed gRNA *piggyBac* transposon system facilitates efficient induction of CRISPRi and CRISPRa in human pluripotent stem cells. *Sci. Rep.* **10**, 1–10 (2020).
102. Y. Y. Zhang, J. Wang, Z. Wang, Y. Y. Zhang, S. Shi, J. Nielsen, Z. Liu, A gRNA-tRNA array for CRISPR-Cas9 based rapid multiplexed genome editing in *Saccharomyces cerevisiae*. *Nat. Commun.* **10**, 1053 (2019).
103. M. Deaner, A. Holzman, H. S. Alper, Modular Ligation Extension of Guide RNA Operons (LEGO) for Multiplexed dCas9 Regulation of Metabolic Pathways in *Saccharomyces cerevisiae*. *Biotechnol. J.* **13**, 1700582 (2018).
104. E. D. Jensen, R. Ferreira, T. Jakočiūnas, D. Arsovska, J. Zhang, L. Ding, J. D. Smith, F. David, J. Nielsen, M. K. Jensen, J. D. Keasling, Transcriptional reprogramming in yeast using dCas9 and combinatorial gRNA strategies. *Microb. Cell Fact.* **16**, 46 (2017).
105. K. R. Kildegaard, L. Ribeiro, R. Tramontin, T. J. Goedecke, I. Borodina, CRISPR / Cas9 - RNA interference system for combinatorial metabolic engineering of Saccharomyces cerevisiae. **36**, 237–247 (2019).
106. K. Nishida, T. Arazoe, N. Yachie, S. Banno, M. Kakimoto, M. Tabata, M. Mochizuki, A. Miyabe, M. Araki, K. Y. Hara, Z. Shimatani, A. Kondo, Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science*. **353**, aaf8729 (2016).
107. P. C. Després, A. K. Dubé, M. Seki, N. Yachie, C. R. Landry, Systematic perturbation of yeast essential genes using base editing. *bioRxiv*. **5242**, 677203 (2019).
108. A. V. Anzalone, L. W. Koblan, D. R. Liu, Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38** (2020), pp. 824–844.
109. A. V. Anzalone, P. B. Randolph, J. R. Davis, A. A. Sousa, L. W. Koblan, J. M. Levy, P. J. Chen, C. Wilson, G. A. Newby, A. Raguram, D. R. Liu, Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*. **5** (2019), doi:10.1038/s41586-018-0864-x.
110. R. Verwaal, N. Buiting-Wiessenhaan, S. Dalhuijsen, J. A. Roubos, CRISPR/Cpf1 enables fast and simple genome editing of Saccharomyces cerevisiae. *Yeast*. **35**, 201–211 (2017).
111. M. A. Swiat, S. Dashko, M. den Ridder, M. Wijsman, J. van der Oost, J.-M. M. Daran, P. Daran-Lapujade, M. A. Świat, S. Dashko, M. den Ridder, M. Wijsman, J. van der Oost, J.-M. M. Daran, P. Daran-Lapujade, FnCpf1: a novel and efficient genome editing tool for Saccharomyces cerevisiae. *Nucleic Acids Res*. **45**, 12585–12598 (2017).
112. K. Ciurkot, T. E. Gorochowski, J. A. Roubos, R. Verwaal, Efficient multiplexed gene regulation in *Saccharomyces cerevisiae* using dCas12a. *Nucleic Acids Res.* **49**, 7775–7790 (2021).
113. H. Raschmanová, A. Weninger, A. Glieder, K. Kovar, T. Vogl, Implementing CRISPR-Cas technologies in conventional and non-conventional yeasts: Current state and future prospects. *Biotechnol. Adv.* **36**, 641–665 (2018).
114. P. Cai, J. Gao, Y. Zhou, CRISPR-mediated genome editing in non-conventional yeasts for biotechnological applications. *Microb. Cell Fact.* **18**, 1–12 (2019).
115. Y. Ninomiya, K. Suzuki, C. Ishii, H. Inoue, Highly efficient gene replacements in *Neurospora strains* deficient for nonhomologous end-joining. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 12248–53 (2004).
116. S. H. Kung, A. C. Retchless, J. Y. Kwan, R. P. P. Almeida, Effects of DNA size on transformation and recombination efficiencies in *Xylella fastidiosa*. *Appl. Environ. Microbiol.* **79**, 1712–1717 (2013).

117. F. Easmin, N. Hassan, Y. Sasano, K. Ekino, H. Taguchi, S. Harashima, gRNA-transient expression system for simplified gRNA delivery in CRISPR/Cas9 genome editing. *J. Biosci. Bioeng.* **128**, 373–378 (2019).

118. V. T. Chu, T. Weber, B. Wefers, W. Wurst, S. Sander, K. Rajewsky, R. Kühn, Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol. 2015 335*. **33**, 543–548 (2015).

119. S. Gao, Y. Tong, Z. Wen, L. Zhu, M. Ge, D. Chen, Y. Jiang, S. Yang, Multiplex gene editing of the *Yarrowia lipolytica* genome using the CRISPR-Cas9 system. *J Ind Microbiol Biotechnol*. **43**, 1085–1093 (2016).

120. H. Juergens, J. A. Varela, A. R. G. G. de Vries, T. Perli, V. J. M. M. Gast, N. Y. Gyurchev, A. S. Rajkumar, R. Mans, J. T. Pronk, J. P. Morrissey, J. G. M. G. Daran, Genome editing in *Kluyveromyces* and *Ogataea* yeasts using a broad-host-range Cas9/gRNA co-expression plasmid. *FEMS Yeast Res.* **18**, 12 (2018).

121. I.-S. Jang, B. J. Yu, J. Y. Jang, J. Jegal, J. Y. Lee, Improving the efficiency of homologous recombination by chemical and biological approaches in *Yarrowia lipolytica*. *PLoS One*. **13**, e0194954 (2018).

122. B. Zetsche, J. S. Gootenberg, O. O. Abudayyeh, I. M. Slaymaker, K. S. Makarova, P. Essletzbichler, S. E. Volz, J. Joung, J. van der Oost, A. Regev, E. V Koonin, F. Zhang, Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*. **163**, 759–771 (2015).

123. B. Zetsche, M. Heidenreich, P. Mohanraju, I. Fedorova, J. Kneppers, E. M. DeGennaro, N. Winblad, S. R. Choudhury, O. O. Abudayyeh, J. S. Gootenberg, W. Y. Wu, D. A. Scott, K. Severinov, J. van der Oost, F. Zhang, Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat Biotechnol*. **35**, 31–34 (2017).

124. T. Yamano, H. Nishimasu, B. Zetsche, H. Hirano, I. M. Slaymaker, Y. Li, I. Fedorova, T. Nakane, K. S. Makarova, E. V. Koonin, R. Ishitani, F. Zhang, O. Nureki, Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell*. **165**, 949–962 (2016).

125. P. Gao, H. Yang, K. R. Rajashankar, Z. Huang, D. J. Patel, Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell Res. 2016 268*. **26**, 901–913 (2016).

126. D. C. Swarts, J. van der Oost, M. Jinek, Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Mol. Cell*. **66**, 221-233.e4 (2017).

127. S. Shmakov, A. Smargon, D. Scott, D. Cox, N. Pyzocha, W. Yan, O. O. Abudayyeh, J. S. Gootenberg, K. S. Makarova, Y. I. Wolf, K. Severinov, F. Zhang, E. V. Koonin, Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.* **15**, 169–182 (2017).

128. T. Jacobsen, F. Ttofali, C. Liao, S. Manchalu, B. N. Gray, C. L. Beisel, Characterization of Cas12a nucleases reveals diverse PAM profiles between closely-related orthologs. *Nucleic Acids Res.* **48**, 5624–5638 (2020).

129. W. X. Yan, P. Hunnewell, L. E. Alfonse, J. M. Carte, E. Keston-Smith, S. Sothiselvam, A. J. Garrity, S. Chong, K. S. Makarova, E. V. Koonin, D. R. Cheng, D. A. Scott, Functionally diverse type V CRISPR-Cas systems. *Science (80-.).* **363**, 88–91 (2019).

130. L. B. Harrington, D. Burstein, J. S. Chen, D. Paez-Espino, E. Ma, I. P. Witte, J. C. Cofsky, N. C. Kyrpides, J. F. Banfield, J. A. Doudna, Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science (80-.).* **362**, 839–842 (2018).

131. Z. Li, H. Zhang, R. Xiao, R. Han, L. Chang, Cryo-EM structure of the RNA-guided ribonuclease Cas12g. *Nat. Chem. Biol.* **17**, 387–393 (2021).

132. D. Y. Kim, J. M. Lee, S. Bin Moon, H. J. Chin, S. Park, Y. Lim, D. Kim, T. Koo, J.-H. Ko, Y.-S. Kim, Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. *Nat. Biotechnol.* **40**, 94–102 (2022).

133. S. N. Takeda, R. Nakagawa, S. Okazaki, K. Yamashita, H. Nishimasu, O. N. Correspondence, H. Hirano, K. Kobayashi, T. Kusakizako, T. Nishizawa, O. Nureki, Structure of the miniature type V-F CRISPR-Cas effector enzyme. *Mol. Cell*. **81**, 558–570 (2021).

134. S. Y. Lee, H. U. Kim, T. U. Chae, J. S. Cho, J. W. Kim, J. H. Shin, D. I. Kim, Y.-S. Ko, W. D. Jang, Y.-S. Jang, A comprehensive metabolic map for production of bio-based chemicals. *Nat. Catal.* **2**, 18–33 (2019).

135. Z. Dai, J. Nielsen, Advancing metabolic engineering through systems biology of industrial microorganisms. *Curr. Opin. Biotechnol.* **36**, 8–15 (2015).

136. K.-K. Hong, J. Nielsen, Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. *Cell. Mol. Life Sci.* **69**, 2671–2690 (2012).

137. P. D. Donohoue, R. Barrangou, A. P. May, Advances in Industrial Biotechnology Using CRISPR-Cas Systems. *Trends Biotechnol.* **36**, 134–146 (2018).

138. K. R. Choi, W. D. Jang, D. Yang, J. S. Cho, D. Park, S. Y. Lee, Systems Metabolic Engineering Strategies: Integrating Systems and Synthetic Biology with Metabolic Engineering. *Trends Biotechnol.* **37**, 817–837 (2019).

139. K. S. Makarova, E. V. Koonin, (Humana Press, New York, NY, 2015; http://link.springer.com/10.1007/978-1-4939-2687-9_4), pp. 47–75.

140. K. S. Makarova, Y. I. Wolf, O. S. Alkhnbashi, F. Costa, S. A. Shah, S. J. Saunders, R. Barrangou, S. J. J. Brouns, E. Charpentier, D. H. Haft, P. Horvath, S. Moineau, F. J. M. Mojica, R. M. Terns, M. P. Terns, M. F. White, A. F. Yakunin, R. A. Garrett, J. van der Oost, R. Backofen, E. V. Koonin, An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).

141. S. Shmakov, O. O. Abudayyeh, K. S. Makarova, Y. I. Wolf, J. S. Gootenberg, E. Semenova, L. Minakhin, J. Joung, S. Konermann, K. Severinov, F. Zhang, E. V. Koonin, Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol. Cell*. **60**, 385–397 (2015).

142. K. R. Choi, S. Y. Lee, CRISPR technologies for bacterial systems: Current achievements and future directions. *Biotechnol. Adv.* **34**, 1180–1209 (2016).

143. F. Alberti, C. Corre, Editing streptomycete genomes in the CRISPR/Cas9 age. *Nat. Prod. Rep.* **36**, 1237–1248 (2019).

144. T.-Q. Shi, G.-N. Liu, R.-Y. Ji, K. Shi, P. Song, L.-J. Ren, H. Huang, X.-J. Ji, CRISPR/Cas9-based genome editing of the filamentous fungi: the state of the art. *Appl. Microbiol. Biotechnol.* **101**, 7435–7443 (2017).

145. M. I. S. Naduthodi, M. J. Barbosa, J. van der Oost, Progress of CRISPR-Cas Based Genome Editing in Photosynthetic Microbes. *Biotechnol. J.* **13**, 1700591 (2018).

146. R. Ferreira, F. David, J. Nielsen, Advancing biotechnology with CRISPR/Cas9: recent applications and patent landscape. *J. Ind. Microbiol. Biotechnol.* **45**, 467–480 (2018).

147. G. Gasiunas, R. Barrangou, P. Horvath, V. Siksnys, Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E2579-86 (2012).

148. H. Deveau, R. Barrangou, J. E. Garneau, J. Labonte, C. Fremaux, P. Boyaval, D. A. Romero, P. Horvath, S. Moineau, Phage Response to CRISPR-Encoded Resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400 (2008).

149. B. Zetsche, M. Heidenreich, P. Mohanraju, I. Fedorova, J. Kneppers, E. M. DeGennaro, N. Winblad, S. R. Choudhury, O. O. Abudayyeh, J. S. Gootenberg, W. Y. Wu, D. A. Scott, K. Severinov, J. van der Oost, F. Zhang, Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35**, 31–34 (2017).

150. D. C. Swarts, M. Jinek, Cas9 versus Cas12a/Cpf1: Structure-function comparisons and implications for genome editing. *Wiley Interdiscip. Rev. RNA*. **9**, e1481 (2018).

151. I. Fonfara, H. Richter, M. BratoviÄ, A. Le Rhun, E. Charpentier, M. Bratovic, A. Le Rhun, E. Charpentier, The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*. **532**, 517–521 (2016).

152. J. Strecker, S. Jones, B. Koopal, J. Schmid-Burgk, B. Zetsche, L. Gao, K. S. Makarova, E. V Koonin, F. Zhang, Engineering of CRISPR-Cas12b for human genome editing. *Nat. Commun.* **10** (2019), doi:10.1038/s41467-018-08224-4.

153. J.-J. Liu, N. Orlova, B. L. Oakes, E. Ma, H. B. Spinner, K. L. M. Baney, J. Chuck, D. Tan, G. J. Knott, L. B. Harrington, B. Al-Shayeb, A. Wagner, J. Brötzmann, B. T. Staahl, K. L. Taylor, J. Desmarais, E. Nogales, J. A. Doudna, CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature*. **566**, 218–223 (2019).

154. A. Sfeir, L. S. Symington, Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? **40**, 701–714 (2015).

155. X. Yao, X. Wang, X. Hu, Z. Liu, J. Liu, H. Zhou, X. Shen, Y. Wei, Z. Huang, W. Ying, Y. Wang, Y.-H. Nie, C.-C. Zhang, S. Li, L. Cheng, Q. Wang, Y. Wu, P. Huang, Q. Sun, L. Shi, H. Yang, Homology-mediated end joining-based targeted integration using CRISPR/Cas9. *Cell Res.* **27**, 801–814 (2017).

156. R. Chayot, B. Montagne, D. Mazel, M. Ricchetti, An end-joining repair mechanism in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2141–6 (2010).

157. S. McGovern, S. Baconnais, P. Roblin, P. Nicolas, P. Drevet, H. Simonson, O. Piétrement, J.-B. Charbonnier, E. Le Cam, P. Noirot, F. Lecointe, C-terminal region of bacterial Ku controls DNA bridging, DNA threading and recruitment of DNA ligase D for double strand breaks repair. *Nucleic Acids Res.* **44**, 4785–4806 (2016).

158. R. Bowater, A. J. Doherty, Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. *PLoS Genet.* **2**, e8 (2006).

159. D. D. Nayak, W. W. Metcalf, Cas9-mediated genome editing in the methanogenic archaeon *Methanosarcina acetivorans*. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 2976–2981 (2017).

160. T.-Q. Shi, H. Huang, E. J. Kerkhoven, X.-J. Ji, Advancing metabolic engineering of *Yarrowia lipolytica* using the CRISPR/Cas system. *Appl. Microbiol. Biotechnol.* **102**, 9541–9548 (2018).

161. A. Berlec, K. Škrlec, J. Kocjan, M. Olenic, B. Štrukelj, Single plasmid systems for inducible dual protein expression and for CRISPR-Cas9/CRISPRi gene regulation in lactic acid bacterium *Lactococcus lactis*. *Sci. Rep.* **8**, 1009 (2018).

162. L. S. Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin, W. A. Lim, Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell*. **152**, 1173–1183 (2013).

163. D. Bikard, W. Jiang, P. Samai, A. Hochschild, F. Zhang, L. A. Marraffini, Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* **41**, 7429–7437 (2013).

164. M. K. Jensen, Design principles for nuclease-deficient CRISPR-based transcriptional regulators. *FEMS Yeast Res.* **18**, 39 (2018).

165. J. Lian, M. HamediRad, S. Hu, H. Zhao, Combinatorial metabolic engineering using an orthogonal tri-functional CRISPR system. *Nat. Commun.* **8**, 1688 (2017).

166. 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, L. S. Qi, CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell*. **154**, 442–451 (2013).

167. A. Chavez, J. Scheiman, S. Vora, B. W. Pruitt, M. Tuttle, E. P R Iyer, S. Lin, S. Kiani, C. D. Guzman, D. J. Wiegand, D. Ter-Ovanesyan, J. L. Braff, N. Davidsohn, B. E. Housden, N. Perrimon, R. Weiss, J. Aach, J. J. Collins, G. M. Church, Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods*. **12**, 326–328 (2015).

168. C. Schwartz, N. Curtis, A.-K. Löbs, I. Wheeldon, Multiplexed CRISPR Activation of Cryptic Sugar Metabolism Enables *Yarrowia Lipolytica* Growth on Cellobiose. *Biotechnol. J.* **13**, 1700584 (2018).

169. M. F. La Russa, L. S. Qi, The New State of the Art: Cas9 for Gene Activation and Repression. *Mol. Cell. Biol.* **35**, 3800–9 (2015).

170. W. Jiang, D. Bikard, D. Cox, F. Zhang, L. A. Marraffini, RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).

171. P. Mali, L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville, G. M. Church, RNA-Guided Human Genome Engineering via Cas9. *Science (80-.).* **339**, 823–826 (2013).

172. S. Mertens, B. Gallone, J. Steensels, B. Herrera-Malaver, J. Cortebeek, R. Nolmans, V. Saels, V. K. Vyas, K. J. Verstrepen, V. K. Vyas Id, K. J. Verstrepen Id, Reducing phenolic off-flavors through CRISPR-based gene editing of the FDC1 gene in *Saccharomyces cerevisiae* x *Saccharomyces eubayanus* hybrid lager beer yeasts. *PLoS One*. **14**, e0209124 (2019).

173. O. W. Ryan, J. M. Skerker, M. J. Maurer, X. Li, J. C. Tsai, S. Poddar, M. E. Lee, W. DeLoache, J. E. Dueber, A. P. Arkin, J. H. Cate, Selection of chromosomal DNA libraries using a multiplex CRISPR system. *Elife*. **3**, e03703 (2014).

174. X. Feng, D. Zhao, X. Zhang, X. Ding, C. Bi, CRISPR/Cas9 Assisted Multiplex Genome Editing Technique in *Escherichia coli*. *Biotechnol. J.* **13**, 1700604 (2018).

175. Y. Jiang, B. Chen, C. Duan, B. Sun, J. Yang, S. Yang, Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* **81**, 2506–14 (2015).

176. C. Ronda, L. E. Pedersen, M. O. A. Sommer, A. T. Nielsen, CRMAGE: CRISPR Optimized MAGE Recombineering. *Sci. Rep.* **6**, 19452 (2016).

177. Y. Li, Z. Lin, C. Huang, Y. Zhang, Z. Wang, Y. Tang, T. Chen, X. Zhao, Metabolic engineering of *Escherichia coli* using CRISPR–Cas9 meditated genome editing. *Metab. Eng.* **31**, 13–21 (2015).

178. R. E. Cobb, Y. Wang, H. Zhao, High-efficiency multiplex genome editing of *Streptomyces species* using an engineered CRISPR/Cas system. *ACS Synth. Biol.* **4**, 723–728 (2015).

179. H. Huang, G. Zheng, W. Jiang, H. Hu, Y. Lu, One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. *Acta Biochim. Biophys. Sin. (Shanghai).* **47**, 231–243 (2015).

180. X. Ao, Y. Yao, T. Li, T.-T. Yang, X. Dong, Z.-T. Zheng, G.-Q. Chen, Q. Wu, Y. Guo, A Multiplex Genome Editing Method for *Escherichia coli* Based on CRISPR-Cas12a. *Front. Microbiol.* **9** (2018), doi:10.3389/fmicb.2018.02307.

181. L. Li, K. Wei, G. Zheng, X. Liu, S. Chen, W. Jiang, Y. Lu, CRISPR-Cpf1-Assisted Multiplex Genome Editing and Transcriptional Repression in *Streptomyces*. *Appl. Environ. Microbiol.* **84**, e00827-18 (2018).

182. W. C. Generoso, M. Gottardi, M. Oreb, E. Boles, Simplified CRISPR-Cas genome editing for *Saccharomyces cerevisiae*. *J Microbiol Methods*. **127**, 203–205 (2016).

183. Z.-H. Li, F.-Q. Wang, D.-Z. Wei, Self-cloning CRISPR/Cpf1 facilitated genome editing in *Saccharomyces cerevisiae*. *Bioresour. Bioprocess.* **5**, 36 (2018).

184. M. E. Lee, W. C. DeLoache, B. Cervantes, J. E. Dueber, A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.* **4**, 975–986 (2015).

185. C. Ronda, J. Maury, T. Jakociunas, S. A. Jacobsen, S. M. Germann, S. J. Harrison, I. Borodina, J. D. Keasling, M. K. Jensen, A. T. Nielsen, T. Jakočiūnas, S. A. Baallal Jacobsen, S. M. Germann, S. J. Harrison, I. Borodina, J. D. Keasling, M. K. Jensen, A. T. Nielsen, CrEdit: CRISPR mediated multi-loci gene integration in *Saccharomyces cerevisiae*. *Microb Cell Fact*. **14**, 97 (2015).

186. M. Wijsman, M. A. Świat, W. L. Marques, J. K. Hettinga, M. van den Broek, P. de la Torre Cortés, R. Mans, J. T. Pronk, J.-M. M. Daran, P. Daran-Lapujade, M. A. Swiat, W. L. Marques, J. K. Hettinga, M. van den Broek, P. Torre Cortes, R. Mans, J. T. Pronk, J.-M. M. Daran, P. Daran-Lapujade, A toolkit for rapid CRISPR-SpCas9 assisted construction of hexose-transport-deficient Saccharomyces cerevisiae strains. *FEMS Yeast Res*. **19**, foy107 (2019).

187. L. Wang, A. Deng, Y. Zhang, S. Liu, Y. Liang, H. Bai, D. Cui, Q. Qiu, X. Shang, Z. Yang, X. He, T. Wen, Efficient CRISPR-Cas9 mediated multiplex genome editing in yeasts. *Biotechnol Biofuels*. **11**, 277 (2018).

188. C. Pohl, J. A. Kiel, A. J. Driessen, R. A. Bovenberg, Y. Nygard, CRISPR/Cas9 Based Genome Editing of *Penicillium chrysogenum*. *ACS Synth Biol*. **5**, 754–764 (2016).

189. C. S. Nodvig, J. B. Hoof, M. E. Kogle, Z. D. Jarczynska, J. Lehmbeck, D. K. Klitgaard, U. H. Mortensen, Efficient oligo nucleotide mediated CRISPR-Cas9 gene editing in *Aspergilli*. *Fungal Genet Biol*. **115**, 78–89 (2018).

190. R. Liu, L. Chen, Y. Jiang, Z. Zhou, G. Zou, Efficient genome editing in filamentous fungus *Trichoderma reesei* using the CRISPR/Cas9 system. *Cell Discov*. **1**, 15007 (2015).

191. M. Cao, M. Gao, D. Ploessl, C. Song, Z. Shao, CRISPR-Mediated Genome Editing and Gene Repression in *Scheffersomyces stipitis*. *Biotechnol J*. **13**, e1700598 (2018).

192. Q. Liu, R. Gao, J. Li, L. Lin, J. Zhao, W. Sun, C. Tian, Development of a genome-editing CRISPR/Cas9 system in thermophilic fungal Myceliophthora species and its application to hyper-cellulase production strain engineering. *Biotechnol Biofuels*. **10**, 1–14 (2017).

193. A. R. G. Gorter de Vries, P. A. de Groot, M. van den Broek, J. G. M. G. Daran, A. R. Gorter de Vries, P. A. de Groot, M. van den Broek, J. G. M. G. Daran, CRISPR-Cas9 mediated gene deletions in lager yeast *Saccharomyces pastorianus*. *Microb Cell Fact*. **16**, 222 (2017).

194. A. Weninger, A.-M. Hatzl, C. Schmid, T. Vogl, A. Glieder, Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast Pichia pastoris. *J. Biotechnol.* **235**, 139–149 (2016).

195. M. Serif, G. Dubois, A.-L. Finoux, M.-A. Teste, D. Jallet, F. Daboussi, One-step generation of multiple gene knock-outs in the diatom *Phaeodactylum tricornutum* by DNA-free genome editing. *Nat. Commun.* **9**, 3924 (2018).

196. A. J. Foster, M. Martin-Urdiroz, X. Yan, H. S. Wright, D. M. Soanes, N. J. Talbot, CRISPR-Cas9 ribonucleoprotein-mediated co-editing and counterselection in the rice blast fungus. *Sci. Rep.* **8**, 14355 (2018).

197. T.-Q. Q. Shi, J. Gao, W.-J. J. Wang, K.-F. F. Wang, G.-Q. Q. Xu, H. Huang, X.-J. J. Ji, CRISPR/Cas9-Based Genome Editing in the Filamentous Fungus *Fusarium fujikuroi* and Its Application in Strain Engineering for Gibberellic Acid Production. *ACS Synth. Biol.* **8**, 445–454 (2019).

198. T. Si, R. Chao, Y. Min, Y. Wu, W. Ren, H. Zhao, Automated multiplex genome-scale engineering in yeast. *Nat. Commun.* **8** (2017), doi:10.1038/ncomms15187.

199. C. M. Schwartz, M. S. Hussain, M. Blenner, I. Wheeldon, Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR–Cas9-Mediated Genome Editing in *Yarrowia lipolytica*. *ACS Synth. Biol.* **5**, 356–359 (2016).

200. L. Song, J.-P. Ouedraogo, M. Kolbusz, T. T. M. Nguyen, A. Tsang, Efficient genome editing using tRNA promoter-driven CRISPR/Cas9 gRNA in *Aspergillus niger*. *PLoS One*. **13**, e0202868 (2018).

201. Y. Gao, 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.* **56**, 343–349 (2014).

202. C. S. Nødvig, J. B. Nielsen, M. E. Kogle, U. H. Mortensen, C. S. Nodvig, J. B. Nielsen, M. E. Kogle, U. H. Mortensen, A CRISPR-Cas9 System for Genetic Engineering of Filamentous Fungi. *PLoS One*. **10**, e0133085 (2015).

203. L. Wong, J. Engel, E. Jin, B. Holdridge, P. Xu, YaliBricks, a versatile genetic toolkit for streamlined and rapid pathway engineering in *Yarrowia lipolytica*. *Metab. Eng. Commun.* **5**, 68–77 (2017).

204. N. J. Morse, J. M. Wagner, K. B. Reed, M. R. Gopal, L. H. Lauffer, H. S. Alper, T7 Polymerase Expression of Guide RNAs in vivo Allows Exportable CRISPR-Cas9 Editing in Multiple Yeast Hosts. *ACS Synth. Biol.* **7**, 1075–1084 (2018).

205. C. R. Reisch, K. L. J. Prather, The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in *Escherichia coli*. *Sci. Rep.* **5**, 15096 (2015).

206. C. Pohl, L. Mozsik, A. J. M. Driessen, R. A. L. Bovenberg, Y. I. Nygard, Genome Editing in *Penicillium chrysogenum* Using Cas9 Ribonucleoprotein Particles. *Methods Mol Biol*. **1772**, 213–232 (2018).

207. J. Behler, D. Vijay, W. R. Hess, M. K. Akhtar, CRISPR-Based Technologies for Metabolic Engineering in Cyanobacteria. *Trends Biotechnol.* **36**, 996–1010 (2018).

208. L. Qi, R. E. Haurwitz, W. Shao, J. A. Doudna, A. P. Arkin, RNA processing enables predictable programming of gene expression. *Nat. Biotechnol.* **30**, 1002–1006 (2012).

209. F. Song, K. Stieger, Optimizing the DNA Donor Template for Homology-Directed Repair of Double-Strand Breaks. *Mol. Ther. Nucleic Acids*. **7**, 53–60 (2017).

210. C. D. Richardson, G. J. Ray, M. A. DeWitt, G. L. Curie, J. E. Corn, Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344 (2016).

211. A. D. Garst, M. C. Bassalo, G. Pines, S. A. Lynch, A. L. Halweg-Edwards, R. Liu, L. Liang, Z. Wang, R. Zeitoun, W. G. Alexander, R. T. Gill, Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat Biotechnol*. **35**, 48–55 (2017).

212. K. R. Roy, J. D. Smith, S. C. Vonesch, G. Lin, C. S. Tu, A. R. Lederer, A. Chu, S. Suresh, M. Nguyen, J. Horecka, A. Tripathi, W. T. Burnett, M. A. Morgan, J. Schulz, K. M. Orsley, W. Wei, R. S. Aiyar, R. W. Davis, V. A. Bankaitis, J. E. Haber, M. L. Salit, R. P. St Onge, L. M. Steinmetz, Multiplexed precision genome editing with trackable genomic barcodes in yeast. *Nat Biotechnol*. **36**, 512–520 (2018).

213. N. G. A. Kuijpers, S. Chroumpi, T. Vos, D. Solis-Escalante, L. Bosman, J. T. Pronk, J. M. Daran, P. Daran-Lapujade, One-step assembly and targeted integration of multigene constructs assisted by the I-SceI meganuclease in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **13**, 769–781 (2013).

214. J.-H. Oh, J.-P. van Pijkeren, CRISPR–Cas9-assisted recombineering in *Lactobacillus reuteri*. *Nucleic Acids Res.* **42**, e131–e131 (2014).

215. T. Aparicio, V. de Lorenzo, E. Martínez-García, CRISPR/Cas9-Based Counterselection Boosts Recombineering Efficiency in *Pseudomonas putida*. *Biotechnol. J.* **13**, 1700161 (2018).

216. I. Mougiakos, E. F. Bosma, J. Ganguly, J. van der Oost, R. van Kranenburg, Hijacking CRISPR-Cas for high-throughput bacterial metabolic engineering: advances and prospects. *Curr. Opin. Biotechnol.* **50**, 146–157 (2018).

217. V. K. Vyas, G. G. Bushkin, D. A. Bernstein, M. A. Getz, M. Sewastianik, M. I. Barrasa, D. P. Bartel, G. R. Fink, New CRISPR Mutagenesis Strategies Reveal Variation in Repair Mechanisms among Fungi. *mSphere*. **3** (2018), doi:10.1128/mSphere.00154-18.

218. C. Gao, S. Wang, G. Hu, L. Guo, X. Chen, P. Xu, L. Liu, Engineering *Escherichia coli* for malate production by integrating modular pathway characterization with CRISPRi-guided multiplexed metabolic tuning. *Biotechnol. Bioeng.* **115**, 661–672 (2018).

219. X. Zhang, J. Wang, Q. Cheng, X. Zheng, G. Zhao, J. Wang, Multiplex gene regulation by CRISPR-ddCpf1. *Cell Discov.* **3**, 17018 (2017).

220. Y. Zhao, L. Li, G. Zheng, W. Jiang, Z. Deng, Z. Wang, Y. Lu, CRISPR/dCas9-Mediated Multiplex Gene Repression in *Streptomyces*. *Biotechnol. J.* **13**, 1800121 (2018).

221. S. Z. Tan, C. R. Reisch, K. L. J. Prather, A Robust CRISPR Interference Gene Repression System in *Pseudomonas*. *J. Bacteriol.* **200**, e00575-17 (2018).

222. Y. Wu, T. Chen, Y. Liu, X. Lv, J. Li, G. Du, R. Ledesma-Amaro, L. Liu, CRISPRi allows optimal temporal control of N-acetylglucosamine bioproduction by a dynamic coordination of glucose and xylose metabolism in *Bacillus subtilis*. *Metab. Eng.* **49**, 232–241 (2018).

223. L. Yao, I. Cengic, J. Anfelt, E. P. Hudson, Multiple Gene Repression in Cyanobacteria Using CRISPRi. *ACS Synth. Biol.* **5**, 207–212 (2016).

224. D. Kaczmarzyk, I. Cengic, L. Yao, E. P. Hudson, Diversion of the long-chain acyl-ACP pool in *Synechocystis* to fatty alcohols through CRISPRi repression of the essential phosphate acyltransferase PlsX. *Metab. Eng.* **45**, 59–66 (2018).

225. C. Schwartz, K. Frogue, A. Ramesh, J. Misa, I. Wheeldon, CRISPRi repression of nonhomologous end-joining for enhanced genome engineering via homologous recombination in *Yarrowia lipolytica*. *Biotechnol. Bioeng.* **114**, 2896–2906 (2017).

226. J.-L. Zhang, Y.-Z. Peng, D. Liu, H. Liu, Y.-X. Cao, B.-Z. Li, C. Li, Y.-J. Yuan, Gene repression via multiplex gRNA strategy in *Y. lipolytica*. *Microb. Cell Fact.* **17**, 62 (2018).

227. A.-K. Löbs, C. Schwartz, S. Thorwall, I. Wheeldon, Highly Multiplexed CRISPRi Repression of Respiratory Functions Enhances Mitochondrial Localized Ethyl Acetate Biosynthesis in *Kluyveromyces marxianus*. *ACS Synth. Biol.* **7**, 2647–2655 (2018).

228. C. Liao, F. Ttofali, R. A. Slotkowski, S. R. Denny, T. D. Cecil, R. T. Leenay, A. J. Keung, C. L. Beisel, Modular one-pot assembly of CRISPR arrays enables library generation and reveals factors influencing crRNA biogenesis. *Nat. Commun.* **10** (2019), doi:10.1038/s41467-019-10747-3.

229. M. Wang, Y. Mao, Y. Lu, X. Tao, J.-K. Zhu, Multiplex Gene Editing in Rice Using the CRISPR-Cpf1 System. *Mol. Plant*. **10**, 1011–1013 (2017).

230. C. Dong, J. Fontana, A. Patel, J. M. Carothers, J. G. Zalatan, Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. *Nat. Commun.* **9**, 2489 (2018).

231. J. G. Zalatan, M. E. Lee, R. Almeida, L. A. Gilbert, E. H. Whitehead, M. La Russa, J. C. Tsai, J. S. Weissman, J. E. Dueber, L. S. Qi, W. A. Lim, Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. *Cell*. **160**, 339–350 (2015).

232. M. Deaner, J. Mejia, H. S. Alper, Enabling Graded and Large-Scale Multiplex of Desired Genes Using a Dual-Mode dCas9 Activator in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* **6**, 1931–1943 (2017).

233. R. Chari, P. Mali, M. Moosburner, G. M. Church, Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat. Methods*. **12**, 823–826 (2015).

234. M. Labuhn, F. F. Adams, M. Ng, S. Knoess, A. Schambach, E. M. Charpentier, A. Schwarzer, J. L. Mateo, J.-H. Klusmann, D. Heckl, Refined sgRNA efficacy prediction improves large- and small-scale CRISPR–Cas9 applications. *Nucleic Acids Res.* **46**, 1375–1385 (2018).

235. M. A. Moreno-Mateos, C. E. Vejnar, J.-D. Beaudoin, J. P. Fernandez, E. K. Mis, M. K. Khokha, A. J. Giraldez, CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat. Methods*. **12**, 982–988 (2015).

236. D. B. Graham, D. E. Root, Resources for the design of CRISPR gene editing experiments. *Genome Biol.* **16**, 260 (2015).

237. G. Zhu, S. Wang, Z. Huang, S. Zhang, Q. Liao, C. Zhang, T. Lin, M. Qin, M. Peng, C. Yang, X. Cao, X. Han, X. Wang, E. van der Knaap, Z. Zhang, X. Cui, H. Klee, A. R. Fernie, J. Luo, S. Huang, Rewiring of the Fruit Metabolome in Tomato Breeding. *Cell*. **172**, 249-261.e12 (2018).

238. K. Lee, V. A. Mackley, A. Rao, A. T. Chong, M. A. Dewitt, J. E. Corn, N. Murthy, Synthetically modified guide RNA and donor DNA are a versatile platform for CRISPR-Cas9 engineering. *Elife*. **6** (2017), doi:10.7554/eLife.25312.

239. B. F. Cress, Ö. D. Toparlak, S. Guleria, M. Lebovich, J. T. Stieglitz, J. A. Englaender, J. A. Jones, R. J. Linhardt, M. A. G. Koffas, CRISPathBrick: Modular Combinatorial Assembly of Type II-A CRISPR Arrays for dCas9-Mediated Multiplex Transcriptional Repression in *E. coli*. *ACS Synth. Biol.* **4**, 987–1000 (2015).

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

241. B. P. Kleinstiver, A. A. Sousa, R. T. Walton, Y. E. Tak, J. Y. Hsu, K. Clement, M. M. Welch, J. E. Horng, J. Malagon-Lopez, I. Scarfò, M. V. Maus, L. Pinello, M. J. Aryee, J. K. Joung, Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat. Biotechnol.* **37**, 276–282 (2019).

242. J. H. Hu, S. M. Miller, M. H. Geurts, W. Tang, L. Chen, N. Sun, C. M. Zeina, X. Gao, H. A. Rees, Z. Lin, D. R. Liu, Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*. **556**, 57–63 (2018).

243. D. Burstein, L. B. Harrington, S. C. Strutt, A. J. Probst, K. Anantharaman, B. C. Thomas, J. A. Doudna, J. F. Banfield, New CRISPR–Cas systems from uncultivated microbes. *Nature*. **542**, 237–241 (2017).

244. J. Kweon, A.-H. Jang, D.-E. Kim, J. W. Yang, M. Yoon, H. Rim Shin, J.-S. Kim, Y. Kim, Fusion guide RNAs for orthogonal gene manipulation with Cas9 and Cpf1. *Nat. Commun.* **8**, 1723 (2017).

245. M. Breinig, A. Y. Schweitzer, A. M. Herianto, S. Revia, L. Schaefer, L. Wendler, A. Cobos Galvez, D. F. Tschaharganeh, Multiplexed orthogonal genome editing and transcriptional activation by Cas12a. *Nat. Methods*. **16**, 51–54 (2019).

246. A. Eid, S. Alshareef, M. M. Mahfouz, CRISPR base editors: genome editing without double-stranded breaks. *Biochem. J.* **475**, 1955–1964 (2018).

247. W. Y. Wu, J. H. G. Lebbink, R. Kanaar, N. Geijsen, J. Van Der Oost, Genome editing by natural and engineered CRISPR-associated nucleases. *Nat. Chem. Biol.* **14**, 642–651 (2018).

248. V. Marx, Base editing a CRISPR way. *Nat. Methods*. **15**, 767–770 (2018).

249. Y. Wang, Y. Liu, J. Liu, Y. Guo, L. Fan, X. Ni, X. Zheng, M. Wang, P. Zheng, J. Sun, Y. Ma, MACBETH: Multiplex automated *Corynebacterium glutamicum* base editing method. *Metab. Eng.* **47**, 200–210 (2018).

250. S. Banno, K. Nishida, T. Arazoe, H. Mitsunobu, A. Kondo, Deaminase-mediated multiplex genome editing in *Escherichia coli*. *Nat. Microbiol.* **3**, 423–429 (2018).

251. T. Arazoe, A. Kondo, K. Nishida, Targeted Nucleotide Editing Technologies for Microbial Metabolic Engineering. *Biotechnol. J.* **13**, 1700596 (2018).

252. J. van der Oost, E. R. Westra, R. N. Jackson, B. Wiedenheft, Unravelling the structural and mechanistic basis of CRISPR–Cas systems. *Nat. Publ. Gr.* **12**, 479–492 (2014).

253. G. J. Knott, J. A. Doudna, CRISPR-Cas guides the future of genetic engineering. *Science*. **361**, 866–869 (2018).

254. P. Mohanraju, K. S. Makarova, B. Zetsche, F. Zhang, E. V. Koonin, J. Van Der Oost, Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science (80-.).* **353** (2016), doi:10.1126/science.aad5147.

255. K. S. Makarova, Y. I. Wolf, J. Iranzo, S. A. Shmakov, O. S. Alkhnbashi, S. J. J. Brouns, E. Charpentier, D. Cheng, D. H. Haft, P. Horvath, S. Moineau, F. J. M. Mojica, D. Scott, S. A. Shah, V. Siksnys, M. P. Terns, Č. Venclovas, M. F. White, A. F. Yakunin, W. Yan, F. Zhang, R. A. Garrett, R. Backofen, J. van der Oost, R. Barrangou, E. V. Koonin, Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).

256. E. Schunder, K. Rydzewski, R. Grunow, K. Heuner, First indication for a functional CRISPR/Cas system in *Francisella tularensis*. *Int. J. Med. Microbiol.* **303**, 51–60 (2013).

257. T. Karvelis, G. Bigelyte, J. K. Young, Z. Hou, R. Zedaveinyte, K. Budre, S. Paulraj, V. Djukanovic, S. Gasior, A. Silanskas, Č. Venclovas, V. Siksnys, PAM recognition by miniature CRISPR-Cas12f nucleases triggers programmable double-stranded DNA target cleavage. *Nucleic Acids Res.* **48**, 5016–5023 (2020).

258. J. Strecker, A. Ladha, Z. Gardner, J. L. Schmid-Burgk, K. S. Makarova, E. V. Koonin, F. Zhang, RNA-guided DNA insertion with CRISPR-associated transposases. *Science (80-.).* **364**, 48–53 (2019).

259. P. Pausch, B. Al-Shayeb, E. Bisom-Rapp, C. A. Tsuchida, Z. Li, B. F. Cress, G. J. Knott, S. E. Jacobsen, J. F. Banfield, J. A. Doudna, Crispr-casF from huge phages is a hypercompact genome editor. *Science (80-.).* **369**, 333–337 (2020).

260. J. S. Chen, E. Ma, L. B. Harrington, M. Da Costa, X. Tian, J. M. Palefsky, J. A. Doudna, CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science (80-.).* **360**, 436–439 (2018).

261. H. Altae-Tran, S. Kannan, F. E. Demircioglu, R. Oshiro, S. P. Nety, L. J. McKay, M. Dlakić, W. P. Inskeep, K. S. Makarova, R. K. Macrae, E. V. Koonin, F. Zhang, The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science (80-.).* **374**, 57–65 (2021).

262. T. Karvelis, G. Druteika, G. Bigelyte, K. Budre, R. Zedaveinyte, A. Silanskas, D. Kazlauskas, Č. Venclovas, V. Siksnys, R. Im Ante Zedaveinyte, A. Silanskas, D. Kazlauskas, Č. Venclovas, V. Siksnys, R. Zedaveinyte, A. Silanskas, D. Kazlauskas, Č. Venclovas, V. Siksnys, Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature*. **599**, 692–696 (2021).

263. C. Liao, C. L. Beisel, The tracrRNA in CRISPR Biology and Technologies. *Annu. Rev. Genet.* **55**, 161–181 (2021).

264. T. yuan Chyou, C. M. Brown, Prediction and diversity of tracrRNAs from type II CRISPR-Cas systems. *RNA Biol.* **16**, 423–434 (2019).

265. A. East-Seletsky, M. R. O'Connell, S. C. Knight, D. Burstein, J. H. D. Cate, R. Tjian, J. A. Doudna, Two Distinct RNase Activities of CRISPR-C2c2 Enable Guide RNA Processing and RNA Detection. *Nature*. **21**, 129–139 (2017).

266. A. Biswas, R. H. J. Staals, S. E. Morales, P. C. Fineran, C. M. Brown, CRISPRDetect: A flexible algorithm to define CRISPR arrays. *BMC Genomics*. **17**, 1–14 (2016).

267. M. J. Soto-Giron, L. M. Rodriguez-R, C. Luo, M. Elk, H. Ryu, J. Hoelle, J. W. Santo Domingo, K. T. Konstantinidis, Biofilms on hospital shower hoses: Characterization and implications for nosocomial infections. *Appl. Environ. Microbiol.* **82**, 2872–2883 (2016).

268. F. J. M. Mojica, C. Díez-Villaseñor, J. García-Martínez, C. Almendros, Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology.* **155**, 733–740 (2009).

269. A. Biswas, J. N. Gagnon, S. J. J. Brouns, P. C. Fineran, C. M. Brown, CRISPRTarget: Bioinformatic prediction and analysis of crRNA targets. *RNA Biol.* **10**, 817–827 (2013).

270. R. T. Leenay, K. R. Maksimchuk, R. A. Slotkowski, R. N. Agrawal, A. A. Gomaa, A. E. Briner, R. Barrangou, C. L. Beisel, Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Mol. Cell*. **62**, 137–147 (2016).

271. T. Su, F. Liu, Y. Chang, Q. Guo, J. Wang, Q. Wang, Q. Qi, The phage T4 DNA ligase mediates bacterial chromosome DSBs repair as single component non-homologous end joining. *Synth. Syst. Biotechnol.* **4**, 107–112 (2019).

272. H. Yang, P. Gao, K. R. Rajashankar, D. J. Patel, PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. *Cell*. **167**, 1814-1828.e12 (2016).

273. R. Marshall, C. S. Maxwell, S. P. Collins, T. Jacobsen, M. L. Luo, M. B. Begemann, B. N. Gray, E. January, A. Singer, Y. He, C. L. Beisel, V. Noireaux, Rapid and Scalable Characterization of CRISPR Technologies Using an E. coli Cell-Free Transcription-Translation System. *Mol. Cell.* **69**, 146-157.e3 (2018).

274. S. C. A. Creutzburg, W. Y. Wu, P. Mohanraju, T. Swartjes, F. Alkan, J. Gorodkin, R. H. J. Staals, J. van der Oost, Good guide, bad guide: spacer sequence-dependent cleavage efficiency of Cas12a. *Nucleic Acids Res.* **48**, 3228–3243 (2020).

275. D. Bikard, L. A. Marraffini, Control of gene expression by CRISPR-Cas systems. *F1000Prime Rep.* **5**, 1–8 (2013).

276. B. P. Kleinstiver, S. Q. Tsai, M. S. Prew, N. T. Nguyen, M. M. Welch, J. M. Lopez, Z. R. Mccaw, M. J. Aryee, J. K. Joung, J. K. J. Conceived, N. B. Author, Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells HHS Public Access Author manuscript. *Nat Biotechnol*. **34**, 869–874 (2016).

277. E. Semenova, M. M. Jore, K. A. Datsenko, A. Semenova, E. R. Westra, B. Wanner, J. Van Der Oost, S. J. J. Brouns, K. Severinov, Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 10098–10103 (2011).

278. S. Sugiura, S. Ohkubo, K. Yamaguchi, Minimal essential origin of plasmid pSC101 replication: Requirement of a region downstream of iterons. *J. Bacteriol.* **175**, 5993–6001 (1993).

279. J. Wiktor, C. Lesterlin, D. J. Sherratt, C. Dekker, CRISPR-mediated control of the bacterial initiation of replication. *Nucleic Acids Res.* **44**, 3801–3810 (2016).

280. E. Ma, L. B. Harrington, M. R. O'Connell, K. Zhou, J. A. Doudna, Single-Stranded DNA Cleavage by Divergent CRISPR-Cas9 Enzymes. *Mol. Cell*. **60**, 398–407 (2015).

281. Y. Zhang, R. Rajan, H. S. Seifert, A. Mondragón, E. J. Sontheimer, DNase H Activity of *Neisseria meningitidis* Cas9. *Mol. Cell*. **60**, 242–255 (2015).

282. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*. **533**, 420–424 (2016).

283. X. Li, Y. Wang, Y. Liu, B. Yang, X. Wang, J. Wei, Z. Lu, Y. Zhang, J. Wu, X. Huang, L. Yang, J. Chen, Base editing with a Cpf1–cytidine deaminase fusion. *Nat. Biotechnol. 2018 364*. **36**, 324–327 (2018).

284. L. Casalino, Ł. Nierzwicki, M. Jinek, G. Palermo, Catalytic Mechanism of Non-Target DNA Cleavage in CRISPR-Cas9 Revealed by Ab Initio Molecular Dynamics. *ACS Catal.* **10**, 13596–13605 (2020).

285. G. Sheng, H. Zhao, J. Wang, Y. Rao, W. Tian, D. C. Swarts, J. Van Der Oost, D. J. Patel, Y. Wang, Structure-based cleavage mechanism of Thermusthermophilus argonaute DNA guide strand-mediated DNA target cleavage. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 652–657 (2014).

286. I. M. Slaymaker, L. Gao, B. Zetsche, D. A. Scott, W. X. Yan, F. Zhang, Rationally engineered Cas9 nucleases with improved specificity. *Science (80-.).* **351**, 84–88 (2016).

287. M. Saito, A. Ladha, J. Strecker, G. Faure, E. Neumann, H. Altae-Tran, R. K. Macrae, F. Zhang, Dual modes of CRISPR-associated transposon homing. *Cell*. **184**, 2441-2453.e18 (2021).

288. C. J. Huang, B. A. Adler, J. A. Doudna, *bioRxiv*, in press, doi:10.1101/2021.12.06.471469.

289. C. A. Lino, J. C. Harper, J. P. Carney, J. A. Timlin, Delivering crispr: A review of the challenges and approaches. *Drug Deliv.* **25**, 1234–1257 (2018).

290. Z. Wu, H. Yang, P. Colosi, Effect of genome size on AAV vector packaging. *Mol. Ther.* **18**, 80–86 (2010).

291. X. Xu, A. Chemparathy, L. Zeng, H. R. Kempton, S. Shang, M. Nakamura, L. S. Qi, Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. *Mol. Cell*. **81**, 4333–4345 (2021).

292. D. J. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

293. S. Lu, J. Wang, F. Chitsaz, M. K. Derbyshire, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, G. H. Marchler, J. S. Song, N. Thanki, R. A. Yamashita, M. Yang, D. Zhang, C. Zheng, C. J. Lanczycki, A. Marchler-Bauer, CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Res.* **48**, D265–D268 (2020).

294. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

295. D. Couvin, A. Bernheim, C. Toffano-Nioche, M. Touchon, J. Michalik, B. Néron, E. P. C. Rocha, G. Vergnaud, D. Gautheret, C. Pourcel, CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* **46**, W246–W251 (2018).

296. S. A. Shmakov, Y. I. Wolf, E. Savitskaya, K. V. Severinov, E. V. Koonin, Mapping CRISPR spaceromes reveals vast host-specific viromes of prokaryotes. *Commun. Biol.* **3**, 1–9 (2020).

297. P. Mohanraju, J. Oost, M. Jinek, D. Swarts, Heterologous Expression and Purification of the CRISPR-Cas12a/Cpf1 Protein. *Bio-protocol*. **8**, e2842 (2018).

298. K. G. Wandera, S. P. Collins, F. Wimmer, R. Marshall, V. Noireaux, C. L. Beisel, An enhanced assay to characterize anti-CRISPR proteins using a cell-free transcription-translation system. *Methods*. **172**, 42–50 (2020).

299. J. Söding, Protein homology detection by HMM-HMM comparison. *Bioinformatics*. **21**, 951–960 (2005).

300. W. Dawson, K. Fujiwara, G. Kawai, Y. Futamura, K. Yamamoto, A method for finding optimal RNA secondary structures using a new entropy model (vsfold). *Nucleosides, Nucleotides and Nucleic Acids*. **25**, 171–189 (2006).

301. M. G. Kluesner, D. A. Nedveck, W. S. Lahr, J. R. Garbe, J. E. Abrahante, B. R. Webber, B. S. Moriarity, EditR: A Method to Quantify Base Editing from Sanger Sequencing. *Cris. J.* **1**, 239–250 (2018).

302. Y. Ji, B. Zhang, S. F. Van Horn, P. Warren, G. Woodnutt, M. K. R. Burnham, M. Rosenberg, Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science (80-.).* **293**, 2266–2269 (2001).

303. A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, C. C. Mello, Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nat. 1998 3916669*. **391**, 806–811 (1998).

304. S. K. Kim, W. Seong, G. H. Han, D. H. Lee, S. G. Lee, CRISPR interference-guided multiplex repression of endogenous competing pathway genes for redirecting metabolic flux in *Escherichia coli*. *Microb. Cell Fact.* **16**, 1–15 (2017).

305. J. Alonso-Gutierrez, R. Chan, T. S. Batth, P. D. Adams, J. D. Keasling, C. J. Petzold, T. S. Lee, Metabolic engineering of *Escherichia coli* for limonene and perillyl alcohol production. *Metab. Eng.* **19**, 33–41 (2013).

306. M. Kampmann, CRISPRi and CRISPRa screens in mammalian cells for precision biology and medicine HHS Public Access. *ACS Chem Biol*. **13**, 406–416 (2018).

307. J. M. Peters, M. R. Silvis, D. Zhao, J. S. Hawkins, C. A. Gross, L. S. Qi, Bacterial CRISPR: Accomplishments and Prospects. *Curr. Opin. Microbiol.* **27**, 121 (2015).

308. E. Cámara, I. Lenitz, Y. Nygård, A CRISPR activation and interference toolkit for industrial *Saccharomyces cerevisiae* strain KE6-12. *Sci. Reports 2020 101*. **10**, 1–13 (2020).

309. M. Deaner, H. S. Alper, Systematic testing of enzyme perturbation sensitivities via graded dCas9 modulation in *Saccharomyces cerevisiae*. *Metab. Eng.* **40**, 14–22 (2017).

310. A. Chavez, M. Tuttle, B. W. Pruitt, B. Ewen-Campen, R. Chari, D. Ter-Ovanesyan, S. J. Haque, R. J. Cecchi, E. J. K. Kowal, J. Buchthal, B. E. Housden, N. Perrimon, J. J. Collins, G. Church, Comparison of Cas9 activators in multiple species. *Nat. Methods 2016 137*. **13**, 563–567 (2016).

311. M. A. Horlbeck, L. A. Gilbert, J. E. Villalta, B. Adamson, R. A. Pak, Y. Chen, A. P. Fields, C. Y. Park, J. E. Corn, M. Kampmann, J. S. Weissman, Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife*. **5** (2016), doi:10.7554/ELIFE.19760.

312. L. A. Gilbert, M. A. Horlbeck, B. Adamson, J. E. Villalta, Y. Chen, E. H. Whitehead, C. Guimaraes, B. Panning, H. L. Ploegh, M. C. Bassik, L. S. Qi, M. Kampmann, J. S. Weissman, Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*. **159**, 647–661 (2014).

313. A. Momen-Roknabadi, P. Oikonomou, M. Zegans, S. Tavazoie, An inducible CRISPR interference library for genetic interrogation of *Saccharomyces cerevisiae* biology. *Commun. Biol. 2020 31*. **3**, 1–12 (2020).

314. J. Lian, C. Schultz, M. Cao, M. HamediRad, H. Zhao, Multi-functional genome-wide CRISPR system for high throughput genotype–phenotype mapping. *Nat. Commun. 2019 101*. **10**, 1–10 (2019).

315. F. A. Ran, L. Cong, W. X. Yan, D. A. Scott, J. S. Gootenberg, A. J. Kriz, B. Zetsche, O. Shalem, X. Wu, K. S. Makarova, E. V. Koonin, P. A. Sharp, F. Zhang, In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature*. **520**, 186–191 (2015).

316. W. Y. Wu, P. Mohanraju, C. Liao, B. Adiego-Pérez, S. C. A. Creutzburg, K. S. Makarova, K. Keessen, T. A. Lindeboom, T. S. Khan, S. H. P. Prinsen, R. Joosten, W. X. Yan, A. Migur, C. Laffeber, D. A. Scott, J. H. G. Lebbink, E. Koonin, C. L. Beisel, J. van der Oost, The Miniature CRISPR-Cas12m Effector Binds DNA To Block Transcription. *SSRN Electron. J.* (2021), doi:10.2139/SSRN.3991079.

317. W. Y. Wu, Right tool for the right job: Exploring the diversity of type V CRISPR-Cas systems (2021).

318. N. Schreiber-Agus, L. Chin, K. Chen, R. Torres, G. Rao, P. Guida, A. I. Skoultchi, R. A. DePinho, An amino-terminal domain of Mxi1 mediates anti-myc oncogenic activity and interacts with a homolog of the Yeast Transcriptional Repressor SIN3. *Cell*. **80**, 777–786 (1995).

319. S. Cho, D. Choe, E. Lee, S. C. Kim, B. Palsson, B. K. Cho, High-Level dCas9 Expression Induces Abnormal Cell Morphology in *Escherichia coli*. *ACS Synth. Biol.* **7**, 1085–1094 (2018).

320. H. H. Lee, N. Ostrov, B. G. Wong, M. A. Gold, A. S. Khalil, G. M. Church, Functional genomics of the rapidly replicating bacterium *Vibrio natriegens* by CRISPRi. *Nat. Microbiol. 2019 47*. **4**, 1105–1113 (2019).

321. W. Jiang, A. J. Brueggeman, K. M. Horken, T. M. Plucinak, D. P. Weeks, Successful transient expression of Cas9 and single guide RNA genes in Chlamydomonas reinhardtii. *Eukaryot. Cell*. **13**, 1465–1469 (2014).

322. J. D. Smith, S. Suresh, U. Schlecht, M. Wu, O. Wagih, G. Peltz, R. W. Davis, L. M. Steinmetz, L. Parts, R. P. St.Onge, Quantitative CRISPR interference screens in yeast identify chemical-genetic interactions and new rules for guide RNA design. *Genome Biol.* **17** (2016), doi:10.1186/S13059-016-0900-9.

323. M. W. Gander, J. D. Vrana, W. E. Voje, J. M. Carothers, E. Klavins, Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. *Nat. Commun. 2017 81*. **8**, 1–11 (2017).

324. F. Farzadfard, S. D. Perli, T. K. Lu, Tunable and Multifunctional Eukaryotic Transcription Factors Based on CRISPR/Cas. *ACS Synth. Biol.* **2**, 604 (2013).

325. K. J. A. Martens, S. P. B. van Beljouw, S. van der Els, J. N. A. Vink, S. Baas, G. A. Vogelaar, S. J. J. Brouns, P. van Baarlen, M. Kleerebezem, J. Hohlbein, Visualisation of dCas9 target search in vivo using an open-microscopy framework. *Nat. Commun.* **10**, 1–11 (2019).

326. H. Ma, L. C. Tu, A. Naseri, M. Huisman, S. Zhang, D. Grunwald, T. Pederson, CRISPR-Cas9 nuclear dynamics and target recognition in living cells. *J. Cell Biol.* **214**, 529–537 (2016).

327. C. Verduyn, E. Postma, W. A. Scheffers, J. P. Van Dijken, Effect of benzoic acid on metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast*. **8**, 501–517 (1992).

328. J. T. Pronk, Auxotrophic yeast strains in fundamental and applied research. **68**, 2095–2100 (2002).

329. K. Labun, T. G. Montague, M. Krause, Y. N. Torres Cleuren, H. Tjeldnes, E. Valen, CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.* **47**, W171–W174 (2019).

330. J. McMillan, Z. Lu, J. S. Rodriguez, T.-H. Ahn, Z. Lin, YeasTSS: an integrative web database of yeast transcription start sites. *Database J. Biol. Databases Curation*. **2019**, baz048 (2019).

331. F. D. Urnov, E. J. Rebar, M. C. Holmes, H. S. Zhang, P. D. Gregory, Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **11**, 636–646 (2010).

332. T. Gaj, C. A. Gersbach, C. F. Barbas, ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31** (2013), pp. 397–405.

333. M. Jinek, A. East, A. Cheng, S. Lin, E. Ma, J. Doudna, RNA-programmed genome editing in human cells. *Elife*. **2**, e00471 (2013).

334. M. Finger-Bou, E. Orsi, J. van der Oost, R. H. J. Staals, CRISPR with a Happy Ending: Non-Templated DNA Repair for Prokaryotic Genome Engineering. *Biotechnol. J.* **15** (2020), p. 1900404.

335. C. Bertrand, A. Thibessard, C. Bruand, F. Lecointe, P. Leblond, Bacterial NHEJ: a never ending story. *Mol. Microbiol.* **111** (2019), pp. 1139–1151.

336. N. Hustedt, D. Durocher, The control of DNA repair by the cell cycle. *Nat. Cell Biol.* **19** (2017), pp. 1–9.

337. M. Shrivastav, L. P. De Haro, J. A. Nickoloff, Regulation of DNA double-strand break repair pathway choice. *Cell Res.* **18**, 134–147 (2008).

338. K. C. Murphy, Use of bacteriophage λ recombination functions to promote gene replacement in *Escherichia coli*. *J. Bacteriol.* **180**, 2063–2071 (1998).

339. N. M. Gaudelli, A. C. Komor, H. A. Rees, M. S. Packer, A. H. Badran, D. I. Bryson, D. R. Liu, Programmable base editing of T to G C in genomic DNA without DNA cleavage. *Nature*. **551**, 464–471 (2017).

340. T. L. Cheng, S. Li, B. Yuan, X. Wang, W. Zhou, Z. Qiu, Expanding C–T base editing toolkit with diversified cytidine deaminases. *Nat. Commun.* **10**, 1–10 (2019).

341. H. E. Krokan, M. Bjørås, Base Excision Repair. *Cold Spring Harb. Perspect. Biol.* **5**, a012583 (2013).

342. L. Wang, W. Xue, L. Yan, X. Li, J. Wei, M. Chen, J. Wu, B. Yang, L. Yang, J. Chen, Enhanced base editing by co-expression of free uracil DNA glycosylase inhibitor. *Cell Res. 2017 2710*. **27**, 1289–1292 (2017).

343. H. A. Rees, D. R. Liu, Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).

344. M. F. Richter, K. T. Zhao, E. Eton, A. Lapinaite, G. A. Newby, B. W. Thuronyi, C. Wilson, L. W. Koblan, J. Zeng, D. E. Bauer, J. A. Doudna, D. R. Liu, Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol. 2020 387*. **38**, 883–891 (2020).

345. K. S. Makarova, Y. I. Wolf, E. V. Koonin, Classification and Nomenclature of CRISPR-Cas Systems : Where from Here ? *Cris. J.* **1**, 325–336 (2018).

346. B. Tong, H. Dong, Y. Cui, P. Jiang, Z. Jin, D. Zhang, The Versatile Type V CRISPR Effectors and Their Application Prospects. *Front. Cell Dev. Biol.* **8** (2021), , doi:10.3389/fcell.2020.622103.

347. V. Bharathi, A. Girdhar, A. Prasad, M. Verma, V. Taneja, B. K. Patel, Use of ade1 and ade2 mutations for development of a versatile red/white colour assay of amyloid-induced oxidative stress in *Saccharomyces cerevisiae*. *Yeast*. **33**, 607–620 (2016).

348. C. Dong, L. Jiang, S. Xu, L. Huang, J. Cai, J. Lian, Z. Xu, A Single Cas9-VPR Nuclease for Simultaneous Gene Activation, Repression, and Editing in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* **9**, 2252–2257 (2020).

349. V. Stovicek, I. Borodina, J. Forster, CRISPR–Cas system enables fast and simple genome editing of industrial Saccharomyces cerevisiae strains. *Metab. Eng. Commun.* **2**, 13–22 (2015).

350. R. Ferreira, thesis, Chalmers Tekniska Hogskola (Sweden), Gothenburg (2019).

351. P. C. Després, A. K. Dubé, L. Nielly-Thibault, N. Yachie, C. R. Landry, Double selection enhances the efficiency of target-AID and Cas9-based genome editing in yeast. *G3 Genes, Genomes, Genet.* **8**, 3163–3171 (2018).

352. J. Tan, F. Zhang, D. Karcher, R. Bock, Engineering of high-precision base editors for site-specific single nucleotide replacement. *Nat. Commun. 2019 101*. **10**, 1–10 (2019).

353. S.-J. Bae, B. G. Park, B.-G. Kim, J.-S. Hahn, Multiplex Gene Disruption by Targeted Base Editing of *Yarrowia lipolytica* Genome Using Cytidine Deaminase Combined with the CRISPR/Cas9 System. *Biotechnol. J.* **15**, 1900238 (2020).

354. X. Jing, B. Xie, L. Chen, N. Zhang, Y. Jiang, H. Qin, H. Wang, P. Hao, S. Yang, X. Li, Implementation of the CRISPR-Cas13a system in fission yeast and its repurposing for precise RNA editing. *Nucleic Acids Res.* **46**, 90 (2018).

355. S. Chen, Y. Jia, Z. Liu, H. Shan, M. Chen, H. Yu, L. Lai, Z. Li, Robustly improved base editing efficiency of Cpf1 base editor using optimized cytidine deaminases. *Cell Discov. 2020 61*. **6**, 1–4 (2020).

356. P. C. Després, A. K. Dubé, M. Seki, N. Yachie, C. R. Landry, Perturbing proteomes at single residue resolution using base editing. *Nat. Commun.* **11** (2020), doi:10.1038/s41467-020-15796-7.

357. R. D. Gietz, R. A. Woods, Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* **350**, 87–96 (2002).

358. K. Entian, P. S. Kötte, 25 yeast genetic strain and plasmid collections. *Methods Microbiol.* **36**, 629–666 (2007).

359. V. F. Holmes, K. R. Benjamin, N. J. Crisona, N. R. Cozzarelli, Bypass of heterology during strand transfer by *Saccharomyces cerevisiae* Rad51 protein. *Nucleic Acids Res.* **29**, 5052 (2001).

360. A. Paix, A. Folkmann, D. H. Goldman, H. Kulaga, M. J. Grzelak, D. Rasoloson, S. Paidemarry, R. Green, R. R. Reed, G. Seydoux, Precision genome editing using synthesis-dependent repair of Cas9-induced DNA breaks. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E10745–E10754 (2017).

361. J. R. Chapman, M. R. G. Taylor, S. J. Boulton, Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. *Mol. Cell*. **47**, 497–510 (2012).

362. K. M. Vasquez, K. Marburger, Z. Intody, J. H. Wilson, Manipulating the mammalian genome by homologous recombination. *Proc. Natl. Acad. Sci.* **98**, 8403–8410 (2001).

363. M. Lieberman-Lazarovich, A. A. Levy, Homologous Recombination in Plants: An Antireview. *Methods Mol. Biol.* **701**, 51–65 (2011).

364. T. Gutschner, M. Haemmerle, G. Genovese, G. F. Draetta, L. Chin, Post-translational Regulation of Cas9 during G1 Enhances Homology-Directed Repair. *Cell Rep.* **14**, 1555–1566 (2016).

365. M. E. Pyne, M. Moo-Young, D. A. Chung, C. P. Chou, Coupling the CRISPR/Cas9 System with Lambda Red Recombineering Enables Simplified Chromosomal Gene Replacement in *Escherichia coli*. *Appl. Environ. Microbiol.* **81**, 5103–5114 (2015).

366. B. Su, D. Song, H. Zhu, Homology-dependent recombination of large synthetic pathways into *E. coli* genome via λ-Red and CRISPR/Cas9 dependent selection methodology. *Microb. Cell Fact.* **19**, 1–11 (2020).

367. J. Sun, Q. Wang, Y. Jiang, Z. Wen, L. Yang, J. Wu, S. Yang, Genome editing and transcriptional repression in *Pseudomonas putida* KT2440 via the type II CRISPR system. *Microb. Cell Fact.* **17**, 1–17 (2018).

368. Y. Zhou, L. Lin, H. Wang, Z. Zhang, J. Zhou, N. Jiao, Development of a CRISPR/Cas9n-based tool for metabolic engineering of *Pseudomonas putida* for ferulic acid-to-polyhydroxyalkanoate bioconversion. *Commun. Biol. 2020 31*. **3**, 1–13 (2020).

369. C. Patinios, S. C. A. Creutzburg, A. Q. Arifah, B. Adiego-Pérez, E. A. Gyimah, C. J. Ingham, S. W. M. Kengen, J. van der Oost, R. H. J. Staals, Streamlined CRISPR genome engineering in wild-type bacteria using SIBR-Cas. *Nucleic Acids Res.* **49**, 11392–11404 (2021).

370. T. Maruyama, S. K. Dougan, M. C. Truttmann, A. M. Bilate, J. R. Ingram, H. L. Ploegh, Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* **33**, 538–542 (2015).

371. J. H. Choo, C. Han, J. Y. Kim, H. A. Kang, Deletion of a *KU80* homolog enhances homologous recombination in the thermotolerant yeast Kluyveromyces marxianus. *Biotechnol. Lett.* **36**, 2059–2067 (2014).

372. A. S. Rajkumar, J. A. Varela, H. Juergens, J. M. G. Daran, J. P. Morrissey, Biological parts for *Kluyveromyces marxianus* synthetic biology. *Front. Bioeng. Biotechnol.* **7**, 97 (2019).

373. A. Kretzschmar, C. Otto, M. Holz, S. Werner, L. Hübner, G. Barth, Increased homologous integration frequency in *Yarrowia lipolytica* strains defective in non-homologous end-joining. *Curr. Genet.* **59**, 63–72 (2013).

374. B. E. Hew, R. Sato, D. Mauro, I. Stoytchev, J. B. Owens, RNA-guided piggyBac transposition in human cells. *Synth. Biol.* **4**, ysz018 (2019).

375. J. E. Peters, N. L. Craig, Tn7: smarter than we thought. *Nat. Rev. Mol. Cell Biol. 2001 211.* **2**, 806–814 (2001).

376. M. Rogers, N. Ekaterinaki, E. Nimmo, D. Sherratt, Analysis of Tn 7 transposition. *Mol. Gen. Genet. MGG.* **205**, 550–556 (1986).

377. K. Y. Choi, Y. Li, R. Sarnovsky, N. L. Craig, Direct interaction between the TnsA and TnsB subunits controls the heteromeric Tn7 transposase. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E2038–E2045 (2013).

378. R. J. Sarnovsky, E. W. May, N. L. Craig, The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. *EMBO J.* **15**, 6348–6361 (1996).

379. E. W. May, N. L. Craig, Switching from Cut-and-Paste to Replicative Tn7 Transposition. *Science (80-.).* **272**, 401–404 (1996).

380. P. Gamas, N. L. Craig, Purification and characterization of TnsC, a Tn7 transposition protein that binds ATP and DNA. *Nucleic Acids Res.* **20**, 2525–2532 (1992).

381. A. E. Stellwagen, N. L. Craig, Analysis of gain-of-function mutants of an ATP-dependent regulator of Tn7 transposition. *J. Mol. Biol.* **305**, 633–642 (2001).

382. K. Y. Choi, J. M. Spencer, N. L. Craig, The Tn7 transposition regulator TnsC interacts with the transposase subunit TnsB and target selector TnsD. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E2858–E2865 (2014).

383. C. S. Waddell, N. L. Craig, Tn7 transposition: two transposition pathways directed by five Tn7-encoded genes. *Genes Dev.* **2**, 137–149 (1988).

384. J. E. Peters, K. S. Makarova, S. Shmakov, E. V. Koonin, Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E7358–E7366 (2017).

385. G. Faure, S. A. Shmakov, W. X. Yan, D. R. Cheng, D. A. Scott, J. E. Peters, K. S. Makarova, E. V. Koonin, CRISPR–Cas in mobile genetic elements: counter-defence and beyond. *Nat. Rev. Microbiol. 2019 178*. **17**, 513–525 (2019).

386. P. A. Rice, N. L. Craig, F. Dyda, Comment on “RNA-guided DNA insertion with CRISPR-associated transposases.” *Science*. **368** (2020), doi:10.1126/science.abb2022.

387. S.-C. Hsieh, J. E. Peters, *bioRxiv*, in press, doi:10.1101/2021.02.06.429022.

388. J. Strecker, A. Ladha, K. S. Makarova, E. V. Koonin, F. Zhang, Response to Comment on “RNA-guided DNA insertion with CRISPR-associated transposases.” *Science*. **368** (2020), doi:10.1126/science.abb2920.

389. G. Y. Kholodii, O. V. Yurieva, O. L. Lomovskaya, Z. M. Gorlenko, S. Z. Mindlin, V. G. Nikiforov, Tn5053, a Mercury Resistance Transposon with Integron’s Ends. *J. Mol. Biol.* **230**, 1103–1107 (1993).

390. G. Y. Kholodii, S. Z. Mindlin, I. A. Bass, O. V Yurieva, S. V Minakhina, V. G. Nikiforov, Four genes, two ends, and a res region are involved in transposition of Tn5053: a paradigm for a novel family of transposons carrying either a mer operon or an integron. **17**, 1189–1200 (1995).

391. P. L. H. Vo, C. Acree, M. L. Smith, S. H. Sternberg, Unbiased profiling of CRISPR RNA-guided transposition products by long-read sequencing. *Mob. DNA*. **12**, 4–11 (2021).

392. P. L. H. Vo, C. Ronda, S. E. Klompe, E. E. Chen, C. Acree, H. H. Wang, S. H. Sternberg, CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering. *Nat. Biotechnol.* **39**, 480–489 (2021).

393. B. E. Rubin, S. Diamond, B. F. Cress, A. Crits-Christoph, Y. C. Lou, A. L. Borges, H. Shivram, C. He, M. Xu, Z. Zhou, S. J. Smith, R. Rovinsky, D. C. J. Smock, K. Tang, T. K. Owens, N. Krishnappa, R. Sachdeva, R. Barrangou, A. M. Deutschbauer, J. F. Banfield, J. A. Doudna, Species- and site-specific genome editing in complex bacterial communities. *Nat. Microbiol.* **7**, 34–47 (2022).

394. J. Lu, T. Wu, B. Zhang, S. Liu, W. Song, J. Qiao, H. Ruan, Types of nuclear localization signals and mechanisms of protein import into the nucleus. *Cell Commun. Signal.* **19**, 1–10 (2021).

395. J. P. S. Makkerh, C. Dingwall, R. A. Laskey, Comparative mutagenesis of nuclear localization signals reveals the importance of neutral and acidic amino acids. *Curr. Biol.* **6**, 1025–1027 (1996).

396. E. M. Young, Z. Zhao, B. E. M. Gielesen, L. Wu, D. B. Gordon, J. A. Roubos, C. A. Voigt, Iterative algorithm-guided design of massive strain libraries , applied to itaconic acid production in yeast. *Metab. Eng.* **48**, 33–43 (2018).

397. G. Bigelyte, J. K. Young, T. Karvelis, K. Budre, R. Zedaveinyte, V. Djukanovic, E. Van Ginkel, S. Paulraj, S. Gasior, S. Jones, L. Feigenbutz, G. S. Clair, P. Barone, J. Bohn, A. Acharya, G. Zastrow-Hayes, S. Henkel-Heinecke, A. Silanskas, R. Seidel, V. Siksnys, Miniature type V-F CRISPR-Cas nucleases enable targeted DNA modification in cells. *Nat. Commun.* **12**, 1–8 (2021).

398. M. I. S. Naduthodi, P. Mohanraju, C. Südfeld, S. D’Adamo, M. J. Barbosa, J. Van Der Oost, CRISPR-Cas ribonucleoprotein mediated homology-directed repair for efficient targeted genome editing in microalgae *Nannochloropsis oceanica* IMET1. *Biotechnol. Biofuels*. **12**, 1–11 (2019).

399. E. Roggenkamp, R. M. Giersch, M. N. Schrock, E. Turnquist, M. Halloran, G. C. Finnigan, Tuning CRISPR-Cas9 gene drives in *Saccharomyces cerevisiae*. *G3 Genes, Genomes, Genet.* **8**, 999–1018 (2018).

400. A. I. Garrido-godino, F. Gutiérrez-santiago, F. Navarro, Biogenesis of RNA Polymerases in Yeast. **8**, 1–10 (2021).

401. P. Cameron, M. M. Coons, S. E. Klompe, A. M. Lied, S. C. Smith, B. Vidal, P. D. Donohoue, T. Rotstein, B. W. Kohrs, D. B. Nyer, R. Kennedy, L. M. Banh, C. Williams, M. S. Toh, M. J. Irby, L. S. Edwards, C. H. Lin, A. L. G. Owen, T. Künne, J. van der Oost, S. J. J. Brouns, E. M. Slorach, C. K. Fuller, S. Gradia, S. B. Kanner, A. P. May, S. H. Sternberg, Harnessing type I CRISPR–Cas systems for genome engineering in human cells. *Nat. Biotechnol. 2019 3712*. **37**, 1471–1477 (2019).

402. J. K. Young, S. L. Gasior, S. Jones, L. Wang, P. Navarro, B. Vickroy, R. Barrangou, The repurposing of type I-E CRISPR-Cascade for gene activation in plants. *Commun. Biol.* **2**, 383 (2019).

403. P. Liu, K. Luk, M. Shin, F. Idrizi, S. Kwok, B. Roscoe, E. Mintzer, S. Suresh, K. Morrison, J. B. Frazão, M. F. Bolukbasi, K. Ponnienselvan, J. Luban, L. J. Zhu, N. D. Lawson, S. A. Wolfe, Enhanced Cas12a editing in mammalian cells and zebrafish. *Nucleic Acids Res.* **47**, 4169–4180 (2019).

404. J. U. Park, A. W. L. Tsai, E. Mehrotra, M. T. Petassi, S. C. Hsieh, A. Ke, J. E. Peters, E. H. Kellogg, Structural basis for target site selection in RNA-guided DNA transposition systems. *Science (80-.).* **373**, 768–774 (2021).

405. I. Querques, M. Schmitz, S. Oberli, C. Chanez, M. Jinek, M. Schmitz, Target site selection and remodelling by type V CRISPR-transposon systems. *Nat. 2021 5997885*. **599**, 497–502 (2021).

406. A. B. Hickman, Y. Li, S. V. Mathew, E. W. May, N. L. Craig, F. Dyda, Unexpected Structural Diversity in DNA Recombination: The Restriction Endonuclease Connection. *Mol. Cell*. **5**, 1025–1034 (2000).

407. C. J. Tou, B. Orr, B. P. Kleinstiver, *bioRxiv*, in press, doi:10.1101/2022.01.07.475005.

408. R. Xiao, S. Wang, R. Han, Z. Li, C. Gabel, I. A. Mukherjee, L. Chang, Structural basis of target DNA recognition by CRISPR-Cas12k for RNA-guided DNA transposition. *Mol. Cell*. **81**, 4457-4466.e5 (2021).

409. S. C. Dillon, C. J. Dorman, Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol. 2010 83*. **8**, 185–195 (2010).

410. A. V. Anzalone, X. D. Gao, C. J. Podracky, A. T. Nelson, L. W. Koblan, A. Raguram, J. M. Levy, J. A. M. Mercer, D. R. Liu, Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. *Nat. Biotechnol.* **40**, 731–740 (2022).

411. R. Green, E. J. Rogers, Chemical Transformation of *E. coli*. *Methods Enzymol.* **529**, 329 (2013).

412. K. AU - Ciurkot, B. AU - Vonk, T. E. AU - Gorochowski, J. A. AU - Roubos, R. AU - Verwaal, CRISPR/Cas12a Multiplex Genome Editing of *Saccharomyces cerevisiae* and the Creation of Yeast Pixel Art. *JoVE*. **28**, e59350 (2019).

413. M. P. Lacerda, E. J. Oh, C. Eckert, The Model System *Saccharomyces cerevisiae* Versus Emerging Non-Model Yeasts for the Production of Biofuels. *Life*. **10**, 1–20 (2020).

414. S. K. Nandy, R. K. Srivastava, A review on sustainable yeast biotechnological processes and applications. *Microbiol. Res.* **207**, 83–90 (2018).

415. J. A. Varela, L. Gethins, C. Stanton, P. Ross, J. P. Morrissey, in *Yeast Diversity in Human Welfare*, T. Satyanarayana, G. Kunze, Eds. (Springer Singapore, Singapore, 2017; https://doi.org/10.1007/978-981-10-2621-8_17), pp. 439–453.

416. A. Karim, N. Gerliani, M. Aïder, *Kluyveromyces marxianus*: An emerging yeast cell factory for applications in food and biotechnology. *Int. J. Food Microbiol.* **333**, 108818 (2020).

417. M. M. Lane, J. P. Morrissey, *Kluyveromyces marxianus*: A yeast emerging from its sister's shadow. *Fungal Biol. Rev.* **24**, 17–26 (2010).

418. C. H. Emerson, A. A. Bertuch, Consider the workhorse: Nonhomologous end joining in budding yeast. *Biochem. Cell Biol.* **94**, 396–406 (2016).

419. J. M. Daley, P. L. Palmbos, D. Wu, T. E. Wilson, Nonhomologous end joining in yeast. *Annu. Rev. Genet.* **39**, 431–451 (2005).

420. E. Shor, C. A. Fox, J. R. Broach, The Yeast Environmental Stress Response Regulates Mutagenesis Induced by Proteotoxic Stress. *PLOS Genet.* **9**, e1003680 (2013).

421. A.-K. Löbs, R. Engel, C. Schwartz, A. Flores, I. Wheeldon, CRISPR-Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in *Kluyveromyces marxianus*. *Biotechnol Biofuels*. **10**, 164 (2017).

422. K. M. Davis, V. Pattanayak, D. B. Thompson, J. A. Zuris, D. R. Liu, Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat. Chem. Biol.* **11**, 316–318 (2015).

423. K. I. Liu, M. N. Bin Ramli, C. W. A. Woo, Y. Wang, T. Zhao, X. Zhang, G. R. D. Yim, B. Y. Chong, A. Gowher, M. Z. H. Chua, J. Jung, J. H. J. Lee, M. H. Tan, A chemical-inducible CRISPR-Cas9 system for rapid control of genome editing. *Nat. Chem. Biol.* **12**, 980–987 (2016).

424. D. J. J. Truong, K. Kühner, R. Kühn, S. Werfel, S. Engelhardt, W. Wurst, O. Ortiz, Development of an intein-mediated split-Cas9 system for gene therapy. *Nucleic Acids Res.* **43**, 6450–6458 (2015).

425. Y. Nihongaki, F. Kawano, T. Nakajima, M. Sato, Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat. Biotechnol.* **33**, 755–760 (2015).

426. B. Zetsche, S. E. Volz, F. Zhang, A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* **33**, 139–142 (2015).

427. Y. Wang, Y. Liu, F. Xie, J. Lin, L. Xu, Photocontrol of CRISPR/Cas9 function by site-specific chemical modification of guide RNA. *Chem. Sci.* **11**, 11478–11484 (2020).

428. P. K. Jain, V. Ramanan, A. G. Schepers, N. S. Dalvie, A. Panda, H. E. Fleming, S. N. Bhatia, Development of Light-Activated CRISPR Using Guide RNAs with Photocleavable Protectors. *Angew. Chemie - Int. Ed.* **55**, 12440–12444 (2016).

429. E. V. Moroz-Omori, D. Satyapertiwi, M. C. Ramel, H. Høgset, I. K. Sunyovszki, Z. Liu, J. P. Wojciechowski, Y. Zhang, C. L. Grigsby, L. Brito, L. Bugeon, M. J. Dallman, M. M. Stevens, Photoswitchable gRNAs for Spatiotemporally Controlled CRISPR-Cas-Based Genomic Regulation. *ACS Cent. Sci.* **6**, 695–703 (2020).

430. W. Tang, J. H. Hu, D. R. Liu, Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. *Nat. Commun.* **8**, 1–8 (2017).

431. J. Zhao, R. Inomata, Y. Kato, M. Miyagishi, Development of aptamer-based inhibitors for CRISPR/Cas system. *Nucleic Acids Res.* **49**, 1330–1344 (2021).

432. K. M. Thompson, H. A. Syrett, S. M. Knudsen, A. D. Ellington, Group I aptazymes as genetic regulatory switches. *BMC Biotechnol.* **2**, 1–12 (2002).

433. M. Belfort, J. Pedersen-Lane, K. Ehrenman, F. K. Chu, G. F. Maley, F. Maley, D. S. McPheeters, L. Gold, RNA splicing and in vivo expression of the intron-containing td gene of bacteriophage T4. *Gene*. **41**, 93–102 (1986).

434. H. Ge, M. A. Marchisio, Aptamers, riboswitches and ribozymes in *S. cerevisiae* synthetic biology. *Life*. **11** (2021), p. 248.

435. A. H. Babiskin, C. D. Smolke, Engineering ligand-responsive RNA controllers in yeast through the assembly of RNase III tuning modules. *Nucleic Acids Res.* **39**, 5299–5311 (2011).

436. L. Sandegren, B.-M. Sjöberg, Self-Splicing of the Bacteriophage T4 Group I Introns Requires Efficient Translation of the Pre-mRNA In Vivo and Correlates with the Growth State of the Infected Bacterium. *J. Bacteriol.* **189**, 980–990 (2007).

437. W. S. Jung, S. Shin, I. K. Park, Novobiocin inhibits the self-splicing of the primary transcripts of T4 phage thymidylate synthase gene. *Mol. Cell. Biochem.* **314**, 143–149 (2008).

438. C. Jung, S. Shin, I. K. Park, Pyridoxal phosphate inhibits the group I intron splicing. *Mol. Cell. Biochem.* **280**, 17–23 (2005).

439. C. Waldsich, K. Semrad, R. Schroeder, Neomycin B inhibits splicing of the td intron indirectly by interfering with translation and enhances missplicing in vivo. *Rna*. **4**, 1653–1663 (1998).

440. U. von Ahsen, J. Davies, R. R. Schroeder, U. Von Ahsen, J. Davies, R. R. Schroeder, Non-competitive inhibition of group I intron RNA self-splicing by aminoglycoside antibiotics. *J. Mol. Biol.* **226**, 935–941 (1992).

441. U. Von Ahsen, J. Davies, R. Schroeder, Antibiotic inhibition of group I ribozyme function. *Nature*. **353**, 368–370 (1991).

442. I. K. Park, J. Y. Kim, E. H. Lim, S. Shin, Spectinomycin inhibits the self-splicing of the group 1 intron RNA. *Biochem. Biophys. Res. Commun.* **269**, 574–579 (2000).

443. Y. LIU, R. R. TIDWELL, M. J. LEIBOWITZ, Inhibition of In Vitro Splicing of a Group I Intron of *Pneumocystis carinii*. *J. Eukaryot. Microbiol.* **41**, 31–38 (1994).

444. I. K. Park, Effects of deamido-NAD + on self-splicing of primary transcripts of phage T4 thymidylate synthase gene . *Korean J. Biol. Sci.* **4**, 141–144 (2000).

445. I. K. Park, J. Y. Kim, NAD+ inhibits the self-splicing of the group I intron. *Biochem. Biophys. Res. Commun.* **281**, 206–211 (2001).

446. J. H. Kim, I. K. Park, Inhibition of the group I ribozyme splicing by NADP+. *Mol. Cell. Biochem.* **252**, 285–293 (2003).

447. E. Ford, M. Ares, Synthesis of circular RNA in bacteria and yeast using RNA cyclase ribozymes derived from a group I intron of phage T4. *Proc Natl Acad Sci U S A*. **91**, 3117–3121 (1994).

448. S. Meaux, A. van Hoof, K. E. Baker, Nonsense-Mediated mRNA Decay in Yeast Does Not Require PAB1 or a Poly(A) Tail. *Mol. Cell*. **29**, 134–140 (2008).

449. N. Kuperwasser, S. Brogna, K. Dower, M. Rosbash, Nonsense-mediated decay does not occur within the yeast nucleus. *RNA*. **10**, 1907–1915 (2004).

450. A. Celik, R. Baker, F. He, A. Jacobson, High-resolution profiling of NMD targets in yeast reveals translational fidelity as a basis for substrate selection. *Rna.* **23**, 735–748 (2017).

451. S. W. Peltz, A. H. Brown, A. Jacobson, mRNA destabilization triggered by premature translational termination depends on at least three cis-acting sequence elements and one trans-acting factor. *Genes Dev.* **7**, 1737–1754 (1993).

452. J. P. B. Lloyd, The evolution and diversity of the nonsense-mediated mRNA decay pathway. *F1000Research*. **7**, 1–28 (2018).

453. J. Wen, M. He, M. Petric, L. Marzi, J. Wang, K. Piechocki, T. McLeod, A. Singh, V. Dwivedi, S. Brogna, An intron proximal to a PTC enhances NMD in *Saccharomyces cerevisiae*. *bioRxiv [Preprint]*, 1–33 (2020).

454. C. Patinios, thesis, Wageningen University (2022).

455. S. Brogna, T. McLeod, M. Petric, The Meaning of NMD: Translate or Perish. *Trends Genet.* **32**, 395–407 (2016).

456. P. J. Lopez, B. Séraphin, YIDB: The Yeast Intron DataBase. *Nucleic Acids Res.* **28**, 85–86 (2000).

457. Z. Fatma, J. C. Schultz, H. Zhao, Recent advances in domesticating non-model microorganisms. *Biotechnol. Prog.* **36**, e3008 (2020).

458. A. M. B. Johns, J. Love, S. J. Aves, Four inducible promoters for controlled gene expression in the oleaginous yeast *Rhodotorula toruloides*. *Front. Microbiol.* **7**, 1666 (2016).

459. R. F. Bruns, J. W. Daly, S. H. Snyder, Adenosine receptor binding: Structure-activity analysis generates extremely potent xanthine antagonists. *Proc. Natl. Acad. Sci. U. S. A.* **80**, 2077–2080 (1983).

460. Z. Benko, M. Sipiczki, Caffeine tolerance in *Schizosaccharomyces pombe*: Physiological adaptation and interaction with theophylline. *Can. J. Microbiol.* **39**, 551–554 (1993).

461. A. Sarachek, L. A. Henderson, Recombinagenicity of caffeine for *Candida albicans*. *Mycopathologia*. **110**, 63–76 (1990).

462. J. Pourquie, Antagonism by adenine in the nutrition of *Schizosaccharomyces pombe* mutants Inhibition at the level of guanine uptake. *Biochim. Biophys. Acta - Nucleic Acids Protein Synth.* **209**, 269–277 (1970).

463. L. L. Ruta, I. C. Farcasanu, *Saccharomyces cerevisiae* and caffeine implications on the eukaryotic cell. *Nutrients*. **12**, 1–22 (2020).

464. I. Borodina, J. Nielsen, Advances in metabolic engineering of yeast *Saccharomyces cerevisiae* for production of chemicals. *Biotechnol. J.* **9**, 609–620 (2014).

465. B. Adiego-Pérez, P. Randazzo, J. M. Daran, R. Verwaal, J. A. Roubos, P. Daran-Lapujade, J. Van Der Oost, Multiplex genome editing of microorganisms using CRISPR-Cas. *FEMS Microbiol. Lett.* **366**, 1–19 (2019).

466. P. Randazzo, N. X. Bennis, J. M. Daran, P. Daran-Lapujade, GEL DNA: A Cloning-and Polymerase Chain Reaction-Free Method for CRISPR-Based Multiplexed Genome Editing. *Cris. J.* **4**, 896–913 (2021).

467. T. L. Orr-Weaver, J. W. Szostak, R. J. Rothsteint, R. J. Rothstein, Yeast transformation: A model system for the study of recombination. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 6354–6358 (1981).

468. B. Chaikind, J. L. Bessen, D. B. Thompson, J. H. Hu, D. R. Liu, A programmable Cas9-serine recombinase fusion protein that operates on DNA sequences in mammalian cells. *Nucleic Acids Res.* **44**, 9758–9770 (2016).

469. A. Kovač, C. Miskey, M. Menzel, E. Grueso, A. Gogol-Döring, Z. Ivics, RNA-guided retargeting of *Sleeping Beauty* transposition in human cells. *Elife*. **9** (2020), doi:10.7554/ELIFE.53868.

470. M. G. Durrant, A. Fanton, J. Tycko, M. Hinks, S. S. Chandrasekaran, N. T. Perry, J. Schaepe, P. P. Du, P. Lotfy, M. C. Bassik, L. Bintu, A. S. Bhatt, P. D. Hsu, F. G. Program, I. Bioengineering, *bioRxiv [Preprint]*, in press (available at https://doi.org/10.1101/2021.11.05.467528).

471. H. Nielsen, S. D. Johansen, Group I introns: Moving in new directions. *RNA Biol.* **6** (2009), doi:10.4161/rna.6.4.9334.

472. G. Hausner, M. Hafez, D. R. Edgell, Bacterial group I introns: Mobile RNA catalysts. *Mob. DNA*. **5**, 1–12 (2014).

473. T. Swartjes, P. Shang, D. van den Berg, T. A. Künne, N. Geijsen, S. J. J. Brouns, J. van der Oost, R. H. J. Staals, R. A. Notebaart, *bioRxiv*, in press, doi:10.1101/2022.01.15.475214.

474. 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, E. V. Koonin, Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **9**, 467–477 (2011).

475. N. Kurihara, R. Nakagawa, H. Hirano, S. Okazaki, A. Tomita, K. Kobayashi, T. Kusakizako, T. Nishizawa, K. Yamashita, D. A. Scott, H. Nishimasu, O. Nureki, N. Kurihara, R. Nakagawa, H. Hirano, S. Okazaki, A. Tomita, K. Kobayashi, T. Kusakizako, T. Nishizawa, K. Yamashita, D. A. Scott, H. Nishimasu, O. Nureki, Structure of the type V-C CRISPR-Cas effector enzyme. *Mol. Cell.* **82**, 1–13 (2022).

476. H. Yang, D. J. Patel, CasX: a new and small CRISPR gene-editing protein. *Cell Res.* **29**, 345–346 (2019).

477. Z. X. Zhang, L. R. Wang, Y. S. Xu, W. T. Jiang, T. Q. Shi, X. M. Sun, H. Huang, Recent advances in the application of multiplex genome editing in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* **105**, 3873–3882 (2021).

478. H. Zhang, Z. Li, R. Xiao, L. Chang, Mechanisms for target recognition and cleavage by the Cas12i RNA-guided endonuclease. *Nat. Struct. Mol. Biol.* **27**, 1069–1076 (2020).

479. X. Huang, W. Sun, Z. Cheng, M. Chen, X. Li, J. Wang, G. Sheng, W. Gong, Y. Wang, Structural basis for two metal-ion catalysis of DNA cleavage by Cas12i2. *Nat. Commun.* **11** (2020), doi:10.1038/s41467-020-19072-6.

480. A. Carabias, A. Fuglsang, P. Temperini, T. Pape, N. Sofos, S. Stella, S. Erlendsson, G. Montoya, N. Sofos, Structure of the mini-RNA-guided endonuclease CRISPR-Cas12j3. *Nat. Commun.* **12**, 1–12 (2021).

481. F. Teng, J. Li, T. Cui, K. Xu, L. Guo, Q. Gao, G. Feng, Enhanced mammalian genome editing by new Cas12a orthologs with optimized crRNA scaffolds. **20**, 3–8 (2019).

482. G. Palermo, L. Casalino, M. Jinek, Two-Metal Ion Mechanism of DNA Cleavage in CRISPR-Cas9. *Biophys. J.* **118**, 64a (2020).

483. C. Huai, G. Li, R. Yao, Y. Zhang, M. Cao, L. Kong, C. Jia, H. Yuan, H. Chen, D. Lu, Q. Huang, Structural insights into DNA cleavage activation of CRISPR-Cas9 system. *Nat. Commun.* **8**, 1–9 (2017).

484. H. Tang, H. Yuan, W. Du, G. Li, D. Xue, Q. Huang, Active-Site Models of *Streptococcus pyogenes* Cas9 in DNA Cleavage State. *Front. Mol. Biosci.* **8**, 1–12 (2021).

485. D. C. Swarts, Making the cut(s): How Cas12a cleaves target and non-target DNA. *Biochem. Soc. Trans.* **47**, 1499–1510 (2019).

486. W. X. Yan, S. Chong, H. Zhang, K. S. Makarova, E. V. Koonin, D. R. Cheng, D. A. Scott, Cas13d Is a Compact RNA-Targeting Type VI CRISPR Effector Positively Modulated by a WYL-Domain-Containing Accessory Protein. *Mol. Cell.* **70**, 327-339.e5 (2018).

487. S. Kannan, H. Altae-Tran, X. Jin, V. J. Madigan, R. Oshiro, K. S. Makarova, E. V. Koonin, F. Zhang, Compact RNA editors with small Cas13 proteins. *Nat. Biotechnol.* **40**, 194–197 (2022).

488. B. Zhang, Y. Ye, W. Ye, V. Perčulija, H. Jiang, Y. Chen, Y. Li, J. Chen, J. Lin, S. Wang, Q. Chen, Y. S. Han, S. Ouyang, Two HEPN domains dictate CRISPR RNA maturation and target cleavage in Cas13d. *Nat. Commun.* **10** (2019), doi:10.1038/s41467-019-10507-3.

489. L. Li, S. Li, N. Wu, J. Wu, G. Wang, G. Zhao, J. Wang, HOLMESv2: A CRISPR-Cas12b-Assisted Platform for Nucleic Acid Detection and DNA Methylation Quantitation. *ACS Synth. Biol.* **8**, 2228–2237 (2019).

490. X. Ding, K. Yin, Z. Li, R. V. Lalla, E. Ballesteros, M. M. Sfeir, C. Liu, Ultrasensitive and visual detection of SARS-CoV-2 using all-in-one dual CRISPR-Cas12a assay. *Nat. Commun.* **11**, 1–10 (2020).

491. Z. Wang, C. Zhong, Cas12c-DETECTOR: A specific and sensitive Cas12c-based DNA detection platform. *Int. J. Biol. Macromol.* **193**, 441–449 (2021).

492. R. Xiao, Z. Li, S. Wang, R. Han, L. Chang, Structural basis for substrate recognition and cleavage by the dimerization-dependent CRISPR-Cas12f nuclease. *Nucleic Acids Res.* **49**, 4120–4128 (2021).

493. Y. Zhang, Q. Ren, X. Tang, S. Liu, A. A. Malzahn, J. Zhou, J. Wang, D. Yin, C. Pan, M. Yuan, L. Huang, H. Yang, Y. Zhao, Q. Fang, X. Zheng, L. Tian, Y. Cheng, Y. Le, B. McCoy, L. Franklin, J. D. Selengut, S. M. Mount, Q. Que, Y. Zhang, Y. Qi, Expanding the scope of plant genome engineering with Cas12a orthologs and highly multiplexable editing systems. *Nat. Commun.* **12**, 1–11 (2021).

494. Z. Wu, Y. Zhang, H. Yu, D. Pan, Y. Wang, Y. Wang, F. Li, C. Liu, H. Nan, W. Chen, Q. Ji, Programmed genome editing by a miniature CRISPR-Cas12f nuclease. *Nat. Chem. Biol.* **17**, 1132–1138 (2021).

495. M. B. Begemann, B. N. Gray, E. January, A. Singer, D. C. Kesler, Y. He, H. Liu, H. Guo, A. Jordan, T. P. Brutnell, T. C. Mockler, M. Oufattole, Characterization and Validation of a Novel Group of Type V, Class 2 Nucleases for in vivo Genome Editing. *bioRxiv [Preprint]* (2017), doi:10.1101/192799.

496. S. E. Klompe, P. L. H. Vo, T. S. Halpin-Healy, S. H. Sternberg, Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature*. **571**, 219–225 (2019).

497. A. Jansen, K. J. Verstrepen, Nucleosome Positioning in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **75**, 301–320 (2011).

498. N. Fleck, C. Grundner, A Cas12a-based CRISPR interference system for multigene regulation in mycobacteria. *J. Biol. Chem.* **297**, 100990 (2021).

499. A. Vigouroux, D. Bikard, CRISPR Tools To Control Gene Expression in Bacteria. *Microbiol. Mol. Biol. Rev.* **84** (2020), doi:10.1128/MMBR.00077-19.

500. J. N. A. Vink, K. J. A. Martens, M. Vlot, R. E. McKenzie, C. Almendros, B. Estrada Bonilla, D. J. W. Brocken, J. Hohlbein, S. J. J. Brouns, Direct Visualization of Native CRISPR Target Search in Live Bacteria Reveals Cascade DNA Surveillance Mechanism. *Mol. Cell*. **77**, 39-50.e10 (2020).

ABOUT THE AUTHOR

Belén Adiego Pérez was born on March 29th, 1992 in Barcelona, Spain. She grew up in the coast city Vilanova i la Geltrú. She conducted her BSc studies of Biotechnology at the Universitat de Barcelona (UB, Spain). For her BSc thesis in 2014, she had the opportunity to join the Erasmus program and moved to Ghent, Belgium. There, she found her way in the lab and performed her BSc thesis on metabolic engineering of a sophorolipid production yeast. In 2015, she started her MSc studies in Cellular and Molecular Biotechnology at Wageningen University Research. During her MSc studies, she conducted her MSc thesis at the Bioprocess Engineering Department (BPE) of the same University where she worked on the development of genome editing tools for the bacterium *Rhodobacter sphaeroides*. She conducted her internship at the Yeast Metabolic Engineering of the Technical University of Denmark (DTU) on the establishment of carotenoid biosynthetic pathways in the yeast *Yarrowia lipolytica*. After that, she worked as research assistant in the same group for the production of gibberelins in yeast.

In October 2017, she started her PhD studies in the Bacterial Genetics group under the supervision of Prof Dr John van der Oost. During her PhD studies she enjoyed most of her time in the lab and engaged with many educational activities that woke her mentoring and teaching enthusiasm. In parallel with her PhD studies and following her interest in education and teaching, she joined a part-time MSc program that allowed her to teach Biology in Spanish secondary schools. Currently, she is employed as Research Assistant in the group of Bacterial Genetics of the Laboratory of Microbiology where she participates in multiple research projects and coordinates the activity of the lab.

LIST OF PUBLICATIONS

Wen Y. Wu*, Prarthana Mohanraju*, Chunyu Liao, **Belén Adiego-Pérez**, Sjoerd C. A. Creutzburg, Kira S. Makarova, Karlijn Keessen, Timon A. Lindeboom, Tahseen S. Khan, Stijn Prinsen, Rob Joosten, Winston X. Yan, Anzhela Migur, Charlie Laffeber, David A. Scott, Joyce H.G. Lebbink, Eugene V. Koonin, Chase L. Beisel and John van der Oost. The miniature CRISPR-Cas12m effector binds DNA to block transcription. *(Manuscript submitted for publication)*

Patinios, Constantinos, Sjoerd CA Creutzburg, Adini Q. Arifah, **Belén Adiego-Pérez**, Evans A. Gyimah, Colin J. Ingham, Servé WM Kengen, John van der Oost, and Raymond HJ Staals. Streamlined CRISPR genome engineering in wild-type bacteria using SIBR-Cas. *Nucleic acids research* 49, no. 19 (2021): 11392-11404.

Kildegaard, Kanchana R., Jonathan A. Arnesen, **Belén Adiego-Pérez**, Daniela Rago, Mette Kristensen, Andreas K. Klitgaard, Esben H. Hansen, Jørgen Hansen, and Irina Borodina. Tailored biosynthesis of gibberellin plant hormones in yeast. *Metabolic engineering* 66 (2021): 1-11.

Mougiakos, Ioannis, Enrico Orsi, Mohammad Rifqi Ghiffary, Wilbert Post, Alberto de Maria, **Belén Adiego-Perez**, Servé WM Kengen, Ruud A. Weusthuis, and John van der Oost. Efficient Cas9-based genome editing of Rhodobacter sphaeroides for metabolic engineering. *Microbial cell factories* 18, no. 1 (2019): 1-13.

Belén Adiego-Pérez, Paola Randazzo, Jean Marc Daran, Rene Verwaal, Johannes A. Roubos, Pascale Daran-Lapujade, and John Van Der Oost. Multiplex genome editing of microorganisms using CRISPR-Cas. *FEMS microbiology letters* 366, no. 8 (2019): fnz086.

Kildegaard Kanchana Rueksomtawin, **Belén Adiego-Pérez**, David Doménech Belda, Jaspreet Kaur Khangura, Carina Holkenbrink, and Irina Borodina. Engineering of Yarrowia lipolytica for production of astaxanthin. *Synthetic and systems biotechnology* 2, no. 4 (2017): 287-294.

OVERVIEW OF COMPLETED TRAINING ACTIVITIES

Name of the course/meeting	Organizing institute (s)	Year
	Discipline specific activities	
Genome Editing and synthetic biology	VIB	2018
MIB Centenial Symposium	Laboratory of Microbiology (WUR)	2017
Yeasterday	Delft University of Technology	2018
Yeasterday	Groningen University	2019
Genome Engineering: CRISPR Frontiers	Cold Spring Harbor Laboratory	2020
Consortium meetings (NWO-BBoL)	Delft University/DSM/WUR	2017, 2018, 2019
Training period at laboratory Delft University	Industrial Microbiology Research Section	2020
wAGOningen 2019	Laboratory of Biochemistry (WUR)	2019
FEMS Yeast Research Webinar on Genomic Insights into Yeast Diversity and Evolution, FEMS Microbiology Letters in Educating in a pandemic and beyond	FEMS	2021
5th Applied Synthetic Biology Europe	TUD	2020
1st International BioDesign Research Conference	Stanford University, the University of Warwick, and *BioDesign Research* (BDR)	2020
GE helathcare AKTA introduction	GE Healthcare - WUR	2019
CRISPR meeting	Institut Pasteur	2021
	General courses	
VLAG PhD week	VLAG	2018
PhD competence assessment	WGS	2018
Brain training	WGS	2018
PhD carousel	WGS	2018
Reviewing a Scientific Manuscript	WGS	2019
Infographics and Iconography	WGS	2019
Scientific Writing	WGS	2021
Presenting with Impact	WGS	2021
Bioinformatics with Linux and Python	WIMEK, WUR	2020
Introduction to Python	Datacamp	2020
	Other activities	
Preparation of research proposal	Laboratory of Microbiology (WUR)	2018
Weekly group meetings	Laboratory of Microbiology (WUR)	2018-2022
PhD meetings	Laboratory of Microbiology (WUR)	2018-2022
MSc in teaching	UNIR	2019-2021
Organization of student Thesis Rings	Laboratory of Microbiology (WUR)	2018-2021

ACKNOWLEDGEMENTS

Over the development of this thesis I believe I have grown both as a person and as a scientist. But this could have not been possible without all the great scientists and great people I encountered in these last years and the impact they had on me. In the following paragraphs I would like to thank each of them and apology for the ones I surely miss.

First of all, **John**, thank you for accepting me as a PhD student in the Bacterial Genetics group under your supervision. Although during my process as PhD candidate I have had multiple ups and downs, your door was always open and I appreciated your open attitude. Although I do not consider myself a “yeast expert”, thank you for trusting my previous knowledge and bringing up new questions that allowed me to learn more from the eukaryotic world in a prokaryotic lab. You have taught me the importance of resilience in science and gave me wonderful opportunities to develop myself. I believe what you have built in BacGen is a great example of how science can be both impactful and fun! Your curiosity and interest for fundamental questions are truly inspiring. Thank you for bringing this group together and for allowing me to continue in it some more years. I wish you all the best for you, your family and for the lovely Charlie.

Sarah, you came into this story half-way and I am very happy you did. I approached you without knowing you at all (although Enrico had only good words) in an effort to change the direction of my thesis. I could not thank you more for accepting this adventure and for all your valuable input in the last years. The meetings we had during pandemic times were very important for me and helped me to go through a tough period. Thank you for all your support during the development of the thesis and I hope we can keep collaborating in the future. I wish you all the best in your scientific and personal life.

Raymond, you have also been a very big piece of this thesis, first with your involvement in the SIBR-Cas project but also throughout all my thesis in the meetings with Carina. First of all, thank you for involving me in the project and for valuing my opinion. I value a lot your critical analysis of experimental designs and results and I enjoyed a lot our meetings. I also love your irony, your passion for science and jokes (even when I can be the target of that).

Ruud, I also need to thank you for introducing me to John and for giving me the opportunity to reach out when I was not doing okey. I consider myself very lucky to have had you as Master thesis supervisor. I am very glad we had the opportunity to talk before my defence, I wish you all the best in your personal and professional life. **Nico**, I met you as a PhD candidate and I have seen you growing through your PostDoc and starting your tenure. It has been very funny to participate and cycle together in the Veluweloop. I wish you a brilliant future! **Servé**, I mostly interacted with you through the organization of the thesis ring and the teaching. It has been a pleasure to be involved a bit more in the educational side of Microbiology. **Clara** and **Caroline**, thank you for trusting me organizing RMM for four years. I had a lot of fun preparing and teaching the course.

To my collaborators at TuDelft, **Pascale**, **Paola** and **Jean Marc**, thank you for hosting me at your department and teaching me techniques I introduced in my work. Also thanks for the nice discussions in the meetings and your contribution on the review we wrote together. To **René** and **Hans**, thanks for your input in the meetings and your collaboration during the project. **Klaudia**, it was very nice to work with you in the development of chapter 6. Thank you to the thesis committee: for the time taken to read and evaluate my thesis.

To office 6073, thank you for creating such a nice environment to work and have fun. Although I became a wild card, I still feel part of the office and I enjoy that we keep doing dinners and stuff together. **Janneke**, you are the engine of that office and of many things in MIB, many initiatives would have not taken place without you. You are a very good and determined scientist and I like that we might work a bit together. I hope you enjoyed Copenhagen a lot and come back full of energy for the last stretch of your thesis! Thank you for organizing the Veluweloop and all the events you did throughout the years. I hope you forgive us for giving you as present peanut butter two years in a row (it is completely on you that I got addicted to it though)! **Joep**, we started almost at the same time and shared teaching time, lunches (lots of Friesdays), projects and office time together. Thank you for listening to me many times and I wish you all the best for you, Liza, Palm and Pils. **Thijs**, as you wrote in your thesis, thank you for balancing the fun and the working time in the office. I admire your organization skills and practical approach to stuff. You have also been there since the beginning of my thesis and I am happy you stayed around a bit more after. I wish you all the best for you and Tanya at your new place. **Isma**, va ser genial estar a la mateixa office i poder parlar de tantes coses! Tot i que vam compartir penes digestives, me n’alegro molt que ho hagis pogut solucionar. Et desitjo tot el millor! **Jurre**, you were the last incorporation to the office while I was still there. I admire how determined you are in whatever you do, not only your work, but your cooking activities, everything. I am aware I still need to find a good bread recipe to give you! I wish you and Scope all the best. **Carl**, welcome to the group! It has been very nice to have you around in the last dinners we did! I loved the flan! I hope we can enjoy way more and I wish you all the best in your PhD! Thank you, Office 6073 for being such a strong support!

To office 5032, it was very nice to move to your office half way my PhD. **Despoina**, your enthusiasm and hard work is admirable! You are very energetic and determined and I am sure you will get wherever you wish. It was great to share time with you inside and outside the office, and lots of fun going to Belgium or Terschelling, together! **Sjoerd**, although we shared little time in the office, I’m very thankful you were always open to discuss science and different experimental approaches. **Wen**, you are the soul of BacGen and I am very happy to have collaborated with you. You are a brilliant scientist, mentor and a vitamin person. I believe it took a while to understand each other but you are a wonderful person and I am very thankful you were always there for me. **Maartje**, it was great to have you around. You had a lot of experience

in different fields and a lot of great ideas! It was very nice to have you in the office. **Isabelle**, you are great addition to BacGen and a very caring person. I admire your critical thinking and strong opinions, I wish you a brilliant future!

Special thanks to my paranymphs Max and Thomas. **Max**, com a part de l'office 6073, ha estat genial passar aquests anys al teu costat. Moltes gràcies per escoltar-me tantes i tantes vegades i gràcies per aportar el teu toc d'humor a l'oficina (el Google doc en el que registraves la teva temperatura periòdicament, va superar tots els límits). Gràcies per tots els sopars i chills al teu jardí de Droev. **Thomas**, we shared a lot of teaching together and it has always been a pleasure to work with you. Your sense of humor and use of irony keeps me always alert. Besides, your talent and dedication to learn Spanish is admirable (¡hablas muy bien!). Thanks for organizing office events in the last year.

Carina, thank you for being my yeast colleague and being always there to discuss our problems. You are always full of ideas and initiatives and it is very nice to share the bench with you (sorry for being a bit messy...!). Thank you for caring for me and listening all my problems and ideas. **Anneleen**, it is very nice to work with you. I love all your travelling plans and stories. Looking forward to future projects together! **Hanne**, it is great to keep working with you! Thanks for being always so energetic and caring! **Miguel**, has traído la locura al laboratorio. Junto con **Ricardo**, hacéis un dúo maravilloso. Gracias por el buen rollo a los dos! **Eric**, it was very chill to teach with you, thank you also for helping us with the moving! **Costas**, do you know how did the skeleton know that it was going to rain? Now I will create the suspense by telling you that it's a pleasure to work with you and you are a super great mentor and person. It is admirable the will you put in all you do. Even when things do not go as expected (many times in the SIBR project) it is always very interesting to discuss results with you and look for explanations. Thank you for all your great work and by the way, the answer was that it could feel it in its bones. **Rob**, thanks for all the support you offer to everyone in BacGen, you are a key element in the functioning of the lab and I am very happy to work together with you now. Your knowledge and ability in multiple fields together with your ideas and ways to improve the lab make always the difference. As you once told me, it is important you know how to read the lab, and you do that very well! Thanks for everything! Finally **Guus**, thanks for all your input during my thesis, you are a great contributor to this work. Although I suffered being the target of many of your projectiles and got some nicknames from you, it was fun to work together.

To everyone at Bacgen: **Adini, Alex, Catarina, Christian, Eugenios, Ioannis , James, Jeroen, Jorrit, Joyshree, Mihris, Olufemi, Prarthana, Suzan, Stijn, Teunke, Vittorio, Yifan.** Thanks the amazing environment in Bacgen, for all the discussions, meetings and science hours. Thanks to all the fellow **technicians: Steven, Laura, Philippe, Ton, Yehor, Caglar, Hans, Iame, Ineke** and **Merlijn,** it is great to be part of the group, your work is key to maintain a group as big as MIB. To **Hannie**, **Anja** and **Heidi**, thanks for all the help in all the bureaucratic issues I encountered during these years. To **Carrie, Jannie, Sharon, Patrick, Ivette** and all the ones that shared a timeline in MIB, it has been great to share this time together, join PhD parties together and

enjoy FriesDays! To all the people at MIB, thank you for keeping such a nice environment to work in. The diversity of topics and multidisciplinary techniques allows for a lot of interactions and a rich working environment.

To the students that contributed to the work in this thesis: **Koen**, **Kirsten**, **Kirsten**, **Carmen** and **Katerina**. Thank you for all the great work you put into your thesis. You all came in very different moments of my PhD and you all worked on different projects. Thank you for helping me learn while teaching. I hope you are all doing great!

Lyon, I met you at the beginning of my PhD and you have become a very good friend for me. Thank you for listening to me soooo many times and for being such a positive person. Thank you also for all the time shared together, all the dinners, parties, and nice time together. You even took care of our groceries all the times I got COVID...! I look forward to many more parties and trips together! **María**, eres una gran científica y te admiro un montón! Gracias por el buen rollo todos estos años, desde que nos conocimos en BacGen y ahora en SSB. Gracias por aguantar mis penas también y por todas las cosas que hemos organizado fuera del lab. Me encantan nuestros encuentros en el pasillo que pueden derivar en largas charlas. **Enrique**, desde que nos conocimos durante el máster, eres una persona que me aporta mucha paz. Muchas gracias por tus consejos y tus ánimos durante todos estos años y gracias por ayudarme en mis momentos de crisis. Tu amistad ha aportado mucho a este camino. **Pilar**, muchas gracias por ser la mejor compañera de salsa y por todas las charlas que hemos mantenido estos años. ¡Me lo paso genial contigo siempre! Te deseo todo lo mejor para esta última fase del doctorado, tú puedes con todo! **Yahia** and **Dani**, it has been great to share all the nice times we shared together in the last years. (I secretly feel very grateful that you were there in Houston when our flight was cancelled). Thank you for all the good moments we shared together. Without all these memories I would not have make it! Estoy muy feliz de haber creado tantos recuerdos felices con vosotros! Muchas gracias.

Albert, moltes gràcies per ser-hi sempre. Tot i que només ens veiem pocs cops a l'any, sempre em fa molta il·lusió tornar a Vilanova i anar a prendre algo o convidar-nos a sopar. Estic molt feliç de considerar-te amic meu i crec que ets una persona brillant i molt bona persona. Cada cop que torno, és com si no hagués passat el temps, ens coinexem des de fa tant...! Moltes gràcies per poder parlar de tot amb tú, m'encantaria que visquéssim més a prop i poder trobar-nos més sovint!

Anna, Wageningen em recorda molt sobint el temps que vam compartir les dues aquí. No saps quant de bé m'han fet les trucades de duració il·limitada en aquests anys on hem compartit penes, alegries i de tot. Gràcies a tu, també tinc una mica de trauma cada cop que veig una caixa de mudances... Ets una persona super positiva i me n'alegro molt de tenir-te com amiga. **Sandra**, estic molt feliç per tu i per la familia que estàs formant amb el Mark i la Maeve. Sé que puc parlar de tot amb tu i, tot i que ara sí que ens hem de veure més sovint, agraeixo tots els moments que hem compartit aquests anys. Tens una visió de la vida molt sana i m'agrada molt

escoltar-te. **Maya**, crazy mallorquina, m'encanta el teu entusiasme per tot el que fas. Estic molt feliç d'anar a la teva boda i celebrar amb tu un dia tant especial. **Marina**, amb tu he compartit molts moments durant la carrera (molt de tren, sobretot...) i estic molt contenta de poder tenir una sitgetana com a amiga. Em fa molta il·lusió que seguim quedant cada Nadal i conservem la tradició dels mojitos. Sense tu (i l'Anna), el món s'hauria perdut una gran despertadora de menta. Us estimo molt noies i us agraeixo haver estat al meu costat aquests anys.

Irving and **Chiara**, you were there in the beginning of my PhD and I share very good memories at Rijnveste 99. Thank you for being such great housemates. I will never forget our trip to Mexico and all the great moments we shared together. I hope we can meet together again soon.

Maite, no tengo palabras, muchas gracias por ayudarme tanto todos estos años. Hemos compartido mucho y me alegro mucho de poder mantener nuestra amistad durante años. **Kathe**, muchas gracias a ti también por todo lo compartido, me encantó compartir esta última fase del doctorado contigo. Te deseo mucha suerte en tu defensa. **Thibault**, thank you for being such a good person and friend. I still find very fun that we defend one day after the other. Good luck! **Gauthier**, all the best to you too! It has been great to share this cycle with the both of you and being able to cope with it throughout the salsa lessons, dinners and trips together. **Francesco** and **Paula**, thank you for hosting us in Italy almost every Pasqua, you and your family really made us feel at home. Giancarlo, Cristina e tutta la famiglia, grazzie per la boccata d'aria fresca ogni anno. Looking forward to more trips and meetings together!

Vicen, yo creo que no estaría escribiendo esto de no ser por tí. Gracias por confiar en mi y valorar-me tanto. Cuántas cosas han pasado desde el día en el que nos conocimos (aquella reunión del máster...). Me conoces muy bien y estoy feliz de tenerte como amigo! **Fer**, quién nos iba a decir que nos conoceríamos en Wageningen! Eres genial y aunque me cueste aceptarlo, me hacen demasiada gracia tus bromas. Nuestra amistad basada en las deudas, me parece algo que tenemos que mantener. Os quiero mucho, chicos y os echo de menos! **Anne Lise**, eres una buenísima amiga para mí y agradezco todos los cafés y comidas que hemos compartido juntas. Te deseo todo lo mejor para el doctorado que vas a comenzar. Me pareces una persona brillante y conseguirás todo lo que te proptongas! **Claudio,** as you said in Vicen's defence, it feels as if time did not pass every time we meet. Thanks for your friendship and all the good moments together! **Enrico**, thanks for being an awesome supervisor and for helping me during the PhD. Good luck in all your endevours!

Tía Ángeles, **Tío Miguel**, **Mireia** y **Noelia**, muchas gracias por ser familia para mí. Me encanta que aunque estemos lejos, podamos hablar y vernos en Navidades. **Mari**, **Maria** y **Manola**, gracias por ser parte de la familia. Vosotras me habéis visto crecer y convertirme en quien soy. David, Rosa Maria, Montse i Neus, moltes gràcies pels Nadals i vacances junts. Us estimo molt!

Giuliano e **Grazia**, grazie per farmi sentire a casa con voi. Andare in Italia tante volte gli ultimi anni mi ha fatto sempre molto bene.

Papa, espero que estuvieras orgulloso de mi. Muchas gracias por haberme hecho parte de la persona que soy. **Abuelito**, echo de menos pasar tiempo contigo. **Yayos**, os fuisteis sin poderme despedir de vosotros. Muchas gracias a todos por hacerme la persona que soy y transmitirme siempre tanto amor y generosidad. Os quiero mucho a todos y me hubiera gustado compartir este momento con vosotros.

Raef, gràcies per cuidar de la Judith i de ma mare i fer-me super abraçades sempre que em veus. **Judith**, gracias por aguantarme al teléfono, aguantar todos mis tics, manías y desorden. Aunque somos muy diferentes, sé que puedo contar contigo siempre. Gracias por aportar tu granito de arena en esta tesis en el diseño de la portada. Te admiro mucho, eres una persona súper creativa y llena de carácter y eso te va a abrir puertas siempre. Te quiero mucho Judy y espero que sepas que siempre voy a estar aquí para ti.

Mama, no estaría escribiendo esto de no ser por ti. Te tengo que dar las gracias por muchas cosas y tu apoyo durante todos los momentos de esta tesis es sin duda una de ellas. Muchas gracias por estar siempre al otro lado del teléfono o de la cámara y por mostrarme que las cosas se pueden conseguir. Eres un ejemplo de mujer para mí. Muchas gracias por ser un pilar tanto para mi como para Judith y por apoyarnos en todo lo que nos proponemos. Te quiero mucho y espero estar para ti en todo lo que necesites.

Lore, grazie per avermi completato. Hai sopportato gran parte del mio stress e della mia frustrazione negli ultimi anni, ma la tua visione positiva e costruttiva degli eventi mi ha sempre aiutato tanto. Nel bene e nel male, hai sempre le parole giuste per darmi energia, tranquillità e forza. Spesso ti fidi di me più di me stessa. La tua passione per la scienza e il lavoro ben fatto mi hanno spesso spinto a voler migliorare. Grazie per avermi aiutato a costruire questo libro e non vedo l'ora di scrivere una lunga storia con te. Te quiero.

The research described in this thesis was financially supported by Building Blocks of Life (BBoL) NWO program.

Provided by thesis specialist Ridderprint, ridderprint.nl

Layout and cover design: Wiebke Keck, persoonlijkproefschrift.nl
Printing: Ridderprint

Printed by Ridderprint on FSC-certified paper