Microbial Treasure Trove

Unravelling the Potential of Class 2 CRISPR-Cas systems

Prarthana Mohanraju

Microbial Treasure Trove

Unravelling the Potential of Class 2 CRISPR–Cas Systems

Prarthana Mohanraju

Thesis committee

Promotor
Prof. Dr John van der Oost
Professor of Microbial Genetics
Wageningen University & Research

Co-promotor
Dr Daan Swarts
Assistant Professor at the Laboratory of Biochemistry
Wageningen University & Research

Other members
Prof. Dr Michiel Kleerebezem, Wageningen University & Research
Prof. Dr Edze Westra, University of Exeter, Cornwall, UK
Prof. Dr Pascale Daran-Lapujade, Delft University of Technology
Prof. Dr Niels Geijsen, Hubrecht Institute, Utrecht

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

Microbial Treasure Trove

Unravelling the Potential of Class 2 CRISPR–Cas Systems

Prarthana Mohanraju

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 Friday 25th October 2019
at 4 p.m. in the Aula.

Prarthana Mohanraju
Microbial Treasure Trove - Unravelling the Potential of Class 2 CRISPR–Cas Systems
320 pages.

PhD thesis, Wageningen University, Wageningen, NL (2019)
With references, with summary in English

ISBN: 978-94-6395-125-8
DOI: 10.18174/500416

Ma, for you.

Propositions

1. CRISPR–Cas technologies are by no means the first or last example of ground-breaking applications that have been derived from bacterium-phage interaction studies.
   (this thesis)

2. Multifunction proteins are ingenious designs created by evolution's tinkering.
   (this thesis)

3. Directed evolution holds tremendous potential to solve mankind's chemical and pharmaceutical problems.

4. Infusion of funds for fundamental research is a force multiplier for the emergence of cutting-edge technological applications.

5. Flexible working hours offer tremendous productivity improvement.

6. Real creativity lies in how we can simplify things.

Propositions belonging to the thesis entitled:

Microbial Treasure Trove - Unravelling the Potential of Class 2 CRISPR–Cas Systems

Prarthana Mohanraju
Wageningen, 25th October 2019

Contents

Chapter 1

General introduction and thesis outline

Prokaryotes versus viruses

Life on earth is comprised of three domains; the Bacteria, the Archaea, and the Eukarya. Survival of all forms of life is continuously confronted by their constantly changing surroundings. Prokaryotes (bacteria and archaea) dominate many natural habitats, including some highly competitive and harsh environments. They thrive in such environments despite facing a constant threat of invasions by selfish Mobile Genetic Elements (MGEs) such as transposons, plasmids and viruses.

Bacteriophages (phages), are the viruses of bacteria that have been estimated to be the most abundant biological entities on the planet with population numbers in the order of 10^{31} (*1*, *2*). Phages are divided into two groups, virulent and temperate. Virulent phages follow a strictly productive lytic life cycle for propagation which ultimately kills the host for the release of new viral particles, each of which can engage in another round of infection. Temperate phages have the choice to switch between productive and dormant states. They either multiply in their host cells leading to cell lysis similar to virulent phages or integrate their phage genome into the bacterial chromosome to become a 'dormant' prophage. Prophages are propagated passively by the replication machinery of the bacterial cell (*3*). Microscopic phage–bacteria interactions have quite a profound effect on the macroscopic scale as they strongly influence global biogeochemical cycles (*4*), global climate (*5*), incidence of human diseases (*6*), as well as patterns of microbial genome diversity (*7*).

The Red Queen hypothesis (**Box 1**) posits that antagonistic biotic interactions, such as those displayed by hosts and parasites, generates a selective pressure on host organisms to continuously evolve systems that neutralize infections leading to adaptations of the host and counter-adaptations on the parasite's end (*8*). The continuous arms race of developing infection and resistance strategies results in a rapid co-evolution of the parasite's offence systems and host defence systems. Such an evolutionary trend is extremely prominent in phage-microbe interactions. This is mainly due to the rapid evolution and turnover of phage particles (*4*) leading to acute selective pressure on microbial communities to evade infection by developing immunity.

Bacterial immune arsenals

To combat the rapid turnover and evolution of phages, bacteria have evolved a diverse arsenal of defence mechanisms that act on the various stages of the viral life cycle (*9*–*11*). The combination of these defence mechanisms allows bacteria to fight off invading phages. Analogous to the immune systems in higher eukaryotes, defence mechanisms in bacteria can be divided into innate and adaptive immune systems. Innate immune systems are generic defence mechanisms that respond to invaders in a non-specific way, whereas adaptive immune systems are highly specific to a particular pathogen (*12*).

Box 1| The Red Queen hypothesis

The Red Queen hypothesis is related to the co-evolution of species. It is often used to describe evolutionary biological arms race based on the hypothesis originally proposed by Leigh Van Valen in 1973. The name is derived from a statement that the Red Queen tells Alice in Lewis Carroll's *Through the Looking Glass* in her explanation of the nature of Looking-Glass Land: "Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!"(*13*).

Van Valen coined the hypothesis "Red Queen" to explain the "law of extinction" where species must "run" – *i.e.* continuously adapt and evolve to survive rather than from going extinct. In tightly coevolved interactions, an adaptation in a population of the host or parasite species may change the selection pressure on a population of the parasite or host species, giving rise to an antagonistic co-evolution. The original theory suggested that in prey-predator relationship changes (*e.g.*, running faster (evolve) on the prey's end may lead to the predator to go extinct. The only way the predator can maintain its fitness is by counter-adaptation (running even faster). This will lead to a perturbed balance between prey and predator, where both species must constantly evolve to stay at the same place.

The Red Queen is teaching Alice, as depicted by John Tenniel in Chapter two – The Garden of Live Flowers of Lewis Carrol's famous book

In bacteria, defence mechanisms provide innate immunity that interfere at many different levels of the invader's infection cycle: **(i)** by receptor masking and/or mutation to prevent phage adsorption to the surface of the cells, **(ii)** by preventing the phage genome from entering the cytoplasm through Superinfection exclusion (Sie) systems, **(iii)** by cleaving the phage genome using Restriction–Modification (R–M systems) or prokaryotic Argonaute (pAgo) proteins, **(iv)** by restricting phage replication through the Bacteriophage Exclusion (BREX) or Defence Island System Associated with Restriction–Modification systems (DISARM), **(v)** by inducing Toxin–Antitoxin (TA) modules that trigger premature cell death, and **(vi)** by Abortive infection (Abi) mechanisms in which the host cell undergoes programmed cell death to prevent phage propagation (*14*–*19*). Together these innate immune mechanisms provide the first line of defence against infecting phages and other MGEs.

The above-mentioned innate immunity strategies are complemented by an adaptive immune function of the prokaryotic systems of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) proteins (*20*, *21*). The CRISPR–Cas system serves to specifically recognize and destroy genomes of MGEs. An overview of the discovery, biology and applications of the prokaryotic CRISPR–Cas defence systems is provided in this chapter. The chapter is concluded with an outline of this thesis, which describes the unravelling of unique mechanisms and developing applications of novel class 2 CRISPR–Cas systems.

Overview and nomenclature of CRISPR–Cas systems

Components and basic biology of CRISPR–Cas systems

The molecular machinery involved in CRISPR–Cas immunity is encoded by the CRISPR–Cas locus. A typical CRISPR–Cas locus consists of two genetic components that are often located next to each other in microbial genomes: a CRISPR array and an operon of multiple *cas* genes. The *cas* genes encode for the Cas proteins, which are necessary for the acquisition of new spacers, the maturation and binding of the CRISPR RNAs (crRNAs), and the targeting of MGEs.

The CRISPR array is composed of a series of short, partially palindromic repeating sequences (repeats) interspaced by variable sequences (spacers), originating from genomes of invading MGE. The spacers act as a 'memory bank' of prior infections and enable recognition of the invaders upon infection. Spacers complementary to the host genome (self-targeting spacers) have also been observed in some instances, albeit at low frequencies (*22*). Typically, new spacers are inserted at a specific end of the CRISPR array adjacent to a "leader" region. The leader usually contains promoter elements to drive transcription of the array as well as binding sites for other regulatory elements, some of which play a role in integration of new spacers (*23*–*26*). The integration of new spacers at the leader end enhances defence against recently encountered MGEs (*27*). Using these basic components, CRISPR–Cas systems mediate adaptive immunity through three general phases: adaptation, expression, and interference (*28*). During the adaptation phase, a subset of Cas proteins called the 'adaptation module' obtain and insert fragments of an invading MGE in the CRISPR array of the host genome. A sequence on the MGE (termed the protospacer) is selected, processed and added to the CRISPR array as a new spacer. Selection of protospacers occurs on the basis of a short Protospacer Adjacent Motif (PAM). This expands the memory of the immune system. In the expression phase, the *cas* genes are transcribed as a poly-cistronic mRNA and the CRISPR array is transcribed as a long precursor CRISPR RNA (pre-crRNA) and processed into individual small crRNAs (the guides). Each mature crRNA bears a segment complementary to the previously encountered MGE as well as a part of the CRISPR repeat. During interference, the crRNAs guide the 'effector complex' to complementary invading nucleic acids. Only in case a PAM motif is present next to the target sequence, Cas nuclease activity is induced to cleave the matching DNA and to neutralize the invasion (*29*).

Classification of CRISPR–Cas systems

A plethora of distinct variants of CRISPR–Cas systems have been identified in 87% of the archaeal genomes and 45% of the bacterial genomes sequenced to date (crispr.i2bc.paris-saclay.fr, accessed April 2019) (*30*, *31*). However, a recent study analysing the genomes of uncultivable microbes from environmental samples revealed a CRISPR–Cas abundance of only 9-10% in the sequenced organisms (*32*). This might suggest that the real abundance of CRISPR–Cas systems in nature is lower than previously thought. Nevertheless, the natural diversity of CRISPR–Cas systems is remarkably extensive. Whereas all systems use crRNA guides, there is variation concerning the nature of the target nucleic acid: many systems target DNA, some systems target RNA, and some systems target both DNA and RNA (*33*). To capture the diverse evolutionary origins and mechanisms of action, the current CRISPR-Cas classification includes two broad classes, each of which is subdivided into three types and many subtypes (*33*). Class 1 systems (types I, III and IV) utilize multi-subunit Cas protein complexes for the recognition and interference of target nucleic acids, while the less common class 2 systems (types II, V and VI) employ a single multi-domain effector complex that performs target recognition and cleavage. This classification has evolved during the last decennium as new genomic and metagenomic information has become available at an incredible pace, resulting in discovery and functional characterisation of new types and sub-types.

1

Brief history CRISPR–Cas systems

Discovery of CRISPR–Cas

An array of alternating repeats with interspersed spacers downstream of the *iap* gene was first observed in 1987 in the genome of *Escherichia coli* K12 (*34*) (**Figure 1.1**). This region was later given the CRISPR acronym (*21*). The first clues on the function of CRISPR–Cas systems came in early 2002 from two bioinformatics studies, both based on the principle of 'guilt by association'. One of these studies reported the presence of conserved operons that appeared to encode a novel DNA repair system [on the basis of predicted helicases, nucleases (*e.g.* HD, now known as Cas3/Cas10 in type I systems), polymerases] (*35*). The other reported that the spatial clustering of CRISPR arrays to specific CRISPR-associated *cas* genes indicates a functional relationship between them (*21*). A major subsequent breakthrough was the finding that some of the spacer sequences possess homology with DNA fragments of phages, leading to the hypothesis that the CRISPR–Cas could function as an immune system against the corresponding phages (*36*, *37*). Around the same time, another study on *Streptococcus thermophilus* reported more spacers that matched phage sequences and identified a large CRISPR-associated protein containing the DNA-cleaving HNH domain, which is now known as Cas9, the signature protein of type II systems (*38*). The detection of defined CRISPR transcripts and the *in silico* prediction of *cas* gene functions suggested that CRISPR–Cas was an RNA-based interference system analogous to RNAi mechanism found in eukaryotes (*39*). Despite the link between CRISPR–Cas and phage infection, the specific function of CRISPR spacers in immunity remained unclear.

Elucidation of the role of CRISPR–Cas systems

The first experimental evidence of CRISPR–Cas mediated adaptive immunity emerged in 2007 with the isolation of phage-resistant lactic acid bacterium *Streptococcus thermophilus* possessing new CRISPR spacers after a phage challenge. Acquisition of specific spacers from the phage genome into the bacterial CRISPR locus was shown to be causative for the phage resistant phenotype of the cells (*20*). Moreover, *cas* genes were required for both immunization as well as immunity (*20*).

In 2008, the first insight into the molecular mechanism of CRISPR–Cas came from the work on the type I-E system in *E. coli* (*40*). After transcription of the CRISPR array as a pre-crRNA, processing of mature crRNAs was demonstrated to be catalysed by a ribonuclease subunit of the CRISPR-associated complex for anti-viral defence (Cascade) complex. After transplantation of Cascade and a predicted nuclease Cas3 to an appropriate *E. coli* strain, the introduction of a designed CRISPR array with spacers of phage lambda genes resulted in efficient and specific protection of the bacterial host against transfection by phage lambda (*40*). Not only was it shown that the Cascade effector complex can be directed to multiple targets by adjusting the crRNA sequence, it was also demonstrated that both orientations of the crRNA guides were effective, thus for the first time indicating that DNA was the most likely target. This was in agreement with the predicted DNA targeting by the HD-containing Cas3 nuclease (*40*).

These exciting results established CRISPR–Cas as a novel microbial adaptive immune system. Unlike the adaptive immune system in humans (*12*), the CRISPR–Cas system is inheritable (*20*, *28*). In subsequent studies it was demonstrated that Cas3 activity resulted in severe damage of the target DNA, through exonuclease activity of the displaced DNA strand (*41*, *42*). Analyses of distinct Cascade-like complexes of type III systems reported that the transformation of plasmids carrying sequences matching to the CRISPR spacers was blocked by the type III-A

CRISPR–Cas system in *Staphylococcus epidermidis,* thereby limiting horizontal gene transfer (*43*). Interestingly, in 2009, it was reported that the type III-B CRISPR–Cas system in *Pyrococcus furiosus* was shown to target RNA instead of DNA (*44*). This study underscored the substantial mechanistic differences between the CRISPR–Cas systems. The studies mentioned above established and kicked off further mechanistic exploration of the function of CRISPR–Cas.

Molecular requirements of CRISPR–Cas systems

The role of CRISPR–Cas in microbial cell immunity sparked intense research interest in the understanding of its underlying molecular mechanisms. Discrimination between self and non-self is a universal necessity of immune systems. In 2009, it was demonstrated that a short well-conserved sequence motif adjacent to CRISPR targets on the invader DNA is required for recognizing the target for successful Cas9-mediated interference (*45*). This motif has been called the PAM (*46*) and it is now known that different systems have different PAM requirements (*47*). Typically, Cas proteins use PAM directly adjacent to their target to differentiate between the spacer found in the host CRISPR array, which lacks the PAM, and a viral protospacer. In this way, the requirement of a PAM facilitates discrimination between self and non-self DNA and thereby prevents autoimmunity.

In 2010, an important study demonstrated that *S. thermophilus* Cas9 is guided by crRNAs to generate double-stranded breaks (DSBs) at the targeted sites in invading phage and plasmid DNA. Moreover, Cas9 was shown to be the only protein required for crRNA-mediated target DNA cleavage (*48*). In 2011, small-RNA sequencing of *Streptococcus pyogenes* revealed the presence of an additional small RNA associated with the CRISPR array of type II systems (Cas9) (*49*). This additional RNA, termed the trans-activating CRISPR RNA (tracrRNA), forms a dsRNA duplex with the direct repeat sequences on the pre-crRNA that is recognized by a non-Cas ribonuclease (RNaseIII), resulting in a mature, dual RNA (crRNA:tracrRNA) (*49*). In the same year, another important step in understanding the components required for the CRISPR–Cas system came from the reconstitution of the CRISPR–Cas9 locus from *S. thermophilus* in *E. coli*. It was found that *E. coli* cells harbouring the CRISPR–Cas9 loci (including the gene encoding for tracrRNA) were resistant to plasmids containing DNA fragment that match with the spacers in the CRISPR-array (*50*). Finally, in 2012, a breakthrough study revealed that recombinant purified *S. pyogenes* Cas9 protein with the crRNA and tracrRNA can be used for targeted DNA cleavage *in vitro* (*51*). It was demonstrated that *S. pyogenes* Cas9-crRNA-tracrRNA complex cleaves DNA 3 bp upstream of the PAM, in agreement with previous *in vivo* results with the *S. thermophilus* system (*48*). Furthermore, it showed that the HNH domain and the RuvC domain is responsible for cleaving the target DNA strand and the non-target DNA strand, respectively. Inactivation of either domain turns the Cas9:crRNA:tracrRNA RibonucleoProtein (RNP) complex into a DNA nickase. Finally, the study also showed that the crRNA and tracrRNA can be fused to create a chimeric single guide RNA (sgRNA), leading to a simple two component system for targeted DNA cleavage (*51*). Similar findings using a purified RNP complex of *S. thermophilus* Cas9 reported in parallel (*52*). However, since this study was based on purified RNP complex from bacteria without analysing the components of the complex, it failed to identify the requirement of the tracrRNA for Cas9 function (*52*).

Early applications of CRISPR–Cas systems

The key milestones in the discovery and characterisation of the function of CRISPR–Cas systems (**Figure 1.1**) has provided the basis for the utilization of CRISPR–Cas system for diverse applications. However, some applications were applied even before elucidation of all the details

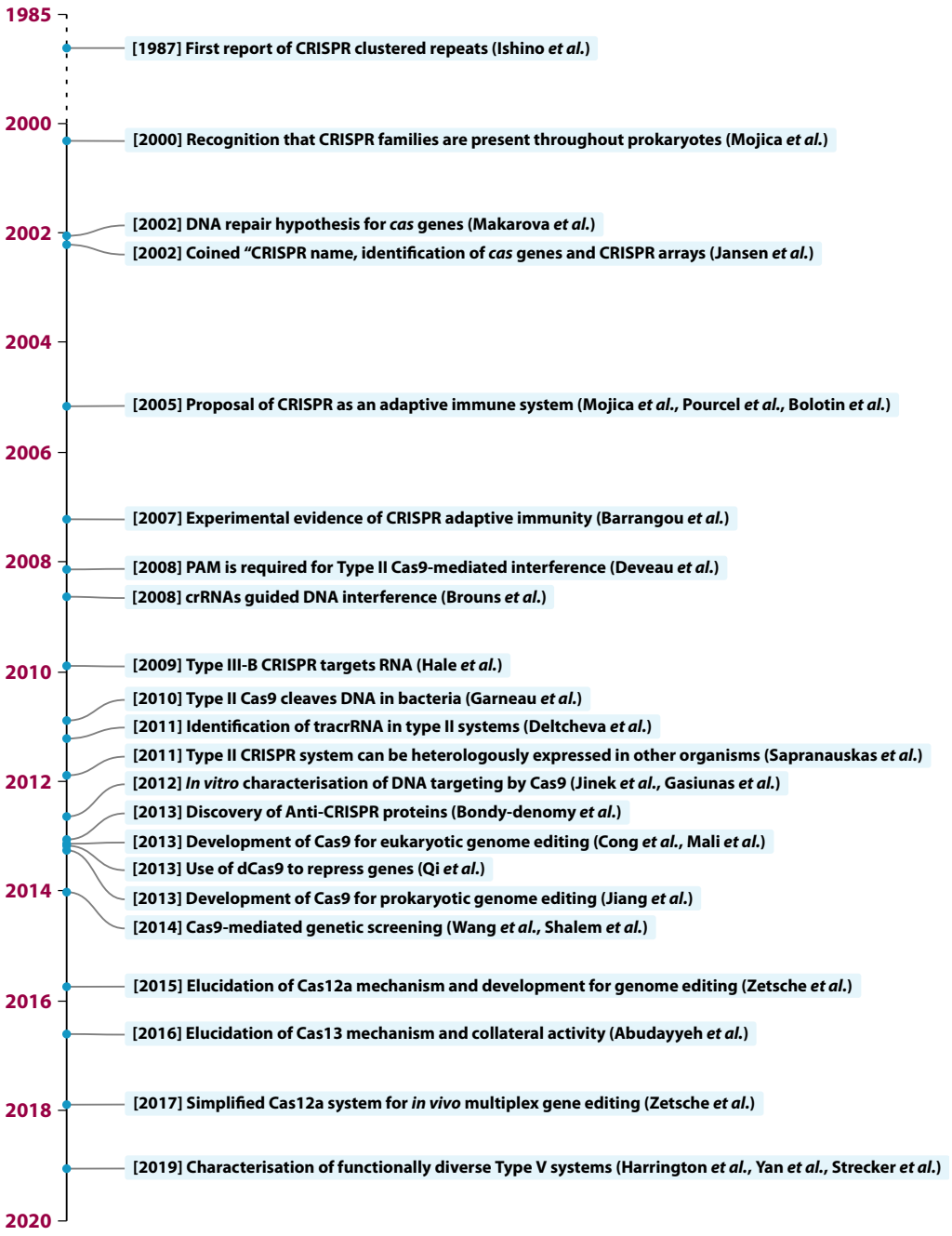


Figure 1.1 | Key CRISPR milestones and seminal discoveries. Select publications of fundamental CRISPR research as well as development of CRISPR into a genome editing tool are shown. This list is not exhaustive.

of the CRISPR–Cas defence mechanism. The heterogeneity of CRISPRs among isolates that are otherwise isogenic was used as a basis for bacterial strain genotyping (*37*, *53*, *54*). Further-

more, it was used in its native context where *S. thermophilus* cells harbouring the CRISPR–Cas systems were inoculated with phages to generate phage-resistant strains that can be deployed in industrial dairy applications, such as yogurt and cheese making (*55*). After the elucidation of its function, some of the proposed early applications for the use of CRISPR–Cas include transcriptional silencing in microbes (*56*), combating antibiotic resistance, vaccination against plasmid DNA uptake, and targeted DNA destruction (*43*, *48*).

Adaptation of CRISPR–Cas effectors to the molecular biology toolbox

In 2013, CRISPR–Cas was harnessed for mammalian genome editing for the first time. The human codon-optimised *S. pyogenes* Cas9 tagged with a nuclear localization signal and chimeric sgRNA was used for gene editing in human and murine cell lines (*57*, *58*). The functionality of Cas9 in eukaryotic cells set the stage for a range of applications in biotechnology and biomedicine. Subsequently, additional studies also reported the use of Cas9 in human and animal cells (*59*, *60*) and a catalytically inactivated variant of Cas9 was used to achieve targeted silencing of gene expression (*61*). Although CRISPR–Cas system was first used in mammalian and bacterial cells only six years ago, it has already been used for a myriad of applications, including but not limited to: genome engineering, pathway component screens, transcriptional regulation of gene expression, imaging of specific DNA sequences, epigenetic modification of specific sequences, and detection of specific nucleotide sequences for diagnostics (*57*, *58*, *61*–*73*) (*References not exhaustive*). The finding that Cas effectors can be repurposed as programmable enzymes for technological applications has shaped the CRISPR–Cas field as we know it to date.

The success of Cas9-based technologies has stimulated extensive exploration of bacterial genomes for discovering new class 2 systems with the potential to expand the capabilities of CRISPR-based technologies (*74*–*76*). In just the last four years, a veritable explosion in the identified number of class 2 types and sub-types has taken place, with many novel systems likely remaining to be discovered. Further investigations into the microbial diversity has revealed many additional subtypes of CRISPR–Cas systems (*74*, *77*–*81*) and provided insight into the origin, evolution, and function of these sophisticated systems. By mining the microbial diversity for signatures of CRISPR–Cas systems (*e.g.*, conserved genes and CRISPR-like arrays) (*76*, *82*) and elucidation of the functions, type V systems containing the signature DNA-targeting Cas12 (formerly Cpf1 or C2c1) effectors (*83*, *84*) and type VI containing the RNA-targeting Cas13 (formerly C2c2) effectors (*85*, *86*) have been added to the CRISPR tool chest. The class 2 effectors represent the enormous diversity of CRISPR–Cas systems offer an extensive toolkit for a broad array of technologies. The focus of this thesis is on the diverse types of the class 2 CRISPR–Cas systems, therefore specific molecular mechanistic details on the different types will be detailed in the following chapters.

Phages fight back

As bacteria find ways to evade the infection by phages through evolving new defence systems, the phages evolve to counter these hurdles. Relatively high rates of mutation and recombination in phages helps them to rapidly counter the microbial antiviral defence mechanisms. Phages can modify their receptor-binding proteins to recognize altered receptors or acquire enzymes to access masked receptors and overcome adsorption inhibition (*16*, *87*). Phages protect themselves from R–M systems by expressing proteins to modify nucleotides in their genome (*88*–*90*), by expressing proteins that mimic DNA to bind and block R–M effectors (*91*), or by

degrading R–M cofactors (*92*). To counteract the TA systems, some phages express antitoxin molecules that silence resident toxin-antitoxin modules (*16*, *87*).

Phage evasion from CRISPR–Cas systems was first found to rely on mutational drifts, predominantly occurring in regions that require perfect complementarity between the crRNA and the protospacer (so-called seed region) for interference or in the PAM sequences (*46*, *93*, *94*). Depending on the location of the mismatch (between the crRNA and the protospacer), a single mutation can be sufficient to abolish CRISPR–Cas immunity (*46*, *95*). Deletion of the protospacer sequence and/or the PAM have also been shown to provide an effective way for phages to escape CRISPR–Cas targeting, despite the potential of imposing a fitness cost due to loss of gene function (*46*). Moreover, some phages encode their own CRISPR locus that encodes crRNAs that target host antiviral genomic regions, such as chromosomal (defence) islands (*96*).

A more refined countermeasure of phages against CRISPR–Cas mediated immunity is the use of anti-CRISPR (Acr) proteins, that were first discovered in phages to silence class 1 type I-F and I-E systems of *Pseudomonas aeruginosa* (*97*, *98*). Acr proteins have distinct sequences, structures (*99*, *100*) and mechanisms (*94*) and they provide phages with a direct and specific means to inhibit targeting by the CRISPR–Cas system. To date, 45 unique families of Acr proteins have been discovered, and categorized into class 1 and class 2 CRISPR–Cas inhibitors (*101*). These highly diverse and small (typically 50 – 330 amino acids) proteins do not share considerable sequence or protein domain similarity to each other or to any protein of known function (*102*). Anti-CRISPR proteins impede CRISPR–Cas systems using a wide variety of mechanisms, including but not limited to: steric occlusion of DNA binding (*94*), mimicking DNA and sgRNA (*94*, *103*, *104*), leading to dimerization of Cas9 (*105*), and acetylation of the residues involved in the recognition of the PAM (*106*). Several AcrIIA and AcrIIC proteins have been used as "off-switches" for Cas9 activity during genome editing experiments in human cell lines (*107*, *108*).

The Acr proteins that have been identified so far underscore the sequence and structural diversity of Acrs. Therefore, the discovery of new Acrs, especially those targeting CRISPR–Cas (sub)types for which Acr proteins have not yet been found, is expected to expand the numbers of distinct Acr protein families. Elucidation of the mechanisms of multiple Acr proteins is expected to fuel the Acr-based fine-tuning of CRISPR–Cas applications, such as gene editing. On the basis of the everlasting arms-race between bacteria and phages, it is anticipated that bacteria have also developed mechanisms to counteract Acr protein activity. Future research will increase our understanding on phage-bacterium coevolution and expand the CRISPR toolbox for biotechnological applications.

CRISPR–Cas: beyond defence

Apart from the canonical role in host defence, certain studies have suggested additional involvement of the CRISPR–Cas systems in other cellular processes. Nevertheless, this is not the first time that a protein complex has been associated in more than one function. Methyltransferases of the type III R-M system are known for their role in defence, but they also play a key function in phase variation (*109*). The eukaryotic RNAi system has been implicated in host defence as well as in gene regulation (*110*).

Although the exact mechanism is not known, the *cas* genes of the CRISPR–Cas type I-C system in the soil bacterium *Myxococcus xanthus* appear to be implicated in the regulation of the gene circuits that control fruiting-body formation during sporulation in its life-cycle (*111*). In *Pseu-*

domonas aeruginosa, the type I-F system is involved in phage-mediated inhibition of swarming and biofilm formation (*112*, *113*). Cas2 is requisite for infection of amoebae by *Legionella pneumophila* (*114*). In *Listeria monocytogenes*, a *cas* genes-independent CRISPR targets the host genome, and the crRNAs increase the level of target RNA by stabilizing it, thereby involving in direct gene regulation (*115*). Another study based on bioinformatics revealed that resistance of *Enterococcus faecalis* to antibiotics is inversely correlated with the presence of CRISPR–Cas systems (*116*). Suggestive of a role in DNA repair, association of Cas1 of type I-E system in *E. coli* with DNA repair components (RecB, RecC and RuvB), as well as defects in chromosome segregation and DNA repair in Δ*cas1* strains have been reported (*117*). Consistent with this report, *cas* gene expression in *Pyrococcus furiosus* is upregulated in response to gamma irradiation (*118*).

Comparative genomic analysis revealed high prevalence of CRISPR–Cas9 systems in bacterial pathogens (*119*). Notably, in intracellular bacterial pathogens such as *Francisella novicida*, *Neisseria meningitidis,* and *Camplyobacter jejuni*, Cas9 activity correlates with induced death of infected human cells (*120*) or infected laboratory animals (*121*). Overproduction of a specific bacterial lipoprotein (BLP) in the membrane of *F. novicida* was found to significantly decrease its survival in host macrophages (*122*). It was later demonstrated that *F. novicida* is able to silence the production of this BLP via a CRISPR–Cas9 based antisense RNA mechanism that is activated upon invasion of eukaryotic cells (*121*). In *C. jejuni*, Cas9 has been demonstrated to be translocated across the intestinal epithelial barrier, as well as in cytotoxicity (*120*). However, the molecular mechanism and the (in)direct involvement of CjeCas9 in inducing cell death awaits characterisation. It is, however, not clear if these proposed additional roles developed prior to the function as an immune system, in parallel with it, or from it. Evolutionary tinkering is particularly evident in the microbial world (*123*), so perhaps, these "additional roles" are an outcome of that.

1

Thesis Outline

As introduced in **Chapter 1**, the eternal biological arms-race between phages and bacteria has led to the relatively fast evolution of diverse bacterial defence systems. The prokaryotic CRISPR–Cas adaptive defence systems are an elegant complementation to the bacterial innate immune systems that combat invaders such as phages and plasmids. In addition to providing insight into the molecular underpinnings of life in the microbial world, CRISPR–Cas systems have successfully been adapted into powerful technologies for broad applications in life sciences, medicine, and industrial biotechnology. The development of crucial molecular biology technologies such as restriction enzymes and green fluorescent proteins have benefitted significantly from explorations of natural diversity. Inspired by these spectacular fundamental discoveries, the overall aim of this thesis is to unravel molecular mechanisms of hitherto uncharacterised class 2 CRISPR–Cas systems, and to harness the fundamental mechanistic features of different class 2 systems for the design and development of innovative genome engineering tools.

Chapter 2 | Diverse evolutionary roots and mechanistic variations of CRISPR–Cas systems

The second chapter reviews the recent advances in the field of CRISPR–Cas research. The latest CRISPR–Cas classification scheme delineates two classes that are each subdivided into three types. Integration of biochemistry, molecular genetics, structural analysis and single molecule studies has revealed many unique features of the variant CRISPR–Cas types, which are described in this chapter. This chapter also provides a plausible scenario of CRISPR–Cas evolution, and briefly describes the latest developments of a wide range of CRISPR-based applications.

Chapter 3 | Heterologous expression and purification of CRISPR–Cas12a/Cpf1

Chapter 3 provides a brief background of CRISPR–Cas12a based genome editing in diverse species and provides a step-by-step protocol for heterologous expression of *Francisella novicida* Cas12a (previously known as Cpf1) in *Escherichia coli*. In addition, it includes a protocol for high-purity purification as well as instructions on how to verify the activity of the purified proteins. These protocols can additionally be used for the purification of other Cas12a orthologues for use in biochemical and genome editing experiments. This purification protocol has been applied in the other chapters of this thesis.

Chapter 4 | Multiplex gene editing by CRISPR-Cas12a (Cpf1) using a single crRNA array

In this chapter, the mechanism of crRNA biogenesis in type V-A CRISPR–Cas systems is revealed. Unlike Cas9, which utilizes a tracrRNA as well as the endogenous RNaseIII for maturation of its dual crRNA/tracrRNA guides, crRNA processing of the Cas12a system proceeds in the absence of tracrRNA or other Cas proteins. It is demonstrated that Cas12a nucleases possess a previously unknown RNase domain that is responsible for cleaving the precursor-crRNA to generate the mature crRNAs. Thus, Cas12a is sufficient for pre-crRNA maturation in type V systems. The ability to autonomously process crRNA has significant implications from a genome editing standpoint, as it would provide a simple route to editing multiple genomic loci at a time (multiplex editing). It is demonstrated that using a single customized CRISPR array up to four genes in mammalian cells and up to three genes in mouse brain cells can be

edited simultaneously.

Chapter 5 | Characterising the CRISPR–Cas type V-U1 C2c4 effector enzyme

The characterisation of a novel, diminutive type V-U1 Cas effector from *Mycolicibacterium mucogenicum* (MmuC2c4) is described in **Chapter 5**. Type V-U effectors are highly similar to the typical transposon-encoded TnpB and appear to have originated independently from distinct TnpB families. Akin to most type V effectors characterised so far, MmuC2c4 recognizes a 5'-TTN-3' PAM on a double-stranded target DNA. Unexpectedly, it was found that this enzyme does not cleave dsDNA; in fact MmuC2c4 can be used for single- and multiplex- gene silencing in *E. coli*. Additionally, experiments are presented that hint towards a unique targeting mechanism where binding to a transcription-associated dsDNA activates MmuC2c4 to conduct mRNA cleavage.

Chapter 6 | Characterising a thermostable Cas9 for bacterial genome editing and silencing

This chapter focuses on the characterisation of ThermoCas9, a thermostable Cas9 orthologue from the type II-C CRISPR–Cas system of *Geobacillus thermodenitrificans*. It is demonstrated that ThermoCas9 can be programmed to cleave DNA *in vitro* up to 70 ℃. This ability of ThermoCas9 provided new inroads for the development of versatile and robust Cas9-based genome engineering tools, applied on *Bacillus smithii* at 55 ℃ as well as on *Pseudomonas putida* at 37 ℃. This study expands the potential of Cas9-based genome editing and opens up new avenues for using Cas9 technologies in novel applications requiring activity over a wide temperature range.

Chapter 7 | Development of a CRISPR-Cas12a based genome editing tool for facultative thermophiles

This chapter describes the biochemical characterisation of four Cas12a orthologues in order to assess their salt tolerance as well as pH- and temperature- stability. In addition, it describes the first application of *Francisella tularensis subsp. novicida* (FnCas12a) and *Eubacterium eligens* (EeCas12a) for genome editing in a moderate thermophilic bacterium, *Bacillus smithii*. Plasmid-borne editing templates introduce the desired mutations into the *B. smithii* genome at higher culturing temperatures via homologous recombination (HR), while consequent culturing at lower temperatures allows for Cas12a-based introduction of lethal double-stranded DNA breaks to the genomes of the non-edited cells, acting as a counter-selection mechanism. It is shown that using this process, a mutant can be generated within a short span of 2-3 days. This protocol can be easily adapted for gene editing in a wide variety of both mesophilic- and moderate thermophilic- organisms.

Chapter 8 | Characterising an anti-CRISPR based on/off switch for bacterial genome engineering

This chapter presents the effort to harness the activity of an anti-CRISPR protein for controllable bacterial genome engineering. It is demonstrated that AcrIIC1$_{Nme}$ anti-CRISPR protein from *Neisseria meningitidis* can bind to two thermo-active Cas9 orthologues (ThermoCas9 and GeoCas9) and transform them into their catalytically inactive forms *in vivo* in *E. coli* at 37 ℃. The Cas9/AcrIIC1$_{Nme}$ complexes possess a transcriptional silencing effect with an efficiency that is comparable to the catalytically "dead" ThermodCas9 and GeodCas9 variants. Finally, a controllable and efficient Cas9/AcrIIC1$_{Nme}$-based tool for coupling silencing and targeting in

bacteria is described.

Chapter 9 | Guide-free Cas9 from pathogenic *Campylobacter jejuni* bacteria causes severe DNA damage

A novel biological role for the CRISPR–Cas9 system of the pathogen *Campylobacter jejuni* (CjeCas9) is presented in **Chapter 9**. It is demonstrated that upon *C. jejuni* infection of human cells CjeCas9 is secreted into the cytoplasm of the infected cells, and it can autonomously enter the nucleus. Inside the nucleus, it catalyses metal-dependent and sequence-independent nicking of double stranded DNA, eventually leading to cell death. Genome editing using CjeCas9 is compared with the widely used Cas9 from *Streptococcus pyogenes* (SpyCas9), and the latter is shown to be superior. It is concluded that the unique catalytic features make CjeCas9 nuclease less suitable for genome editing applications.

Chapter 10 | Summary and general discussion

The final chapter contains a summary of the work detailed in this thesis. It further discusses the general scope of the thesis, describes additional experiments not reported in the preceding chapters, and presents the described results in perspective. In addition, pertinent recent developments and outstanding questions in terms of fundamental mechanisms as well as applications of class 2 CRISPR–Cas systems are discussed.

Chapter 2

Diverse evolutionary roots and mechanistic variations of CRISPR–Cas systems

Prarthana Mohanraju, Kira S. Makarova, Bernd Zetsche, Feng Zhang, Eugene V. Koonin, John van der Oost

Adapted from:
Diverse evolutionary roots and mechanistic variations of the CRISPR–Cas systems.
2016, Science, *353*(6299), aad5147.

Abstract

Adaptive immunity had been long thought of as an exclusive feature of animals. However, the discovery of the CRISPR–Cas defence system, present in almost half of prokaryotic genomes, proves otherwise. Because of the everlasting parasite-host arms race, CRISPR–Cas has rapidly evolved through horizontal transfer of complete loci or individual modules, resulting in extreme structural and functional diversity. CRISPR–Cas systems are divided into two distinct classes that each consist of three types and multiple subtypes. We discuss recent advances in CRISPR–Cas research that reveal elaborate molecular mechanisms and provide for a plausible scenario of CRISPR–Cas evolution. We also briefly describe the latest developments of a wide range of CRISPR-based applications.

Introduction

Bacteria and archaea suffer constant predation by viruses, which are extremely abundant in almost all environments (*2*). Accordingly, bacteria and archaea have evolved a wide range of antivirus defence mechanisms (*124*). Because viruses generally have high rates of mutation and recombination, they have the potential to rapidly escape these prokaryotic defence systems. Thus, the hosts' defences must also adjust and evolve rapidly, leading to an ongoing virus-host arms race. Protective systems provide innate immunity at all stages of the parasite's infection cycle via receptor masking, restriction-modification (R-M) systems, DNA interference [prokaryotic Argonaute proteins protect the host against mobile genetic elements (MGEs) through DNA-guided DNA interference], bacteriophage exclusion (BREX systems allow phage adsorption but block phage DNA replication; PGL systems have been hypothesized to modify the phage progeny DNA to inhibit their growth upon reinfection), and abortive infection (*14*, *16*–*18*, *125*, *126*).

The innate immunity strategies are complemented by an adaptive immune function of the systems of prokaryotic clustered regularly interspaced short palindromic repeats (CRISPR) and the associated Cas proteins (*20*, *21*). Diverse variants of the CRISPR–Cas systems are present in the examined genomes of most archaea and almost half of the bacteria (*15*). Here, we discuss insights into the evolution and functionality of class 1 and class 2 CRISPR–Cas systems. This progress has enabled the development of sophisticated tools for genetic engineering in molecular biology, biotechnology, and molecular medicine.

CRISPR–Cas defense

The CRISPR–Cas systems provide protection against MGEs — in particular, viruses and plasmids — by sequence-specific targeting of foreign DNA or RNA (*20*, *40*, *43*, *44*, *48*, *127*). A CRISPR–Cas locus generally consists of an operon of CRISPR-associated (*cas*) genes and a CRISPR array composed of a series of direct repeats interspaced by variable DNA sequences (known as spacers) (**Figure 2.1A**). The repeat sequences and lengths as well as the number of repeats in CRISPR arrays vary broadly, but all arrays possess the characteristic arrangement of alternating repeat and spacer sequences. The spacers are key elements of adaptive immunity, as they store the "memory" of an organism's encounters with specific MGEs acquired as a result of a previous unsuccessful infection (*36*–*39*). This memory enables the recognition and neutralization of the invaders upon subsequent infections (*20*).

CRISPR-mediated adaptive immunity involves three steps: adaptation, expression, and interference (*28*, *48*, *128*, *129*) (**Figure 2.1B**). During the adaptation step, fragments of foreign DNA (known as protospacers) from invading elements are processed and incorporated as new spacers into the CRISPR array. The expression step involves the transcription of the CRISPR array, which is followed by processing of the precursor transcript into mature CRISPR RNAs (crRNAs). The crRNAs are assembled with one or more Cas proteins into CRISPR ribonucleoprotein (crRNP) complexes. The interference step involves crRNA-directed cleavage of invading cognate virus or plasmid nucleic acids by Cas nucleases within the crRNP complex (*28*, *48*, *130*). The multifaceted and modular architecture of the CRISPR–Cas systems also allows it to play non-defense roles, such as biofilm formation, cell differentiation, and pathogenicity (*131*–*133*).

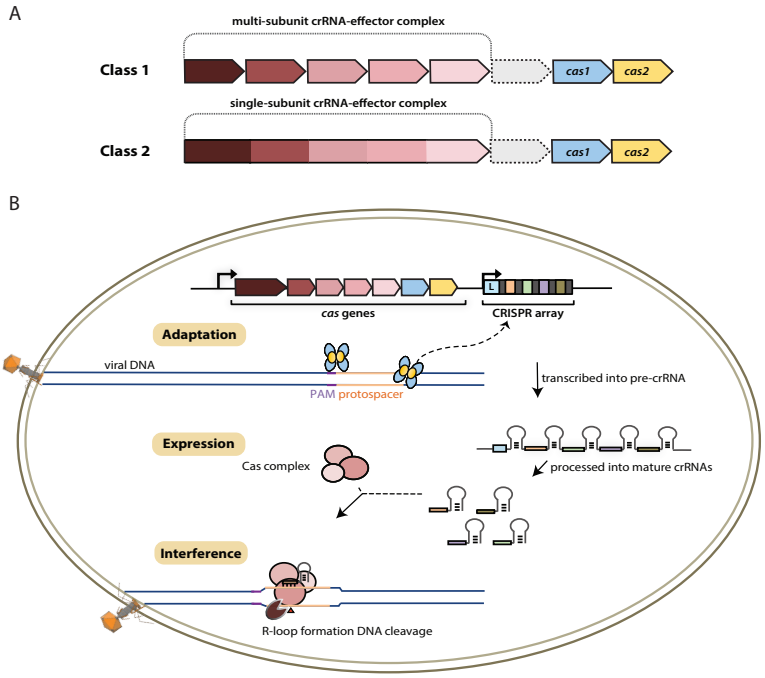


Figure 2.1 | Overview of the Clustered Regularly Interspaced Short palindromic repeats–CRISPR-associated proteins (CRISPR-Cas) systems. (A) Architecture of class 1 (multiprotein effector complexes) and class 2 (single-protein effector complexes) CRISPR–Cas systems. **(B)** CRISPR–Cas adaptive immunity is mediated by CRISPR RNAs (crRNAs) and Cas proteins, which form multicomponent CRISPR ribonucleoprotein (crRNP) complexes. The first stage is adaptation, which occurs upon entry of an invading mobile genetic element (in this case, a viral genome). Cas1 (blue) and Cas2 (yellow) proteins select and process the invading DNA, and thereafter, a protospacer (orange) is integrated as a new spacer at the leader end of the CRISPR array [repeat sequences (grey) that separate similar-sized, invader-derived spacers (multiple colours)]. During the second stage, expression, the CRISPR locus is transcribed and the pre-crRNA is processed into mature crRNA guides by Cas (*e.g.*, Cas6) or non-Cas proteins (*e.g.*, RNase III). During the final interference stage, the Cas-crRNA complex scans invading DNA for a complementary nucleic acid target, after which the target is degraded by a Cas nuclease.

CRISPR–Cas diversity, classification and evolution

The rapid evolution of highly diverse CRISPR–Cas systems is thought to be driven by the continuous arms race with the invading MGEs (*134*, *135*). The latest classification scheme for CRISPR–Cas systems, which takes into account the repertoire of *cas* genes and the sequence similarity between Cas proteins and the locus architecture, includes two classes that are currently subdivided into six types and 19 subtypes (*31*, *76*). The key feature of the organization and evolution of the CRISPR–Cas loci is their pronounced modularity. The module responsible for the adaptation step is largely uniform among the diverse CRISPR–Cas systems and consists of the *cas1* and *cas2* genes, both of which are essential for the acquisition of spacers. In many CRISPR–Cas variants, the adaptation module also includes the *cas4* gene. By contrast, the CRISPR–Cas effector module, which is involved in the maturation of the crRNAs as well as in target recognition and cleavage, shows a far greater versatility (**Figure 2.2A**) (*31*).

The two classes of CRISPR–Cas systems differ fundamentally with respect to the organization of the effector module (*31*). Class 1 systems (including types I, III, and IV) are present in bac-

teria and archaea, and encompass effector complexes composed of four to seven Cas protein subunits in an uneven stoichiometry [*e.g.*, the CRISPR-associated complex for antiviral defence (Cascade) of type I systems, and the Csm/Cmr complexes of type III systems]. Most of the subunits of the class 1 effector complexes — in particular, Cas5, Cas6, and Cas7 — contain variants of the RNA-binding RNA recognition motif (RRM) domain (*124*, *136*). Although the sequence similarity between the individual subunits of type I and type III effector complexes is generally low, the complexes share strikingly similar overall architectures that suggest a common origin (*76*, *136*–*138*). The ancestral CRISPR–Cas effector complex most likely resembled the extant type III complexes, as indicated by the presence of the archetypal type III protein, the large Cas10 subunit, which appears to be an active enzyme of the DNA polymerase–nucleotide cyclase superfamily, unlike its inactive type I counterpart (Cas8) (*76*, *124*, *136*).

In the less common class 2 CRISPR–Cas systems (types II, V, and VI), which are almost completely restricted to bacteria, the effector complex is represented by a single multidomain protein (*31*). The best-characterised class 2 effector is Cas9 (type II), the RNA-dependent endonuclease that contains two unrelated nuclease domains, HNH and RuvC, that are responsible for the cleavage of the target and the displaced strand, respectively, in the crRNA–target DNA complex (*51*). The type II loci also encode a trans-acting CRISPR RNA (tracrRNA) that evolved from the corresponding CRISPR repeat and is essential for pre-crRNA processing and target recognition in type II systems (*49*, *139*). The prototype type V effector Cas12a (previously known as Cpf1) (subtype V-A) contains only one nuclease domain (RuvC-like) that is identifiable by sequence analysis (*83*). However, analysis of the recently solved structure of Cas12a complexed with the crRNA and target DNA has revealed a second nuclease domain, the fold of which is unrelated to HNH or any other known nucleases. In analogy to the HNH domain in Cas9, the novel nuclease domain in Cas12a is inserted into the RuvC domain, and it is responsible for cleavage of the target strand (*140*).

Screening of microbial genomes and metagenomes for undiscovered class 2 systems (*76*) has resulted in the identification of three novel CRISPR–Cas variants. These include subtypes V-B and V-C, which resemble Cas12a in that their predicted effector proteins contain a single, RuvC-like nuclease domain. Cleavage of target DNA by the type V-B effector, denoted Cas12b (previously known as C2c1), has been experimentally demonstrated (*76*). Type VI is unique in that its effector protein contains two conserved HEPN domains that possess ribonuclease (RNase) activity (**Figure 2.2A**).

Recent comparative genomic analyses of variant CRISPR–Cas systems (**Figure 2.2B**) (*76*) have revealed a strong modular evolution with multiple combinations of adaptation modules and effector modules, as well as a pivotal contribution of mobile genetic elements to the origin and diversification of the CRISPR–Cas systems. The ancestral prokaryotic adaptive immune system could have emerged via the insertion of a casposon (a recently discovered distinct class of self-synthesizing transposons that appear to encode a Cas1 homolog) next to an innate immunity locus (probably consisting of genes encoding a Cas10 nuclease and possibly one or more RNA binding proteins). Apart from providing the Cas1 nuclease/integrase that is required for recombination during spacer acquisition (*141*–*143*), the casposon may also have contributed the prototype CRISPR repeat unit that could have evolved from one of the inverted terminal repeats of the casposon (*144*). An additional toxin-antitoxin module that inserted either in the ancestral casposon or in the evolving adaptive immunity locus probably provided the *cas2* gene, thus completing the adaptation module. The Cas10 nuclease and one or more additional proteins with an RRM fold (the ultimate origin of which could be a polymerase or cyclase that

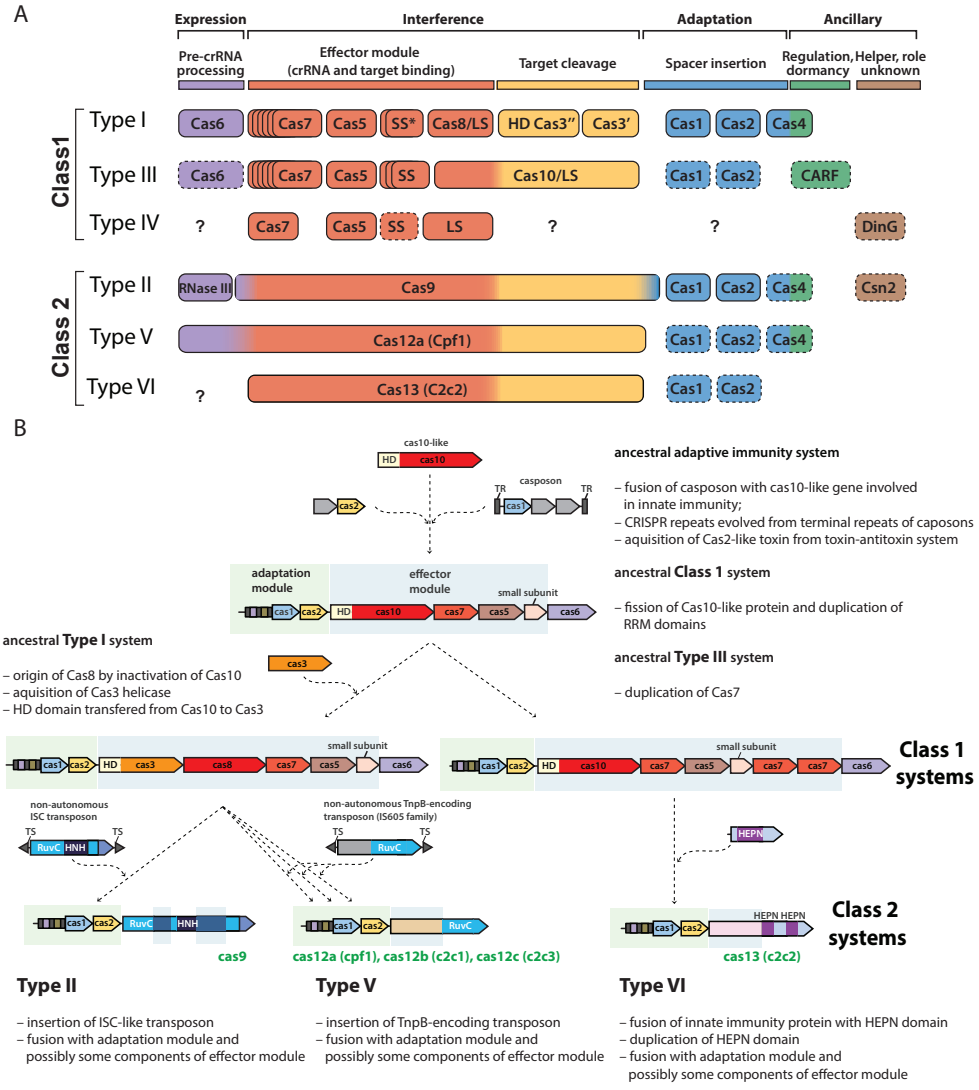


Figure 2.2|CRISPR diversity and evolution. (A) Modular organization of the CRISPR–Cas systems. LS, large subunit; SS, small subunit. A putative small subunit that might be fused to the large subunit in several type I subtypes is indicated by an asterisk. Cas3 is shown as fusion of two distinct genes encoding the helicase Cas3′ and the nuclease HD Cas3′′; in some type I systems, these domains are encoded by separate genes. Functionally dispensable components are indicated by dashed outlines. Cas6 is shown with a thin solid outline for type I because it is dispensable in some systems, and by a dashed line for type III because most systems lack this gene and use the Cas6 provided in trans by other CRISPR–Cas loci. The two colours for Cas4 and Cas13 (C2c2) and three colours for Cas9 and Cas12a reflect the contributions of these proteins to different stages of the CRISPR–Cas response (see text). The question marks indicate currently unknown components. Modified with permission from (*31*). **(B)** Evolutionary scenario for the CRISPR–Cas systems. TR, terminal repeats; TS, terminal sequences; HD, HD-family endonuclease; HNH, HNH-family endonuclease; RuvC, RuvC-family endonuclease; HEPN, putative endoribonuclease of HEPN superfamily. Genes and portions of genes shown in grey denote sequences that are thought to have been encoded in the respective mobile elements but were eliminated in the course of evolution of CRISPR–Cas systems. Modified with permission from (*76*).

gave rise to Cas10) of the hybrid locus could have subsequently evolved to become the ancestral CRISPR–Cas effector module (*76*, *124*, *136*, *144*).

The widespread occurrence of class 1 systems in archaea and bacteria, together with the proliferation of the ancient RRM domain in class 1 effector proteins, strongly suggests that the ancestral CRISPR–Cas belonged to class 1. Most likely, the multiple class 2 variants then evolved via several independent replacements of the class 1 effector locus with nuclease genes that were derived from distinct MGEs (**Figure 2.2B**). In particular, type V effector variants (Cas12a) seem to have evolved from different families of the TnpB transposase genes that are widespread in transposons (*76*), whereas the type II effector (Cas9) may have evolved from IscB, a protein with two nuclease domains that belongs to a recently identified distinct transposon family (*145*). Notably, class 2 CRISPR–Cas systems, in their entirety, appear to have been derived from different MGEs: Cas1 from a casposon, Cas2 from a toxin-antitoxin module, and the different effector proteins (such as Cas9 and Cas12a) from respective transposable elements (*76*).

CRISPR adaptation

The spacers of a CRISPR array represent a chronological archive of previous invader encounters. The captured spacer sequences are integrated into the CRISPR loci after exposure to MGEs, at the leader end of the array that contains the start site of CRISPR transcription (*20*, *48*, *146*). Analysis of invader target sequences (also called protospacers) has revealed a short motif directly adjacent to the target sequence, called the protospacer adjacent motif (PAM) (*46*). This PAM motif allows self/nonself discrimination by the host in two ways: (i) because its presence in alien targets is required for nonself interference, and (ii) because its absence in the host's CRISPR array avoids self-targeting (*45*). In class 1 type I and class 2 type II systems, the PAM is not only involved in interference, but also plays a role in spacer selection during the adaptation stage, implying the acquisition of functional spacers only (*147*, *148*). The PAM is a short [2 to 7 nucleotides (nt)], partially redundant sequence that in itself cannot preclude incorporation of spacers from the host DNA because of the low information content of the motif. The short PAM appears to be the result of an evolutionary trade-off between efficient incorporation of spacers from nonself DNA and preventing an autoimmune reaction.

Although host chromosomal fragments can be incorporated as new CRISPR spacers, detection of such events obviously implies that this did not result in a lethal phenotype, either due to a modified PAM and/or to an inactivated CRISPR–Cas effector module (*22*). Indeed, in the absence of the effector module, elevated frequencies of self-spacer acquisition occur in *Escherichia coli* (*149*). Similarly, *Streptococcus thermophilus* with a catalytically inactive Cas9 results in a major increase of spacers derived from the host genome (*150*). In addition, there is a strong preference for the integration of plasmid over chromosomal spacer sequences (*149*, *151*, *152*), with plasmid sequences incorporated more frequently than host DNA by two to three orders of magnitude (*153*). Spacer acquisition in *E. coli* requires active replication of the protospacer-containing DNA (*153*). Thus, small, fast-replicating plasmid genomes are a much better source of spacers than the large host DNA, and such findings are consistent with acquisition of spacers from an infecting virus genome in the archaeon *Sulfolobus islandicus* requiring its active replication (*154*). In *E. coli*, the CRISPR–Cas system derives the spacers primarily from products of RecBCD-catalysed DNA degradation that are formed during the repair of double-stranded breaks associated with stalled replication forks (*155*). Other possible sources of substrates for CRISPR adaptation include DNA fragments generated either by other defence systems, such

as restriction-modification systems (*156*), or by the CRISPR–Cas system itself (*147*).

Cas1 and Cas2 play crucial roles in spacer acquisition in all CRISPR–Cas systems (*148*, *149*). In addition, these proteins can function in trans, provided that the repeats involved are sufficiently similar in size and structure. Accordingly, *cas1* and *cas2* genes are missing in many active CRISPR–Cas loci – in particular, of type III as well as types IV and VI (*31*). Overexpression of Cas1 and Cas2 from the *E. coli* type I-E system has been shown to be sufficient for the extension of the CRISPR array (*149*). Mutations in the active site of Cas1 abolish spacer integration in *E. coli* (*149*), whereas the nuclease activity of Cas2 is dispensable (*152*). In *E. coli*, a central Cas2 dimer and two flanking Cas1 dimers form a complex that binds and processes PAM-containing DNA fragments (**Figure 2.3A**) (*152*, *157*), after which the newly generated spacers can be integrated into a CRISPR array via a recombination mechanism akin to that of retroviral integrases and transposases (*158*) (**Figure 2.3B**).

In several type III CRISPR–Cas systems, Cas1 is fused to reverse transcriptase (*28*), and it was recently shown that these systems are capable of acquisition of RNA spacers by direct incorporation of an RNA segment into the CRISPR array followed by reverse transcription and replacement of the RNA strand by DNA (*159*). Although the biological function of this process remains to be elucidated, these findings demonstrate remarkable versatility of adaptation pathways.

Spacer acquisition (adaptation) in type I systems proceeds along two distinct paths: (i) naïve acquisition, which occurs during an initial infection, and (ii) primed acquisition, when the CRISPR contains a previously integrated spacer that is complementary to the invading DNA (*160*). According to the proposed model, naïve spacer adaptation involves five steps (**Figure 2.3B**).

1. Fragmentation of (mainly) invasive nucleic acids by non-Cas systems [*e.g.*, by RecBCD after stalling a replication fork, or by restriction enzymes (restriction-modification systems)] (*153*, *156*), or by CRISPR-associated nucleases (*147*). Although this step may be non-essential, it probably enhances the efficiency of the overall process and its specificity toward invading DNA.

2. Selection of DNA fragments for (proto)spacers by scanning for potential PAMs (after partial target unwinding) by one of the four Cas1 subunits of the Cas1-Cas2 complex (*161*).

3. Measuring of the selected protospacer generating fragments of the correct size with 3′ hydroxyl groups by Cas1 nuclease.

4. Nicking of both strands of the leader-proximal repeat of the CRISPR array at the 5′ ends through a direct nucleophilic attack by the generated 3′ OH groups, resulting in covalent links of each of the strands of the newly selected spacer to the single-stranded repeat ends.

5. Second-strand synthesis and ligation of the repeat flanks by a non-Cas repair system (*146*, *158*).

Primed spacer adaptation so far has been demonstrated only in type I systems (*148*, *162*, *163*). This priming mechanism constitutes a positive feedback loop that facilitates the acquisition of new spacers from formerly encountered genetic elements (*164*). Priming can occur even with

spacers that contain several mismatches, making them incompetent as guides for targeting the cognate foreign DNA (*164*). Based on PAM selection, functional spacers are preferentially acquired during naïve adaptation. This initial acquisition event triggers a rapid priming response after subsequent infections. Priming appears to be a major pathway of CRISPR adaptation,

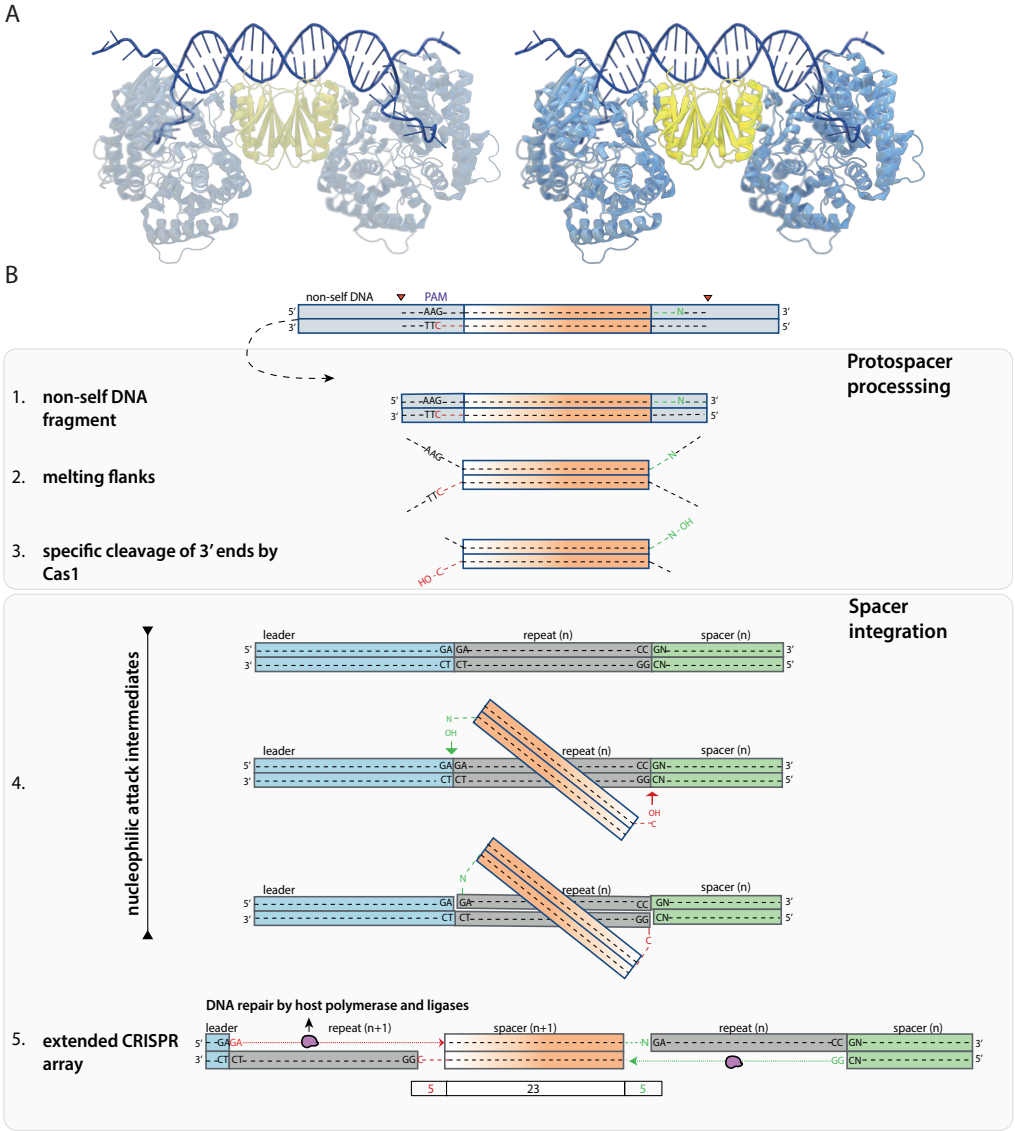


Figure 2.3 | Spacer acquisition. (A) Crystal structure of the complex of Cas1-Cas2 bound to the dual-forked DNA (PDB accession 5DQZ). The target DNA is shown in dark blue; the Cas1 and Cas2 dimers of the complex are indicated in blue and yellow, respectively. **(B)** Model explaining the capture of new DNA sequences from invading nucleic acid and the subsequent DNA integration into the host CRISPR array. The numbers on the left correspond to the order of events as described in the text. The dashed lines indicate nucleotides; the nucleotides C and N on the two sides of the protospacer are shown in red and green to clarify the orientation.

at least for some type I systems (*162*). Primed adaptation strongly depends on the spacer sequence (*165*), and the acquisition efficiency is highest in close proximity to the priming site. In addition, the orientation of newly inserted spacers indicates a strand bias, which is consistent with the involvement of single-stranded adaption intermediates (*166*). According to one proposed model (*167*), replication forks in the invader's DNA are blocked by the Cascade complex bound to the priming crRNA, enabling the RecG helicase and the Cas3 helicase/nuclease proteins to attack the DNA. The ends at the collapsed forks then could be targeted by RecBCD, which provides DNA fragments for new spacer generation (*167*). Given that the use of crRNA for priming has much less strict sequence requirements than direct targeting of the invading DNA, priming is a powerful strategy that might have evolved in the course of the host-parasite arms race to reduce the escape by viral mutants, to provide robust resistance against invading DNA, and to enhance self/nonself discrimination. Naïve as well as primed adaptation in the subtype I-F system of *Pseudomonas aeruginosa* CRISPR–Cas require both the adaptation and the effector module (*166*)

In the type II-A system, the Cas9-tracrRNA complex and Csn2 are involved in spacer acquisition along with the Cas1-Cas2 complex (*150*, *168*); the involvement of Cas9 in adaptation is likely to be a general feature of type II systems. Although the key residues of Cas9 involved in PAM recognition are dispensable for spacer acquisition, they are essential for the incorporation of new spacers with the correct PAM sequence (*168*). The involvement of Cas9 in PAM recognition and protospacer selection (*168*) suggests that in type II systems Cas1 may have lost this role. Similarly, Cas4 that is present in subtypes IA-D and II-B has been proposed to be involved in the CRISPR adaptation process, and this prediction has been validated experimentally for type I-B (*162*). Cas4 is absent in the subtype II-C system of *Campylobacter jejuni*. Nonetheless, a conserved Cas4-like protein found in *Campylobacter* bacteriophages can activate spacer acquisition to use host DNA as an effective decoy to bacteriophage DNA. Bacteria that acquire self-spacers and escape phage infection must either overcome CRISPR-mediated autoimmunity by loss of the interference functions, leaving them susceptible to foreign DNA invasions, or tolerate changes in gene regulation (*169*). Furthermore, in subtypes I-U and V-B, Cas4 is fused to Cas1, which implies cooperation between these proteins during adaptation. In type I-F systems, Cas2 is fused to Cas3 (*39*), which suggests a dual role for Cas3 (*28*): involvement in adaptation as well as in interference. These findings support the coupling between the adaptation and interference stages of CRISPR–Cas defence during priming.

Biogenesis of crRNAs

The short mature crRNAs contain spacer sequences, which are the guides that are responsible for the specificity of CRISPR–Cas immunity (*40*). They associate with one or more Cas proteins to form effector complexes that target invading MGEs through crRNA:target sequence–specific recognition. The CRISPR arrays are transcribed as long precursors, known as pre-crRNA, that may contain secondary structured elements (hairpins) in those cases where the CRISPR contains palindromic repeats. The processing of the pre-crRNA typically yields 30- to 65-nt mature crRNAs that consist of a single spacer flanked by a partial repeat at either one or both ends (*40*, *170*).

The pathways of crRNA biogenesis differ among the different CRISPR–Cas types. In class 1 systems, the Cas6 protein is critical for the primary processing of pre-crRNA. Cas6 is a metal-independent endoribonuclease that recognizes and cleaves a single phosphodiester bond in the repeat sequences of a pre-crRNA transcript (*40*, *171*, *172*). Members of the Cas6 family contain

two RRM-type RNA-binding domains. The primary cleavage by Cas6 results in crRNAs containing a repeat-derived 5′ "handle" of 8 nt with a 5′ hydroxyl group, followed by the complete spacer sequence and a repeat-derived 3′ handle of variable size that in some subtypes forms a hairpin structure with either a 3′-phosphate or a cyclic 2′3′-phosphate (*40*, *171*, *173*). The Cas6 family proteins show considerable structural variation that might reflect the cleavage specificity (*170*, *174*, *175*).

In type I-E and I-F systems, the Cas6 ribonuclease is a single-turnover enzyme that remains attached to the crRNA cleavage product. In these cases, Cas6 is a subunit of a multisubunit Cascade complex (*40*, *176*) (**Figure 2.4A**). In the type I-F systems, the crRNP complex consists of the crRNA, Cas6f, and Csy1, Csy2, and Csy3 proteins (*177*–*179*). In other systems (subtypes I-A, I-B, I-D, and III-A to III-D), Cas6 is not associated with the crRNA-processing complex. The absence of a Cas6 subunit in the complex correlates with the lack of a hairpin structure of the 3′ handle and a variable 3′ end. The absence of a *cas6* gene in type I-C is complemented by another double RRM-fold subunit, Cas5d, which has adopted the role of the endoribonuclease that in other subtypes is carried out by Cas6 (*180*). Some systems coexisting in the same species have been demonstrated to share the same set of guides; examples include type III-A (Csm) and type III-B (Cmr) of *Thermus thermophilus* (*181*), and type III-B (Cmr), type I-A (Csa), and type I-G (Cst) of *Pyrococcus furiosus* (*182*). Given that the type III loci usually lack *cas6* genes, a single stand-alone Cas6 nuclease is likely to be responsible for the supply of crRNAs to the type III complexes in *T. thermophilus* (*181*). In *P. furiosus*, Cas6 nuclease of type I generates the crRNAs from all CRISPR loci for the different coexisting complexes (*182*). Cas6-based processing of pre-crRNA in type III systems is typically followed by a sequence-unspecific trimming at the 3′ end (by RNases yet to be identified) to yield mature crRNAs with a defined 8-nt 5′ end and a variable 3′ end (*137*, *183*, *184*).

Type II systems use a unique mechanism for crRNA biogenesis whereby processing depends on Cas9, a host RNase III, and a tracrRNA that forms base pairs with the repeats of the pre-crRNA (*31*, *49*, *51*) (**Figure 2.4B**). The cleaved crRNA-tracrRNA hybrid is bound and stabilized by Cas9, triggering a conformational change toward a state compatible with target scanning, recognition, and interference (*49*, *51*, *52*). Trimming of the 5′ end of the crRNA probably occurs by a non-Cas RNase. The absence of type II systems in archaea is consistent with the absence of RNase III genes in most archaeal genomes (*185*). In the type II-C system of *Neisseria meningitidis*, short intermediate crRNA guides are transcribed from multiple promoters embedded within the repeats of the CRISPR array, implying that the system does not require RNase III (*186*) (**Figure 2.4C**). Expression of tracrRNA has also been demonstrated for the subtype V-B system, suggestive of a crRNA processing pathway analogous to that in type II. By contrast, in subtype V-A and type VI systems, no tracrRNA is co-expressed with the pre-crRNA (*76*, *83*). Class 2 CRISPR–Cas systems lacking tracrRNA can be expected to function using novel mechanisms of crRNA biogenesis, including processing by other host RNases or by the effector proteins themselves.

A third variant of guide maturation has recently been described for the Cas12a effector complex, a class 2 system that (unlike Cas9) does not associate with a tracrRNA. It has been demonstrated that Cas12a has an intrinsic RNase activity that allows for the primary processing of the pre-crRNA to crRNA guides with a 5′ hairpin (*187*). The biosynthesis of crRNAs by Cas12a system is metal-, sequence-, and structure-dependent (*187*). Secondary processing of CRISPR guides probably occurs via a non-Cas RNase; maturation of Cas9-associated guides occurs by trimming at the 5′ end (**Figure 2.4B**), whereas in Cas12a the 3′ flanks of the crRNA are

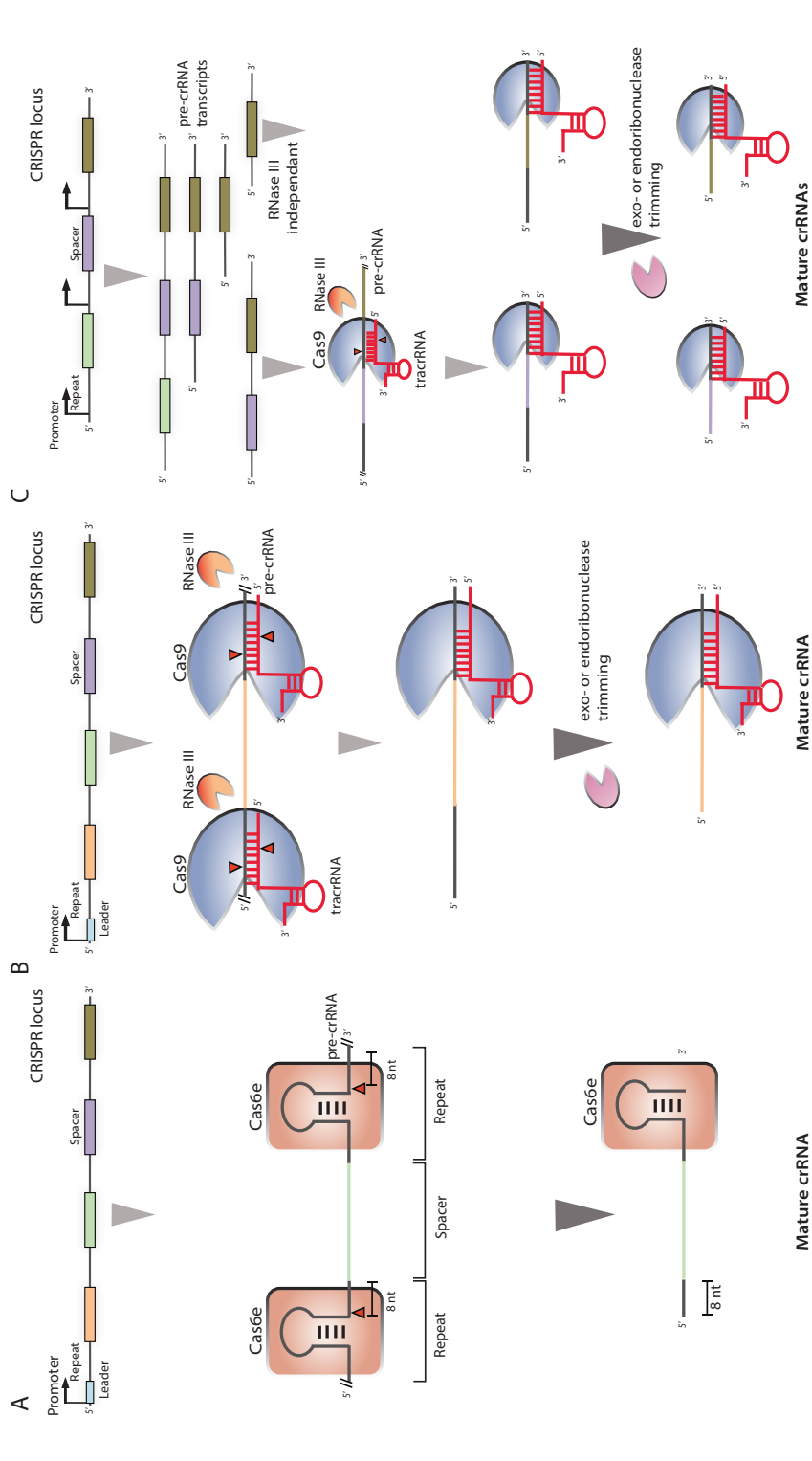


Figure 2.4 | Guide expression and processing. **(A)** Generation of CRISPR RNA (crRNA) guides in type I and type III CRISPR–Cas systems. Primary processing of the pre-crRNA is catalysed by Cas6, which typically results in a crRNA with a 5′ handle of 8 nt, a central spacer sequence, and (in some subtypes) a longer 3′ handle. Shown here is the guide processing (red triangles) for subtype I-E by Cas6e. The occasional secondary processing of the 3′ end of crRNA is catalysed by one or more unknown RNases. **(B)** In type II-A CRISPR–Cas systems, the repeat sequences of the pre-crRNA hybridize with complementary sequences of transactivating CRISPR RNA (tracrRNA). The double-stranded RNA is cleaved by RNase III (red triangles); further trimming of the 5′ end of the spacer is carried out by unknown RNase(s) (pink). **(C)** CRISPR with transcriptional start site (TSS) in repeats, as observed in type II-B CRISPR–Cas systems.

removed.

Target interference

Selection of CRISPR–Cas targets is a stepwise process that relies on recognition of a nonself sequence, a complementary spacer of which is stored in the CRISPR locus. In most cases, with the exception of the RNA-targeting type III systems, cognate protospacer sequences flanked by a PAM sequence are recognized by a CRISPR ribonucleoprotein (crRNP) complex [type I Cascade, type II Cas9, type V Cas12a (Cas12a) (**Figure 2.5**) and specifically degraded (*40*, *48*, *83*). In addition, selection of an appropriate target sequence depends on a so-called seed sequence on the guide (*95*, *176*). The seed is a sequence of seven or eight base pairs in close proximity to the PAM. Matching PAM and seed sequences are crucial for target interference (*95*, *176*, *188*) and act as a quality control step that is required for the complete displacement of the noncomplementary strand of the target DNA by the crRNA guide, the so-called R-loop conformation. Downstream of the seed region, mismatches between spacer and protospacer are tolerated to some extent (see below) (*95*).

In type I systems, the Cascade RNP complex scans DNA for complementary target sites, initially by identifying an appropriate PAM motif, followed by partial melting and base pairing by the guide's seed sequence, and eventually by formation of a complete R-loop structure (*173*, *189*). Upon reaching a PAM-proximal mismatch, the R-loop propagation stalls and the interference is aborted (*190*). When base pairing between guide and protospacer is complete, the R-loop structure appears to be locked in a state to license DNA degradation by the Cas3 nuclease/helicase (*39*, *40*, *190*).

Single-molecule experiments with *E. coli* Cascade demonstrate that crRNA-guided Cascade exhibits two distinct binding modes for matching and mismatched targets, which trigger either interference (matching target) or primed spacer acquisition (mismatched target). Unlike the interference of matching targets, mismatched targets are recognized with low fidelity, as indicated by a short-lived binding. The latter association is PAM- and seed-independent and can involve base pairing by any part of the crRNA spacer. In this case, the Cascade complex does not adopt a conformation that allows docking of Cas3 (*191*), precluding DNA interference. Instead, this Cascade-target complex primes the formation of a spacer acquisition complex that consists of Cas3 and Cas1-Cas2 and generates DNA fragments that are integrated as new spacers in the CRISPR array (*189*). These dual roles of Cascade allow for efficient degradation of bona fide targets and priming the acquisition of new spacers from mismatched targets (*e.g.*, from viral escape mutants) as an update of the CRISPR memory (*191*).

Although type III systems are structurally related to the type I system (**Figure 2.5**) (*137*, *138*, *192*–*197*), they show some substantial mechanistic variations. Initial analyses indicated that Csm (III-A) complexes target DNA (*43*), whereas Cmr (III-B) complexes target RNA (*44*, *198*, *199*). However, it has recently been demonstrated that both type III complexes are transcription-dependent DNA nucleases (*181*, *200*); that is, they initially recognize their target through specific interaction of the crRNA guide with a complementary nascent mRNA, after which cleavage of the flanking DNA sequences occurs (*201*–*206*). Robust interference by these systems relies on the concerted cleavage of the transcript RNA and the transcribed DNA. The Cas7-like backbone subunits (Csm3, Cmr4) are responsible for the RNase activity, typically resulting in cleavage of the target RNA at 6-nt intervals (*181*, *196*, *199*, *200*, *207*–*209*). Binding of the Cmr complex to its complementary RNA target induces a conformational change (*138*, *196*) that

results in activation of the Cas10 DNA-cleaving subunit (Csm1/Cmr2) (*202*, *203*, *205*). Disruption of the RNase active sites (in Csm3/Cmr4), at least in some cases, does not hamper the activation of the DNA nuclease activity of the complexes (*200*, *202*). Exonucleolytic cleavage of single-stranded DNA and RNA by recombinant *Staphylococcus epidermidis* Csm1 (Cas10) and by *Thermotoga maritima* and *P. furiosus* Cmr2 has been demonstrated *in vitro* (*202*, *203*, *210*). In the *S. epidermidis* system, a Csx1 ortholog (Csm6) provides an auxiliary RNA-targeting activity that operates in conjunction with the RNA- and DNA-targeting endonuclease activities of the Csm effector complex (*211*–*213*); in the *P. furiosus* Cmr system, Csx1 appears not to be an essential component (*200*). The relative contribution of the different nuclease subunits appears to vary in the different type III systems and under different conditions and awaits further characterisation.

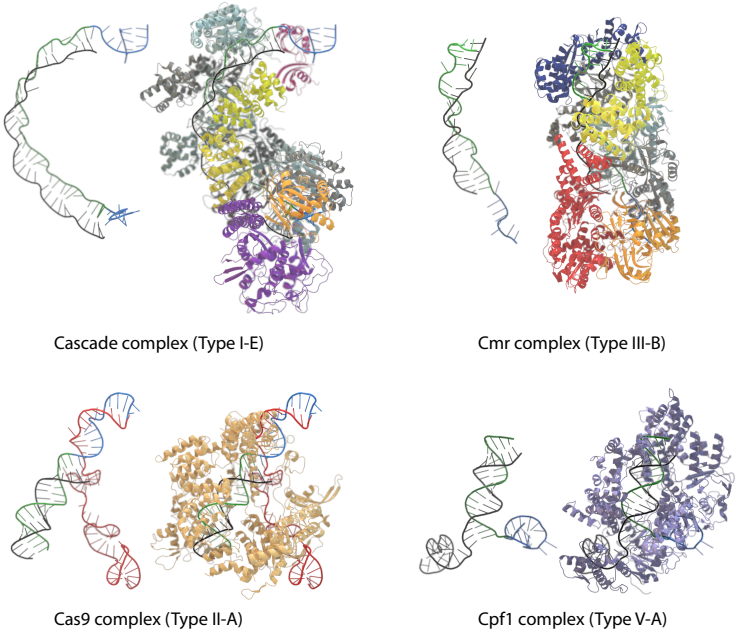


Figure 2.5 | CRISPR RNP complexes. Crystal structures of the CRISPR ribonucleoprotein (crRNP) complexes responsible for target interference. Shown are the type I-E Cascade complex (PDB accession 4QYZ) and type III-B Cmr complex (PDB accession 3X1L) from class 1, and the type II-A Cas9 complex (PDB accession 4OO8) and type V-A Cas12a (Cas12a) complex (PDB accession 5B43) from class 2. Colours of nucleic acid fragments are the same as in **Figure 2.6**.

Another unique feature of type III system concerns the mechanism of self/nonself discrimination. Genetic analyses have revealed that type III systems do not use the PAM-based "nonself-activation" mechanism of type I (Cascade), type II (Cas9), and type V (Cas12a). The mechanism used by the *S. epidermidis* Csm system apparently involves crRNA- or protein-based recognition of the repeats in the CRISPR locus, resulting in "self-inactivation" (*214*, *215*). However, the DNA cleavage activity of the *P. furiosus* Cmr complex was recently reported to require the presence of a short sequence adjacent to the target sequence within the activating target RNA (*i.e.*, an RNA PAM) (*203*). Additional analysis is required to reveal whether the reported motifs are typical features that distinguish the two subtypes.

Class 2 systems require only a single protein for interference. In type II, the crRNP complex involved in target recognition and degradation consists of Cas9 bound to the crRNA guide base-paired with the tracrRNA (*49*). The crystal structures of Cas9 reveal two distinct lobes that are involved in target recognition and nuclease activity (**Figure 2.5**). The positively charged groove at the interface of the two lobes accommodates the crRNA-DNA heteroduplex (*216*, *217*). A major step in Cas9 activation is the reorientation of the structural lobes upon crRNA/tracrRNA

loading, which results in the formation of a central channel that accommodates the target DNA (*216*). Binding and cleavage of the target DNA by the Cas9-crRNA effector complex depend on the recognition of an appropriate PAM located at the 3′ end of the protospacer (*188*), which serves as a licensing element in subsequent DNA strand displacement and R-loop formation. The PAM motif resides in a base-paired DNA duplex. Sequence-specific PAM readout by Arg^{1333} and Arg^{1335} in Cas9 positions the DNA duplex such that the +1 phosphate group of the target strand interacts with the phosphate lock loop (*218*). This promotes local duplex melting, allowing the Cas9-RNA complex to probe the identity of the nucleotides immediately upstream of the PAM. Base pairing between a 12-nt seed sequence of the guide RNA and the target DNA strand (*188*) drives further stepwise destabilization of the target DNA duplex and directional formation of the guide RNA–target DNA heteroduplex (*218*). This R-loop triggers a conformational change of the two nuclease domains (HNH and RuvC) of Cas9, which adopt an active state that allows for the completion of interference by cleavage of both target strands (*217*, *219*). Cas9 generates a blunt-end double-strand break, typically located 3 nt from the 3′ end of the protospacer (*48*, *220*). Recently, however, PAM-independent single-stranded targeting by *N. meningitidis* Cas9 has been described (*221*).

Similar to type II, the effector modules of type V systems consist of a large multidomain protein complex (Cas12a and Cas12b in subtypes V-A and V-B, respectively). Like Cas9, these proteins encompass a RuvC-like nuclease domain and an arginine-rich bridging helix. However, in contrast to Cas9, the RuvC-like domain of and Cas12b is more compact and the HNH domain is missing (**Figure 2.6**). Subtype V-B systems resemble type II with respect to the requirement for a tracrRNA, both for processing and for interference. In contrast, Cas12a-crRNA (type V-A) complexes are single RNA-guided endonucleases that cleave target DNA molecules in the absence of a tracrRNA (*83*). A model is proposed for a stepwise cleavage of the target DNA by (*i.e.*, initial RuvC-dependent cleavage of the displaced strand, followed by cleavage of the target strand by the novel nuclease domain) (*83*, *187*) suggests that the novel nuclease is allosterically activated by the RuvC cleavage event. Although allosteric control has also been demonstrated in interference by Cas9 (*219*), details appear to differ (*140*). Both and Cas12b from different bacteria efficiently cleave target DNA containing a well-defined T-rich PAM at the 5′ end of the protospacer (5′-PAM) (*76*, *83*), in contrast to the more variable, G-rich 3′-PAM sequence of Cas9 (*222*). Structural analysis has shown that recognizes its PAM through a combination of base and shape readout, in which several PAM-interacting amino acid residues that are conserved in the family are involved (*140*). Another unique feature of the endonuclease is the generation of staggered double-stranded DNA breaks with 4- or 5-nt 5′ overhangs (*83*); Another unique feature of the Cas12a endonuclease is the generation of staggered double-stranded DNA breaks with 4- or 5-nt 5′ overhangs (*140*) (**Figure 2.6**).

The type VI systems contain a unique effector protein (Cas13, formerly known as C2c2) with two HEPN domains. The *Leptotrichia shahii* Cas13 protein provides efficient interference against the RNA phage MS2. Cas13 is guided by a single crRNA and can be programmed to cleave ssRNA targets carrying complementary protospacers (*223*) (**Figure 2.6**). Spacers with a G immediately flanking the 3′ end of the protospacer are less fit relative to all other nucleotides at this position, which suggests that the 3′ protospacer flanking site (PFS) affects the efficacy of Cas13-mediated targeting (*224*) (**Figure 2.6**). Remarkably, once primed with the cognate target RNA, the Cas13 protein turns into a sequence-nonspecific RNase that causes a toxic effect in bacteria (*224*). Thus, the defence strategy of type VI systems appears to couple adaptive immunity with programmed cell death or dormancy induction.

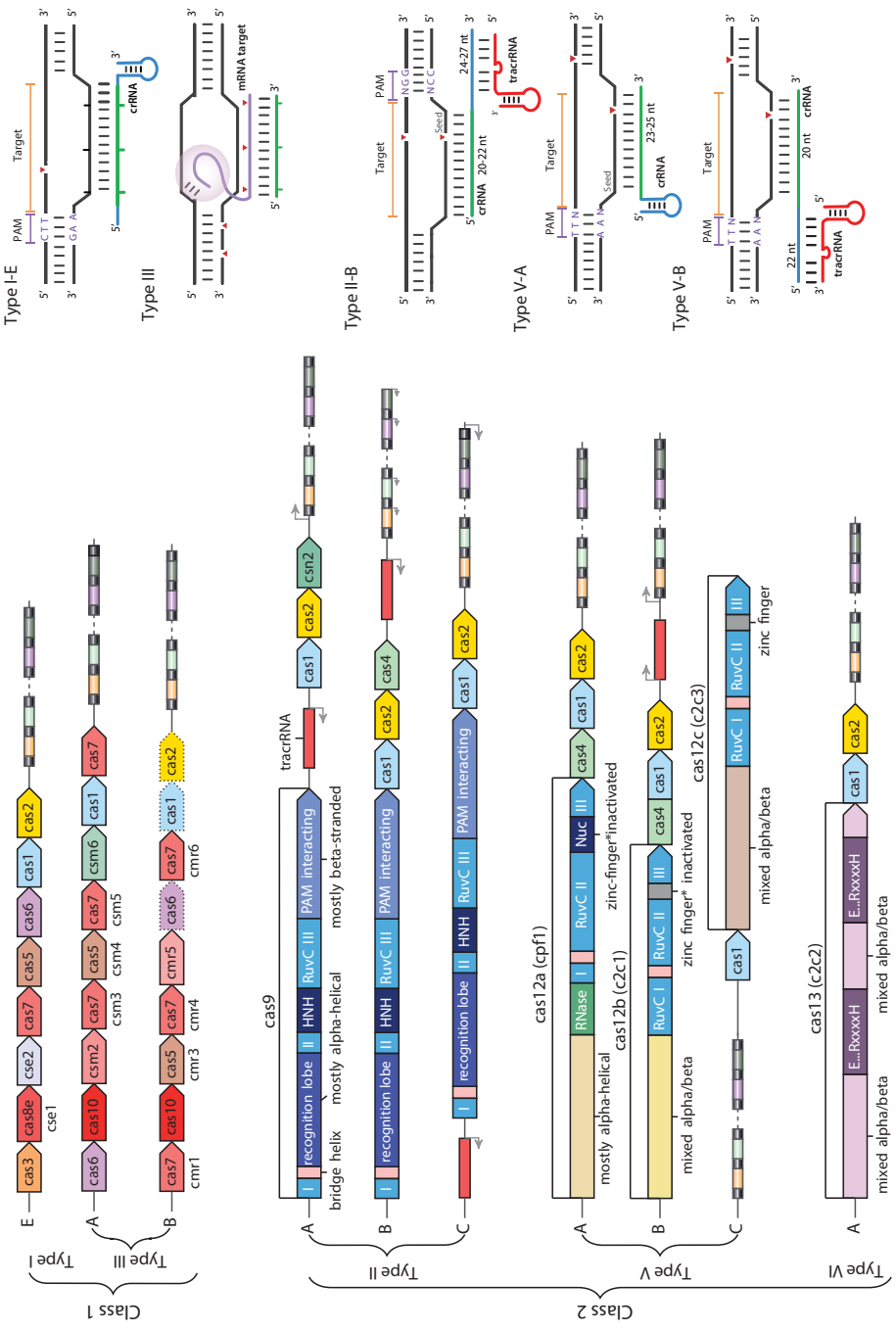


Figure 2.6 | Target interference. Genomic loci architecture of the components of class 1 and class 2 CRISPR–Cas systems and schematic representation of target interference for the different subtypes. The double-stranded DNA (target) is shown in black, the target RNA in grey, the CRISPR RNA (crRNA) repeat in blue, the spacer region of the crRNA in green, and the transactivating CRISPR RNA (tracrRNA) in red.

Phages are constantly evolving multiple tactics to avoid, circumvent, or subvert prokaryotic defence mechanisms (*16*). Phages can evade CRISPR interference through single-nucleotide substitution in the protospacer region or in the conserved protospacer-adjacent motif (*46*). Additionally, *P. aeruginosa* phages encode several proteins affecting the activity of type I-E and I-F systems (*94*). Diverse sequences of these proteins and mechanisms of action, coupled with the strong selection imposed by different antiviral systems, suggest an abundance of anti-CRISPR proteins yet to be discovered. Strikingly, some bacteriophages themselves encode a CRISPR–Cas system that in this case functions as an antidefense device targeting an anti-phage island of the bacterial host and thus enabling productive infection (*96*). Together, these findings emphasize the complexity of the virus-host arms race in which CRISPR–Cas systems are involved and suggest that many important aspects of this race remain to be characterised.

Very recently, an unexpected claim has been published on the existence of a CRISPR-like defence system in a giant mimivirus infecting unicellular eukaryotes (amoeba) (*225*). This system, named "mimivirus virophage resistance element" (MIMIVIRE), has been proposed to protect certain mimivirus strains from the Zamilon virophage, a small virus that parasitizes on mimiviruses. However, the MIMIVIRE locus lacks CRISPR-like repeats or a Cas1 homolog and encodes only very distant, generic homologs of two Cas proteins (a helicase and a nuclease that belong to the same protein super families as Cas3 and Cas4, respectively, but lack any specific relationship with these Cas proteins). Thus, any analogy between this putative eukaryotic giant virus defence system and CRISPR–Cas should be perceived with caution.

Genome editing applications

The molecular features of CRISPR–Cas systems, particularly class 2 systems with single-protein effectors, have made them attractive starting points for researchers interested in developing programmable genome editing tools. In 2013, the first reports of harnessing Cas9 for multiplex gene editing in human cells appeared (*57*, *58*, *226*, *227*). These studies have demonstrated that Cas9 could efficiently create indels at precise locations and that by supplying exogenous repair templates, insertion of a new sequence at target sites could be achieved via homologous recombination. A "dead" Cas9 (dCas9) variant with inactivating mutations in the HNH and RuvC domains binds DNA without cutting, providing a programmable platform for recruiting different functional moieties to target sites. The dCas9 has been used for transcriptional activation and repression (*61*, *73*, *228*, *229*), localizing fluorescent protein labels (*66*), and recruiting histone modifying enzymes (*67*, *230*). Other applications of Cas9 include building gene circuits (*231*–*233*), creating new anti-microbials (*234*) and antivirals (*235*–*237*), and large-scale gain- and loss-of-function screening (*69*, *70*, *238*, *239*).

The genome editing toolbox has been expanding through the discovery of novel class 2 effector proteins, such as Cas12a (*83*). The Cas12a nuclease possesses on-target efficiencies in human cells that are comparable with that of Cas9. Besides, Cas12a is also highly specific in its targeting, as minimal or no off-target cleavage has been detected (*240*, *241*). Cas12a does not require a tracrRNA, further simplifying the system for genome editing applications. In addition, it generates sticky ends, which could potentially increase the efficiency of insertion of new DNA sequences relative to the blunt ends created by Cas9 (*83*).

Central to the success of any Cas-based genome editing tool is the specificity of the enzyme, and many approaches to increase specificity have been reported. For example, "double-nicking," which uses dimers of two Cas9 variants, each mutated to create a nick in one strand of the

DNA, improves specificity by requiring two target matches to create the double-strand break (*242*, *243*). Another tactic is to control the amount of Cas9 in the cell via an inducible system that expresses a low level of Cas9 (*83*, *244*). Shortening the region of complementarity in the guide RNA also reduces off-target cleavage (*245*). Finally, structure-guided engineering has been used to mutate specific residues in Cas9, to weaken the interaction with the nontarget strand or to decrease nonspecific interactions with the target DNA site, favouring cleavage at sites that are perfectly complementary to the guide RNA and reducing off-target effects to undetectable levels at many sites (*246*, *247*).

A major outstanding challenge for realizing the full potential of Cas-based genome editing, including its use as a therapeutic, is efficient and tissue-specific delivery. Some progress has been made in this area, including the use of a smaller Cas9 ortholog (*248*), which is more amenable to packaging into viral vectors. Other approaches are also being pursued, including non-viral methods for delivery of DNA or mRNA by nanoparticles (*249*) and electroporation (*250*), or direct delivery of Cas9 protein (*251*). Additionally, the long-term effects of Cas9 expression in heterologous eukaryotic cells remain unexplored. Finally, the potential for editing the human genome as well as the possibility of using Cas-based gene drives for ecosystem engineering (*252*) raise ethical concerns that must be fully considered.

Outlook

The intensive research over the past few years on structural and functional features of variant CRISPR–Cas systems has revealed that they encompass many homologous components and share common mechanistic principles but also show enormous variability. A key aspect of this variability is module shuffling, which involves frequent recombination of adaptation and effector modules coming from different types of CRISPR–Cas within the same locus. Apart from major differences in the architectures of the effector complexes, functional diversity of CRISPR–Cas includes versatile mechanisms of crRNA guide processing, self/nonself discrimination, and target cleavage. The versatility of class 2 systems in particular, where distinct subtypes apparently evolved via independent recombination of adaptation modules with widely different effectors, is notable, given the potential of these systems as genome editing tools. The in-depth analysis of a few well-characterised CRISPR systems has revealed key structural and mechanistic features. However, the continuing discovery of novel CRISPR–Cas variants and new molecular mechanisms implies that our current insights have limited power for predicting functional details of distantly related variants. Hence, such new CRISPR–Cas systems need to be meticulously analysed to understand the biology of prokaryotic adaptive immunity and harness its potential for biotechnology. In this Review, we could not cover in any detail several fascinating aspects of CRISPR–Cas biology, such as coevolution of immune systems with viruses, the interplay between CRISPR–Cas activity and horizontal gene transfer, or nonimmune functions of CRISPR–Cas. The complexity and extreme variability of the CRISPR–Cas systems ensure that researchers in this field will have much to do for many years to come.

Acknowledgments

E.V.K. and K.S.M. are supported by the intramural program of the U.S. Department of Health and Human Services (to the National Library of Medicine). J.v.d.O. and P.M. are supported by the NWO/TOP grant 714.015.001. F.Z. is a New York Stem Cell Foundation–Robertson Investigator. F.Z. is supported by the NIH through NIMH (5DP1-MH100706 and 1R01-MH110049) and NIDDK (5R01DK097768-03), the New York Stem Cell, Simons, Paul G. Allen Family, and

Vallee Foundations, and D. R. Cheng, T. Harriman, and R. Metcalfe. F.Z. is a founder of Editas Medicine and scientific advisor for Editas Medicine and Horizon Discovery.

Author information

Correspondence should be addressed to john.vanderoost@wur.nl.

Chapter 3

Heterologous expression and purification of CRISPR–Cas12a/Cpf1

Prarthana Mohanraju, John van der Oost, Martin Jinek, Daan C. Swarts

Adapted from:
Heterologous Expression and Purification of CRISPR-Cas12a/Cpf1.
2018, Bio-Protocol, 8(9).

Abstract

This protocol provides step by step instructions for heterologous expression of Francisella novicida *Cas12a (previously known as Cpf1) in* Escherichia coli. *It additionally includes a protocol for high-purity purification and briefly describes how activity assays can be performed. These protocols can also be used for purification of other Cas12a homologs and the purified proteins can be used for subsequent genome editing experiments.*

Introduction

Prokaryotic CRISPR–Cas immune systems provide protection against viruses and plasmids by using CRISPR RNAs (crRNAs) as a guide for sequence-specific targeting of foreign DNA or RNA (*215*, *253*). Class 1 CRISPR–Cas systems (comprising types I, III, and IV) typically form multi-subunit protein-crRNA effector complexes, while the class 2 systems (comprising types II, V, and VI) rely on single crRNA-guided effector nucleases for target interference (*29*).

Effector nuclease enzymes from the Class 2 CRISPR–Cas systems have emerged as efficient and precise tools for genome editing and gene expression control (*222*, *242*, *254*). The widely used Cas9, which is the signature protein of type II systems, utilizes a dual guide RNA structure consisting of crRNA and a *trans*-activating crRNA (tracrRNA) for target recognition (*49*). For genome editing purposes, the dual guide RNA is often replaced by a synthetic fusion of the mature crRNA and tracrRNA, resulting in a long single-molecule guide RNA (sgRNA) in which the individual RNAs are fused by a short linker sequence (*51*). The sequence of the guide RNA allows binding of complementary DNA targets by base pairing with the target strand, while the other strand of the DNA is displaced. Upon finding a cognate DNA target, the HNH and RuvC nuclease domains of Cas9 mediate cleavage of the target and the displaced strand, respectively (*51*, *255*).

More recently, another novel class 2 CRISPR–Cas nuclease with distinctive features has been identified in bacterial genomes: Cas12a (also known as Cpf1) (*31*, *82*, *83*). Cas12a utilizes a single crRNA guide for DNA targeting; it does not require a tracrRNA, resulting in a shorter gRNA sequence compared to the chimeric single-molecule guide RNAs (sgRNA) used by Cas9. While Cas9 requires RNase III-mediated processing of pre-crRNA or individual expression of sgRNAs for the formation of mature guide RNAs, Cas12a can process its own pre-crRNA. This pre-crRNA processing activity allows for simple multiplexing in Cas12a-mediated genome editing (*256*, *257*). Whereas Cas9 generates double stranded DNA breaks (DSBs) that are blunt ended, Cas12a generates staggered-end DSBs (*83*). Such overhangs can be utilized for overhang-based cloning (*258*, *259*). Moreover, Cas9 typically recognizes a G-rich PAM sequence, while all Cas12a orthologues characterised to date recognize a T-rich PAM sequence (*83*). Taken together, these features make Cas12a a valuable addition to the genome editing toolbox.

Cas12a has been successfully repurposed for genome editing applications in mammalian cells (*83*, *241*), mice (*260*, *261*), rice (*262*, *263*), yeast (*264*, *265*), zebrafish, *xenopus* (*266*), microalga (*267*) and plant cells (*268*, *269*). The high efficiency and specificity of Cas12a in human cells, coupled with fewer off-target cleavage events compared to Cas9 (*246*), makes Cas12a a robust and reliable tool for genome editing.

For its *in vitro* characterisation and crystallization (*270*), Cas12a from *Francisella novicida* U112 was purified after heterologous expression in *Escherichia coli*. The expression strain *E. coli* Rosetta™ 2 (DE3) carries a chromosomal T7 RNA polymerase gene under control of an IPTG inducible *lacUV5* promoter. The *cas12a* gene is expressed using a pET vector (*271*–*273*) with a *lacI*-controlled T7 promoter. Here we describe the steps (**Figure 3.1**) required for controlled expression and purification of FnCas12a. The protocol can also be used for the expression and purification of Cas12a homologs from *Acidaminococcus* sp. and *Lachnospiraceae bacterium*.

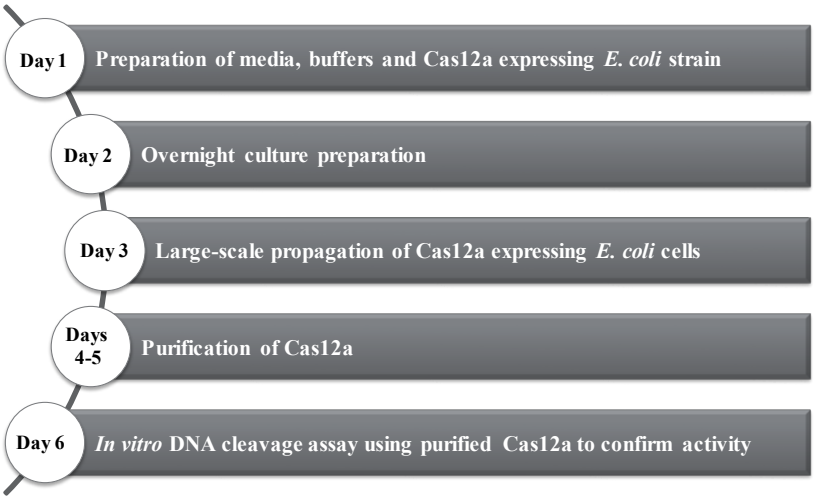


Figure 3.1 | Timeline of activities for the heterologous expression and purification of *Francisella novicida* Cas12a (FnCas12a) from *Escherichia coli*.

Materials and Reagents

Note: Equivalent materials and reagents may be used as substitutes.

Expression of FnCas12a in *E. coli* Rosetta™ 2(DE3)

1. 100-mL Erlenmeyer flask (DWK Life Sciences, DURAN®, catalogue number: 21 213 24)
2. 2-L Erlenmeyer flask (DWK Life Sciences, DURAN®, catalogue number: 21 216 63)
3. 5-L Erlenmeyer flasks (DWK Life Sciences, DURAN®, catalogue number: 21 216 73)
4. 50-mL conical centrifuge tubes (Sigma-Aldrich, catalogue number: T2318-500EA)
5. 2-mL screw top tube (Corning, catalogue number: 430659)
6. Nalgene™ PPCO Centrifuge Bottles with Sealing Closure (Thermo Fisher Scientific, Thermo Scientific™, catalogue number: 3141-0500) or equivalent 500-mL centrifuge bottles
7. Pipette tips (DeckWorks™ standard pipet tips, Corning, catalogue numbers: 4110; 4112; 4867)
8. 10-mL syringe (BD, catalogue number: 309604)
9. 0.22 µm syringe filter (Mdi, catalogue number: SYPL0601MNXX204)
10. 250-mL bottle (Greiner Bio One International, catalogue number: 227261)
11. *Escherichia coli* Rosetta™ 2(DE3) cells (Merck, Novagen, catalogue number: 71400) [encodes a T7 RNA polymerase gene under control of a *lacUV5* promoter]
12. Plasmid pDS015[a] [pET His6 TEV LIC cloning vector (Addgene, catalogue number: 29653), with *F. novicida* U112 *cas12a* gene insert fused to an N-terminal His-tag; expression under the control of a *lacI*-controlled T7 promoter]
13. Tryptone (Thermo Fisher Scientific, Thermo Scientific™, catalogue number: LP0042B)
14. Yeast extract (BD, Bacto™, catalogue number: 212720)

[a]Acidaminococcus sp. *BV3L6 Cas12a (AsCas12a) and* Lachnospiraceae bacterium *ND2006 Cas12a (LbCas12a) proteins can also be purified using this protocol with expression vectors 6His-MBP-TEV-huAsCpf1 (Addgene, catalogue number: 90095) and 6His-MBP-TEV-huLbCpf1 (Addgene, catalogue number: 90096)*

15. Sodium chloride (NaCl) (Fisher Scientific, catalogue number: S271-10)
16. Sodium hydroxide (NaOH) (Merck, EMD Millipore, catalogue number: 106462)
17. Ethanol (Fisher Scientific, catalogue number: BP2818500)
18. Chloramphenicol (Fisher Scientific, catalogue number: BP904100)
19. Kanamycin sulphate (Thermo Fisher Scientific, catalogue number: 11815024)
20. Glycerol (Fisher Scientific, catalogue number: BP229-4)
21. IPTG (Fisher Scientific, catalogue number: BP1755-1)
22. Agar (Acros Organics, catalogue number: 400400050)
23. LB medium (see Section Recipes)
24. 1,000x chloramphenicol solution (34 mg mL^{-1}) (see Section Recipes)
25. 1,000x kanamycin solution (50 mg mL^{-1}) (see Section Recipes)
26. 1 M IPTG (IsoPropyl-1-Thio-β-D-Galactopyranoside) (see Section Recipes)
27. Glycerol stock (50% solution) (see Section Recipes)

Purification of FnCas12a

1. 5 mL HisTrap HP (GE Healthcare, catalogue number: 17524701)
2. Dialysis tubing, high retention seamless cellulose tubing, avg. flat width 23 mm (0.9 in.), MWCO 12,400, 99.99% retention (Sigma-Aldrich, catalogue number: D0405)
3. Dialysis tubing clamps (Sigma-Aldrich, catalogue number: Z371092)
4. 5 mL HiTrap Heparin HP (GE Healthcare, catalogue number: 17040601)
5. Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-100 membrane (Merck, EMD Millipore, catalogue number: UFC9100)
6. HiLoad 16/600 Superdex 200 pg (GE Healthcare, catalogue number: 28989335)
7. Gosselin™ Round-Base 10-mL Test Tubes (Thermo Fisher Scientific, Fisher Scientific, catalogue number: 11723284) or other equivalent fraction collection tubes
8. Nalgene™ Oak Ridge High-Speed Centrifuge Tubes (Thermo Fisher Scientific, Thermo Scientific™, catalogue number: 3114-0050) or equivalent 50-mL centrifuge tubes
9. Membrane Filter, mixed cellulose esters (Merck, MF-Millipore, catalogue number: HAWP04700)
10. Membrane Filter, mixed cellulose esters (Merck, MF-Millipore, catalogue number: GSWP04700)
11. Cell pellet from overnight culture in which FnCas12a was expressed (from Procedure A)
12. cOmplete™, EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich, Roche Diagnostics, catalogue number: 11873580001)
13. Lysozyme from chicken egg white (Sigma-Aldrich, catalogue number: L6876-5G)
14. β-Mercaptoethanol (Sigma-Aldrich, catalogue number: M6250)
15. TEV protease (Sigma-Aldrich, catalogue number: T4455)
16. 12% Mini-PROTEAN® TGX™ Precast Protein Gels (Bio-Rad Laboratories, catalogue number: 4561043)
17. 4x Laemmli protein sample buffer for SDS-PAGE (Bio-Rad Laboratories, catalogue number: 1610747)
18. Bio-Safe™ Coomassie Stain (Bio-Rad Laboratories, catalogue number: 1610786)
19. PageRuler Prestained Protein Ladder, 10 to 250 kDa (Thermo Fisher Scientific, Thermo Scientific™, catalogue number: 26619)

20. Dithiothreitol (DTT) (Sigma-Aldrich, catalogue number: D0632)
21. Ethylenedinitrilotetraacetic acid (EDTA) (Sigma-Aldrich, catalogue number: E9884)
22. Sodium chloride (NaCl) (Fisher Scientific, catalogue number: S271-10)
23. Tris (Thermo Fisher Scientific, Thermo Scientific™, catalogue number: 17926)
24. Imidazole (Sigma-Aldrich, catalogue number: I0250)
25. Hydrochloric acid (HCl) (Sigma-Aldrich, catalogue number: 258148)
26. Potassium chloride (KCl) (Merck, EMD Millipore, catalogue number: 104933)
27. HEPES (Sigma-Aldrich, catalogue number: H3375)
28. Potassium hydroxide (KOH) (Sigma-Aldrich, catalogue number: 757551)
29. Glycine (Sigma-Aldrich, catalogue number: G8898)
30. Sodium dodecyl sulphate (SDS) (Sigma-Aldrich, catalogue number: L3771)
31. 1 M DTT (Dithiothreitol) stock (see Section Recipes)
32. 0.5 M EDTA (Disodium Ethylene Diamine Tetra-Acetate) stock (pH 8) (see Section Recipes)
33. Lysis Buffer (see Section Recipes)
34. Wash Buffer (see Section Recipes)
35. Elution Buffer (see Section Recipes)
36. Dialysis Buffer (see Section Recipes)
37. Dilution Buffer (see Section Recipes)
38. IEX-A Buffer (see Section Recipes)
39. IEX-B Buffer (see Section Recipes)
40. SEC Buffer (see Section Recipes)
41. 10x SDS-PAGE Electrophoresis Running Buffer (see Section Recipes)

Activity assay using purified FnCas12a

1. Purified Cas12a Nuclease (from Procedure B)
2. Nuclease-free water
3. Proteinase K, Molecular Biology Grade (New England Biolabs, catalogue number: P8107S)
4. crRNA containing the targeting sequence complementary to the target DNA
   Note: The RNA can be ordered as a desalted RNA oligonucleotide or as PAGE-purified RNA oligonucleotide from an RNA synthesis company such as Sigma-Aldrich or IDT.
5. DNA substrate containing the target sequence and a 5'-TTTN- 3' PAM sequence. *Note: The substrate DNA can be circular or linearized plasmid, PCR products, or synthesized oligonucleotides). As an example, the DNA substrate and crRNA used in the activity assay is shown in* ***Figure 3.2***
6. GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific, Thermo Scientific™, catalogue number: SM0311) or equivalent
7. DNA gel Loading Dye [*e.g.*, 6x DNA Loading Dye (Thermo Fisher Scientific, Thermo Scientific™, catalogue number: R0611)]
8. Invitrogen™ SYBR™ Safe DNA Gel Stain (Thermo Fisher Scientific, Thermo Scientific™, catalogue number: S33102)
9. Sodium chloride (NaCl) (Fisher Scientific, catalogue number: S271-10)
10. Magnesium chloride hexahydrate ($MgCl_2 \cdot 6H_2O$)
11. HEPES (Sigma-Aldrich, catalogue number: H3375)
12. Ethylenedinitrilotetraacetic acid (EDTA) (Sigma-Aldrich, catalogue number: E9884)

13. Hydrochloric acid (HCl) (Sigma-Aldrich, catalogue number: 258148)
14. 10x Nuclease Reaction Buffer (see Section Recipes)

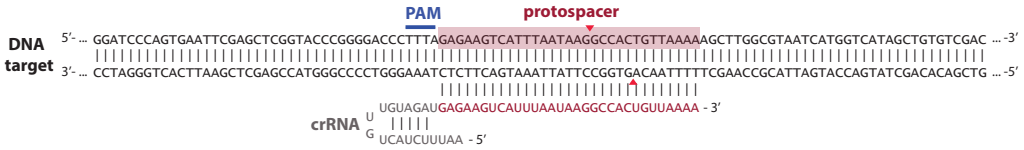


Figure 3.2 | Schematic of the Cas12a crRNA-DNA-targeting complex. The expected cleavage sites are indicated by red arrows.

3

Equipment

Expression of FnCas12a in *E. coli* Rosetta™ 2

Note: Equivalent equipment can be used.

1. Pipettes (Corning, model: Lambda™ Plus Single-Channel Pipettor, catalogue numbers: 4070; 4074; 4075)
2. New Brunswick™ Innova® 42 incubator (Eppendorf, New Brunswick™, model: Innova® 42, catalogue number: M1335-0002) or an equivalent incubator that can be set at 37 ℃
3. Sorvall LYNX 4000 Superspeed Centrifuge (Thermo Fisher Scientific, Thermo Scientific™, model: Sorvall LYNX 4000, catalogue number: 75006580) or an equivalent centrifuge that can be cooled down to 4 ℃ and can perform up to 6,000xg
4. New Brunswick™ Innova® 44/44R (Eppendorf, New Brunswick™, model: Innova® 44/44R, catalogue number: M1282-0002) or any equivalent shaker incubator where the temperature can be set at 37 ℃ and 18 ℃
5. Cell density meter (GE Healthcare, model: Ultrospec™ 10, catalogue number: 80-2116-30), or equivalent spectrophotometer that can measure the density of cells in suspension at 600 nm
6. Ice-water bath (water and ice mixed)

Purification of FnCas12a

1. SONOPULS HD (Bandelin electronic, model: HD 3200) with VS 70 T Sonotrode (Bandelin) or equivalent ultrasonic homogenizer/Sonifier, or alternatively a French Pressure Cell (French Press) for cell lysis
2. Peristaltic pump P-1 with connectors for 5 mL HisTrap HP (GE Healthcare, model: Peristaltic Pump P-1, catalogue number: 18111091) or an equivalent peristaltic pump
3. Tubing Connectors for Use with Peristaltic Pump P-1 (GE Healthcare, catalogue number: 11300082)
4. ÄKTApurifier 10 FPLC system (GE Healthcare, model: ÄKTApurifier 10, catalogue number: 28406264) or an equivalent FPLC system
5. Sorvall LYNX 4000 Superspeed Centrifuge (Thermo Fisher Scientific, Thermo Scientific™, model: Sorvall LYNX 4000, catalogue number: 75006580) or an equivalent centrifuge that can be cooled down to 4 ℃ and can perform up to 30,000xg
6. pH meter (QiS, model number: B210)

7. Filter holder assembly for filtration (Merck, catalogue number: XX1014700 or Nalgene, Thermo Fisher Scientific, Thermo Scientific™, catalogue number: DS0320-2545), or equivalent filter holder assembly
8. Diaphragm Vacuum Pumps LABOPORT®N 820 (ABM van Zijl B.V, catalogue number: ABMK N8203FT18), or an equivalent vacuum pump
9. Nanodrop (Thermo Fisher Scientific, Thermo Scientific™, model: NanoDrop™ 2000, catalogue number: ND-2000)
10. Mini-PROTEAN Tetra cell (Bio-Rad Laboratories, model: Mini-PROTEAN Tetra Cell, catalogue number: 1658004EDU), or an equivalent vertical electrophoresis system
11. Epson Perfection V850 Pro scanner (Epson, model: Perfection V850 Pro) or equivalent scanner or imager suitable for SDS-PAGE gel imaging

Activity assay using purified FnCas12a

1. Eppendorf™ 5424 Microcentrifuge (Eppendorf, model: 5425, catalogue number: 022620498)
2. MUPID One Horizontal Electrophoresis System (Bulldog Bio, catalogue number: MU2) or an equivalent horizontal electrophoresis system
3. G:BOX F3 (Syngene, model: G:BOX F3, catalogue number: 05-GBOX-F3) gel doc system or equivalent DNA agarose gel imaging equipment

Procedure

Transformation of *E. coli* Rosetta™ 2(DE3) with pDS015 plasmid and preparation of a glycerol stock

1. Add 1 ng of pDS015 plasmid DNA directly to 50 µL of chemically competent *E. coli* Rosetta™ 2(DE3) cells.
2. Stir gently to mix and place the tubes on ice for 5 min.
3. Heat the tubes for exactly 30 sec in a 42 ℃ water bath; do not shake.
4. Immediately place the tube on ice for 2 min.
5. Add 250 µL of room temperature sterile SOC medium (provided by the manufacturer) to the tube. Incubate at 37 ℃ while shaking at 250 rpm for 60 min.
6. Spread 5-50 µL of the transformation mixture on LB agar plates containing 34 µg mL^{-1} and 50 µg mL^{-1} of chloramphenicol and kanamycin, respectively. If plating less than 25 µL of the transformation, we recommend adding 50 µl of sterile SOC medium to the transformation mixture before plating to facilitate even colony distribution on the LB agar plate surface.
7. Incubate the LB agar plates overnight at 37 ℃.
8. The next day, pick a single colony from the transformation plates using a sterile pipette tip and inoculate 10 mL LB in a 50-mL tube.
9. Incubate the 50-mL tube overnight in a 37 ℃ shaking incubator shaking at 160 rpm.
10. The next day, add 500 µL of the overnight culture to 500 µL of 50% sterile glycerol in a 2-mL screw-top tube or cryovial and mix gently.
11. Store the glycerol stock at -80 ℃ for future use.

Large-scale expression of FnCas12a in *E. coli* Rosetta™ 2(DE3)

Day 1. Preparation of media, buffers and single colonies

1. Prepare 20 mL of LB in a 100-mL Erlenmeyer flask (for starting overnight cultures) (Recipe A1).
2. Prepare three 5-L Erlenmeyer flasks, each containing 1.5 L of LB medium (for large-scale propagation and protein purification) (Recipe A1).
3. Prepare antibiotic and stock solutions (Recipes A2 and A3).
4. Prepare the buffers needed for the purification (Recipes B1-B8).
5. Streak out a glycerol stock of *E. coli* Rosetta™ 2(DE3) transformed with pDS015 on a LB agar plate containing 50 μg mL^{-1} kanamycin and 34 μg mL^{-1} chloramphenicol.
6. Incubate the LB agar plate overnight in a 37 ℃ incubator.

Day 2. Overnight culture preparation

1. Add 20 μL of the 50 mg mL^{-1} kanamycin stock solution and 20 μL of the 34 mg mL^{-1} chloramphenicol stock solution to the 20 mL of autoclaved LB medium in a 100-mL Erlenmeyer flask from *Day 1*.
2. With a sterile pipette tip, pick a single colony of *E. coli* Rosetta™ 2(DE3) transformed with pDS015 from the LB agar plate from *Day 1*.
3. Use the colony to inoculate the medium containing the kanamycin and chloramphenicol.
4. Loosely close the 100-mL Erlenmeyer flask with a cotton plug.
5. Incubate the bacterial culture at 37 ℃ for 16-20 h in a shaking incubator (set at 160 rpm).

Day 3. Large-scale propagation of cells overexpressing FnCas12a

1. Take three autoclaved 5-L Erlenmeyer flasks[b] each containing 1.5 L LB medium from *Day 1*.
2. To each flask, add 1.5 mL of 50 mg mL^{-1} kanamycin solution.
3. To each flask, add 1.5 mL of 34 mg mL^{-1} chloramphenicol solution.
4. To each flask, add 15 mL of the overnight culture prepared on *Day 2*.
5. Incubate the culture flasks at 37 ℃ in a shaking incubator at 160 rpm[c].
6. Monitor the $OD_{600\ nm}$ of the culture every half an hour. Once an $OD_{600\ nm}$ of 0.5-0.6 is reached (this normally takes ~3-4 h), transfer the Erlenmeyer containing the culture to the ice-water bath and incubate (cold-shock) it for 15 min. This step slows down the metabolism of *E. coli* and triggers expression of cold-shock proteins which may aid FnCas12a folding during expression.
7. To each flask, add 200 μL of filter-sterilized 1 M IPTG solution to the culture to induce expression of FnCas12a.
8. Transfer the culture to an 18 ℃ shaking incubator (set at 120 rpm) for overnight expression (~16 h).

[b] *It is also possible to express smaller volumes of cell culture (e.g., a single 1.5 L culture or one or more 750 mL cultures in 2-L Erlenmeyer flasks)*

[c] *When using a baffled Erlenmeyer flask, reduce the shaking incubator speed to 120 rpm to prevent the formation of foam.*

Day 4–Part I. Large-scale propagation of cells overexpressing FnCas12a (continued)

1. Transfer the overnight culture from *Day 3* to centrifuge bottles.
2. Harvest the cells by centrifuging the culture for 15 min at 6,000xg at 4 ℃.
3. Discard supernatant and store the pelleted cells at -20 ℃ (for use within a week for optimal purification) or at -80 ℃ (for long-term storage) or proceed directly to purification. The expected yield is ~5 g of cell pellet per liter of cell culture.

Day 4–Part II. Purification of FnCas12a–Part I

1. If continuing with a frozen cell pellet, thaw the cell pellet from *Day 4–Part I* on ice for 30-60 min. If proceeding directly after protein expression, skip this step. *Note: All subsequent steps should be performed on ice or at 4 ℃.*
2. Resuspend the entire cell pellet in Lysis Buffer (~2.5-5 mL Lysis Buffer per gram of cell pellet).
3. Add 1 tablet cOmplete™ protease inhibitor for every 50 mL.
4. Add lysozyme to a final concentration of 1 mg mL^{-1}.
5. Incubate the sample on ice for 30 minutes (min).
6. If using a French Press for cell lysis: after the lysozyme treatment, pass cell suspension through French Press twice at 16,000 psi.
7. If using sonication for cell lysis: after the lysozyme treatment, lyse the cell suspension by using a sonicator with an appropriate tip and a protocol suitable for lysis of large volume cell suspensions. For our setup (Bandelin SONOPULS HD with VS 70 T tip), we use the following settings: 10 min total time, 1 sec on, 0.7 sec off, and 20% amplitude. The cell suspension often has a brownish tinge after lysis as shown in **Figure 3.3**.

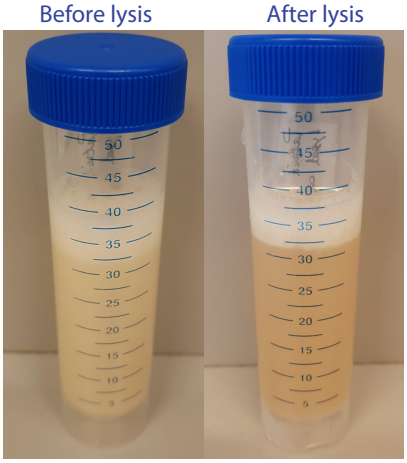


Figure 3.3 | After lysis, the cell suspension becomes tinted brown and less viscous.

Note: Using a French Pressure cell or sonicator gives approximately the same yield – FnCas12a is very stable and little to no protein will be lost during sonication. Keep in mind, however, that sonication is usually less suitable for large volumes, and therefore a protocol suitable for lysis of a large volume of cell suspension should be applied.

8. Pour the lysate into (a) centrifugation tube(s) and centrifuge for 45 min, 4 ℃, at 30,000xg.
9. Transfer the supernatant to (a) clean 50-mL tube(s). This is the 'cell-free extract'.
Note: A sample of the lysed cell pellet may be stored at 4 ℃ and analyzed later by SDS-PAGE

analysis for the presence of FnCas12a to assess its solubility and to determine if cell lysis was successful.

10. Pass the cell free extract through a 0.22 µm membrane filter and save the filtrate into a sterile tube.
11. Using a peristaltic pump, wash a HisTrap HP column with 3-5 column volumes of distilled water to remove the solution in which the resin is stored.
12. Using a peristaltic pump, equilibrate the HisTrap HP with at least 5 column volumes of Lysis Buffer with a flow rate of 2 mL/min.
13. Using the peristaltic pump[d], pass the filtered cell free extract with a flow rate of 1 mL/min through the HisTrap HP column and collect the flow-through in a 50-mL tube(s) labelled 'flow-through'[e].

The next chromatography steps are performed using an ÄKTA FPLC system.

14. Equilibrate the ÄKTA FPLC system with Wash Buffer until the absorbance at 280 nm reaches a steady baseline. Transfer the HisTrap HP column to the ÄKTA FPLC and wash the column using Wash Buffer with a flow rate of 2 mL/min for 10-15 times the column volume or until the absorbance at 280 nm becomes near[f] stable with the Wash Buffer. Collect the first 3-4 10-mL wash fractions[g] in separate tubes labelled 'wash-through #'.
15. Elute the protein using the Elution Buffer with a flow rate of 2 mL/min while fractionating to 1 mL samples, collect the eluate and save the fractions in separate tubes labelled 'Elution fraction #'. An example of a typical elution chromatogram of FnCas12a purified by Histrap HP (5 mL) affinity purification is shown in **Figure 3.4**.
16. Dilute 10 µl of the collected fractions with 4x Laemmli protein sample buffer, heat for 5 min at 95 ℃ and resolve the samples on SDS-PAGE gel to assess the purity of the sample.
17. Combine the elution fractions in which the protein is present and check the absorbance at 280 nm to estimate the protein concentration. The extinction coefficient of FnCas12a is 145,820 M^{-1} cm^{-1}. Typical yield at this step is ~25 mg per litre of expression culture[h].
18. Add 2 mL (final 1 mM) of 1 M DTT stock and 4 mL of 0.5 M EDTA stock to 2 L of Dialysis Buffer before use. Increase the volume of the combined protein fractions to 25 mL using the Dialysis Buffer. Add 1 mg TEV protease per 100 mg of protein.
19. Take a dialysis membrane with an MWCO of 12,400 and soak it in the Dialysis Buffer for 1 min. Use a clamp to close the dialysis membrane on one end to make a bag.
20. Pipet the combined protein sample into the dialysis membrane bag and close the other end with another clamp. Dialyze the sample overnight at 4 ℃ with slow stirring against 2 L of the Dialysis Buffer.

[d] *For sample loading, a superloop can be used instead of the peristaltic pump.*
[e] *A sample of the flow-through may be stored at 4 ℃ and analysed later by SDS-PAGE analysis for the presence of the FnCas12a to determine if the protein bound to the column (if a large fraction of the protein remains in the flow-through, regenerate or replace your column). In some cases, the high amount of proteins can saturate the column. To recover the protein in the flow-through, the flow-through can be (re)loaded onto another or regenerated HisTrap HP column.*
[f] *Even at its low concentration, imidazole in the washing buffer can remove small amounts of the protein of interest. Therefore, start eluting the protein as the absorbance at 280 nm is near stable to avoid unnecessary loss of the protein of interest.*
[g] *The wash-through may be saved and checked later using SDS-PAGE analysis for the presence of FnCas12a to determine if it was eluted off the column during washing.*
[h] *The yield is most likely overestimated at this point due to protein and nucleic acid contaminations.*

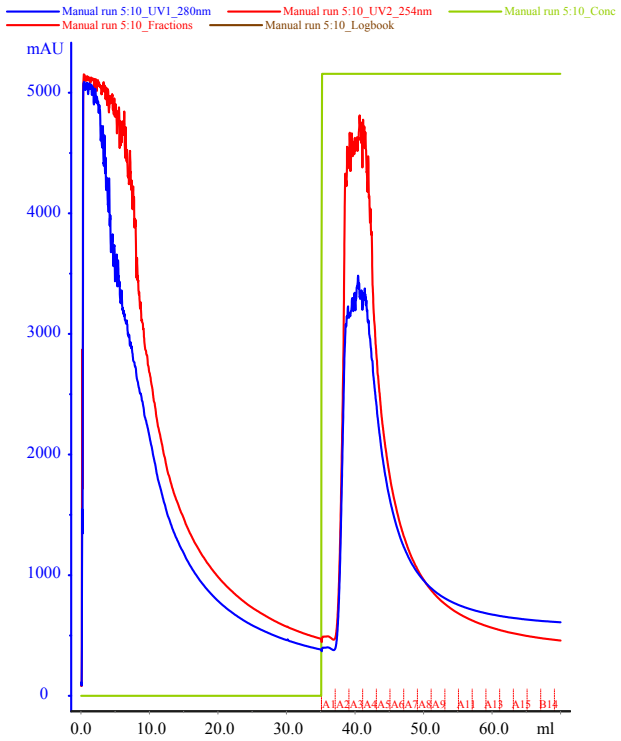


Figure 3.4 | Representative elution chromatogram of FnCas12a purified by Histrap HP (5 mL) affinity purification. 75 mL of cell-free extract was loaded on the column. Elution fractions were 1 mL and the flow rate was set at 2 mL/min. Absorbance at 280 nm is expressed in milli-absorbance units for the A_{280} (blue) and A_{254} (red). Please note that the A_{254} is not very informative after niNTA purification, as at this stage, the sample is contaminated with various nucleic acids. The green line indicates the concentration of Elution Buffer (0% at the start of the chromatogram and 100% at the final stage of the chromatogram).

Day 5. Purification of FnCas12a–Part II

1. Transfer the protein sample from the dialysis membrane into a 50 mL tube and centrifuge it for 10 min at 4,500 *x g* at 4 ℃ to remove potential precipitated proteins.
2. Dilute the sample 1:1 using Sample Dilution Buffer[i].
3. Using a peristaltic pump, wash the column with 3-5 column volumes of distilled water to remove the solution in which the column resin is stored.
4. Using a peristaltic pump, equilibrate the Heparin FF column with at least 5 column volumes of IEX-A Buffer with a flow rate of 2 mL min^{-1}.
5. Using a peristaltic pump, load the protein sample onto the Heparin FF column.
6. Equilibrate the ÄKTA FPLC with IEX-A Buffer until the absorbance at 280 nm reaches a steady baseline.
7. Transfer the Heparin FF column to the ÄKTA FPLC and wash it with 10 mL IEX-A Buffer at 2 mL min^{-1}. Collect the flow-through in appropriately labelled clean tubes.
8. Elute the protein using a linear gradient from 0 to 50% IEX-B Buffer over 60 mL at a flow rate of 2 mL min^{-1}, collect the eluate as 1 mL fractions in appropriately labelled clean tubes. An example of a typical elution chromatogram of FnCas12a purified by Heparin FF (5 mL) affinity purification is shown in **Figure 3.5**.
9. Dilute 10 µl of the collected fractions with 4x Laemmli protein sample buffer, heat for 5 min at 95 ℃, and analyse on SDS-PAGE gel to assess the purity of the sample. An example of a Coomassie Brilliant Blue stained 10% SDS-PAGE gel on which FnCas12a

[i] *For LbCas12a, dilute in a 2:1 ratio (protein sample: Dilution Buffer) due to the instability of LbCas12a at lower salt concentrations.*

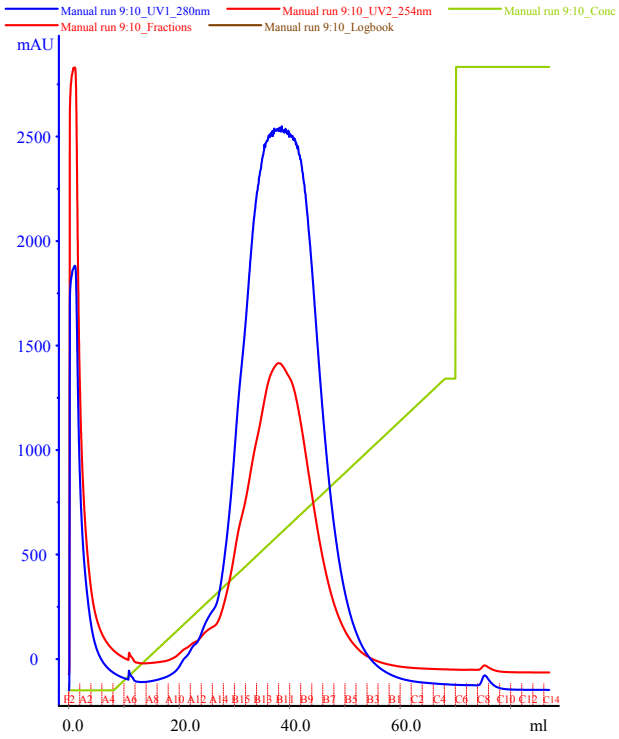


Figure 3.5 | Representative elution chromatogram of FnCas12a purified by Heparin FF (5 mL) affinity purification. 50 mL of sample was loaded. The flow rate was set at 2 mL $min^{-1}$and elution fractions were 1 mL. Absorbance is expressed in milli-absorbance units for the A_{280} (blue) and A_{254} (red). The green line indicates the concentration of IEX-B (0% at the start of the chromatogram with the concentration raising to 50% over 60 mL, and at 100% at the final stage of the chromatogram to wash the column).

Heparin FF elution fractions were resolved is shown in **Figure 3.6**.

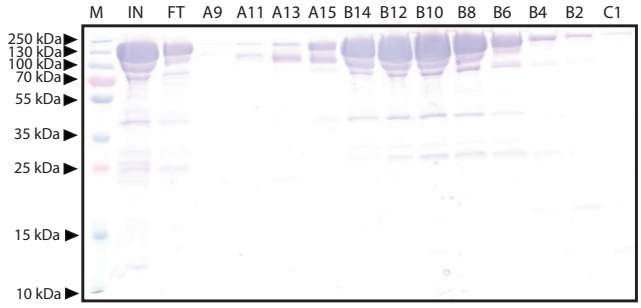


Figure 3.6 | Representative Coomassie Brilliant Blue stained 12% SDS-PAGE gel on which FnCas12a Heparin FF elution fractions were resolved. M: PageRuler™ Plus Prestained Protein Ladder. Marker band sizes are indicated in kDa. IN: concentrated input sample of TEV protease treated Histrap HP elution fractions after overnight dialysis. FT: Flow through from the column. Besides the large band formed by FnCas12a, other (contamination) bands can be observed. A9-C1: Elution fractions near the protein absorbance peak. The FnCas12a protein appears as a band with a size slightly larger than 130 kDa. In this case, fractions B14-C1 were combined.

10. Combine the elution fractions containing pure protein (and as little contaminants as possible) and concentrate the sample by transferring the sample to Amicon Ultra-15 Centrifugal Filter Units with a membrane MWCO of 100 kD and centrifuging the samples at 4,500 *x g* at 4 ℃ until a final volume of < 1 mL is reached.
11. Transfer the sample to an Eppendorf tube and centrifuge the sample for 2 min at maxi-

mum speed in a pre-cooled (4 ℃) microcentrifuge to remove potential precipitation.

12. Equilibrate a 2-mL injection loop and the HiLoad 16/600 Superdex 200 pg column with 10 mL and 240 mL SEC Buffer, respectively, on the ÄKTA FPLC.
13. Load the protein concentrate on the HiLoad 16/600 Superdex 200 pg column using the 2-mL injection loop and resolve the sample on the column using SEC Buffer with a flow rate of 1 mL/min.
14. Collect 1 mL fractions. An example of a typical elution chromatogram of FnCas12a purified by HiLoad 16/600 Superdex 200 pg column size exclusion purification is shown in **Figure 3.7**.

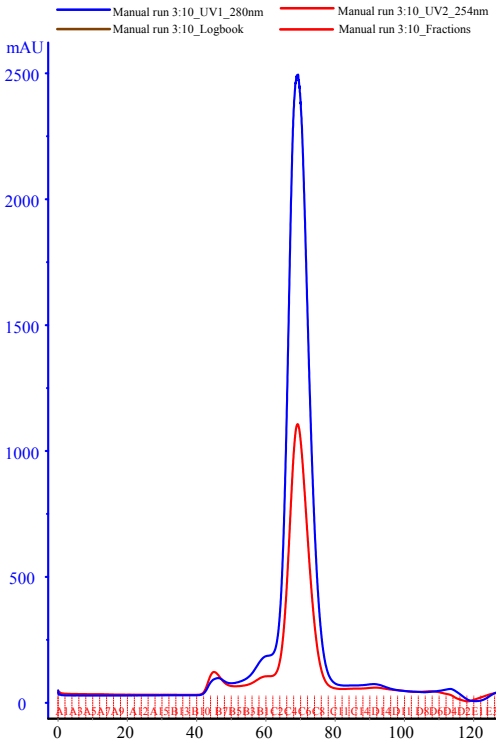


Figure 3.7 | Representative elution chromatogram of FnCas12a resolved on a HiLoad 16/600 Superdex 200 pg column. 1 ml of sample was loaded. Elution fractions were 1 mL and the flow rate was set at 1 mL/min. Absorbance is expressed in milli-absorbance units for the A_{280} (blue) and A_{254} (red).

15. Dilute 10 µl of the collected fractions with 4x Laemmli protein sample buffer, heat for 5 min at 95 ℃, and resolve on a 10% SDS-PAGE gel to assess the purity of the sample. An example of a Coomassie Brilliant Blue stained 10% SDS-PAGE gel on which FnCas12a SEC elution fractions were resolved is shown in **Figure 3.8**.
16. Combine and concentrate the fractions that contain the pure protein and no (or a negligible amount of) contaminants.
17. For analysis of purity and final protein yield, see the Data analysis section. The typical protein yield at this step is ~5-10 mg of FnCas12a per litre of *E. coli* expression culture. Expression and purification of AsCas12a or LbCas12a typically results in slightly higher Cas12a yields.

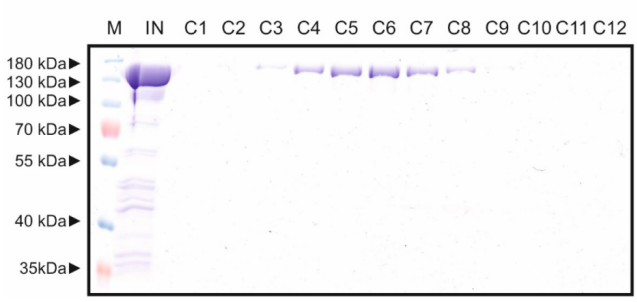


Figure 3.8 | Representative Coomassie Brilliant Blue stained 12% SDS-PAGE gel on which FnCas12a SEC elution fractions were resolved. M: PageRuler Plus Prestained Protein Ladder. Marker band sizes are indicated in kDa. IN: concentrated input sample. Besides the large band formed by FnCas12a, other (contamination) bands can be observed. C1-C12: Elution fractions near the protein absorbance peak. The FnCas12a protein appears as a band with a size slightly larger than 130 kDa. In this case, fractions C3-C9 were combined.

18. Dilute the samples to a concentration suitable for subsequent experiments (*e.g.*, 10 µM)[j] and aliquot the protein at a desired concentration and store at -80 ℃.

Day 6. *In vitro* cleavage assay for confirming the activity of purified FnCas12a

Notes:

a. *We strongly recommend wearing gloves and using nuclease-free tubes and reagents to avoid RNase contamination.*
b. *The reaction volume is typically 20 µL but can be scaled up as needed. Reactions should be assembled in nuclease-free 1.5 mL Eppendorf tubes or in 200 µl PCR (strip) tubes.*
c. *Prepare a 1 µM crRNA solution by diluting the stock with nuclease-free water on ice.*
d. *Prepare a 0.1 µM substrate plasmid or linear DNA solution by diluting the stock with nuclease-free water on ice.*

1. Prepare the following two-step reaction (at a molar ratio Cas12a:crRNA:substrate = 10:20:1) at room temperature:

Component	20 µl reaction
Nuclease-free water	14 µl
10x Nuclease Reaction Buffer	2 µl
1 µM crRNA	2 µl (~100 nM final)
1 µM Cas12a	1 µl (~50 nM final)
Reaction volume	19 µl
Pre-incubate at room temperature for 20 min	
0.1 µM (substrate DNA fragment)	1 µl (~5 nM final)
Total reaction volume	20 µl

[j] *It is recommended to store the protein at ≥ 10 µM and dilute it to the right concentration only just before use. At lower concentrations, a relative high fraction of the protein can be lost due freezing/thawing and non-specific adsorption to the surface of the tube/container used for storage.*

2. Mix thoroughly and pulse-spin in a micro-centrifuge.
3. Incubate at 37 ℃ for 30 min.
4. Add 1 µl of Proteinase K, mix thoroughly and pulse-spin in a micro-centrifuge.
5. Incubate at room temperature for 10 min.
6. Add 4 µl of 6x DNA loading dye.
7. Resolve 20 µl of the sample on an 1% agarose gel pre-stained with SYBR™ Safe DNA Gel Stain.
8. Visualize the gel using an imaging system equipped with an excitation source in the UV range or between 470-530 nm. An example of an *in vitro* cleavage assay using FnCas12a:crRNA and a linear DNA substrate is shown in **Figure 3.9**.

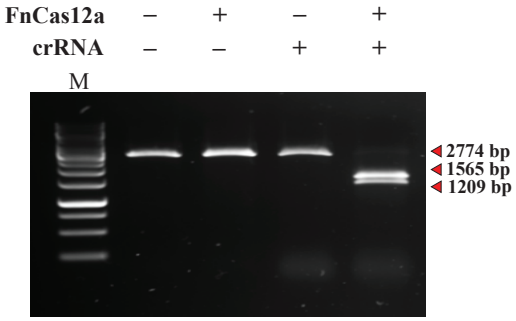


Figure 3.9 | A 2,774 bp linear target DNA substrate is cleaved by the FnCas12:crRNA complex, yielding products of 1,209 bp and 1,565 bp. M: GeneRuler 1 kb DNA Ladder.

Data analysis

Note: This section explains how to determine the yield and purity of your protein after the final step of the purification protocol (i.e., after size exclusion chromatography).

1. Use a NanoDrop Spectrophotometer to measure the protein concentration. When using the NanoDrop software method 'Protein A_{280}', both the absorption at 260 nm and the absorption at 280 nm are measured.
2. The 260/280 ratio of the purified protein sample can be used to determine if the complex is free from nucleic acids. Typically, a pure protein sample has a 260/280 ratio of ~0.57. We typically achieve a 260/280 ratio between 0.54 and 0.61. Nucleic acid contamination rapidly increases the 260/280 ratio to above 1.
3. Use the measured absorbance at 280 nm to determine the final protein concentration.
   (a) When using the 'Other protein (ε + MW)' option of the NanoDrop software, provide the molar extinction coefficient of FnCas12a (144,330 M^{-1} cm^{-1}) and its molecular weight (151 kDa). For AsCas12a, these values are 15,780 M^{-1} cm^{-1} and 156 kDa. For LbCas12a, these values are 181,690 M^{-1} cm^{-1} and 149 kDa.
   (b) When using the '1 Abs = 1 mg mL^{-1} option of the NanoDrop software, use a correction factor to determine the real protein concentration. The correction factors for FnCas12a, AsCas12a, and LbCas12a are 0.951, 1.009 and 1.221, respectively. For example, if the NanoDrop measurement gives a protein concentration of 5.0 mg

mL^{-1} of FnCas12a (assuming 1 Abs = 1 mg mL^{-1}), your real protein concentration is 5/0.951 = 5.26 mg mL^{-1}. To calculate the protein concentration in mM, use the following formula: protein concentration (in mg mL^{-1})/molecular weight (in kDa; given above). For example, if you have a FnCas12a sample with 5.26 mg mL^{-1} protein, the protein concentration is 5.26/151 = 0.035 mM = 35 µM.

Recipes

Media, antibiotics and stock solutions

1. LB medium (1 L)
    a. Weigh out 10 g tryptone, 5 g yeast extract, and 10 g NaCl.
    b. Fill up to 800 mL with demi-water.
    c. Adjust pH to 7.5 with NaOH.
    d. Fill up to 1 L with demi-water.
    e. Autoclave at 121 ℃ for 30 min.
    f. Store at room temperature.
2. 1,000x chloramphenicol stock (34 mg mL^{-1} stock)
    a. Weigh out 0.34 g chloramphenicol and dissolve it in 10 mL of 100% ethanol.
    b. Pass through a 0.22 µm syringe filter.
    c. Store at -20 ℃.
3. 1,000x kanamycin stock (50 mg mL^{-1} stock)
    a. Weigh out 0.5 g kanamycin sulphate and dissolve it in 10 mL of sterile water.
    b. Pass through a 0.22 µm syringe filter.
    c. Store at -20 ℃.
4. Glycerol stock (50% solution)
    a. Add 50 mL of 100% glycerol solution into a 250-mL bottle. When pipetting glycerol, use ethanol sterilized scissors to cut off the end of a pipette to make pipetting easier.
    b. Add 50 mL demi-water.
    c. Autoclave at 121 ℃ for 30 min.
    d. Store at room temperature.
5. 1 M IPTG (IsoPropyl-1-Thio-β-D-Galactopyranoside)
    a. Weigh out 2.38 g of IPTG and dissolve in 10 mL of sterile water.
    b. Pass through a 0.22 µm syringe filter.
    c. Store at -20 ℃.
6. 1 M DTT (Dithiothreitol) stock
    a. Weigh out 1.5 g DTT and dissolve it in 10 mL of sterile water.
    b. Pass through a 0.22 µm syringe filter.
    c. Store in the dark at -20 ℃.
7. 0.5 M EDTA (Disodium Ethylene Diamine Tetra-Acetate) stock (pH 8.0)

a. Weigh out 18.16 g of $Na_2EDTA{\cdot}2H_2O$ and dissolve in 80 mL of demi-water[k].
b. Adjust to pH 8.0 with pellets of NaOH (~2 g of NaOH is required).
c. Fill up to 100 mL with demi-water.
d. Sterilize by autoclaving Autoclave at 121 ℃ for 30 min.
e. Store at room temperature.

Buffers

1. Lysis Buffer (1 L)
    a. Weigh out 29.22 g NaCl (final 500 mM), 2.42 g Tris (final 20 mM) and 0.68 g imidazole (final 10 mM) and dissolve in 900 mL of demi-water.
    b. Adjust pH to 8.0 using HCl.
    c. Fill up to 1 L with demi-water.
    d. Filter using 0.22 µm membrane filter.
    e. Store at 4 ℃.
2. Wash Buffer (1 L)
    a. Weigh out 29.22 g NaCl (final 500 mM), 2.42 g Tris (final 20 mM) and 1.36 g imidazole (final 20 mM) and dissolve in 900 mL of demi-water.
    b. Adjust pH to 8.0 using HCl.
    c. Fill up to 1 L with demi-water.
    d. Filter using 0.22 µm membrane filter.
    e. Store at 4 ℃.
3. Elution Buffer (1 L)
    a. Weigh out 29.22 g NaCl (final 500 mM), 2.42 g Tris (final 20 mM) and 17 g imidazole (final 250 mM) and dissolve in 900 mL of demi-water.
    b. Adjust pH to 8.0 using HCl.
    c. Fill up to 1 L with demi-water.
    d. Filter using 0.22 µm membrane filter.
    e. Store at 4 ℃.
4. Dialysis Buffer (2 L)
    a. Weigh out 37.27 g KCl (final 250 mM) and 9.53 g HEPES (final 20 mM) and dissolve in 1,900 mL of demi-water.
    b. Adjust pH to 8.0 using KOH.
    c. Fill up to 2 L with demi-water.
    d. Filter using 0.22 µm membrane filter.
    e. Store at 4 ℃.
5. Dilution Buffer (200 mL)
    a. Weigh out 0.96 g HEPES (final 20 mM) and dissolve in 150 mL of demi-water.
    b. Adjust pH to 8.0 using KOH.

[k] *The disodium salt of EDTA will not go into solution until the pH of the solution is adjusted to approximately 8.0 by the addition of NaOH.*

c. Fill up to 200 mL with demi-water.
d. Filter using 0.22 µm membrane filter.
e. Store at 4 ℃.

6. IEX-A Buffer (1 L)

a. Weigh out 11.18 g KCl (final 150 mM) and 4.77 g HEPES (final 20 mM) and dissolve in 900 mL of demi-water.
b. Adjust pH to 8.0 using KOH.
c. Fill up to 1 L with demi-water.
d. Filter using 0.22 µm membrane filter.
e. Store at 4 ℃.

7. IEX-B Buffer (1 L)

a. Weigh out 149.10 g KCl (final 2 M) and 4.77 g HEPES (final 20 mM) and dissolve in 900 mL of demi-water.
b. Adjust pH to 8.0 using KOH.
c. Fill up to 1 L with demi-water.
d. Filter using 0.22 µm membrane filter.
e. Store at 4 ℃.

8. SEC Buffer (1 L)

a. Weigh out 37.27 g KCl (final 500 mM), 4.77 g HEPES (final 20 mM) and dissolve in 900 mL of demi-water. Add 1 mL (final 1 mM) of 1 M DTT stock just before use.
b. Adjust pH to 8.0 using KOH.
c. Fill up to 1 L with demi-water.
d. Filter using 0.22 µm membrane filter.
e. Store at 4 ℃.

Note: Setting the pH of the buffers at different temperatures will influence the final pH as the pH is temperature dependent. However, in our experience, the pH of the buffers can be set at 4 ℃ or at RT without having a significant impact on the purification procedure.

9. 10x SDS-PAGE Electrophoresis Running Buffer (1 L)

a. Weigh out 30.0 g Tris (final 250 mM), 144 g glycine (final 1920 mM), 10 g SDS [final 1% (w/v)] and dissolve in 900 mL of demi-water.
b. Fill up to 1 L with demi-water.

10. 10x Cas9 Nuclease Reaction Buffer (10 mL)

a. Weight out 0.58 g NaCl (final 1 M), 0.1 g $MgCl_2 \cdot 6H_2O$ (final 50 mM), 0.476 HEPES (final 200 mM) and 3.72 mg EDTA (final 1 mM) and dissolve in 8 mL of nuclease-free water.
b. Adjust pH to 6.5 and fill up to 10 mL with nuclease-free water.

Acknowledgments

This protocol is adapted from Swarts *et al.* (2017; Reference (*270*)). This work was financially supported by the grant from the Netherlands Organization of Scientific Research (NWO) to J.v.d.O (NWO-TOP, 714.015.001.), by a Swiss National Science Foundation (SNSF) Project Grant to M.J. (SNSF 31003A_149393), and by fellowships of the European Molecular Biology Organization (EMBO) to D.C.S. (EMBO ALTF 179-2015 and EMBO aALTF 509-2017). The authors declare no conflicts of interest or competing interests.

Author information

Correspondence should be addressed to daan.swarts@wur.nl.

Chapter 4

Multiplex gene editing by CRISPR-Cas12a (Cpf1) using a single crRNA array

Bernd Zetsche*, Matthias Heidenreich*, Prarthana Mohanraju*, Iana Fedorova, Jeroen Kneppers, Ellen M. DeGennaro, Nerges Winblad, Sourav R. Choudhury, Omar O. Abudayyeh, Jonathan S. Gootenberg, Wen Y. Wu, David A. Scott, Konstantin Severinov, John van der Oost, Feng Zhang

* contributed equally
Adapted from:
Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array.
2017, Nature Biotechnology, 35(1), 31-34.

Abstract

Targeting of multiple genomic loci with Cas9 is limited by the need for multiple or large expression constructs. Here we show that the ability of Cas12a to process its own CRISPR RNA (crRNA) can be used to simplify multiplexed genome editing. Using a single customized CRISPR array, we edit up to four genes in mammalian cells and three in the mouse brain, simultaneously.

Results

Although multiplex gene editing is possible with Cas9 nuclease, it requires relatively large constructs or simultaneous delivery of multiple plasmids (*233*, *274–277*), both of which are problematic for multiplex screens or *in vivo* applications. By contrast, the Cas12a nuclease requires only one Pol III promoter to drive several small crRNAs (39 nucleotide (nt) per crRNA).

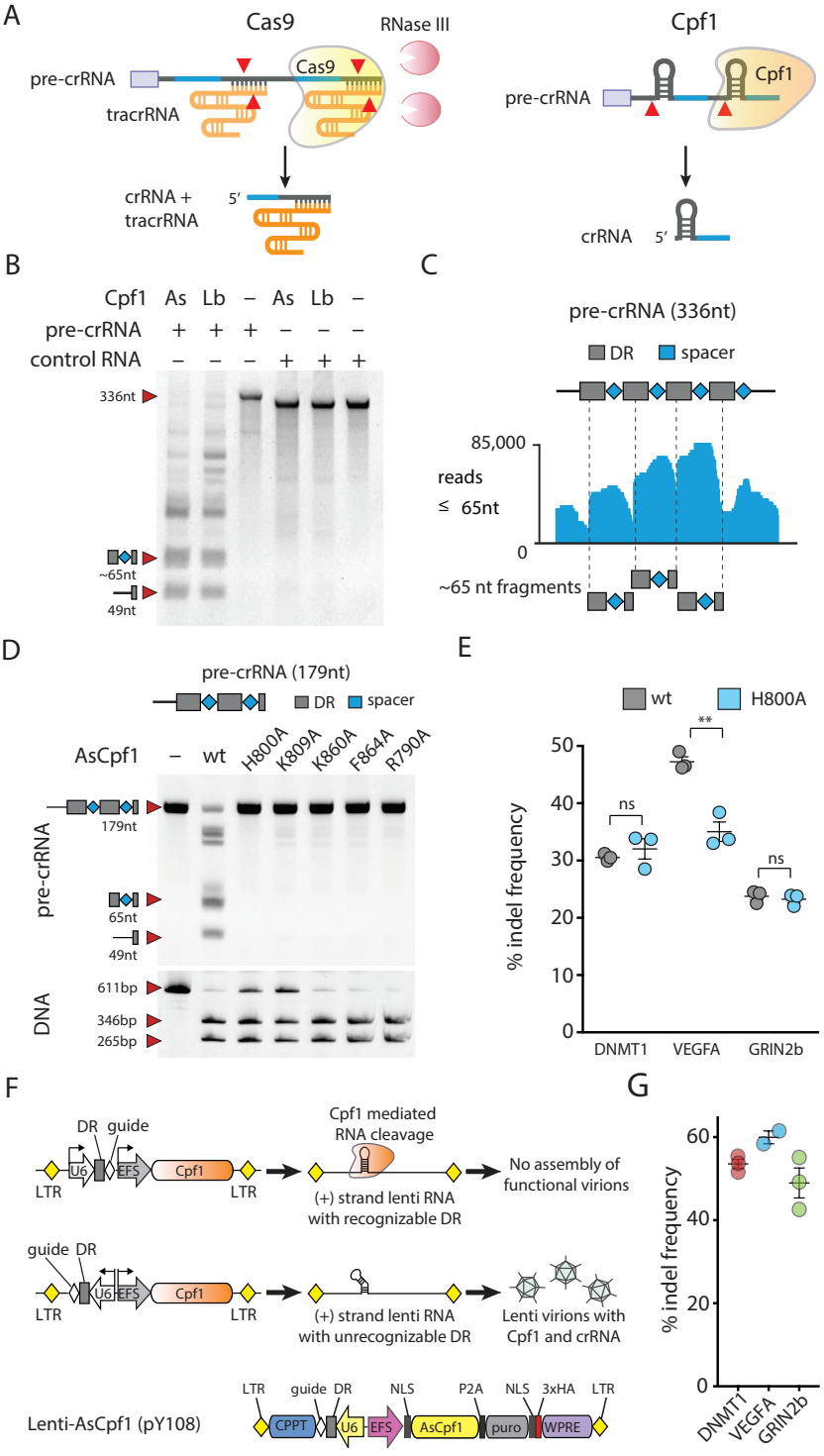
We confirmed *in vitro* that Cas12a (hereafter referred to by its former name, Cpf1) alone is sufficient for maturation of crRNAs (*83*, *187*) (**Figure 4.1A**) using an artificial CRISPR pre-crRNA array consisting of four spacers separated by direct repeats from the CRISPR locus of *Francisella novicida* (FnCpf1) and two Cpf1 orthologs with activity in mammalian cells, *Acidaminococcus* Cpf1 (AsCpf1) and *Lachnospiraceae* Cpf1 (LbCpf1) (**Figure 4.1B**). Small RNA-seq showed that AsCpf1 cleavage products correlate to fragments resulting from cuts at the 5′ end of direct repeat hairpins, identical to the cleavage pattern we observed in *Escherichia coli* heterologously expressing FnCpf1 CRISPR systems (*83*) (**Figure 4.1C**).

We further validated these results by generating AsCpf1 mutants that are unable to process arrays. Guided by the crystal structure of AsCpf1 (*140*), we mutated five conserved amino acid residues likely to disrupt array processing (H800A, K809A, K860A, F864A, and R790A) (*140*). All mutations interfered with pre-crRNA processing but not DNA cleavage activity *in vitro* (**Figure 4.1D** and **Figure S4.1A,B**), an effect that was also observed for FnCpf1 (*187*). AsCpf1 recognizes specific nucleotides at the 5′ flank of the direct repeat stem loop. Substitution of these nucleotides weakened or abolished RNA cleavage (**Figure S4.1A**). Dosage tests with the five AsCpf1 mutants revealed that mutants K809A, K860A, F864A, and R790A show pre-crRNA processing when used at high concentration (**Figure S4.1B**) or for extended incubation times (**Figure S4.1C**), but H800A was inactive regardless of dose and incubation time.

We next tested whether this mutant retains DNase activity in human embryonic kidney (HEK) 293T cells using three guides. Insertion/deletion (indel) frequency at the *DNMT1* and *GRIN2b* loci were identical for wild-type and H800A AsCpf1, and only slightly higher at the *VEGFA* locus in cells transfected with wild-type AsCpf1, demonstrating that the RNA and DNA cleavage activity can be separated in mammalian cells (**Figure 4.1E**).

Cpf1-mediated RNA cleavage needs to be considered when designing lentivirus vectors for simultaneous expression of nuclease and guide (**Figure 4.1F**). Lentiviruses carry a (+) strand RNA copy of the DNA sequence flanked by long terminal repeats, including the pre-crRNA, allowing Cpf1 to bind and cleave at the direct repeat sequence. Hence, reversing the orientation of the direct repeat is expected to result in (+) strand lentivirus RNAs not susceptible to Cpf1-mediated cleavage. We designed a lentivirus encoding AsCpf1 and a crRNA expression cassette. We transduced HEK293T cells with a MOI (multiplicity of infection) of <0.3 and analysed indel frequencies in puromycin-selected cells 10 days after infection. Using guides encoded on a reversed expression cassette targeting *DNMT1*, *VEGFA*, or *GRIN2b* resulted in robust indel formation for each targeted gene (**Figure 4.1G**).

We leveraged the simplicity of Cpf1 crRNA maturation to achieve multiplex genome editing in HEK293T cells using customized CRISPR arrays. We chose four guides targeting different genes (*DNMT1*, *EMX1*, *VEGFA*, and *GRIN2b*) and constructed three arrays with variant direct repeat and guide lengths for expression of pre-crRNAs (array 1, 19 DR with 23 nt guide; array 2, 19 nt DR with 30 nt guide; array 3, 35 nt DR with 30 nt guide; **Figure 4.2A**). Indel events


A
Cas9
RNase III
pre-crRNA
tracrRNA
Cas9
crRNA +
tracrRNA
5′
Cpf1
pre-crRNA
Cpf1
crRNA
5′
B
Cpf1 As Lb – As Lb –
pre-crRNA + + + – – –
control RNA – – – + + +
336nt
~65nt
49nt
C
pre-crRNA (336nt)
DR
spacer
85,000
reads
≤ 65nt
0
~65 nt fragments
D
pre-crRNA (179nt)
DR
spacer
AsCpf1 – wt H800A K809A K860A F864A R790A
179nt
65nt
49nt
pre-crRNA
611bp
346bp
265bp
DNA
E
wt
H800A
**
ns
ns
% indel frequency
50
40
30
20
10
0
DNMT1
VEGFA
GRIN2b
F
DR guide
U6
EFS
Cpf1
LTR
LTR
Cpf1 mediated
RNA cleavage
(+) strand lenti RNA
with recognizable DR
No assembly of
functional virions
guide DR
U6
EFS
Cpf1
LTR
LTR
(+) strand lenti RNA
with unrecognizable DR
Lenti virions with
Cpf1 and crRNA
Lenti-AsCpf1 (pY108)
LTR
guide
DR
NLS
P2A
NLS
3xHA
LTR
CPPT
U6
EFS
AsCpf1
puro
WPRE
G
% indel frequency
60
40
20
0
DNMT1
VEGFA
GRIN2b

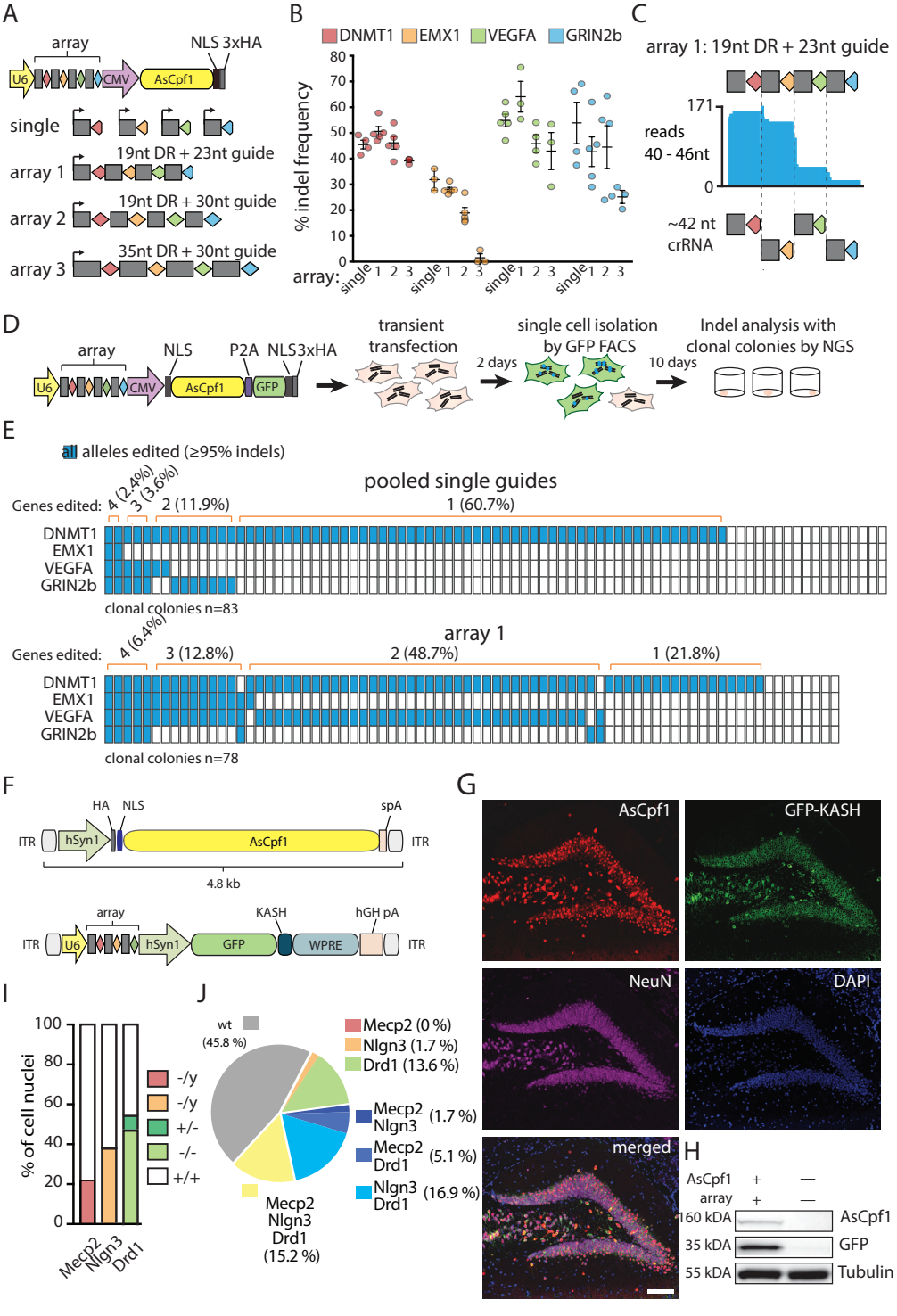
were detected at each targeted locus in cells transfected with array 1 or array 2. However, the crRNA targeting *EMX1* resulted in indel frequencies of <2% when expressed from array 3. Overall, array 1 performed best, with all guides showing indel levels comparable to those mediated by single crRNAs (**Figure 4.2B**). Furthermore, small RNA-seq confirmed that autonomous, Cpf1-mediated pre-crRNA processing occurs in mammalian cells (**Figure 4.2C**). Using arrays with guides in different orders resulted in similar indel frequencies, suggesting that positioning within an array is not crucial for activity (**Figure S4.2A,B**).

To confirm that multiplex editing occurs within single cells, we generated AsCpf1-P2A-GFP constructs to enable fluorescence-activated cell sorting (FACS) of transduced single cells (**Figure 4.2D**) and clonal expansion. We used next-generation deep sequencing (NGS) to compare edited loci within clonal colonies derived from cells transfected with either pooled single guides or array 1. Focusing on targeted genes edited at every locus (indels ≥95%) shows that multiplex editing occurs more frequently in colonies transfected with array 1 (6.4% all targets, 12.8% three targets, 48.7% two targets) than in pooled transfection (2.4% all targets, 3.6% three targets, 11.9% two targets) (**Figure 4.2E**).

We next tested multiplex genome editing in neurons using AsCpf1. We designed a gene-delivery system based on adeno-associated viral vectors (AAVs) for expression of AsCpf1. We generated a dual vector system in which AsCpf1 and the CRISPR-Cpf1 array were cloned separately (**Figure 4.2F**). We constructed a U6-promoter-driven Cpf1 array targeting the neuronal genes *Mecp2*, *Nlgn3*, and *Drd1*. This plasmid also included a green fluorescent protein (GFP), fused to KASH nuclear transmembrane domain (*278*), in order to enable FACS of targeted cell nuclei (*279*).

We first transduced mouse primary cortical neurons *in vitro* and observed robust expression of AsCpf1 and GFP-KASH 1 week after viral delivery. A SURVEYOR nuclease assay run on purified neuronal DNA confirmed indel formations in all three targeted genes (**Figure S4.3**). Next, we tested whether AsCpf1 could be expressed in the brains of living mice for multiplex genome editing *in vivo*. We stereotactically injected our dual vector system in a 1:1 ratio into the hippocampal dentate gyrus (DG) of adult male mice. Four weeks after viral delivery we observed robust expression of AsCpf1 and GFP-KASH in the DG (**Figure 4.2G,H**). Consistent with previous studies (*228*, *279*), we observed ~75% co-transduction efficiency of the dual viral vectors (**Figure S4.3C**). We isolated targeted DG cell nuclei by FACS (**Figure S4.4**) and

Figure 4.1 ***(preceding page)*****|Cpf1-mediated processing of pre-crRNA is independent of DNA cleavage. (A)** Schematic of pre-crRNA processing for Cas9 and Cpf1. Cleavage sites indicated with red triangles. Trans-activating crRNA (tracrRNA). **(B)** *In vitro* processing of FnCpf1 pre-crRNA transcript (80 nM) with purified AsCpf1 or LbCpf1 protein (~320 nM). **(C)** RNA-seq analysis of FnCpf1 pre-crRNA cleavage products, as shown in b. A high fraction of sequence reads smaller than 65 nt are cleavage products of spacers flanked by direct repeat sequences, cropped gel images. **(D)** Pre-crRNA (top) and DNA cleavage (bottom) mediated by AsCpf1 point mutants. H800A, K809A, K860A, F864A, and R790A fail to process pre-crRNA but retain DNA cleavage activity *in vitro*. 330 nM pre-crRNA was cleaved with 500 nM Cpf1 in 15 min and 25 nM DNA was cleaved with 165 nM Cpf1 in 30 min. **(E)** Indel frequencies mediated by AsCpf1H800A are comparable to wt AsCpf1, bars are mean of 3 technical replicates from one experiment, error bars are s.e.m. (Student t-test; n.s., not significant; **P = 0.003). **(F)** Schematic of lentivirus Cpf1 construct with the U6::direct repeat cassette in different orientations (top and middle), (+)-strand RNA copy with recognizable direct repeats are susceptible to Cpf1-mediated degradation, preventing functional virion formation. Schematic of AsCpf1 (pY108) construct (bottom). **(G)** Indel frequencies analysed by SURVEYOR nuclease assay after puromycin selection 10 d after transduction with lentivirus AsCpf1 in HEK cells. Horizontal bars are mean of 2 or 3 individual infections; error bars are mean ± s.e.m. U6, Pol III promoter; CMV, cytomegalovirus promoter; NLS, nuclear localization signal; HA, hemagglutinin tag; DR, direct repeat sequence; P2A, porcine teschovirus-1 2A self-cleaving peptide; LTR, long terminal repeat; WPRE, woodchuck hepatitis virus post-transcriptional regulatory element.


A
array
NLS 3xHA
U6
CMV
AsCpf1
single
array 1
19nt DR + 23nt guide
array 2
19nt DR + 30nt guide
array 3
35nt DR + 30nt guide
B
DNMT1 EMX1 VEGFA GRIN2b
% indel frequency
array: single 1 2 3 single 1 2 3 single 1 2 3 single 1 2 3
C
array 1: 19nt DR + 23nt guide
171
reads
40 - 46nt
0
~42 nt
crRNA
D
array
NLS
P2A
NLS3xHA
GFP
transient
transfection
2 days
single cell isolation
by GFP FACS
10 days
Indel analysis with
clonal colonies by NGS
E
all alleles edited (≥95% indels)
pooled single guides
Genes edited:
4 (2.4%)
3 (3.6%)
2 (11.9%)
1 (60.7%)
DNMT1
EMX1
VEGFA
GRIN2b
clonal colonies n=83
array 1
Genes edited:
4 (6.4%)
3 (12.8%)
2 (48.7%)
1 (21.8%)
DNMT1
EMX1
VEGFA
GRIN2b
clonal colonies n=78
F
HA
NLS
spA
ITR
hSyn1
AsCpf1
ITR
4.8 kb
array
KASH
hGH pA
ITR
U6
hSyn1
GFP
WPRE
ITR
G
AsCpf1
GFP-KASH
NeuN
DAPI
merged
I
% of cell nuclei
-/y
-/y
+/-
-/-
+/+
Mecp2
Nlgn3
Drd1
J
wt
(45.8 %)
Mecp2 (0 %)
Nlgn3 (1.7 %)
Drd1 (13.6 %)
Mecp2
Nlgn3 (1.7 %)
Mecp2
Drd1 (5.1 %)
Nlgn3
Drd1 (16.9 %)
Mecp2
Nlgn3
Drd1
(15.2 %)
H
AsCpf1 + —
array + —
160 kDA
AsCpf1
35 kDA
GFP
55 kDA
Tubulin

quantified indel formation using NGS. We found indels in all three targeted loci with ~23%, ~38%, and ~51% indel formation in *Mecp2*, *Nlgn3*, and *Drd1*, respectively (**Figure S4.3D,E**). We quantified the effectiveness of bi-allelic disruption of the autosomal gene *Drd1* and found that ~47% of all sorted nuclei (*i.e.*, ~87% of all *Drd1*-edited cells) harboured bi-allelic modifications (**Figure 4.2I**). Next, we quantified the multiplex targeting efficiency in single neuronal nuclei. Our results show that ~15% of all transduced neurons were modified in all three targeted loci (**Figure 4.2J**). Taken together, our results demonstrate the effectiveness of AAV-mediated delivery of AsCpf1 into the mammalian brain and simultaneous multi-gene targeting *in vivo* using a single array transcript.

Taken together, these data highlight the utility of Cpf1 array processing in designing simplified systems for *in vivo* multiplex gene editing. This system should simplify guide RNA delivery for many genome editing applications in which targeting of multiple genes is desirable.

Experimental procedures

Cpf1 protein purification

Humanized Cpf1 were cloned into a bacterial expression vector (6-His-MBP-TEV-Cpf1, a pET-based vector kindly given to us by Doug Daniels). Two litres of Terrific Broth growth media with 100 µg mL^{-1} ampicillin was inoculated with 10 mL of an overnight culture of Rosetta (DE3) pLyseS (EMD Millipore) cells containing the Cpf1 expression construct. Growth media plus inoculant was grown at 37 ℃ until the cell density reached 0.2 OD_{600}, then the temperature was decreased to 21 ℃. Growth was continued until OD_{600} reached 0.6 when a final concentration of 500 µM IPTG was added to induce MBP-Cpf1 expression. The culture was induced for 14–18 h before harvesting cells and freezing at –80 ℃ until purification. Cell paste was resuspended in 200 mL of lysis buffer (50 mM HEPES pH 7, 2 M NaCl, 5 mM $MgCl_2$, 20 mM imidazole) supplemented with protease inhibitors (Roche cOmplete, EDTA-free) and lysozyme. Once homogenized, cells were lysed by sonication (Branson Sonifier 450), then centrifuged at 10,000xg for 1 h to clear the lysate. The lysate was filtered through 0.22-µm filters (Millipore, Stericup) and applied to a nickel column (HisTrap FF, 5 mL), washed, and then eluted with a gradient of imidazole. Fractions containing protein of the expected size were pooled,

Figure 4.2 ***(preceding page)*****|Cpf1-mediated multiplex gene editing in mammalian cells and mouse brain.** **(A)** Schematic of multiplex gene editing with AsCpf1, using a single plasmid approach. **(B)** Genome editing at four different genomic loci mediated by AsCpf1 with different versions of artificial CRISPR arrays (array 1, crRNAs in their mature form (19-nt DR with 23-nt guide); array 2, crRNAs are in an intermediate form (19-nt DR with 30-nt guide); array 3 crRNAs are in their unprocessed form (35-nt DR with 30-nt guides)). Indels were analysed by SURVEYOR nuclease assay 3 d after transfection. Horizontal bars are the means of two individual experiments with three to five technical replicates; error bars are mean ± s.e.m. **(C)** Small RNA-seq reads from HEK cells transfected with AsCpf1 and array 1 show fragments corresponding to mature crRNA for each of the four guides. **(D)** Schematic for analysis of indel events in clonal colonies 48 h after transient transfection. **(E)** Quantification of indel events measured by NGS in clonal colonies from HEK cells transiently transfected with pooled single-guide-RNA plasmids or plasmid carrying array 1. Colonies were expanded for 10 d after sorting. Each column represents one clonal colony; blue rectangles indicate target genes with all alleles edited. **(F)** Schematic of AAV vector design for multiplex gene editing. Bottom: grey rectangles, direct repeat; diamonds, spacer (red: Mecp2, orange: Nlgn3, green: Drd1). **(G)** Immunostaining of dorsal DG 4 weeks after stereotactic AAV injection (representative image of n = 4 mice). Brain sections were co-stained with anti-HA (red), anti-GFP (green) and anti-NeuN (magenta) antibodies. Nuclei were labelled with DAPI (blue). Scale bar, 100 µm. **(H)** Western blot analysis of DG expressing HA-AsCpf1 and GFP-KASH (representative blot from n = 4 mice). **(I)** Fraction of mono- (–/+), bi- (–/–) or maternal (–/y) allele editing for Drd1 (autosomal), Mecp2 and Nlgn3 (x-chromosomal). **(J)** Analysis of multiplexing efficiency in individual cells. ITR, inverted terminal repeat; spA, synthetic polyadenylation signal; hSyn1, human synapsin 1 promoter; KASH, Klarsicht ANC1 Syne1 homology nuclear transmembrane domain; hGH pA, human growth hormone polyadenylation signal.

TEV protease (Sigma) was added, and the sample was dialyzed overnight into TEV buffer (500 mM NaCl, 50 mM HEPES pH 7, 5 mM $MgCl_2$, 2 mM DTT). After dialysis, TEV cleavage was confirmed by SDS-PAGE, and the sample was concentrated to 500 µL before loading on a gel filtration column (HiLoad 16/600 Superdex 200) via FPLC (fast protein liquid chromatography, AKTA Pure). Fractions from gel filtration were analysed by SDS-PAGE; fractions containing Cpf1 were pooled and concentrated to 200 µL (50 mM Tris-HCl pH 7.5, 2 mM DTT, 5% glycerol, 500 mM NaCl) and either used directly for biochemical assays or frozen at –80 ℃ for storage.

In vitro synthesis of pre-crRNA arrays

Pre-crRNA arrays were synthesized using the HiScribe T7 High Yield RNA Synthesis Kit (NEB). PCR fragments coding for arrays, with a short T7-priming sequence on the 5′ end, were used as templates for *in vitro* transcription reaction (**Table S4.1**). T7 transcription was performed for 4 h and then RNA was purified using the MEGAclear Transcription Clean-Up Kit (Ambion).

In vitro cleavage assay

In vitro cleavage was performed with purified recombinant proteins for AsCpf1 and LbCpf1. Cpf1 protein and *in vitro*-transcribed pre-crRNA arrays were incubated at 37 ℃ in cleavage buffer (20 mM Tris HCl, 50 mM KCl supplemented with RNase Inhibitor Murine (NEB)) for 5 min to 1 h, as indicated in figure legends. Each cleavage reaction contained 20–630 nM of Cpf1 protein and 165 or 330 nM of synthetized pre-crRNA array, as indicated in figure legends. For DNA cleavage, 25 nM of target was cleaved with 165 nM Cpf1 and 340 nM crRNA for 30 min at 37 ℃. Reactions were stopped with proteinase K (Qiagen), heat denaturation and run on 10% TBE-Urea polyacrylamide gels. Gels were stained with SYBR Gold DNA stain (Life Technologies) for 10 min and imaged with a Gel Doc EZ gel imaging system (Bio-Rad).

Pre-crRNA array design and cloning

Guide sequences targeting human genes are listed in **Table S4.2**. crRNAs were designed as four oligos (IDT) consisting of direct repeats, each one followed by a crRNA (**Table S4.3**). The oligos favoured a one-directional annealing through their sticky-end design. The oligonucleotides (final concentration 10 µM) were annealed in 10×T4 ligase buffer (final concentration 1×; NEB) and T4 PNK (5 units; NEB). Thermocycler conditions were adjusted to 37 ℃ for 30 min, 95 ℃ for 5 min followed by a –5 ℃/min ramp down to 25 ℃. The annealed oligonucleotides were diluted 1:10 (final concentration 1 µM) and ligated into BsmBI-cut pcDNA-huAsCpf1-U6 (pY26), using T7 DNA ligase (Enzymatics), in room temperature for 30 min. The constructs were transformed into STBL3 bacteria and plated on ampicillin-containing (100 g mL^{-1}) agar plates. Single colonies were grown in standard LB media (Broad Facilities) for 16 h. Plasmid DNA was harvested from bacteria according to QIAquick Spin Miniprep protocol (QIAGEN).

Cell culture and transfection

Human embryonic kidney 293T (HEK293T) cell line (Life Technologies) were maintained in Dulbecco's modified Eagle's Medium (DMEM) + GLUTAMAX (Gibco) supplemented with 10% FBS (HyClone) at 37 ℃ with 5% CO_2 incubation. HEK293FT cells were seeded onto 24-well plates (Corning) 24 h before transfection. Cells were transfected using Lipofectamine 2000 (Life Technologies) at 70–80% confluency following the manufacturer's recommended protocol. For each well of a 24-well plate, a total of 500 ng plasmid DNA was used; each well represents one technical replicate.

Surveyor nuclease assay for genome modification

HEK293T cells were transfected with DNA, as described above. Cells were incubated at 37 ℃ for 72 h after transfection before genomic DNA extraction. Genomic DNA was extracted using the QuickExtract DNA Extraction Solution (Epicentre) following the manufacturer's protocol. Briefly, pelleted cells were suspended in QuickExtract solution and incubated at 65 ℃ for 15 min, 68 ℃ for 15 min, and 98 ℃ for 10 min. The genomic region flanking the CRISPR target site for each gene was PCR amplified (primers listed in **Table S4.4**), and products were purified using QIAQuick PCR purification Kit (Qiagen) following the manufacturer's protocol. 200 ng total of the purified PCR products were mixed with 1 µL 10× Taq DNA Polymerase PCR buffer (Enzymatics) and ultrapure water to a final volume of 10 µl, and subjected to a re-annealing process to enable heteroduplex formation: 95 ℃ for 10 min, 95 ℃ to 85 ℃ ramping at 2 ℃/s, 85 ℃ to 25 ℃ at 0.25 ℃/s, and 25 ℃ hold for 1 min. After re-annealing, products were treated with Surveyor nuclease and Surveyor enhancer S (IDT) following the manufacturer's recommended protocol and analysed on 10% Novex TBE polyacrylamide gels (Life Technologies). Gels were stained with SYBR Gold DNA stain (Life Technologies) for 10 min and imaged with a Gel Doc gel imaging system (Bio-Rad). Quantification was based on relative band intensities. Indel percentage was determined by the formula, $100 \times (1-(1-(b+c)/(a+b+c))1/2)$, where a is the integrated intensity of the undigested PCR product, and b and c are the integrated intensities of each cleavage product.

Small RNA extraction from cells

HEK293T cells were harvested 48 h after transfection and the total RNA was extracted with the miRNeasy mini kit (Qiagen) according to manufacturer's conditions. rRNA was removed using the bacterial Ribo-Zero rRNA removal kit (Illumina).

NGS analysis of *in vitro* and *in vivo* cleavage pattern

RNA-seq libraries were prepared using a derivative of a previously described method (*280*). Briefly, after PNK treatment in the absence and presence of ATP (enrichment of 5′OH and 3′P, respectively) RNA cleavage products were poly-A tailed with *E. coli* Poly(A) Polymerase (NEB), ligated to 5′ RNA adapters using T4 RNA ligase I (NEB) and reverse transcribed with AffinityScript Multiple Temperature Reverse Transcriptase (Agilent Technologies). cDNA was amplified by a fusion PCR method to attach the Illumina P5 adapters as well as unique sample-specific barcodes to the target amplicons (*281*). PCR products were purified by gel-extraction using QiaQuick PCR purification Kit (Qiagen) following the manufacturer's recommended protocol. DNA samples from single nuclei were pre-amplified with SURVEYOR primers (**Table S4.4**) and nested-PCR was performed with NGS primers (**Table S4.5**) before Illumina barcodes were added. Finally, barcoded and purified DNA samples were quantified by Qubit 2.0 Fluorometer (Life Technologies) and pooled in an equi-molar ratio. Sequencing libraries were then sequenced with the Illumina MiSeq Personal Sequencer (Life Technologies).

RNA-sequencing analysis

The prepared cDNA libraries were pooled and sequenced on a MiSeq (Illumina). Pooled sequencing reads were assigned to their respective samples on the basis of their corresponding barcodes and aligned to the proper CRISPR array template sequence using BWA 3. Interval lists were generated using the paired-end alignment coordinates and the intervals were used to extract entire transcript sequences using Galaxy tools (usegalaxy.org/) (*282*). The extracted transcript sequences were analysed using Geneious 9.

AAV DNA constructs

The AAV hSyn1-HA-NLS-AsCpf1-spA vector was generated by PCR amplifying the AsCpf1 encoding sequence using forward PCR primer including HA-NLS (5′-aacacaggaccggtgccaccatgta cccatacgatgttccagattacgcttcgccgaagaaaaagcgcaaggtcgaag cgtccacacagttcgagggctttaccaacctg tatcaggtgagc-3′) and reverse PCR primer including a short poly A signal (spA) (5′-gcggccgcac acaaaaaaccaacacacagatctaatgaaaataaagatcttttatt gaattcttagttgcgcagctcctggatgtaggccagcc-3′) (*279*), and cloning of the resulting PCR template into AAV backbone under control of the human *Synapsin* 1 promoter (hSyn1). For the generation of AAV U6-DR(SapI)-hSyn1-GFP-KASH-hGH (not shown) and U6-*Mecp2-Nlgn3-Drd1* array-hSyn1-GFP-KASH-hGH vectors, gene blocks (Integrated DNA Technologies) encoding U6-DR(SapI) and U6-*Mecp2-Nlgn3-Drd1* array, respectively, have been cloned into AAV hSyn-GFP-KASH-hGH backbone (*279*). All constructs were verified by Sanger sequencing.

4

Production of AAV vectors

AAV1 particles in DMEM culture medium were produced as described previously (*228*). Briefly, HEK293FT cells were transfected with transgene plasmid, AAV1 serotype plasmid and pDF6 helper plasmid using polyethyleneimine (PEI). DMEM culture medium containing low-titre AAV1 particles was collected after 48 h and sterile filtered. For high-titre AAV1/2 production, HEK293FT cells were transfected with AAV1 and AAV2 serotype plasmids in equal ratios, transgene plasmid and pDF6 helper plasmid. 48 h after transfection, cells were harvested, and high-titre AAV1/2 virus was purified on heparin affinity column (*228*). The titre of AAV vectors was determined by real-time quantitative PCR (qPCR) using probe and primers specific for the hSyn1 promoter sequence (Integrated DNA Technologies).

Primary cortical neuron culture

Mice used to obtain neurons for tissue cultures were euthanized according to the protocol approved by the Broad's Institutional Animal Care and Use Committee (IACUC). Primary neurons were prepared from postnatal day P0.5 mouse brains and plated on laminin/poly-D-lysine-coated coverslips (VWR). Briefly, cortices were dissected in ice-cold HBSS (Sigma) containing 50 µg mL^{-1} penicillin/streptomycin (Thermo Fisher) and incubated for 10 min at 37 ℃ with HBSS containing 125 Units papain (Worthington Biochemical) and 400 Units DNase I (Sigma). After enzymatic digestion, the tissues were washed twice in HBSS and gently triturated with a fire-polished Pasteur pipette. Cells were then transferred into neuronal growth medium (Neurobasal A medium, supplemented with B27, Glutamax (Life Technologies) and penicillin/streptomycin) and grown at 37 ℃ and 5% CO_2. For inhibition of glia cell proliferation, cytosine-beta-D-arabinofuranoside (AraC, Sigma) at a final concentration of 10 µM was added to the culture medium after 48 h and replaced by fresh culture medium after 72 h. For AAV1 transduction, cultured neurons were infected with low-titre AAV1 as described previously (*228*). One week after transduction, neurons were harvested for isolating genomic DNA [QuickExtract DNA extraction buffer (Epicentre)] or fixed in 4% paraformaldehyde (PFA) for immunofluorescence staining.

Stereotactic injection of AAV1/2 into the mouse brain

The Broad's Institutional Animal Care and Use Committee (IACUC) approved all animal procedures described here. Craniotomy was performed on adult (12–16 weeks) male C57BL/6N mice according to approved procedures, and 1 µL of 1:1 AAV mixture (AAV hSyn1-HA-NLS-AsCpf1-spA: 2.25×10^{12} Vg mL^{-1}; AAV U6-*Mecp2-Nlgn3-Drd1* array-hSyn1-GFP-KASH-

hGH: 9.7×10^{12} Vg mL^{-1}) was injected into the dorsal dentate gyrus (anterior/posterior: −1.7; mediolateral: +/−0.6; dorsal/ventral: −2.15). The pipette was held in place for 3–5 min after injection to prevent leakage. After injection, the incision was sutured, and post-operative analgesics were administered according to approved IACUC protocol for 3 d following surgery.

Purification of cell nuclei from intact brain tissue

Cell nuclei from AAV1/2-injected hippocampal tissue were purified as described previously (*279*). Briefly, dissected tissue was homogenized in ice-cold homogenization buffer (HB) (320 mM sucrose, 5 mM $CaCl_2$, 3 mM $Mg(Ac)_2$, 10 mM Tris pH7.8, 0.1 mM EDTA, 0.1% NP40, 0.1 mM PMSF, 1 mM β-mercaptoethanol) using 2 ml type A and B Dounce homogenizer (Sigma). For gradient centrifugation, OptiPrep density gradient medium (Sigma) was used. Samples were centrifuged at 10,100*g* (7,500 r.p.m.) for 30 min at 4 ℃ (Beckman Coulter, SW28 rotor). Cell nuclei pellets were resuspended in 65 mM β-glycerophosphate (pH 7.0), 2 mM $MgCl_2$, 25 mM KCl, 340 mM sucrose, and 5% glycerol. Number and quality of purified nuclei was examined using bright-field microscopy.

FACS of cell nuclei

Purified cell nuclei were co-labeled with Vybrant DyeCycle Ruby Stain (1:500, Life Technologies) and sorted using a Beckman Coulter MoFlo Astrios EQ cell sorter (Broad Institute Flow Cytometry Core). Single and population (250–500 nuclei) GFP-KASH$^+$ and GFP-KASH$^-$ nuclei were collected in 96-well plates containing 5 µL of QuickExtract DNA extraction buffer (Epicentre) and spun down at 2,000*g* for 2 min. Each 96-well plate included two empty wells as negative control.

Western blot analysis

AAV-injected dentate gyrus tissues were lysed in 100 µL of ice-cold RIPA buffer (Cell Signalling Technologies) containing 0.1% SDS and protease inhibitors (Roche, Sigma) and sonicated in a Bioruptor sonicator (Diagenode) for 1 min. Protein concentration was determined using the BCA Protein Assay Kit (Pierce Biotechnology, Inc.). Protein samples were separated under reducing conditions on 4–15% Tris-HCl gels (Bio-Rad) and analysed by western blotting using primary antibodies: mouse anti-HA (Cell Signalling Technologies 1:500), mouse anti-GFP (Roche, 1:500), rabbit anti-Tubulin (Cell Signalling Technologies, 1:10,000) followed by secondary anti-mouse and anti-rabbit HRP antibodies (Sigma-Aldrich, 1:10,000). Blots were imaged with Amersham Imager 600.

Immuno-fluorescent staining

4 weeks after viral delivery, mice were transcardially perfused with PBS followed by PFA according to approved IACUC protocol. 30 µm free-floating sections (Leica, VT1000S) were boiled for 2 min in sodium citrate buffer (10 mM tri-sodium citrate dehydrate, 0.05% Tween20, pH 6.0) and cooled down at RT for 20 min. Sections were blocked with 4% normal goat serum (NGS) in TBST (137 mM NaCl, 20 mM Tris pH 7.6, 0.2% Tween-20) for 1 h. Primary antibodies were diluted in TBST with 4% NGS and sections were incubated overnight at 4 ℃. After three washes in TBST, samples were incubated with secondary antibodies for 1 hour at RT. After washing three times with TBST, sections were mounted using VECTASHIELD HardSet Mounting Medium including DAPI and visualized with confocal microscope (Zeiss LSM 710, Ax10 ImagerZ2, Zen 2012 Software). Following primary antibodies were used: mouse anti-NeuN (Millipore, 1:400); chicken anti-GFP (Aves Labs, 1:200–1:400); rabbit anti-HA (Cell Signalling Technologies, 1:100). Anti-HA signalling was amplified using biotinylated anti-rabbit (1:200) followed by strepta-

vidin AlexaFluor 568 (1:500) (Life Technologies). Anti-chicken AlexaFluor488 and anti-mouse AlexaFluor647 secondary antibodies (Life Technologies) were used at 1:1,000.

Accession codes

SRA: PRJNA354073

Randomization and blinding

Neither randomization nor blinding were used in these experiments.

Acknowledgements

We would like to thank F.A. Ran for helpful discussions and overall support, and B. Cartigny and J. van den Bogaerde for technical assistance, and the entire Zhang laboratory for support and advice. 6-His-MBP-TEV, a pET-based vector, was kindly given to us by Doug Daniels of the Broad Institute. M.H. was supported by the Human Frontiers Scientific Program. O.A.A. is supported by a Paul and Daisy Soros Fellowship and a Friends of the McGovern Institute Fellowship. J.S.G. is supported by a D.O.E. Computational Science Graduate Fellowship. E.M.D.G. is supported by the National Institute of Biomedical Imaging and Bioengineering (NIBIB), of the National Institutes of Health (5T32EB1680). K.S. is supported by an NIH grant GM10407, Russian Science Foundation grant 14-14-00988, and Skoltech. J.v.d.O. is supported by Netherlands Organization for Scientific Research (NWO) through a TOP grant (714.015.001). F.Z. is supported by the NIH through NIMH (5DP1-MH100706 and 1R01-MH110049); by the New York Stem Cell, Poitras, Simons, Paul G. Allen Family, and Vallee Foundations; and by David R. Cheng, Tom Harriman, and B. Metcalfe. F.Z. is a New York Stem Cell Foundation Robertson Investigator. The authors plan to make the reagents widely available to the academic community through Addgene and to provide software tools via the Zhang lab website (http://www.genome-engineering.org/).

Author contributions

B.Z., M.H., J.v.d.O., and F.Z. conceived this study and designed the experiments. B.Z., M.H., P.M., I.F., J.K., E.M.D., N.W., S.R.C., O.O.A., J.S.G., W.Y.W. and D.A.S. conducted the experiments. K.S., J.v.d.O., and F.Z. supervised this project. B.Z., M.H., J.v.d.O., and F.Z. wrote the manuscript with input from all authors.

Author information

Correspondence should be addressed to john.vanderoost@wur.nl or zhang@broadinstitute.org.

Competing interests

A patent has been filed relating to the presented data. F.Z. is a founder and scientific advisor for Editas Medicine, and a scientific advisor for Horizon Discovery.

Supplementary Figures and Tables

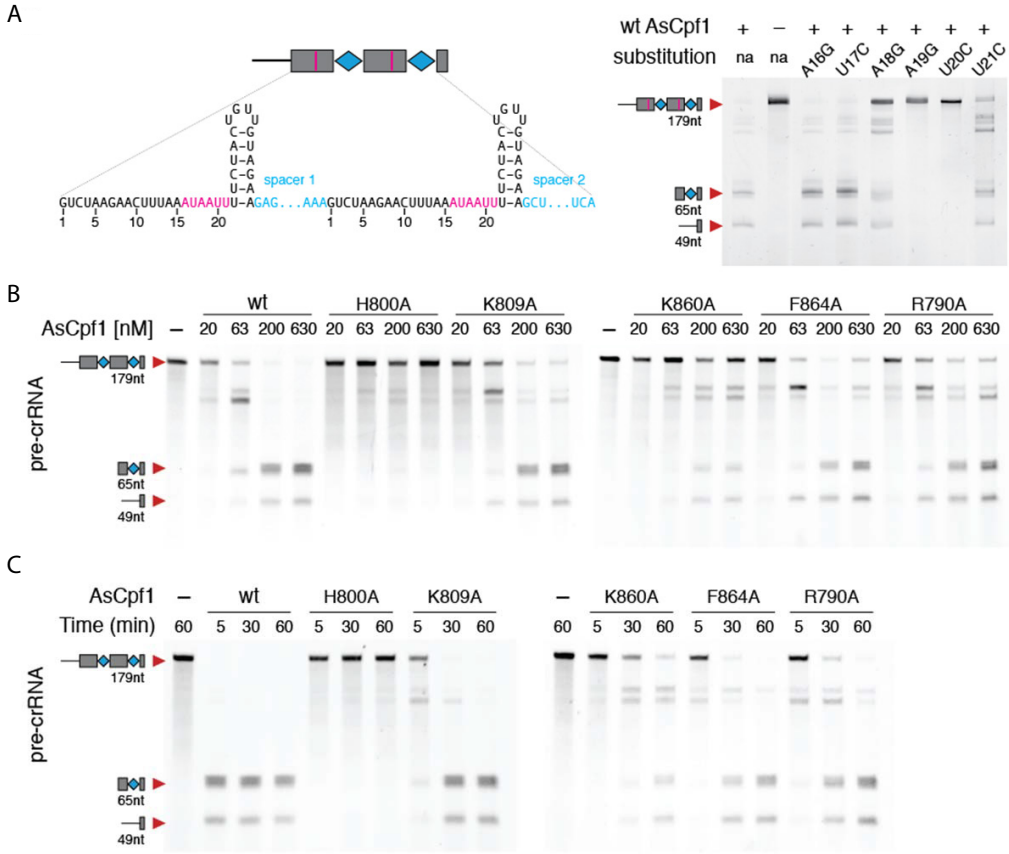


Figure S4.1|Cpf1 mediated pre-crRNA cleavage is sequence and dose dependent. **(A)** Cpf1 mediated pre-crRNA processing is sequence dependent. Single nucleotide substitutions at position A19 and U20 abolish RNA cleavage *in vitro*. 200 nM pre-crRNA was cleaved with 500 nM Cpf1 in 1 hour. **(B, C)** AsCpf1 point mutants, with the exception of H800A, are active at high dose. **(C)** Titration of AsCpf1 mutants reveals pre-crRNA processing at high AsCpf1 protein concentration. Prolonged incubation time allows pre-crRNA processing by AsCpf1 point mutants. Only H800A does not process pre-crRNA to mature crRNA at any time point. 165 nM pre-crRNA was incubated with the indicated concentration.

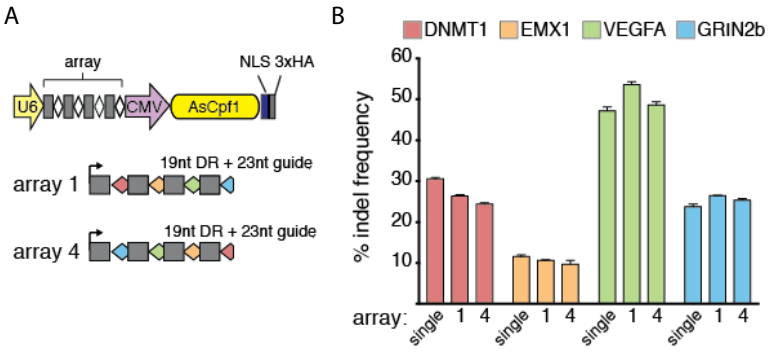


Figure S4.2 | Indel levels are not influenced by guide order. (A) Schematic of multiplex gene editing with AsCpf1, using a single plasmid approach. Two arrays with guides in reversed order are compared (array-1 and array-4). **(B)** Quantification of indel frequencies measured by Surveyor nuclease assay. Guides expressed from array-1 and array-4 result in similar indel frequencies for each targeted gene.

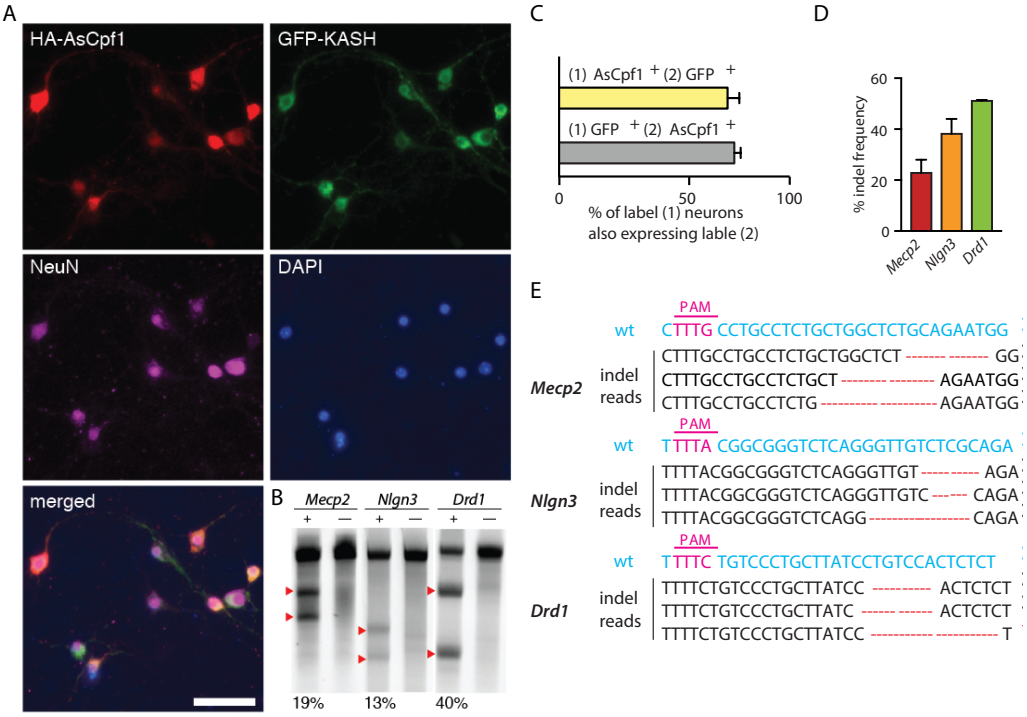


Figure S4.3 | AAV delivery of AsCpf1 and multiplex gene editing in primary neurons and mouse brain.(A) Immunostaining of AsCpf1 (anti-HA antibody, red) and GFP-KASH (anti-GFP antibody, green) in primary cortical neurons (anti-NeuN antibody, magenta) 7 days after viral infection with dual vector system. Nuclei were labelled with DAPI (blue). Scale bar: 25 µm. **(B)** SURVEYOR nuclease assay showing indel formations (+) in all 3 targeted loci. Control neurons (-) were infected with AsCpf1 only (Bottom: Indel percentage; representative images from *n* = 3 independent experiments). **(C)** Quantification of dentate gyrus neurons (DG) efficiently transduced by the dual-vector system *in vivo* (*n* = 581 nuclei from 3 mice). **(D)** NGS indel analysis of modified *Mecp2*, *Nlgn3* and *Drd1* loci in single DG nuclei (*n* = 59 cells from 2 male mice, error bars represent mean ± SEM). **(E)** Representative mutation patterns detected by NGS. Blue, wild-type (wt) sequence; red dashes, deleted bases; PAM sequence marked in magenta.

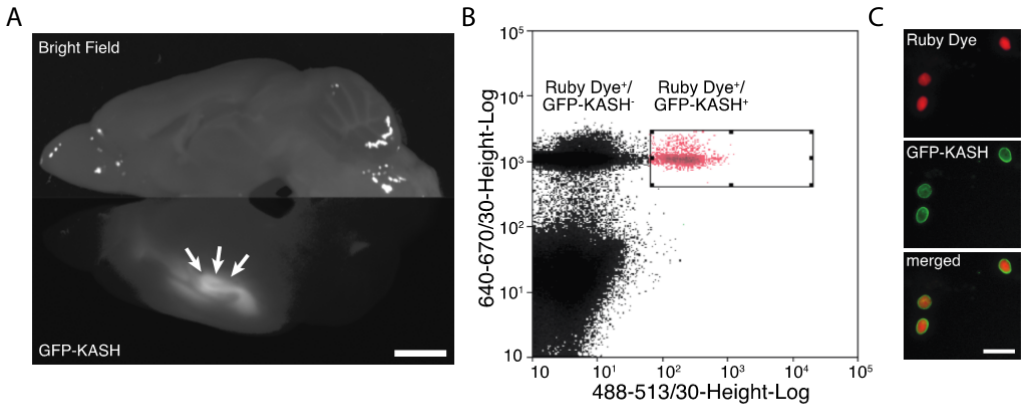


Figure S4.4 | *In vivo* delivery of AAV dual vector system and sorting of targeted cell nuclei from intact brain. **(A)** Sagittal dissection of adult mouse brain 4 weeks after viral delivery shows infected hippocampal formation (bottom). **(B)** Representative FACS plot showing Ruby Dye$^+$/GFP-KASH$^-$ and Ruby Dye$^+$/GFP-KASH$^+$ nuclei populations. **(C)** Representative images of sorted Ruby Dye$^+$/GFP-KASH$^+$ nuclei used for NGS indel analysis. Scale bars: 2 mm in (a), 25 µm in (c).

Table S4.1 | Sequences of pre-crRNA arrays used for *in vitro* cleavage reaction.

pre-crRNA	Sequence (5' - 3')
4 Spacer pre-crRNA	GGGGGUCUUUUUUUGCUGAUUUAGGCAAAAACGGGUCUAAGAACUUUAAAUAAUUUCUACU GUUGUAGAUGAGAAGUCAUUUAAUAAGGCCACUGUUAAAAGUCUAAGAACUUUAAAUAAUU UCUACUGUUGUAGAUGCUACUAUUCCUGUGCCUUCAGAUAAUUCAGUCUAAGAACUUUAAA UAAUUUCUACUGUUGUAGAUGUCUAGAGCCUUUUGUAUUAGUAGCCGGUCUAAGAACUUUA AAUAAUUUCUACUGUUGUAGAUUAGCGAUUUAUGAAGGUCAUUUUUUUGUCUAGCUUUAAU GCGGUAGUUUAUCACAGUUAAAUUGCUAACG
2 Spacer pre-crRNA	UAGGUCUUUUUUUGCUGAUUUAGGCAAAAACGGGUCUAAGAACUUUAAAUAAUUUCUACUG UUGUAGAUGAGAAGUCAUUUAAUAAGGCCACUGUUAAAAGUCUAAGAACUUUAAAUAAUUU CUACUGUUGUAGAUGCUACUAUUCCUGUGCCUUCAGAUAAUUC
Control RNA	UACGCCAGCUGGCGAAAGGGGGAUGUGCUGCAAGGCGAUUAAGUUGGGUAACGCCAGGGUU UUCCCAGUCACGACGUUGUAAAACGACGGCCAGUGAAUUCGAGCUCGGUACCCGGGNNNNN NNNGAGAAGUCAUUUAAUAAGGCCACUGUUAAAAAGCUUGGCGUAAUCAUGGUCAUAGCUG UUUCCUGUGUGAAAUUGUUAUCCGCUCACAAUUCCACACAACAUACGAGCCGGAAGCAUAA AGUGUAAAGCCUGGGGUGCCUAAUGAGUGAGCUAACUCACAUUAAUUGCGUU

Table S4.2 | Cpf1 guide sequences used for single and pre-crRNA array expression.

Guide	Sequence (5' - 3')
DNMT1 23 nt guide	CTGATGGTCCATGTCTGTTACTC
EMX1 23 nt guide	TGGTTGCCCACCCTAGTCATTGG
VEGFA 23 nt guide	CTAGGAATATTGAAGGGGGCAGG
GRIN2b 23 nt guide	GTGCTCAATGAAAGGAGATAAGG
DNMT1 30 nt guide	CTGATGGTCCATGTCTGTTACTCGCCTGTC
EMX1 30 nt guide	TGGTTGCCCACCCTAGTCATTGGAGGTGAC
VEGFA 30 nt guide	CTAGGAATATTGAAGGGGGCAGGGGAAGGC
GRIN2b 30 nt guide	GTGCTCAATGAAAGGAGATAAGGTCCTTGA

Table S4.3 | DNA oligonucleotides for array cloning.

Array	Sequence (5' - 3')
array 1 T1	AGATCTGATGGTCCATGTCTGTTACTCAATTTCTACTCTTGTAGATTGGTTGCCCAC
array 1 T2	CCTAGTCATTGGAATTTCTACTCTTGTAGATCTAGGAATATTGAAGGGGGCAGGAATTTCTACTCTTGTAGATGTGCTCAATGAAAGGAGATAAGG
array 1 B1	AAAACCTTATCTCCTTTCATTGAGCACATCTACAAGAGTAGAAATTCCTGCCCCCTT
array 1 B2	CAATATTCCTAGATCTACAAGAGTAGAAATTCCAATGACTAGGGTGGGCAACCAATCTACAAGAGTAGAAATTGAGTAACAGACATGGACCATCAG
array 2 T1	AGATCTGATGGTCCATGTCTGTTACTCGCCTGTCAATTTCTACTCTTGTAGATTGGTTGCCCACCCTAGTC
array 2 T2	TGAAGGGGGCAGGGGAAGGCAATTTCTACTCTTGTAGATGTGCTCAATGAAAGGAGATAAGGTCCTTGA
array 2 B1	AAAATCAAGGACCTTATCTCCTTTCATTGAGCACATCTACAAGAGTAGAAATTGCCTTCCCCTGCCCCCTT
array 2 B2	CAATATTCCTAGATCTACAAGAGTAGAAATTGTCACCTCCAATGACTAGGGTGGGCAACCAATCTACAAGAGTAGAAATTGACAGGCGAGTAACAGACATGGACCATCAG
array 3 T1	AGATGTCAAAAGACCTTTTTAATTTCTACTCTTGTAGATCTGATGGTCCATGTCTGTTACTCGCCTGTCGTCAAAAGACCTTTTTAATTTCTACTCTTGTAGATTGGTTGCCCACCCTAGTCATTGGAGGTGACGTCAAAAGACCTTTTTAATTTCTACTCTTGTAGATCTAGGAATATT
array 3 T2	GAAGGGGGCAGGGGAAGGCGTCAAAAGACCTTTTTAATTTCTACTCTTGTAGATGTGCTCAATGAAAGGAGATAAGGTCCTTGAGTCAAAAGACCTTTTTAATTTCTACTCTTGTAGAT
array 3 B1	AGAAATTAAAAAGGTCTTTTGACGCCTTCCCCTGCCCCCTTCAATATTCCTAGATCTACAAGAGTAGAAATTAAAAAGGTCTTTTGACGTCACCTCCAA
array 3 B2	TGACTAGGGTGGGCAACCAATCTACAAGAGTAGAAATTAAAAAGGTCTTTTGACGACAGGCGAGTAACAGACATGGACCATCAGATCTACAAGAGTAGAAATTAAAAAGGTCTTTTGAC
array 4 T1	AGATGTGCTCAATGAAAGGAGATAAGGAATTTCTACTCTTGTAGATCTAGGAATATT
array 4 T2	GAAGGGGGCAGGAATTTCTACTCTTGTAGATTGGTTGCCCACCCTAGTCATTGGAATTTCTACTCTTGTAGATCTGATGGTCCATGTCTGTTACTC
array 4 B1	AAAAGAGTAACAGACATGGACCATCAGATCTACAAGAGTAGAAATTCCAATGACTAG
array 4 B2	GGTGGGCAACCAATCTACAAGAGTAGAAATTCCTGCCCCCTTCAATATTCCTAGATCTACAAGAGTAGAAATTCCTTATCTCCTTTCATTGAGCAC

Table S4.4 | PCR primers for amplification of DNA regions for SURVEYOR nuclease assay.

Primers	Sequence (5' - 3')
DNMT1 FW	CTGGGACTCAGGCGGGTCAC
DNMT1 RV	CCTCACACAACAGCTTCATGTCAGC
EMX1 FW	CCATCCCCTTCTGTGAATGT
EMX1 RV	GGAGATTGGAGACACGGAGA
VEGFA FW	CTCAGCTCCACAAACTTGGTGCC
VEGFA RV	AGCCCGCCGCAATGAAGG
GRIN2b FW	GCATACTCGCATGGCTACCT
GRIN2b RV	CTCCCTGCAGCCCCTTTTTA
Mecp2 FW	GGTCTCATGTGTGGCACTCA
Mecp2 RV	TGTCCAACCTTCAGGCAAGG
Nlgn3 FW	GTAACGTCCTGGACACTGTGG
Nlgn3 RV	TTGGTCCAATAGGTCATGACG
Drd1 FW	TGGCTAAGCCTGGCCAAGAACG
Drd1 RV	TCAGGATGAAGGCTGCCTTCGG

Table S4.5 | PCR primers for amplification of DNA regions for next generation sequencing.

Primers	Sequence (5' - 3')
NGS DNMT1 FW	CCATCTCATCCCTGCGTGTCTCCTGAACGTTCCCTTAGCACTCTGCC
NGS DNMT1 RV	CCTCTCTATGGGCAGTCGGTGATGCCTTAGCAGCTTCCTCCTCC
NGS EMX1 FW	CCATCTCATCCCTGCGTGTCTCCGGGCTCCCATCACATCAACCG
NGS EMX1 RV	CCTCTCTATGGGCAGTCGGTGATGCCAGAGTCCAGCTTGGGCCC
NGS VEGFA FW	CCATCTCATCCCTGCGTGTCTCCCAGGGGTCACTCCAGGATTCCA
NGS VEGFA RV	CCTCTCTATGGGCAGTCGGTGATGCATTGGCGAGGAGGGAGCAG
NGS GRIN2b FW	CCATCTCATCCCTGCGTGTCTCCGTTCAAGGATTTCTGAGGCTTTTGAAAG
NGS GRIN2b RV	CCTCTCTATGGGCAGTCGGTGATGGGGCTTCATCTTCAACTCGTCGAC
NGS Mecp2 FW	CCATCTCATCCCTGCGTGTCTCCGGAAAAGTCAGAAGACCAGG
NGS Mecp2 RV	CCTCTCTATGGGCAGTCGGTGATGGTGGGGTCATCATACATAGG
NGS Nlgn3 FW	CCATCTCATCCCTGCGTGTCTCCACCCCGAGGATGGTGTCTCG
NGS Nlgn3 RV	CCTCTCTATGGGCAGTCGGTGATGGGTAGAAGGCGTAGAAGTAGG
NGS Drd1 FW	CCATCTCATCCCTGCGTGTCTCCAAGCCACCGGAAGTGCTTTCC
NGS Drd1 RV	CCTCTCTATGGGCAGTCGGTGATGCACAGCTTTCCAGGGCATGACC

Chapter 5

Characterising the CRISPR–Cas type V-U1 C2c4 effector enzyme

Prarthana Mohanraju*, Wen Y. Wu*, Sjoerd C. A. Creutzburg, Karlijn Keessen, Tahseen S. Khan, Stijn Prinsen, Winston X. Yan, David A. Scott, Kira S. Makarova, Eugene V. Koonin, John van der Oost

* contributed equally
Adapted from:
Characterising the CRISPR–Cas type V-U1 C2c4 effector enzyme.
Manuscript in preparation

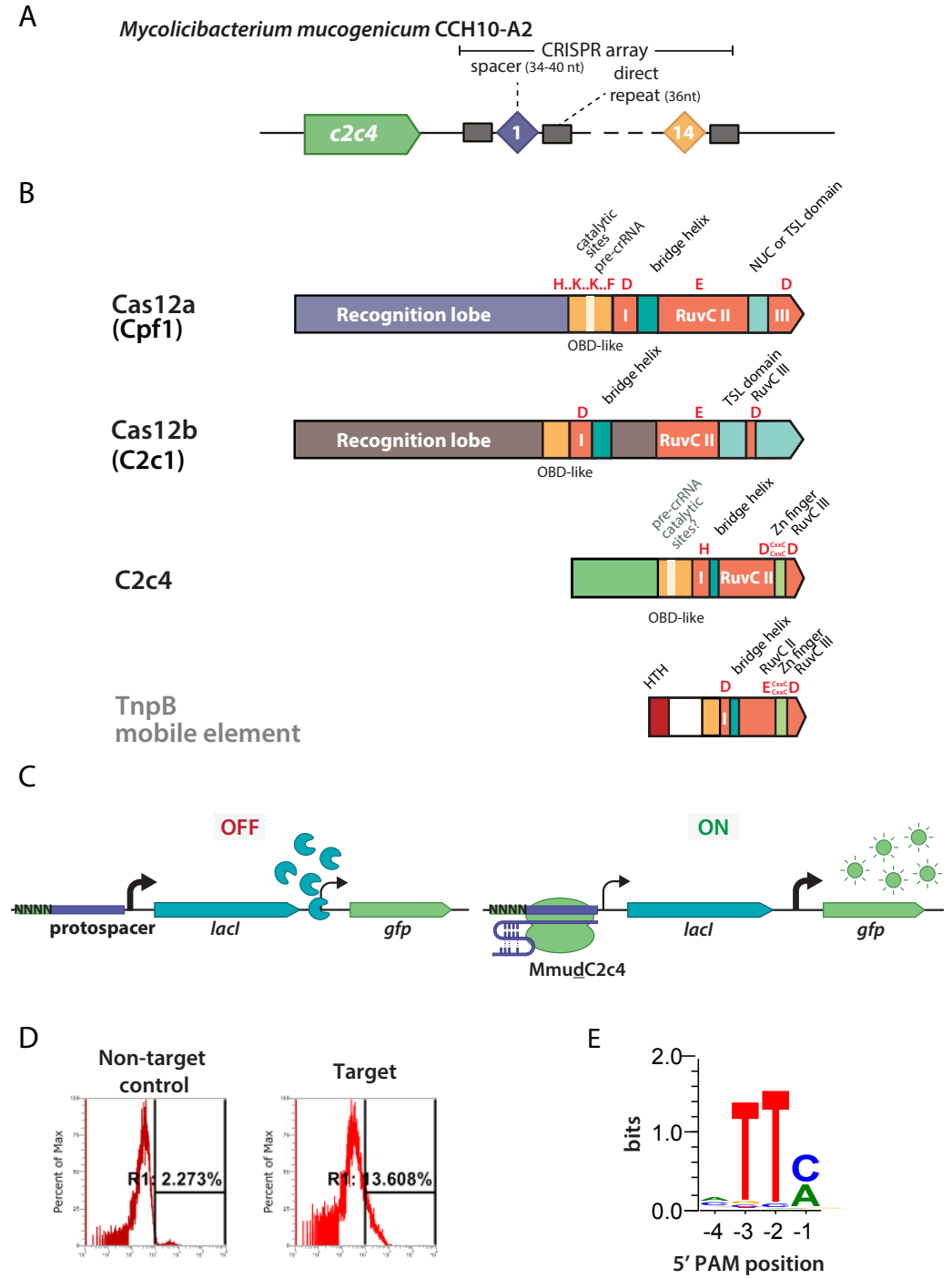
Abstract

CRISPR–Cas systems are prokaryotic heritable adaptive immune systems that have been repurposed as powerful genome editing tools in a wide range of organisms. These tools use RNA-guided Cas nucleases whose large size (928 to 1400 amino acids) has been considered indispensable to their specific DNA- or RNA-targeting activities. Here we present a novel Cas protein from the CRISPR–Cas type V-U1 system from Mycolicibacterium mucogenicum *CCH10 (596 amino acids). Despite its small size, MmuC2c4 is capable of targeting and binding to double-stranded DNA (dsDNA). Akin to most type V proteins, MmuC2c4 recognizes a 5'-TTN-3' PAM on a double-stranded target DNA. Unexpectedly, MmuC2c4 enzyme does not cleave dsDNA and analysis in* E. coli *indicates a crRNA-guided MmuC2c4-mediated transcriptional silencing. By leveraging this property, MmuC2c4 has been used for single- and multiplex- transcriptional silencing in* E. coli. *Finally, experiments suggest that the RuvC-dependent ribonuclease activity of MmuC2c4 appears to enhance the silencing effect.*

Results

The everlasting biological arms-race between prokaryotes (bacteria and archaea) and viruses has resulted in the evolution of remarkably diverse CRISPR–Cas defence systems in prokaryotes against their invaders (*20*, *283*, *284*). The key players of the CRISPR–Cas systems are the Cas protein complexes, consisting of the RNA-guided DNA or RNA cleaving nucleases (*29*). Based on the unique Cas effector complexes, CRISPR–Cas systems are currently grouped into two broad classes and subdivided into six types. Class 1 systems use a multi-protein effector complex to achieve target recognition and interference, while the class 2 systems use a single protein with multiple functional domains for target recognition and interference (*31*, *82*). The facile programmability of class 2 CRISPR–Cas nucleases has allowed for their repurposing for genome editing, transcriptional regulation, and diagnostic tools (*285*, *286*). Class 2 consists of types II, V and VI, represented by the signature nucleases Cas9, Cas12 and Cas13, respectively. Cas9 cleaves double-stranded (ds) DNA using its HNH and RuvC nuclease domains while the first characterised Cas12 variants (subtypes V-A and V-B) have been demonstrated to cleave dsDNA specifically and single-stranded (ss) DNA non-specifically using a single RuvC domain (*51*, *52*, *83*, *287*). Both Cas9 and Cas12a cleave dsDNA adjacent to a short sequence, termed the Protospacer Adjacent Motif (PAM) (*51*, *83*). Cas13 is the only known Cas nuclease to exclusively cleave RNA using two HEPN domains (*224*, *288*). Although these nucleases have been widely used for genome engineering, the large size of Cas9, Cas12a and Cas13 (928-1400 amino acids) places constraints on some cellular delivery approaches that may limit certain applications including therapeutics (*289*, *290*). By screening rapidly growing genomic and metagenomic databases, partly as a quest for potential novel genome editing tools, eight new functionally diverse Cas12-like systems have recently been identified and characterised: type V-C to V-K (Cas12c-k) (*74*, *76*, *77*, *80*, *81*). Some of these effectors are nearly half the size of the smallest Cas9 or Cas12 proteins potentially making them highly appealing for packaging in FDA approved safe-to-use Adeno-Associated Viruses (AAVs) for use *in vivo* genome engineering applications and therapeutics (*289*, *290*). Thus, discovery and unravelling the mechanism of novel and compact CRISPR–Cas systems holds great potential for new and improved additional technological advancements.

Using CRISPR arrays as the search seed in the computational class 2 discovery pipeline (*82*) yielded several variants of type V loci, tentatively called uncharacterised (U) subtypes V-U1, -U2, U3, U-4 and -U5 (*82*, *284*). The putative effector proteins of the type V-U loci are much smaller than the archetypal class 2 effectors, comprising between 500 and 700 amino acids (between the size of transposon-encoded TnpB proteins and the typical size of class 2 effectors). These type V-U proteins show highly significant similarity to the TnpB-like proteins and appear to have evolved independently from distinct TnpB families (*82*). The resemblance of type V-U1 proteins to type V nucleases suggests that they may have existed as an ancestral class 2 CRISPR system. They most likely evolved from a distinct, 'domesticated' TnpB-like transposase that gained domains over time, resulting in Cas12 variants with different features, and leading to the large signature type V nucleases like Cas12a (*82*). Despite the occurrence of the characteristic bacterial RuvC-like domains found in the type V-U1 proteins, their small size and the absence of other *cas* genes near the CRISPR array made it unlikely for these systems to function as stand-alone CRISPR effectors without additional partners (*82*). Nonetheless, at least some of them were predicted to be active based on their respective CRISPR arrays which contain spacers homologous to phage genome sequences (*82*). Recently, the former type V-U5 effector, Cas12k (formerly, C2c5), containing a naturally inactivated RuvC-like nuclease domain was


A
Mycolicibacterium mucogenicum CCH10-A2
CRISPR array
spacer (34-40 nt)
direct repeat (36nt)
c2c4
1
14
B
Cas12a (Cpf1)
catalytic sites
pre-crRNA
bridge helix
NUC or TSL domain
H..K..K..F D E D
Recognition lobe
I
RuvC II
III
OBD-like
Cas12b (C2c1)
bridge helix
TSL domain
RuvC III
D E D
Recognition lobe
I
RuvC II
OBD-like
C2c4
pre-crRNA catalytic sites?
bridge helix
Zn finger
RuvC III
H
I
RuvC II
OBD-like
TnpB mobile element
HTH
bridge helix
RuvC II
Zn finger
RuvC III
D
I
C
OFF
ON
NNNN
protospacer
lacI
gfp
NNNN
lacI
gfp
MmudC2c4
D
Non-target control
Target
Percent of Max
R1: 2.273%
R1: 13.608%
E
bits
2.0
1.0
0.0
-4 -3 -2 -1
5′ PAM position

shown to be hijacked by Tn7-like transposons for directed DNA transposition via crRNA-guided targeting (*81*). However, the functionality of the other four subtype V-U systems remains to be uncovered.

Of the five V-U variants, subtype V-U1 is the most prevalent in different bacteria, whereas the remaining subtypes are largely limited in their spread to particular bacterial taxa (*82*). The evolutionary stability in terms of sequence conservation and consistent association with CRISPR arrays with diverse spacers (*82*), led us to hypothesize that these type V-U1 loci encode biologically functional enzymes with nucleic acid targeting activity despite their small size. To test the hypothesis, we studied the type V-U1 CRISPR-associated nuclease, C2c4 from *Mycolicibacterium mucogenicum* CCH10-A2 (MmuC2c4) (**Figure 5.1A**). MmuC2c4 contains a RuvC-like nuclease domain near the C-terminal end, with an organization reminiscent to that found in other type V nucleases (**Figure 5.1B**, **S5.1**).

We initially performed an *in-silico* prediction for the presence or absence of a tracrRNA-like sequence in the MmuC2c4 loci using a previously described approach (*248*). Using this approach, no tracrRNA-like sequences have been detected in the adjacent DNA sequences. This is in line with the fact that the C2c4 CRISPR arrays have partial palindromic sequences (**Figure S5.2**) a feature that appears to correlate with tracrRNA-independent guide processing systems.

Next, to functionally characterise the type V-U1 protein, we transformed *Escherichia coli* cells with a plasmid containing the codon-harmonized *mmuc2c4* gene and a minimal CRISPR array (repeat-spacer-repeat). After purification of the MmuC2c4 protein to homogeneity (**Figure S5.3A**), subsequent analysis revealed the presence of co-purified RNAs, that are presumably around the size of the mature crRNAs (**Figure S5.3B**). This strongly suggests that MmuC2c4 associates with a crRNA (see below, determination of PAM sequence). However, sequencing of the small RNAs must be performed to corroborate this finding and to map the cleavage sites on the crRNA (**Figure S5.3B**). The absence of a predicted tracrRNA and the size of the co-purified RNAs, suggested a potential for crRNA biogenesis by the effector protein itself, as has been reported for type V-A, V-I and VI (*80*, *187*, *256*, *288*). Therefore, we performed an *in vitro* pre-crRNA processing assay using purified recombinant MmuC2c4 protein and a minimal pre-crRNA (repeat-spacer-repeat-spacer-repeat). Processing of the pre-crRNA to intermediates and seemingly mature guides was observed (**Figure S5.3C**), which was independent of the presence of an active RuvC-like domain (data not shown). The conservation of catalytic residues similar to those found to be responsible for pre-crRNA processing by Cas12a and the presence of a beta-stranded domain similar to the OB-fold like domain in MmuC2c4 (**Figure 5.1B**, **S5.1**)

Figure 5.1 *(preceding page)*| The *Mycolicibacterium mucogenicum* CCH10-A2 C2c4 protein recognizes ds-DNA targets flanked by a 5′-TTM PAM. (A) Organization of the CRISPR–Cas locus on the genome of *Mycolicibacterium mucogenicum* CCH10-A2. Dark grey boxes are perfect direct repeats; the light grey box at the right side of the CRISPR array indicates a slightly degenerated repeat, generally indicative of the 3' end of the transcribed precursor-crRNA. **(B)** Domain architectures of Cas12a (PDB: 5B43), Cas12b (PDB:5WQE), C2c4 (WP_061006603.1) and TnpB (mobile element) proteins are compared. Protein lengths are drawn to scale. **(C)** Schematic illustrating the *in vivo* PAM screen achieved by PAM-SCANR. It consists of a library of randomized 5' PAM sequences (4N) cloned upstream of the *lacI* promoter. Immediately downstream of *lacI* is the LacI-dependent *lacZ* promoter controlling expression of GFP. A catalytically dead MmuC2c4 (Mmu<u>d</u>C2c4) protein is targeted to a protospacer within the *lacI* promoter, resulting in GFP fluorescence only in the presence of a functional PAM. **(D)** Cells harbouring 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. **(E)** Plasmids from the FACS-sorted cells were extracted and sequenced to determine functional PAM sequences. Sequence logo for the MmuC2c4 PAM as determined by NGS sequencing of plasmids from sorted fluorescent cells.

are most likely responsible for pre-crRNA processing by MmuC2c4. Further experiments are currently underway to assess the cleavage sites and investigate the active site residues responsible for the pre-crRNA processing using site-directed mutagenesis.

The PAM sequence plays a central role in target selection in dsDNA cleaving CRISPR systems. In the absence of a PAM, the Cas nucleases cannot recognize the target, even if it is perfectly complementary to the spacer (*45*). To test whether MmuC2c4 requires a PAM and can conduct crRNA-guided dsDNA interference, we adapted the previously developed PAM-SCANR assay (*291*), a high-throughput *E. coli*-based positive and tuneable screen for assessment of PAMs (**Figure 5.1C**). It is based on a catalytically inactive crRNA-guided Cas effector blocking the –35 element within the promoter upstream of *lacI*. In the absence of binding (due to a non-functional PAM) by the inactive Cas effector, the expressed LacI repressor blocks the lac operator in the promoter of the green fluorescent protein (*gfp*) gene. In the case of binding of the inactive Cas nuclease (due to a functional PAM), *lacI* expression will be inhibited, hence resulting in expression of GFP (*291*). We generated an effector plasmid (pCas-MmudC2c4) encoding a catalytically inactive *mmuc2c4* gene [single mutant of one of the RuvC-II active site residues: Asp 485 --> Ala (D485A)], a CRISPR array plasmid (pCRISPR-PS), with a spacer targeting a 5'-NNNN-3' PAM library placed upstream of the –35 element of the promoter of *lacI* in the target PAM-SCANR plasmid (pTarget-PS). A CRISPR array plasmid with a non-targeting spacer (CRISPR-NT) was used as a control. Gene repression of the LacI repressor by crRNA-guided MmudC2c4 binding of the dsDNA containing a functional PAM would lead to the expression of the GFP reporter. *E. coli* cells were transformed with the pCas-MmudC2c4, either pCRISPR-PS or -NT and pTarget-PS plasmids and subsequently the GFP fluorescent cells were isolated through fluorescence-activated cell sorting (FACS) (**Figure 5.1D**). Comprehensive screening based on next-generation sequencing of the pre-sorted and post-sorted PAM libraries and analyses of the target-flanking sequences revealed that the binding of target dsDNA by MmudC2c4 depends on a 5'-NTTM-3' PAM (**Figure 5.1E**). Weak functional PAMs were also detected by tuning the Isopropyl β-D-1-thiogalactopyranoside (IPTG) levels to titrate active LacI repressors (**Figure S5.4**). The presence of a T nucleotide at the -2 position of the 5'- PAM is most crucial. Thus, the PAM recognized by MmuC2c4 is similar to that of the other known type V effector proteins (*80*, *83*, *292*, *293*).

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 S5.5A**). *E. coli* cells harbouring the CRISPR array plasmid with a spacer targeting the promoter (pCRISPR-promoter) and pCas-MmudC2c4 were transformed with the pTarget-GFP plasmids and assessed for repression of GFP fluorescence, as a result of efficient dsDNA binding (**Figure S5.5A**). For comparison, we also analysed the catalytically inactive type V-A effector (dCas12a) of *Francisella novicida* U112 (pCas-FndCas12a), with its corresponding crRNA guide targeting the same protospacer (**Figure S5.5B**). Efficient GFP repression was observed for all the tested PAM variants, confirming the PAM sequence as 5'-TTN-3' together with robust in vivo crRNA-guided dsDNA binding by MmudC2c4 (**Figure S5.5B**).

Characteristic to most DNA-targeting class 2 interference complexes is their ability to recognize and cleave both dsDNA and single-stranded DNA (ssDNA) substrates (*51*, *83*, *221*, *294*, *295*). Therefore, to test crRNA-guided dsDNA interference, a target plasmid containing a 5' CTTA PAM adjacent to the previously used PAMSCANR protospacer (pTarget-CTTA) was generated. It was transformed into *E. coli* cells harbouring the effector plasmid encoding the wild-

type MmuC2c4 protein (pCas-MmuC2c4) 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 harbouring the pCas-MmuC2c4 and pCRISPR-PS, as compared with the strain harbouring the pCas-MmuC2c4 and the control pCRISPR-NT plasmids (**Figure S5.6A**). In contrast, the dsDNA targeting Cas12a control did result in substantially lower number of transformants (**Figure S5.6A**). This indicates that, at least under the tested conditions, CRISPR-MmuC2c4 does not cleave dsDNA in the heterologous *E. coli* host (**Figure S5.6A**).

To ascertain the inability of MmuC2c4 to cleave dsDNA, we repeated the same experiment, but with a plasmid (pCRISPR-GFP) containing a different spacer targeting the end of the *gfp* gene. As seen previously, we did not observe any decrease in the number of the transformants as compared to the non-target control. Interestingly, we observed a drop in the GFP fluorescence signal for the cells harbouring the pCas-MmuC2c4, pCRISPR-GFP and pTarget-GFP plasmid (**Figure S5.6B**). This GFP repression activity was much lower or undetectable in the cells harbouring the pCas-MmudC2c4, pCRISPR-GFP and pTarget-GFP plasmids. This suggests that activation of the MmuC2c4 results in increased gene repression, most likely due to cleavage of the mRNA transcript (**Figure S5.6B**). The recently characterised Cas12g nuclease mediates metal-ion dependent RNA and ssDNA cleavage using an intact RuvC domain that was previously known to cleave only DNA (*80*).

Thus, on the basis of this observation, we hypothesized that MmuC2c4 might possess target activated nonspecific ssRNA cleavage activity. To test this possibility, we incubated MmuC2c4:crRNA complex with a complementary dsDNA plasmid target and a target or non-target RNA. However, under these conditions MmuC2c4 appeared to be incapable of cleaving (non)target RNA *in vitro* (**Figure S5.7**), suggesting that additional endogenous factors might be required for *in vivo* RNA targeting.

To further investigate the dsDNA target-dependent RNA cleavage activity by MmuC2c4, we cloned a target plasmid (pTarget-Operon) containing a bi-cistronic operon with two fluorescence reporter genes, *rfp and gfp* (**Figure 5.2A**). *E. coli* cells harbouring either the pCas-MmudC2c4 or pCas-MmuC2c4 and the pTarget-operon were transformed with a set of different CRISPR array plasmids (pCRISPR-A1_F2) containing spacers targeting either the coding or the non-coding strand at different locations throughout the entire operon (**Figure 5.2A**). As expected, for the crRNA–target dsDNA pairs on and nearby the promoter region, low GFP and RFP fluorescent signal was attained, indicating high transcriptional silencing of both the genes (**Figure 5.2B and C**, crRNAs A1/A2). Strikingly, although the transcriptional silencing of the fluorescent reporter genes by the MmudC2c4 protein was weak for the crRNA–target dsDNA pairs towards the end of the operon (crRNAs D2/E1/E2), relatively strong repression of both the red and green fluorescence signal was observed for the cells expressing the wild-type MmuC2c4 (**Figure 5.2B and C**, crRNAs D2/E1). The loss of red as well as the green florescence upon binding to the latter *gfp* gene indicates that transcription and/or translation of the whole mRNA is being affected by MmuC2c4 (**Figure 5.2B and C**, crRNAs D1/D2). Moreover, the crRNA-target dsDNA pairs outside the transcription region (crRNAs E2/F1/F2) resulted in reduced or undetectable loss of fluorescence, hinting towards a transcription-associated *trans* cleavage of nascent mRNA activity of MmuC2c4.

To test the specificity of the mRNA cleavage, we cloned a target plasmid (pTarget-divergent) with two fluorescence reporter genes, *rfp and gfp*, under the transcriptional control of two

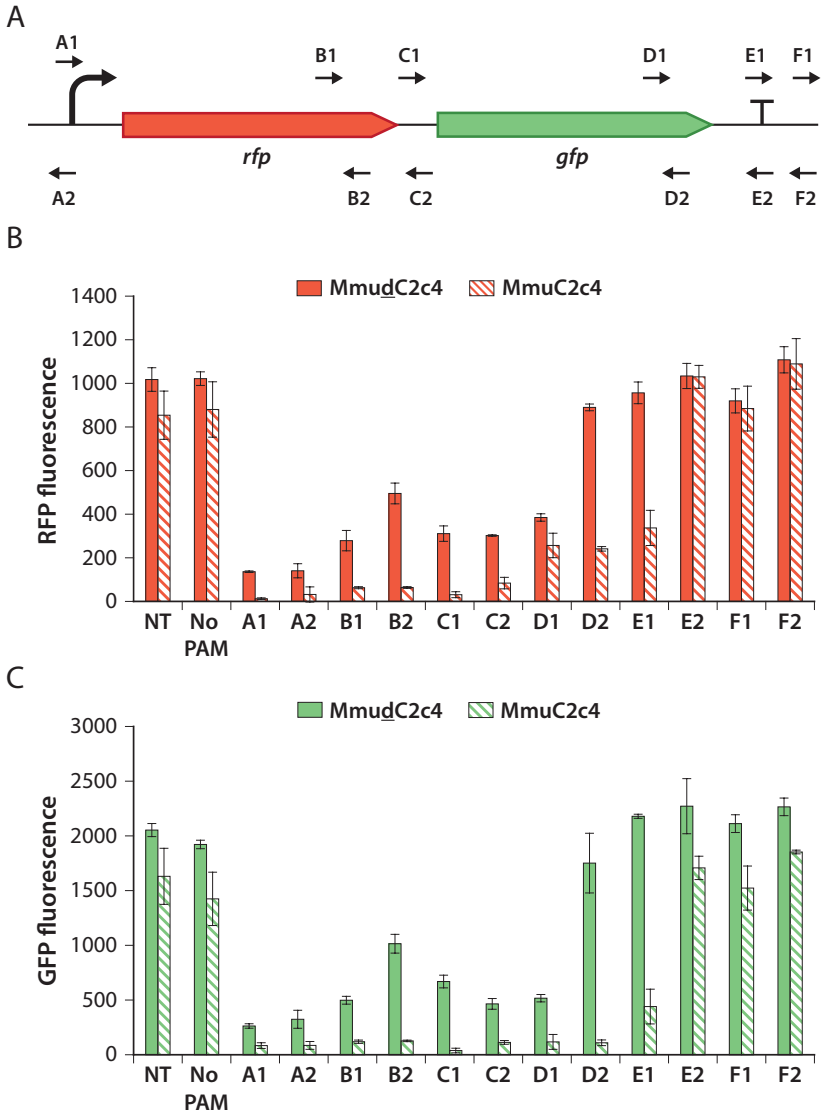


Figure 5.2 | Comparison of transcriptional silencing by MmuC2c4 and MmudC2c4. **(A)** Schematic of the pTarget-operon, including the bi-cistronic operon encoding the *rfp* and *gfp* genes. The arrows indicate the crRNAs used for targeting by MmuC2c4 and MmudC2c4 proteins (A1 to F2). **(B)** RFP fluorescence detected in the cells upon MmudC2c4 and MmuC2c4 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 protospacer next to a GGGC PAM. **(C)** GFP fluorescence detected in the cells upon MmudC2c4 and MmuC2c4 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 protospacer next to GGGC motif (non-functional PAM).

independent constitutive promoters, P_{taq} and P_{laclq} respectively (**Figure 5.3A**). *E. coli* cells harbouring either the pCas-MmudC2c4 or pCas-MmuC2c4 and the pTarget-divergent were transformed with a set of different CRISPR array plasmids (pCRISPR-a_f) containing spacers targeting different locations on the promoters and on the coding strand of either *rfp* or *gfp* (**Fig-**

ure 5.3A). Specific repression of only the targeted reporter gene was observed, and yet again, the wild-type MmuC2c4 performed better than the MmudC2c4 in silencing the expression of the reporter gene (**Figure 5.3B and C**). The 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 (**Figure 5.3B**).

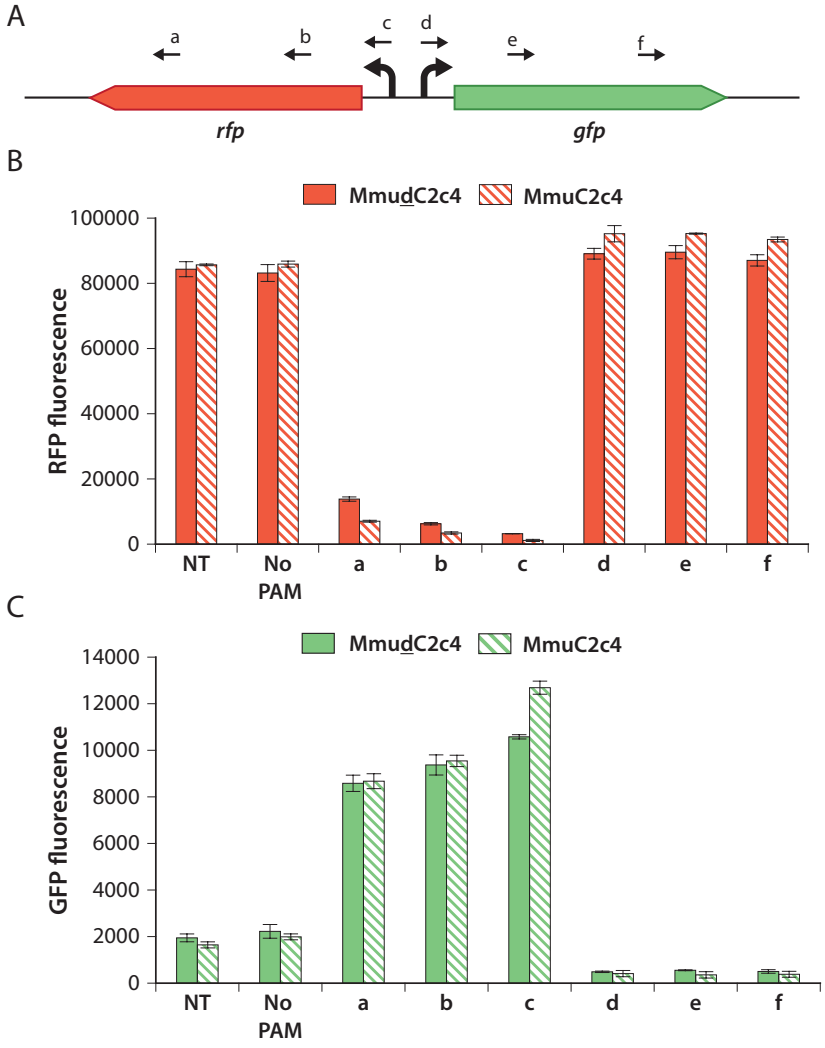


Figure 5.3|Mmuc2c4 is activated by dsDNA binding to cleave nascent RNA transcripts. (A) Schematic of the pTarget-divergent, including the *rfp* and *gfp* genes under the transcriptional control of two different constitutive promoters, P_{taq} and P_{laclq}. The arrows indicate crRNAs used for targeting by MmuC2c4 and MmudC2c4 proteins using the respective spacers (a to f). **(B)** RFP fluorescence detected in the cells upon MmudC2c4 and MmuC2c4 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 protospacer next to a GGGC PAM. **(C)** GFP fluorescence detected in the cells upon MmudC2c4 and MmuC2c4 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 protospacer next to a GGGC PAM (non-functional PAM).

Subsequently, to investigate MmuC2c4 mismatch tolerance, we introduced single mismatches,

tiled 2-nucleotide and tiled 4-nucleotide mismatches across the protospacer located in the target *gfp* gene. MmuC2c4 appeared to be relatively tolerant to most single mismatches (**Figure S5.8A**), except for the mismatch at the PAM-proximal position 8th. This is in contrast to Cas12a which is highly sensitive to single or double substitution in most positions between 1 and 18 (*246*). Although double and quadruple mismatches at PAM-proximal positions 1 to 11 severely impaired the Mmuc2c4 silencing activity (**Figure S5.8B and C**), a clear seed-like sequence (*95*) was not observed. Notably, even though some mismatches impaired the activity of MmudC2c4 in GFP repression, the effect of the mismatches on GFP silencing by the wild-type MmuC2c4 was much less pronounced (**Figure S5.8A to C**). Collectively, these observations point toward a novel mechanism where crRNA-guided binding of MmuC2c4 to a transcriptionally active dsDNA triggers it to cleave nascent mRNA. Cleavage of the mRNA appears to be confined to the transcript of the target DNA, rather than collateral cleavage activity that has been reported for some of the type V and type VI effectors (*80*, *288*).

Moreover, successful *in vivo* multiplex gene silencing was achieved using the MmuC2c4 and MmudC2c4 effectors and a single crRNA array resulting in two mature crRNA guides, one targeting the *rfp* and the other the *gfp,* on the pTarget-divergent plasmid (**Figure S5.9B,C and D**). Notably, one of the spacers used for targeting the *gfp* gene was able to repress the *gfp* when used as a single crRNA, whereas when used in a crRNA array, it did not show high GFP repression (**Figure S5.9C**). This suggests sequence- and context-dependent loss of RNA-directed nuclease activity, most likely due to hindering RNA secondary structures, similar to what has been observed for Cas12a (*296*). However, future studies are necessary to get into the depth of this phenomenon.

The characterisation of the novel type V-U1, MmuC2c4 protein, described here has revealed a unique mechanism. Akin to several other type V systems (Cas12a, b, c, e, k), specific crRNA-guide dependent binding has been demonstrated to dsDNA. However, instead of DNA cleavage, MmuC2c4 appears to target the nascent RNA during transcription of the targeted DNA. Although no direct evidence of RNA cleavage is currently available, RuvC-dependent ribonuclease activity seems most likely given the observed difference in silencing efficiency between the wild-type MmuC2c4 and its catalytically inactivated variant, MmudC2c4 (**Figure 5.2**, **5.3**, **S5.5**, **S5.7** and **S5.8**). The implications of such a DNA binding RNA cleavage CRISPR–Cas system and the biological relevance of Mmuc2c4 is currently unclear and will be addressed in future research. In the absence of an MmuC2c4 crystal structure the molecular basis of the mismatch tolerance, crRNA binding and dsDNA recognition mechanisms remain elusive. Moreover, the small size, multiplexing capability and potential activity of MmuC2c4 in mammalian cells, which is currently being assessed, might facilitate delivery for applications in therapeutics and biomedical research (*78*, *297*). The PAM–dependent DNA-targeting ability of MmuC2c4-like Cas12 variants can be utilized to recruit transcriptional activators or repressors (*229*), as well as base editing enzymes (*298*). These are particularly interesting applications for understanding the molecular pathology of a range of human diseases as well as to develop novel therapeutic strategies to treat these diseases. The finding that these miniature CRISPR–Cas effectors can accommodate crRNA and conduct targeted DNA binding and nascent mRNA cleavage underscores the rich natural functional diversity of CRISPR–Cas systems. We anticipate that the ongoing combination of biochemical and structural studies will reveal insights into the molecular mechanisms of MmuC2c4 in the near future.

Experimental Procedures

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. The *E. coli* strains were routinely cultured at 37 ℃ 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. All fluorescence loss experiments were carried out in the derivative of *E. coli* BW25113 strain lacking the *lacI, lacZ* genes and the type I-E CRISPR–Cas system.

Plasmid construction

The plasmids constructed and the oligonucleotides (IDT) used for cloning and sequencing are listed in **Table S5.1**.

E. coli codon-harmonized *mmuC2c4* 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 MmuC2c4 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-SCNR screening platform were based on the previously published protocol (*291*). The *mmuC2c4* and *MmudC2c4* genes were inserted into the pBAD33 vector backbone under the control of the constitutive J23108 promoter to generate the pCas-MmuC2c4 and pCas-MmudC2c4 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-MmudC2c4 and pCas-MmuC2c4 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 catalytically inactive MmuC2c4 (MmudC2c4) gene fragment was created by site-directed mutagenesis of the aspartic acid in the RuvC domain to an alanine (D485A).

The pCRISPR plasmids were constructed by restriction-digestion and ligation. By PCR 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 aforementioned 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. The pTarget-GFP plasmid was constructed using BamHI restriction and ligation of a linear P_{lacIq} 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.

MmuC2c4 protein expression and purification

The purification protocol was adapted from established Cas12a purification methods previously (*299*). Briefly, the *mmuC2c4* gene was heterologous expressed in *E. coli* 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 Rosetta (DE3) (EMD Millipore) 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 ℃ 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 ℃ at 30,000xg 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 ℃ 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 analysed by SDS-PAGE and fractions containing the C2c4 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 ℃ for storage.

Pre-crRNA processing

The pre-crRNA processing assay was conducted with ~varying amounts of MmuC2c4 nuclease and ~100 nM pre-crRNA. The assay was conducted in Cas9 Nuclease Reaction Buffer (NEB), in a total volume of 15µL, at 37 ℃ for an hour and quenched with 2 µL proteinase K (NEB) at 30 ℃ for 30 minutes. Subsequently, the samples were analyzed on a 10% urea-PAGE gel stained with SYBR™ Gold Nucleic Acid Stain (Invitrogen).

PAM-SCNR assay

A day prior to sorting, *E. coli* cells harbouring 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.

Fluorescence repression assays

For the silencing assays, *E. coli* cells harbouring either the pCas-MmudC2c4 or the pMmuC2c4 and the corresponding target plasmids were made chemically competent and transformed with the pCRISPR library. For the 5'-NTTN PAM determination assays, cells harbouring either the pCas-MmudC2c4 or the pMmuC2c4 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 harbouring either targeting plasmid pCRISPR-MM-[WT] or non-targeting plasmid pCRISPR-BbsI, and either pCas-MmuC2c4 or pCas-MmudC2c4 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 ℃ 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 ℃. Overnight cultures were then used for fluorescence measurements.

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

$$\frac{\left[\frac{Fl_{x_{targeting}}-Fl_{Blank}}{OD600_{x_{targeting}}-OD600_{Blank}}\right]_{avg}-\left[\frac{Fl_{FS}-Fl_{Blank}}{OD600_{FS}-OD600_{Blank}}\right]_{avg}}{\left[\frac{Fl_{x_{non\text{-}targeting}}-Fl_{Blank}}{OD600_{x_{non\text{-}targeting}}-OD600_{Blank}}\right]_{avg}-\left[\frac{Fl_{FS}-Fl_{Blank}}{OD600_{FS}-OD600_{Blank}}\right]_{avg}}$$

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 NWO/TOP grant 714.015.001. W.X.Y and D.A.S are employees and shareholders of Arbor Biotechnologies, Inc. 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).

Author contributions

P.M., W.W., and J.v.d.O., conceived and designed the study. W.W., P.M., S.C.A.C., K.K., T.S.K., conducted all the experimental work and analysed the data. K.S.M. and E.V.K., provided input on the computational and phylogenetic analysis. W.X.Y and D.A.S performed the NGS sequencing experiments. P.M., W.W., and J.v.d.O. wrote the manuscript with input from all authors.

Author Information

Correspondence should be addressed to john.vanderoost@wur.nl

Competing interests

A patent application has been filed related to this work.

Supplementary Figures and Tables

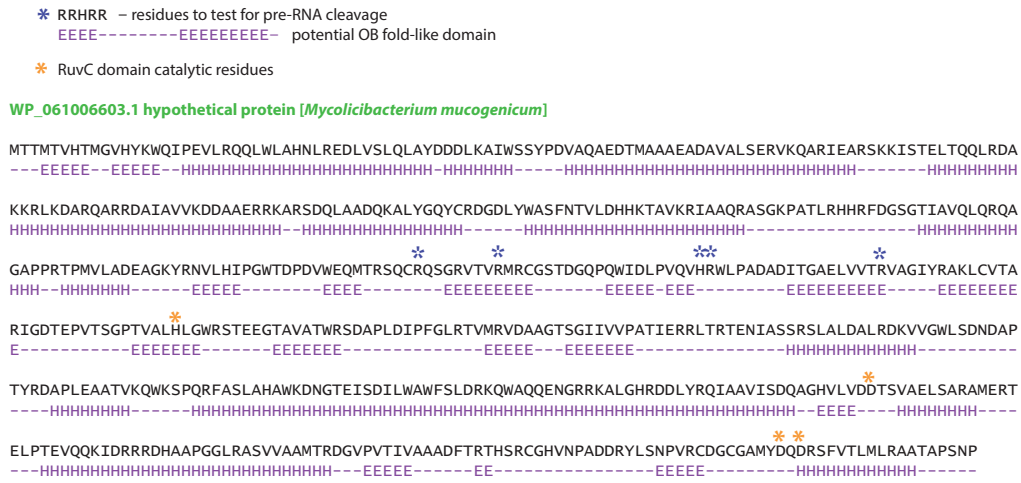


Figure S5.1 | Multiple sequence alignment of type V-U1 orthologues.

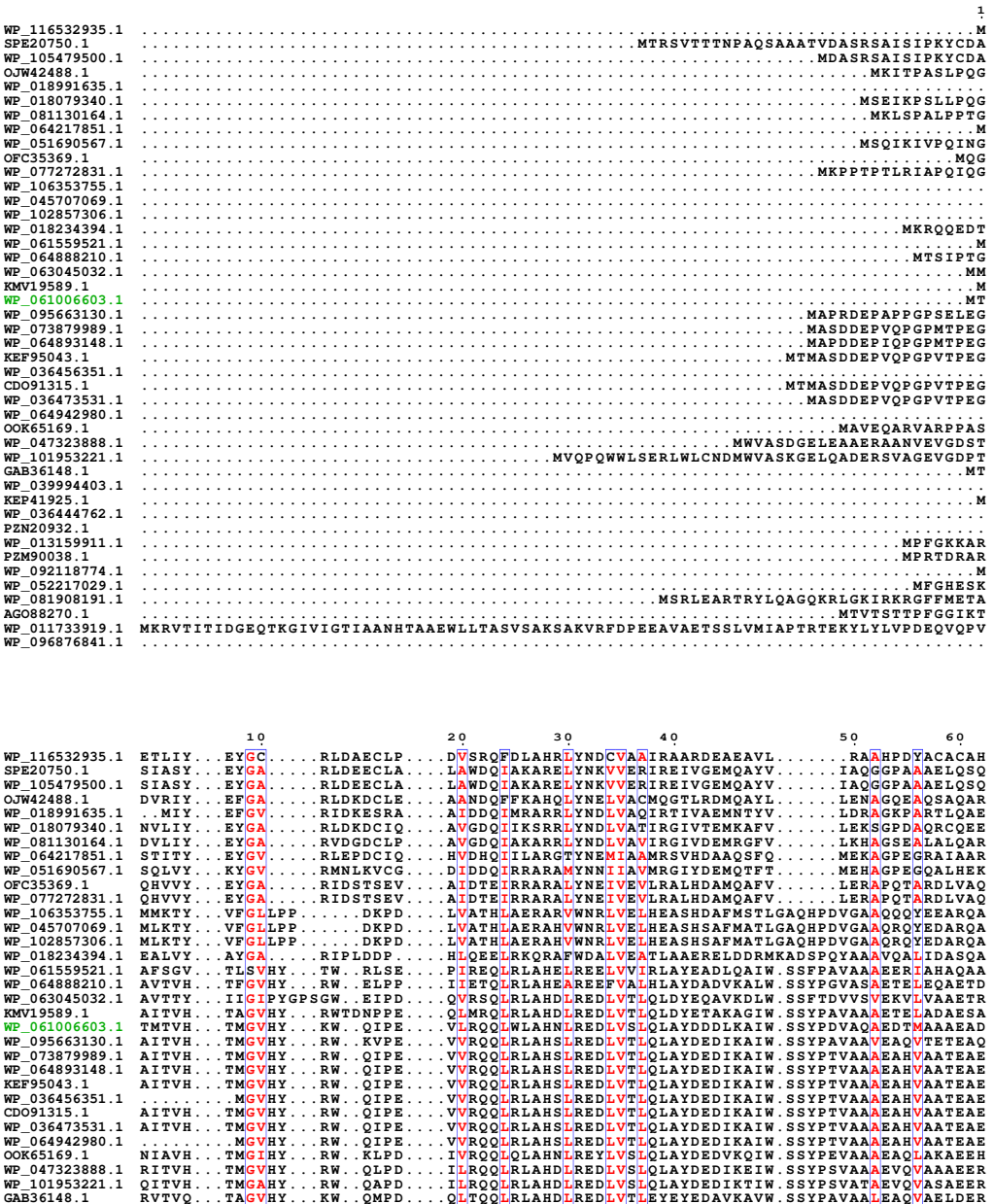


Figure S5.1 | Multiple sequence alignment of type V-U1 orthologues. (continued)

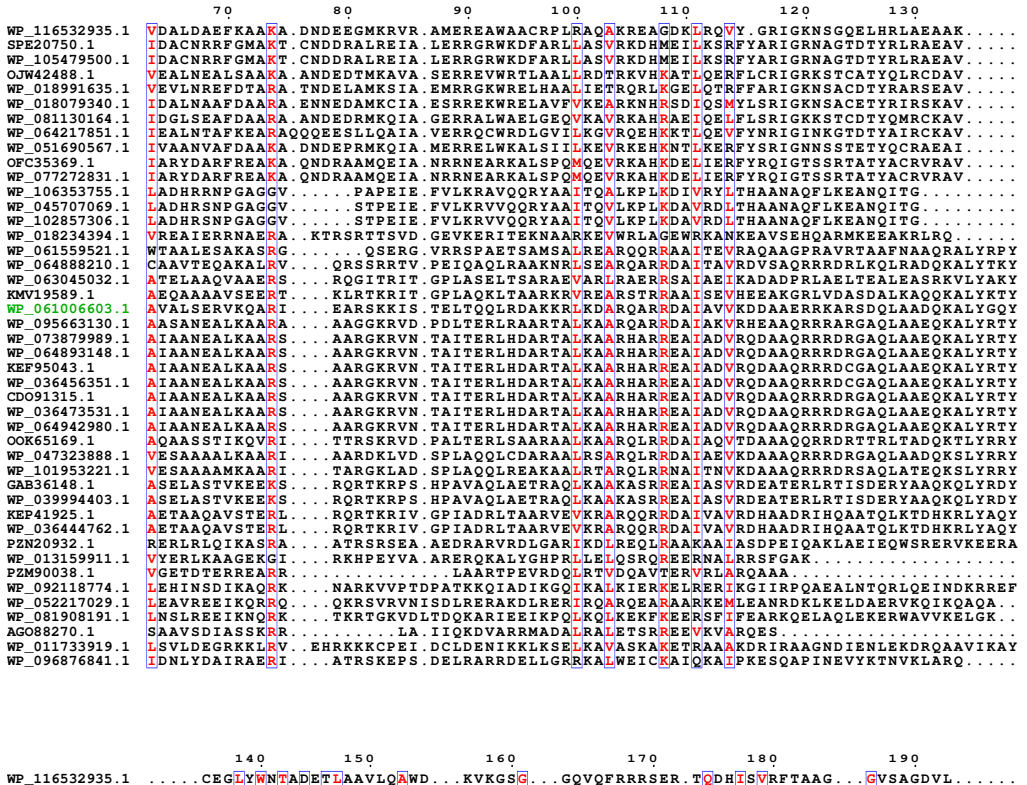


Figure S5.1 | Multiple sequence alignment of type V-U1 orthologues. (continued)

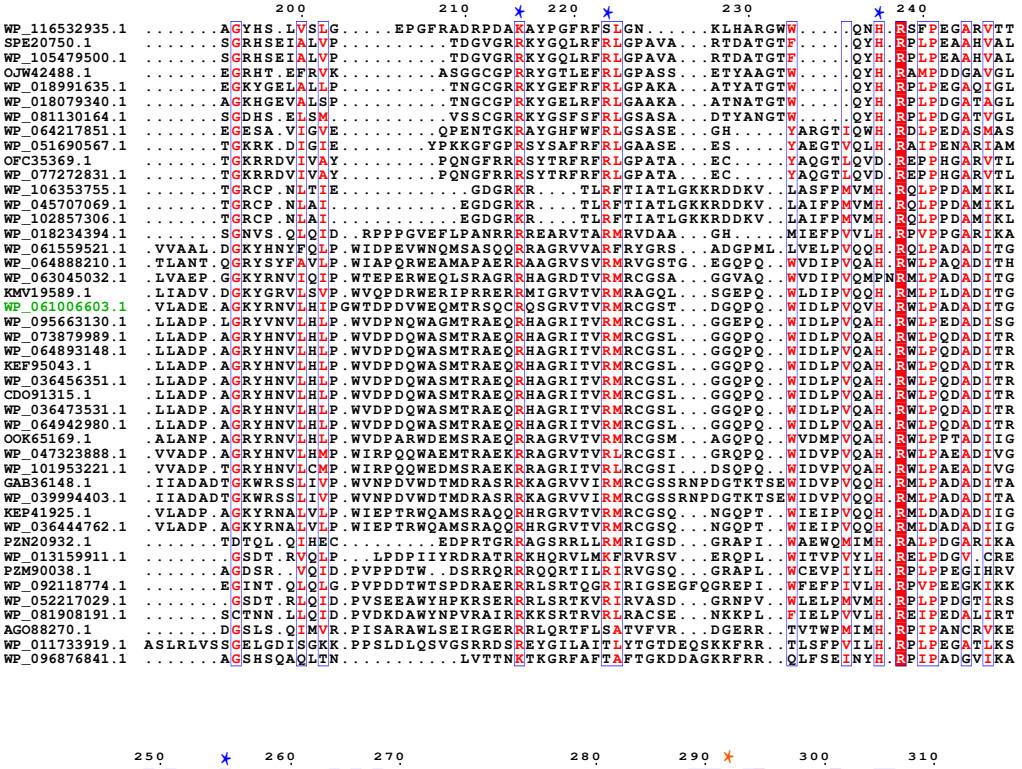


Figure S5.1 | Multiple sequence alignment of type V-U1 orthologues. (continued)

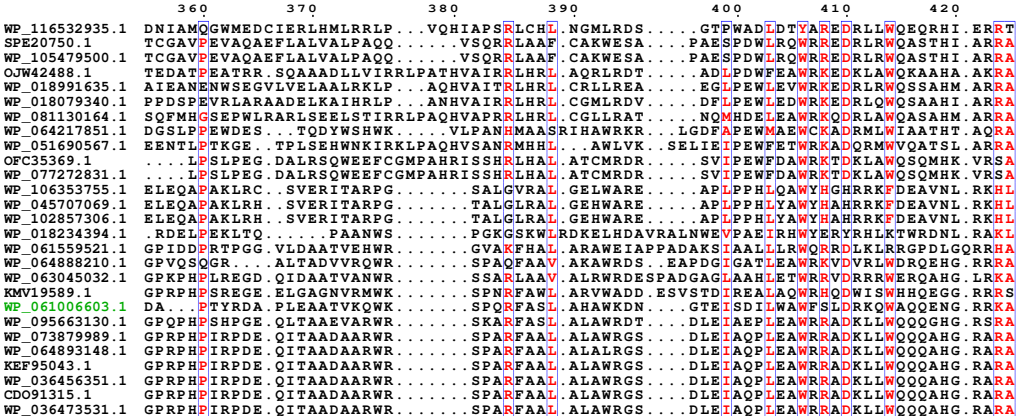


Figure S5.1 | Multiple sequence alignment of type V-U1 orthologues. (continued)

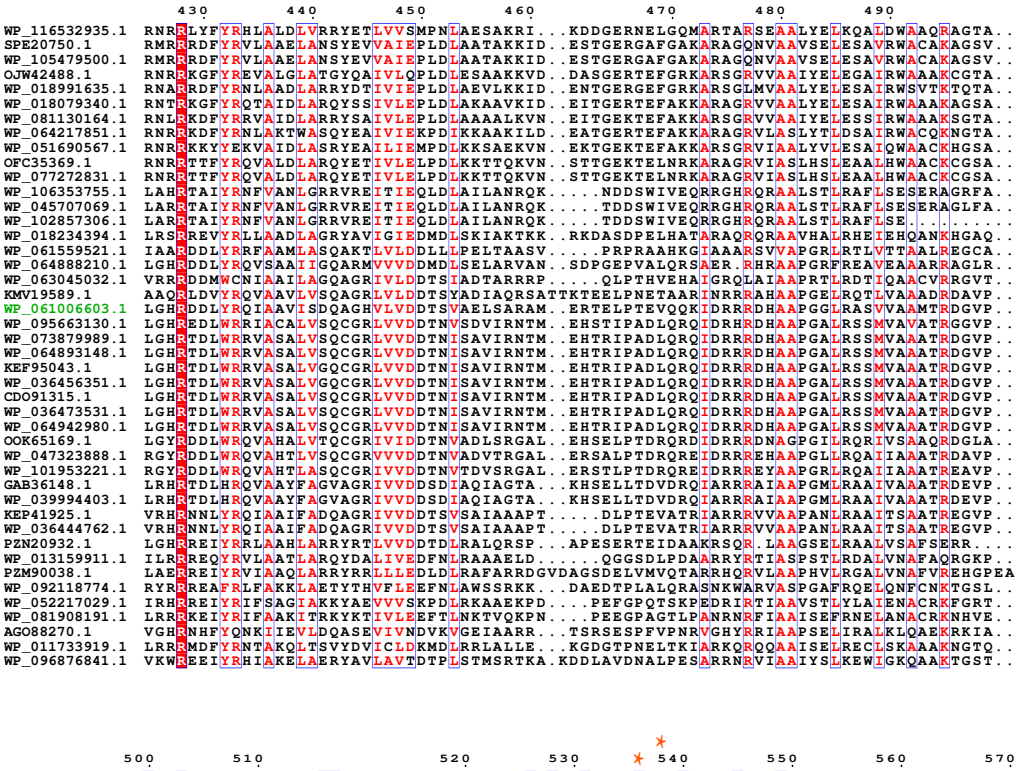


Figure S5.1 | Multiple sequence alignment of type V-U1 orthologues. (continued)

```
                    580         590         600
                     .           .           .
WP_116532935.1  AAEAKKKRLALMQAKRAEVRAKAEKNEGESTRCK............
SPE20750.1      REVNAVARKTKMAAARNAKRQALQEASIAAKETQAGEKAPTCRTGR
WP_105479500.1  REVNAVARKTKMAAARNAKRQALQEASIAAKETQAGEKAPTCRTGR
OJW42488.1      AAAKRKEKREKVAEARRASRVVE.....................
WP_018991635.1  RSAKKADRLARMTDGRRQARGANSSKAP................
WP_018079340.1  AENEQAEKKQKMAEGRRKARTPIGGENTEVSRDSGNGANA.....
WP_081130164.1  HAEKTREKKAKMAEGRRLARTLSAGVSAVGSRNV...........
WP_064217851.1  AAEAKASRLEKMQAARRAKREPALAD....................
WP_051690567.1  KAETKRLKSEKMAEARRLKRQAASQASAGA...............
OFC35369.1      MATLRAQKASGRAQARRASAAAKEKNRAARIAALDAKSEP.....
WP_077272831.1  MATLRAQKASGRAQARRASAAAKEKNRAARIAALDAKSEP.....
WP_106353755.1  .............................................
WP_045707069.1  .............................................
WP_102857306.1  .............................................
WP_018234394.1  TQPNFRERSKTGSRASARA........................... 
WP_061559521.1  .............................................
WP_064888210.1  .............................................
WP_063045032.1  .............................................
KMV19589.1      .............................................
WP_061006603.1  .............................................
WP_095663130.1  .............................................
WP_073879989.1  .............................................
WP_064893148.1  .............................................
KEF95043.1      .............................................
WP_036456351.1  .............................................
CDO91315.1      .............................................
WP_036473531.1  .............................................
WP_064942980.1  .............................................
OOK65169.1      .............................................
WP_047323888.1  .............................................
WP_101953221.1  .............................................
GAB36148.1      .............................................
WP_039994403.1  .............................................
KEP41925.1      .............................................
WP_036444762.1  .............................................
PZN20932.1      RSERLRRGRRKAA................................
WP_013159911.1  .............................................
PZM90038.1      .............................................
WP_092118774.1  .............................................
WP_052217029.1  .............................................
WP_081908191.1  .............................................
AGO88270.1      .............................................
WP_011733919.1  .............................................
WP_096876841.1  AENDFARKIGDNASPLVT..........................
```

Figure S5.1|Multiple sequence alignment of type V-U1 orthologues. (continued) C2c4 protein from *Mycolicibacterium mucogenicum* [WP_061006603.1]and other type V-U1 proteins are aligned.

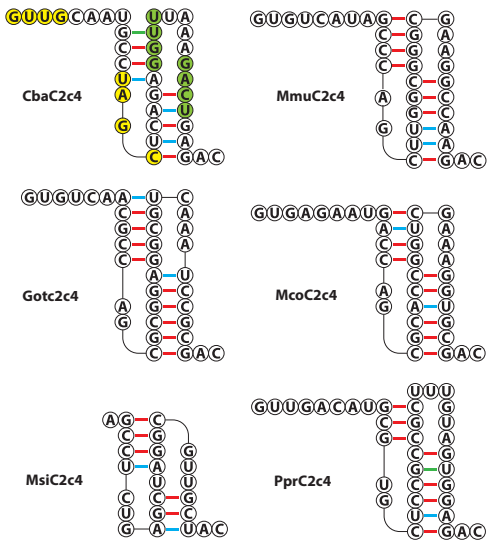


Figure S5.2|Type V-U1 repeats from different bacteria. Vsfold (*300*) prediction of base-pairing of the direct repeat sequence including pseudoknots are shown in coloured bases if they deviate from the basepairing shown here. CbaC2c4: *Clostridiales bacterium* DRI 13 C2c4; MmuC2c4: *Mycobacterium mucogenicum* CCH10 A2 C2c4; GotC2c4: *Gordonia otitidis* NBRC 100426 C2c4; McoC2c4: *Mycobacterium conceptionense* MLE C2c4; MsiC2c4: *Meiothermus silvanus* DSM 9946 C2c4; *Pelobacter propionicus* DSM 2379 C2c4.

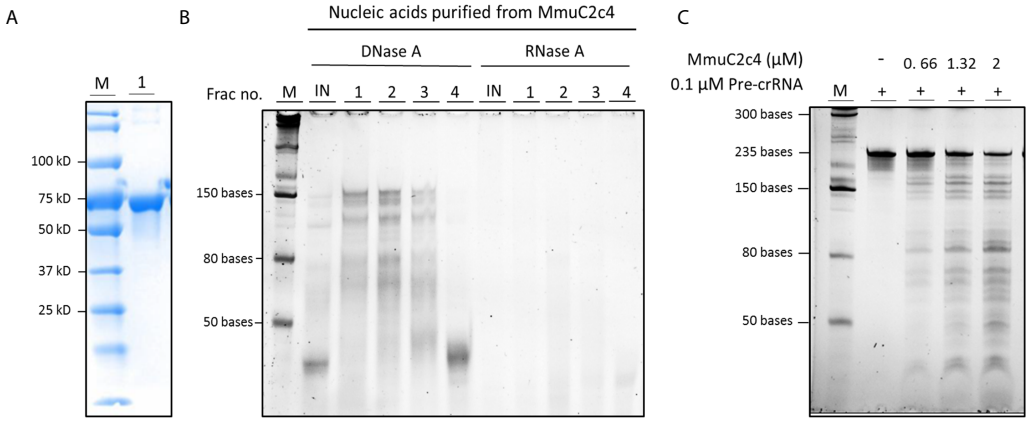


Figure S5.3 | Co-purified nucleic acids and pre-crRNA processing by MmuC2c4. (A) Coomassie blue stained SDS-PAGE gel in which the purified MmuC2c4 protein (66.2 kD) is visualized. **(B)** Co-purified nucleic acids from MmuC2c4 treated with enzymes as indicated. M: low range ssRNA ladder (NEB), IN: input 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).

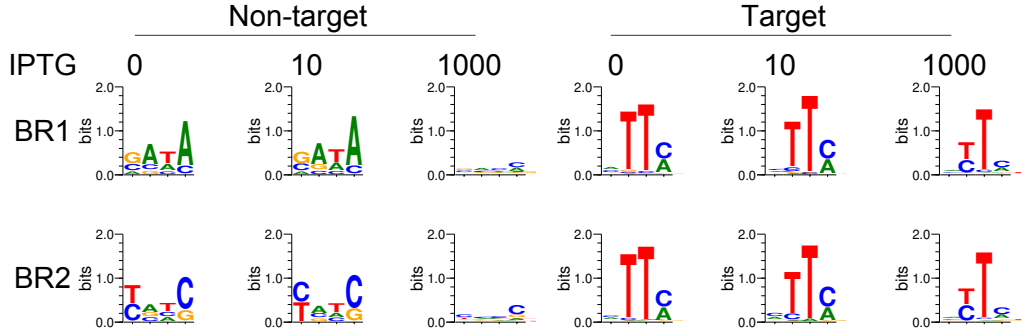


Figure S5.4 | MmuC2c4 PAM determination. Plasmids from the FACS-sorted cells were extracted and sequenced to determine functional PAM sequences. Sequence logo for the MmuC2c4 PAM at different IPTG concentrations (0, 10, 1000 μM) as determined by NGS sequencing. NT: non-targeting, T: targeting, BR1 and BR2 are two independent biological replicates. Letter height at each position is measured by information content.

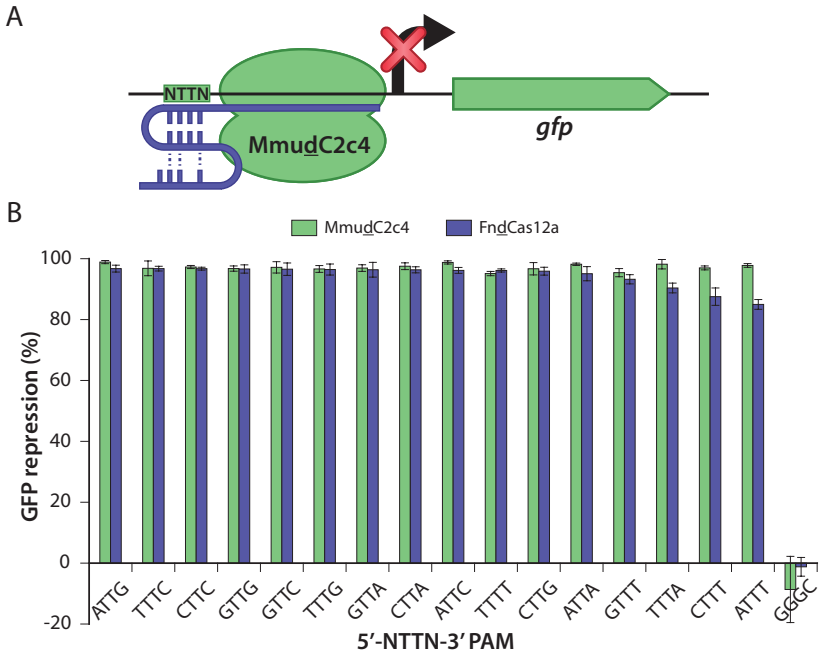


Figure S5.5 | Comparison of transcriptional silencing by MmuC2c4 and FndCas12a. (A) Schematic of the pTarget-GFP encoding the *gfp* gene. The protospacer flanked by 5-NTTN-3' PAM upstream of the promoter is targeted by the MmudC2c4 and FndCas12a proteins using the respective crRNAs. **(B)** GFP repression detected in the cells upon MmudC2c4 and FndCas12a targeting is shown on the Y-axis and the different PAM sequences used are shown in the X-axis (n = 3; error bars represent mean ± SEM).

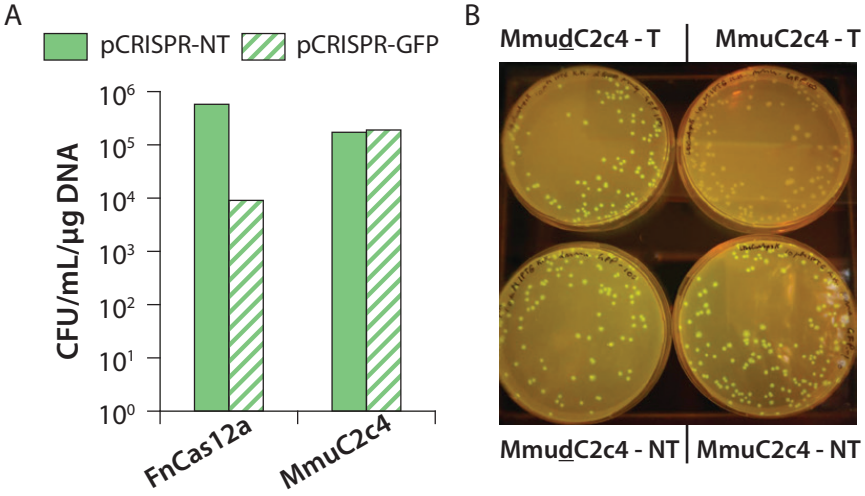


Figure S5.6 | Transformation of MmuC2c4- and FnCas12a- mediated targeting plasmids in *E. coli*. (A) Results of the *in vivo* dsDNA targeting experiment showing OD_{600} measurements from cultures of *E. coli* harbouring the pTarget-PS plasmid transformed with pCas-MmuC2c4 and pCRISPR-PS compared to cells transformed with pCas-FnCas12a and pCRISPR–Cas12a-PS plasmid. **(B)** Qualitative comparison of GFP fluorescence in the cells harbouring pTarget-GFP transformed with pCas-MmudC2c4 with either pCRISPR-GFP (T) or pCRISPR-NT (NT) versus the cells harbouring pTarget-GFP transformed with pCas-MmuC2c4 with either pCRISPR-GFP (T) or pCRISPR-NT (NT).

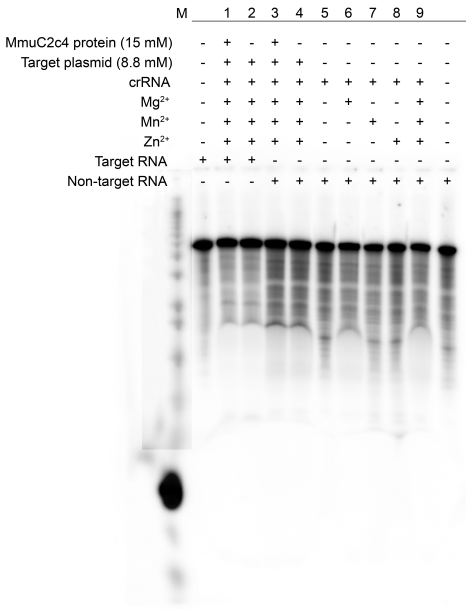


Figure S5.7|*In vitro* dsDNA activated RNA cleavage by MmuC2c4. Urea-PAGE assessing the ability of MmuC2c4 protein incubated with a crRNA and an activator plasmid target DNA to check whether primed MmuC2c4 protein can cleave a [γ-^{32} P] ATP labelled target or a non-target substrate RNA.

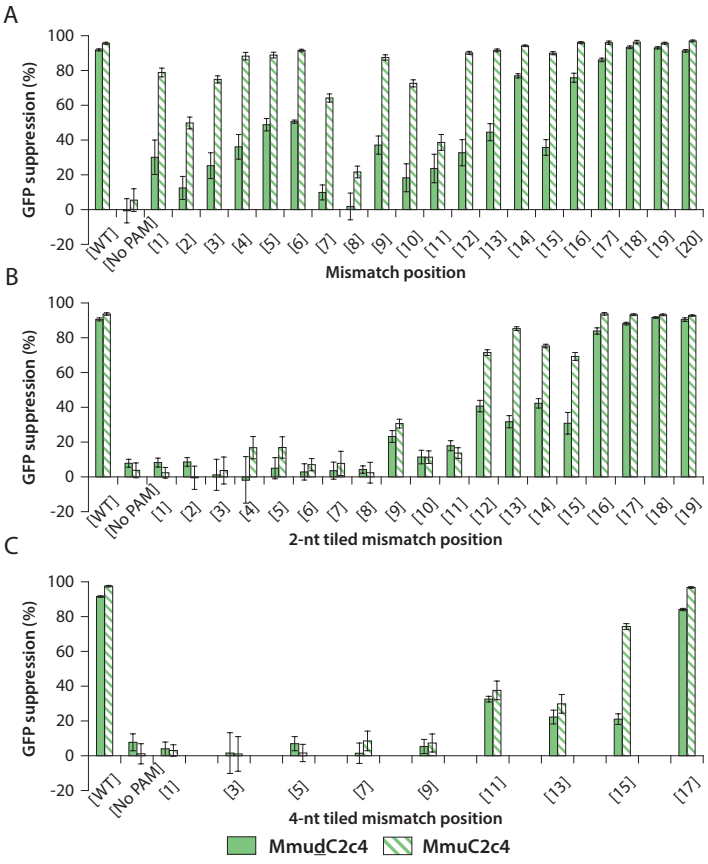


Figure S5.8|Tolerance of MmuC2c4 to mismatched crRNAs. (A) Comparison of the mismatch tolerance of MmuC2c4 with MmudC2c4 for single mismatches across the protospacer sequence. **(B)** Comparison of the mismatch tolerance of MmuC2c4 with MmudC2c4 for 2-nucleotide mismatches tiled across in the target sequence. **(C)** Comparison of the mismatch tolerance of MmuC2c4 with MmudC2c4 for 4-nucleotide mismatches tiled across in the target sequence. GFP repression detected in the cells upon MmudC2c4 or MmuC2c4 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).

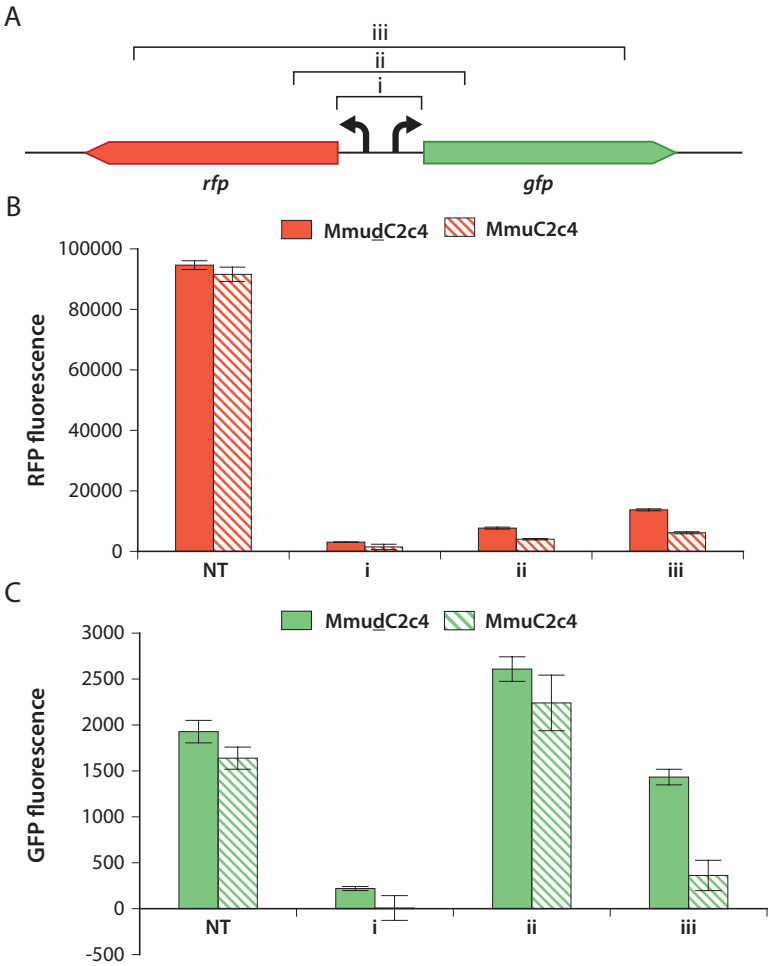


Figure S5.9|MmuC2c4 can be used for multiplex transcriptional silencing. (A) Schematic of the pTarget-divergent including the *rfp* and *gfp* genes under the transcriptional control of two different constitutive promoters, P_{taq} and P_{laclq}. i to iii indicate the crRNA spacer pairs used in the pCRISPR array plasmid to target the *rfp* and *gfp* using the MmuC2c4 and MmudC2c4 proteins. **(B)** RFP fluorescence detected in the cells upon MmudC2c4 or MmuC2c4 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 MmudC2c4 or MmuC2c4 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.

Table S5.1 | List of oligonucleotides used in the study.

Oligo ID	Sequence (5' - 3')	Description
	Construction of the pMmuC2c4 & pMmudC2c4 plasmids	
BG14064	GATGTCCTCCTGAGCTCGC	FW for amplification of the plasmid backbone for construction of the pMmuC2c4 & pMmudC2c4
BG14065	AAGCTTGGCTGTTTTGGCG	FW for amplification of the plasmid backbone for construction of the pMmuC2c4 & pMmudC2c4
BG14070	GCGAGCTCAGGAGGACATCATGACAACAATGACAGTACATACAATGG	FW for amplification of the mmuC2c4 & mmudC2c4gene for construction of the pMmuC2c4 & pMmudC2c4
BG14073	CGCCAAAACAGCCAAGCTTCTAGGGGTTCGAGGGGGC	RV for amplification of the mmuC2c4 gene for construction of the pMmuC2c4 & pMmudC2c4
BG14402	GATAATTCTGCTACCGATGTATCTGCAACTAAAACATGTCCTGCCTGATC	RV for amplification of the mmudC2c4 gene for construction of the pMmudC2c4
BG14403	CAGGCAGGACATGTTTTAGTTGCAGATACATCGGTAGCAGAATTATCGGC	FW for amplification of the mmudC2c4 gene for construction of the pMmudC2c4
	Construction of the pCRISPR plasmids	
BG14103	GGAACTCGAGGTGGTACCG	FW for amplification of the vector for the construction of the pCRISPR
BG14158	GATCGAAGACTAGTGTCATAGCCCAGCTTGGCGGGCGAAGGCCAAGACGTTTTGGCGGATGAGAGAAG	FW for amplification of the vector for the construction of the pCRISPR
BG14086	ACACTGCCATACCGCGAAAGGTTTTGCACTCGACGTCTTGGCCTTCGCCCGCCAAGCTGGGCTATGACACGGTAC	FW oligo for pCRISPR-PS
BG14087	CGTTTCATCGGCCATCGCGGCGGCCTCGTAGCTGCGACGTCGAGTGCAAAAACCTTTCGCGGTATGGCA	RV oligo for pCRISPR-PS
BG15637	CGTGTCATAGCCCAGCTTGGCGGGCGAAGGCCAAGACTGGTCTTCGCATCTTGCCGTTAGAAGACAA	FW oligo for introducing BbsI sites to the pCRISPR, creating pCRISPR-NT plasmid
BG15638	ACACTTGTCTTCTAACGGCAAGATGCGAAGACCAGTCTTGGCCTTCGCCCGCCAAGCTGGGCTATGACACGGTAC	RV oligo for introducing BbsI sites to the pCRISPR, creating pCRISPR-NT plasmid
BG15106	CGTGTCATAGCCCAGCTTGGCGGGCGAAGGCCAAGACGTCGAGTGCAAAACCTTTCG	FW oligo for construction of pCRISPR-A1 & pCRISPR-d
BG15107	ACACCGAAAGGTTTTGCACTCGACGTCTTGGCCTTCGCCCGCCAAGCTGGGCTATGACACGGTAC	RV oligo for construction of pCRISPR-A1 & pCRISPR-d
BG17087	AGACGCTATCATGCCATACCGCGA	FW oligo for construction of pCRISPR-No PAM
BG17088	ACACTCGCGGTATGGCATGATAGC	RV oligo for construction of pCRISPR-No PAM
BG16559	AGACACTCTCTTCCGGGCGCTATC	FW oligo for construction of pCRISPR-A2
BG16560	ACACGATAGCGCCCGGAAGAGAGT	FW oligo for construction of pCRISPR-A2
BG16299	AGACCAAAACCGACATCAAACTGG	FW oligo for construction of pCRISPR-B1 & pCRISPR-a
BG16300	ACACCCAGTTTGATGTCGGTTTTG	RV oligo for construction of pCRISPR-B1 & pCRISPR-a

FW: Forward primer; RV: Reverse primer; MM: Mismatch

Table S5.1 | List of oligonucleotides used in the study. (continued)

Oligo ID	Sequence (5' - 3')	Description
BG16301	AGACGTTGTGGGAGGTGATGTCCA	FW oligo for construction of pCRISPR-B2
BG16302	ACACTGGACATCACCTCCCACAAC	RV oligo for construction of pCRISPR-B2
BG16303	AGACACCTCTAGATTTAAGAAGGA	FW oligo for construction of pCRISPR-C1
BG16304	ACACTCCTTCTTAAATCTAGAGGT	RV oligo for construction of pCRISPR-C1
BG16305	AGACAATCTAGAGGTTAAACAAAA	FW oligo for construction of pCRISPR-C2
BG16306	ACACTTTTGTTTAACCTCTAGATT	RV oligo for construction of pCRISPR-C2
BG16858	AGACCCTGTCCACACAATCTGCCC	FW oligo for construction of pCRISPR-D1 & pCRISPR-f
BG16859	ACACGGGCAGATTGTGTGGACAGG	RV oligo for construction of pCRISPR-D1 & pCRISPR-f
BG16096	AGACGAAAGGGCAGATTGTGTGGA	FW oligo for construction of pCRISPR-D2
BG16097	ACACTCCACACAATCTGCCCTTTC	RV oligo for construction of pCRISPR-D2
BG16886	AGACGTTTTATCTGTTGTTTGTCG	FW oligo for construction of pCRISPR-E1
BG16887	ACACCGACAAACAACAGATAAAAC	RV oligo for construction of pCRISPR-E1
BG16563	AGACTCCTACTCAGGAGAGCGTTC	FW oligo for construction of pCRISPR-E2
BG16564	ACACGAACGCTCTCCTGAGTAGGA	RV oligo for construction of pCRISPR-E2
BG16385	AGACTGGTGTTGCTAGTTTGTTAT	FW oligo for construction of pCRISPR-F1
BG16386	ACACATAACAAACTAGCAACACCA	RV oligo for construction of pCRISPR-F1
BG16561	AGACTGATAACAAACTAGCAACAC	FW oligo for construction of pCRISPR-F2
BG16562	ACACGTGTTGCTAGTTTGTTATCA	RV oligo for construction of pCRISPR-F2
BG16860	AGACGTATGGAAGGTTCCGTTAAC	FW oligo for construction of pCRISPR-b
BG16861	ACACGTTAACGGAACCTTCCATAC	RV oligo for construction of pCRISPR-b
BG16888	AGACAAGTTGACAATTAATCATCG	FW oligo for construction of pCRISPR-c
BG16889	ACACCGATGATTAATTGTCAACTT	RV oligo for construction of pCRISPR-c
BG16088	AGACTTGTTGAATTAGATGGTGAT	FW oligo for construction of pCRISPR-e
BG16089	ACACATCACCATCTAATTCAACAA	RV oligo for construction of pCRISPR-e
BG16890	AGACAAGTTGACAATTAATCATCGGTGTCATAGCCCAGCTTGGCGGGC GAAGGCCAAGACGTCGAGTGCAAAACCTTTCG	FW oligo for construction of pCRISPR-i
BG16891	ACACCGAAAGGTTTTGCACTCGACGTCTTGGCCTTCGCCCGCCAAGCT GGGCTATGACACCGATGATTAATTGTCAACTT	RV oligo for construction of pCRISPR-i
BG16892	AGACGTATGGAAGGTTCCGTTAACGTGTCATAGCCCAGCTTGGCGGGC GAAGGCCAAGACTTGTTGAATTAGATGGTGAT	FW oligo for construction of pCRISPR-ii
BG16893	ACACATCACCATCTAATTCAACAAGTCTTGGCCTTCGCCCGCCAAGCT GGGCTATGACACGTTAACGGAACCTTCCATAC	RV oligo for construction of pCRISPR-ii
BG16894	AGACCAAAACCGACATCAAACTGGGTGTCATAGCCCAGCTTGGCGGGC GAAGGCCAAGACCCTGTCCACACAATCTGCCC	FW oligo for construction of pCRISPR-iii
BG16895	ACACGGGCAGATTGTGTGGACAGGGTCTTGGCCTTCGCCCGCCAAGCT GGGCTATGACACCCAGTTTGATGTCGGTTTTG	RV oligo for construction of pCRISPR-iii

FW: Forward primer; RV: Reverse primer; MM: Mismatch

Table S5.1 | List of oligonucleotides used in the study. (continued)

Oligo ID	Sequence (5' - 3')	Description
BG16520	AGACACCAGTCCATCATTGTAGTG	FW for construction of pCRISPR-MM
BG16521	ACACCACTACAATGATGGACTGGT	RV for construction of pCRISPR-MM
	Construction of pTarget plasmids	
BG15568	ATACTCGGATCCCCTGAATTGACTCTCTTC	FW for construction of pTarget-GFP
BG15569	GGGATCCTCTAGATTTAAG	RV for construction of pTarget-GFP
BG17549	CTTCTTTAGTCGAGTGCAAAACCTTTCGCG	FW for construction of pTarget-GFP with a TTTA PAM
BG16843	ACGAAAGGGCCTCGACGC	RV for construction of pTarget-GFP with different PAMs
BG17550	CTTCGGGCGTCGAGTGCAAAACCTTTCGCG	FW for construction of pTarget-GFP with a GGCC PAM
BG16844	CTTCGTTAGTCGAGTGCAAAACCTTTCGCG	FW for construction of pTarget-GFP with a GTTA PAM
BG16845	CTTCATTAGTCGAGTGCAAAACCTTTCGCG	FW for construction of pTarget-GFP with a ATTA PAM
BG16846	CTTCCTTCGTCGAGTGCAAAACCTTTCGCG	FW for construction of pTarget-GFP with a CTTC PAM
BG16847	CTTCTTTCGTCGAGTGCAAAACCTTTCGCG	FW for construction of pTarget-GFP with a TTTC PAM
BG16848	CTTCGTTCGTCGAGTGCAAAACCTTTCGCG	FW for construction of pTarget-GFP with a GTTC PAM
BG16849	CTTCATTCGTCGAGTGCAAAACCTTTCGCG	FW for construction of pTarget-GFP with a ATTC PAM
BG16850	CTTCCTTTGTCGAGTGCAAAACCTTTCGCG	FW for construction of pTarget-GFP with a CTTT PAM
BG16851	CTTCTTTTGTCGAGTGCAAAACCTTTCGCG	FW for construction of pTarget-GFP with a TTTT PAM
BG16852	CTTCGTTTGTCGAGTGCAAAACCTTTCGCG	FW for construction of pTarget-GFP with a GTTT PAM
BG16853	CTTCATTTGTCGAGTGCAAAACCTTTCGCG	FW for construction of pTarget-GFP with a ATTT PAM
BG16854	CTTCCTTGGTCGAGTGCAAAACCTTTCGCG	FW for construction of pTarget-GFP with a CTTG PAM
BG16855	CTTCTTTGGTCGAGTGCAAAACCTTTCGCG	FW for construction of pTarget-GFP with a TTTG PAM
BG16856	CTTCGTTGGTCGAGTGCAAAACCTTTCGCG	FW for construction of pTarget-GFP with a GTTGPAM
BG16857	CTTCATTGGTCGAGTGCAAAACCTTTCGCG	FW for construction of pTarget-GFP with a ATTGPAM
BG16134	AGAGTCAATTCAGGGGGAGACCACAACGGTTTCCC	FW for construction of the pTarget-operon
BG16135	TTCTTAAATCTAGAGGTTAAACAAAATTATTTCTAGTTTAAGCACCGG	RV for construction of the pTarget-operon
	Construction of pTarget plasmids for mismatch tolerance assays	
BG16430	TATGTTTAACCAGTCCATCATTGTAGTG	FW for construction of pTarget-MM-[WT]
BG16431	TACTCACTACAATGATGGACTGGTTAAA	RV for construction of pTarget-MM-[WT]
BG16432	TATGAAATACCAGTCCATCATTGTAGTG	FW for construction of pTarget-MM-[No PAM]
BG16433	TACTCACTACAATGATGGACTGGTATTT	RV for construction of pTarget-MM-[No PAM]
BG16434	TATGTTTATCCAGTCCATCATTGTAGTG	FW for construction of pTarget-MM-[1]
BG16435	TACTCACTACAATGATGGACTGGATAAA	RV for construction of pTarget-MM-[1]
BG16436	TATGTTTAAGCAGTCCATCATTGTAGTG	FW for construction of pTarget-MM-[2]
BG16437	TACTCACTACAATGATGGACTGCTTAAA	RV for construction of pTarget-MM-[2]
BG16438	TATGTTTAACGAGTCCATCATTGTAGTG	FW for construction of pTarget-MM-[3]
BG16439	TACTCACTACAATGATGGACTCGTTAAA	RV for construction of pTarget-MM-[3]

FW: Forward primer; RV: Reverse primer; MM: Mismatch

Table S5.1 | List of oligonucleotides used in the study. (continued)

Oligo ID	Sequence (5' - 3')	Description
BG16440	TATGTTTAACCTGTCCATCATTGTAGTG	FW for construction of pTarget-MM-[4]
BG16441	TACTCACTACAATGATGGACAGGTTAAA	RV for construction of pTarget-MM-[4]
BG16442	TATGTTTAACCACTCCATCATTGTAGTG	FW for construction of pTarget-MM-[5]
BG16443	TACTCACTACAATGATGGAGTGGTTAAA	RV for construction of pTarget-MM-[5]
BG16444	TATGTTTAACCAGACCATCATTGTAGTG	FW for construction of pTarget-MM-[6]
BG16445	TACTCACTACAATGATGGTCTGGTTAAA	RV for construction of pTarget-MM-[6]
BG16446	TATGTTTAACCAGTGCATCATTGTAGTG	FW for construction of pTarget-MM-[7]
BG16447	TACTCACTACAATGATGCACTGGTTAAA	RV for construction of pTarget-MM-[7]
BG16448	TATGTTTAACCAGTCGATCATTGTAGTG	FW for construction of pTarget-MM-[8]
BG16449	TACTCACTACAATGATCGACTGGTTAAA	RV for construction of pTarget-MM-[8]
BG16450	TATGTTTAACCAGTCCTTCATTGTAGTG	FW for construction of pTarget-MM-[9]
BG16451	TACTCACTACAATGAAGGACTGGTTAAA	RV for construction of pTarget-MM-[9]
BG16452	TATGTTTAACCAGTCCAACATTGTAGTG	FW for construction of pTarget-MM-[10]
BG16453	TACTCACTACAATGTTGGACTGGTTAAA	RV for construction of pTarget-MM-[10]
BG16454	TATGTTTAACCAGTCCATGATTGTAGTG	FW for construction of pTarget-MM-[11]
BG16455	TACTCACTACAATCATGGACTGGTTAAA	RV for construction of pTarget-MM-[11]
BG16456	TATGTTTAACCAGTCCATCTTTGTAGTG	FW for construction of pTarget-MM-[12]
BG16457	TACTCACTACAAAGATGGACTGGTTAAA	RV for construction of pTarget-MM-[12]
BG16458	TATGTTTAACCAGTCCATCAATGTAGTG	FW for construction of pTarget-MM-[13]
BG16459	TACTCACTACATTGATGGACTGGTTAAA	RV for construction of pTarget-MM-[13]
BG16460	TATGTTTAACCAGTCCATCATAGTAGTG	FW for construction of pTarget-MM-[14]
BG16461	TACTCACTACTATGATGGACTGGTTAAA	RV for construction of pTarget-MM-[14]
BG16462	TATGTTTAACCAGTCCATCATTCTAGTG	FW for construction of pTarget-MM-[15]
BG16463	TACTCACTAGAATGATGGACTGGTTAAA	RV for construction of pTarget-MM-[15]
BG16464	TATGTTTAACCAGTCCATCATTGAAGTG	FW for construction of pTarget-MM-[16]
BG16465	TACTCACTTCAATGATGGACTGGTTAAA	RV for construction of pTarget-MM-[16]
BG16466	TATGTTTAACCAGTCCATCATTGTTGTG	FW for construction of pTarget-MM-[17]
BG16467	TACTCACAACAATGATGGACTGGTTAAA	RV for construction of pTarget-MM-[17]
BG16468	TATGTTTAACCAGTCCATCATTGTACTG	FW for construction of pTarget-MM-[18]
BG16469	TACTCAGTACAATGATGGACTGGTTAAA	RV for construction of pTarget-MM-[18]
BG16470	TATGTTTAACCAGTCCATCATTGTAGAG	FW for construction of pTarget-MM-[19]
BG16471	TACTCTCTACAATGATGGACTGGTTAAA	RV for construction of pTarget-MM-[19]
BG16472	TATGTTTAACCAGTCCATCATTGTAGTC	FW for construction of pTarget-MM-[20]
BG16473	TACTGACTACAATGATGGACTGGTTAAA	RV for construction of pTarget-MM-[20]
BG17030	TATGTTTATGCAGTCCATCATTGTAGTG	FW for construction of pTarget-2MM-[1]
BG17031	TACTCACTACAATGATGGACTGCATAAA	RV for construction of pTarget-2MM-[1]

FW: Forward primer; RV: Reverse primer; MM: Mismatch

Table S5.1 | List of oligonucleotides used in the study. (continued)

Oligo ID	Sequence (5' - 3')	Description
BG17032	TATGTTTAAGGAGTCCATCATTGTAGTG	FW for construction of pTarget-2MM-[2]
BG17033	TACTCACTACAATGATGGACTCCTTAAA	RV for construction of pTarget-2MM-[2]
BG17034	TATGTTTAACGTGTCCATCATTGTAGTG	FW for construction of pTarget-2MM-[3]
BG17035	TACTCACTACAATGATGGACACGTTAAA	RV for construction of pTarget-2MM-[3]
BG17036	TATGTTTAACCTCTCCATCATTGTAGTG	FW for construction of pTarget-2MM-[4]
BG17037	TACTCACTACAATGATGGAGAGGTTAAA	RV for construction of pTarget-2MM-[4]
BG17038	TATGTTTAACCACACCATCATTGTAGTG	FW for construction of pTarget-2MM-[5]
BG17039	TACTCACTACAATGATGGTGTGGTTAAA	RV for construction of pTarget-2MM-[5]
BG17040	TATGTTTAACCAGAGCATCATTGTAGTG	FW for construction of pTarget-2MM-[6]
BG17041	TACTCACTACAATGATGCTCTGGTTAAA	RV for construction of pTarget-2MM-[6]
BG17042	TATGTTTAACCAGTGGATCATTGTAGTG	FW for construction of pTarget-2MM-[7]
BG17043	TACTCACTACAATGATCCACTGGTTAAA	RV for construction of pTarget-2MM-[7]
BG17044	TATGTTTAACCAGTCGTTCATTGTAGTG	FW for construction of pTarget-2MM-[8]
BG17045	TACTCACTACAATGAACGACTGGTTAAA	RV for construction of pTarget-2MM-[8]
BG17046	TATGTTTAACCAGTCCTACATTGTAGTG	FW for construction of pTarget-2MM-[9]
BG17047	TACTCACTACAATGTAGGACTGGTTAAA	RV for construction of pTarget-2MM-[9]
BG17048	TATGTTTAACCAGTCCAAGATTGTAGTG	FW for construction of pTarget-2MM-[10]
BG17049	TACTCACTACAATCTTGGACTGGTTAAA	RV for construction of pTarget-2MM-[10]
BG17050	TATGTTTAACCAGTCCATGTTTGTAGTG	FW for construction of pTarget-2MM-[11]
BG17051	TACTCACTACAAACATGGACTGGTTAAA	RV for construction of pTarget-2MM-[11]
BG17052	TATGTTTAACCAGTCCATCTATGTAGTG	FW for construction of pTarget-2MM-[12]
BG17053	TACTCACTACATAGATGGACTGGTTAAA	RV for construction of pTarget-2MM-[12]
BG17054	TATGTTTAACCAGTCCATCAAAGTAGTG	FW for construction of pTarget-2MM-[13]
BG17055	TACTCACTACTTTGATGGACTGGTTAAA	RV for construction of pTarget-2MM-[13]
BG17056	TATGTTTAACCAGTCCATCATACTAGTG	FW for construction of pTarget-2MM-[14]
BG17057	TACTCACTAGTATGATGGACTGGTTAAA	RV for construction of pTarget-2MM-[14]
BG17058	TATGTTTAACCAGTCCATCATTCAAGTG	FW for construction of pTarget-2MM-[15]
BG17059	TACTCACTTGAATGATGGACTGGTTAAA	RV for construction of pTarget-2MM-[15]
BG17060	TATGTTTAACCAGTCCATCATTGATGTG	FW for construction of pTarget-2MM-[16]
BG17061	TACTCACATCAATGATGGACTGGTTAAA	RV for construction of pTarget-2MM-[16]
BG17062	TATGTTTAACCAGTCCATCATTGTTCTG	FW for construction of pTarget-2MM-[17]
BG17063	TACTCAGAACAATGATGGACTGGTTAAA	RV for construction of pTarget-2MM-[17]
BG17064	TATGTTTAACCAGTCCATCATTGTACAG	FW for construction of pTarget-2MM-[18]
BG17065	TACTCTGTACAATGATGGACTGGTTAAA	RV for construction of pTarget-2MM-[18]
BG17066	TATGTTTAACCAGTCCATCATTGTAGAC	FW for construction of pTarget-2MM-[19]
BG17067	TACTGTCTACAATGATGGACTGGTTAAA	RV for construction of pTarget-2MM-[19]

FW: Forward primer; RV: Reverse primer; MM: Mismatch

Table S5.1 | List of oligonucleotides used in the study. (continued)

Oligo ID	Sequence (5' - 3')	Description
BG17068	TATGTTTATGGTGTCCATCATTGTAGTG	RV for construction of pTarget-4MM-[1]
BG17069	TACTCACTACAATGATGGACACCATAAA	FW for construction of pTarget-4MM-[1]
BG17070	TATGTTTAACGTCACCATCATTGTAGTG	RV for construction of pTarget-4MM-[3]
BG17071	TACTCACTACAATGATGGTGACGTTAAA	FW for construction of pTarget-4MM-[3]
BG17072	TATGTTTAACCACAGGATCATTGTAGTG	RV for construction of pTarget-4MM-[5]
BG17073	TACTCACTACAATGATCCTGTGGTTAAA	FW for construction of pTarget-4MM-[5]
BG17074	TATGTTTAACCAGTGGTACATTGTAGTG	RV for construction of pTarget-4MM-[7]
BG17075	TACTCACTACAATGTACCACTGGTTAAA	FW for construction of pTarget-4MM-[7]
BG17076	TATGTTTAACCAGTCCGAGTTTGTAGTG	RV for construction of pTarget-4MM-[9]
BG17077	TACTCACTACAAACTCGGACTGGTTAAA	FW for construction of pTarget-4MM-[9]
BG17078	TATGTTTAACCAGTCCATGGAAGTAGTG	RV for construction of pTarget-4MM-[11]
BG17079	TACTCACTACTTCCATGGACTGGTTAAA	FW for construction of pTarget-4MM-[11]
BG17080	TATGTTTAACCAGTCCATCAAACAAGTG	FW for construction of pTarget-4MM-[13]
BG17081	TACTCACTTGTTTGATGGACTGGTTAAA	RV for construction of pTarget-4MM-[13]
BG17082	TATGTTTAACCAGTCCATCATTCATCTG	FW for construction of pTarget-4MM-[15]
BG17083	TACTCAGATGAATGATGGACTGGTTAAA	RV for construction of pTarget-4MM-[15]
BG17084	TATGTTTAACCAGTCCATCATTGTTCAC	FW for construction of pTarget-4MM-[17]
BG17085	TACTGTGAACAATGATGGACTGGTTAAA	RV for construction of pTarget-4MM-[17]

FW: Forward primer; RV: Reverse primer; MM: Mismatch

Chapter 6

Characterising a thermostable Cas9 for bacterial genome editing and silencing

Ioannis Mougiakos*, Prarthana Mohanraju*, Elleke F. Bosma*, Valentijn Vrouwe, Max Finger Bou, Mihris I. S. Naduthodi, Alex Gussak, Rudolf B. L. Brinkman, Richard van Kranenburg, John van der Oost

* contributed equally

Adapted from:
Characterising a thermostable Cas9 for bacterial genome editing and silencing
2017, Nature Communications, 8(1):1647

Abstract

CRISPR–Cas9 based genome engineering tools have revolutionized fundamental research and biotechnological exploitation of both eukaryotes and prokaryotes. However, the mesophilic nature of the established Cas9 systems does not allow for applications that require enhanced stability, including engineering at elevated temperatures. Here, we identify and characterise ThermoCas9 from the thermophilic bacterium Geobacillus thermodenitrificans *T12. We show that ThermoCas9 is active* in vitro *between 20 ℃ and 70 ℃, a temperature range much broader than that of currently used Cas9 orthologues. Additionally, we demonstrate that ThermoCas9 activity at elevated temperatures is strongly associated with the structure of the employed sgRNA. Subsequently, we develop ThermoCas9-based engineering tools for gene deletion and transcriptional silencing at 55 ℃ in* Bacillus smithii *and for gene deletion at 37 ℃ in* Pseudomonas putida. *Altogether, our findings provide fundamental insights into a thermophilic CRISPR–Cas family member and establish the first Cas9-based bacterial genome editing and silencing tool with a broad temperature range.*

Introduction

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and the CRISPR-associated (Cas) proteins provide adaptive and heritable immunity in prokaryotes against invading genetic elements (*20*, *29*, *40*). CRISPR–Cas systems are subdivided into two classes (1 and 2) and six types (I-VI), depending on their complexity and signature proteins (*301*). Class 2 systems, including type-II CRISPR–Cas9 and type V CRISPR–Cas12a (previously called CRISPR- Cpf1) have recently been exploited as genome engineering tools for both eukaryotes (*302–305*) and prokaryotes (*72*, *306*, *307*). These systems are among the simplest CRISPR–Cas systems known as they introduce targeted double stranded DNA breaks (DSBs) based on a ribonucleoprotein (RNP) complex formed by a single Cas endonuclease and an RNA guide.

The guide of Cas9 consists of a crRNA (CRISPR RNA):tracrRNA (trans-activating-CRISPR-RNA) duplex. For engineering purposes, the crRNA:tracrRNA duplex has been simplified by generating a chimeric, single guide RNA (sgRNA) to guide Cas9 upon co-expression (*51*). In addition, cleavage of the target DNA requires a protospacer adjacent motif (PAM): a 3-8 nucleotide (nt) long sequence located next to the targeted protospacer that is highly variable between different Cas9 proteins (*45*, *46*, *308*). Cas9 endonucleases contain two catalytic domains, denoted as RuvC and HNH. Substituting catalytic residues in one of these domains results in Cas9 nickase variants, and in both domains in an inactive variant (*52*, *61*, *73*). The inactive or dead Cas9 (dCas9) has been instrumental as an efficient gene silencing system and for modulating the expression of essential genes (*72*, *309*, *310*).

To date, *Streptococcus pyogenes* Cas9 (SpCas9) is the best characterised and most widely employed Cas9 for genome engineering. Although a few other type-II systems have been exploited for bacterial genome engineering purposes, none of them is derived from a thermophilic organism. Characterization of such CRISPR–Cas systems would be interesting to gain fundamental insights as well as to develop novel applications.

Although basic genetic tools are available for a number of thermophiles (*311–314*), the efficiency of these tools is still too low to enable full exploration and exploitation of this interesting group of organisms. Based on our finding that SpCas9 is not active *in vivo* at or above 42 ℃, we have previously developed a SpCas9-based engineering tool for facultative thermophiles, combining homologous recombination at elevated temperatures and SpCas9-based counter-selection at moderate temperatures (*315*). However, a Cas9-based editing and silencing tool for obligate thermophiles is not yet available as SpCas9 is not active at elevated temperatures (*315*, *316*), and to date no thermophilic Cas9 has been adapted for such purpose. Here, we describe the characterisation of ThermoCas9: an RNA-guided DNA-endonuclease from the CRISPR–Cas type-IIC system of the thermophilic bacterium *Geobacillus thermodenitrificans* T12 (*317*). We show that ThermoCas9 is active *in vitro* between 20 and 70 ℃ and demonstrate the effect of the sgRNA-structure on its thermostability. We apply ThermoCas9 for *in vivo* genome editing and silencing of the industrially important thermophile *Bacillus smithii* ET 138 (*318*) at 55 ℃, creating the first Cas9-based genome engineering tool readily applicable to thermophiles. In addition, we apply ThermoCas9 for *in vivo* genome editing of the mesophile *Pseudomonas putida* KT2440, for which to date no CRISPR–Cas9-based editing tool had been described confirming the wide temperature range and broad applicability of this novel Cas9 system.

Results

ThermoCas9 identification and purification

We recently isolated and sequenced *Geobacillus thermodenitrificans* strain T12, a Gram positive, moderately thermophilic bacterium with an optimal growth temperature at 65 ℃. Contrary to previous claims that type II CRISPR–Cas systems are not present in thermophilic bacteria (*319*), the sequencing results revealed the existence of a type-IIC CRISPR–Cas system in the genome of *G. thermodenitrificans* T12 (**Figure 6.1A**) (*320*). The Cas9 endonuclease of this system (ThermoCas9) was predicted to be relatively small (1082 amino acids) compared to other Cas9 orthologues, such as SpCas9 (1368 amino acids). The size difference is mostly due to a truncated REC lobe, as has been demonstrated for other small Cas9 orthologues (**Figure S6.1**)(*248*). Furthermore, ThermoCas9 was expected to be active at least around the temperature optimum of *G. thermodenitrificans* T12 (*317*). Using the ThermoCas9 sequence as query, we performed BLAST-P searches in the NCBI/non-redundant protein sequences dataset, and found a number of highly identical Cas9 orthologues (87-99% identity at amino acid level, **Table S6.1**), mostly within the *Geobacillus* genus, supporting the idea that ThermoCas9 is part of a highly conserved defense system of thermophilic bacteria (**Figure 6.1B**). These characteristics suggested it may be a potential candidate for exploitation as a genome editing and silencing tool for thermophilic microorganisms, and for conditions at which enhanced protein robustness is required.

We initially performed *in-silico* prediction of the crRNA and tracrRNA modules of the *G. thermodenitrificans* T12 CRISPR–Cas system using a previously described approach (*72*, *248*). Based on this prediction, a 190 nt sgRNA chimera was designed by linking the predicted full-size crRNA (30 nt long spacer followed by 36 nt long repeat) and tracrRNA (36 nt long anti-repeat followed by an 88 nt sequence with three predicted hairpin structures). ThermoCas9 was heterologously expressed in *E. coli* and purified to homogeneity. Hypothesizing that the loading of the sgRNA to the ThermoCas9 would stabilize the protein, we incubated purified apo-ThermoCas9 and ThermoCas9 loaded with *in vitro* transcribed sgRNA at 60 ℃ and 65 ℃, for 15 and 30 min. SDS-PAGE analysis showed that the purified ThermoCas9 denatures at 65 ℃ but not at 60 ℃, while the denaturation temperature of ThermoCas9-sgRNA complex is above 65 ℃ (**Figure 6.1C**). The demonstrated thermostability of ThermoCas9 implied its potential as a thermo-tolerant CRISPR–Cas9 genome editing tool and encouraged us to analyse some relevant molecular features in more detail.

ThermoCas9 PAM determination

The first step towards the characterisation of ThermoCas9 was the *in-silico* prediction of its PAM preferences for successful cleavage of a DNA target. We used the 10 spacers of the *G. thermodenitrificans* T12 CRISPR locus to search for potential protospacers in viral and plasmid sequences using CRISPRtarget (*323*). As only two hits were obtained with phage genomes (**Figure S6.2A**), it was decided to proceed with an *in vitro* PAM determination approach. The predicted sgRNA sequence was generated by *in vitro* transcription, including a spacer that should allow for ThermoCas9-based targeting of linear dsDNA substrates with a matching protospacer. The protospacer was flanked at its 3'-end by randomized 7-base pair (bp) sequences. After performing ThermoCas9-based cleavage assays at 55 ℃, the cleaved sequences of the library (together with a non-targeted library sample as control) were separated from uncleaved sequences, by gel electrophoresis, and analysed by deep-sequencing in order to identify the ThermoCas9 PAM preference (**Figure 6.2A**). The sequencing results revealed that ThermoCas9 introduces double stranded DNA breaks that, in analogy with the mesophilic Cas9 vari-

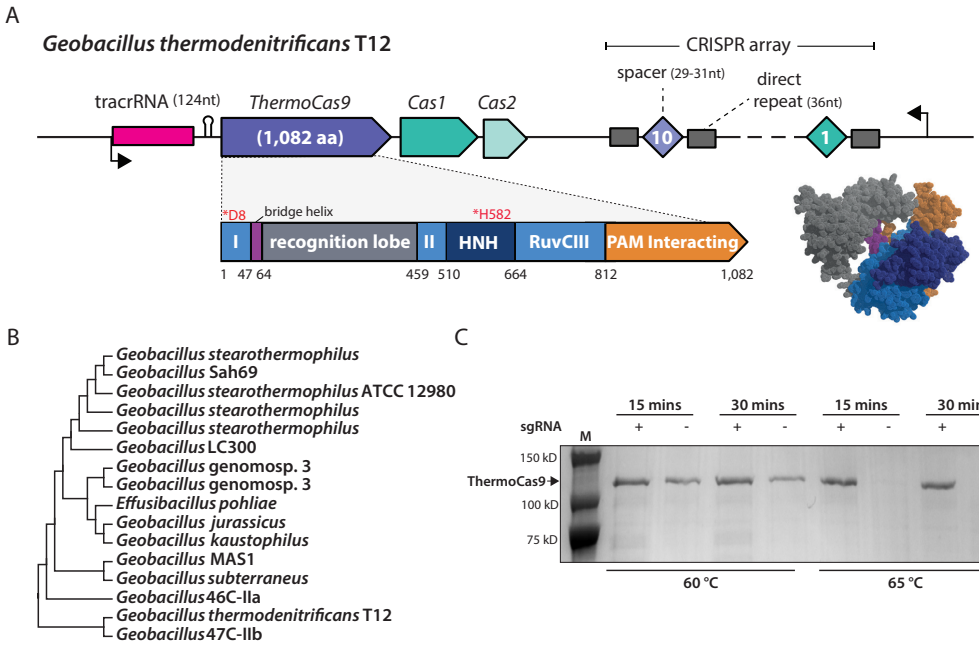


Figure 6.1 | The *Geobacillus thermodenitrificans* T12 type-IIC CRISPR–Cas locus encoding ThermoCas9. (A) Schematic representation of the genomic locus encoding ThermoCas9. The domain architecture of ThermoCas9 based on sequence comparison, with predicted active sites residues highlighted in magenta. A homology model of ThermoCas9 generated using Phyre 2 (*321*) is shown, with different colours for the domains. **(B)** Phylogenetic tree of Cas9 orthologues those are highly identical to ThermoCas9. Evolutionary analysis was conducted in MEGA7 (*322*). **(C)** SDS-PAGE of ThermoCas9 after purification by metal-affinity chromatography and gel filtration. The migration of the obtained single band is consistent with the theoretical molecular weight of 126 kD of the apo-ThermoCas9.

ants, are located mostly between the 3rd and the 4th PAM proximal nucleotides, at the 3' end of the protospacer. Moreover, the cleaved sequences revealed that ThermoCas9 recognizes a 5'-NNNNCNR-3' PAM, with subtle preference for cytosine at the 1st, 3rd, 4th and 6th PAM positions (**Figure 6.2B**). Recent studies have revealed the importance of the 8th PAM position for target recognition of some Type IIC Cas9 orthologues (*308*, *324*) For this purpose and taking into account the results from the *in silico* ThermoCas9 PAM prediction (**Figure S6.2**), we performed additional PAM determination assays. This revealed optimal targeting efficiency in the presence of an adenine at the 8th PAM position (**Figure 6.2C**). Interestingly, despite the limited number of hits, the aforementioned *in silico* PAM prediction (**Figure S6.2B**) also suggested the significance of a cytosine at the 5th and an adenine at the 8th PAM positions.

To further clarify the ambiguity of the PAM at the 6th and 7th PAM positions, we generated a set of 16 different target DNA fragments in which the matching protospacer was flanked by 5'-CCCCCNNA-3'PAMs. Cleavage assays of these fragments (each with a unique combination of the 6th and 7th nucleotide) were performed in which the different components (ThermoCas9, sgRNA guide, dsDNA target) were pre-heated separately at different temperatures (20, 30, 37, 45, 55 and 60 ℃) for 10 min before combining and incubating them for 1 hour at the corresponding assay temperature.

When the assays were performed at temperatures between 37 ℃ and 60 ℃, all the different DNA substrates were cleaved (**Figure 6.2D**, **S6.3**). However, the most digested target frag-

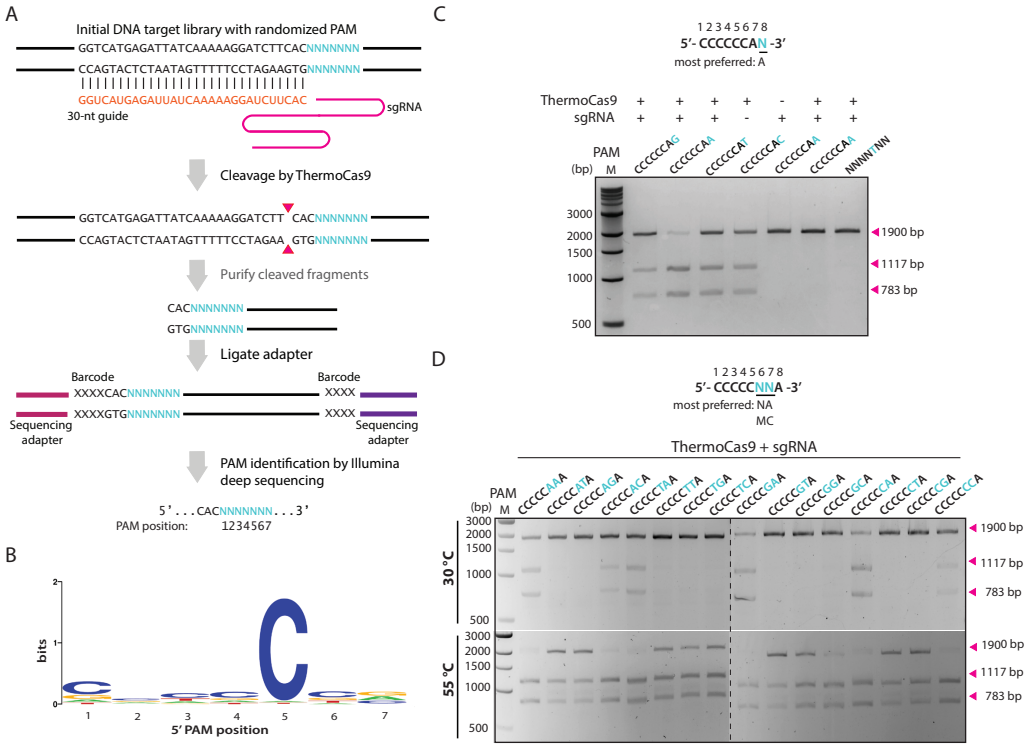


Figure 6.2 | ThermoCas9 PAM analysis. (A) Schematic illustrating the *in vitro* cleavage assay for discovering the position and identity (5′-NNNNNNN-3′) of the protospacer adjacent motif (PAM). Magenta triangles indicate the cleavage position. **(B)** Sequence logo of the consensus 7nt long PAM of ThermoCas9, obtained by comparative analysis of the ThermoCas9-based cleavage of target libraries. Letter height at each position is measured by information content. **(C)** Extension of the PAM identity to the 8th position by *in vitro* cleavage assay. Four linearized plasmid targets, each containing a distinct 5′-CCCCCCAN-3 PAM, were incubated with ThermoCas9 and sgRNA at 55˚C for 1 hour, then analysed by agarose gel electrophoresis. **(D)** *In vitro* cleavage assays for DNA targets with different PAMs at 30˚C and 55˚C. Sixteen linearized plasmid targets, each containing one distinct 5'-CCCCCNNA-3' PAM, were incubated with ThermoCas9 and sgRNA, then analysed for cleavage efficiency by agarose gel electrophoresis. See also **Figure S6.3**.

ments consisted of PAM sequences (5th to 8th PAM positions) 5'-CNAA-3' and 5'-CMCA-3', whereas the least digested targets contained a 5'-CAKA-3' PAM. At 30 ℃, only cleavage of the DNA substrates with the optimal PAM sequences (5th to 8th PAM positions) 5'-CNAA-3' and 5'-CMCA-3' was observed (**Figure 6.2D**). Lastly, at 20 ℃ only the DNA substrates with (5th to 8th PAM positions) 5'-CVAA-3' and 5'-CCCA-3' PAM sequences were targeted (**Figure S6.3**), making these sequences the most preferred PAMs. Our findings demonstrate that at its lower temperature limit, ThermoCas9 only cleaves fragments with a preferred PAM. This characteristic could be exploited during *in vivo* editing processes, for example to avoid off-target effects.

Thermostability and truncations

The predicted tracrRNA consists of the anti-repeat region followed by three hairpin structures (**Figure 6.3A**). Using the tracrRNA along with the crRNA to form a sgRNA chimera resulted in successful guided cleavage of the DNA substrate. It was observed that a 41-nt long deletion of the spacer distal end of the full-length repeat-anti-repeat hairpin (**Figure 6.3A**), most

likely better resembling the dual guide's native state, had little to no effect on the DNA cleavage efficiency. The effect of further truncation of the predicted hairpins (**Figure 6.3A**) on the cleavage efficiency of ThermoCas9 was evaluated by performing a cleavage time-series in which all the components (sgRNA, ThermoCas9, substrate DNA) were pre-heated separately at different temperatures (37-65 ℃) for 1, 2 and 5 min before combining and incubating them for 1 hour at various assay temperatures (37-65 ℃). The number of predicted stem-loops of the tracrRNA scaffold seemed to play a crucial role in DNA cleavage; when all three loops were present, the cleavage efficiency was the highest at all tested temperatures, whereas the efficiency decreased upon removal of the 3' hairpin (**Figure 6.3B**). Moreover, the cleavage efficiency drastically dropped upon removal of both the middle and the 3' hairpins (**Figure S6.4**).

Whereas pre-heating ThermoCas9 at 65 ℃ for 1 or 2 min resulted in detectable cleavage, the cleavage activity was abolished after 5 min incubation. The thermostability assay showed that sgRNA variants without the 3'stem-loop result in decreased stability of the ThermoCas9 protein at 65 ℃, indicating that a full length tracrRNA is required for optimal ThermoCas9-based DNA cleavage at elevated temperatures. Additionally, we also varied the lengths of the spacer sequence (from 25 to 18 nt) and found that spacer lengths of 23, 21, 20 and 19 cleaved the targets with the highest efficiency. The cleavage efficiency drops significantly when a spacer of 18 nt is used (**Figure 6.3C**).

In vivo, the ThermoCas9:sgRNA RNP complex is probably formed within minutes. Together with the above findings, this motivated us to evaluate the activity and thermostability of the RNP. Pre-assembled RNP complex was heated at 60, 65 and 70 ℃ for 5 and 10 min before adding pre-heated DNA and subsequent incubation for 1 hour at 60, 65 and 70 ℃. Strikingly, we observed that the ThermoCas9 RNP was active up to 70 ℃, in spite of its pre-heating for 5 min at 70 ℃ (**Figure 6.3D**). This finding confirmed our assumption that the ThermoCas9 stability strongly correlates with the association of an appropriate sgRNA guide (*295*).

Proteins of thermophilic origin generally retain activity at lower temperatures. Hence, we set out to compare the ThermoCas9 temperature range to that of the *Streptococcus pyogenes* Cas9 (SpCas9). Both Cas9 homologues were subjected to *in vitro* activity assays between 20 and 65 ℃. Both proteins were incubated for 5 min at the corresponding assay temperature prior to the addition of the sgRNA and the target DNA molecules. In agreement with previous analysis (*315*, *316*) the mesophilic SpCas9 was active only between 25 and 44 ℃ (**Figure 6.3E**); above this temperature SpCas9 activity rapidly decreased to undetectable levels. In contrast, ThermoCas9 cleavage activity could be detected between 25 and 65 ℃ (**Figure 6.3E**). This indicates the potential to use ThermoCas9 as a genome editing tool for both thermophilic and mesophilic organisms.

Metal ion dependency

Previously characterised, mesophilic Cas9 endonucleases employ divalent cations to catalyse the generation of DSBs in target DNA (*51*, *325*). To determine the ion dependency of ThermoCas9 cleavage activity, plasmid cleavage assays were performed in the presence of one of the following divalent cations: Mg^{2+}, Ca^{2+}, Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, and Zn^{2+}; an assay with the cation-chelating agent EDTA was included as negative control. As expected, target dsDNA was cleaved in the presence of divalent cations and remained intact in the presence of EDTA (**Figure S6.5A**). The DNA cleavage activity of ThermoCas9 was the highest when Mg^{2+} and Mn^{2+} was added to the reaction consistent with other Cas9 variants (*51*, *52*). Addition of Fe^{2+}, Co^{2+}, Ni^{2+}, or Zn^{2+} ions also mediated cleavage. Ca^{2+} only supported plasmid nicking, suggesting that

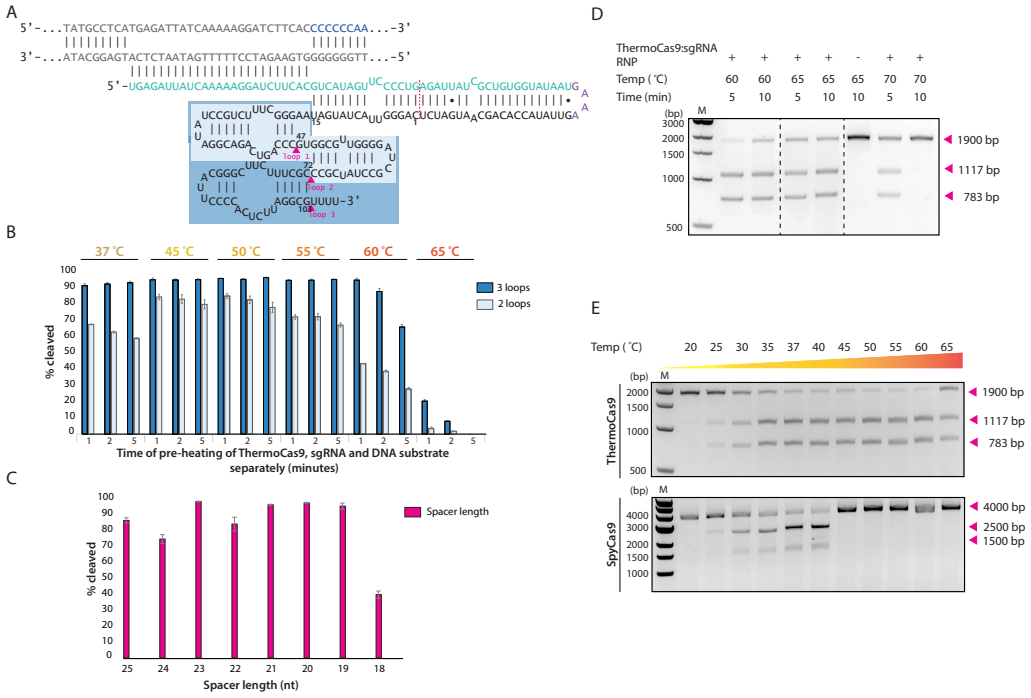


Figure 6.3 | ThermoCas9 temperature range and effect of sgRNA-binding. (A) Schematic representation of the sgRNA and a matching target DNA. The target DNA, the PAM and the crRNA are shown in grey, blue and green, respectively. The site where the crRNA is linked with the tracrRNA is shown in purple. The dark blue and light blue boxes indicate the predicted three and two loops of the tracrRNA, respectively. The 41-nt truncation of the repeat-antirepeat region and the three loops of the sgRNA are indicated by the magenta dotted line and magenta triangles, respectively. **(B)** The importance of the predicted three stem-loops of the tracrRNA scaffold was tested by transcribing truncated variants of the sgRNA and evaluating their ability to guide ThermoCas9 to cleave target DNA at various temperatures. Average values of three replicates are shown, with error bars representing S.D. **(C)** The importance of the length of the spacer was tested by transcribing truncated variants of the initial spacer in the sgRNA and evaluating their ability to guide ThermoCas9 to cleave target DNA at 55 ℃. Average values of three replicates are shown, with error bars representing S.D. **(D)** To identify the maximum temperature, endonuclease activity of ThermoCas9:sgRNA RNP complex was assayed after incubation at 60˚C, 65˚C and 70˚C for 5 or 10 min. The pre-heated DNA substrate was added and the reaction was incubated for 1 hour at the corresponding temperature. **(E)** Comparison of active temperature range of ThermoCas9 and SpCas9 by activity assays conducted after 5 min of incubation at the indicated temperature. The pre-heated DNA substrate was added, and the reaction was incubated for 1 hour at the same temperature.

with this cation only one of the endonuclease domains is functional.

Based on previous reports that certain type-IIC systems were efficient single stranded DNA cutters (*221*, *295*), we tested the activity of ThermoCas9 on ssDNA substrates. However, no cleavage was observed, indicating that ThermoCas9 is a dsDNA nuclease (**Figure S6.5B**).

Spacer-protospacer mismatch tolerance

The targeting specificity and spacer-protospacer mismatch tolerance of a Cas9 endonuclease provide vital information for the development of the Cas9 into a genome engineering tool. To investigate the targeting specificity of ThermoCas9 towards a selected protospacer, we constructed a target plasmid library by introducing either single- or multiple-mismatches to the previously employed protospacer (**Figure 6.4A**). Each member of the plasmid library, and its

PCR-linearized derivative, was separately used as substrate for *in vitro* ThermoCas9 cleavage assays in three independent experiments, both at 37 and 55 ℃.

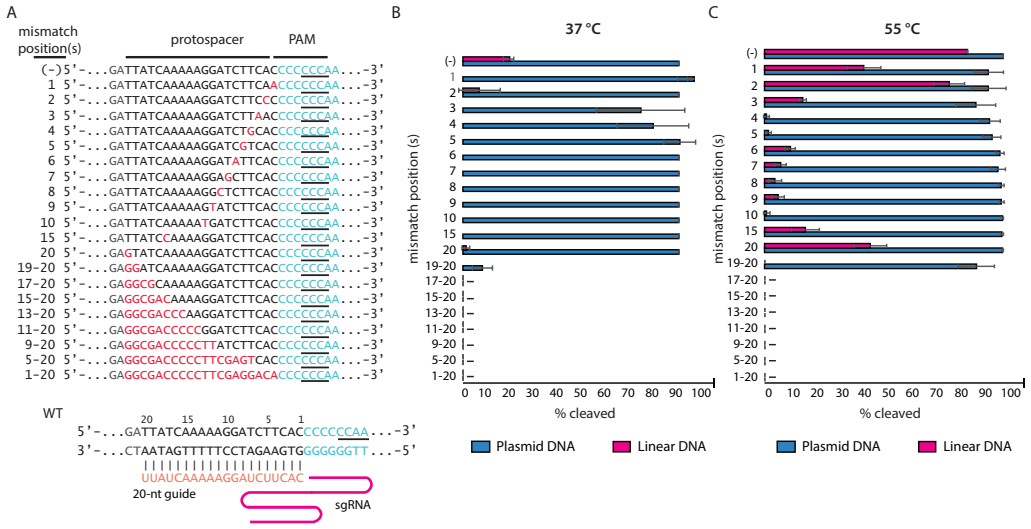


Figure 6.4 | Protospacer targeting Specificity of ThermoCas9. (A) Scheme of the generated mismatch protospacers library, employed for evaluating the ThermoCas9:sgRNA targeting specificity *in vitro*. The mismatch spacer-protospacer positions are shown in red, the PAM in light blue with the 5th to 8th positions underlined. **(B)** Graphical representation of the ThermoCas9:sgRNA cleavage efficiency over linear or plasmid targets with different mismatches at 37 ℃. The percentage of cleavage was calculated based on integrated band intensities after gel electrophoresis. Average values from three biological replicates are shown, with error bars representing S.D.. Average values of three replicates are shown, with error bars representing S.D. **(C)** Graphical representation of the ThermoCas9:sgRNA cleavage efficiency over linear or plasmid targets with different mismatches at 55 ℃. The percentage of cleavage was calculated based on integrated band intensities after gel electrophoresis. Average values from three biological replicates are shown, with error bars representing S.D. Average values of three replicates are shown, with error bars representing S.D.

The ThermoCas9:sgRNA activity on linear dsDNA targets was abolished at 37 ℃ for most of the single-mismatch targets (**Figure 6.4B**). Noteworthy exceptions were the targets with single mismatches at the PAM proximal position 2 and PAM distal position 20, which allowed for weak cleavage (**Figure 6.4B**). At 55 ℃, the cleavage efficiency for single-mismatch linear targets was higher than at 37 ℃, however it was strongly hampered for most of the tested targets, especially for single-mismatches at the PAM proximal positions 4, 5 and 10 (**Figure 6.4C**). On the contrary, single mismatches at positions 1, 2 and 20 were the most tolerated for cleavage (**Figure 6.4C**).

In complete contrast to the linear targets with single-mismatches, all the corresponding plasmid targets were cleaved by the ThermoCas9:sgRNA complex at 37 ℃, regardless the position of the mismatch, with preference for the targets with single-mismatches at the PAM proximal positions 2, 6 to 10, 15 and 20 (**Figure 6.4B**). At 55 ℃, a similar trend was observed, with elevated cleavage rates (**Figure 6.4C**). Remarkably, the ThermoCas9:sgRNA activity was impeded for both linear and plasmid targets with multiple-mismatches as there was no detectable cleavage for most of these targets at the tested temperatures (**Figure 6.4B**, **6.4C**). Notable exception was the target with a double mismatch at positions 19 and 20 which was cleaved at both tested temperatures but more prominently at 55 ℃ (**Figure 6.4B**, **6.4C**).

ThermoCas9-based gene deletion in the thermophile *Bacillus smithii*

We set out to develop a ThermoCas9-based genome editing tool for thermophilic bacteria. This group of bacteria is of great interest both from a fundamental as well as from an applied perspective. For biotechnological applications, their thermophilic nature results in for example less cooling costs, higher reaction rates and less contamination risk compared to the widely used mesophilic industrial work horses such as *E. coli* (*311*, *313*, *326*, *327*). Here, we show a proof of principle study on the use of ThermoCas9 as genome editing tool for thermophiles, employing *Bacillus smithii* ET 138 cultured at 55 ℃. Its wide substrate utilization range, thermophilic and facultative anaerobic nature, combined with its genetic amenability make this an organism with high potential as platform organism for the production of green chemicals in a biorefinery (*311*, *315*, *328*). In order to use a minimum of genetic parts, we followed a single plasmid approach. We constructed a set of pNW33n-based pThermoCas9 plasmids containing the *thermocas9* gene under the control of the native *xylL* promoter (P_{xylL}), a homologous recombination template for repairing Cas9-induced double stranded DNA breaks within a gene of interest, and a sgRNA expressing module under control of the constitutive *pta* promoter (P_{pta}) from *Bacillus coagulans* (**Figure 6.5A**).

The first goal was the deletion of the full length *pyrF* gene from the genome of *B. smithii* ET 138. The pNW33n-derived plasmids pThermoCas9_bsΔ*pyrF*1 and pThermoCas9_bsΔ*pyrF*2 were used for expression of different ThermoCas9 guides with spacers targeting different sites of the *pyrF* gene, while a third plasmid (pThermoCas9_ctrl) contained a random non-targeting spacer in the sgRNA expressing module. Transformation of *B. smithii* ET 138 competent cells at 55 ℃ with the control plasmids pNW33n (no guide) and pThermoCas9_ ctrl resulted in the formation of ~200 colonies each. Out of 10 screened pThermoCas9_ ctrl colonies, none contained the Δ*pyrF* genotype, confirming findings from previous studies that, in the absence of appropriate counter-selection, homologous recombination in *B. smithii* ET 138 is not sufficient to obtain clean mutants (*315*, *328*). In contrast, transformation with the pThermoCas9_bsΔ*pyrF*1 and pThermoCas9_bsΔ*pyrF*2 plasmids resulted in 20 and 0 colonies respectively. Out of the ten pThermoCas9_Δ*pyrF*1 colonies screened, one was a clean Δ*pyrF* mutant whereas the rest had a mixed wild type/Δ*pyrF* genotype (**Figure 6.5B**), proving the applicability of the system, as the designed homology directed repair of the targeted *pyrF* gene was successful. Contrary to eukaryotes, most prokaryotes including *B. smithii* do not possess a functional NHEJ system, and hence DSBs induced by Cas9 have been shown to be lethal in the absence of a functional HR system and/or of an appropriate HR template (*72*, *329*). Hence, Cas9 functions as stringent counter-selection system to kill cells that have not performed the desired HR prior to Cas9 cleavage (*72*, *315*, *330*). The combination of lack of NHEJ and low HR-frequencies found in most prokaryotes provides the basis for the power of Cas9-based editing but also creates the need for tight control of Cas9 activity (*72*, *315*, *330*). As the promoter we use here for *thermocas9*-expression is not sufficiently controllable and HR is inefficient in *B. smithii* (*315*, *328*), the low number (*pyrF*1) or even complete lack (*pyrF*2) of colonies we observed here in the presence of an HR template confirms the high *in vivo* activity of ThermoCas9 at 55 ℃. In the SpCas9-based counter-selection system we previously developed for *B. smithii*, the activity of Cas9 was very tightly controlled by the growth temperature rather than by gene expression. This allowed for extended time for the cells to perform HR prior to Cas9 counter-selection, resulting in a higher *pyrF* deletion efficiency (*315*). We anticipate that the use of a tightly controlled promoter will increase efficiencies of the ThermoCas9-system.

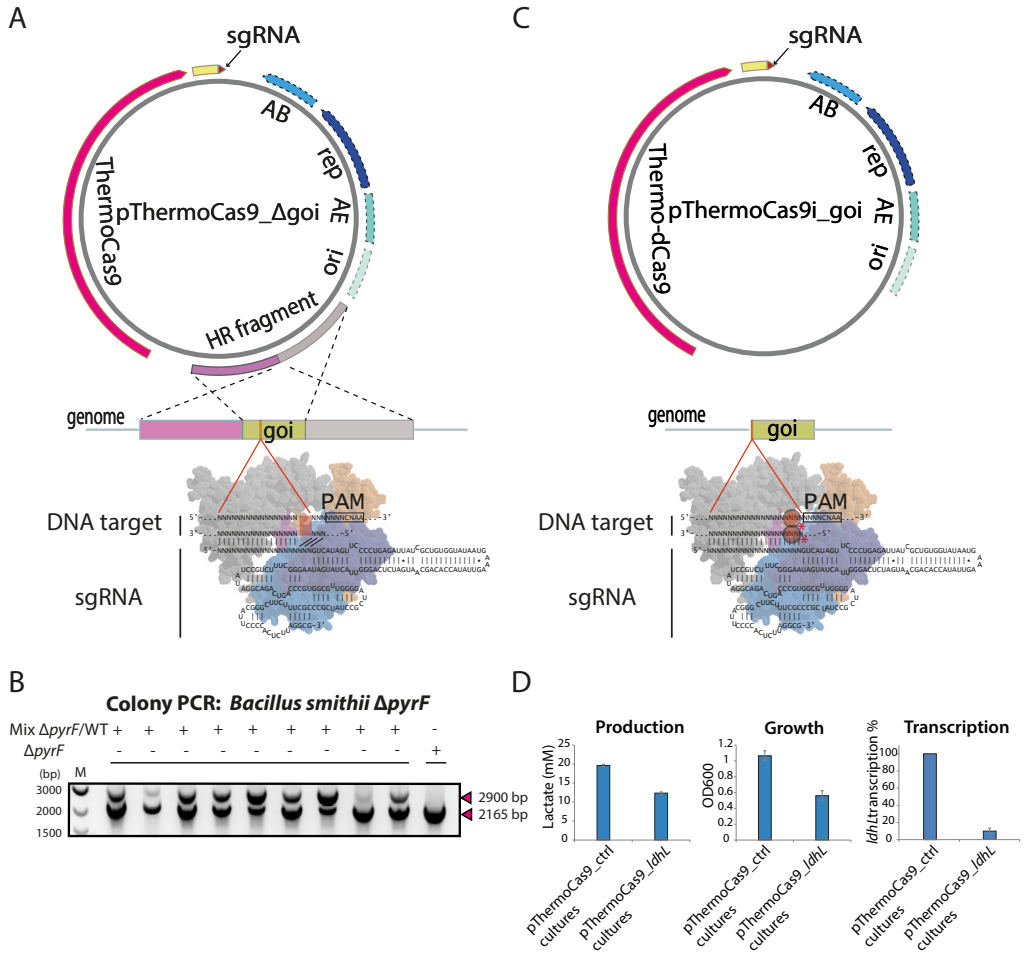


Figure 6.5 | ThermoCas9-based genome engineering in thermophiles. (A) Schematic overview of the basic pThermoCas9_Δgene-of-interest (goi) construct. The *thermocas9* gene was introduced either to the pNW33n (*B. smithii*) or to the pEMG (*P. putida*) vector. Homologous recombination flanks were introduced upstream *thermocas9* and encompassed the 1kb (*B. smithii*) or 0.5kb (*P. putida*) upstream and 1kb or 0.5 kb downstream region of the gene of interest (goi) in the targeted genome. A sgRNA-expressing module was introduced downstream the *thermocas9* gene. As the origin of replication (ori), replication protein (rep), antibiotic resistance marker (AB) and possible accessory elements (AE) are backbone specific, they are represented with dotted outline. **(B)** Agarose gel electrophoresis showing the resulting products from genome-specific PCR on ten colonies from the ThermoCas9-based *pyrF* deletion process from the genome of *B. smithii* ET 138. All ten colonies contained the Δ*pyrF* genotype and one colony was a clean Δ*pyrF* mutant, lacking the wild type product. **(C)** Schematic overview of the basic pThermoCas9i_goi construct. Aiming for the expression of a catalytically inactive ThermoCas9 (ThermodCas9: D8A, H582A mutant), the corresponding mutations were introduced to create the *thermodcas9* gene. The *thermodcas9* gene was introduced to the pNW33n vector. A sgRNA-expressing module was introduced downstream the *thermodcas9*. Graphical representation of the production, growth and RT-qPCR results from the *ldhL* silencing experiment using ThermodCas9. The graphs represent the lactate production, optical density at 600nm and percentage of *ldhL* transcription in the repressed cultures compared to the control cultures. Average values from three biological replicates are shown, with error bars representing S.D.

ThermoCas9-based gene deletion in the mesophile *Pseudomonas putida*

To broaden the applicability of the ThermoCas9-based genome editing tool and to evaluate whether our *in vitro* results could be confirmed *in vivo*, we next evaluated its activity in the

mesophilic Gram-negative bacterium *P. putida* KT2440. This soil bacterium is well-known for its unusual metabolism and biodegradation capacities, especially of aromatic compounds. Recently, interest in this organism has further increased due to its potential as platform host for biotechnology purposes using metabolic engineering (*331*, *332*). However, to date no CRISPR–Cas9-based editing system has been reported for *P. putida* whereas such a system would greatly increase engineering efficiencies and enhance further study and use of this organism (*333*, *334*). Once again, we followed a single plasmid approach and combined homologous recombination and ThermoCas9-based counter-selection. We constructed the pEMG-based pThermoCas9_ppΔ*pyrF* plasmid containing the *thermocas9* gene under the control of the 3-methylbenzoate-inducible Pm-promoter, a homologous recombination template for deletion of the *pyrF* gene and a sgRNA expressing module under the control of the constitutive P3 promoter. After transformation of *P. putida* KT2440 cells and PCR confirmation of plasmid integration, a colony was inoculated in selective liquid medium for overnight culturing at 37 ℃. The overnight culture was used for inoculation of selective medium and ThermoCas9 expression was induced with 3-methylbenzoate. Subsequently, dilutions were plated on non-selective medium, supplemented with 3-methylbenzoate. For comparison, we performed a parallel experiment without inducing ThermoCas9 expression with 3-methylbenzoate. The process resulted in 76 colonies for the induced culture and 52 colonies for the non-induced control culture. For the induced culture, 38 colonies (50%) had a clean deletion genotype and 6 colonies had mixed wild type/deletion genotype. On the contrary, only 1 colony (2%) of the non-induced culture had the deletion genotype and there were no colonies with mixed wild type/deletion genotype retrieved (**Figure S6.6**). These results show that ThermoCas9 can be used as an efficient counter-selection tool in the mesophile *P. putida* KT2440 when grown at 37 ℃.

ThermoCas9-based gene silencing

An efficient thermoactive transcriptional silencing CRISPRi tool is currently not available. Such a system could greatly facilitate metabolic studies of thermophiles. A catalytically dead variant of ThermoCas9 could serve this purpose by steadily binding to DNA elements without introducing dsDNA breaks. To this end, we identified the RuvC and HNH catalytic domains of ThermoCas9 and introduced the corresponding D8A and H582A mutations for creating a dead (d) ThermoCas9. After confirmation of the designed sequence, ThermodCas9 was heterologously produced, purified and used for an *in vitro* cleavage assay with the same DNA target as used in the aforementioned ThermoCas9 assays; no cleavage was observed confirming the catalytic inactivation of the nuclease.

Towards the development of a ThermodCas9-based CRISPRi tool, we aimed for the transcriptional silencing of the highly expressed *ldhL* gene from the genome of *B. smithii* ET 138. We constructed the pNW33n-based vectors pThermoCas9i_*ldhL* and pThermoCas9i_ctrl. Both vectors contained the *thermodCas9* gene under the control of P_{xylL} promoter and a sgRNA expressing module under the control of the constitutive P_{pta} promoter (**Figure 6.5C**). The pThermoCas9i_*ldhL* plasmid contained a spacer for targeting the non-template DNA strand at the 5' end of the 138 *ldhL* gene in *B. smithii* ET 138 (**Figure S6.7**). The position and targeted strand selection were based on previous studies (*73*, *335*), aiming for the efficient down-regulation of the *ldhL* gene. The pThermoCas9i_ctrl plasmid contained a random non-targeting spacer in the sgRNA-expressing module. The constructs were used to transform *B. smithii* ET 138 competent cells at 55 ℃ followed by plating on LB2 agar plates, resulting in equal amounts of colonies. Two out of the approximately 700 colonies per construct were selected for culturing under microaerobic lactate-producing conditions for 24 hours (*328*). The growth of the pTher-

moCas9i_*ldhL* cultures was 50% less than the growth of the pThermoCas9i_ctrl cultures (**Figure 6.5D**). We have previously shown that deletion of the *ldhL* gene leads to severe growth retardation in *B. smithii* ET 138 due to a lack of Ldh-based NAD^+-regenerating capacity under micro-aerobic conditions (*328*). Thus, the observed decrease in growth is likely caused by the transcriptional inhibition of the *ldhL* gene and subsequent redox imbalance due to loss of NAD^+-regenerating capacity. Indeed, HPLC analysis revealed 40% reduction in lactate production of the *ldhL* silenced cultures, and RT-qPCR analysis showed that the transcription levels of the *ldhL* gene were significantly reduced in the pThermoCas9i_*ldhL* cultures compared to the pThermoCas9i_ctrl cultures (**Figure 6.5D**).

Discussion

Most CRISPR–Cas applications are based on RNA-guided DNA interference by Class 2 CRISPR–Cas proteins, such as Cas9 and Cas12a (*72*, *256*, *302*–*304*, *306*, *336*, *337*). Prior to this work, there were only a few examples of Class 1 CRISPR–Cas systems present in thermophilic bacteria and archaea (*31*, *338*), which have been used for genome editing of thermophiles (*319*). As a result, the application of CRISPR–Cas technologies was mainly restricted to temperatures below 42 ℃, due to the mesophilic nature of the employed Cas-endonucleases (*315*, *316*). Hence, this has excluded application of these technologies in obligate thermophiles and in experimental approaches that require elevated temperatures and/or improved protein stability.

In the present study, we have characterised ThermoCas9, a Cas9 orthologue from the thermophilic bacterium *G. thermodenitrificans* T12 that has been isolated from compost (*317*). Data mining revealed additional Cas9 orthologues in the genomes of other thermophiles, which were nearly identical to ThermoCas9, showing that CRISPR–Cas type II systems do exist in thermophiles, at least in some branches of the *Bacillus* and *Geobacillus* genera. We showed that ThermoCas9 is active *in vitro* in a wide temperature range of 20-70 ℃, which is much broader than the 25–44 ℃ range of its mesophilic orthologue SpCas9. The extended activity and stability of ThermoCas9 allows for its application in molecular biology techniques that require DNA manipulation at temperatures of 20-70 ℃, as well as its exploitation in harsh environments that require robust enzymatic activity. Furthermore, we identified several factors that are important for conferring the thermostability of ThermoCas9. Firstly, we showed that the PAM preferences of ThermoCas9 are very strict for activity in the lower part of the temperature range (≤ 30 ℃), whereas more variety in the PAM is allowed for activity at the moderate to optimal temperatures (37-60 ℃). Secondly, we showed that ThermoCas9 activity and thermostability strongly depends on the association with an appropriate sgRNA guide. This stabilization of the multi-domain Cas9 protein is most likely the result of a major conformational change from an open/flexible state to a rather compact state, as described for SpCas9 upon guide binding (*216*). Additionally, we showed that the ThermoCas9 activity on linear DNA targets is very sensitive to spacer-protospacer mismatches. At 55 ℃, cleavage is affected of all linear fragments with single mismatches ranging from position 1 (PAM proximal) to position 20 (PAM distal). Interestingly, positions 4, 5 and 10 are most important, whereas base pairing at position 2 appears less important. The same cleavage pattern is observed with plasmid targets. The elevated cleavage efficiencies of these targets suggest that the negatively supercoiled configuration of the used plasmids improves the target accessibility, as has been described before for the *E. coli* Type I-E CRISPR–Cas system (*42*). Interestingly, despite overall lower activities, similar trends are observed in the case of cleavage assays of both linear fragments and plasmids with single mismatches at 37 ℃. The analysis of multiple mismatches reveals that the ThermoCas9

nuclease is extremely sensitive to four or more mismatches at the PAM distal end: at both temperatures cleavage is completely abolished of all tested targets. Even with two mismatches at positions 19-20, a severe drop of cleavage efficiency is observed. These results indicate a lower *in vitro* spacer-protospacer mismatch tolerance of ThermoCas9 compared to SpCas9 (*51*) and highlight its potential as a genome editing tool for eukaryotic cells with enhanced target specificity. Elucidation of the ThermoCas9 crystal structure is required to gain insight on the molecular basis of the ThermoCas9 cleavage mechanism.

Based on the here described characterisation of the novel ThermoCas9, we successfully developed genome engineering tools for strictly thermophilic prokaryotes. We showed that ThermoCas9 is active *in vivo* at 55 ℃ and 37 ℃, and we adapted the current Cas9-based engineering technologies for the thermophile *B. smithii* ET 138 and the mesophile *P. putida* KT2440. Due to the wide temperature range of ThermoCas9, it is anticipated that the simple, effective and single plasmid based ThermoCas9 approach will be suitable for a wide range of thermophilic and mesophilic microorganisms that can grow at temperatures from 37 ℃ up to 70 ℃. This complements the existing mesophilic technologies, allowing their use for a large group of organisms for which these efficient tools were thus far unavailable

Screening natural resources for novel enzymes with desired traits is unquestionably valuable. Previous studies have suggested that the adaptation of a mesophilic Cas9 orthologue to higher temperatures, with directed evolution and protein engineering, would be the best approach towards the construction of a thermophilic Cas9 protein (*319*). Instead, we identified a clade of Cas9 in some thermophilic bacteria and transformed one of these thermostable ThermoCas9 variants into a powerful genome engineering tool for both thermophilic and mesophilic organisms. With this study, we further stretched the potential of the Cas9-based genome editing technologies and open new possibilities for using Cas9 technologies in novel applications under harsh conditions or requiring activity over a wide temperature range.

Experimental Procedures

Bacterial strains and growth conditions

The moderate thermophile *B. smithii* ET 138 Δ*sigF* Δ*hsdR* (*315*) was used for the gene editing and silencing experiments using ThermoCas9. It was grown in LB2 medium (*328*) at 55 ℃. For plates, 30 g of agar (Difco) per liter of medium was used in all experiments. If needed chloramphenicol was added at the concentration of 7 µg mL^{-1}. For microaerobic lactate-producing conditions, *B. smithii* strains were grown in 25 mL TVMY medium (*328*) in a 50 mL Greiner tube for 24 h at 55 ℃ and 120 rpm after transfer from an overnight culture (*328*). For protein expression, *E. coli* Rosetta (DE3) was grown in LB medium in flasks at 37 ℃ in a shaker incubator at 120 rpm until an $OD_{600\ nm}$ of 0.5 was reached after the temperature was switched to 16 ℃. After 30 min, expression was induced by addition of isopropyl-1-thio-β-D-gal-actopyranoside (IPTG) to a final concentration of 0.5 mM, after which incubation was continued at 16 ℃. For cloning PAM constructs for 6^{th} and 7^{th}, and 8^{th} positions, DH5-alpha competent *E. coli* (NEB) was transformed according to the manual provided by the manufacturer and grown overnight on LB agar plates at 37 ℃. For cloning degenerate 7-nt long PAM library, electro-competent DH10B *E. coli* cells were transformed according to standard procedures (*339*) and grown on LB agar plates at 37 ℃ overnight. *E. coli* DH5α *λpir* (Invitrogen) was used for *P. putida* plasmid construction using the transformation procedure described by Ausubel *et al.* For all *E. coli* strains, if required chloramphenicol was used in concentrations of 25 mg L^{-1} and kanamycin

in 50 mg L^{-1}. *P. putida* KT2440 (DSM 6125) strains were cultured at 37 ℃ in LB medium unless stated otherwise. If required, kanamycin was added in concentrations of 50 mg L^{-1} and 3-methylbenzoate in a concentration of 3 mM.

ThermoCas9 expression and purification

ThermoCas9 was PCR-amplified from the genome of *G. thermodenitrificans* T12, then cloned and heterologously expressed in *E. coli* Rosetta (DE3) and purified using FPLC by a combination of Ni^{2+}-affinity, ion exchange and gel filtration chromatographic steps. The gene sequence was inserted into plasmid pML-1B (obtained from the UC Berkeley MacroLab, Addgene #29653) by ligation-independent cloning using oligonucleotides (**Table S6.2**) to generate a protein expression construct encoding the ThermoCas9 polypeptide sequence (residues 1-1082) fused with an N-terminal tag comprising a hexahistidine sequence and a Tobacco Etch Virus (TEV) protease cleavage site. To express the catalytically inactive ThermoCas9 protein (ThermodCas9), the D8A and H582A point mutations were inserted using PCR and verified by DNA sequencing.

The proteins were expressed in *E. coli* Rosetta 2 (DE3) strain. Cultures were grown to an OD_{600nm} of 0.5-0.6. Expression was induced by the addition of IPTG to a final concentration of 0.5 mM and incubation was continued at 16 ℃ overnight. Cells were harvested by centrifugation and the cell pellet was resuspended in 20 mL of Lysis Buffer (50 mM sodium phosphate pH 8, 500 mM NaCl, 1 mM DTT, 10 mM imidazole) supplemented with protease inhibitors (Roche cOmplete, EDTA-free) and lysozyme. Once homogenized, cells were lysed by sonication (Sonoplus, Bandelin) using a using an ultrasonic MS72 microtip probe (Bandelin), for 5-8 minutes consisting of 2s pulse and 2.5s pause at 30% amplitude and then centrifuged at 16,000*g* for 1 hour at 4 ℃ to remove insoluble material. The clarified lysate was filtered through 0.22-micron filters (Mdi membrane technologies) and applied to a nickel column (Histrap HP, GE Lifesciences), washed and then eluted with 250 mM imidazole. Fractions containing ThermoCas9 were pooled and dialyzed overnight into the dialysis buffer (250 mM KCl, 20 mM HEPES/KOH, and 1 mM DTT, pH 8). After dialysis, sample was diluted 1:1 in 10 mM HEPES/KOH pH 8 and loaded on a heparin FF column pre-equilibrated in IEX-A buffer (150 mM KCl, 20 mM HEPES/KOH pH 8). Column was washed with IEX-A and then eluted with a gradient of IEX-C (2M KCl, 20 mM HEPES/KOH pH 8). The sample was concentrated to 700 µL prior to loading on a gel filtration column (HiLoad 16/600 Superdex 200) via FPLC (AKTA Pure). Fractions from gel filtration were analysed by SDS-PAGE; fractions containing ThermoCas9 were pooled and concentrated to 200 µL (50 mM sodium phosphate pH 8, 2 mM DTT, 5% glycerol, 500 mM NaCl) and either used directly for biochemical assays or frozen at -80 ℃ for storage.

In vitro synthesis of sgRNA

The sgRNA module was designed by fusing the predicted crRNA and tracrRNA sequences with a 5'-GAAA-3' linker. The sgRNA-expressing DNA sequence was put under the transcriptional control of the T7 promoter. It was synthesized (Baseclear, Leiden, The Netherlands) and provided in the pUC57 backbone. All sgRNAs used in the biochemical reactions were synthesized using the HiScribe™ T7 High Yield RNA Synthesis Kit (NEB). PCR fragments coding for sgRNAs, with the T7 sequence on the 5' end, were utilized as templates for *in vitro* transcription reaction. T7 transcription was performed for 4 hours. The sgRNAs were run and excised from urea-PAGE gels and purified using ethanol precipitation.

In vitro cleavage assay

In vitro cleavage assays were performed with purified recombinant ThermoCas9. ThermoCas9 protein, the *in vitro* transcribed sgRNA and the DNA substrates (generated using PCR amplification using primers described in **Table S6.2**) were incubated separately (unless otherwise indicated) at the stated temperature for 10 min, followed by combining the components together and incubating them at the various assay temperatures in a cleavage buffer (100 mM sodium phosphate buffer (pH=7), 500 mM NaCl, 25 mM $MgCl_2$, 25 (V/V%) glycerol, 5 mM dithiothreitol (DTT)) for 1 hour. Each cleavage reaction contained 160 nM of ThermoCas9 protein, 4 nM of substrate DNA, and 150 nM of synthetized sgRNA. Reactions were stopped by adding 6x loading dye (NEB) and run on 1.5% agarose gels. Gels were stained with SYBR safe DNA stain (Life Technologies) and imaged with a Gel DocTM EZ gel imaging system (Bio-rad).

Library construction for *in vitro* PAM screen

For the construction of the PAM library, a 122-bp long DNA fragment, containing the protospacer and a 7-bp long degenerate sequence at its 3'-end, was constructed by primer annealing and Klenow fragment (exo-) (NEB) based extension. The PAM-library fragment and the pNW33n vector were digested by BspHI and BamHI (NEB) and then ligated (T4 ligase, NEB). The ligation mixture was transformed into electro-competent *E. coli* DH10B cells and plasmids were isolated from liquid cultures. For the 7nt-long PAM determination process, the plasmid library was linearized by SapI (NEB) and used as the target. For the rest of the assays the DNA substrates were linearized by PCR amplification.

PAM screening assay

The PAM screening of thermoCas9 was performed using *in vitro* cleavage assays, which consisted of (per reaction): 160 nM of ThermoCas9, 150 nM *in vitro* transcribed sgRNA, 4 nM of DNA target, 4 µL of cleavage buffer (100 mM sodium phosphate buffer pH 7.5, 500 mM NaCl, 5 mM DTT, 25% glycerol) and MQ water up to 20 µL final reaction volume. The PAM containing cleavage fragments from the 55 ℃ reactions were gel purified, ligated with Illumina sequencing adaptors and sent for Illumina HiSeq 2500 sequencing (Baseclear). Equimolar amount of non-ThermoCas9 treated PAM library was subjected to the same process and sent for Illumina HiSeq 2500 sequencing as a reference. HiSeq reads with perfect sequence match to the reference sequence were selected for further analysis. From the selected reads, those present more than 1000 times in the ThermoCas9 treated library and at least 10 times more in the ThermoCas9 treated library compared to the control library were employed for WebLogo analysis (*340*).

Library construction for *in vitro* mismatch tolerance screen

For the construction of the spacer-protospacer mismatch target library, twenty pairs of 40-nt long complementary ssDNA fragments, containing the mismatch-protospacers, were annealed. The annealing products were designed to have overhangs compatible for their directional ligation (T4 ligase, NEB) into the pNW33n backbone, upon BspHI and BamHI (NEB) digestion of the vector. The ligation mixtures were transformed into chemically competent *E. coli* DH5α cells (NEB), plasmids were isolated from liquid cultures and verified by sequencing. Both plasmids and PCR linearized DNA substrates were employed for the mismatch-tolerance assays.

B. smithii and *P. putida* editing and silencing constructs

All the primers and plasmids used for plasmid construction were designed with appropriate overhangs for performing NEBuilder HiFi DNA assembly (NEB), and they are listed in **Table S6.1** and **Table S6.2**, respectively. The fragments for assembling the plasmids were obtained through PCR with Q5 Polymerase (NEB) or Phusion Flash High-Fidelity PCR Master Mix (ThermoFisher Scientific), the PCR products were subjected to 1% agarose gel electrophoresis and they were purified using Zymogen gel DNA recovery kit (Zymo Research). The assembled plasmids were transformed to chemically competent *E. coli* DH5α cells (NEB), or to *E. coli* DH5α λpir (Invitrogen) in the case of *P. putida* constructs, the latter to facilitate direct vector integration. Single colonies were inoculated in LB medium, plasmid material was isolated using the GeneJet plasmid miniprep kit (ThermoFisher Scientific) and sequence verified (GATC-biotech) and 1μg of each construct was transformed to *B. smithii* ET 138 electrocompetent cells (*328*). The MasterPure™ Gram Positive DNA Purification Kit (Epicentre) was used for genomic DNA isolation from *B. smithii* and *P. putida* liquid cultures. For the construction of the pThermoCas9_ctrl, pThermoCas9_bsΔ*pyrF*1 and pThermoCas9_bsΔ*pyrF*2 vectors, the pNW33n backbone together with the Δ*pyrF* homologous recombination flanks were PCR amplified from the pWUR_Cas9sp1_hr vector (*315*) (BG8191and BG8192). The native P_{xylA} promoter was PCR amplified from the genome of *B. smithii* ET 138 (BG8194 and BG8195). The *thermocas9* gene was PCR amplified from the genome of *G. thermodenitrificans* T12 (BG8196 and BG8197). The P_{pta} promoter was PCR amplified from the pWUR_Cas9sp1_hr vector (*315*) (BG8198 and BG8261_2/BG8263_nc2/ BG8317_3). The spacers followed by the sgRNA scaffold were PCR amplified from the pUC57_T7t12sgRNA vector (BG8266_2/BG8268_nc2/8320_3 and BG8210).

A four-fragment assembly was designed and executed for the construction of the pThermoCas9i_ldhL vectors. Initially, targeted point mutations were introduced to the codons of the *thermocas9* catalytic residues (mutations D8A and H582A), through a two-step PCR approach using pThermoCas9_ctrl as template. During the first PCR step (BG9075, BG9076), the desired mutations were introduced at the ends of the produced PCR fragment and during the second step (BG9091, BG9092) the produced fragment was employed as PCR template for the introduction of appropriate assembly-overhangs. The part of the *thermocas9* downstream the second mutation along with the *ldhL* silencing spacer was PCR amplified using pThermoCas9_ctrl as template (BG9077 and BG9267). The sgRNA scaffold together with the pNW33n backbone was PCR amplified using pThermoCas9_ctrl as template (BG9263 and BG9088). The promoter together with the part of the *thermocas9* upstream the first mutation was PCR amplified using pThermoCas9_ctrl as template (BG9089, BG9090)

A two-fragment assembly was designed and executed for the construction of pThermoCas9i_ctrl vector. The spacer sequence in the pThermoCas9i_ldhL vector was replaced with a random sequence containing BaeI restriction sites at both ends. The sgRNA scaffold together with the pNW33n backbone was PCR amplified using pThermoCas9_ctrl as template (BG9548, BG9601). The other half of the construct consisting of *thermodcas9* and promoter was amplified using pThermoCas9i_*ldhL* as template (BG9600, BG9549).

A five-fragment assembly was designed and executed for the construction of the *P. putida* KT2440 vector pThermoCas9_ppΔ*pyrF*. The replicon from the suicide vector pEMG was PCR amplified (BG2365, BG2366). The flanking regions of *pyrF* were amplified from KT2440 genomic DNA (BG2367, BG2368 for the 576-bp upstream flank, and BG2369, BG2370 for the

540-bp downstream flank). The flanks were fused in an overlap extension PCR using primers BG2367 and BG2370 making use of the overlaps of primers BG2368 and BG2369. The sgRNA was amplified from the pThermoCas9_ctrl plasmid (BG2371, BG2372). The constitutive P3 promoter was amplified from pSW_I-SceI (BG2373, BG2374). This promoter fragment was fused to the sgRNA fragment in an overlap extension PCR using primers BG2372 and BG2373 making use of the overlaps of primers BG2371 and BG2374. ThermoCas9 was amplified from the pThermoCas9_ctrl plasmid (BG2375, BG2376). The inducible Pm-XylS system, to be used for 3-methylbenzoate induction of ThermoCas9 was amplified from pSW_I-SceI (BG2377, BG2378).

Editing protocol for *P. putida*

Transformation of the plasmid to *P. putida* was performed according to Choi *et al.* (*341*). After transformation and selection of integrants, overnight cultures were inoculated. 10 µL of overnight culture was used for inoculation of 3 mL fresh selective medium and after 2 hours of growth at 37 ℃ ThermoCas9 was induced with 3-methylbenzoate. After an additional 6h, dilutions of the culture were plated on non-selective medium supplemented with 3-methylbenzoate. For the control culture the addition of 3-methylbenzoate was omitted in all the steps. Confirmation of plasmid integration in the *P. putida* chromosome was done by colony PCR with primers BG2381 and BG2135. Confirmation of *pyrF* deletion was done by colony PCR with primers BG2381 and BG2382.

RNA isolation

RNA isolation was performed by the phenol extraction based on a previously described protocol (*342*). Overnight 10 mL cultures were centrifuged at 4 ℃ and 4816*g* for 15 min and immediately used for RNA isolation. After removal of the medium, cells were suspended in 0.5 mL of ice-cold TE buffer (pH 8.0) and kept on ice. All samples were divided into two 2 mL screw-capped tubes containing 0.5 g of zirconium beads, 30 µL of 10% SDS, 30 µL of 3 M sodium acetate (pH 5.2), and 500 µL of Roti-Phenol (pH 4.5–5.0, Carl Roth GmbH). Cells were disrupted using a FastPrep-24 apparatus (MP Biomedicals) at 5500 rpm for 45 s and centrifuged at 4 ℃ and 9400*g* for 5 min. 400 µL of the water phase from each tube was transferred to a new tube, to which 400 µL of chloroform–isoamyl alcohol (Carl Roth GmbH) was added, after which samples were centrifuged at 4 ℃ and 18 400*g* for 3 min. 300 µL of the aqueous phase was transferred to a new tube and mixed with 300 µL of the lysis buffer from the high pure RNA isolation kit (Roche). Subsequently, the rest of the procedure from this kit was performed according to the manufacturer's protocol, except for the DNase incubation step, which was performed for 45 min. The concentration and integrity of cDNA was determined using Nanodrop-1000 Integrity and concentration of the isolated RNA was checked on a NanoDrop 1000.

Quantification of mRNA by RT-qPCR

First-strand cDNA synthesis was performed for the isolated RNA using SuperScriptTM III Reverse Transcriptase (Invitrogen) according to manufacturer's protocol. qPCR was performed using the PerfeCTa SYBR Green Supermix for iQ from Quanta Biosciences. 40 ng of each cDNA library was used as the template for qPCR. Two sets of primers were used; BG9665:BG9666 amplifying a 150-nt long region of the *ldhL* gene and BG9889:BG9890 amplifying a 150-nt long sequence of the *rpoD* (RNA polymerase sigma factor) gene which was used as the control for the qPCR. The qPCR was run on a Bio-Rad C1000 Thermal Cycler.

HPLC

A high-pressure liquid chromatography (HPLC) system ICS-5000 was used for lactate quantification. The system was operated with Aminex HPX 87H column from Bio-Rad Laboratories and equipped with a UV1000 detector operating on 210 nm and a RI-150 40 ℃ refractive index detector. The mobile phase consisted of 0.16 N H_2SO_4 and the column was operated at 0.8 mL min^{-1}. All samples were diluted 4:1 with 10 mM DMSO in 0.01 N H_2SO_4.

Acknowledgements

We would like to thank Koen Weenink, Steven Aalvink and Bastienne Vriesendorp for their technical assistance. R.v.K. and R.B. are employed by Corbion. I.M. and E.F.B. are supported by the Dutch Technology Foundation STW, which is part of The Netherlands Organization for Scientific Research (NWO) and which is partly funded by the Ministry of Economic Affairs. J.v.d.O. and P.M. are supported by the NWO/TOP grant 714.015.001.

Author contributions

I.M., P.M., E.F.B., R.v.K., and J.v.d.O., conceived this study and design of experiments. I.M., P.M., E.F.B., M.F., V.V., M.N., A.G., and R.B. conducted the experiments. R.v.K. and J.v.d.O. supervised this project. I.M., P.M., E.F.B., R.v.K., and J.v.d.O. wrote the manuscript with input from all authors.

Author information

Correspondence should be addressed to john.vanderoost@wur.nl.

Competing interesting

A patent application has been filed based on the work in the paper.

Data availability

Plasmids expressing ThermoCas9 or ThermodCas9, together with the corresponding sgRNA, are available on Addgene (#100981 & #100982).

Supplementary Figures and Tables

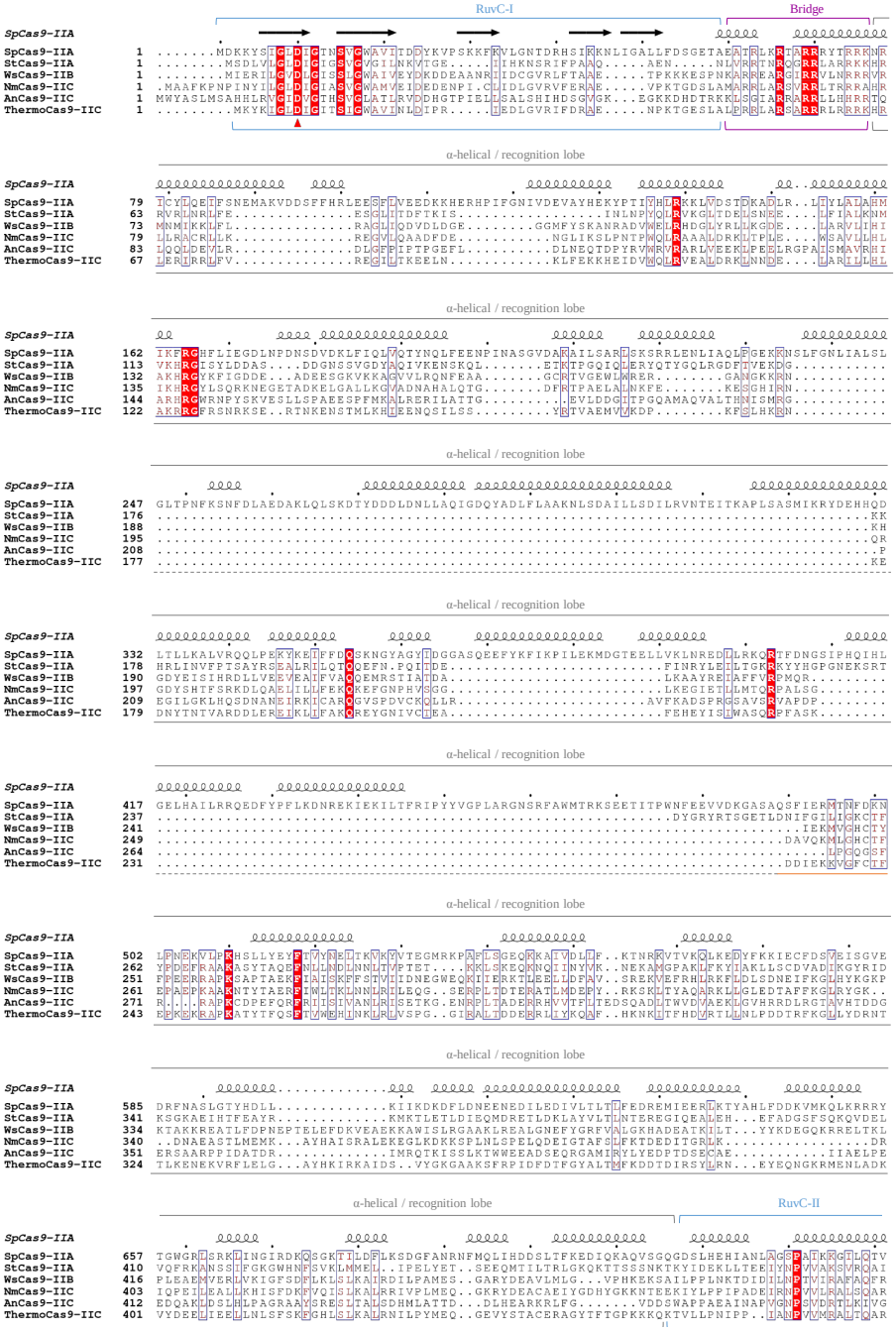


Figure S6.1 | Multiple sequence alignment of Type II-A, B and C Cas9 orthologues.

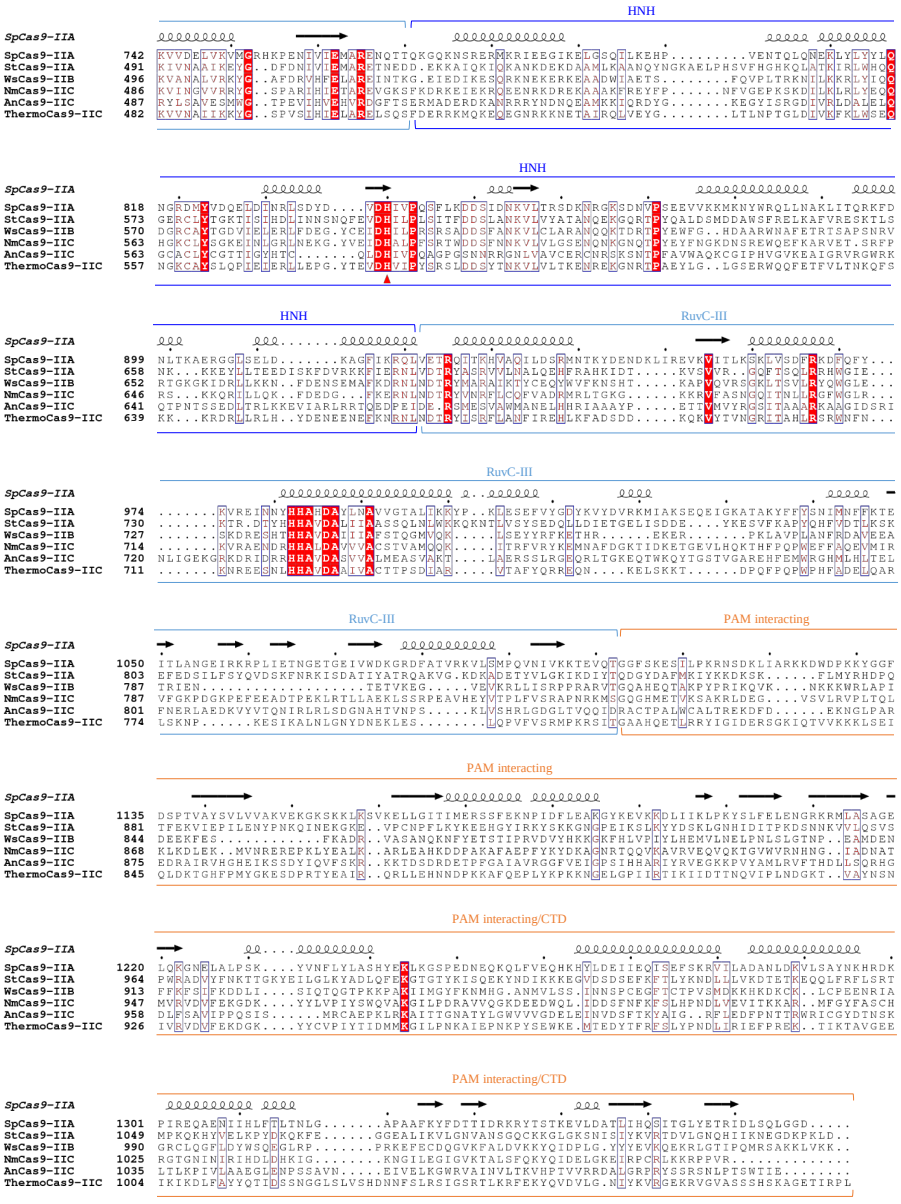


Figure S6.1 | Multiple sequence alignment of Type II-A, B and C Cas9 orthologues. (continued) Cas9 protein sequences of *Streptococcus pyogenes* (Sp), *Streptococcus thermophilus* (St), *Wolinella succinogenes* (Ws), *Neisseria meningitides* (Nm), *Actinomyces naeslundii* (An), and *Geobacillus thermodenitrificans* (Thermo) were aligned using ClustalW (*343*) in MEGA7 (*322*) with default settings; ESPript (*344*) was used to generate the visualization. Strictly conserved residues are shown in white text on red background; similar residues are shown in red text on white background. Red pyramids indicate the two conserved nuclease domains in all sequences. Horizontal blacks arrow and curls indicate β-strands and α-helices, respectively, in the SpCas9 secondary structure (protein database nr 4CMP (*216*)). Structural domains are indicated for SpCas9 and ThermoCas9 using the same colour scheme as in **Figure 6.1A**.

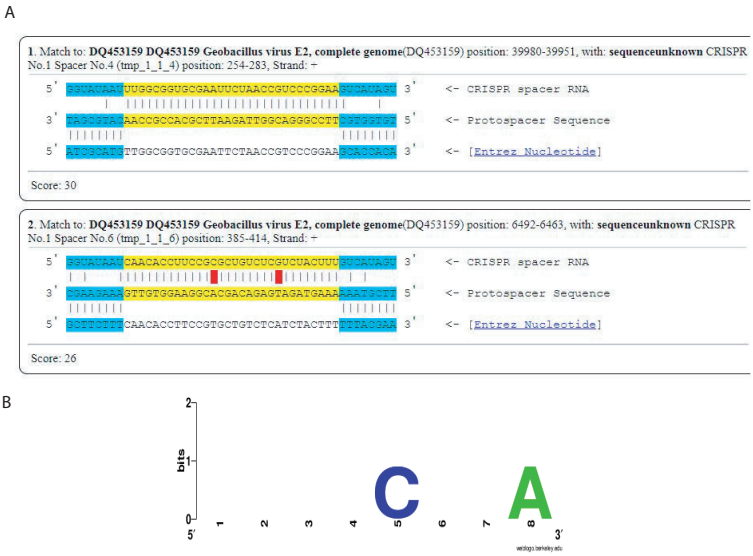


Figure S6.2 | *In silico* PAM determination results. (A) The two hits obtained with phage genomes using CRISPRtarget (*323*). **(B)** Sequence logo of the consensus 7-nt long PAM of ThermoCas9, obtained by in silico PAM analysis. Letter height at each position is measured by information content (*340*).

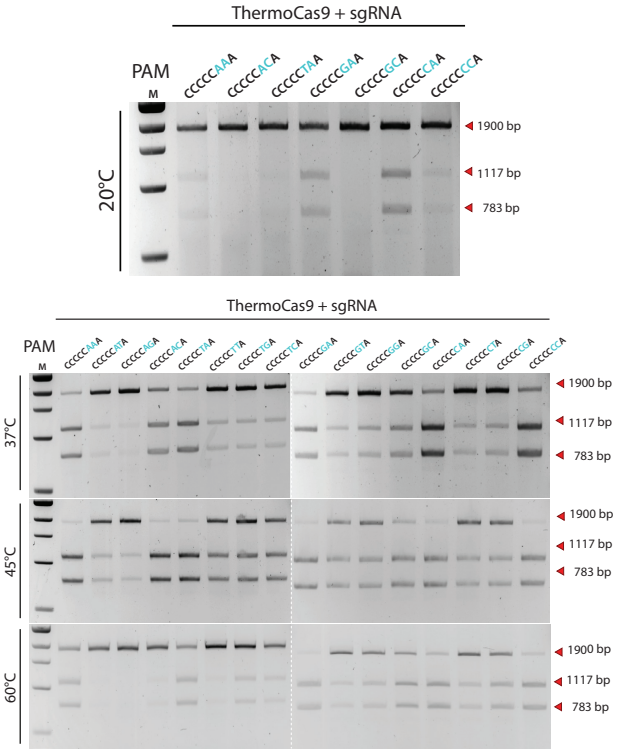


Figure S6.3 | ThermoCas9 PAM discovery. *In vitro* cleavage assays for DNA targets with different PAMs at 20˚C, 37˚C, 45˚C and 60˚C. 7 (20˚C) or 16 (37˚C, 45˚C, 60˚C) linearized plasmid targets, each containing a distinct 5'-CCCCCNNA-3' PAM, were incubated with ThermoCas9 and sgRNA, then analysed by agarose gel electrophoresis.

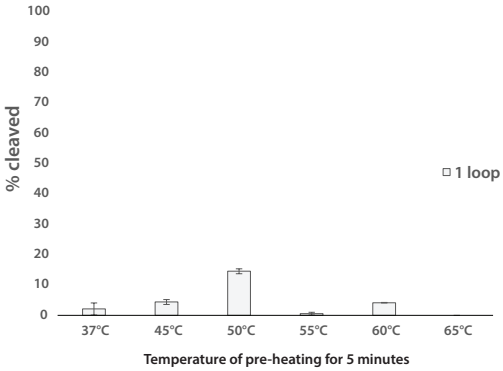


Figure S6.4 | Activity of ThermoCas9 at a wide temperature range using sgRNA containing one stem-loop. The importance of the predicted three stem loops of the tracrRNA scaffold was tested by transcribing truncated variations of the sgRNA and evaluating their ability to guide ThermoCas9 to cleave target DNA at various temperatures. Shown above is the effect of one stem-loop on the activity of ThermoCas9 at various temperatures. Average values from at least two biological replicates are shown, with error bars representing S.D.

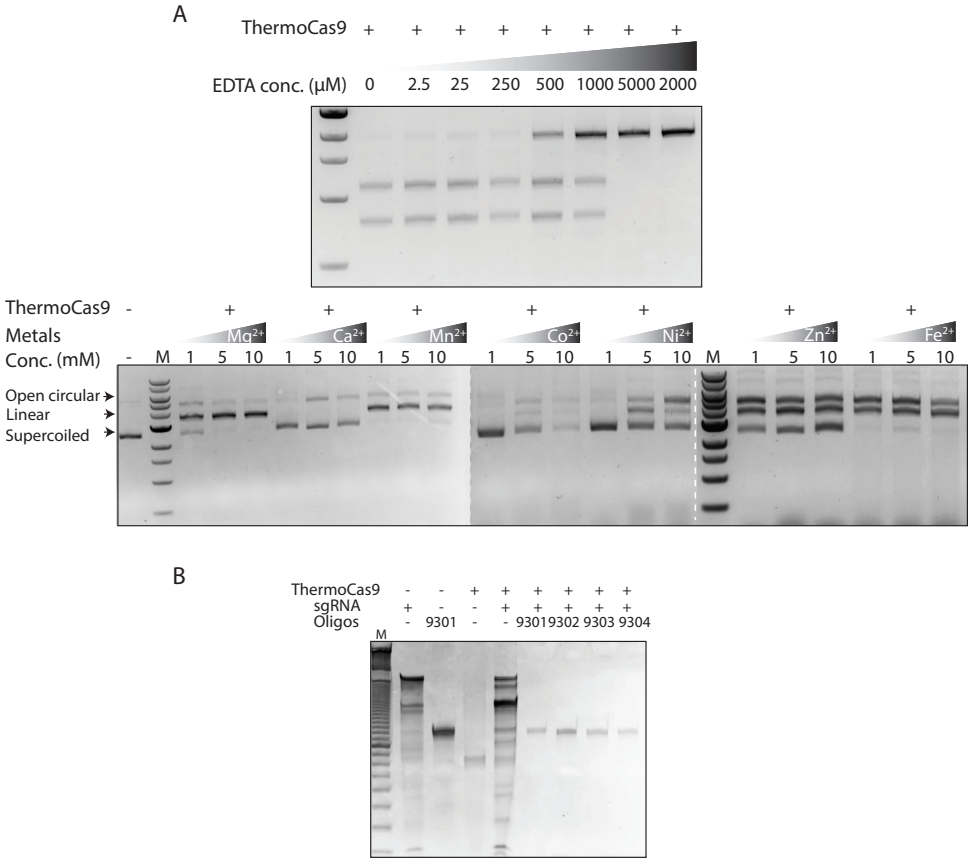


Figure S6.5 | ThermoCas9 mediates dsDNA targeting using divalent cations as catalysts and does not cleave ssDNA. (A) *In vitro* DNA cleavage by ThermoCas9 with EDTA and various metal ions. M = 1 kb DNA ladder. **(B)** Activity of ThermoCas9 on ssDNA substrates. M= 10 bp DNA ladder.

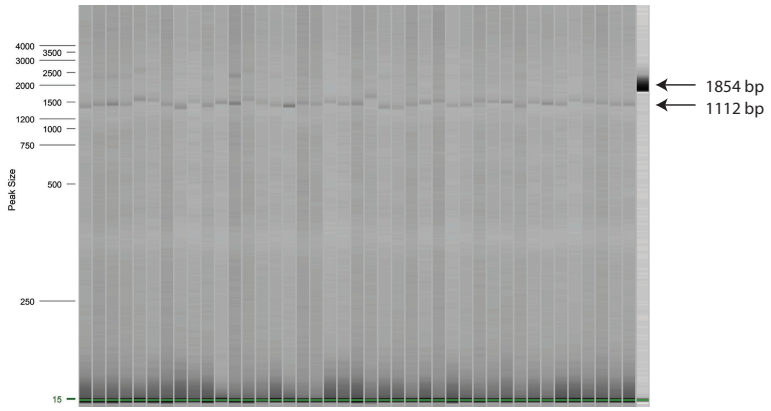


Figure S6.6 | Colony PCR of *P. putida* ThermoCas9-based *pyrF* deletion. Capillary gel electrophoresis showing the resulting products from genome-specific PCR on the obtained colonies from the ThermoCas9-based *pyrF* deletion process from the genome of *Pseudomonas putida*. The 1854 bp band and the 1112 bp band corresponds to the *pyrF* and Δ*pyrF* genotype, respectively.

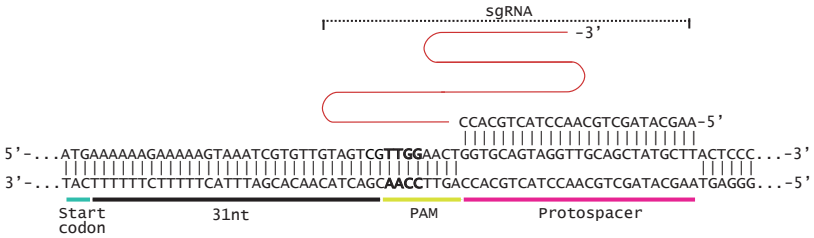


Figure S6.7 | Spacer selection for the *ldhL* silencing experiment. Schematic representation of the spacer (sgRNA)-protospacer annealing during the *ldhL* silencing process; the selected protospacer resides on the non- template strand and 39nt downstream the start codon of the *ldhL* gene. The PAM sequence is shown in bold.

Table S6.1 | pBLAST results of Cas9 protein sequences from Figure 6.1B compared to ThermoCas9.

Species	% identity[a]
Geobacillus 47C-IIb	99
Geobacillus 46C-IIa	89
Geobacillus LC300	89
Geobacillus jurassicus	89
Geobacillus MAS1	88
Geobacillus stearothermophilus	88
Geobacillus stearothermophilus ATCC 12980	88
Geobacillus Sah69	88
Geobacillus stearothermophilus	88
Geobacillus kaustophilus	88
Geobacillus stearothermophilus	88
Geobacillus genomosp. 3	87
Geobacillus genomosp. 3	87
Geobacillus subterraneus	87
Effusibacillus pohliae	86

[a] Query coverage was 100% in all cases.

Table S6.2 | Oligonucleotides used in this study, related to Figures 6.1-6.5.

Oligo ID	Sequence (5' - 3')	Description
	PAM Library construction	
BG6494	TATGCCTCATGAGATTATCAAAAAGGATCTTCACNNNNNNNCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with 7-nt long random PAM sequence
BG6495	TATGCCGGATCCTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAG	RV for construction of *in vitro* target DNA sequences
BG7356	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-T-	Adaptor when annealed with BG7357, ligates to A-tailed ThermoCas9 cleaved fragments
BG7357	CTGTCTCTTATACACATCTGACGCTGCCGACGA	Adaptor when annealed with BG7356, ligates to A-tailed ThermoCas9 cleaved fragments
BG7358	TCGTCGGCAGCGTCAG	FW sequencing adaptor for PCR amplification of the ThermoCas9 cleaved fragments
BG7359	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACCATGATTACGCCAAGC	RV sequencing adapter for PCR amplification of the ThermoCas9 cleaved fragments
BG7616	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTCATGAGATTATCAAAAAGGATCTTC	RV sequencing adaptor for PCR amplification of the control fragments
BG8157	TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCCAGCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM CCCCCCAG
BG8158	TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCCAACTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM CCCCCCAA
BG8159	TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCCATCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM CCCCCCAT
BG8160	TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM CCCCCCAC
BG8161	TATGCCTCATGAGATTATCAAAAAGGATCTTCACNNNNTNNCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM NNNNTNN
BG8363	ACGGTTATCCACAGAATCAG	FW for PCR linearization of PAM identification libraries
BG8364	CGGGATTGACTTTTAAAAAAGG	RV for PCR linearization of PAM identification libraries
BG8763	TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCAAACTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6 & 7 "AA"
BG8764	TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCATACTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "AT"
BG8765	TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCAGACTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "AG"

FW: Forward primer; RV: Reverse primer
Nucleotides in lowercase letters correspond to primer overhangs for HiFi DNA Assembly.
Restriction sites are double underlined. The PAMs are underlined. Spacer regions are shown in ***bold and gray***

Table S6.2 | Oligonucleotides used in this study, related to Figures 6.1-6.5. (continued)

Oligo ID	Sequence (5' - 3')	Description
BG8766	TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCACACTAGATCCTTTTA AATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "AC"
BG8767	TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCTAACTAGATCCTTTTA AATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "TA"
BG8768	TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCTTACTAGATCCTTTTA AATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "TT"
BG8769	TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCTGACTAGATCCTTTTA AATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "TG"
BG8770	TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCTCACTAGATCCTTTTA AATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "TC"
BG8771	TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCGAACTAGATCCTTTTA AATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "GA"
BG8772	TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCGTACTAGATCCTTTTA AATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "GT"
BG8773	TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCGGACTAGATCCTTTTA AATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "GG"
BG8774	TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCGCACTAGATCCTTTTA AATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "GC"
BG8775	TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCCAACTAGATCCTTTTA AATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "CA"
BG8776	TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCCTACTAGATCCTTTTA AATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "CT"
BG8777	TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCCGACTAGATCCTTTTA AATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "CG"
BG8778	TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCCCACTAGATCCTTTTA AATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "CC"
	sgRNA module for *in vitro* transcription	
BG6574	AAGCTTGAAATAATACGACTCACTATAGG	FW for PCR amplification of the sgRNA template for the first PAM identification process (30nt long spacer)
BG6576	AAAAAAGACCTTGACGTTTTCC	FW for PCR amplification of the sgRNA template for the first PAM identification process
BG9307	AAGCTTGAAATAATACGACTCACTATAGGTGAGATTATCAAAAAGGATCTTCACG TC	RV for PCR amplification of the sgRNA template for all the PAM identification processes except the first one (25nt long spacer)

FW: Forward primer; RV: Reverse primer

Nucleotides in lowercase letters correspond to primer overhangs for HiFi DNA Assembly.

Restriction sites are double underlined. The PAMs are underlined. Spacer regions are shown in ***bold and gray***

Table S6.2 | Oligonucleotides used in this study, related to Figures 6.1-6.5. (continued)

Oligo ID	Sequence (5' - 3')	Description
BG9309	AAAACGCCTAAGAGTGGGGAATG	RV for PCR amplification of the 3-hairpins long sgRNA template for all the PAM identification processes except the first one
BG9310	AAAAGGCGATAGGCGATCC	RV for PCR amplification of the 2-hairpins long sgRNA template for all the PAM identification processes except the first one
BG9311	AAAACGGGTCAGTCTGCCTATAG	RV for PCR amplification of the 1-hairpin long sgRNA template for all the PAM identification processes except the first one
BG9308	TGAGATTATCAAAAAGGATCTTCACGTC	FW 25nt spacer sgRNA
BG9312	AAAACGCCTAAGAGTGGGGAATGCCCGAAGAAAGCGGGCGATAGGCGATCC	RV 3 loops sgRNA
	Construction of targets for mismatch tolerance assays	
BG10561	CATGAGATTATCAAAAAGGATCTTCAACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with a mismatch at PAM proximal position 1
BG10562	GATCCCTAGTTGGGGGGTTGAAGATCCTTTTTGATAATCT	RV for construction of *in vitro* target DNA with a mismatch at PAM proximal position 1
BG10563	CATGAGATTATCAAAAAGGATCTTCCCCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with a mismatch at PAM proximal position 2
BG10564	GATCCCTAGTTGGGGGGGGGAAGATCCTTTTTGATAATCT	RV for construction of *in vitro* target DNA with a mismatch at PAM proximal position 2
BG10565	CATGAGATTATCAAAAAGGATCTTCAACCCCCCAACTAGG	FW for construction of *in vitro* target DNA with a mismatch at PAM proximal position 3
BG10566	GATCCCTAGTTGGGGGGTTGAAGATCCTTTTTGATAATCT	RV for construction of *in vitro* target DNA with a mismatch at PAM proximal position 3
BG10567	CATGAGATTATCAAAAAGGATCTTCCCCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with a mismatch at PAM proximal position 4
BG10568	GATCCCTAGTTGGGGGGGGGAAGATCCTTTTTGATAATCT	RV for construction of *in vitro* target DNA with a mismatch at PAM proximal position 4
BG10569	CATGAGATTATCAAAAAGGATCTTAACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with a mismatch at PAM proximal position 5
BG10570	GATCCCTAGTTGGGGGGGTTAAGATCCTTTTTGATAATCT	RV for construction of *in vitro* target DNA with a mismatch at PAM proximal position 5
BG10571	CATGAGATTATCAAAAAGGATCTGCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with a mismatch at PAM proximal position 6

FW: Forward primer; RV: Reverse primer
Nucleotides in lowercase letters correspond to primer overhangs for HiFi DNA Assembly.
Restriction sites are double underlined. The PAMs are underlined. Spacer regions are shown in ***bold and gray***

Table S6.2 | Oligonucleotides used in this study, related to Figures 6.1-6.5. (continued)

Oligo ID	Sequence (5' - 3')	Description
BG10572	GATCCCTAGTTGGGGGGGTGCAGATCCTTTTTGATAATCT	RV for construction of *in vitro* target DNA with a mismatch at PAM proximal position 6
BG10573	CATGAGATTATCAAAAAGGATCGTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with a mismatch at PAM proximal position 7
BG10574	GATCCCTAGTTGGGGGGGTGACGATCCTTTTTGATAATCT	RV for construction of *in vitro* target DNA with a mismatch at PAM proximal position 7
BG10575	CATGAGATTATCAAAAAGGATATTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with a mismatch at PAM proximal position 8
BG10576	GATCCCTAGTTGGGGGGGTGAATATCCTTTTTGATAATCT	RV for construction of *in vitro* target DNA with a mismatch at PAM proximal position 8
BG10577	CATGAGATTATCAAAAAGGAGCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with a mismatch at PAM proximal position 9
BG10578	GATCCCTAGTTGGGGGGGTGAAGCTCCTTTTTGATAATCT	RV for construction of *in vitro* target DNA with a mismatch at PAM proximal position 9
BG10579	CATGAGATTATCAAAAAGGCTCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with a mismatch at PAM proximal position 10
BG10580	GATCCCTAGTTGGGGGGGTGAAGAGCCTTTTTGATAATCT	RV for construction of *in vitro* target DNA with a mismatch at PAM proximal position 10
BG10581	CATGAGATTATCAAAAAGTATCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with a mismatch at PAM proximal position 15
BG10582	GATCCCTAGTTGGGGGGGTGAAGATACTTTTTGATAATCT	RV for construction of *in vitro* target DNA with a mismatch at PAM proximal position 15
BG10583	CATGAGATTATCAAAAATGATCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with a mismatch at PAM proximal position 20
BG10584	GATCCCTAGTTGGGGGGGTGAAGATCATTTTTGATAATCT	RV for construction of *in vitro* target DNA with a mismatch at PAM proximal position 20
BG10585	CATGAGATTATCCAAAAGGATCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatches at PAM distal positions 19-20
BG10586	GATCCCTAGTTGGGGGGGTGAAGATCCTTTTGGATAATCT	RV for construction of *in vitro* target DNA with mismatches at PAM distal positions 19-20
BG10587	CATGAGAGTATCAAAAAGGATCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatches at PAM distal positions 17-20
BG10588	GATCCCTAGTTGGGGGGGTGAAGATCCTTTTTGATACTCT	RV for construction of *in vitro* target DNA with mismatches at PAM distal positions 17-20

FW: Forward primer; RV: Reverse primer
Nucleotides in lowercase letters correspond to primer overhangs for HiFi DNA Assembly.
Restriction sites are double underlined. The PAMs are underlined. Spacer regions are shown in ***bold and gray***

Table S6.2 | Oligonucleotides used in this study, related to Figures 6.1-6.5. (continued)

Oligo ID	Sequence (5' - 3')	Description
BG10589	CATGAGAGGATCAAAAAGGATCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatches at PAM distal positions 15-20
BG10590	GATCCCTAGTTGGGGGGGTGAAGATCCTTTTTGATCCTCT	RV for construction of *in vitro* target DNA with mismatches at PAM distal positions 15-20
BG10591	CATGAGAGGCGCAAAAAGGATCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatches at PAM distal positions 13-20
BG10592	GATCCCTAGTTGGGGGGGTGAAGATCCTTTTTGCGCCTCT	RV for construction of *in vitro* target DNA with mismatches at PAM distal positions 13-20
BG10593	CATGAGAGGCGACAAAAGGATCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatches at PAM distal positions 11-20
BG10594	GATCCCTAGTTGGGGGGGTGAAGATCCTTTTGTCGCCTCT	RV for construction of *in vitro* target DNA with mismatches at PAM distal positions 11-20
BG10595	CATGAGAGGCGACCCAAGGATCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatches at PAM distal positions 9-20
BG10596	GATCCCTAGTTGGGGGGGTGAAGATCCTTGGGTCGCCTCT	RV for construction of *in vitro* target DNA with mismatches at PAM distal positions 9-20
BG10597	CATGAGAGGCGACCCCCGGATCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatches at PAM distal positions 5-20
BG10598	GATCCCTAGTTGGGGGGGTGAAGATCCGGGGGTCGCCTCT	RV for construction of *in vitro* target DNA with mismatches at PAM distal positions 5-20
BG10599	CATGAGAGGCGACCCCCTTATCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatches at PAM distal positions 0-20
BG10600	GATCCCTAGTTGGGGGGGTGAAGATAAGGGGGTCGCCTCT	RV for construction of *in vitro* target DNA with mismatches at PAM distal positions 0-20
	Construction of editing and silencing constructs	
BG8194	tatggcgaatcacaacatgggaactcatgaGAACATCCTCTTTCTTAG	For the construction of the pThermoCas9_ctrl plasmid & pThermoCas9_bsΔ*pyrF*1/2
BG8195	gccgatatcaagaccgattttatacttcatTTAAGTTACCTCCTCGATTG	For the construction of the pThermoCas9_ctrl plasmid & pThermoCas9_bsΔ*pyrF*1/2
BG8196	ATGAAGTATAAAATCGGTCTTG	For the construction of the pThermoCas9_ctrl plasmid & pThermoCas9_bsΔ*pyrF*1/2
BG8191	AAGCTTGGCGTAATCATGGTC	For the construction of the pThermoCas9_ctrl plasmid & pThermoCas9_bsΔpyrF1/2

FW: Forward primer; RV: Reverse primer
Nucleotides in lowercase letters correspond to primer overhangs for HiFi DNA Assembly.
Restriction sites are double underlined. The PAMs are underlined. Spacer regions are shown in ***bold and gray***

Table S6.2 | Oligonucleotides used in this study, related to Figures 6.1-6.5. (continued)

Oligo ID	Sequence (5' - 3')	Description
BG8192	TCATGAGTTCCCATGTTGTG	For the construction of the pThermoCas9_ctrl plasmid & pThermoCas9_bsΔ*pyrF*1/2
BG8197	TAACGGACGGATAGTTTC	For the construction of the pThermoCas9_ctrl plasmid & pThermoCas9_bsΔ*pyrF*1/2
BG8198	gaaagccggggaaactatccgtccgttataAATCAGACAAAATGGCCTGCTTATG	For the construction of the pThermoCas9_ctrl plasmid & pThermoCas9_bsΔ*pyrF*1/2
BG8263	gaactatgacactttattttcagaatggacGTATAACGGTATCCATTTTAAGAATAATCC	For the construction of the pThermoCas9_ctrl plasmid
BG8268	accgttatacgtccattctgaaaataaagtGTCATAGTTCCCCTGAGAT	For the construction of the pThermoCas9_ctrl plasmid
BG8210	aacagctatgaccatgattacgccaagcttCCCTCCCATGCACAATAG	For the construction of the pThermoCas9_ctrl plasmid & pThermoCas9_bsΔ*pyrF*1/2
BG8261	gaactatgacatcatggagttttaaatccaGTATAACGGTATCCATTTTAAGAATAATCC	For the construction of the pThermoCas9_bsΔ*pyrF*1
BG8266	accgttatactggatttaaaactccatgatGTCATAGTTCCCCTGAGAT	For the construction of the pThermoCas9_bsΔ*pyrF*2
BG8317	gaactatgaccacccagcttacatcaacaaGTATAACGGTATCCATTTTAAGAATAATCC	For the construction of the pThermoCas9_bsΔ*pyrF*2
BG8320	accgttatacttgttgatgtaagctgggtgGTCATAGTTCCCCTGAGAT	For the construction of the pThermoCas9_bsΔ*pyrF*2
BG9075	CTATCGGCATTACGTCTATC	For the construction of the pThermoCas9i_ctrl
BG9076	GCGTCGACTTCTGTATAGC	For the construction of the pThermoCas9i_ctrl
BG9091	TGAAGTATAAAATCGGTCTTGCTATCGGCATTACGTCTATC	For the construction of the pThermoCas9i_ctrl
BG9092	CAAGCTTCGGCTGTATGGAATCACAGCGTCGACTTCTGTATAGC	For the construction of the pThermoCas9i_ctrl
BG9077	GCTGTGATTCCATACAG	For the construction of the pThermoCas9i_ctrl
BG9267	GGTGCAGTAGGTTGCAGCTATGCTTGTATAACGGTATCCAT	For the construction of the pThermoCas9i_ctrl
BG9263	AAGCATAGCTGCAACCTACTGCACCGTCATAGTTCCCCTGAGATTATCG	For the construction of the pThermoCas9i_ctrl
BG9088	TCATGACCAAAATCCCTTAACG	For the construction of the pThermoCas9i_ctrl
BG9089	TTAAGGGATTTTGGTCATGAGAACATCCTCTTTCTTAG	For the construction of the pThermoCas9i_ctrl
BG9090	GCAAGACCGATTTTATACTTCATTTAAG	For the construction of the pThermoCas9i_ctrl
BG9548	GGATCCCATGACGCTAGTATCCAGCTGGGTCATAGTTCCCCTGAGATTATCG	For the construction of the pThermoCas9i_*ldhL*
BG9601	TTCAATATTTTTTTTGAATAAAAAATACGATACAATAAAAATGTCTAGAAAAAGATAAAAATG	For the construction of the pThermoCas9i_*ldhL*
BG9600	TTTTTTATTCAAAAAAAATATTGAATTTTAAAAATGATGGTGCTAGTATGAAG	For the construction of the pThermoCas9i_*ldhL*
BG9549	CCAGCTGGATACTAGCGTCATGGGATCCGTATAACGGTATCCATTTTAAGAATAATCC	For the construction of the pThermoCas9i_*ldhL*
BG8552	TCGGGGGTTCGTTTCCCTTG	FW to check genomic *pyrF* deletion KO check

FW: Forward primer; RV: Reverse primer

Nucleotides in lowercase letters correspond to primer overhangs for HiFi DNA Assembly.

Restriction sites are double underlined. The PAMs are underlined. Spacer regions are shown in ***bold and gray***

Table S6.2 | Oligonucleotides used in this study, related to Figures 6.1-6.5. (continued)

Oligo ID	Sequence (5' - 3')	Description
BG8553	CTTACACAGCCAGTGACGGAAC	RV to check genomic *pyrF* deletion KO check
BG2365	GCCGGCGTCCCGGAAAACGA	For the construction of the pThermoCas9_ppΔ*pyrF*
BG2366	GCAGGTCGGGTTCCTCGCATCCATGCCCCCGAACT	For the construction of the pThermoCas9_ppΔ*pyrF*
BG2367	ggcttcggaatcgttttccgggacgccggcACGGCATTGGCAAGGCCAAG	For the construction of the pThermoCas9_ppΔ*pyrF*
BG2368	gacacaggcatcggtGCAGGGTCTCTTGGCAAGTC	For the construction of the pThermoCas9_ppΔ*pyrF*
BG2369	gccaagagaccctgCACCGATGCCTGTGTCGAACC	For the construction of the pThermoCas9_ppΔ*pyrF*
BG2370	cttggcggaaaacgtcaaggtctttttacACGCGCATCAACTTCAAGGC	For the construction of the pThermoCas9_ppΔ*pyrF*
BG2371	atgacgagctgttcaccagcagcgcTATTATTGAAGCATTTATCAGGG	For the construction of the pThermoCas9_ppΔ*pyrF*
BG2372	GTAAAAAAGACCTTGACGTTTTC	For the construction of the pThermoCas9_ppΔ*pyrF*
BG2373	tatgaagcgggccatTTGAAGACGAAAGGGCCTC	For the construction of the pThermoCas9_ppΔ*pyrF*
BG2374	taatagcgctgctggtgaacagctcGTCATAGTTCCCCTGAGATTATCG	For the construction of the pThermoCas9_ppΔ*pyrF*
BG2375	tggagtcatgaacatATGAAGTATAAAATCGGTCTTG	For the construction of the pThermoCas9_ppΔ*pyrF*
BG2376	ccctttcgtcttcAAATGGCCCGCTTCATAAGCAG	For the construction of the pThermoCas9_ppΔ*pyrF*
BG2377	gattttatacTTCATATGTTCATGACTCCATTATTATTG	For the construction of the pThermoCas9_ppΔ*pyrF*
BG2378	gggggcatggatgCGAGGAACCCGACCTGCATTGG	For the construction of the pThermoCas9_ppΔ*pyrF*
BG2381	ACACGGCGGATGCACTTACC	FW for confirmation of plasmid integration and pyrF deletion in *P. putida*
BG2382	TGGACGTGTACTTCGACAAC	RV for confirmation of pyrF deletion in *P. putida*
BG2135	ACACGGCGGATGCACTTACC	RV for confirmation of plasmid integration in *P. putida*
	Sequencing primers	
BG8196	ATGAAGTATAAAATCGGTCTTG	*thermocas9* seq. 1
BG8197	TAACGGACGGATAGTTTC	*thermocas9* seq. 2
BG6850	GCCTCATGAATGCAGCGATGGTCCGGTGTTC	pyrF US
BG6849	GCCTCATGAGTTCCCATGTTGTGATTC	pyrF DS
BG6769	CAATCCAACTGGGCTTGAC	*thermocas9* seq. 3
BG6841	CAAGAACTTTATTGGTATAG	*thermocas9* seq. 4
BG6840	TTGCAGAAATGGTTGTCAAG	*thermocas9* seq. 5
BG9215	GAGATAATGCCGACTGTAC	pNW33n backbone seq. 1
BG9216	AGGGCTCGCCTTTGGGAAG	pNW33n backbone seq. 2
BG9505	GTTGCCAACGTTCTGAG	*thermocas9* seq. 6
BG9506	AATCCACGCCGTTTAG	*thermocas9* seq. 7
BG9505	GTTGCCAACGTTCTGAG	*thermocas9* seq. 8
BG9506	AATCCACGCCGTTTAG	*thermocas9* seq. 9

FW: Forward primer; RV: Reverse primer

Nucleotides in lowercase letters correspond to primer overhangs for HiFi DNA Assembly.

Restriction sites are double underlined. The PAMs are underlined. Spacer regions are shown in ***bold and gray***

Table S6.2 | Oligonucleotides used in this study, related to Figures 6.1-6.5. (continued)

Oligo ID	Sequence (5' - 3')	Description
	Cleavage assays	
BG8363	ACGGTTATCCACAGAATCAG	FW for PCR linearization of DNA target
BG8364	CGGGATTGACTTTTAAAAAAGG	RV for PCR linearization of DNA target
BG9302	AAACTTCATTTTTAATTTAAAAGGATCTAGAACCCCCCGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAA	Non-template strand oligonucleotide for ssDNA cleavage assays
BG9303	TTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCCCCCCAACTAGATCCTTTTAAATTAAAAATGAAGTTT	Template strand oligonucleotide for ssDNA cleavage assays
BG9304	TTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACGGGGGGTTCTAGATCCTTTTAAATTAAAAATGAAGTTT	Template strand oligonucleotide for ssDNA cleavage assays
	ThermoCas9 expression	
BG7886	TACTTCCAATCCAATGCAAAGTATAAAATCGGTCTTGATATCG	FW LIC_*thermocas9*
BG7887	TTATCCACTTCCAATGTTATTATAACGGACGGATAGTTTCCCCGGCTTTCATGACGAAAGGAGTTTCTTATTATG	RV LIC_*thermocas9*
	RT-qPCR	
BG9665	ATGACGAAAGGAGTTTCTTATTATG	RV qPCR check *ldhl*
BG9666	AACGGTATTCCGTGATTAAG	FW qPCR check *ldhl*

FW: Forward primer; RV: Reverse primer

Nucleotides in lowercase letters correspond to primer overhangs for HiFi DNA Assembly.

Restriction sites are double underlined. The PAMs are underlined. Spacer regions are shown in ***bold and gray***

Table S6.3 | Plasmids used in this study, related to Figures 6.1-6.5.

Plasmid	Description	Restriction sites used	Primers	Source
pNW33n	E. coli-Bacillus shuttle vector, cloning vector, CamR	-	-	BGSC
pUC57_T7sgRNAfull	pUC57 vector containing DNA encoding the sgRNA under the control of T7 promoter; serves as a template for *in vitro* transcription of full-length Repeat/Antirepeat sgRNAs	-	-	Baseclear

LIC: Ligase Independent cloning; FW: Forward primer; RV: Reverse primer

Table S6.3 | Plasmids used in this study, related to Figures 6.1-6.5. (continued)

Plasmid	Description	Restriction sites used	Primers	Source
pMA2_T7sgRNAtruncated R/AR	Vector containing DNA encoding the truncated Repeat/Antirepeat part of the sgRNA under the control of T7 promoter; serves as a template for *in vitro* transcription of truncated Repeat/Antirepeat sgRNAs	-	-	Gen9
pRARE	T7 RNA polymerase-based expression vector, KanR	-	-	EMD Millipore
pML-1B	E. coli Rosetta™ (DE3) plasmid, encodes rare tRNAs, CamR	-	-	Macrolab, Addgene
pEMG	P. putida suicide vector, used as template for replicon and KanR		See **Table S6.2**	(*345*)
pSW_I-SceI	P. putida vector containing I-SceI, used as template for xylS and PPm		See **Table S6.2**	(*345*)
pWUR_Cas9sp1_hr	pNW33n with spCas9-module containing spacer targeting the *pyrF* gene. This plasmid was used as a template for constructing the ThermoCas9 based constructs	-	-	(*315*)
pThermo_Cas9	thermocas9 with N-term. His-tag and TEV cleavage site in pML-1B. Expression vector for ThermoCas9	SspI and Ligase Independent Cloning	BG7886 and BG7887	This study
pThermo_dCas9	cas9dthermocas9 with N-term. His-tag and TEV cleavage site in pML-1B. Expression vector for catalytically inactive (dead) ThermodCas9	SspI and Ligase Independent Cloning	BG7886 and BG7888	This study
pNW-PAM7nt	Target sequence in pNW33n vector containing a 7-nt degenerate PAM for *in vitro* PAM determination assay	BamHI and BspHI	See **Table S6.2**	This study
pNW63-pNW78	Target sequence in pNW33n vector containing distinct nucleotides at the 6th and 7th positions of the PAM (CCCCCNNA)	BamHI and BspHI	See **Table S6.2**	This study
pThermoCas9_ctrl	pNW33n with ThermoCas9-module1 containing a non-targeting spacer. Used as a negative control	-	See **Table S6.2**	This study
pThermoCas9_ppΔ*pyrF*1	pNW33n with ThermoCas9-module1 containing spacer 1 targeting the *pyrF* gene and the fused us+ds pyrF-flanks	-	See **Table S6.2**	This study
pThermoCas9_ppΔ*pyrF*2	pNW33n with ThermoCas9-module1 containing spacer 2 targeting the *pyrF* gene and the fused us+ds pyrF-flanks	-	See **Table S6.2**	This study
pThermoCas9i_ctrl	pNW33n with Thermo-dCas9-module2 containing a non-targeting spacer. Used as a wild-type control	-	See **Table S6.2**	This study
pThermoCas9i_ldhL	pNW33n with Thermo-dCas9-module2 containing spacer 2 targeting the ldhL gene	-	See **Table S6.2**	This study

LIC: Ligase Independent cloning; FW: Forward primer; RV: Reverse primer

Table S6.3 | Plasmids used in this study, related to Figures 6.1-6.5. (continued)

Plasmid	Description	Restriction sites used	Primers	Source
pThermoCas9_ppΔpyrF	pEMG with ThermoCas9-module3 for Pseudomonas putida containing a spacer targeting a spacer targeting the *pyrF* gene and the fused us+ds pyrF-flanks	-	See **Table S6.2**	This study

LIC: Ligase Independent cloning; FW: Forward primer; RV: Reverse primer

Chapter 7

Development of a CRISPR-Cas12a based genome editing tool for facultative thermophiles

Prarthana Mohanraju*, Ioannis Mougiakos*, Justin Albers, Megumu Mabuchi, Ryan T. Fuchs, Richard van Kranenburg, G. Brett Robb, John van der Oost

* contributed equally

Adapted from:
Development of a CRISPR-Cas12a based genome editing tool for facultative thermophiles
2019, Manuscript in preparation

Abstract

The ability of CRISPR–Cas12a nuclease to function reliably in a wide range of species has been key to its rapid adoption as a genome engineering tool. However, so far Cas12a has been limited for use in organisms that can grow below 42 ℃. Here, we characterise four Cas12a orthologs for their salt tolerance, as well as their pH- and temperature- stability. We demonstrate that Francisella novicida *Cas12a (FnCas12a) and* Eubacterium eligens *Cas12a (EeCas12a) have great potential for applications that require enhanced stability, including use at elevated temperatures. Furthermore, we exploit the temperature stability of these nucleases to develop a simple, single-plasmid Cas12a-based genome editing tool for* Bacillus smithii*, an industrially relevant moderate thermophile. We combined the CRISPR–Cas12a targeting system with plasmid-borne homologous recombination (HR) templates for developing a highly efficient gene-editing tool in* B. smithii *that yields gene-knockout mutants within 2 days. Altogether, our findings provide biochemical insights of the widely used Cas12a proteins and establish the first Cas12a-based bacterial genome editing tool that can be easily adapted for both mesophilic- and moderate thermophilic non-model organisms.*

Introduction

Bacterial cell factories can convert renewable feedstocks, like plant biomass-derived sugars, into green fuels or chemicals. However, the microbial fermentation-based bio-refinery production costs of fuels and chemicals are still relatively high and not competitive to the traditional fossil fuel-based industries. Thermophilic organisms are a valuable addition to the bio-refinery concept due to their multiple advantages over mesophilic organisms (*311*, *312*, *346*) such as (a) lower operational costs (cooling (*347*, *348*) and contamination costs (*326*, *349*, *350*)), (b) higher substrate and product solubility (*351*), (c) higher reaction and production rates (*352*) (d) reduced risk of enzyme product inhibition (*353*) and (e) reduced enzyme dosage, which usually represents 40–50% of the production costs (*352*). Typically, genetic manipulation of microbial cell factories, including thermophilic bacteria, is required when aiming to enhance production capacities. The traditional methods, based on recombination systems, are usually laborious and time-consuming, especially for non-model organisms, limiting both fundamental studies of these organisms and their development into industrial production hosts.

A wide range of CRISPR–Cas (Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR associated proteins) systems has been repurposed from prokaryotic immune systems to universal genetic engineering tools. Some Cas nucleases have revolutionized molecular biological research based on their efficiency, affordability, versatility and accessibility (13–21). The majority of the current tools exploit the RNA-guided DNA endonuclease from the type II CRISPR–Cas system of *Streptococcus pyogenes* (SpCas9). The CRISPR–Cas9 system consists of a Cas nuclease (Cas9), a trans-activating CRISPR RNA (tracrRNA), and a crRNA guide (*49*, *255*). Cas9 can generate double-stranded DNA breaks (DSBs) at the target sequence, also referred to as protospacer, when loaded and guided by a chimeric crRNA-tracrRNA single RNA molecule, denoted as single guide RNA (sgRNA) (*51*). The target-DNA binding, which precedes the Cas9 cleavage activity, is dependent on the initial recognition of a protospacer adjacent motif (PAM), a short (3 to 5 bp long) DNA sequence at the 3'end of the protospacer (*45*, *53*). The simplicity and facile programmability of CRISPR–Cas-derived DNA targeting systems have greatly facilitated efficient genetic manipulation of model and non-model bacteria. However, the use of Cas9 is limited in some organisms due to reported toxicity in the recipient cells (*306*, *354*, *355*).

Cas12a (formerly Cpf1), a type V-A endonuclease of the class 2 CRISPR–Cas system, is a dual nuclease that is involved in crRNA processing, target-site recognition, and DNA cleavage (*83*, *187*, *256*). Cas12a recognizes AT-rich PAM sequences and can be guided by a single crRNA without an additional tracrRNA (*83*). Similar to Cas9, Cas12a has successfully been used for genome editing in a repertoire of industrially relevant microorganisms (*256*, *307*, *356–358*).

While both Cas9 and Cas12a provide a robust set of tools, they all originate from mesophilic hosts, making them unsuitable for genome editing applications in thermophiles. We have previously developed a SpCas9-based engineering tool for facultative thermophiles, combining homologous recombination (HR) at elevated temperatures and SpCas9-based counter-selection at moderate temperatures (*315*). We employed the wide temperature growth range of the moderate thermophile *B. smithii* ET 138 as an induction system for SpCas9 expression. HR with plasmid-borne editing templates was performed at 45-55 ℃, where SpCas9 was shown to be inactive. Subsequent transfer to 37 ℃ allowed for counter-selection through production of active SpCas9 that introduces lethal double-stranded DNA breaks to the non-edited cells. The whole process took 4-7 days to obtain a gene deletion. However, SpCas9 requires a G-rich PAM sequence (5'-NGG-3') which may not always be available at the desired site of en-

gineering, more particularly in the AT-rich regions of the genome. On the other hand, Cas12a recognizes a AT-rich PAM sequence (5'-TTTN-3') for target cleavage. Therefore, addition of a thermostable Cas12a system would enable facile genome engineering and extend the toolbox in thermophilic organisms.

For that reason, we assessed the temperature tolerance and stability of four different Cas12a orthologs, and subsequently we explored if the *Francisella novicida* U112 Cas12a (FnCas12a) and *Eubacterium eligens* Cas12a (EeCas12a) nucleases could be used for genome editing in facultative thermophiles. As a proof of concept, we chose to engineer the industrially-relevant *Bacillus smithii* ET 138. Our results show that HR-CRISPRCas12a-based tool can be easily, rapidly and efficiently used for gene deletions in *B. smithii.* Moreover, we have further shortened the time of the process to 1-2 days to obtain a *B. smithii* ET 138 mutant. This study describes, to our knowledge, the first Cas12a-based genome editing tool for moderate thermophiles. Additionally, the functional temperature range of FnCas12a and EeCas12a complements that of the previously developed Cas12a systems, greatly expanding the temperatures that Cas12a can be used for both *in vitro* cleavage and genome-editing applications.

Results

In vitro characterisation of Cas12a nucleases

To adapt Cas12a nucleases for their use as genome editing tools in thermophilic organisms, we biochemically assessed different Cas12a variants for their stability and temperature sensitivity. To this end, we expressed in *E. coli* and purified recombinant Cas12a from *Francisella tularensis subsp. novicida U112* (FnCas12a), *Eubacterium eligens* (EeCas12a), *Acidaminococcus* sp. BV3L6 (AsCas12a) and *Lachnospiraceae bacterium ND2006* (LbCas12a). For comparison, the widely used Cas9 nuclease from *Streptococcus pyogenes* (SpCas9) was taken.

To assess the activity of the selected Cas12a and Cas9 nucleases at higher temperatures, we executed i*n vitro* target cleavage assays between 30 and 68 ℃. To facilitate easy monitoring of DNA cleavage, we generated a linear dsDNA substrate with the 5' end labelled with 6-carboxyfluorescein (FAM) using PCR. The respective guide RNA and the nucleases were separately pre-equilibrated at the designated temperature prior to the addition of the FAM labelled DNA substrate. The guide RNA loading to the protein to form the ribonucleoprotein (RNP) complex was also done at the same temperature and the cleavage reactions were initiated by the addition of the FAM labelled target-containing DNA substrate. The cleaved DNA fragments were resolved by capillary electrophoresis. We observed high cleavage activity by SpCas9 from 30 °C to 42 °C (**Figure 7.1A**); above this temperature SpCas9 activity rapidly decreased to undetectable levels, corroborating previous findings (*315*, *316*). The LbCas12a and AsCas12a nucleases exhibited reduced cleavage activity at temperatures above 30 °C (**Figure 7.1A**). Interestingly, we observed complete cleavage of DNA substrates with EeCas12a and FnCas12a at temperatures from 30 °C to 45 °C and even up to 50 °C for EeCas12a (**Figure 7.1A**). This indicates the potential to use FnCas12a and EeCas12a as a genome editing tool for both mesophilic as well as facultative thermophilic organisms.

In cells, the Cas12a:crRNA ribonucleoprotein RNP complex probably takes a few minutes to be formed. Therefore, we also evaluated the activity and thermostability of the recombinant proteins and RNP complexes using nano differential scanning fluorimetry (nanoDSF). NanoDSF monitors the intrinsic fluorescence of tryptophan or tyrosine residues as a thermal gradient is applied to the samples. The melting temperature (Tm) of the protein is calculated using the

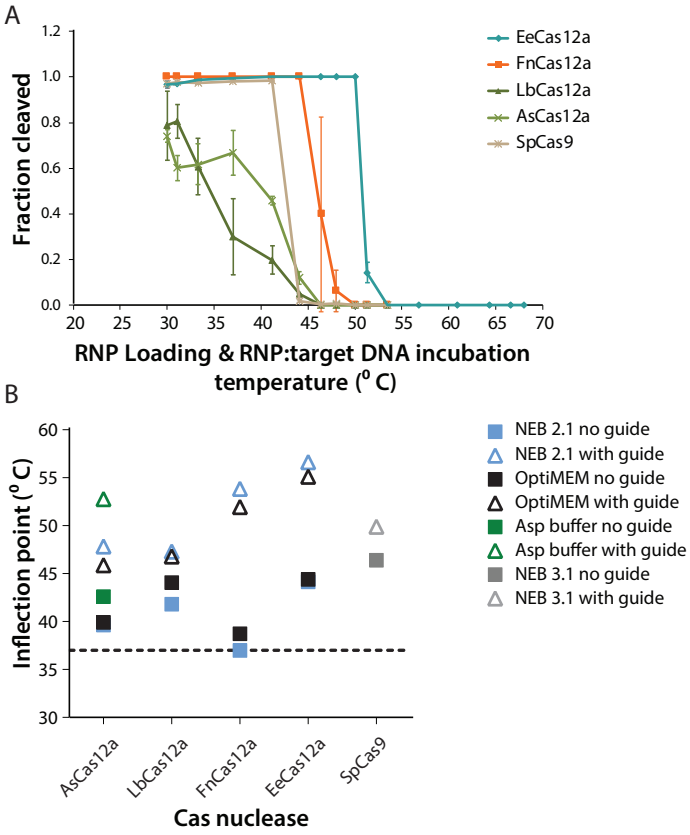


Figure 7.1|Activity of Cas12a orthologs at varied temperature. (A) Cas12a orthologs, SpCas9, with their respective guide RNAs and other reaction components were equilibrated to the loading and incubation temperatures as shown. Recombinant Cas protein, buffers (same as indicated in panel (B) and gRNAs were combined to program the RNPs for 5 minutes at the temperatures indicated. The dsDNA cleavage reactions were initiated by adding the 5'-FAM-labeled dsDNA containing the target site. Digested fragments were resolved by capillary electrophoresis and peaks corresponding to cleaved and intact substrates were quantified. Data is shown as mean +/- SD of at least two experimental replicates. **(B)** Recombinant guide-free Cas proteins or guide-loaded RNPs were subjected to thermal melting analysis using Nano differential scanning fluorimetry. RNPs were formed using a 2:1 guide RNA to protein ratio for >5 minutes in buffers as indicated. For the guide-free condition, recombinant protein was incubated in a mock loading conditions in the buffers incubated but without guide RNA. Data are shown as the inflection points of representative melting curves.

ratio of the intrinsic fluorescence of the protein at 330 and 350 nm, plotted against temperature (*359*).

We observed that SpCas9 has a melting temperature of 45 °C for the guide-free recombinant protein and ~50 °C for the RNP (**Figure 7.1B**). Both the LbCas12a RNPs and the guide-free recombinant protein had a Tm in the range of 42-47 °C (**Figure 7.1B**). AsCas12a had a melting temperature between 45 and 52 °C for loaded RNPs and between 40 and 42 °C for the guide-free recombinant protein (**Figure 7.1B**). The FnCas12a and EeCas12a RNPs have melting temperatures above 50°C. This is in contrast with the guide-free recombinant protein which had melting temperatures of ~37 °C and ~45 °C, respectively (**Figure 7.1B**). Of note, these determinations were performed both in the recommended *in vitro* reaction buffers, and in Opti-MEM (ThermoFisher) which is a commonly used diluent for transfection or electroporation of CRISPR nucleases into mammalian cells. Interestingly, AsCas12a-specific buffer had a stabilizing effect on both the RNP and protein, however no differences in thermal stability among the different buffer conditions was observed for FnCas12a and EeCas12a (**Figure 7.1B**). This reflects the overall trend that the thermal stability of the Cas12a orthologues strongly correlates with the association of an appropriate crRNA guide.

To evaluate the tolerance to variation in pH and salt concentration, we performed the above-

described *in vitro* cleavage assays over a range of pH and NaCl concentrations (**Figure S7.1A**). All the tested nucleases except AsCas12a were highly active between a wide range of pH 6 to pH 9. AsCas12a displayed reduced activity in buffer pH above ~7.5 (**Figure S7.1A**). SpCas9 and AsCas12a were tolerant to a wide range of salt concentrations, from 50 to 300 mM NaCl and 50-500 mM NaCl, respectively (**Figure S7.1B**). On the other hand, LbCas12a and FnCas12a lost their activity at NaCl concentrations greater than 150mM NaCl, and EeCas12a was not active at NaCl concentrations over 300 mM (**Figure S7.1B**).

FnCas12a expression is not toxic for *B. smithii*

The thermal and pH stability of FnCas12a and EeCas12a motivated us to test the utility of these nucleases as genome editing tools for facultative thermophilic bacteria. For the assessment of FnCas12a as genome editing tool for thermophiles, we employed *B. smithii* ET 138 cultured at 55 ℃. *B. smithii* ET 138 is a thermophilic and facultative anaerobe that can utilize a wide range of substrates which combined with its genetic amenability makes it an ideal platform organism for use in a bio-refinery (*318*, *326*).

Initially, to discount any potential toxicity of FnCas12a in the sporulation- and restriction-deficient *B. smithii* ET 138 Δ*sigF* Δ*hsdR* cells, we designed a single plasmid approach and constructed a non-targeting (NT) pFnCas12a_NT plasmid (**Figure 7.2A**) by cloning *Francisella novicida* U112 *cas12a* (*fncas12a*) gene and a guide RNA (gRNA) comprised of 36 nucleotides (nt) repeat and a 25 nt spacer sequence in the *Bacillus - E. coli* shuttle vector pNW33n. The target sequence is flanked by a non-canonical PAM (5'-CTCG), therefore, FnCas12a should not be able to target the *B. smithii* genome. The expression of the *fncas12a* gene and the gRNA is under the control of the inducible promoter of *xynA* gene (P_{xynA}) from *Thermoanaerobacterium saccharolyticum* and the phosphotransacetylase *(pta)* promoter (P_{pta}) from *Bacillus coagulans* (*328*, *360*), respectively. During all the culturing experiments in this study, both xylose and glucose were added to the LB2 media ($LB2_{xg}$) to ensure constant induction of FnCas12a expression from the P_{xynA} promoter.

The non-targeting pFnCas12a_NT plasmid and an empty pNW33n control plasmid were individually electroporated into a single batch of competent *B. smithii* cells (**Figure 7.2B**). The transformed cells were plated on pre-warmed LB2 plates supplemented with chloramphenicol and incubated overnight at 55 ℃ (**Figure 7.2B**). The next day, four single colonies per transformation were used for sequential transfers in fresh $LB2_{xg}$ liquid media from 55 ℃ to 37 ℃ every 24 hours, with an intermediate step at 45 ℃ (**Figure 7.2B**). No significant difference in growth was observed between the pFnCas12a_NT cultures and the pNW33n control cultures (**Figure 7.2C**), indicating that the expression of FnCas12a is not toxic for the *B. smithii* cells at any of the above tested temperatures.

FnCas12a is capable of targeted DNA cleavage in *B. smithii* genome

It has been repeatedly reported that the double-strand breaks (DSB) induced by Cas nucleases are lethal to bacterial cells because many microorganisms lack an effective and functional endogenous non-homologous end joining (NHEJ) mechanism. Consequently, CRISPR–Cas is usually used as a lethality based counter-selection tool (*72*, *305*, *330*, *361*) in bacterial cells to get rid of the cells that have not performed the desired HR prior to Cas9 cleavage (*72*, *305*, *330*, *361*). To enable the use of HR-Cas12a mediated genome editing, we first had to evaluate the lethality of Cas12a-induced DSB in the genome of *B. smithii.*

To this end, we designed and constructed three targeting plasmids, pFnCas12a_Sp1-3, each

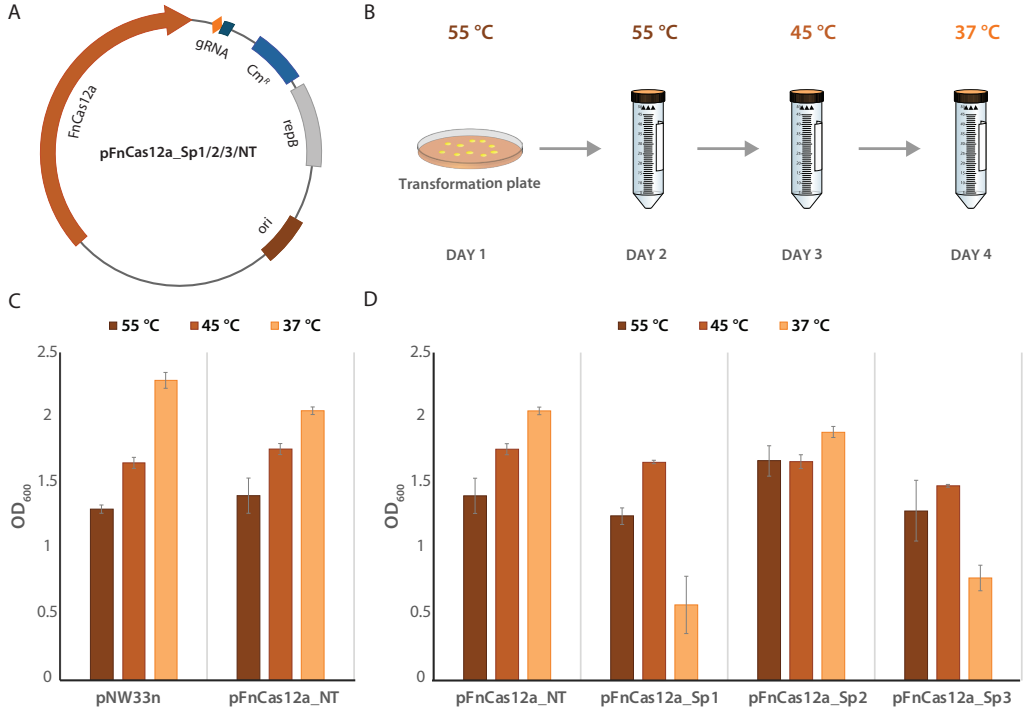


Figure 7.2 | Transformation of FnCas12a mediated targeting plasmids in *B. smithii* ET 138.
(A) Schematic representation of the basic pFnCas12a plasmid. The *fncas12a* gene was introduced in the pNW33n vector backbone. A gRNA-expressing module was introduced downstream the *fncas12a* gene. Other elements on the plasmid are: origin of replication (*ori*), replication protein (*repB*), chloramphenicol-resistance marker (CmR). **(B)** Sequential transfer scheme of wild-type *B. smithii* cultures to evaluate FnCas12a expression and targeting efficiency at different temperatures; detailed description of the protocol can be found in the Material and Methods section. **(C)** Results of the targeting experiment showing OD_{600} measurements from cultures of *B. smithii* transformed with pFnCas12a_NT compared to an empty pNW33n plasmid. **(D)** Results of the targeting experiment showing OD_{600} measurements from cultures of *B. smithii* transformed with the targeting plasmids harbouring different spacer (pFnCas12a_Sp1-3) compared to the non-targeting plasmid (pFnCas12a_NT). Average values of four replicates are shown, with error bars representing standard deviations.

containing gRNA with a unique spacer [Spacer 1 (Sp1), Spacer 2 (Sp2) and Spacer 3 (Sp3)] corresponding to a different target sequence (**Table S7.3**) within *pyrF* gene in the *B. smithii* genome. The *pyrF* gene encodes orotidine 5'-monophosphate (OMP) decarboxylase, which is required for pyrimidine biosynthesis. This enzyme is also able to convert the uracil analogue 5-fluorootic acid (5-FOA) to the toxic product 5-fluorodeoxyuridine 5'-monophosphate (*362*). Therefore, Δ*pyrF* strains are uracil auxotrophs and resistant to 5-FOA (*363*, *364*). It is expected that if FnCas12a is functionally expressed in *B. smithii*, the growth of the cultures harbouring the targeting plasmids will be substantially lower compared to growth of the cultures harbouring the control (pFnCas12a_NT and pNW33n) plasmids.

The three targeting plasmids and the control plasmids were individually electroporated into a single batch of *B. smithii* competent cells. Four single colonies per transformation were subjected to sequential transfers at different temperatures as described before. No striking differences in growth were observed for the control cultures incubated at 55 ℃, 45 ℃ and 37 ℃ (**Figure 7.2D**). Conversely, the growth of the cultures harbouring the pFnCas12a_Sp1

and pFnCas12a_Sp3 plasmids was significantly inhibited compared to cultures harbouring the pFnCas12a_NT plasmid, suggesting that FnCas12a is active at 37 ℃ and can successfully target the *B. smithii* genome (**Figure 7.2D**). On the other hand, no difference in growth for the cultures harbouring the pFnCas12a_Sp2 plasmid was observed (**Figure 7.2D**). This observed fluctuation is most likely due to the reported guide-dependent variation in CRISPR–Cas12a target efficiency (*365*, *366*), similar to that of the CRISPR–Cas9 system (*367*, *368*).

In line with the *in vitro* assay data, the aforementioned results indicate that the FnCas12a is active at 37 ℃, but is inactive and thus does not introduce lethal DSBs to the *B. smithii* genome at temperatures above 45 ℃. To explore the temperature range that can be used for *in vivo* targeting using FnCas12a, four single colonies harbouring the pFnCas12a_Sp1 and pFnCas12a_Sp3 plasmids were grown overnight in $LB2_{xg}$ medium and incubated at temperatures ranging from 37 °C – 42 °C. The most optimal temperature for *in vivo* targeting using FnCas12a appeared to be 37-38 °C, although targeting also occurred at temperatures till 42 °C (**Figure S7.2**). The targeting efficiency was higher for cells harbouring the pFnCas12a_Sp3 plasmid as compared to that of the cells harbouring pFnCas12a_Sp1 plasmid at temperatures 40-42 °C. Nevertheless, the growth of both cultures harbouring pFnCas12a_Sp1 and pFnCas12a_Sp3 were prominently retarded compared to the growth of the cultures harbouring the pFnCas12a_NT plasmid (**Figure S7.2**).

Marker-less chromosomal gene deletion using FnCas12a and plasmid-borne editing template

HR-Cas12a mediated genome editing requires three parts, i) a double stranded DNA fragment [(homologous flanks (HR flanks)] that can undergo HR for editing the genome. (ii) the Cas12a endonuclease, and (iii) the gRNA to lead Cas12a to the target DNA site to select (enrich) mutations introduced through HR by counter-selecting (eliminating) wild-type clones. Based on our previous research (*315*) and by employing the here-developed FnCas12a targeting system, we set out to develop a HR-FnCas12a based counter-selection system for efficient genome editing in *B. smithii.* A sequential culturing process of first equipping the cells with an appropriate plasmid-borne editing template for HR at 55 ℃, and then inducing the expression of active FnCas12a at 37 ℃, is expected to lead to counter-selection against the wild-type genotype.

As a proof of principle, we programmed the designed HR-FnCas12a system for deletion of the *pyrF* gene in *B. smithii.* Two editing plasmids (pFnCas12a_Δ*pyrF*-HR_Sp1 and pFnCas12a _Δ*pyrF*-HR_Sp3) were constructed, both containing the previously described spacers 1 and 3 for efficient *pyrF* targeting together with a 2-kb HR template encompassing 1-kb of the up- and down-stream genomic regions flanking the selected *pyrF* gene fragment (**Figure 7.3A**). A non-targeting plasmid (pFnCas12a_Δ*pyrF*-HR_NT) containing the same HR template as the editing plasmids but with a non-targeting spacer and (non) targeting plasmids without the HR template were taken as a control for recombination.

After transforming the constructs into *B. smithii* ET 138 Δ*sigF* Δ*hsdR* at 55 ℃, one verified transformant per construct was used to inoculate $LB2_{xg}$ and $TVMY_{xg}$ selection medium supplemented with uracil ($LB2_{xgu}$ and $TVMY_{xgu}$) to complement the auxotrophy in case of successful *pyrF* gene deletion. 5-FOA was not added to the medium in order to assess Cas12a-mediated counter-selection efficiency. After growth at 55 ℃ for 24 hours, cells were transferred stepwise every 24 hours to fresh media while gradually lowering the culturing temperature from 55 ℃ to 37 ℃, with an intermediate transfer at 45 ℃ to allow for gradual adjustment of the cellular metabolism (**Figure 7.3B**).

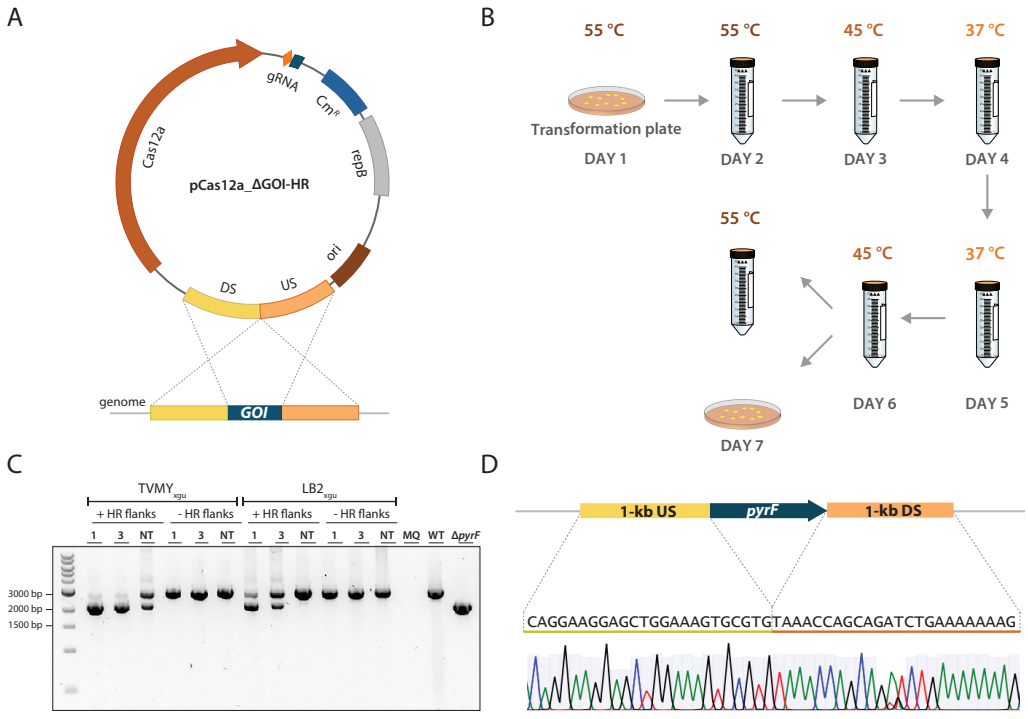


Figure 7.3 | Transformation of HR-FnCas12a mediated editing plasmids in *B. smithii* ET 138.
(A) Schematic representation of the basic pFnCas12a_Δgene-of-interest(goi)-HR construct. The *fncas12a* gene was introduced to the pNW33n vector backbone. Homologous recombination (HR) flanks were introduced upstream *fncas12a* gene and encompassed the 1-kb upstream and 1-kb downstream region of the gene of interest (goi) in the *B. smithii* genome. A gRNA-expressing module was introduced downstream the *fncas12a* gene. Other elements on the plasmid are: origin of replication (*ori*), replication protein (*repB*), chloramphenicol-resistance marker (CmR). **(B)** Sequential transfer scheme of wild-type *B. smithii* cultures to evaluate FnCas12a editing efficiency; detailed description of the protocol can be found in the Material and Methods section. **(C)** Agarose gel electrophoresis showing the results from PCR on the genomic DNA of *B. smithii* cultures transformed with pFnCas12a_Δ*pyrF*-HR_Sp1 (1), pFnCas12a_Δ*pyrF*-HR_Sp3 (3) and pFnCas12a_Δ*pyrF*-HR_NT (NT) in two different selection media ($TVMY_{xgu}$ and $LB2_{xgu}$). The last 2 lanes are the negative (wild type) and positive (Δ*pyrF*) controls, that correspond to 2.9 kb and 2.2 kb long DNA fragments respectively. **(D)** Representative image of the sequence verification of the desired *pyrF* gene deletion by sanger sequencing.

After transforming the constructs into *B. smithii* ET 138 Δ*sigF* Δ*hsdR* at 55 ℃, one verified transformant per construct was used to inoculate $LB2_{xg}$ and $TVMY_{xg}$ selection medium supplemented with uracil ($LB2_{xgu}$ and $TVMY_{xgu}$) to complement the auxotrophy in case of successful *pyrF* gene deletion. 5-FOA was not added to the medium in order to assess Cas12a-mediated counter-selection efficiency. After growth at 55 ℃ for 24 hours, cells were transferred stepwise every 24 hours to fresh media while gradually lowering the culturing temperature from 55 ℃ to 37 ℃, with an intermediate transfer at 45 ℃ to allow for gradual adjustment of the cellular metabolism (**Figure 7.3B**).

As expected, PCR analysis on the genomic DNA extracted from all the cultures harbouring the pFnCas12a_Sp1/Sp3/NT lacking the homologous recombination flanks showed the absence of any *pyrF* knockout bands in these cells (**Figure 7.3B**). Despite the lack of counter-selective pressure, population PCR analysis on the genomic DNA extracted from the $TVMY_{xgu}$ liquid

cultures of cells harbouring the pFnCas12a_Δ*pyrF*-HR_NT plasmid after incubation showed the presence of both wild type and mutant bands (**Figure 7.3C**). Remarkably, population PCR analysis on the genomic DNA extracted from the $TVMY_{xgu}$ liquid cultures harbouring the pFnCas12a_Δ*pyrF*-HR_Sp1 and pFnCas12a_Δ*pyrF*-HR_Sp3 editing plasmids, at least in some cases showed the presence of only the mutant amplicons (**Figure 7.3C**). PCR on genomic DNA extracted from the LB2 liquid medium cultures harbouring the pFnCas12a_Δ*pyrF*-HR_NT plasmid did not show the presence of any *pyrF* knockout bands, as was to be expected (**Figure 7.3C**). On the other hand, PCR on genomic DNA extracted from the cultures harbouring the pFnCas12a_Δ*pyrF*-HR_Sp1 or pFnCas12a_Δ*pyrF*-HR_Sp3 editing plasmids showed the presence of mixed genotype bands (**Figure 7.3C**). The striking difference in the density of the knockout bands between the cultures harbouring the targeting and the non-targeting plasmids underlines the significant contribution of FnCas12a to the efficiency of the genome editing tool.

The abovementioned cultures were plated on corresponding media and incubated overnight at 55 °C to screen for clean deletion mutants. We genotyped 10 random colonies from each plate using colony PCR using primers specific to the *pyrF* genomic region to confirm the desired clean deletions. 9 out of the 10 and 6 out of the 10 tested pFnCas12a_Δ*pyrF*-HR_Sp1 and pFnCas12a_Δ*pyrF*-HR_Sp3 harbouring colonies from the $TVMY_{xgu}$ medium plates, respectively, were clean *pyrF* gene deletion mutants (**Figure S7.3**). 7 out of the 10 and 8 out of the 10 tested pFnCas12a_Δ*pyrF*-HR_Sp1 and pFnCas12a_Δ*pyrF*-HR_Sp3 harbouring colonies from the $LB2_{xgu}$ medium plates, respectively, were clean *pyrF* gene deletion mutants (**Figure S7.3**). As expected, editing efficiencies were significantly lower in the $TVMY_{xgu}$ cultures harbouring the pFnCas12a_Δ*pyrF*-HR_NT plasmid. 1 and 2 out of 10 tested colonies were found to be of Δ*pyrF* and mixed genotype, respectively. The gene deletion was verified by sequencing using specific primers on the genomic DNA isolated from 4 representative colonies (**Figure 7.3D**). The phenotype of three representative Δ*pyrF* mutant colonies was further verified by their ability to grow on plates supplemented with 5-FOA.

To effectively use the HR-FnCas12a tool in *B. smithii* for iterative genome engineering, a plasmid-free strain is desired for gene editing cycles with multiple vectors. For this purpose, a sequence-verified Δ*pyrF* mutant was grown in $TVMY_{xgu}$ medium without antibiotic for three iterative inoculation cycles of 8-16 hours each. The final culture was plated on $TVMY_{xgu}$ plates with and without antibiotic (chloramphenicol). Colony PCR analysis with plasmid-specific primers on 6 out of the 7 tested colonies confirmed loss of the plasmid (**Figure S7.4**). The 6 plasmid-cured *B. smithii* ET 138 Δ*sigF* Δ*hsdR* Δ*pyrF* colonies were also verified to be uracil auxotroph and 5-FOA resistant.

Next, we aimed to further optimize the efficiency of the system, to reduce the number of culturing steps and the total time for the editing process. To this end, a small amount from the glycerol stock of *B. smithii* cells harbouring the pFnCas12a_Δ*pyrF*-HR_Sp1 plasmid was plated on a $LB2_{xgu}$ plate and incubated at 45 ℃ overnight. Single colonies from the plate were used to inoculate $LB2_{xgu}$ and $TVMY_{xgu}$ liquid media and incubated overnight at 37 ℃. The following day, the cultures were plated on corresponding media and incubated overnight at 45 ℃. We genotyped 10 random colonies that presumably survived FnCas12a targeting from each plate, through colony PCR and sequencing. Interestingly, all the tested colonies from the $TVMY_{xgu}$ plate using were Δ*pyrF* mutants (**Figure S7.5**). As always, the Δ*pyrF* mutants were verified by sequencing and for their ability to grow on $TVMY_{xgu}$ plates supplemented with 5-FOA. Interestingly, the total time required from the first transformation to successfully obtaining a *B.*

smithii mutant using the here developed system is as short as 3 days. On the contrary, none of the colonies from the LB2 plate were positive for the Δ*pyrF* genotype. This variance in the use of the two ($LB2_{xgu}$ and $TVMY_{xgu}$) culture medium on the efficiency of the editing system has also been described previously (*315*), although the basis for this phenomenon remains elusive and requires further research.

EeCas12a is active *in vivo* at elevated temperatures

The thermal stability of EeCas12a based on the *in vitro* characterisation of the different Cas12a variants (**Figure 7.1**), motivated us to test it as an alternative for genome editing in our moderate thermophile. Perhaps, a variant with an optimal activity at a higher temperature could be exploited to shorten the time required for editing the genome of *B. smithii* using the aforementioned FnCas12a-HR system. Additionally, it could constitute the basis of an alternative genome editing system for thermophilic bacteria that do not grow at temperatures below 45 ℃.

To evaluate the *in vivo* targeting efficiency of EeCas12a in *B. smithii*, we constructed the pEeCas12a_Sp1 and pEeCas12a_NT plasmids (**Figure 7.4A**) by substituting the *fncas12a* gene by *eecas12a* gene in the previously constructed pFnCas12a_Sp1 and pFnCas12a_NT plasmids, respectively. Since the PAM sequence recognized by the EeCas12a is similar to that of FnCas12a (5'-TTTN -3'), we did not the change the (proto) spacer sequence. The direct repeat sequences for FnCas12a and EeCas12a show strong conservation in the 19 nucleotides at the 3′ of the direct repeat, which is the portion recognized by the nuclease (*83*) (**Figure S7.6**). Hence, we used the orthologous FnCas12a direct repeat for the gRNA on the pEeCas12a plasmids.

The targeting pEeCas12a_Sp1, the non-targeting pEeCas12a_NT and the empty pNW33n control plasmids were individually electroporated into a single batch of competent *B. smithii* cells. Each transformation mix was equally distributed and plated on four pre-warmed LB2 plates supplemented with chloramphenicol. The plates were incubated overnight at different temperatures: 45 ℃, 50 ℃, 55 ℃ and 60 ℃. While a large number of colonies was obtained on the plate containing the cells transformed with the pEeCas12a_NT plasmid at 45 ℃, no colonies were obtained on the plate containing the cells transformed with the pEeCas12a_Sp1 plasmid (**Figure S7.7**), suggesting high targeting activity of EeCas12a at 45 ℃. Minor differences were observed in the transformation efficiency for the cells transformed with the targeting and non-targeting plasmids at 50 ℃ and no difference was observed in the transformation efficiency at 55 and 60 ℃ (**Figure S7.7**).

To confirm targeting by EeCas12a, a small amount of the glycerol stock of the cells harbouring the pEeCas12a_Sp1 and pEeCas12a_NT plasmids was used to inoculate $LB2_{xgu}$ selection medium liquid and incubated overnight at 55 ℃, 50 ℃ and 45 ℃. As seen previously, no growth was observed for the cultures harbouring the targeting plasmid as compared to the non-targeting plasmids at 45 ℃ (**Figure 7.4B**). On the contrary, no significant growth differences were observed for the cultures harbouring the plasmids at 50 ℃ and 55 ℃ (**Figure 7.4B**). This prompted us to explore the temperature window that supports *in vivo* targeting activity of EeCas12a. Four single colonies of each transformant harbouring either pEeCas12a_Sp1 or pEeCas12a_NT were used to inoculate $LB2_{xgu}$ selection medium and were incubated at different temperatures ranging from 41 ℃ to 48 ℃. After 18-24 hours, the growth of the cultures was measured and normalized using the corresponding control cultures harbouring the pEeCas12a_NT plasmid (**Figure 7.4C**). Effective targeting of the genome by EeCas12a was observed up to 46 ℃ (**Figure 7.4C**).

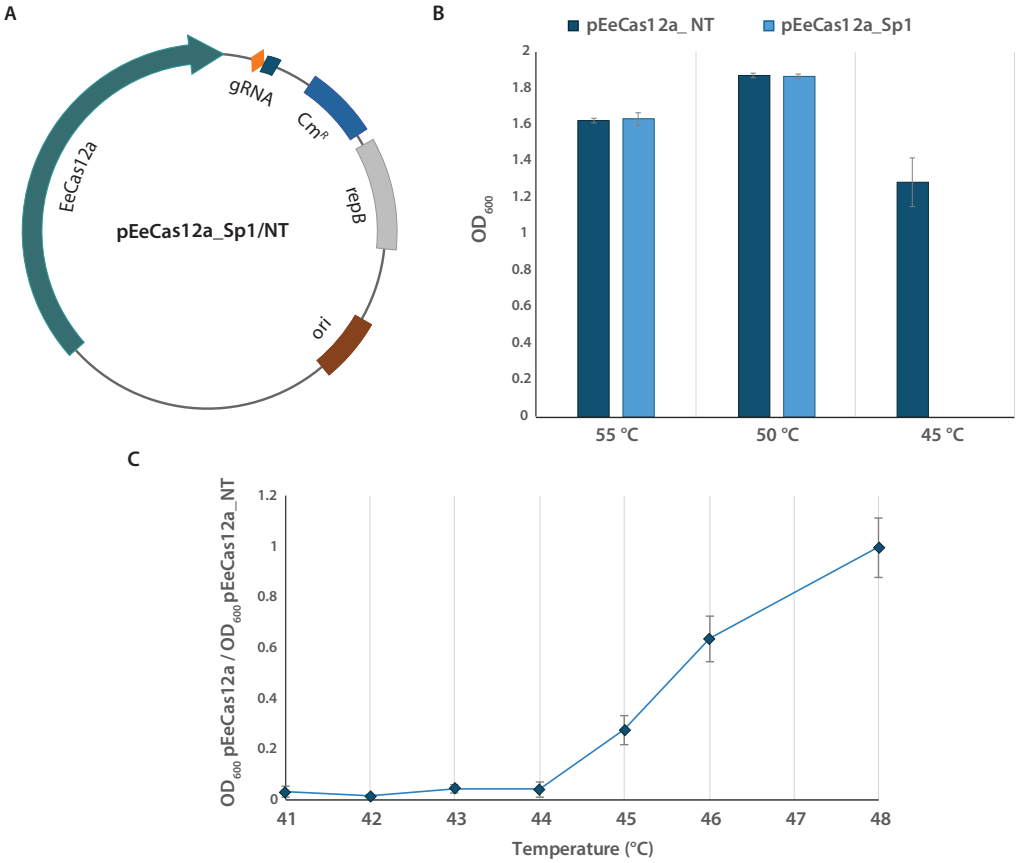


Figure 7.4 | Transformation of HR-EeCas12a mediated editing plasmids in *B. smithii* ET 138.
(A) Schematic representation of the basic pEeCas12a plasmid. The *eecas12a* gene was introduced in the pNW33n vector backbone. A gRNA-expressing module was introduced downstream the *eecas12a* gene. Other elements on the plasmid are: origin of replication (*ori*), replication protein (*repB*), chloramphenicol-resistance marker (CmR). **(B)** Results of the targeting experiment showing OD_{600} measurements from cultures of *B. smithii* transformed with pEeCas12a_Sp1 normalized to a non-targeting pEeCas12a_NT plasmid after sequential transfer at 45 ˚C, 50 ˚C and 55 ˚C. **(C)** Results of the targeting experiment showing OD_{600} measurements from cultures of *B. smithii* transformed with pEeCas12a_Sp1 compared to a non-targeting pEeCas12a_NT and grown between 41-48 ˚C to evaluate the optimal temperature for *in vivo* activity. Average values of four replicates are shown, with error bars representing standard deviations.

EeCas12a can be utilized for fast and efficient *in vivo* genome editing

To exploit the efficient targeting by EeCas12a, we explored the use of HR coupled with EeCas12a-based counter-selection for editing *B. smithii* genome. An editing plasmid (pEeCas12a_Δ*pyrF*-HR_Sp1) was constructed containing the previously tested spacer 1, that exhibited the highest targeting efficiency for the FnCas12a-based system, together the two 1-kb HR flanks (**Figure 7.5A**). A non-targeting plasmid (pEeCas12a_Δ*pyrF*-HR_NT) containing a non-targeting spacer but with the same HR flanks as the editing plasmid was taken as a control for recombination. The two plasmids were individually electroporated into ET 138 Δ*sigF* Δ*hsdR*, plated on LB2 selection medium plates and incubated at 45 ℃ and 55 ℃. The number of colonies on the plates incubated at 45 ℃ were very low, most likely due to strong targeting by EeCas12a at these temperatures. This is in line with the previous observations

(*329*, *330*) that the efficiency of the approach to use Cas nucleases as a counter-selection tool is higher compared to using them as a trigger for induction of the cellular HR mechanism after introduction of targeted double stranded breaks (DSBs). Nevertheless, 5 out of 6 tested colonies from the plate incubated at 45 ° C assessed by colony PCR showed a mixed genotype (**Figure 7.5B**). Re-streaking these colonies on fresh media plates could lead to obtaining clean deletion mutants, thereby simplifying the process to just transformation and plating to obtain mutants.

Furthermore, two single colonies from the plate incubated at 55 ℃ were used to inoculate $LB2_{xgu}$ and $TVMY_{xgu}$ selection medium. The cultures were incubated for 18-24 hours at 43 ℃ and transferred back to 55 ℃. The next day, population PCR using specific primers on the genomic DNA extracted from the cultures grown on both the media indicated an apparent editing efficiency >50% (**Figure 7.5C**).

Next, to enhance the emergence of clean deletion mutants, we repeated the previous transformations and incubated the selection media plates at 43 ℃ and 45 ℃. Again, a low number of colonies were obtained on these plates. All the colonies were genotyped using colony PCR; only wild-type genotype colonies and 1 colony of mixed genotype were found on the plates incubated at 43 ℃ and 45 ℃, respectively (**Figure S7.8**). A random single colony from the transformation plate was used to inoculate the corresponding liquid media and incubated at the respective temperature overnight. The following day, genomic DNA was isolated from the cultures and subjected to population PCR using genome specific-primers. Strikingly, a single knock-out band was obtained, indicating that the majority of the population consisted of *pyrF* mutants (**Figure 7.5D**). The cultures were plated on respective selective media and single colonies were subjected to colony PCR. All the tested colonies were Δ*pyrF* mutants (**Figure S7.9**). Finally, the phenotype of the mutants was verified by their ability to grow on plates containing 5-FOA. In this way we have shortened the process time by using a single culturing step at 43 ℃ or 45 ℃ to obtain *B. smithii* mutants using EeCas12a. The established HR-EeCas12a system is an efficient and fast method that expands the genome editing toolbox for moderate thermophiles.

Discussion

Our work on the temperature- and pH- stability and salt concentration tolerance of Cas12a variants expands our knowledge on the biochemical conditions at which Cas12a can be used. This underscores the significance of mining genome- and metagenome databases for unique CRISPR–Cas systems to uncovering novel protein variants with valuable characteristics. We show that the thermostability of all the tested nucleases strongly depends on the association with an appropriate guide RNA. Although AsCas12a, LbCas12a and FnCas12a have been extensively characterised and used (*83*, *307*, *356*), this is the first time that the cleavage activity of EeCas12a nuclease has been characterised. We anticipate that FnCas12a and EeCas12a will be useful for *in vitro* molecular biology applications requiring targeted cleavage either at elevated temperatures, or under other harsh conditions.

Similar to Cas9 proteins from Class 2 Type II CRISPR systems, Cas12a proteins from Class 2 Type V-A CRISPR systems have been used for genome engineering in different bacterial species (*306*, *307*, *356*, *369*, *370*). Thermophiles have various advantages as production hosts for bio-based industries over current mesophilic model organisms, however their limited genetic toolboxes often hamper their development into efficient cell factories. Previously, SpCas9- and

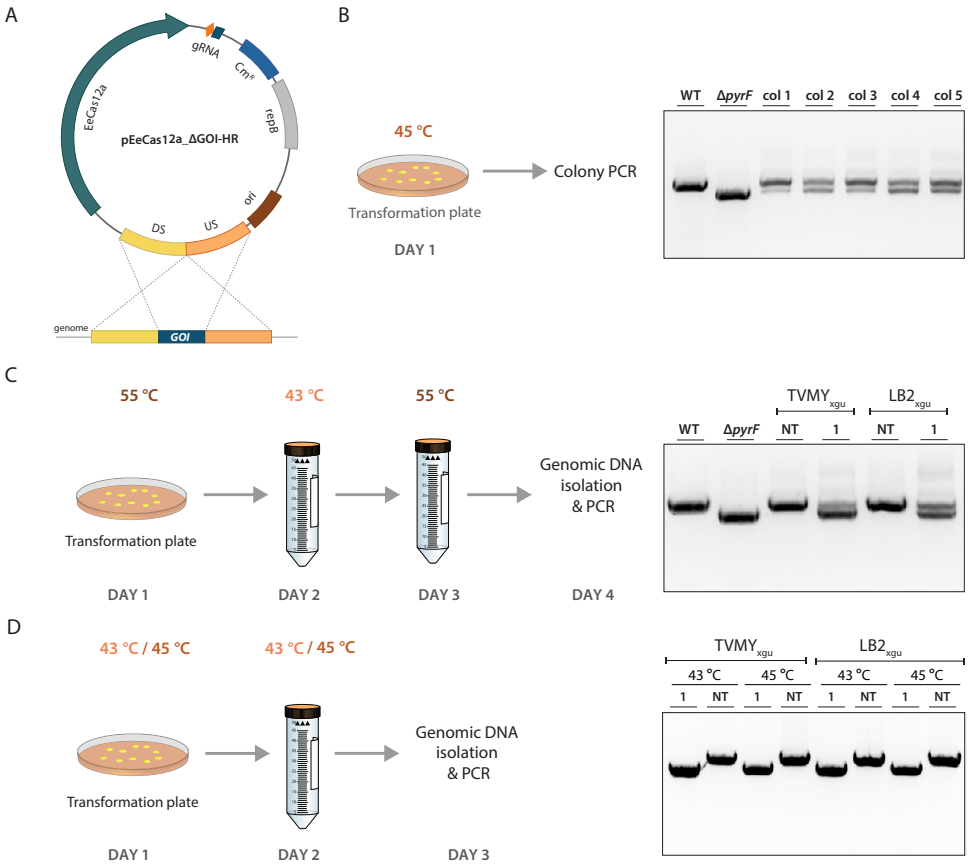


Figure 7.5 | Transformation of HR-EeCas12a mediated editing plasmids in *B. smithii* ET 138.
(A) Schematic representation of the basic pEeCas12a_Δgene-of-interest (goi) construct. The *eecas12*gene was introduced in the pNW33n vector backbone. Homologous recombination (HR) flanks were introduced upstream *eecas12*gene and encompassed the 1-kb upstream and 1-kb downstream region of the gene of interest (goi) in the *B. smithii* genome. A gRNA-expressing module was introduced downstream the *eecas12a* gene. Other elements on the plasmid are: origin of replication (ori), replication protein (repB), chloramphenicol-resistance marker (CmR). **(B)** Schematic representation of the experiment and agarose gel electrophoresis showing the results from PCR on the colonies (col) from *B. smithii*cells transformed with pEeCas12a_Δ*pyrF*-HR_Sp1. The first 2 lanes are the negative (wild type) and positive (Δ*pyrF*) controls, that correspond to 2.9 kb and 2.2 kb long DNA fragments respectively. **(C)** Schematic representation of the experiment and agarose gel electrophoresis showing the results from PCR on the genomic DNA of a *B. smithii*cultures harbouring pEeCas12a_Δ*pyrF*-HR_Sp1 (1) as compared with cultures harbouring pEeCas12a_Δ*pyrF*-HR_NT (NT) in two different selection media. The first 2 lanes are the negative (wild type) and positive (Δ*pyrF*) controls, that correspond to 2.9 kb and 2.2 kb long DNA fragments respectively. **(D)** Schematic representation of the experiment and agarose gel electrophoresis showing the results from PCR on the genomic DNA of a *B. smithii*cultures harbouring pEeCas12a_Δ*pyrF*-HR_Sp1 (1) as compared with cultures harbouring pEeCas12a_Δ*pyrF*-HR_NT (NT) at 45 ˚C and 43 ˚C in two different selection media.

thermostable ThermoCas9-based systems were developed into a Cas9-based engineering tool for thermophilic bacteria (*315*, *329*). So far, all genome engineering applications in thermophilic microorganisms have been based on Cas9.

A Cas12a-based genome editing tool is a complementary addition to the CRISPR-based bacterial genome engineering toolbox and offers distinct advantages over using Cas9-based system.

Firstly, Cas12a recognizes TTTN PAM sequences, whereas Cas9 recognizes a NGG PAM sequence (*291*). This increases the number of target sites within a genome, allowing for precise selection of the cleavage site. This is particularly handy for generation of point mutations and when targeting short intergenic regions. Secondly, in some organisms, inherent toxicity to SpCas9 has been reported, in such cases Cas12a is an excellent alternative (*306*). Finally, the Cas12a cleaves the target sequence between 18 and 23 position distal to the PAM sequence. Due to this, any indel resulting from NHEJ would not disrupt the PAM, and the resulting sequence could be cleaved over-and-over again, increasing the likelihood of recovering homologous recombinants over recovering NHEJ-mediated mutants. While a NHEJ system is generally not present or not active in most prokaryotes (*371*), the *B. smithii* ET 138 genome contains the DNA ligase LigD and DNA-end-binding protein Ku genes composing the basic prokaryotic NHEJ-like system (*372*, *373*). However, we did not detect any evidence for NHEJ activity in *B. smithii* using Cas12a-based recombination. It could be that the NHEJ system in *B. smithii* ET 138 either is inactive at 37 ℃ or it is not active enough to counteract the Cas12a activity.

In this study, we developed a single-plasmid system coupling homologous recombination with Cas12a-based counter selection to allow for efficient gene deletion in an industrially relevant, moderate thermophile, *B. smithii* ET 138. Based on the *in vitro* temperature stability and activity of the tested Cas12a variants, we selected two Cas12a variants for genome editing: FnCas12a and EeCas12a. Accordingly, we developed an editing system where mutants are generated via homologous recombination events at elevated temperatures (>45 ℃) before Cas12-induced counter-selection takes place at 37 ℃ for FnCas12a and 43-45 ℃ for EeCas12a. Thus, the complementary temperature range of FnCas12a and EeCas12a opens up Cas12a-based genome editing for moderate thermophiles. Although, a CRISPRi system using nuclease-deactivated Cas12a from *Eubacterium eligens* (EedCas12a) for gene regulation in *E. coli* has been previously reported (*374*), to the best our knowledge, this is first time EeCas12a has been used for bacterial genome editing.

In the course of our study we observed spacer sequence- and media- dependent differences in targeting by FnCas12a. While the spacer sequence dependency could be attributed to sequence and structural features such as GC content, minimum free energy, melting temperature and position-specific nucleotide composition (*365*), no evident link between gene editing and the medium used for the editing process could be established, and thus this should be addressed in future studies.

Furthermore, the HR-Cas12a system eliminates the need of an antibiotic gene as a selection marker or the presence of chromosome scar sites (*e.g.*, LoxP recognition sites). It is substantially time-effective compared to the previously established lacZ-based and SpCas9-based counter-selection systems (*315*, *328*). It only takes 3-4 days from transformation to generating a clean deletion mutant, including the plasmid curing step, while maintaining a high efficiency. The here developed HR-Cas12a system could be further simplified by exploiting the pre-crRNA processing activity of Cas12a for more rapid and facile multiplexing. A single CRISPR array containing multiple spacers under the transcriptional regulation of a single promoter and terminator can be used to target multiple loci simultaneously (*256*). Consequently, there is no need for supplying multiple targeting expression plasmids for which construction procedure is usually time-consuming and requiring a sophisticated genetic toolbox (*375*). Cas12a has previously been used for multiplexing in *E. coli and Streptomyces coelicolor* for gene insertions and deletions, respectively (*356*, *376*, *377*).

In conclusion, we here demonstrated that FnCas12a and EeCas12a endonucleases are active and can target the genome of the biotechnologically interesting bacterium *B. smithii* ET 138. We further combined the Cas12a targeting activity with the native homologous recombination (HR) system of *B. smithii* for the development of efficient gene editing tools. In addition, these tools are potentially applicable to various model or non-model organisms with an active HR pathway and a growth temperature range covering 37 ℃ to over 45 ℃ expanding the range of organisms in which CRISPR-based editing can be applied and, thereby, expanding the range of possible cell factories for improved production of interesting compounds.

Experimental Procedures

Protein expression

FnCas12a, AsCas12a and EeCas12a were expressed and purified as previously described (*378*). The FnCas12a, AsCas12a and EeCas12a expression constructs were synthesized by GenScript. EeCas12a is fused to an N-terminal tag comprising a hexahistidine (6xHis-tag) sequence and SV40 NLS, and C-terminal SV40 NLS, and FnCas12a and AsCas12a with a N-terminal SV40 NLS, and C-terminal SV40 NLS and 6xHis-tag. Proteins were expressed in NiCo21 (DE3) *Escherichia coli* cells (NEB #C2529) containing the respective expression plasmid by growing in LB-Kan (40 ug mL^{-1}) at 37 ℃ followed by inducing at 23 ℃ for 16 hours in presence of IPTG at 0.4 mM. Cells were lysed using sonication prior to chromatographic purification. Recombinant Cas12a proteins were purified using HiTrap DEAE FF, HisTrap HP (Ni-NTA) and HiTrap Heparin HP columns from GE Healthcare (Pittsburgh, PA). The purified proteins were dialyzed and concentrated into 20 mM Tris-HCl (pH 7.4), 500 mM NaCl, 1 mM DTT, 0.1 mM EDTA and 50% glycerol.

In vitro activity assays

Buffers and chemicals were obtained from NEB, VWR (Radnor, PA) and Sigma-Aldrich (St. Louis, MA). NEBuffers were from New England Biolabs Inc (Ipswich MA). OptiMEM cell culture medium was obtained from Thermofisher (Waltham, MA). The oligonucleotides (IDT, Coralville, IA) and the plasmids used for the enzyme activity assays are listed in **Table S7.1** and **Table S7.2**, respectively. The 515 bp linear DNA substrates were amplified (using NEB1 and NEB2) by PCR using Q5 2X master mix (NEB #M0492S) from the pMiniT-WTAP Exon 8 plasmid which contains a fragment of the human WTAP gene. The substrate contains the target sequence 5'-TTTC- CCACTCACTGCTTTCTCCTC-3', where TTTN is a generic PAM sequence for Cas12a. PCR products were purified using Monarch PCR & DNA Cleanup Kit (NEB #T1030S). Guide RNA for Cas12a was synthesized (NEB3) and the guide for Cas9 was generated using the EnGen sgRNA synthesis kit *S. pyogenes* (NEB #E3322S) with the template for top strand (NEB4) designed using the NEB Bio Calculator. Recombinant SpCas9 and LbaCas12a were obtained from NEB.

Activity of the Cas12a RNPs was determined by monitoring the dsDNA cleavage of 5'-FAM labelled PCR products. DNA containing a target site was digested at varying temperatures as indicated using a ratio of 20:10:1, crRNA:Cas12a:target for 10 minutes in T100 Thermal Cycler (Bio-Rad, Hercules CA). Reactions were quenched and then subjected to capillary electrophoresis on an Applied Biosystems 3730xl Genetic Analyzer instrument (*379*). The cleavage products were quantified using Peakscanner (ThermoFisher).

Thermal stability of Cas12a proteins

The thermal stability of the nucleases was determined in the presence and absence of their respective gRNAs using a Prometheus nano differential scanning fluorimetry (NanoDSF) instrument (NanoTemper Technologies, München, Germany). 200 pmol of Cas12a protein was mixed with either 400 pmol of gRNA or with water in a mixture that contained 1x NEBuffer2.1 (10 mM Tris-HCl (pH 7.9) 50 mM NaCl, 10 mM $MgCl_2$, 100 ug mL^{-1} BSA), 1x NEBuffer3.1 (50 mM Tris-HCl (pH 7.9), 100 mM NaCl, 10 mM $MgCl_2$, 100 ug mL^{-1} BSA), 1x AspCas12a Buffer (20mM HEPES (pH 6.5), 50 mM NaCl, 10 mM $MgCl_2$) or Opti-MEM by ThermoFisher. Unfolding of the protein or complex was monitored according to the manufacturer's instructions and data are reported as the inflection point of the melting curve of the protein or RNP complex.

Strains and growth media

Bacterial strains used in the current study are *E. coli* DH10β and *Bacillus smithii* ET 138 Δ*sigF* Δ*hsdR* (*315*). The *E. coli* and *B. smithii* strains were routinely cultured at 37 ℃ and 55 ℃, respectively, unless stated otherwise. *E. coli* strains were grown in 10 mL Luria-Bertani broth (LB) in 50-mL Greiner tubes or plated on LB with 15 g/L agar (Difco) plates containing 15 μg mL^{-1} chloramphenicol, unless stated otherwise. *B. smithii* strains were cultured in 10 mL LB2 medium (10 g/l tryptone, 5 g L^{-1} yeast extract, 100 mL L^{-1} 10×ESS, pH adjusted to 6.9-7.0) (*328*) in 50-mL Greiner tubes at 150 rpm or plated on LB2 with 30 g L^{-1} agar (Difco) plates supplemented with 7 μg mL^{-1} chloramphenicol, vitamins and metals (*328*). For all the gene editing experiments, glucose (0.5 g L^{-1}), xylose (0.5 g L^{-1}) and uracil (50 mg L^{-1}) were added to the media to induce expression of Cas12a and provide for the uracil auxotrophic *pyrF* deletion mutants. Additionally, Thermophile Vitamin Medium with Yeast extract (TVMY) medium (0.5 g L^{-1} yeast extract, 8.37 g L^{-1} MOPS, 100 mL L^{-1} 10×ESS, pH of 6.9-7.0) (*328*) or plates (made with 30 g L^{-1} agar) with the same supplements was used in the editing experiments. For the evaluation of 5-FOA sensitivity of *B. smithii* ET138 Δ*sigF* Δ*hsdR* Δ*pyrF* cells, $TVMY_{xgu}$, $LB2_{xgu}$ agar plates supplemented with 1 g L^{-1} 5-FOA were used.

Making of competent cells

E. coli DH10B cells were made electro-competent after growing an overnight culture of the strain in 10 mL LB medium at 37 °C. 5 mL of the overnight culture was used to inoculate 500 mL of pre-warmed 2×YP medium (10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, 10 g L^{-1} NaCl) at 37 °C, 180 rpm until an OD600 value of 0.4 was reached. Cells were then centrifuged for 10 minutes (4 °C, 3000xg) and washed with 250 mL of ice-cold MQ. The pelleted cells were suspended in 250 mL 10% glycerol and centrifuged at 3000xg for 10 minutes. The resuspension and subsequent centrifugation steps were repeated using 250 mL of 10% glycerol. The pellet was suspended in 2 mL 10% glycerol. Cells were aliquoted in micro-centrifuge tubes and stored at -80 °C.

B. smithii ET 138 Δ*sigF* Δ*hsdR* cells were made competent as described previously{Bosma, 2015 #148;Bosma, 2015 #799} (*328*). Briefly, an overnight culture was prepared in 10 mL LB2 medium supplemented with vitamins and metals and grown at 55 °C. The cells were diluted to an OD_{600} of 0.08 in LB2 containing vitamins and metals and grown to an OD_{600} of 0.55-0.6 at 55 °C, 150 rpm. Subsequently, the culture was centrifuged for 15 minutes at 4700 rpm (4 °C). The pellet was washed three times with volumes of 50 mL, 25 mL and 12.5 mL ice cold SG-buffer (171.2 g L^{-1} sucrose, 0.2 g L^{-1} MgCl2 and 50 mL L^{-1} glycerol). Centrifugation was performed after each washing step for 15 minutes at 4700 rpm (4 °C). The cells were suspended in 240 μL SG-buffer and aliquoted in micro-centrifuge tubes and stored at -80 °C.

Plasmid construction

The plasmids constructed and the primers (IDT) used for cloning and sequencing are listed in **Table S7.1** and **Table S7.2**, respectively. The fragments for assembling the plasmids were amplified by PCR with Q5® High-Fidelity 2X Master Mix (NEB). The amplicons were run on an 1% agarose gel electrophoresis and purified using Zymogen gel DNA recovery kit (Zymo Research). The HiFi DNA assembly mix was purified using the DNA Clean & Concentrator-5 (Zymo Research) and electroporated into competent *E. coli* DH10B cells using the following settings: 2.5kV, 200Ω and 25µF. Colony PCR with OneTaq® 2X Master Mix with Standard Buffer (NEB) on single colonies was used to screen for the right transformants. Thereafter, the right colonies were used to inoculate 10 mL LB medium, plasmid extractions were performed using the GeneJET Plasmid Miniprep kit (Thermo Fisher Scientific) and verified by Sanger sequencing (Macrogen Europe B.V).

For the construction of the pFnCas12a_Sp1-3 and pFnCas12a_NT plasmids, a 5-part assembly was designed and executed. The parts consist of: a fragment of the pNW33n backbone together with the P_{pta} promoter amplified (using BG14368 and BG14369) from the pWUR_Cas9sp1_hr vector (*315*); a fragment of the pNW33n backbone together with the P_{xynA} promoter amplified (using BG14370 and BG14371) from the pWUR_Cas9sp1_hr vector; a fragment of the pNW33n backbone amplified (using BG14372 and BG14373) from the pWUR_Cas9sp1_hr vector; the *fncas12a* gene was amplified (using BG14374 and BG14375) from the pFnCpf1_min plasmid; the fragment containing the spacer and repeat were DNA oligos annealed together (using BG14376 and BG14377).

For the construction of the pFnCas12a_Δ*pyrF*-HR_Sp1-3 and pFnCas12a_Δ*pyrF*-HR_NT plasmids, a 3-part assembly was designed and executed. The parts consist of: the upstream and downstream flanks of the *pyrF* gene amplified (using BG15279 and BG15280) from the pWUR_Cas9sp1_hr vector (*315*); a fragment of the pNW33n backbone together with the P_{xynA} promoter and the *fncas12a* gene amplified (using BG15281 and BG15282) from the corresponding pFnCas12a_Sp1/Sp3/NT plasmids; a fragment of the pNW33n backbone together with the P_{pta}, the repeat and spacer was amplified (using BG15283 and BG15284) from the corresponding pFnCas12a_Sp1/Sp3/NT plasmids.

For the construction of the pEeCas12a_Sp1 and pEeCas12a_NT plasmids, a 2-part assembly was designed and executed. The parts consist of: The *E. coli* optimised *eecas12a* gene amplified (using BG15839 and BG15840) from the pET28a_EeCas12a vector (NEB) and the pNW33n backbone along-with the spacer, repeat, P_{xynA} and P_{pta} amplified (using BG15833 and BG15834) from the corresponding pFnCas12a_Δ*pyrF*-HR_Sp1/NT plasmids. For the construction of the pEeCas12a_Δ*pyrF*-HR_Sp1 and pEeCas12a_Δ*pyrF*-HR_NT plasmids, a 2-part assembly was designed and executed. The parts consist of: The *E. coli* optimised *eecas12a* gene amplified (using BG15839 and BG15840) from the pET28a_EeCas12a vector (NEB) and the pNW33n backbone along-with the *pyrF* gene homologous flanks, spacer, repeat, P_{xynA} and P_{pta} amplified (using BG15833 and BG15834) from the corresponding pFnCas12a_Δ*pyrF*-HR_Sp1/NT plasmids.

B. smithii transformation, colony PCR, genomic DNA isolation and sequencing

Plasmids for transforming *B. smithii* were extracted from *E. coli* via miniprep isolation (Thermo Scientific). 75 µL electro-competent *B. smithii* cells were combined with 1 µg plasmid and trans-

formed in a 2 mm electroporation cuvettes (2.0 kV, 25 µF and 400 Ω) (*328*). Cells were recovered in 925 µL LB2 medium with vitamins and metals for 3 hours at 52 ℃, 150 rpm. The cells were plated out on LB2 plates incubated at 55 ℃ overnight. Potential *B. smithii* ET 138 Δ*sigF* Δ*hsdR* Δ*pyrF* colonies were randomly selected and subjected to colony PCR using the Q5® High-Fidelity 2X Master Mix (NEB) with the genome specific primers BG15752 and BG15559. Genomic DNA from *B. smithii* strains was isolated using the MasterPure™ Gram Positive DNA Purification Kit (Epicentre). Purification of PCR products was performed using the DNA Clean & Concentrator-5 (Zymo Research) kit. The DNA fragments were subsequently sent for Sanger sequencing (Macrogen Europe B.V) for confirmation of the gene deletion.

Acknowledgements

We would like to thank Tess Hoogeboom for her technical assistance. J.v.d.O. is supported by the NWO/TOP grant 714.015.001. R.v.K. is employed by Corbion.

Author contributions

P.M., I.M., R.v.K., R.T.F, M.M., G.B.R, and J.v.d.O., conceived this study and design of experiments. P.M., I.M., M.M., and J.A., conducted the experiments. P.M., I.M., R. v. K., and J.v.d.O. wrote the manuscript with input from all authors.

Author information

Correspondence should be addressed to john.vanderoost@wur.nl.

Competing interests

No potential conflict of interest is reported by the authors.

Supplementary Figures and Tables

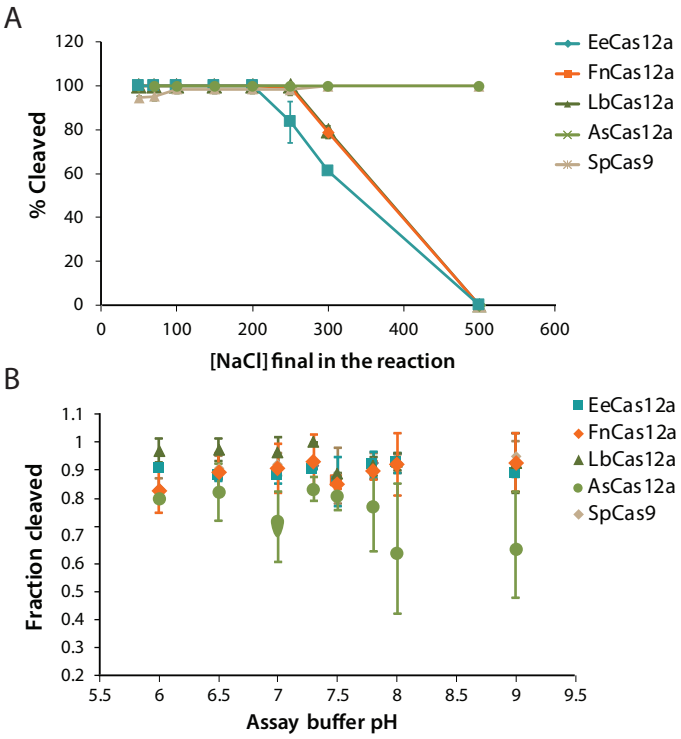


Figure S7.1 | Activity of Cas12a orthologs at varied pH and salt concentrations. **(A)** Cleavage of 5'-FAM-labeled linear dsDNA by Cas12a orthologs was monitored in reaction buffer 20 mM HEPES or TRIS-containing buffer with the pH indicated. The extent of cleavage was measured by resolution of the reaction products by capillary electrophoresis and quantitation of the cleaved and un-cleaved products. Data are shown as the mean +/- SD of at least two experimental replicates. **(B)** Cleavage of 5'-FAM-labeled linear dsDNA by Cas12a orthologs was monitored in reaction buffer 20 mM HEPES pH 6.5 (for AsCas12a) or 20 mM TRIS pH 8.0 (all other orthologs) buffer with 10 mM $MgCl_2$ and the final concentration of NaCl as indicated. The extent of cleavage was measured by resolution of the reaction products by capillary electrophoresis and quantitation of the cut and uncut products. Data are shown as the mean +/- SD of at least 2 experimental replicates.

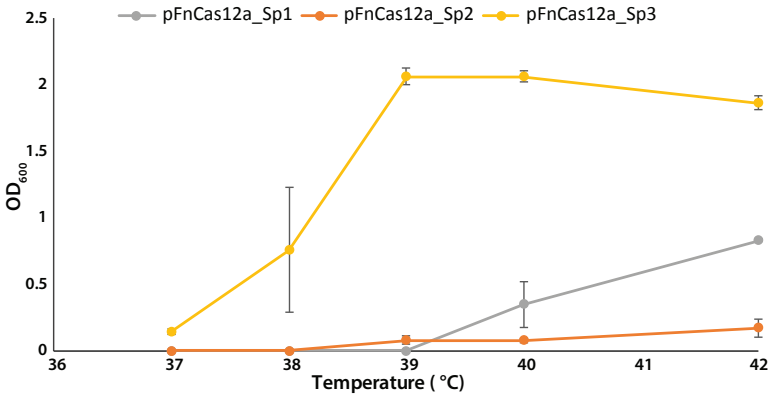


Figure S7.2 | Targeting assays performed with pFnCas12a_Sp1/3/NT plasmids. Results of the targeting experiment showing OD_{600} measurements from cultures of *B. smithii* transformed with pEeCas12a_Sp1 and pEeCas12a_Sp3 plasmids compared to a non-targeting pEeCas12a_NT plasmid and grown between 37-42 °C to evaluate the optimal temperature for *in vivo* activity. Average values of four replicates are shown, with error bars representing standard deviations.

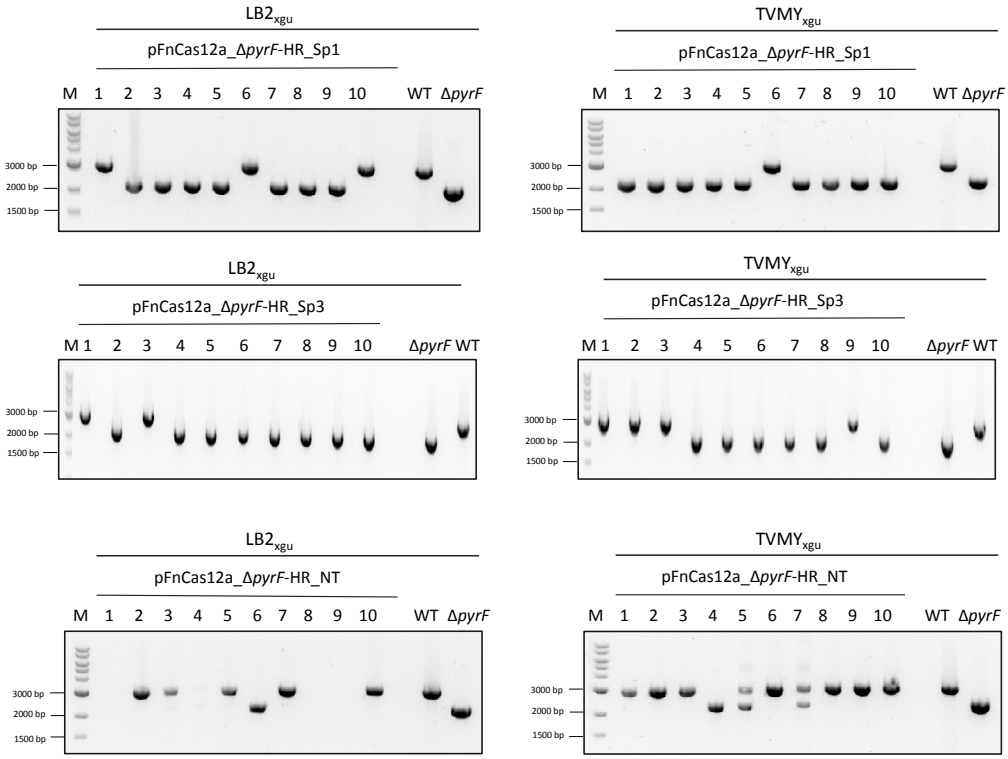


Figure S7.3 | Transformation of FnCas12a editing plasmids in *B. smithii* ET 13. Agarose gel electrophoresis showing the results from PCR on single colonies of *B. smithii* cultures transformed with pFnCas12a_Δ*pyrF*-HR_Sp1, pFnCas12a_Δ*pyrF*-HR_Sp3 and pFnCas12a_Δ*pyrF*-HR_NT) in two different selection media (TVMY$_{xgu}$ and LB2$_{xgu}$). The last 2 lanes are the negative (wild type) and positive (Δ*pyrF*) controls, that correspond to 2.9 kb and 2.2 kb long DNA fragments, respectively.

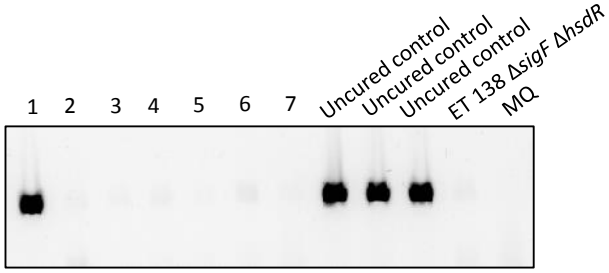


Figure S7.4 | Plasmid curing from *B. smithii* after successful mutant generation. 1-7 are colonies that grew only on the plate without antibiotic subjected to colony PCR using plasmid-specific primers. 3 colonies from TVMY$_{xgu}$ plate with antibiotics were used as an "uncured" negative control, *B. smithii* ET 138 Δ*sigF* Δ*hsdR* and MQ water were used as positive controls.

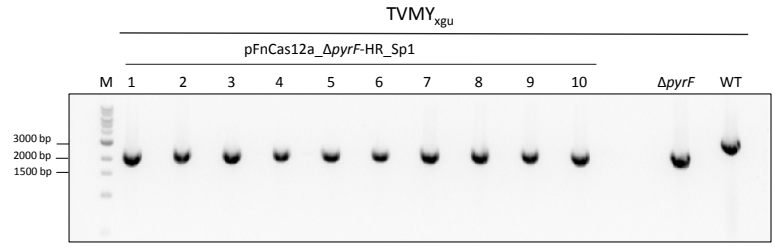


Figure S7.5 | Transformation of FnCas12a editing plasmid in *B. smithii* ET 138 using the fast method. Agarose gel electrophoresis showing the results from PCR on single colonies (1-10) of *B. smithii* cultures transformed with FnCas12a_Δ*pyrF*-HR_Sp1. The last 2 lanes are the positive (Δ*pyrF*) and negative (wild type) controls, that correspond to 2.2 kb and 2.9 kb long DNA fragments, respectively.

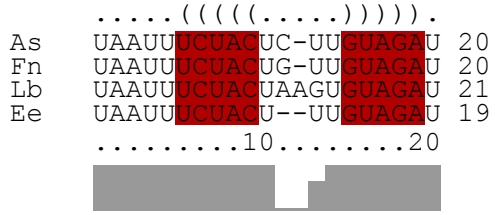


Figure S7.6 | Cas12a repeat sequences. Direct repeat sequences were aligned using LocARNA. Sequences and structures are well conserved among Cas12a orthologues.

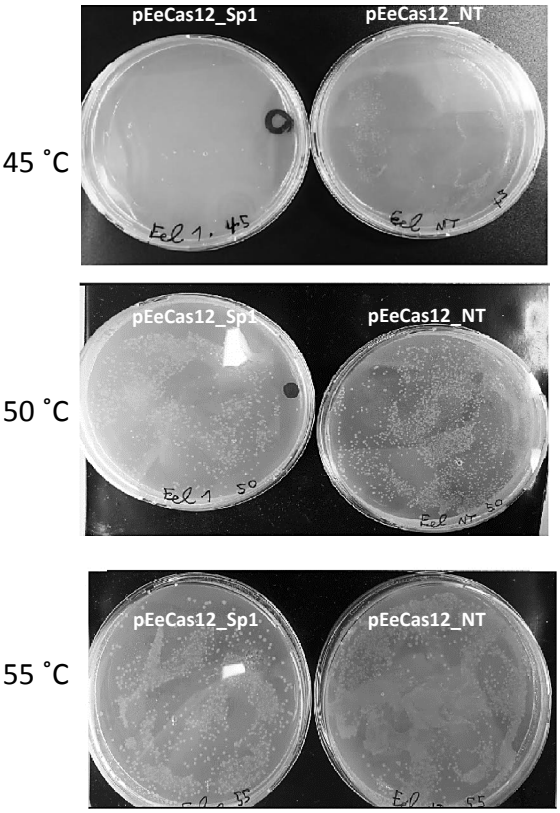


Figure S7.7 | Targeting of *B. smithii* genome using pEeCas12a plasmids. LB2 media plates containing cells transformed with pEeCas12a_Sp1 compared to cells transformed with pEeCas12a_NT after overnight incubation at 45˚ C, 50˚ C and 55˚ C.

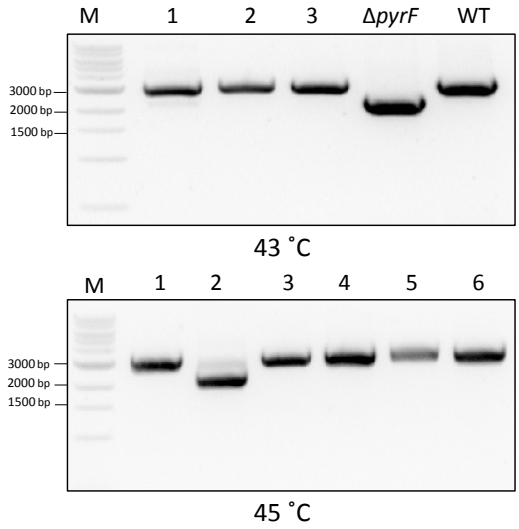


Figure S7.8 | Transformation of EeCas12a editing plasmid in *B. smithii* ET 138 using the fast method. Agarose gel electrophoresis showing the results from PCR on single colonies (1-3) and (1-6) of *B. smithii* cultures transformed with EeCas12a_Δ*pyrF*-HR_Sp1 and incubated at 43 ˚C and 45 ˚C, respectively. The negative (wild type) and positive (Δ*pyrF*) controls correspond to 2.9 kb and 2.2 kb long DNA fragments, respectively.

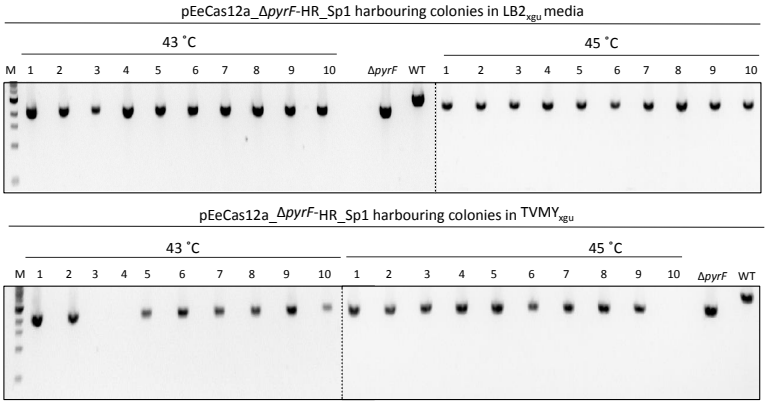


Figure S7.9 | Transformation of EeCas12a editing plasmids in *B. smithii* ET 13. Agarose gel electrophoresis showing the results from PCR on single colonies of *B. smithii* cultures transformed with pEeCas12a_Δ*pyrF*-HR_Sp1 in two different selection media ($TVMY_{xgu}$ and $LB2_{xgu}$) and incubated at 43 ˚C and 45 ˚C. The negative (wild type) and positive (Δ*pyrF*) controls correspond to 2.9 kb and 2.2 kb long DNA fragments, respectively.

Table S7.1 | Oligonucleotides used in this study.

Oligo ID	Sequence (5' - 3')	Description
	***In vitro* activity assays**	
NEB1	AGAGGCCACAAGTAAAGACTGCAG	FW for amplification of the human WTAP target gene with 5' FAM label
NEB2	CCAAAACGTACACCCACAAAAAGGC	RV for amplification of the human WTAP target gene
NEB3	CAAUUUCUACUUUUGUAGAUCCACUCACUGCUUUCUCCUC	Cas12a guide RNA
NEB4	TTCTAATACGACTCACTATAGAAAGGTAATCGAACTGTGTTTTAGAG CTAGA	Template for top strand for generation of Cas9 guide RNA
	Plasmids for FnCas12a-based targeting of *B. smithii* genome	
BG14368	ATCAGACAAAATGGCCTGCTTATG	FW for construction of pFnCas12a_Sp1/2/3/NT plasmids
BG14369	ACTCCTCGTATAACGGTATCC	RV for construction of pFnCas12a_Sp1/2/3/NT plasmids
BG14370	ACGATAGACATATAAAGAGAGCGGCC	FW for construction of pFnCas12a_Sp1/2/3/NT plasmids
BG14371	CCCCTCACATGCATTTAGGAGTTGT	RV for construction of pFnCas12a_Sp1/2/3/NT plasmids
BG14372	ACAACTCCTAAATGCATGTGAG	FW for construction of pFnCas12a_Sp1/2/3/NT plasmids
BG14373	TCTAGATTCCTCCCTCAGT	RV for construction of pFnCas12a_Sp1/2/3/NT plasmids
BG14374	AAATTTACTGAGGGAGGAATCTAGAATGTCAATTTATCAAGAATTTG	FW for construction of pFnCas12a_Sp1/2/3/NT plasmids
BG14375	TCATAAGCAGGCCATTTTGTCTGATTTAGTTATTCCTATTCTGC	RV for construction of pFnCas12a_Sp1/2/3/NT plasmids
BG14376	AAATGGATACCGTTATACGAGGAGTGTCTAAGAACTTTAAATAATTT CTACTGTTGTAGATAGAATGGACCGGCCATCATCGAAGC	FW for construction of pFnCas12a_Sp1/2/3/NT plasmids
BG14377	GCCGCTCTCTTTATATGTCTATCGTGCTTCGATGATGGCCGGTCCAT TCTATCTACAACAGTAGAAATTATTTAAAGTTCTTAGAC	RV for construction of pFnCas12a_Sp1/2/3/NT plasmids
BG14378	AAATGGATACCGTTATACGAGGAGTGTCTAAGAACTTTAAATAATTT CTACTGTTGTAGATATTTTCAGAATGGACCGGCCATCATCGAAG	FW for construction of pFnCas12a_Sp1/2/3/NT plasmids
BG14379	GCCGCTCTCTTTATATGTCTATCGTCTTCGATGATGGCCGGTCCATT CTGAAAATATCTACAACAGTAGAAATTATTTAAAGTTCTTAGAC	RV for construction of pFnCas12a_Sp1/2/3/NT plasmids
BG14380	AAATGGATACCGTTATACGAGGAGTGTCTAAGAACTTTAAATAATTT CTACTGTTGTAGATCATAATGGACCACTGATTCCTCCAG	FW for construction of pFnCas12a_Sp1/2/3/NT plasmids
BG14381	GCCGCTCTCTTTATATGTCTATCGTCTGGAGGAATCAGTGGTCCATT ATGATCTACAACAGTAGAAATTATTTAAAGTTCTTAGAC	RV for construction of pFnCas12a_Sp1/2/3/NT plasmids
BG14382	AAATGGATACCGTTATACGAGGAGTGTCTAAGAACTTTAAATAATTT CTACTGTTGTAGATGAGTGGATATGGTCAACGTGCATGC	FW for construction of pFnCas12a_Sp1/2/3/NT plasmids
BG14383	GCCGCTCTCTTTATATGTCTATCGTGCATGCACGTTGACCATATCCA CTCATCTACAACAGTAGAAATTATTTAAAGTTCTTAGAC	RV for construction of pFnCas12a_Sp1/2/3/NT plasmids
BG14384	GGTCTGATGCTACTAGGT	FW for sequencing the pFnCas12a_Sp1/2/3/NT plasmids
BG14385	ACAGGAAACAGCTATGACC	RV for sequencing the pFnCas12a_Sp1/2/3/NT plasmids
BG14386	TTGAGTTAATCCCACAGGG	FW for sequencing the pFnCas12a_Sp1/2/3/NT plasmids
BG14387	GCATCGTTGTAACTACATC	RV for sequencing the pFnCas12a_Sp1/2/3/NT plasmids

FW: Forward primer; RV: Reverse primer

Table S7.1 | Oligonucleotides used in this study. (continued)

Oligo ID	Sequence (5' - 3')	Description
BG14388	TAATACTATTATTGGTGGTA	FW for sequencing the pFnCas12a_Sp1/2/3/NT plasmids
BG14389	AGTTCTCAAAATTGAGCT	RV for sequencing the pFnCas12a_Sp1/2/3/NT plasmids
BG14390	GCTATCAAGGATCTTTTAGATC	FW for sequencing the pFnCas12a_Sp1/2/3/NT plasmids
BG14391	ATTAGCTATTGCCTCTTTAGCTG	RV for sequencing the pFnCas12a_Sp1/2/3/NT plasmids
BG14392	CAGGGTAAATTGTACCTA	FW for sequencing the pFnCas12a_Sp1/2/3/NT plasmids
BG14393	CTAAAGAACTCTTGAGATTTGC	RV for sequencing the pFnCas12a_Sp1/2/3/NTT plasmids
BG14394	GCTAATTGAGAAACTAAAC	FW for sequencing the pFnCas12a_Sp1/2/3/NT plasmids
BG14395	TACTCACTGCATGATTTC	RV for sequencing the pFnCas12a_Sp1/2/3/NT plasmids
BG8859	GTCTGACGCTCAGTGGAACGA	FW for sequencing the pFnCas12a_ΔpyrF-HR_Sp1/Sp3/NT plasmids
	Plasmids for FnCas12a-based editing of *B. smithii* genome	
BG15279	TCACGTTAAGGGATTTTGGTCATGAATGCAGCGATGGTCCGGTG	FW for construction of pFnCas12a_Δ*pyrF*-HR_Sp1/Sp3/NT plasmids
BG15280	ATTATCCTCAGCTCACTAGCGCCATTATCGGGTTCCCATGTTGTGAT TCGCC	RV for construction of pFnCas12a_Δ*pyrF*-HR_Sp1/Sp3/NT plasmids
BG15281	CCGATAATGGCGCTAGTGAGCTGAGG	FW for construction ofpFnCas12a_Δ*pyrF*-HR_Sp1/Sp3/NT plasmids
BG15282	GCTGCCTGCGGAAGAAAGGGAT	RV for construction of pFnCas12a_Δ*pyrF*-HR_Sp1/Sp3/NT plasmids
BG15283	ATCCCTTTCTTCCGCAGGC	FW for construction of pFnCas12a_Δ*pyrF*-HR_Sp1/Sp3/NT plasmids
BG15284	TCATGACCAAAATCCCTTAACGTGA	RV for construction of pFnCas12a_Δ*pyrF*-HR_Sp1/Sp3/NT plasmids
BG15263	CATAATCGTCGTCAGTAAGC	FW for sequencing of pFnCas12a_Δ*pyrF*-HR_Sp1/Sp3/NT plasmids
BG15264	CTGTCAAACCGATTGCGGTG	FW for sequencing of pFnCas12a_Δ*pyrF*-HR_Sp1/Sp3/NT plasmids
BG15265	GCAATTCATACTGTAATGTG	RV for sequencing of pFnCas12a_Δ*pyrF*-HR_Sp1/Sp3/NT plasmids
BG15334	GACCTACTGCAAATGCGATTAGGC	FW for sequencing of pFnCas12a_Δ*pyrF*-HR_Sp1/Sp3/NT plasmids
BG15833	ATCAGACAAAATGGCCTGC	FW for construction of pEeCas12a-Sp1/NT and pEeCas12a_Δ*pyrF*-HR_Sp1/NT plasmids
	Plasmids for EeCas12a-based targeting and editing of *B. smithii* genome	
BG15834	TCTAGATTCCTCCCTCAG	RV for construction of pEeCas12a-Sp1/NT and pEeCas12a_Δ*pyrF*-HR_Sp1/NT plasmids
BG15839	AAATTTACTGAGGGAGGAATCTAGAATGAACGGTAACCGTAGCATCG	FW for construction of pEeCas12a-Sp1/NT and pEeCas12a_Δ*pyrF*-HR_Sp1/NT plasmids
BG15840	TCATAAGCAGGCCATTTTGTCTGATTTATTCGTAACGTTTGTTCTGG ATG	RV for construction of pEeCas12a-Sp1/NT and pEeCas12a_Δ*pyrF*-HR_Sp1/NT plasmids
BG15869	GCTGGTGTACGCCAGGATTTG	FW for sequencing of pEeCas12a-Sp1/NT and pEeCas12a_Δ*pyrF*-HR_Sp1/NT plasmids
BG15870	GATTAGCGAGAAGGCGCTG	FW for sequencing of pEeCas12a-Sp1/NT and pEeCas12a_ΔpyrF-HR_Sp1/NT plasmids

FW: Forward primer; RV: Reverse primer

Table S7.1 | Oligonucleotides used in this study. (continued)

Oligo ID	Sequence (5' - 3')	Description
BG15871	TTCACGATGTCCTTCTGCG	FW for sequencing of pEeCas12a-Sp1/NT and pEeCas12a_Δ*pyrF*-HR_Sp1/NT plasmids
BG15872	CGATCCCGGACGATATCTAC	FW for sequencing of pEeCas12a-Sp1/NT and pEeCas12a_Δ*pyrF*-HR_Sp1/NT plasmids
BG15873	GGAAGAAAGGGATGCGGCTG	RV for sequencing of pEeCas12a-Sp1/NT and pEeCas12a_Δ*pyrF*-HR_Sp1/NT plasmids
	Check primers	
BG15558	TCGGGGGTTCGTTTCCCTTG	FW for PCR analysis of *B. smithii* colonies or genomic DNA
BG15559	CTTACACAGCCAGTGACGGAAC	RV for PCR analysis of *B. smithii* colonies or genomic DNA
BG15752	CCTCGCTCGATGATATTCTCG	FW #2 for PCR analysis of *B. smithii* colonies or genomic DNA

FW: Forward primer; RV: Reverse primer

Table S7.2 | Plasmids used in this study.

Plasmid	Description	Primers	Source
pNW33n	*E. coli* - *Bacillus* shuttle vector, cloning vector, CamR	-	BGSC
pFnCpf1_min	Expresses WT FnCas12a (Cpf1) and spacers 1-4 of CRISPR array under artificial promoter.	-	Kind gift from Prof. Dr. Feng Zhang
pWUR_Cas9sp1_hr	pNW33n with spCas9-module containing spacer targeting the *pyrF* gene. This plasmid was used as a template for constructing the FnCas12a and EeCas12a based constructs.	-	(*315*)
pET28a_EelCas12a	Expresses EeCas12a from a pET28a vector		New England Biolabs/this study
pMiniT-WTAP Exon 8	Contains a fragment of the human WTAP gene used as the target in the *in vitro* target cleavage assay		New England Biolabs/this study
pET28_FnCas12a	Expresses FnCas12a from a pET28a vector		New England Biolabs/this study
pET28_AsCas12a	Expresses AsCas12a from a pET28a vector		New England Biolabs/this study
pFnCas12a_Sp1	pNW33n with FnCas12a module[1] containing spacer 1 targeting the *pyrF* gene in *B. smithii* ET 138	See **Table S7.1**	This study
pFnCas12a_Sp2	pNW33n with FnCas12a module[1] containing spacer 2 targeting the *pyrF* gene in *B. smithii* ET 138	See **Table S7.1**	This study
pFnCas12a_Sp3	pNW33n with FnCas12a module[1] containing spacer 3 targeting the *pyrF* gene in *B. smithii* ET 138	See **Table S7.1**	This study
pFnCas12a_NT	pNW33n with FnCas12a module[1] containing a non-targeting spacer	See **Table S7.1**	This study

Table S7.2 | Plasmids used in this study. (continued)

Plasmid	Description	Primers	Source
pFnCas12a_Δ*pyrF*-HR_Sp1	pNW33n with FnCas12a module[1] containing spacer 1 targeting the *pyrF* gene in *B. smithii* ET 138 and *pyrF*- HR flanks	See **Table S7.1**	This study
pFnCas12a_Δ*pyrF*-HR_Sp2	pNW33n with FnCas12a module[1] containing spacer 2 targeting the *pyrF* gene in *B. smithii* ET 138 and *pyrF*- HR flanks	See **Table S7.1**	This study
pFnCas12a_Δ*pyrF*-HR_NT	pNW33n with FnCas12a module[1] containing a non-targeting spacer and *pyrF*-HR flanks	See **Table S7.1**	This study
pEeCas12a_Sp1	pNW33n with EeCas12a module[1] containing spacer 1 targeting the *pyrF* gene in *B. smithii* ET 138	See **Table S7.1**	This study
pEeCas12a_NT	pNW33n with EeCas12a module[1] containing a non-targeting spacer	See **Table S7.1**	This study
pEeCas12a_Δ*pyrF*-HR_Sp1	pNW33n with EeCas12a module[1] containing spacer 1 targeting the *pyrF* gene in *B. smithii* ET 138 and *pyrF*- HR flanks	See **Table S7.1**	This study
pEeCas12a_Δ*pyrF*-HR_NT	pNW33n with EeCas12a module[1] containing a non-targeting spacer and *pyrF*-HR flanks	See **Table S7.1**	This study

[1] The FnCas12a and EeCas12a module contains *fncas12a* and *eecas12a* genes, respectively, under the *Thermoanaerobacterium saccharolyticum* P_{xynA} promoter followed by the crRNA under the *B. coagulans* P_{pta} promoter.

Table S7.3 | Spacer sequences used in the Cas12a gRNAs for targeting *Bacillus smithii* genome.

Spacer name	Strand	(Proto) spacer sequence (5'-3')	PAM sequence (5'-3')
		pyrF gene in *B. smithii ET 138* genome	
pyrF spacer **1**	+	AGAATGGACCGGCCATCATCGAAGC	TTTC
pyrF spacer **2**	+	ATTTTCAGAATGGACCGGCCATCATCGAAG	CTTT
pyrF spacer **3**	-	CATAATGGACCACTGATTCCTCCAG	TTTG
pyrF non-targeting (**NT**) spacer	+	GAGTGGATATGGTCAACGTGCATGC	CTCG

Chapter 8

Characterising an anti-CRISPR based on/off switch for bacterial genome engineering

Despoina Trasanidou*, Prarthana Mohanraju*, Richard van Kranenburg, Ioannis Mougiakos, John van der Oost

* contributed equally

Adapted from:
Developing an anti-CRISPR-based toolbox for bacterial genome engineering
Manuscript in preparation

Abstract

CRISPR–Cas9 technologies have enabled unprecedented efficient genome editing and transcriptional regulation in prokaryotes, allowing for their accelerated exploration and exploitation. However, tight control of Cas9 expression requires either multi-step approaches, or synthetic fusions of Cas9 and specific sensor proteins that respond to external stimuli. Here we harness the function of a small anti-CRISPR protein from Neisseria meningitidis (AcrIIC1$_{\mathrm{Nme}}$) as natural and direct "on/off-switch" of two thermo-tolerant Cas9 orthologues, from Geobacillus thermodenitrificans *T12 (ThermoCas9) and* Geobacillus stearothermophilus *(GeoCas9). We demonstrate that both ThermoCas9 and GeoCas9 are* in vivo *active at 37 ℃ and can be used for introducing dsDNA breaks in* E. coli*, in a tuneable and spacer-dependent manner. In addition, we show that AcrIIC1$_{\mathrm{Nme}}$ traps* in vivo *these Cas9 endonucleases in a DNA-bound, catalytically inactive state, robustly inhibiting targeting and resulting in a transcriptional silencing that is comparable to their catalytically "dead" variants (ThermodCas9 and GeodCas9). Moreover, we describe a single-vector, tightly controllable and highly efficient Cas9/AcrIIC1$_{\mathrm{Nme}}$-based tool for silencing in bacteria. Altogether, an anti-CRISPR protein has been used to control Cas9-based genome editing and transcriptional regulation.*

Introduction

The development of genetic engineering tools has been instrumental for fundamental and applied (micro)biology. Initially, a range of seminal discoveries (restriction enzymes, ligases, polymerases) and pivotal technological advances (Sanger sequencing, PCR) spurred recombinant DNA technologies, thereby setting the stage for the establishment of a basic toolkit for precise genetic manipulation (deletion, insertion, substitution) and metabolic engineering (*380*). In case of bacteria, efficient genome engineering traditionally hinges on the endogenous homologous recombination (HR) system and homologous recombination templates combined with counter-selection systems (*pyrF*-system, *lacZ*-system, *tdk-hpt* system) or additional recombinase-based systems (Cre-*lox* system and FLP-*FRT* system) (*311*, *381*, *382*). However, these methods are laborious, time-consuming, require available selectable markers and the latter two leave genomic scars that could trigger undesirable chromosomal rearrangements. Alternatively, endogenous or viral recombineering methods (Rac prophage-RecET system, phage λ-Red system) have been applied in few bacteria, though providing low engineering efficiencies (*252*, *383*). In addition, group II intron retrotransposition (ClosTron system, TargeTron system) has been employed for limited number of bacterial species, albeit often leading to unstable insertions and polar effects (*252*, *384*–*386*). Moreover, the mentioned tools are generally suboptimal for editing the genome of non-model microorganisms with low transformation and recombination efficiencies. Hence, the need for user-friendly, high-throughput and broad-range genome engineering tools remained.

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) and CRISPR-associated (Cas) genes are prokaryotic immune systems that have recently been repurposed as a next-generation, genome engineering toolset. CRISPR–Cas technology has unprecedentedly given the means for streamlined genome editing and stringent transcriptional control for model and non-model industrial workhorses (*72*, *387*, *388*) . CRISPR–Cas applications have also spanned across the detection/tracking/typing and epidemiology of spoilage and pathogenic strains, vaccination of fermentative bacteria against phage threats or undesirable plasmid uptake as well as the development of smart antibiotics and probiotics with improved metabolism and resistance in the gut (*283*, *387*, *389*). The integration of the cutting-edge CRISPR–Cas molecular scissors into the current molecular biology scheme has undeniably provided the impetus for a rapid, and cost-effective manipulation of industrially attractive organisms.

The popularity of the CRISPR–Cas toolbox originates from its impressive efficiency and simplicity. This versatile genome editing tool relies on only two elements: (i) a Cas endonuclease (usually from the type II-A CRISPR–Cas system of *Streptococcus pyogenes* strain SF370, SpCas9) (*72*, *387*, *388*) and (ii) a programmable single-guide RNA (sgRNA) molecule (*51*). The sgRNA guides the Cas9 nuclease to introduce double-stranded DNA breaks (DSBs) to a desired DNA sequence (protospacer), which is complementary to the exchangeable 5'-end of the sgRNA (spacer) (*49*) and flanked downstream by a 3-8nt long conserved motif (protospacer adjacent motif, PAM) (*45*, *46*). Unlike eukaryotes, Cas9-mediated DSBs are lethal for most prokaryotes in the absence of a rescuing template, due to the lack or the conditional expression of an efficient Non-Homologous End Joining (NHEJ) repair mechanism (*368*, *371*, *390*, *391*). Leveraging this lethality, recent genetic engineering methods interlace Cas9 technology with HR or recombineering systems, providing a chromosome/plasmid-borne or ss/ds-DNA editing template, respectively (*72*, *369*, *392*). The Cas9 endonuclease either triggers cellular repair mechanisms, usually resulting in escape mutants and mixed wild type-mutant genotypes (*157*, *393*,

394), or serves as a stringent counter-selection system that allows HR to occur before the Cas9-mediated targeting, though often requiring multiple-plasmid strategies or strictly controlled promoters (*361*, *395*–*397*). Due to the absence of well-characterised inducible promoters in non-model organisms (*398*), Cas9-based targeting often acts as an extended recombination inducer, leading to lower survival rates and editing efficiencies. As such, there is a tangible need for tight control of the Cas9-expression through alternative strategies, as has been recently demonstrated via temperature-based activation of SpCas9 in a moderate thermophile (*315*), and through the development of synthetic Cas9 variants regulated by drugs, light or other external stimuli (*399*–*407*).

Intriguingly, several small (50 - 150 amino acids) proteins naturally neutralizing CRISPR–Cas systems were recently identified in prokaryotic viruses and mobile genetic elements (*97*, *98*, *102*, *107*, *408*–*411*), launching the now rapidly expanding anti-CRISPR field. Out of the currently thirty-seven identified anti-CRISPR proteins, ten have been characterised and shown to exhibit direct interactions with a Cas protein/complex, abolishing DNA cleavage, and thus holding tremendous promise as natural "off-switches". Specifically, most anti-CRISPRs hinder DNA-binding (complete "off-switch") by association with the Cascade (AcrID1, AcrIF1, AcrIF2, AcrIF4, AcrIF10) (*94*, *99*, *104*, *410*, *412*, *413*), by masking of the PAM-interacting domain of Cas9 (AcrIIA2, AcrIIA4) (*337*, *414*–*416*) or by forcing dimerization of the Cas9 (AcrIIC3) (*408*, *417*). Such proteins have been successfully applied to limit off-target editing (*108*, *408*, *415*), control leaky CRISPR interference (CRISPRi) (*408*), and optimize CRISPR-based imaging (*408*) in human cells as well as to inactivate a gene drive in budding yeast (*418*), and to hamper CRISPRi in *E. coli* (*108*). Particularly, two anti-CRISPR proteins, AcrIF3 and AcrIIC1, have attracted interest, due to their ability to impede DNA cleavage but still permit DNA binding (incomplete "off-switch"), thereby switching Cas proteins from the catalytically active (targeting) to the catalytically inactive (silencing) state. AcrIF3 hampers the recruitment of the Cas3 endonuclease to the DNA-associated Cascade (*94*, *100*) in type I CRISPR–Cas systems. AcrIIC1 binds to the HNH nuclease domain in certain type II-C Cas9 systems, obstructing the cleavage of the target DNA strand, thereby restricting conformational changes necessary for the cleavage of the non-target strand by the RuvC domain (*417*). In this way, AcrIIC1 traps Cas9 in a DNA-bound, catalytically inactive state, resulting in transcriptional silencing of a target gene. The robust inhibition of Cas9 endonucleases from *Neisseria meningitidis* (NmeCas9), *Campylobacter jejuni* (CjeCas9) and *Geobacillus stearothermophilus* (GeoCas9) both *in vitro* and in mammalian cells (*408*, *417*) by AcrIIC1 from *N. meningitidis* (AcrIIC1$_{Nme}$) has attracted attention for its potential to become an appealing, broad-range and natural CRISPR/CRISPRi-regulator. However, AcrIIC1$_{Nme}$ has never been employed for bacterial genome engineering.

In this study, we report that the Cas9 orthologues from *Geobacillus thermodenitrificans* (ThermoCas9) and *Geobacillus stearothermophilus* (GeoCas9), which are functional at a wide temperature range, both efficiently introduce lethal DSBs in the *E. coli* genome. We demonstrate that the AcrIIC1$_{Nme}$ robustly inhibits the nuclease activity of both ThermoCas9 and GeoCas9 *in vivo* and can be used to silence a chromosomally integrated *gfp* gene in *E. coli* in a tunable manner. The gene silencing efficiencies of the ThermoCas9/GeoCas9: AcrIIC1$_{Nme}$ complexes is subsequently compared to that of their catalytically inactive ("dead") counterparts (GeodCas9, ThermodCas9). Following on these advances, we develop a one-plasmid based, tuneable silencing system. This sets further directions for the development of a tightly controllable antiCRISPR-based toolkit for sequential silencing and editing in *E. coli* and other industrially relevant non-model organisms. This strategy is expected not only to prolong the HR

phase and thus improve the editing efficiencies in a wide range of microorganisms, but also circumvent the need of time-consuming construction of distinct plasmid variants for editing and silencing.

Results

ThermoCas9 and GeoCas9 are active in *E. coli*

Recently, we characterised *in vitro* the thermo-tolerant and thermo-active ThermoCas9 endonuclease from the type II-C CRISPR–Cas system of the thermophilic *Geobacillus thermodenitrificans* T12 strain (*329*). We further employed ThermoCas9 to develop the first Cas9-based genome editing and silencing tools for thermophilic bacteria, readily applicable to mesophilic bacteria too (*329*). ThermoCas9 and its catalytically inactive mutant ThermodCas9 were applied for gene deletion and transcriptional silencing, respectively, in *Bacillus smithii* ET138 (55 ℃) as well as for gene deletion in *Pseudomonas putida* KT2440 (37 ℃) (*329*). Concurrently, the closely related GeoCas9 orthologue (88% amino acid identity to ThermoCas9) from the type II-C CRISPR–Cas system of another thermophilic species, *Geobacillus stearothermophilus*, was characterised *in vitro* and employed for *in vitro* DNA targeting in human plasma (*419*), for gene disruption in mammalian cells (*419*) and for *in vivo* targeting of viral DNA during *E. coli* infection (*417*). The broad temperature activity range (20-70 ℃ *in vitro*), the recalcitrance and the relatively small sizes of ThermoCas9 and GeoCas9 (**Figure S8.1A**) could serve as basis for the development of various genome engineering applications in a broad range of model and non-model bacteria.

In this study, we started by determining the *in vivo* DNA targeting efficiency of ThermoCas9 and GeoCas9 at 37 ℃. For this purpose, we employed the previously constructed *E. coli* DH10B_*gfp* strain that harbors a genome-integrated *gfp* gene under the transcriptional control of the constitutive *lacUV5* promoter (P_{lacUV5}) (Creutzburg *et al.*, in preparation). The *cas9* (*thermocas9* or *geocas9*) genes were cloned separately into the low copy number plasmid pACYC184, under the transcriptional control of the IPTG-inducible *tet* promoter (P_{tet} combined with *lac* operator, $P_{tet-lac}$). In the same plasmids, we cloned the *lacI* gene, under the transcriptional control of the constitutive *lacI* promoter (P_{lacI}), and a sgRNA-expressing module under the transcriptional control of the synthetic, constitutive promoter P_{J23119}, lacking a ribosomal binding site (RBS) (**Figure 8.1A**, **S8.1B**). Four spacers were introduced separately into the 5'-end of the sgRNA expressing module of the plasmids. Three of the spacers were targeting either a protospacer on the non-coding strand close to the start codon of the *gfp* gene in the genome of the *E. coli* DH10B_*gfp* strain or protospacers in the promoter region of the same gene (**Figure 8.1A**), and the fourth spacer was scrambled, not targeting any region in the genome of the *E. coli* DH10B_*gfp* strain (**Table S8.1**). Two of the spacers (spacers 1 and 2) were corresponding to protospacers with 5'-NNNNCACA-3' as PAM and the remaining spacer (spacer 3) was corresponding to a protospacer with 5'-NNNNCCAA-3' as PAM. Both PAM sequences have been proven to allow ThermoCas9 targeting *in vitro* (*329*), while they have not been tested for GeoCas9. This cloning process resulted in the construction of the pTCas9_Sp1/2/3/scr and pGCas9_Sp1/2/3/scr series of plasmids (**Figure 8.1**, **Table S8.2**). The pACYC184 plasmid and the ThermoCas9 and GeoCas9 plasmids containing the non-targeting, scrambled spacer were used as transformation controls. After transformation of the whole series of plasmids to *E. coli* DH10B_*gfp* cells, the ThermoCas9 plasmid bearing the spacer that corresponds to a protospacer with the optimal PAM for ThermoCas9 (spacer 3) resulted in at least 4 orders of magnitude reduction in the transformation efficiency compared to the control plasmid, even when

ThermoCas9 was not induced (**Figure 8.1B**, **S8.2A**). Moreover, the transformation of the ThermoCas9 plasmid harboring spacer 1 that corresponds to a protospacer with a suboptimal PAM for ThermoCas9-mediated targeting, resulted in transformation efficiency levels similar to the controls in the absence of ThermoCas9 induction, whereas at maximum ThermoCas9 induction there was complete loss of transformation efficiency. Notably, no ThermoCas9 activity was observed for the plasmid with spacer 2, even at the highest ThermoCas9 induction. Intriguingly, even though spacer 1 and 2 target closely located protospacers that have the same PAM, the targeting efficiencies were completely different. This phenomenon suggests that ThermoCas9 cleavage activity depends not only on the PAM or the location of a protospacer within the genome but also on the spacer sequence. Regarding GeoCas9 targeting efficiency, the plasmids carrying spacers 1 and 2 resulted in transformation efficiencies similar to the controls in absence of GeoCas9 induction, whilst high GeoCas9 induction resulted in at least 3 orders of magnitude drop in the transformation efficiency (**Figure 8.1C**, **S8.2B**). However, the GeoCas9 plasmid with the spacer 3 caused less than 1 order of magnitude decrease in the transformation efficiency in all GeoCas9 induction conditions, suggesting that GeoCas9 activity was very limited when this spacer, or this PAM, was employed. Altogether, these findings demonstrate that both ThermoCas9 and GeoCas9 are actively targeting the *E. coli* DH10B_*gfp* genome, in a spacer-dependent manner. Nonetheless, stricter control of their cleavage activity would substantially benefit future genome engineering applications to reach optimal efficiencies.

AcrIIC1$_{Nme}$ inhibits ThermoCas9 and GeoCas9 activity in *E. coli* (dual-vector system)

Recently, various anti-CRISPR systems, encoded by viral genomes and mobile genetic elements, were demonstrated to block the targeting activity of CRISPR–Cas systems. The exploitation of anti-CRISPR proteins for regulating the activity of Cas9 endonucleases during genome engineering applications has the potential to increase the efficiency of these systems. The AcrIIC1 from *Neisseria meningitidis* (AcrIIC1$_{Nme}$) has displayed robust inhibition of GeoCas9-mediated genome editing in human cells (*417*), albeit not yet in prokaryotes. Specifically, *in vitro* studies showed that AcrIIC1$_{Nme}$ binds to the active site of the HNH nuclease domain (D587, H588), trapping GeoCas9 in a DNA-bound but catalytically inactive state (*417*). The crystal structure of the GeoCas9 HNH-AcrIIC1$_{Nme}$ complex showed that AcrIIC1$_{Nme}$ additionally binds to five charged residues around the active site (T549, H551, D598, K603, N616), potentially ensuring the high stability of this inter-protein interaction (*417*). Some of these five charged residues are conserved among diverse type II-C Cas9 orthologues, while others are species-specific. These findings prompted us to align and compare the HNH domains of GeoCas9 and ThermoCas9. We observed that all the crucial residues for the association of the ThermoCas9 HNH domain with AcrIIC1$_{Nme}$ were identical to the GeoCas9 HNH domain, except for the histidine at position 551(H551) that was substituted with the non-charged leucine in the ThermoCas9HNH domain (**Figure S8.3**). These data suggested that not only GeoCas9 but also ThermoCas9 cleavage activity can be sufficiently blocked by AcrIIC1$_{Nme}$.

To evaluate the *in vivo* efficiency of AcrIIC1$_{Nme}$ to block the GeoCas9 and ThermoCas9 activities, we repeated the ThermoCas9- and GeoCas9- targeting assays employing an AcrIIC1$_{Nme}$-expressing *E. coli* DH10B_*gfp* strain. The employed *E. coli* DH10B_*gfp* strain was already transformed with the AcrIIC1$_{Nme}$-encoding pAcr plasmid, resulting in the *E. coli* DH10B_*gfp*(pAcr) strain (**Figure 8.2A**). The pAcr was constructed by cloning the *E. coli* codon optimised *AcrIIC1*$_{Nme}$ gene in the high copy number pUC19 backbone plasmid under the transcriptional control of the L-rhamnose-inducible promoter (P$_{rha}$) (**Figure 8.2A**). This tactic was expected

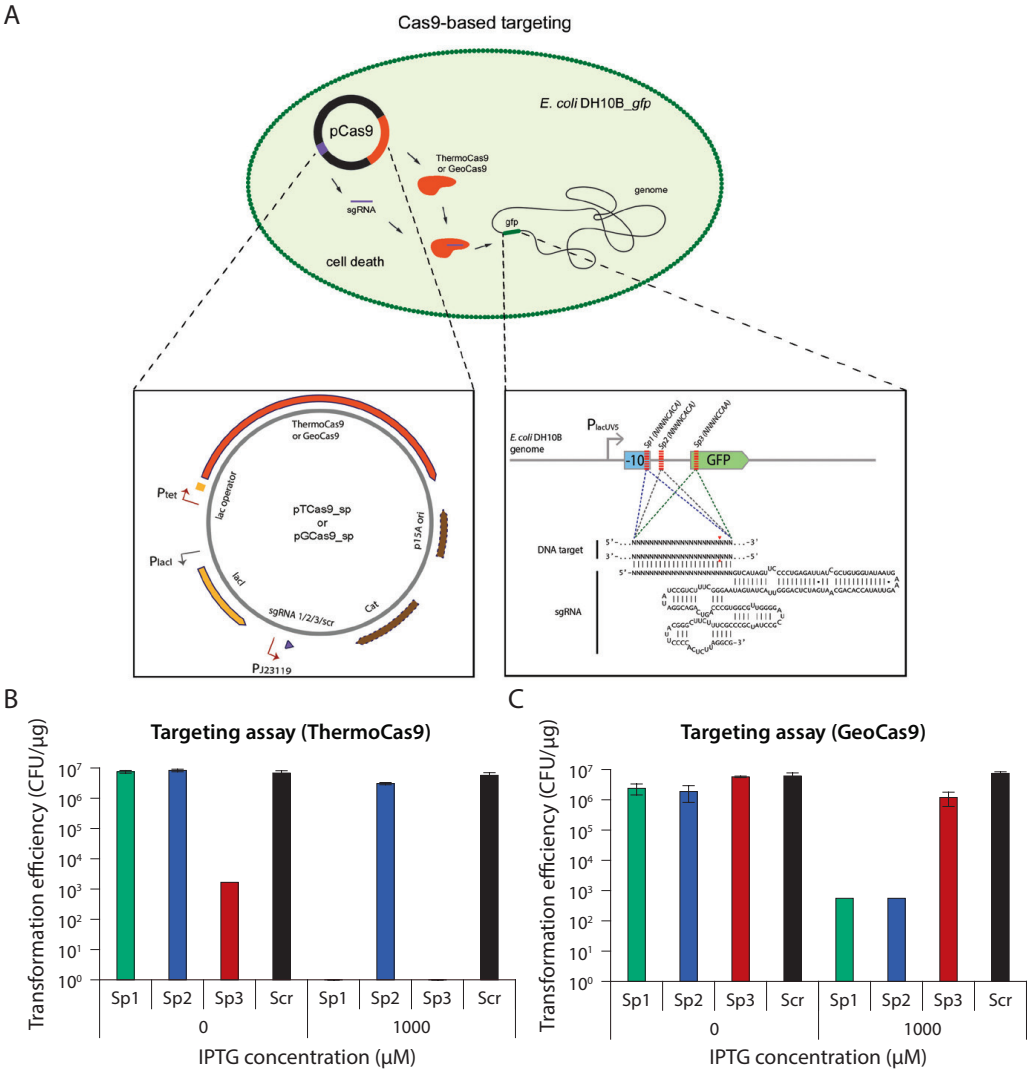


Figure 8.1 | ThermoCas9- and GeoCas9-based targeting in *E. coli* DH10B_*gfp*. **(A)** Schematic illustration of the strategy for ThermoCas9- and GeoCas9-based targeting in *E. coli* DH10B_*gfp*. The pCas9 plasmid was used for the expression of the ThermoCas9/GeoCas9 nuclease and its respective sgRNA to target and cleave specific sites on the genome of *E. coli* DH10B_*gfp*. **(B)** Results of the ThermoCas9-based targeting assays in *E. coli* DH10B_*gfp* cells employing the pTCas9 plasmid series. **(C)** Results of the GeoCas9-based targeting assays in *E. coli* DH10B_*gfp* cells employing the pGCas9 plasmid series.

to allow for high expression of AcrIIC1$_{Nme}$, and thus sufficient inhibition of ThermoCas9 and GeoCas9 activity upon transformation of the *E. coli* DH10B_*gfp*(pAcr) cells with the most effectively targeting plasmids pTCas9_Sp3 and pGCas9 Sp1 and 2 (**Figure 8.2A**), as well as the pTCas9$_{L551H}$_Sp3 plasmid that is similar to the pTCas9_Sp3 plasmid but encompasses the *thermoCas9$_{L551H}$* gene that codes for the ThermoCas9$_{L551H}$ variant, which harbours the L551H substitution (**Figure S8.4A**). Assays were conducted with and without L-rhamnose supplementation (**Figure S8.4A**). Strikingly, complete abolishment of the ThermoCas9, ThermoCas9$_{L551H}$

and GeoCas9 activity was observed in all cases even without L-rhamnose-based induction of the AcrIIC1$_{Nme}$ expression (**Figure 8.2B,C**; **S8.4A,B**; **S8.5A,B**), most probably due to leaky AcrIIC1$_{Nme}$ expression at adequate levels for targeting inhibition. Moreover, the presence of the hydrophobic leucine at the 551 position of the ThermoCas9 sequence, instead of the positively charged histidine that is present in the GeoCas9 sequence, apparently did not inhibit the binding of the AcrIIC1$_{Nme}$ to levels that could reduce the ThermoCas9 inhibition in the current system. To sum up, in the described dual-vector system AcrIIC1$_{Nme}$ presented robust inhibition of the ThermoCas9-, ThermoCas9$_{L551H}$- and GeoCas9-based targeting in *E. coli* DH10B_*gfp* cells, in a non-tuneable manner.

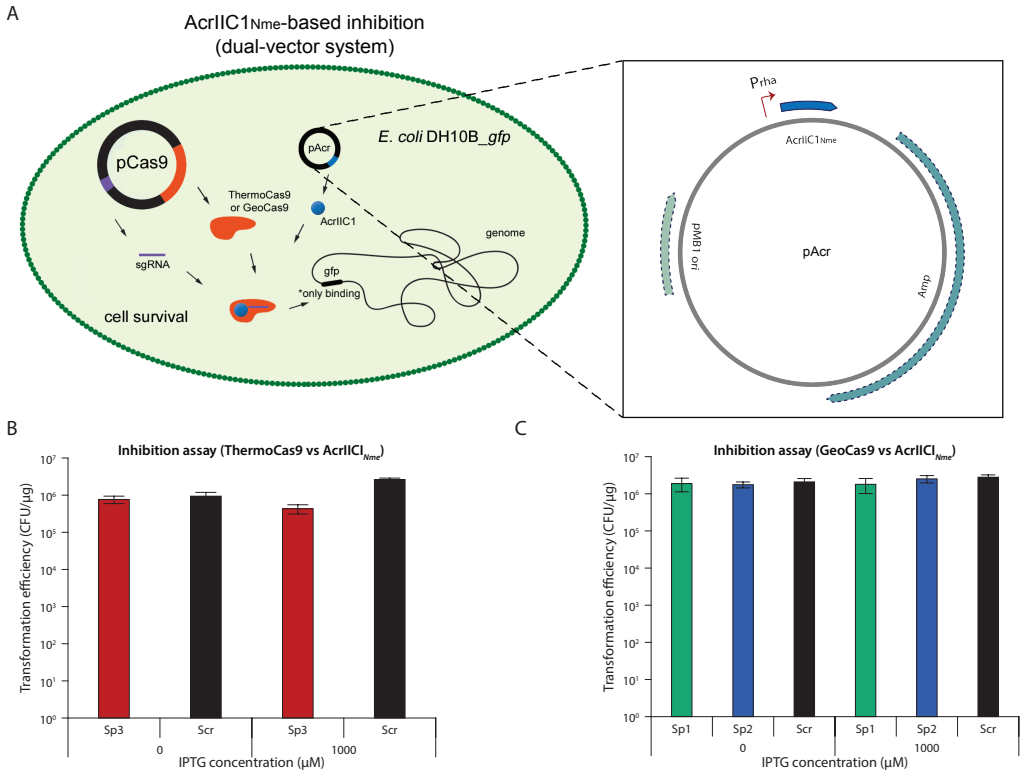


Figure 8.2| AcrIIC1$_{Nme}$-mediated inhibition of ThermoCas9- and GeoCas9-based targeting in *E. coli* DH10B_*gfp* (dual-vector system). (A) Schematic illustration of the dual-vector strategy for AcrIIC1$_{Nme}$-mediated inhibition of ThermoCas9- and GeoCas9-based targeting in *E. coli* DH10B_*gfp*. The pCas9 plasmid was used for the expression of the ThermoCas9/GeoCas9 nuclease and its respective sgRNA to target and cleave specific sites on the genome of *E. coli* DH10B_*gfp* (either at the genome-integrated gfp gene or its promoter, PlacUV5). The Cas9-inhibitor AcrIIC1$_{Nme}$ was expressed from the pAcr plasmid in the same strain. **(B)** Results of the AcrIIC1$_{Nme}$-mediated inhibition assay of ThermoCas9-based targeting. **(C)** Results of the AcrIIC1$_{Nme}$-mediated inhibition assay of GeoCas9-based targeting.

ThermoCas9/GeoCas9:AcrIIC1$_{Nme}$ complexes mediate gene silencing in *E. coli*

In this part of the study, we set off to verify the hypothesis that *in vivo* binding of AcrIIC1$_{Nme}$ to the HNH domain of ThermoCas9, and GeoCas9 transforms the enzymes into a DNA-bound but catalytically inactive form, with similar activity to the catalytically inactive ThermoCas9$_{D8A,H582A}$

(ThermodCas9) and $Geo_{D8A,H582A}$ (GeodCas9) variants. For this purpose, we repeated the transformations of the *E. coli* DH10B_*gfp*(pAcr) strain with the pTCas9_Sp3 and pGCas9_Sp1 and 2 plasmids. We simultaneously constructed variations of the same plasmids containing the *thermodcas9* and *geodCas9* genes instead of the active variants and transformed them to *E. coli* DH10B_*gfp*. We then used the transformation mixtures to inoculate media with a variety of IPTG concentrations, aiming to achieve different induction levels of ThermoCas9, GeoCas9, ThermodCas9 and GeodCas9 expression (**Figure 8.4**), and we executed flow cytometry-based fluorescence loss assays. The *E. coli* DH10B(pUC19, pACYC184) and *E. coli* DH10B(pACYC184) strains were employed as the no-fluorescence control strains of the assays, while the *E. coli* DH10B_*gfp*(pUC19, pACYC184), *E. coli* DH10B_*gfp*(pACYC184), *E. coli* DH10B_*gfp*(pAcr, pTCas9_Spscr), *E. coli* DH10B_*gfp*(pAcr, $pTCas9_{L551H}$_Spscr), *E. coli* DH10B_*gfp*(pAcr, pGCas9-_Spscr), *E. coli* DH10B_*gfp*(pTdCas9_Spscr), and *E. coli* DH10B_*gfp*(pGdCas9_Spscr) strains were used as the fluorescence control strains.

The results of the flow cytometry-based fluorescence loss assays revealed the transcriptional silencing effect of the ThermoCas9-AcrIIC1N_{me} complexes. The fluorescence intensity of the *E. coli* DH10B_*gfp*_AcrIIC1$_{Nme}$ cells transformed with the pTCas9_Sp3 plasmid, under non-induced conditions, was ~77% less than the detected fluorescence intensity for the *E. coli* DH10B_*gfp*(pAcr, pTCas9_Spscr) control strain (**Figure 8.3A**, **S8.6A**). When ThermoCas9 expression was induced with 1 - 50 μM IPTG, a gradual decrease of the GFP signal was observed, while IPTG concentrations ≥ 50 μM resulted in complete loss of fluorescence. Similar results were obtained when the *E. coli* DH10B_*gfp*(pAcr) cells were transformed with the $pTCas9_{L551H}$_Sp3 plasmid, suggesting that the stability of the ThermoCas9-AcrIIC1$_{Nme}$ complex is not significantly impaired by the lack of a positively charged amino-acid at the residue-position 551 of the protein. In comparison, after transformation of *E. coli* DH10B_*gfp* cells with the pTdCas9_Sp3 plasmid under non-induced conditions, absolute fluorescence loss was observed (**Figure 8.3B**, **S8.8A**). Hence, the ThermoCas9-AcrIIC1$_{Nme}$, and the $ThermoCas9_{L551H}$-AcrIIC1$_{Nme}$ complexes mediated silencing of the *gfp* gene with the same efficiency as ThermodCas9 at higher induction levels, while at lower induction levels there was a narrow window of gradual repression.

The results of the flow cytometry-based fluorescence loss assays based on the GeoCas9-related constructs revealed more tunable transcriptional silencing effects compared to the ThermoCas9-related constructs. After transformation of *E. coli* DH10B_*gfp*(pAcr) cells with the pGCas9_Sp1 and 2 plasmids separately and in absence of IPTG-based GeoCas9 expressional induction, the detected fluorescence intensity was reduced by ~21% (spacer 1) and ~29% (spacer 2) compared to the detected fluorescence intensity for the *E. coli* DH10B_*gfp*(pAcr, pGCas9_Spscr) control strain (**Figure 8.3C**, **S8.7A**). When GeoCas9 expression was induced with 1 - 50 μM IPTG, a steady reduction of the fluorescence intensity was observed by ~84% (spacer 1) and ~68% (spacer 2) of the fluorescence intensity obtained by *E. coli* DH10B_*gfp*(pAcr, pGCas9_Spscr) at 50 μM IPTG induction. No further fluorescence loss was observed at >50μM IPTG induction concentrations. In comparison, after transformation of *E. coli* DH10B_*gfp* with the pGdCas9_Sp1 and 2, separately and in absence of induction of GeodCas9 expression, the fluorescence intensity was decreased by ~27% (spacer 1) and ~45% (spacer 2) compared to the fluorescence intensity detected in the *E. coli* DH10B_*gfp*(pGdCas9_Spscr) control strain (**Figure 8.3D**, **S8.8B**). When the GeodCas9 expressionwas induced with 1 - 50 μM IPTG, a progressive reduction of the fluorescence intensity was observed, being reduced by ~89% (spacer 1) and ~91% (spacer 2) of the fluorescence intensity obtained by the *E. coli* DH10B_*gfp*(pGdCas9_Spscr) control

strain at 50 µM IPTG induction. No further fluorescence loss was observed at >50µM IPTG induction concentrations. Therefore, the GeoCas9-AcrIIC1$_{Nme}$ complex mediated silencing of the *gfp* gene with efficiency similar to GeodCas9 for the spacer 1. However, regarding spacer 2, GeodCas9-based silencing resulted in 20% higher loss of fluorescence intensity compared to the GeoCas9-AcrIIC1$_{Nme}$ complex-based silencing. Therefore, the silencing efficiencies of the GeoCas9-AcrIIC1$_{Nme}$ complex and the GeodCas9 were proved to be spacer-dependent but still relatively comparable.

In addition, we repeated the flow cytometry-based fluorescence loss assays for the *E. coli* DH10B_*gfp*(pAcr) strain transformed with the pTCas9_Sp3, pTCas9$_{L551H}$_Sp3, and pGCas9_Sp1 and 2 plasmids, this time inducing the expression of AcrIIC1N_{me} via L-rhamnose supplementation. Notably, no substantial difference was observed in the results of these assays compared to the assays without L-rhamnose supplementation (**Figure S8.6B**, **S8.7B**). Moreover, single cell populations were always observed for all the previously described flow cytometry-based fluorescence loss assays (**Figure S8.9**), indicating the uniform effect of the tool to all the cells. Furthermore, we repeated the described fluorescence loss assays employing a spectrophotometer-based approach obtaining similar outcomes (**Figure S8.10**, **S8.11**, **S8.12**).

ThermoCas9/GeoCas9-AcrIIC1$_{Nme}$-based targeting and targeting inhibition in *E. coli* (single-vector system)

To simplify our system by minimizing the number of employed plasmids, we designed and developed a single-plasmid system for ThermoCas9/GeoCas9-AcrCII1N_{me}-mediated targeting and targeting inhibition in *E. coli.* To this end, we added the *E. coli* codon-optimised *AcrIIC1*$_{Nme}$ gene under the transcriptional control of the L-rhamnose-inducible promoter (P$_{rha}$) to the previously used pTCas9_Sp1/2/3/scr and pGCas9_Sp1/2/3/scr. We repeated the ThermoCas9- and GeoCas9- targeting assays in the *E. coli* DH10B_*gfp* strain using the pAcr_TCas9_Sp1, pAcr_TCas9_Sp1/2/3/scr,and pAcr_GCas9_Sp1/2/3/scr constructs (**Figure 8.4A**). The pAcr_TCas9_Spscr, pAcr_GCas9_Spscr, and the pACYC184 plasmids were used as negative controls.

To assess the efficiency of the ThermoCas9 targeting inhibition using the single plasmid system, experiments similar to the targeting assays were conducted. Initially the experiments were conducted without L-rhamnose supplementation and hence without induction of the AcrIIC1$_{Nme}$ expression (**Figure 8.4B**, **S8.13A**). The transformation efficiency of the *E. coli* DH10B_*gfp* cells was reduced by an order of magnitude when transformed with the pAcr_TCas9_Sp3 plasmid without IPTG-based ThermoCas9 induction. However, when the expression of ThermoCas9 was induced by supplementing 100 µM IPTG, the transformation efficiency was null. The transformation efficiencies of the cells transformed with the pAcr_TCas9_Sp1 and pAcr_TCas9_Spscr plasmid, in the absence of ThermoCas9 induction, were similar. A progressive reduction and an absolute loss of the transformation efficiency was observed when cells transformed with the pAcr_TCas9_Sp1 plasmid were supplemented with 100 µM and 1000 µM IPTG respectively. On the contrary, no loss of transformation efficiency was detected for the cells transformed with pAcr_TCas9_Sp2 plasmid, even upon supplementation with 1000 µM IPTG. Subsequently, the whole set of experiments was repeated by supplementing the culture medium with 0.2% (w/v) L-rhamnose for induction of the AcrIIC1$_{Nme}$ expression (**Figure 8.4B**, **S8.13A**). The transformation efficiency of the cells transformed with the pAcr_TCas9_Sp3 plasmid increased, when compared to the corresponding experiments without L-rhamnose supplementation, by one order of magnitude in the absence of IPTG supplementation and by three orders of magnitude with 100 µM IPTG supplementation (**Figure 8.4B**, **S8.13A**). Interestingly, there was no loss of

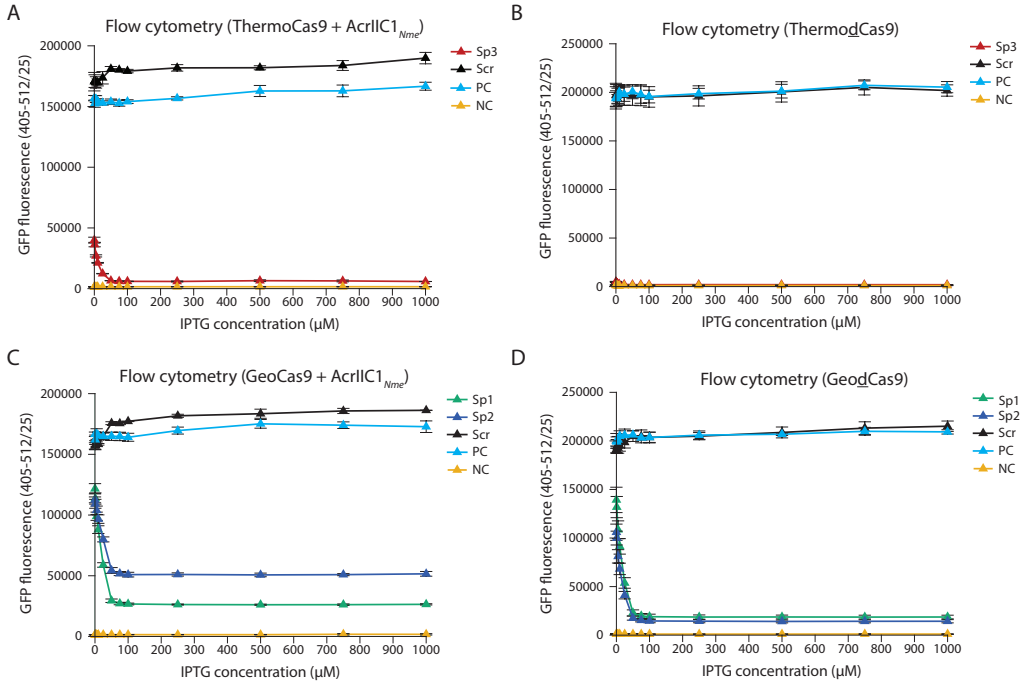


Figure 8.3 | Comparison of the ThermoCas9/GeoCas9:AcrIIC1$_{Nme}$-based and the ThermodCas9/GeodCas9-based silencing in *E. coli* DH10B_*gfp* (flow cytometry-based fluorescence loss assays). (A) ThermoCas9:AcrIIC1$_{Nme}$-mediated silencing/fluorescence downregulation assays in *E. coli* DH10B_*gfp* cells harbouring either the pTCas9_Sp3 or the pTCas9_Scr vectors. **(B)** ThermodCas9-mediated silencing/fluorescence downregulation assays in *E. coli* DH10B_*gfp* cells harbouring either the pTdCas9_Sp3 or the pTdCas9_Scr vectors. **(C)** GeoCas9:AcrIIC1$_{Nme}$-mediated silencing/fluorescence downregulation assays in *E. coli* DH10B_*gfp* cells harbouring either the pGCas9_Sp1 or the pGCas9_Sp2 or the pGCas9_Scr vectors. **(D)** GeodCas9-mediated silencing/fluorescence downregulation assays in *E. coli* DH10B_*gfp* cells harbouring either the pGdCas9_Sp1 or the pGdCas9_Sp2 or the pTdCas9_Scr vectors. The *E. coli* DH10B_*gfp* and the *E. coli* DH10B strains carrying the empty pACYC184 vector were used as the positive -fluorescence- control (PC) and the negative -no-fluorescence- control (NC) respectively for all the above described assays.

transformation efficiency for the cells transformed with pAcr_TCas9_Sp1 and supplemented with L-rhamnose, even when 1000 μM of IPTG was used for the ThermoCas9 induction.

We proceeded with the assessment of the GeoCas9-based targeting inhibition efficiency using the single plasmid system. The transformation efficiencies of *E. coli* DH10B_*gfp* cells with the pAcr_GCas9_Sp1 and 2 plasmids exhibited similar efficiencies as the transformation efficiencies of the cells transformed with the pAcr_GCas9_Spscr (negative control) in the absence of the GeoCas9 induction (**Figure 8.4C**, **S8.13B**). On the contrary, when using 100 μM and 1000 μM of IPTG for induction of the GeoCas9 expression, the transformation efficiencies dropped ~2 orders of magnitude (**Figure 8.4C**, **S8.13B**). When we repeated these experiments by supplementing the culture medium with 0.2% (w/v) L-rhamnose, for induction of the AcrIIC1$_{Nme}$ expression, the transformation efficiencies of the cells transformed with pAcr_GCas9_Sp1 and 2 were similar to the control levels (**Figure 8.4C**, **S8.13B**). These findings imply that it is possible to achieve ThermoCas9/GeoCas9-AcrIIC1$_{Nme}$- mediated targeting and targeting inhibition in a spacer-dependent and tuneable manner using a single plasmid system.

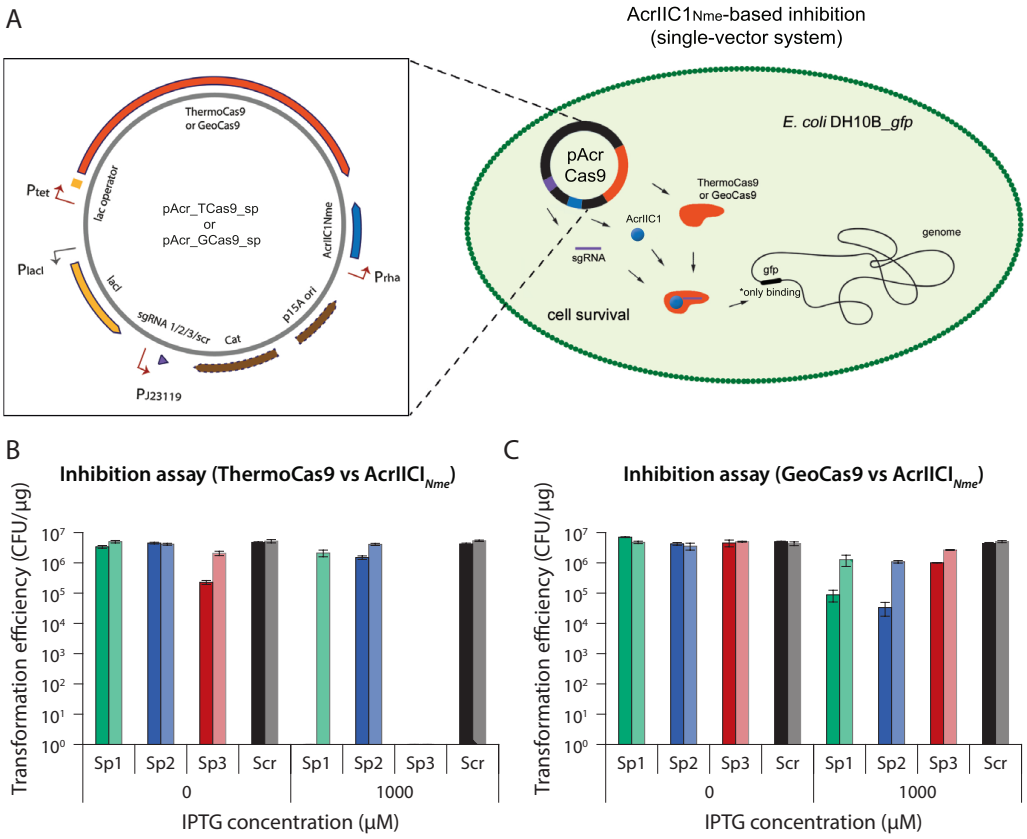


Figure 8.4 | AcrIIC1$_{Nme}$-mediated inhibition of ThermoCas9- and GeoCas9-based targeting in *E. coli* DH10B_*gfp* (single-vector system). **(A)** Schematic illustration of the single-vector strategy for AcrIIC1$_{Nme}$-mediated inhibition of ThermoCas9- and GeoCas9-based targeting in *E. coli* DH10B_*gfp*. The pAcrCas9 plasmid was used for the expression of the ThermoCas9/GeoCas9 endonuclease guided by the sgRNA to target and cleave specific sites on the genome of *E. coli* DH10B_*gfp*. The Cas9-inhibitor, AcrIIC1$_{Nme}$, was additionally expressed from the same plasmid. **(B)** Results of the ThermoCas9:AcrIIC1$_{Nme}$-based targeting-inhibition assays. **(C)** Results of the GeoCas9:AcrIIC1$_{Nme}$-based targeting-inhibition assays. The results for the assays that the AcrIIC1$_{Nme}$ expression was induced with L-rhamnose supplementation are represented with light colors.

8

Discussion

In this study, we demonstrate that the thermostable ThermoCas9 and GeoCas9 endonucleases are highly active *in vivo* at 37 ℃ and introduce lethal DSBs to the *E. coli* genome in a spacer-dependent manner. Moreover, the developed Cas9 systems had tuneable targeting efficiencies, as the general trend for efficiently targeted spacers was that upon ascending ThermoCas9 and GeoCas9 expression levels, the *E. coli* transformation efficiencies were proportionally descending. We also report that the AcrIIC1$_{Nme}$ protein efficiently hinders the DNA targeting activity of ThermoCas9 and GeoCas9, trapping them in a DNA-bound and catalytically inactive condition. It was also shown that the ThermoCas9/GeoCas9-AcrIIC1$_{Nme}$ complexes can efficiently knock-down the *in vivo* transcription of a targeted gene, in this study of the *gfp* gene, and the silencing effect was proved to be spacer-dependent, tuneable and comparable to the silencing effect of the catalytically inactive (”dead”) variants ThermodCas9 and GeodCas9, for the same

targets. Finally, we created a single-vector-based, highly efficient Cas9/AcrIIC1$_{Nme}$-based approach for tuneable bacterial genome targeting and targeting inhibition, and we envision to further explore the potential of this tool by developing a tightly controllable sequential bacterial silencing and editing tool.

The vast majority of the currently developed Cas9-based bacterial genome editing tools are based on the introduction of the desired modifications in a genome via double cross over events of plasmid-borne homologous recombination (HR) templates, followed by Cas9-based counter-selection of the non-mutated cells (*72*, *387*, *388*). The limitation of this approach, for the majority of microorganisms, is that the Cas9-based introduction of lethal DSBs often precedents the incorporation of the desired modifications in a genome, resulting in the dramatic reduction of the surviving mutant number. This can be attributed either to the lack of tightly controlled promoter systems for the microorganisms or to low native HR efficiencies. A solution employed for bacteria with more than 1 plasmid in their genetic toolbox follows a two-step approach. The cells are initially transformed with a plasmid that carries the HR flanks, then cells that have incorporated the plasmid via a single cross over (SCO) event are selected and retransformed with a second plasmid that carries the *cas9* gene and is responsible for the elimination of the non-mutated cells. In this context, a highly controllable and generally applicable Cas9 expression system would be valuable for the development of an "all-in-one plasmid" system, without simultaneously compromising the tool's efficiency. For this purpose, we aim to create a system for which the *thermocas9* or *geocas9* gene, the *AcrIIC1*$_{\text{Nme}}$ gene and the HR flanks that carry the desired modifications will be introduced into the same plasmid. In this system, the *thermocas9* or *geocas9* gene can be set under the transcriptional control of a constitutive, species-specific promoter, while a ribozyme with induced self-splicing capacity can be inserted into the selected *cas9* gene, disrupting its coding sequence. The probable expressional leakiness of the ThermoCas9 or GeoCas9 expression in the absence of the splicing inducer, which can still be enough to efficiently introduce lethal DSBs, will be counteracted by AcrIIC1$_{Nme}$, the gene of which will also be under the transcriptional control of a constitutive promoter. This system will allow for efficient introduction of the desired modifications into the genome via HR of the plasmid-borne HR template. Subsequently, the splicing inducer supplementation will result in the maturation of the *thermocas9* and *geocas9* mRNAs, leading to ThermoCas9/GeoCas9 expression levels that prevail over the antiCRISPR-mediated inhibition, allowing the endonuclease to act as a stringent counter-selection system. We plan to establish the proof of principle approach in *E. coli* and further use the tool for editing other microorganisms.

Harnessing the flexibility of the here described antiCRISPR-based technology, this system may be further programmed for base-editing purposes that enable gene inactivation in prokaryotes, without the need of repair templates or sacrifice of the transformation efficiency. Inspired by the creation of premature stop codons in the genome of *E. coli* (*420*, *421*), *Brucella melitensis* (*421*), *Staphylococcus aureus* (*422*), *Corynebacterium glutamicum* (*423*), and several *Pseudomonas* species (*294*) employing a fusion of catalytically inactivated or nickase SpCas9 and cytidine deaminase (PmCDA1; APOBEC1) that converts cytosine (C) to uracil (U) at spacer-targeted sites without DSBs, active Thermo-Cas9/GeoCas9-AcrIIC1$_{Nme}$ may be applied instead of the catalytically deficient SpCas9 variants to broaden the spectrum of DNA targets (different PAMs) and reduce the off-target effects (stringent PAM recognition at 37 ℃). As such, a ThermoCas9/GeoCas9-cytidine deaminase fusion can be generated and trapped by AcrIIC1$_{Nme}$ in a Cas9-catalytically inactive state on the DNA target, allowing the generation of premature stop codons. As an added value, our system will not require ancillary proteins, as in previous

studies, to enhance the base editing efficiency at screenable levels, as upon induction of the ThermoCas9/GeoCas9 expression at levels that surpass the AcrIIC1$_{Nme}$ counter-acting function, counter-selection in favour of the mutated cells can take efficiently place. Taking an even larger flight in the burgeoning field of genetic tool development, this antiCRISPR-based system may be further customized in a broad-range of microorganisms for multiplexed and sequential editing, silencing, gene inactivation or even base-editing in promoter sequences and small RNA-coding regions.

The demand for tightly controlled and precise CRISPR-based genome editing and regulation tools is higher than ever. The pace and scale at which novel CRISPR–Cas endonucleases are characterised, engineered and exploited as advanced molecular scissors is exponentially increasing. Newly discovered antiCRISPR proteins may serve as powerful natural "off-switches" of these nucleases, essentially diminishing the Achilles' heel of the CRISPR–Cas systems and launching the prokaryotic genome engineering field. The development of an exchangeable genome editing and regulation tool, universally applicable to academically and industrially attractive bacteria, is now closer than ever before.

Experimental Procedures

Bacterial strains and growth conditions

Bacterial strains used in this study are listed in **Table S8.3**. All *E. coli* strains were cultured in Luria-Bertani broth (LB) or on LB agar plates, supplemented with chloramphenicol (15 µg mL^{-1}), and/or ampicillin (100 µg mL^{-1}). L-rhamnose (0.2%), and/or IPTG (0 - 1000 µM) were additionally used for inducing the expression of the anti-CRISPR and ThermoCas9/GeoCas9, respectively. For the fluorescence loss and flow cytometry assays, M9TG (11.28 g 1X M9 salts, 10 g tryptone, 5 g glycerol) medium was used instead of LB.

Construction of plasmids

Plasmids and primers/oligonucleotides used in this study are presented in **Table S8.2** and **S8.4** respectively. The plasmids were constructed using the NEBuilder HiFi DNA Assembly Cloning Kit (NEB). The fragments for assembling the plasmids were obtained through PCR with Q5® High-Fidelity 2X Master Mix (NEB). The PCR products were run on a 0.8 % agarose gel and were subsequently purified using Zymogen gel DNA recovery kit (Zymo Research). The assembled plasmids were transformed to chemically competent *E. coli* DH5α cells (*424*) and plated on LB agar containing chloramphenicol (15 µg mL^{-1}) or ampicillin (100 µg mL^{-1}) and grown overnight at 37 °C The next day, single colonies were inoculated in LB medium, grown overnight at 37 °C and the plasmids were isolated using the GeneJet plasmid Miniprep kit (ThermoFisher Scientific). All the constructs were verified using Sanger sequencing (Macrogen). Description of the assembled fragments used for the construction of each plasmid is detailed in **Table S8.2**. For annealing of oligos to create dsDNA used in the plasmid assembly, 4 µL oligonucleotide pairs (100 µM each) were mixed in Milli-Q water to a final volume of 100 µl, heated at 95 ℃ for 5 min, and slowly cooled to room temperature.

Targeting assays

To target bacterial DNA, chemically competent *E. coli* DH10B cells (*424*) harbouring a genome-integrated *gfp* gene under the transcriptional control of P$_{lacUV5}$ (*E. coli* DH10B_*gfp*) were transformed (*424*) with equal amounts (3 ng) of the different plasmids. These variants include the plasmid expressing ThermoCas9 and a spacer targeting the -10 sequence of P$_{J23119}$ (pT-

Cas9_Sp1), the region immediately after the -10 sequence (pTCas9_Sp2) or the start of the *gfp* gene (pTCas9_Sp3).

As controls, 3 ng of a non-targeting plasmid (pTCas9_Spscr) and the empty vector (pACYC184) were used. Transformed cells were cultured on LB agar supplemented with chloramphenicol (15 μg mL^{-1}) and different IPTG concentrations (0 - 1000 μM) for 17 h at 37 ℃. Colony forming units (CFUs) were counted after plating undiluted biological triplicates and used for calculating the transformation efficiencies. Similar approach was used to assess the targeting efficiency for the GeoCas9 variants. The plasmids pGCas9_Sp1, pGCas9_Sp2, pGCas9_Sp3, pGCas9_Spscr were used for assessing the targeting efficiency by GeoCas9 following the same approach as described above.

Inhibition assays (dual-vector system)

L-rhamnose was added during the preparation of the chemically competent *E. coli* DH10B_*gfp* cells to ensure the induction of the expression of the $AcrIIC1_{Nme}$. To inhibit targeting of the bacterial genomic DNA using the dual-vector system, *E. coli* DH10B_*gfp* cells(pAcr) were transformed (*424*) with equal amounts (5 ng) of plasmids expressing a spacer that targets the start of the *gfp* gene along with an either wild-type ThermoCas9 (pTCas9_Sp3) or the mutant ThermoCas9$_{L551H}$ (pTCas9$_{L551H}$_Sp3). As controls, 5 ng of non-targeting plasmids (pTCas9_Spscr, pTCas9$_{L551H}$_Spscr) and the empty vectors (pACYC184, pUC19) were used. In the cases, where the inducers were used (IPTG, L-rhamnose), they were supplemented in the recovery medium after the heat-shock. Transformed cells were grown on LB agar supplemented with ampicillin (100 μg mL^{-1}), chloramphenicol (15 μg mL^{-1}), different IPTG concentrations (0 - 1000 μM), and L-rhamnose (0 or 0.2%) for 17 h at 37 ℃. Colony forming units (CFUs) were counted after plating undiluted biological triplicates and used for calculating the transformation efficiencies. Similar approach as above was used to assess the target inhibition efficiency for the GeoCas9 variants using the pGCas9_Sp1, pGCas9_Sp2, pGCas9_Sp3, pGCas9_Spscr plasmids.

Spectrophotometry-based fluorescence loss assays

To quantify the fluorescence loss, *E. coli* DH10B_*gfp* cells harbouring an $AcrIIC1_{Nme}$-expressing plasmid (pAcr) were transformed (*424*) with equal amounts (5 ng) of plasmids expressing a spacer that targets the start of the *gfp* gene and either wild-type ThermoCas9 (pTCas9_Sp3) or the mutant ThermoCas9$_{L551H}$ (pTCas9$_{L551H}$_sp 3). As positive control, 5 ng of the empty vectors (pACYC184, pUC19) were used, while as negative controls 5 ng of the non-targeting plasmids (pTCas9_Spscr, pTCas9$_{L551H}$_Spscr) and commercial *E. coli* DH10B cells (NEB) electroporated with 100 ng of empty vectors (pACYC184, pUC19), were used. Post-transformation, 2 μL of recovered cells were cultured in 198 μL M9TG containing ampicillin (100 μg mL^{-1}) and chloramphenicol (15 μg mL^{-1}) in a Masterblock® 96-well deep microplate (Greiner Bio-One) for 22 h at 37 ℃ with vigorous shaking (900 rpm). The second day, 2 μL of the overnight cultures were diluted in 198 μL M9TG containing the same antibiotics, and 2 μL of these were re-diluted in 198 μL M9TG with the antibiotics along with the appropriate concentrations of IPTG (0 - 1000 μM) and L-rhamnose (0 or 0.2%) in a Masterblock® 96-well deep microplate (Greiner Bio-One), and incubated for 22 h at 37 ℃ with vigorous shaking (900 rpm). The third day, 40 μL of overnight cultures were diluted in 160 μL 1X PBS (8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4 $_2H_2O$, 0.24 g KH_2PO_4; pH=6.8) in a Greiner CELLSTAR® 96-well plate (Greiner Bio-One), and 100 μL of these dilutions were transferred 96-well clear bottom black microplate (CorningTM). The OD_{600}s and the fluorescence signal (GFP:395,508) were measured with a SynergyMx Monochromator-Based Multi-Mode Microplate reader (BioTek). Similar approach

as above was used to assess the target inhibition efficiency for the GeoCas9 variants using the pGCas9_Sp1, pGCas9_Sp2, pGCas9_Sp3, pGCas9_Spscr plasmids.

For the comparison of the silencing of the GFP using the Thermo/GeoCas9:AcrIIN$_{me}$ to that of the dead variants, *E. coli* DH10B_*gfp* cells were heat-transformed (*424*) with equal amounts (3 ng) of the different plasmids. These variants include the plasmid expressing Thermo_dCas9 and a spacer targeting the -10 sequence of P_{lacUV5} (pTdCas9_Sp1), the region immediately after the -10 sequence (pTdCas9_Sp2) or the start of the *gfp* gene (pTdCas9_Sp3). As a positive control, 3 ng of empty vector (pACYC184) was used while as negative controls 3 ng of the non-targeting plasmid (pTdCas9_Spscr) and commercial *E. coli* DH10B cells (NEB) harbouring the empty vector (pACYC184) were used. Post-transformation, 2 µL of the recovered cells were cultured in 198 µL M9TG containing chloramphenicol (15 µg mL^{-1}) in a Masterblock® 96-well deep microplate (Greiner Bio-One) for 22 h at 37 ℃ with vigorous shaking (900 rpm). The second day, 2 µL of overnight cultures were diluted in 198 µL M9TG containing the same antibiotic, and 2 µL of these were re-diluted in 198 µL M9TG with antibiotic and IPTG (0 - 1000 µM) in a Masterblock® 96-well deep micro-plate (Greiner Bio-One), and incubated for 22 h at 37 ℃ with vigorous shaking (900 rpm). The third day, 40 µL of overnight cultures were diluted in 160 µL 1X PBS in a Greiner CELLSTAR® 96-well plate (Greiner Bio-One), and 100 µL of these dilutions were transferred 96-well clear bottom black microplate (CorningTM). The OD_{600}s and the fluorescence signal (GFP:395,508) of the cultures were measured with a SynergyMx Monochromator-Based Multi-Mode Micro-plate reader (BioTek). The plasmids, pGeodCas9_Sp1, pGeodCas9_Sp2, pGeodCas9_Sp3, pGeodCas9_Spscr were used for assessing the fluorescence loss by GeodCas9 following the same approach as described above.

Flow cytometry-based fluorescence loss assays

To verify the levels of fluorescence loss and investigate the presence of sub-populations in all samples, the same protocol as the fluorescence loss assays was followed, with the only difference that on the third day 2 µL of overnight cultures were diluted in 998 µL 1X PBS in a Masterblock® 96-well deep microplate (Greiner Bio-One). The fluorescence signal and the presence of subpopulations were examined using the Attune NxT Flow Cytometer device (Thermo Fisher Scientific) (GFP intensity 405-512/25 of 30,000 single cells per sample).

Inhibition assays (single-vector system)

To inhibit targeting of the genomic DNA using the single-vector system, *E. coli* DH10B_*gfp* cells were transformed (*424*) with equal amounts (5 ng) of plasmids expressing AcrIIC1$_{Nme}$, Thermo-Cas9, and a spacer that targets the -10 sequence of P_{lacUV5} (pAcr_TCas9_Sp1), the region immediately after the -10 sequence (pAcr_TCas9_Sp2) or the start of the *gfp* gene (pAcr_TCas9_Sp3). As controls, 5 ng of a non-targeting plasmid (pAcr_TCas9_Spscr) and the empty vector (pACYC184) were used. Transformed cells were grown on LB agar supplemented with chloramphenicol (15 µg mL^{-1}), different IPTG concentrations (0 - 1000 µM), and L-rhamnose (0 or 0.2%) for 17 h at 37 ℃. In the cases, where inducers were used (IPTG, L-rhamnose), they were supplemented in the recovery media after transformation. CFUs were counted after plating undiluted biological triplicates and used for calculating the transformation efficiencies. The single plasmids, pAcr_GCas9_Sp1, pAcr_GCas9_Sp2, pAcr_GCas9_Sp3, pAcr_GCas9_Spscr were used for assessing the inhibition by GeoCas9 following the same approach as described above.

Acknowledgements

We would like to thank Thijs Nieuwkoop for his technical assistance with the flow cytometer. I.M. is supported by the Dutch Technology Foundation STW, which is part of The Netherlands Organization for Scientific Research (NWO) and which is partly funded by the Ministry of Economic Affairs. J.v.d.O. and P.M. are supported by the NWO/TOP grant 714.015.001.

Author contributions

D.T., P.M., R.v.K., I.M. and J.v.d.O. conceived this study and design of experiments. D.T. and P.M. conducted the experiments. R.v.K., I.M. and J.v.d.O. supervised this project. D.T., P.M., R.v.K., I.M. and J.v.d.O. wrote the manuscript.

Author information

Correspondence should be addressed to john.vanderoost@wur.nl.

Competing interests

R.v.K. is employed by the commercial company Corbion (Gorinchem, The Netherlands).

Supplementary Figures and Tables

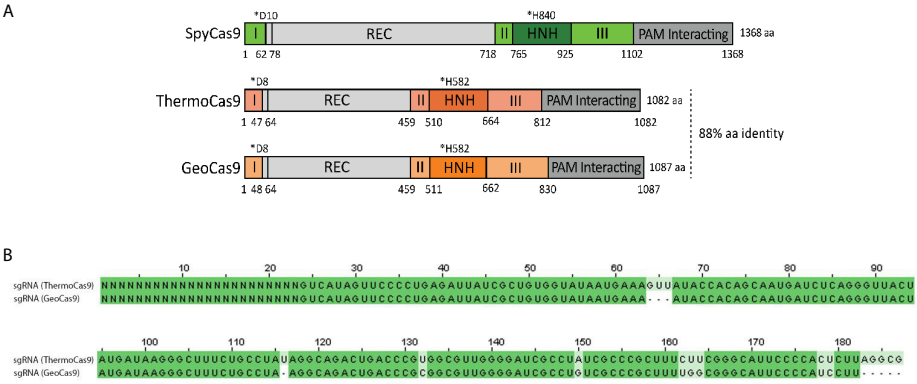


Figure S8.1 | Comparison of the domain architecture and the sgRNA modules of ThermoCas9 and GeoCas9. **(A)** Schematic illustration of the domain architecture of SpyCas9 (green), ThermoCas9 (dark orange), and GeoCas9 (light orange). The active site residues are indicated with asterisks. **(B)** Sequence alignment of the sgRNA modules of ThermoCas9 and GeoCas9, used in previous studies for in vivo experiments. Dark green colour indicates identical nucleotides, while light green colour represents non-identical regions

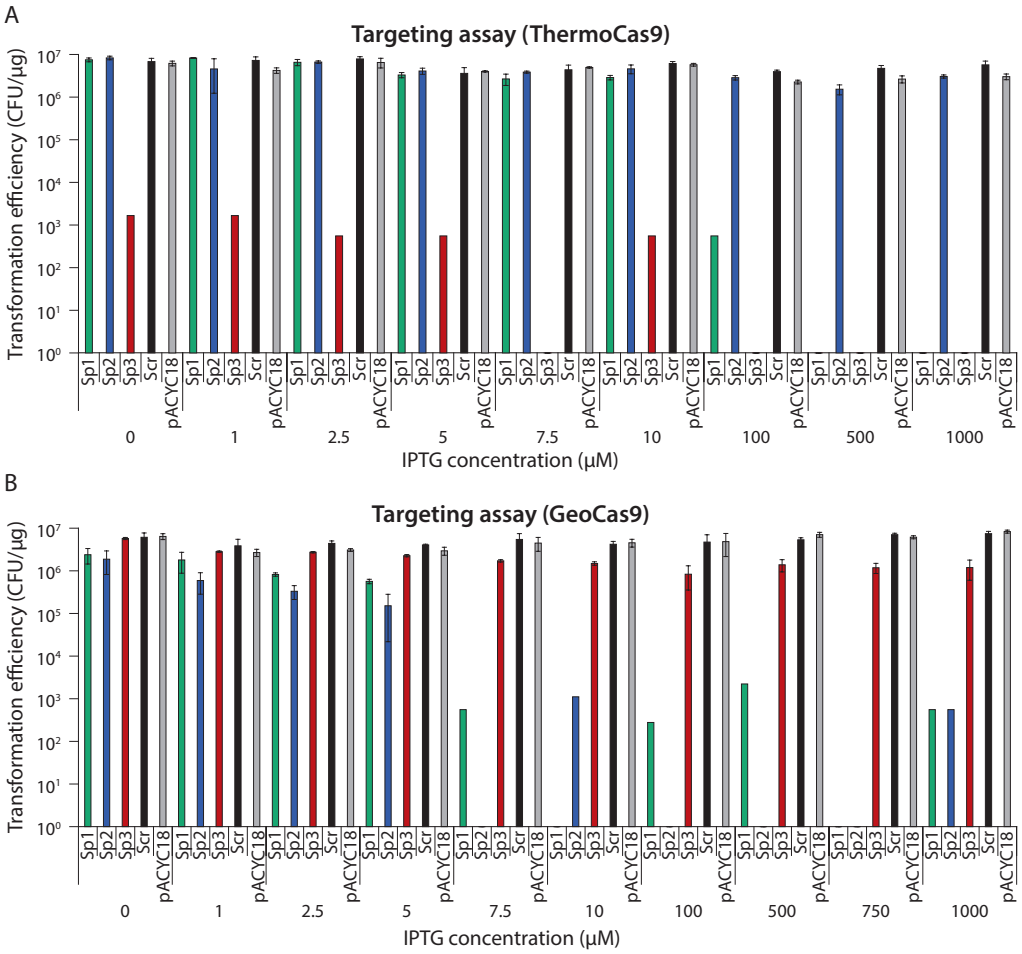


Figure S8.2 | Comparison of ThermoCas9- and GeoCas9-based targeting in *E. coli* DH10B_*gfp*. (A) Complete data for ThermoCas9-based targeting in *E. coli* DH10B_*gfp* including the whole range of IPTG concentrations used for the induction of the ThermoCas9 expression. **(B)** Complete data for GeoCas9-based targeting in *E. coli* DH10B_*gfp* including the whole range of IPTG concentrations used for the induction of the GeoCas9 expression.

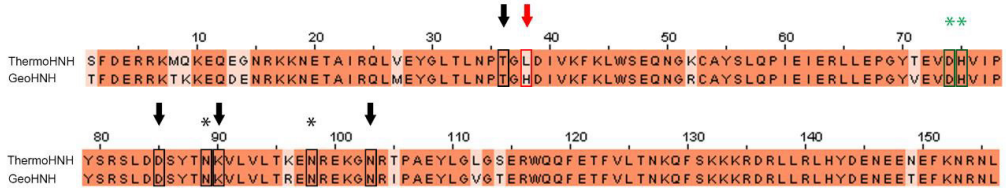


Figure S8.3 | Amino acid sequence alignment of the HNH domain of ThermoCas9 and GeoCas9. Dark orange colour indicates identical amino acid, while light orange colour represents non-identical residues. Asterisks indicate catalytic residues, and, specifically, green asterisks represent catalytic residues that interact with AcrIIC1$_{Nme}$. Arrows indicate surrounding residues possibly involved in the stabilization of the association of AcrIIC1$_{Nme}$ with the GeoCas9 protein. The alignment was performed using ClustalO (default settings) and the results were visualized via Jalview.

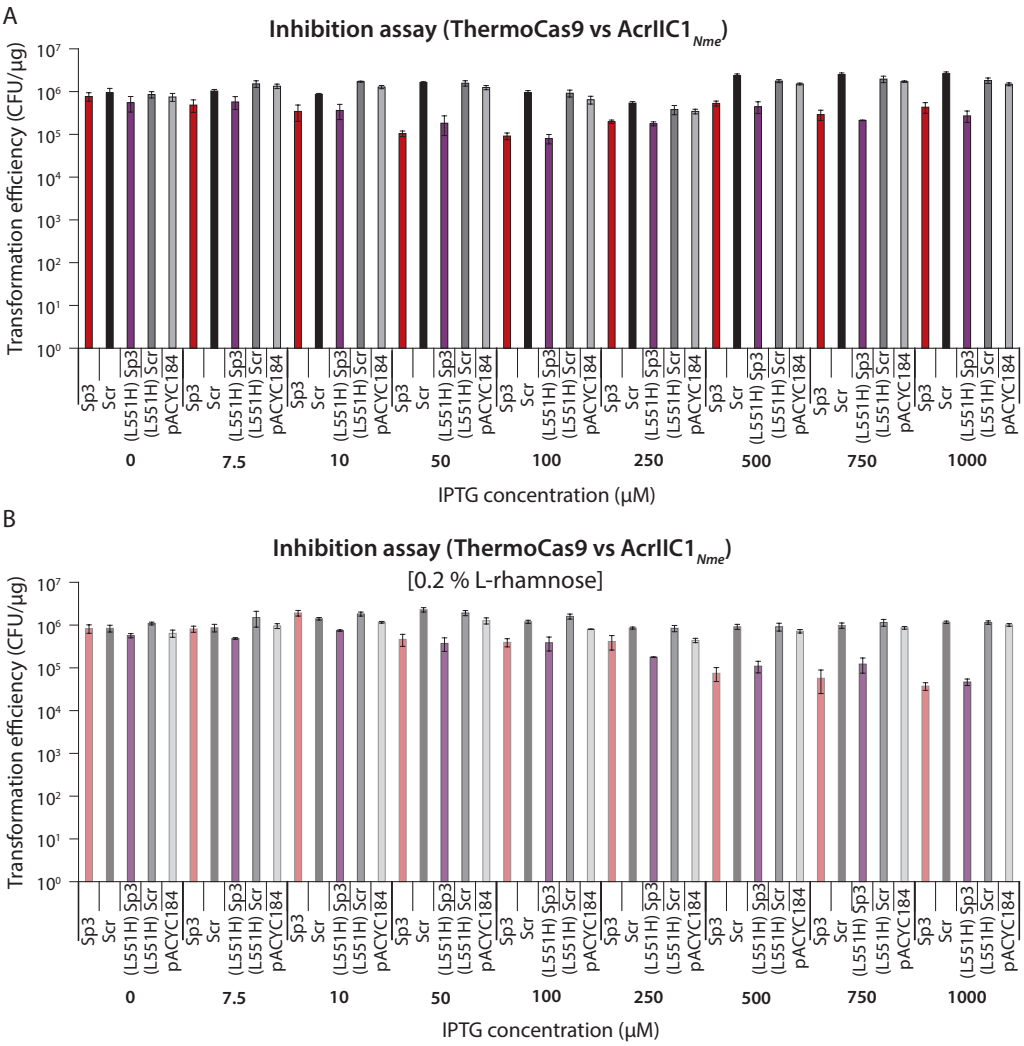


Figure S8.4 | AcrIIC1$_{Nme}$-mediated inhibition of ThermoCas9 and ThermoCas9$_{L551H}$ in *E. coli* DH10B_*gfp* (dual-vector system). (A) Complete data for inhibition of ThermoCas9 and ThermoCas9$_{L551H}$ cleavage activity, when the expression of AcrIIC1$_{Nme}$ is not induced (no supplementation of L-rhamnose). **(B)** Complete data for inhibition of ThermoCas9 and ThermoCas9$_{L551H}$ cleavage activity, when the expression of Acr$_{Nme}$ is induced (supplementation of 0.2% L-rhamnose).

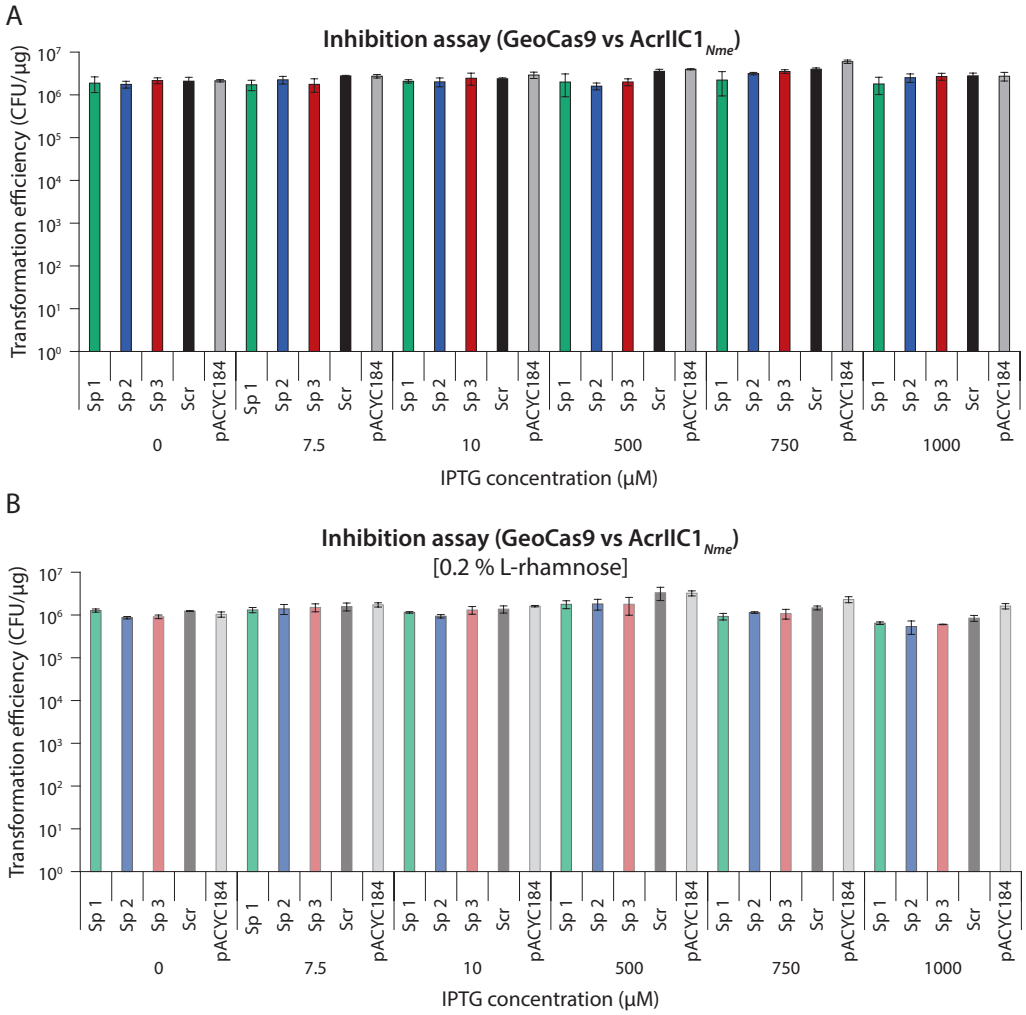


Figure S8.5 | AcrIIC1$_{Nme}$-mediated inhibition of GeoCas9 in *E. coli* DH10B_*gfp* (dual-vector system). (A) Complete data for inhibition of GeoCas9 cleavage activity, when the expression of AcrIIC1$_{Nme}$ is not induced (no supplementation of L-rhamnose). **(B)** Complete data for inhibition of GeoCas9 cleavage activity, when the expression of AcrIIC1$_{Nme}$ is induced (supplementation of 0.2% L-rhamnose).

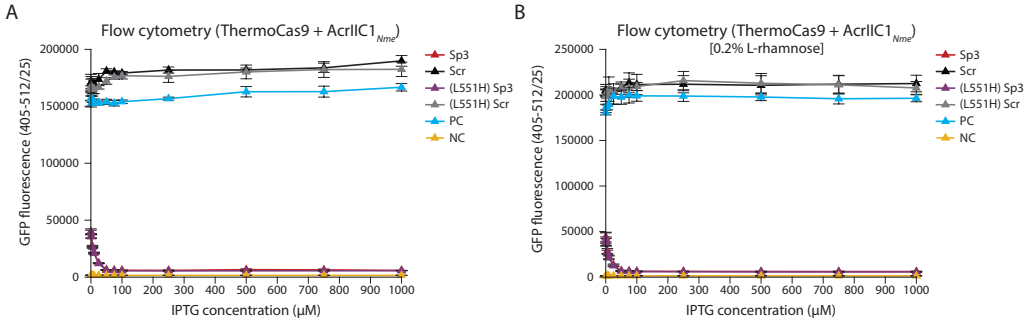


Figure S8.6 | ThermoCas9/ThermoCas9$_{L551H}$:AcrIIC1$_{Nme}$-based silencing in *E. coli* DH10B_*gfp* (flow cytometry-based fluorescence loss assays). (A) Complete data for ThermoCas9/ThermoCas9$_{L551H}$–mediated silencing, when the expression of AcrIIC1$_{Nme}$ is not induced (no supplementation of L-rhamnose). **(B)** Complete data for ThermoCas9/ThermoCas9$_{L551H}$–mediated silencing, when the expression of AcrIIC1$_{Nme}$ is induced (supplementation of 0.2% L-rhamnose).

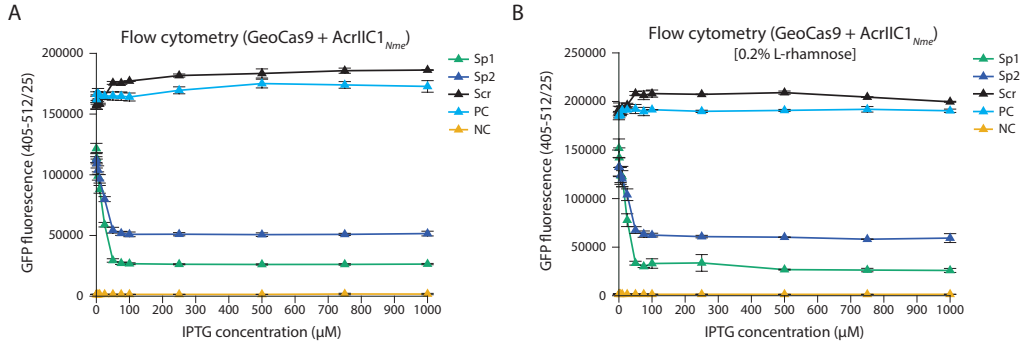


Figure S8.7 | GeoCas9:AcrIIC1$_{Nme}$-based silencing in *E. coli* DH10B_*gfp* (flow cytometry-based fluorescence loss assays).. (A) Complete data for GeoCas9-mediated silencing, when the expression of AcrIIC1$_{Nme}$ is not induced (no supplementation of L-rhamnose). **(B)** Complete data for GeoCas9-mediated silencing, when the expression of Acr$_{Nme}$ is induced (supplementation of 0.2% L-rhamnose).

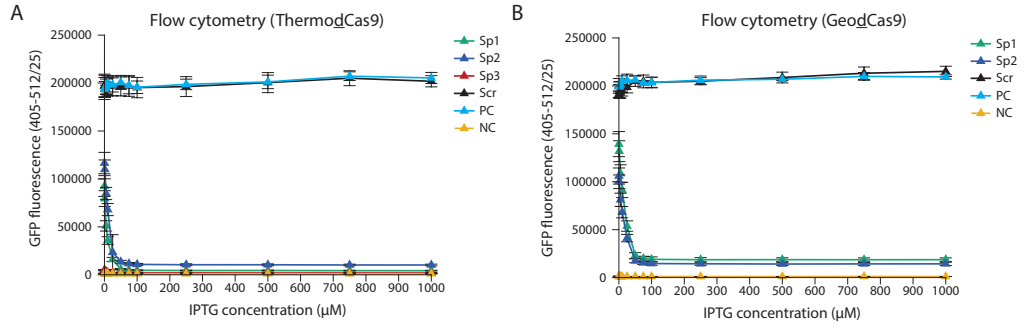


Figure S8.8 | ThermodCas9 and GeodCas9-based silencing in *E. coli* DH10B_*gfp* (flow cytometry-based fluorescence loss assays). (A) Complete data for ThermodCas9-mediated silencing in *E. coli* DH10B_*gfp*. **(B)** Complete data for GeodCas9-mediated silencing in *E. coli* DH10B_*gfp*

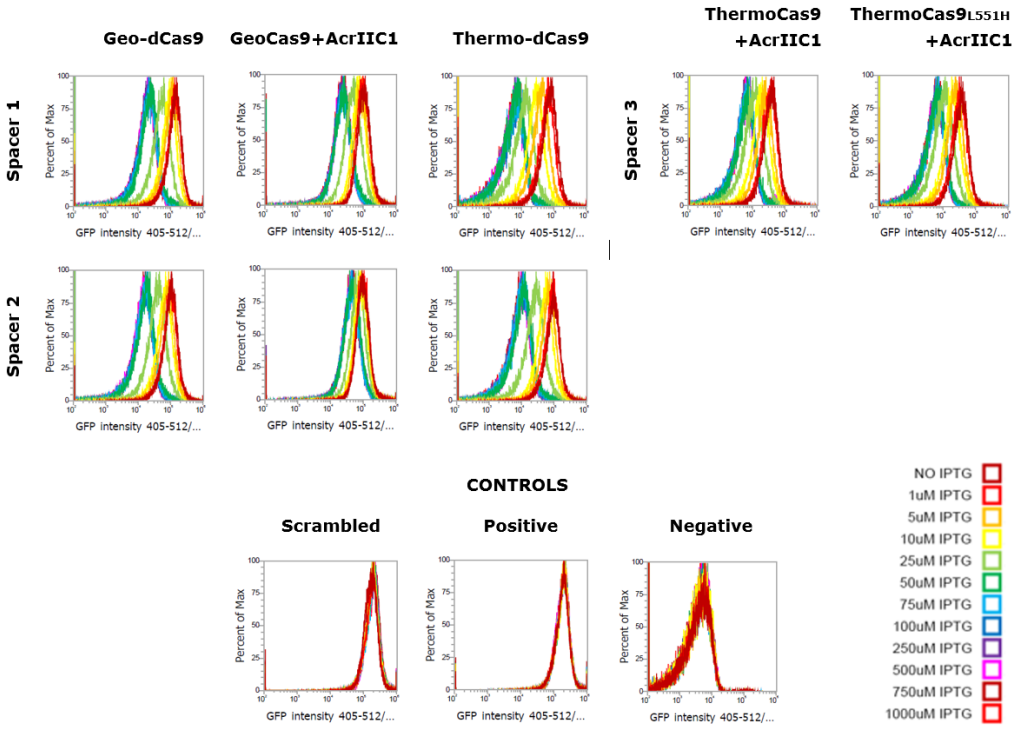


Figure S8.9|GFP intensity (405-512/25) curves indicate single populations in all tested samples (flow cytometry-based fluorescence loss assays). Different colours of the peaks indicate different IPTG concentrations for the induction of the expression of GeoCas9, ThermoCas9, and ThermoCas9$_{L551H}$. All graphs are representative of biological triplicates.

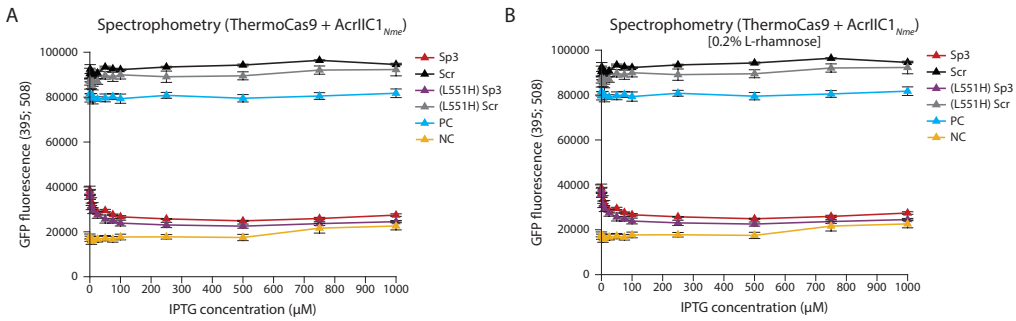


Figure S8.10|ThermoCas9/ThermoCas9$_{L551H}$:AcrIIC1$_{Nme}$-based silencing in *E. coli* DH10B_*gfp* (spectrophotometry-based fluorescence loss assays). (A) Complete data for ThermoCas9/ThermoCas9$_{L551H}$–mediated silencing, when the expression of Acr$_{Nme}$ is not induced (no supplementation of L-rhamnose). **(B)** Complete data for ThermoCas9/ThermoCas9$_{L551H}$–mediated silencing, when the expression of Acr$_{Nme}$ is induced (supplementation of 0.2% L-rhamnose)

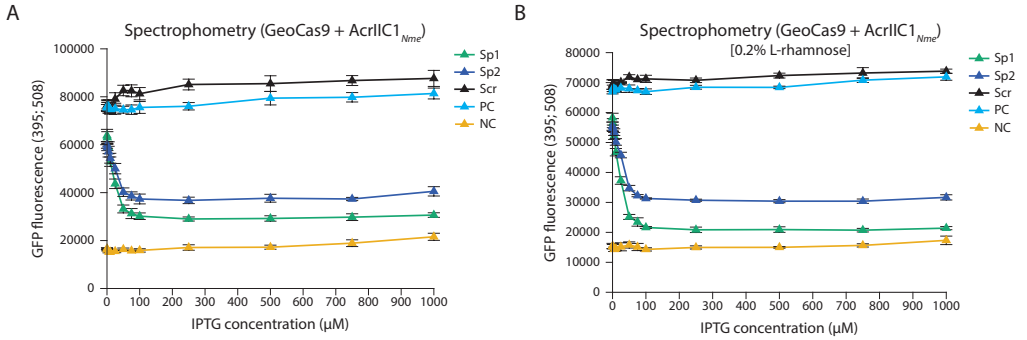


Figure S8.11 | GeoCas9:AcrIIC1$_{Nme}$-based silencing in *E. coli* DH10B_*gfp* (spectrophotometry-based fluorescence loss assays). (A) Complete data for GeoCas9-mediated silencing, when the expression of AcrIIC1$_{Nme}$ is not induced (no supplementation of L-rhamnose). **(B)** Complete data for GeoCas9-mediated silencing, when the expression of AcrIIC1$_{Nme}$ is induced (supplementation of 0.2% L-rhamnose).

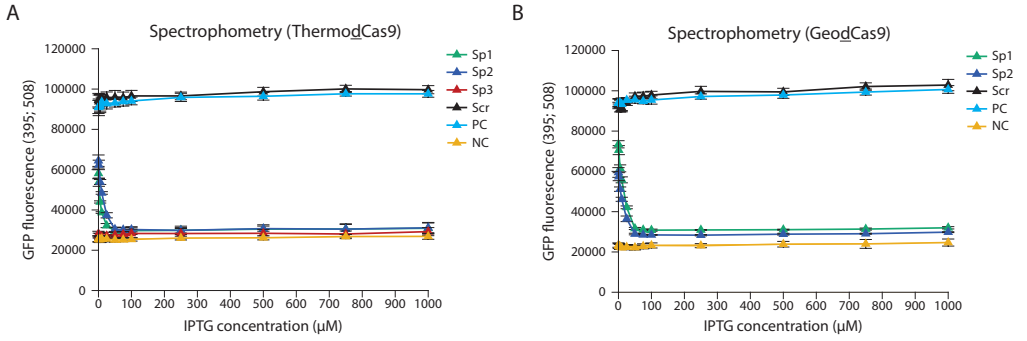


Figure S8.12 | ThermodCas9 and GeodCas9-based silencing in *E. coli* DH10B_*gfp* (spectrophotometry-based fluorescence loss assays). (A) Complete data for ThermodCas9-mediated silencing in *E. coli* DH10B_*gfp* **(B)** Complete data for GeodCas9-mediated silencing in *E. coli* DH10B_*gfp*.

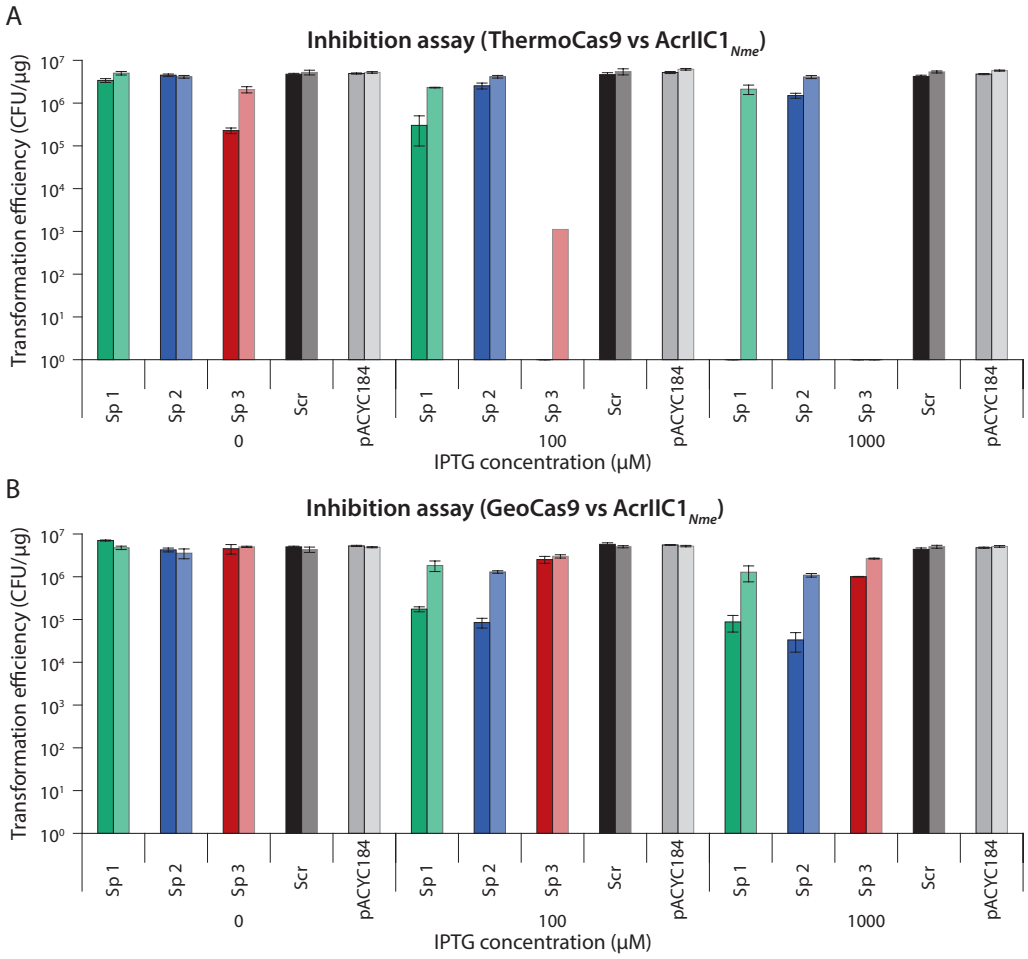


Figure S8.13 | AcrIIC1$_{Nme}$-mediated inhibition of ThermoCas9 and GeoCas9 in *E. coli* DH10B_*gfp* (single-vector system). (A) Complete data for inhibition of ThermoCas9 cleavage activity, when the expression of AcrIIC1$_{Nme}$ is either not induced (no supplementation of L-rhamnose represented in dark colors) or induced (supplementation of 0.2% L-rhamnose represented in light colors). **(B)** Complete data for inhibition of GeoCas9 cleavage activity, when the expression of AcrIIC1$_{Nme}$ is either not induced (no supplementation of L-rhamnose represented in dark colors) or induced (supplementation of 0.2% L-rhamnose represented in light colors)

Table S8.1 | Spacers used in this study for ThermoCas9, ThermoCas9$_{L551H}$, and GeoCas9

Spacer ID	Target strand	Targeting location in P_{lacUV5} (5' -> 3')	PAM (5' -> 3')	Sequence (23 nt) (5' -> 3')
2	-	60 - 38 (P_{lacUV5})	NNNNCACA	TAGAGGGAAACCGTTGTGGTCTC
1	+	16 - 38 (P_{lacUV5})	NNNNCACA	TCCGGCTCGTATAATGTGTGGGG
3	-	82 - 60 (*gfp* gene)	NNNNCCAA	AGAATTTATGCCCATTCACATCC
scr	X	X	X	CTAGATCCGCAGTAACCCCATGG

Table S8.2 | Plasmids used in this study.

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pACYC184	-	-	Negative control in targeting assays, inhibition assays, spectrophotometry- and flow cytometry-based fluorescence loss assays	Lab stock
pUC19	-	-	Negative control in targeting assays, inhibition assays, spectrophotometry- and flow cytometry-based fluorescence loss assays. PCR template for pAcrIIC1$_{Nme}$	Lab stock
pET_MBP_Geo_st	-	-	PCR template for pGCas9_Sp1, pGCas9_Sp2, pGCas9_Sp3, pGCas9_Spscr	[63]
pP_{tet}_Acr_P_{rha}_TCas9	-	-	PCR template for pGCas9_Sp1, pGCas9_Sp2, pGCas9_Sp3, pGCas9_Spscr, pTCas9_Sp1, pTCas9_Sp2, pTCas9_Sp3, pTCas9_Spscr	Lab stock
pP_{tet}_Acr_P_{rha}_TCas9$_{L551H}$	-	-	PCR template for pTCas9$_{L551H}$_Sp3, pTCas9$_{L551H}$_Spscr	Lab stock
pP_{rha}_TCas9_Sp1	-	-	PCR template for pGCas9_Sp1	Lab stock
pP_{rha}_TCas9_Sp2	-	-	PCR template for pGCas9_Sp2	Lab stock
pP_{rha}_TCas9_Sp3	-	-	PCR template for pGCas9_Sp3	Lab stock
pP_{rha}_TCas9_Spscr	-	-	PCR template for pGCas9_Spscr	Lab stock
pP_{rha}_GCas9_Sp1	-	-	PCR template for pAcr	Lab stock
pMK-RQ_Acr_coK12_coT12	-	-	PCR template for pAcr	Lab stock
pThermoCas9i	-	-	PCR template for pTdCas9_Sp1, pTdCas9_Sp2, pTdCas9_Sp3, and pTdCas9_Spscr	[62]

Table S8.2 | Plasmids used in this study. (continued)

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pGCas9_Sp1	Gibson assembly	Fragment 1: *geocas9* gene from pET_MBP_Geo_st amplified with primers BG13307 and BG13310 Fragment 2: 5' UTR, lac *operator*, P_{tet}, P_{lacI}, and *lacI* gene from pP_{tet}_Acr_P_{rha}_TCas9 amplified with primers BG13311 and BG13312 Fragment 3: terminator, sgRNA 1, terminator, backbone, and terminator from pP_{rha}_TCas9_Sp1 amplified with primers BG13313 and BG13309	GeoCas9 and sgRNA (spacer 1) targeting the genome-integrated P_{lacUV5} in targeting assays, inhibition assays (dual-vector system), spectophotometry- and flow cytometry-based fluorescence loss assays. Moreover, PCR template for pTCas9_Sp1, pGdCas9_Sp1, and pAcr_GCas9_Sp1.	This study
pGCas9_Sp2	Gibson assembly	Fragment 1: *geocas9* gene from pET_MBP_Geo_st amplified with primers BG13307 and BG13310 Fragment 2: 5' UTR, lac *operator*, P_{tet}, P_{lacI}, and *lacI* gene from pP_{tet}_Acr_P_{rha}_TCas9 amplified with primers BG13311 and BG13312 Fragment 3: terminator, sgRNA 2, terminator, backbone, and terminator from pP_{rha}_TCas9_Sp2 amplified with primers BG13313 and BG13309	GeoCas9 and sgRNA (spacer 2) targeting the genome-integrated P_{lacUV5} in targeting assays, inhibition assays (dual-vector system), spectophotometry- and flow cytometry-based fluorescence loss assays. Moreover, PCR template for pTCas9_Sp2, pGdCas9_Sp2, and pAcr_GCas9_Sp2.	This study
pGCas9_Sp3	Gibson assembly	Fragment 1: *geocas9* gene from pET_MBP_Geo_st amplified with primers BG13307 and BG13310 Fragment 2: 5' UTR, lac *operator*, P_{tet}, P_{lacI}, and *lacI* gene from pP_{tet}_Acr_P_{rha}_TCas9 amplified with primers BG13311 and BG13312 Fragment 3: terminator, sgRNA 3, terminator, backbone, and terminator from pP_{rha}_TCas9_Sp3 amplified with primers BG13313 and BG13309	GeoCas9 and sgRNA (spacer 3) targeting the genome-integrated P_{lacUV5} in targeting assays, inhibition assays (dual-vector system), spectophotometry- and flow cytometry-based fluorescence loss assays. Moreover, PCR template for pTCas9_Sp3, pTCas9$_{L551H}$_Sp3, pGdCas9_Sp3, and pAcr_GCas9_Sp3.	This study
pGCas9_Spscr	Gibson assembly	Fragment 1: *geocas9* gene from pET_MBP_Geo_st amplified with primers BG13307 and BG13310 Fragment 2: 5' UTR, lac *operator*, P_{tet}, P_{lacI}, and *lacI* gene from pP_{tet}_Acr_P_{rha}_TCas9 amplified with primers BG13311 and BG13312 Fragment 3: terminator, sgRNA scr, terminator, backbone, and terminator from pP_{rha}_TCas9 _Spscr amplified with primers BG13313 and BG13309	GeoCas9 and sgRNA (spacer scr) non-targeting the genome-integrated P_{lacUV5} in targeting assays, inhibition assays (dual-vector system), spectophotometry- and flow cytometry-based fluorescence loss assays. Moreover, PCR template for pTCas9_Spscr, pTCas9$_{L551H}$_Spscr, pGdCas9_Spscr, and pAcr_GCas9_Spscr.	This study

Table S8.2 | Plasmids used in this study. (continued)

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pTCas9_Sp1	Gibson assembly	Fragment 1: 5'UTR, lac *operator*, P_{tet}, P_{lacI}, *lacI* gene, terminator, sgRNA 1, terminator, backbone, and terminator from pGCas9_Sp1 amplified with primers BG13459 and BG13460 Fragment 2: *thermocas9* gene from pP_{tet}_Acr_P_{rha}_TCas9 amplified with primers BG13461 and BG13462	ThermoCas9 and sgRNA (spacer 1) targeting the genome-integrated P_{lacUV5} in targeting assays, inhibition assays (dual-vector system), spectophotometry- and flow cytometry-based fluorescence loss assays. PCR template for pTdCas9_Sp1, and pAcr_TCas9_Sp1.	This study
pTCas9_Sp2	Gibson assembly	Fragment 1: 5'UTR, lac *operator*, P_{tet}, P_{lacI}, *lacI* gene, terminator, sgRNA 2, terminator, backbone, and terminator from pGCas9_Sp2 amplified with primers BG13459 and BG13460 Fragment 2: *thermocas9* gene from pP_{tet}_Acr_P_{rha}_TCas9 amplified with primers BG13461 and BG13462	ThermoCas9 and sgRNA (spacer 2) targeting the genome-integrated P_{lacUV5} in targeting assays, inhibition assays (dual-vector system), spectophotometry- and flow cytometry-based fluorescence loss assays. PCR template for pTdCas9_Sp2, and pAcr_TCas9_Sp2.	This study
pTCas9_Sp3	Gibson assembly	Fragment 1: 5'UTR, lac *operator*, P_{tet}, P_{lacI}, *lacI* gene, terminator, sgRNA 3, terminator, backbone, and terminator from pGCas9_Sp3 amplified with primers BG13459 and BG13460 Fragment 2: *thermocas9* gene from pP_{tet}_Acr_P_{rha}_TCas9 amplified with primers BG13461 and BG13462	ThermoCas9 and sgRNA (spacer 3) targeting the genome-integrated P_{lacUV5} in targeting assays, inhibition assays (dual-vector system), spectophotometry- and flow cytometry-based fluorescence loss assays. PCR template for pTdCas9_Sp3, and pAcr_TCas9_Sp3.	This study
pTCas9_Spscr	Gibson assembly	Fragment 1: 5'UTR, lac *operator*, P_{tet}, P_{lacI}, *lacI* gene, terminator, sgRNA scr, terminator, backbone, and terminator from pGCas9_Spscr amplified with primers BG13459 and BG13460 Fragment 2: *thermocas9* gene from pP_{tet}_Acr_P_{rha}_TCas9 amplified with primers BG13461 and BG13462	ThermoCas9 and sgRNA (spacer scr) non-targeting the genome-integrated P_{lacUV5} in targeting assays, inhibition assays (dual-vector system), spectophotometry- and flow cytometry-based fluorescence loss assays. PCR template for pTdCas9_Spscr, and pAcr_TCas9_Spscr.	This study

Table S8.2 | Plasmids used in this study. (continued)

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pAcr	Gibson assembly	Fragment 1: P_{rha}, and 5'UTR from pP_{rha}_GCas9_Sp1 amplified with primers BG13427 and BG13428 Fragment 2: AcrIIC1$_{Nme}$ from pMK-RQ-_Acr_coK12_coT12 amplified with primers BG13429 and BG13430 Fragment 3: T7 terminator from annealed oligonucleotides BG13431 and BG13432 Fragment 4: pUC19 amplified with primers BG13433 and BG13434	AcrIIC1$_{Nme}$ inhibiting ThermoCas9, ThermoCas9$_{L551H}$ or GeoCas9 in inhibition assays (dual-vector system), spectophotometry- and flow cytometry-based fluorescence loss assays. PCR template for pAcr_TCas9_Sp1, pAcr_TCas9_Sp2, pAcr_TCas9_Sp3, pAcr_TCas9_Spscr, pAcr_GCas9_Sp1, pAcr_GCas9_Sp2, pAcr_GCas9_Sp3, and pAcr_GCas9_Spscr.	This study
pTCas9L511H_Sp3	Gibson assembly	Fragment 1: 5' UTR, lac *operator*, P_{tet}, P_{lacI}, *lacI* gene, terminator, sgRNA 3, terminator, backbone, and terminator from pGCas9_Sp3 amplified with primers BG13459 and BG13460 Fragment 2: *thermocas9$_{L551H}$* gene from pP_{tet}_Acr_P_{rha}_TCas9$_{L551H}$ amplified with primers BG13461 and BG13462	ThermoCas9$_{L551H}$ and sgRNA (spacer 3) targeting the genome-integrated P_{lacUV5} in inhibition assays (dual-vector system), spectophotometry- and flow cytometry-based fluorescence loss assays.	This study
pTCas9L511H_Spscr	Gibson assembly	Fragment 1: 5' UTR, lac *operator*, P_{tet}, P_{lacI}, *lacI* gene, terminator, sgRNA scr, terminator, backbone, and terminator from pGCas9_Spscr amplified with primers BG13459 and BG13460 Fragment 2: *thermocas9$_{L551H}$* gene from pP_{tet}_Acr_P_{rha}_TCas9$_{L551H}$ amplified with primers BG13461 and BG13462	ThermoCas9$_{L551H}$ and sgRNA (spacer scr) non-targeting the genome-integrated P_{lacUV5} in inhibition assays (dual-vector system), spectophotometry- and flow cytometry-based fluorescence loss assays.	This study
pGdCas9_Sp1	Gibson assembly	Fragment 1: *GeodCas9* gene from pGCas9_Sp1 amplified with primers BG13680 and BG13681 Fragment 2: *GeodCas9* gene, 5' UTR, lac *operator*, P_{tet}, P_{lacI}, *lacI* gene, terminator, sgRNA 1, terminator, backbone, and terminator from pGCas9_Sp1 amplified with primers BG13682 and BG13683	GeodCas9 and sgRNA (spacer 1) targeting the genome-integrated P_{lacUV5} in spectophotometry- and flow cytometry-based fluorescence loss assays.	This study

Table S8.2 | Plasmids used in this study. (continued)

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pGdCas9_Sp2	Gibson assembly	Fragment 1: *GeodCas9* gene from pGCas9_Sp2 amplified with primers BG13680 and BG13681 Fragment 2: *GeodCas9* gene, 5' UTR, lac *operator*, P_{tet}, P_{lacI}, *lacI* gene, terminator, sgRNA 2, terminator, backbone, and terminator from pGCas9_Sp1 amplified with primers BG13682 and BG13683	GeodCas9 and sgRNA (spacer 2) targeting the genome-integrated P_{lacUV5} in spectophotometry- and flow cytometry-based fluorescence loss assays.	This study
pGdCas9_Sp3	Gibson assembly	Fragment 1: *GeodCas9* gene from pGCas9_Sp3 amplified with primers BG13680 and BG13681 Fragment 2: *GeodCas9* gene, 5' UTR, lac *operator*, P_{tet}, P_{lacI}, *lacI* gene, terminator, sgRNA 3, terminator, backbone, and terminator from pGCas9_Sp1 amplified with primers BG13682 and BG13683	GeodCas9 and sgRNA (spacer 3) targeting the genome-integrated P_{lacUV5} in spectophotometry- and flow cytometry-based fluorescence loss assays.	This study
pGdCas9_Spscr	Gibson assembly	Fragment 1: *GeodCas9* gene from pGCas9_Spscr amplified with primers BG13680 and BG13681 Fragment 2: *GeodCas9* gene, 5' UTR, lac *operator*, P_{tet}, P_{lacI}, *lacI* gene, terminator, sgRNA scr, terminator, backbone, and terminator from pGCas9_Sp1 amplified with primers BG13682 and BG13683	GeodCas9 and sgRNA (spacer scr) targeting the genome-integrated P_{lacUV5} in spectophotometry- and flow cytometry-based fluorescence loss assays.	This study
pTdCas9_Sp1	Gibson assembly	Fragment 1: *thermodCas9* gene from pThermoCas9i amplified with primers BG12625 and BG13684 Fragment 2: 5' UTR, lac *operator*, P_{tet}, P_{lacI}, *lacI* gene, terminator, sgRNA 1, terminator, backbone, and terminator from pTCas9_Sp1 amplified with primers BG13311 and BG13685	ThermodCas9 and sgRNA (spacer 1) targeting the genome-integrated P_{lacUV5} in spectophotometry- and flow cytometry-based fluorescence loss assays.	This study
pTdCas9_Sp2	Gibson assembly	Fragment 1: *thermodCas9* gene from pThermoCas9i amplified with primers BG12625 and BG13684 Fragment 2: 5' UTR, lac *operator*, P_{tet}, P_{lacI}, *lacI* gene, terminator, sgRNA 2, terminator, backbone, and terminator from pTCas9_Sp2 amplified with primers BG13311 and BG13685	ThermodCas9 and sgRNA (spacer 2) targeting the genome-integrated P_{lacUV5} in spectophotometry- and flow cytometry-based fluorescence loss assays.	This study

Table S8.2 | Plasmids used in this study. (continued)

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pTdCas9_Sp3	Gibson assembly	Fragment 1: *thermodCas9* gene from pThermoCas9i amplified with primers BG12625 and BG13684 Fragment 2: 5' UTR, lac *operator*, P_{tet}, P_{lacI}, *lacI* gene, terminator, sgRNA 3, terminator, backbone, and terminator from pTCas9_Sp3 amplified with primers BG13311 and BG13685	ThermodCas9 and sgRNA (spacer 3) targeting the genome-integrated P_{lacUV5} in spectophotometry- and flow cytometry-based fluorescence loss assays.	This study
pTdCas9_Spscr	Gibson assembly	Fragment 1: *thermodCas9* gene from pThermoCas9i amplified with primers BG12625 and BG13684 Fragment 2: 5' UTR, lac *operator*, P_{tet}, P_{lacI}, *lacI* gene, terminator, sgRNA scr, terminator, backbone, and terminator from pTCas9_Spscr amplified with primers BG13311 and BG13685	ThermodCas9 and sgRNA (spacer scr) targeting the genome-integrated P_{lacUV5} in spectophotometry- and flow cytometry-based fluorescence loss assays.	This study
pAcr_GCas9_Sp2	Gibson assembly	Fragment 1: P_{rha}, and $AcrIIC1_{Nme}$ from pAcr amplified with primers BG14000 and BG14001 Fragment 2: Terminator, GeoCas9, 5' UTR, lac *operator*, P_{tet}, backbone, P_{lacI}, and partly *lacI* gene from pGCas9_Sp2 amplified with primers BG12705 and BG14002 Fragment 3: Partly *lacI* gene, terminator, P_{J23119}, sgRNA 2, terminator, and backbone from pGCas9_Sp2 amplified with primers BG12722 and BG14003	GeoCas9 and sgRNA (spacer 2) targeting the genome-integrated P_{lacUV5} and $AcrIIC1_{Nme}$ inhibiting GeoCas9, in inhibition assays (single-vector system).	This study
pAcr_GCas9_Sp3	Gibson assembly	Fragment 1: P_{rha}, and $AcrIIC1_{Nme}$ from pAcr amplified with primers BG14000 and BG14001 Fragment 2: Terminator, GeoCas9, 5' UTR, lac *operator*, P_{tet}, backbone, P_{lacI}, and partly *lacI* gene from pGCas9_Sp3 amplified with primers BG12705 and BG14002 Fragment 3: Partly *lacI* gene, terminator, P_{J23119}, sgRNA 3, terminator, and backbone from pGCas9_Sp3 amplified with primers BG12722 and BG14003	GeoCas9 and sgRNA (spacer 3) targeting the genome-integrated P_{lacUV5} and $AcrIIC1_{Nme}$ inhibiting GeoCas9, in inhibition assays (single-vector system).	This study

Table S8.2 | Plasmids used in this study. (continued)

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pAcr_GCas9_Spscr	Gibson assembly	Fragment 1: P_{rha}, and AcrIIC1$_{Nme}$ from pAcr amplified with primers BG14000 and BG14001 Fragment 2: Terminator, GeoCas9, 5' UTR, lac *operator*, P_{tet}, backbone, P_{lacI}, and partly *lacI* gene from pGCas9_Spscr amplified with primers BG12705 and BG14002 Fragment 3: Partly *lacI* gene, terminator, P_{J23119}, sgRNA scr, terminator, and backbone from pGCas9_Spscr amplified with primers BG12722 and BG14003	GeoCas9 and sgRNA (spacer scr) targeting the genome-integrated P_{lacUV5} and AcrIIC1$_{Nme}$ inhibiting GeoCas9, in inhibition assays (single-vector system).	This study
pAcr_TCas9_Sp1	Gibson assembly	Fragment 1: P_{rha}, and AcrIIC1$_{Nme}$ from pAcr amplified with primers BG14000 and BG14001 Fragment 2: Terminator, ThermoCas9, 5' UTR, lac *operator*, P_{tet}, backbone, P_{lacI}, and partly *lacI* gene from pTCas9_Sp1 amplified with primers BG12705 and BG14002 Fragment 3: Partly *lacI* gene, terminator, P_{J23119}, sgRNA 1, terminator, and backbone from pTCas9_Sp1 amplified with primers BG12722 and BG14003	ThermoCas9 and sgRNA (spacer 1) targeting the genome-integrated P_{lacUV5} and AcrIIC1$_{Nme}$ inhibiting GeoCas9, in inhibition assays (single-vector system).	This study
pAcr_TCas9_Sp2	Gibson assembly	Fragment 1: P_{rha}, and AcrIIC1$_{Nme}$ from pAcr amplified with primers BG14000 and BG14001 Fragment 2: Terminator, ThermoCas9, 5' UTR, lac *operator*, P_{tet}, backbone, P_{lacI}, and partly *lacI* gene from pTCas9_Sp2 amplified with primers BG12705 and BG14002 Fragment 3: Partly *lacI* gene, terminator, P_{J23119}, sgRNA 2, terminator, and backbone from pTCas9_Sp2 amplified with primers BG12722 and BG14003	ThermoCas9 and sgRNA (spacer 2) targeting the genome-integrated P_{lacUV5} and AcrIIC1$_{Nme}$ inhibiting GeoCas9, in inhibition assays (single-vector system).	This study

Table S8.2 | Plasmids used in this study. (continued)

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pAcr_TCas9_Sp3	Gibson assembly	Fragment 1: P_{rha}, and AcrIIC1$_{Nme}$ from pAcr amplified with primers BG14000 and BG14001 Fragment 2: Terminator, ThermoCas9, 5' UTR, lac *operator*, P_{tet}, backbone, P_{lacI}, and partly *lacI* gene from pTCas9_Sp3 amplified with primers BG12705 and BG14002 Fragment 3: Partly *lacI* gene, terminator, P_{J23119}, sgRNA 3, terminator, and backbone from pTCas9_Sp3 amplified with primers BG12722 and BG14003	ThermoCas9 and sgRNA (spacer 3) targeting the genome-integrated P_{lacUV5} and AcrIIC1$_{Nme}$ inhibiting GeoCas9, in inhibition assays (single-vector system).	This study
pAcr_TCas9_Spscr	Gibson assembly	Fragment 1: P_{rha}, and AcrIIC1$_{Nme}$ from pAcr amplified with primers BG14000 and BG14001 Fragment 2: Terminator, ThermoCas9, 5' UTR, lac *operator*, P_{tet}, backbone, P_{lacI}, and partly *lacI* gene from pTCas9_Spscr amplified with primers BG12705 and BG14002 Fragment 3: Partly *lacI* gene, terminator, P_{J23119}, sgRNA scr, terminator, and backbone from pTCas9_Spscr amplified with primers BG12722 and BG14003	ThermoCas9 and sgRNA (spacer scr) targeting the genome-integrated P_{lacUV5} and AcrIIC1$_{Nme}$ inhibiting GeoCas9, in inhibition assays (single-vector system).	This study

Table S8.3 | Strains used in this study

Strain	Description	Plasmid	Antibiotic resistance	Reference
E. coli DH5α	Wild type	-	-	Lab stock
E. coli DH10B	Wild type	-	-	Lab stock
E. coli DH10B	Promoter lacUV5 and *gfp* gene integrated in the genome	-	-	Lab stock
E. coli DH10B	Promoter lacUV5 and *gfp* gene integrated in the genome	pAcrIIC1$_{Nme}$	Ampicillin	This study

Table S8.4 | Primers/Oligonucleotides used in this study.

Primer/Oligo ID	Sequence (5' -> 3')	Description
	Targeting, inhibition (dual-vector system), spectrophotometry- and flow cytometry-based assays	
BG13307	tcataagcaggccattttgtctgatTTAGTCACGAGTAGATTGCAGT G	Construction of pGCas9_Sp1, pGCas9_Sp2, pGCas9_Sp3, pGCas9_Spscr
BG13310	ttgctaacgcagtcaggcaccgtgtATGCGTTATAAGATTGGCCTGG	
BG13311	ACACGGTGCCTGACTG	
BG13312	ttttttggtccatcagtcgattctgaTCACTGCCCGCTTTCC	
BG13313	TCAGAATCGACTGATGGACC	
BG13309	ATCAGACAAAATGGCCTGC	
BG13459	tatcaagaccgattttatacttcatACACGGTGCCTGACTG	Construction of pTCas9_Sp1, pTCas9_Sp2, pTCas9_Sp3, pTCas9_Spscr
BG13460	cggggaaactatccgtccgttataaATCAGACAAAATGGCCTGCTTA TG	
BG13461	TTATAACGGACGGATAGTTTCCC	
BG13462	ATGAAGTATAAAATCGGTCTTGATATCGG	
	Inhibition (dual-vector approach), spectrophotometry- and flow cytometry-based assays	
BG13427	aagcgggcagtgagcgcaacgcaatCACCACAATTCAGCAAATTGTG	Construction of pAcr
BG13428	ATGTATATCTCCTTCTTATAGTTAAACAAAATTATTTCTAG	
BG13429	ttaactataagaaggagatatacatATGAATAAAACATATAAAATCG GAAAAAATGCGG	
BG13430	gctcagcggtggcggccgctctaggTCATAGTTCAACAAACTCCCAA CATG	
BG13431	CCTAGAGCGGCCGCCACCGCTGAGCAATAACTAGCATAACCCCTTGG GGCCTCTAAACGGGTCTTGAGGGGTTTTTTG	
BG13432	CAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTA GTTATTGCTCAGCGGTGGCGGCCGCTCTAGG	
BG13433	taaacgggtcttgaggggttttttgTTAAGCCAGCCCCGAC	
BG13434	ATTGCGTTGCGCTCAC	
BG13459	tatcaagaccgattttatacttcatACACGGTGCCTGACTG	Construction of pTCas9$_{L551H}$_Sp3 pTCas9$_{L551H}$_Spscr
BG13460	cggggaaactatccgtccgttataaATCAGACAAAATGGCCTGCTTA TG	
BG13461	TTATAACGGACGGATAGTTTCCC	
BG13462	ATGAAGTATAAAATCGGTCTTGATATCGG	
	Spectrophotometry- and flow cytometry-based assays	
BG13680	gatcacaGCatctacttcaacgtaac	Construction of pGdCas9_Sp1, pGdCas9_Sp2, pGdCas9_Sp3, pGdCas9_Spscr
BG13681	gcctggccatcggtattac	
BG13682	gtaataccgatggccaggc	

Table S8.4 | Primers/Oligonucleotides used in this study. (continued)

Primer/Oligo ID	Sequence (5' -> 3')	Description
BG13683	gttacgttgaagtagatGCtgtgatc	
BG12625	atcttattaatcagataaaatatttGTCGACAGGATTAAACAAAAAT GGC	
BG13684	ttgctaacgcagtcaggcaccgtgtATGAAGTATAAAATCGGTCTTG CTATCG	Construction of pTdCas9_Sp1, pTdCas9_Sp2, pTdCas9_Sp3, pTdCas9_Spscr
BG13311	ACACGGTGCCTGACTG	
BG13685	aaatattttatctgattaataagatgatcttcttgagatcg	
	Inhibition assays (single-vector system)	
BG14000	atcttattaatcagataaaatatttCACCACAATTCAGCAAATTGTG	
BG14001	gccatttttgtttaatcctgtcgacTCATAGTTCAACAAACTCCCAA CATG	Construction of pAcr_GCas9_Sp1, pAcr_GCas9_Sp2, pAcr_GCas9_Sp3, pAcr_GCas9_Spscr, pAcr_TCas9_Sp1, pAcr_TCas9_Sp2, pAcr_TCas9_Sp3, pAcr_TCas9_Spscr
BG12705	GTCGACAGGATTAAACAAAAATGGCC	
BG14002	CAATCAGCAACGACTGTTTGC	
BG12722	GCAAACAGTCGTTGCTGATTG	
BG14003	AAATATTTTATCTGATTAATAAGATGATCTTCTTGAGATCG	
	Sequencing	
BG12706	GGTAGCTCAGAGAACCTTCG	
BG12712	ACGCGATGGATATGTTCTG	
BG13157	GCGAATTCGGCAACATGTC	
BG13158	CAGGGTGAAGTCTACTCCTC	
BG13159	GCTACATCAGCCGCTTCTTC	
BG13160	GTCTGCTGGAGCATAACAAC	
BG13161	GCACCCTGAAACGTTTCG	
BG13163	CGAAACGTGTAATCTTCGGTC	Sequencing of pGCas9_Sp1, pGCas9_Sp2, pGCas9_Sp3, pGCas9_Spscr, pGdCas9_Sp1, pGdCas9_Sp2, pGdCas9_Sp3, pGdCas9_Spscr, *pAcr_GCas9_Sp1, *pAcr_GCas9_Sp2, *pAcr_GCas9_Sp3, *pAcr_GCas9_Spscr
BG13164	GGCCAGTTCCTTGTTCTG	
BG13165	CGGTTTTCATCCTGTTCC	
BG13166	CTGGAATGCTTGTTCATACAGC	
BG13167	CGCTCCAGACGGTGTTTAC	
BG12721	CTGCGTTAGCAATTTAACTGTG	
BG12722	GCAAACAGTCGTTGCTGATTG	
BG12723	GTTAACCACCATCAAACAGG	
BG12718	CAGTCGCGTACCGTCTTC	
BG12719	GTGCATGCAAGGAGATGG	
*BG12705	GTCGACAGGATTAAACAAAAATGGCC	
BG12706	GGTAGCTCAGAGAACCTTCG	

Table S8.4 | Primers/Oligonucleotides used in this study. (continued)

Primer/Oligo ID	Sequence (5' -> 3')	Description
BG12707	GTCGTATCGATGATTTTGATTGTTCG	
BG12708	GCGTGAGATATAACGGGTATC	
BG12709	GTAGACTTCGCCTTGTTCC	
BG12710	GTTCCCATATTCGCGCTG	
BG12712	ACGCGATGGATATGTTCTG	
BG12713	CGAAGATTTAGGTGTCCGC	
BG12714	CATACACATTCCAGTCCTTCACC	Sequencing of pTCas9_Sp1, pTCas9_Sp2, pTCas9_Sp3, pTCas9_Spscr, pTCas9$_{L551H}$_Sp3, pTCas9$_{L551H}$_Spscr, pTdCas9_Sp1, pTdCas9_Sp2, and pTdCas9_Sp3, pTdCas9_Spscr, *pAcr_TCas9_Sp1, *pAcr_TCas9_Sp2, *pAcr_TCas9_Sp3, and *pAcr_TCas9_Spscr
BG12715	CACCGGTCTCCATCCATATC	
BG12716	CAACGCCGAGCGATATCG	
BG12717	CCGTACTCTGAGTGGAAGG	
BG12718	CAGTCGCGTACCGTCTTC	
BG12719	GTGCATGCAAGGAGATGG	
BG12721	CTGCGTTAGCAATTTAACTGTG	
BG12722	GCAAACAGTCGTTGCTGATTG	
BG12723	GTTAACCACCATCAAACAGG	
*BG12705	GTCGACAGGATTAAACAAAAATGGCC	
*BG14004	CAGTCGGGAAACCTGTCG	
BG13437	CAATACGCAAACCGCCTC	Sequencing of pAcr
BG13438	GAGACGGTCACAGCTTGTC	

Chapter 9

Guide-free Cas9 from pathogenic *Campylobacter jejuni* bacteria causes severe DNA damage

Chinmoy Saha*, Prarthana Mohanraju*, Andrew Stubbs, Gaurav Dugar, Youri Hoogstrate, Gert-Jan Kremers, Wiggert A. van Cappellen, Deborah Horst-Kreft, Joyce H.G. Lebbink, Serena Bruens, Duncan Gaskin, Dior Beerens, Maarten Klunder, Rob Joosten, Jeroen A. A. Demmers, Dik van Gent, Johan W. Mouton, Peter J. van der Spek, John van der Oost, Peter van Baarlen, Rogier Louwen

* contributed equally

Adapted from:
Guide-free Cas9 from pathogenic Campylobacter jejuni bacteria causes severe DNA damage in cultured human cells.
2019, Manuscript in preparation

Abstract

CRISPR–Cas9 systems are enriched in human pathogenic bacteria, where they have been linked to cytotoxicity and even death of laboratory animals. Here we show that upon infection of human cells, Campylobacter jejuni *secretes its Cas9 (CjeCas9) into the cytoplasm of infected human cells. Next, a native nuclear localization signal enables autonomous entry of the guide-free CjeCas9 into the nucleus, where it catalyses metal-dependent non-specific nicking of double stranded DNA, eventually leading to cell death. Compared to CjeCas9, we confirmed that SpyCas9 of* Streptococcus pyogenes *has an increased efficiency of guide-dependent editing, and in the absence of a guide it does not affect double stranded DNA. Although transduction of the SpyCas9 protein to human cells still may cause some DNA damage, most likely due to its ssDNA cleavage activity, this problem can be completely prevented by saturation with an appropriate guide, whereas this is only partly true for CjeCas9. Overall, we conclude that CjeCas9 has evolved towards a distinct natural function (attack of human cells rather than defense against viruses) and that the unique catalytic features may make this nuclease less suitable for genome editing applications.*

Results

CRISPR–Cas is an RNA-guided adaptive defense system that enables prokaryotes to recognize and destroy nucleic acids of invading genetic elements (viruses, plasmids) by CRISPR-associated Cas nucleases (*29*). In addition, besides providing adaptive immunity, CRISPR–Cas systems have been demonstrated to play other roles as well (*131*), including in bacterial pathogenicity (*133*). Indeed, comparative genomic analysis demonstrated high prevalence of type II CRISPR–Cas systems, characterised by presence of the Cas9 protein, in bacterial pathogens (*119*). In some of these pathogens, such as *Francisella novicida*, *Neisseria meningitidis* and *Campylobacter jejuni,* Cas9 activity was found to correlate with the death of infected human cells (*120*) or infected laboratory animals (*121*). Cas9-mediated silencing of expression of exposed cell wall proteins through an antisense mechanism has been proposed as immune evasion mechanism (*121*). Another important phenomenon is the capacity of several intracellular bacterial pathogens, including *C. jejuni*, to induce double-strand DNA breaks (DSBs) in infected human host cells, either actively by secreted bacterial cytotoxins or passively by activating the host's cell death response (*425*). In a previous study we demonstrated that the CRISPR–Cas9 systems in strains of *C. jejuni* (*Cj*) are typically associated with small or degenerated CRISPR arrays, and that they, at most, play a minor role in anti-viral defense (*120*) (**Figure S9.1A**). In contrast, the presence of CjeCas9 could be linked to cytotoxicity and cell death, although the molecular mechanism remained elusive (*120*). In the current work we therefore studied the role of CjeCas9 in the induction of cytotoxicity in human cells in more detail to elucidate the mechanism behind this phenomenon.

Research on pathogenic bacteria has revealed that many cytotoxins are secreted (*425*), either by dedicated export systems or by outer membrane vesicles (OMVs) (*426*). Whereas *C. jejuni* appears to lack the former transport systems (*427*), it has been demonstrated to produce OMVs (*426*). Interestingly, proteomics analysis of OMVs secreted by *C. jejuni* also revealed the presence of CjeCas9 (*428*). We therefore investigated whether CjeCas9 is also present in the OMVs of our *C. jejuni* model strain GB11 (*429*). To allow for detection of correctly folded CjeCas9, it was fused to a red fluorescent protein (mCherry). Indeed, CjeCas9 could be detected in the OMVs secreted by GB11 bacteria (**Figure S9.1B**). Interestingly, during bacterial infection experiments in human cells, we noticed that CjeCas9-mCherry not only accumulated in the cytoplasm (**Figure 9.1A**), but also in the nucleus (**Figure 9.1A**, **S9.1C**). As infection by *C. jejuni* is accompanied by severe DNA damage in human cells (*430*), we speculated that the observed nuclear localization of the CjeCas9 nuclease, previously demonstrated to catalyse the typical sgRNA-guided generation of DSBs (*431*), might suggest a direct role of CjeCas9 in the pathogenicity of *C. jejuni* (**Figure S9.1D to F**).

The observed nuclear localization of native CjeCas9, produced by *C. jejuni* bacteria during their infection of human cells, suggested that CjeCas9 was somehow able to enter the nuclei of human cells, without prior *in vitro* modification, thus autonomously. To assay for autonomous nuclear entry activity of CjeCas9, we fused the genes coding for CjeCas9 and enhanced green fluorescent protein (eGFP) in a eukaryotic expression vector. After transfection into human cells, we observed nuclear eGFP-CjeCas9 accumulation both by fluorescence microscopy (**Figure 9.1B**, **S9.2A**) and Western blot analysis (**Figure 9.1C**). To elucidate the molecular basis of this autonomous entry of CjeCas9 into the nucleus, an *in-silico* analysis was performed of the CjeCas9 protein sequence of *C. jejuni* GB11. We found a potential nuclear localization signal (NLS) in the Cas9 arginine-rich bridging helix (BH) (*216*) encoding sequence (**Figure S9.2B**),

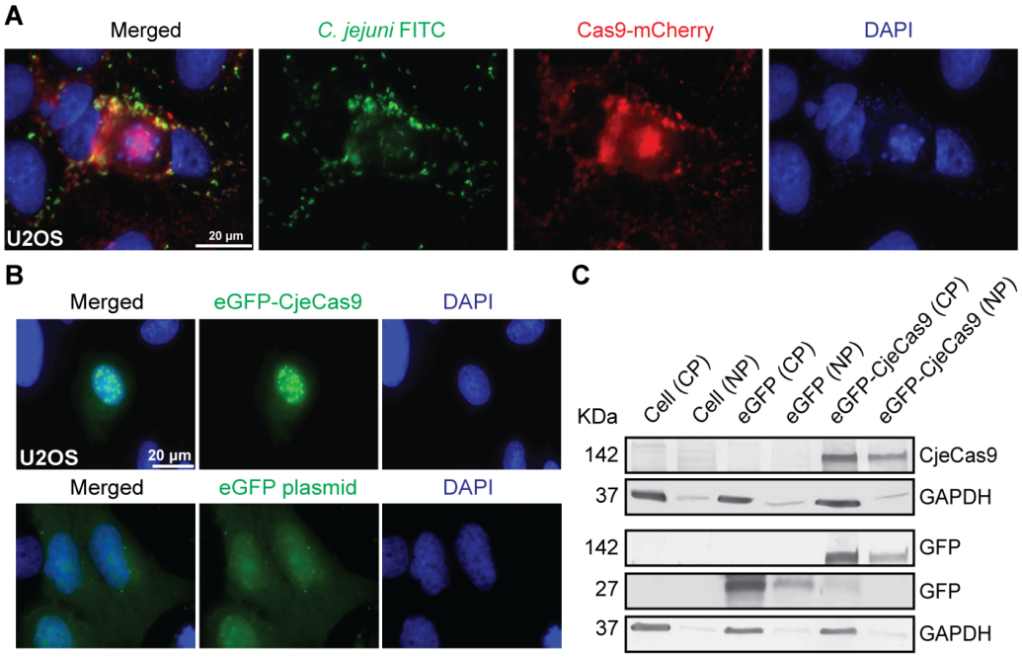


Figure 9.1 | CjeCas9 is released by *C. jejuni* during infection of human cells and localizes into their nucleus. **(A)** Human cells infected with a Cas9-mCherry (red) expressing *C. jejuni* isolate (green). **(B)** Microscopic image of nuclear eGFP-CjeCas9 localization in human cells (green). eGFP transfected cells are shown as a control. **(C)** Nuclear (NP) and cytoplasmic (CP) protein fractions of human cells. GAPDH verified the quality of the separation.

displaying extensive sequence identity to viral arginine-rich domain sequences that occur in dual-function proteins that enable RNA binding as well as nuclear localization (*432*, *433*). To assess whether the predicted CjeCas9 NLS facilitated nuclear entry, the DNA fragment encoding the CjeCas9 NLS region was fused to the *egfp* gene in the same expression vector. After transfection of human cells, correctly expressed eGFP-NLS(CjeCas9) accumulated in the nucleus as efficient as eGFP fused to the well-established NLS of Simian Virus (SV) 40 large T-antigen protein (**Figure S9.2C**).

The observed nuclear localization, together with the established endonuclease activity of CjeCas9 (*431*), led us to hypothesize that CjeCas9 produced and released by *C jejuni* during infection of human cells could potentially alter DNA integrity and cell homeostasis. This was validated by counting foci representing the p53-binding protein 1 (53BP1), an activator of p53 protein function and an important regulator of the cellular response to repairing DSB ends (*434*). Six hours post-infection, microscopic analysis revealed significantly increased numbers of 53BP1 foci in nuclei of human cells infected with either Cas9-producing wild-type GB11 (WT) or a complemented GB11Δ*cas*9::*cas*9 (Δ*cas*9::*cas*9) *C. jejuni* bacteria, but not with the corresponding *cas9* deletion mutant GB11Δ*cas*9 bacteria (Δ*cas*9) (**Figure 9.2A**, **9.2B**). This observation was further validated in infection assays as above, but now using the phosphorylated histone H2A variant X (γ-H2AX), an established histological marker of DSB-associated chromatin decondensation (*435*). Upon infection by CjeCas9-producing bacteria (WT or Δ*cas*9::*cas*9) or upon exposure to the OMVs produced by these bacteria, the human cells revealed bright red nuclei after fluorescent antibody staining of γ-H2AX (**Figure 9.2C,D**, **S9.3A to C**). Compared to

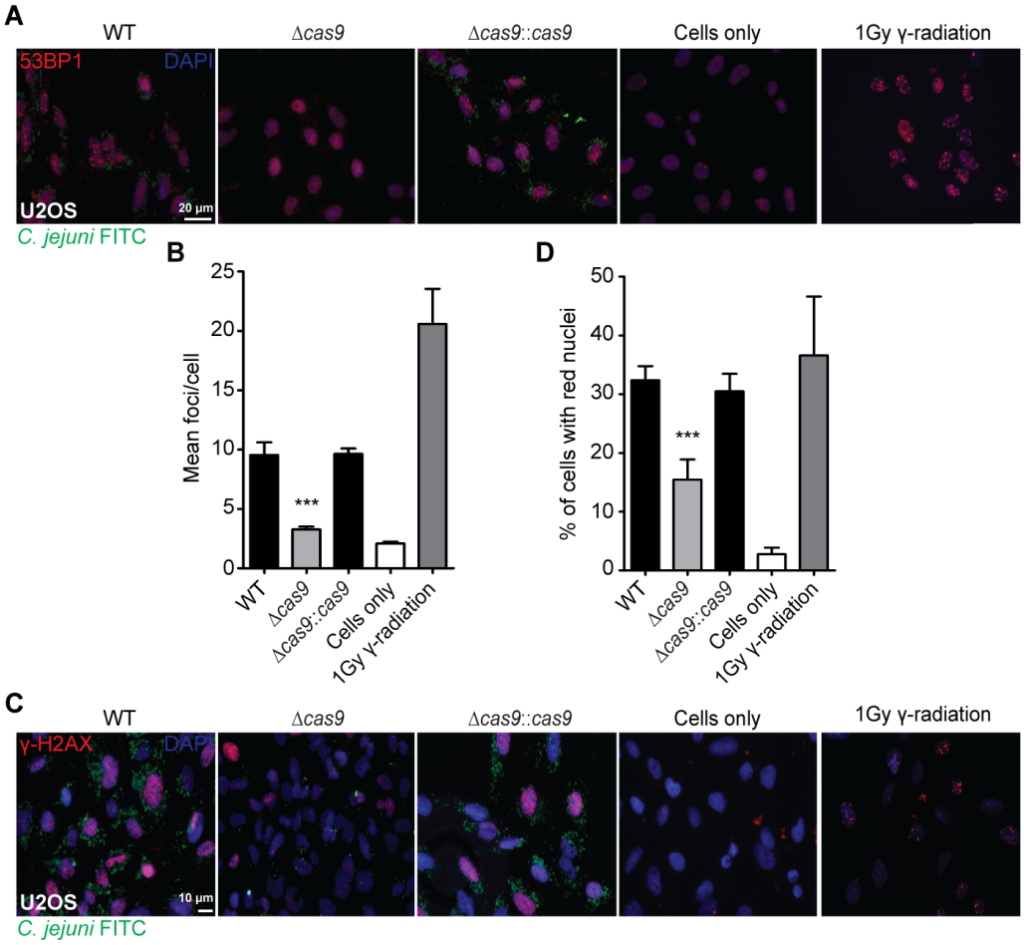


Figure 9.2 | CjeCas9 activates the chromatin- and DNA integrity markers, 53BP1 and γ-H2AX in U2OS cells. *C. jejuni* GB11 (WT) and its variants Δcas9 or Δcas9::cas9 were stained with anti-*C. jejuni* FITC (green), nuclei were stained with DAPI (blue). Gamma (γ)-irradiated cells were used as a positive control and untreated cells as a negative control. **(A)** Human cells infected with WT, Δcas9 or Δcas9::cas9 were stained for 53BP1 (red). **(B)** Number of 53BP1 related foci per cell (mean foci/cell, Y-axis) were counted in human cells. The *C. jejuni* bacteria used are shown on the X-axis. **(C)** Human cells infected with WT, Δcas9 or Δcas9::cas9 were stained for γ-H2AX (red). (D) The percentage of human cells with bright red nuclei is visualized on the Y-axis. The *C. jejuni* bacteria used are shown on the X-axis. Values represent mean ± s.e.m, *** $p < 0.001$ by One-Way ANOVA (Bonferroni's Multiple Comparison test).

human cells infected with CjeCas9-producing bacteria or their OMVs, significantly less bright red nuclei were observed in human cells infected with the Δ*cas*9 bacteria or when exposed to their OMVs (**Figure 9.2C,D**, **S9.3C,D**). Our data suggest a lower infection efficiency of Δ*cas*9 bacteria in human cells; it is tempting to speculate that this may be related to the aforementioned Cas9-mediated silencing phenomenon (*121*). But this apparent difference was not found to influence the overall effect of CjeCas9-induced cytotoxicity earlier (*120*).

Previously, comparative genomics analyses have indicated that different *C. jejuni* strains produce a varying number of virulence factors (*436*, *437*). However, until now, only one of these factors is known for its DNA damaging capacity, *i.e.* Cytolethal Distending Toxin (CDT) (*425*).

As CjeCas9 also known for its DNA-cleaving activity (*51*, *431*), the aforementioned effects on host cell chromatin and DNA integrity was tested in the presence or absence of CDT. The same infection studies as described above were performed using bacteria that did produce or lacked CDT, whereas they both expressed Cas9. By analysing both γ-H2AX and 53BP1, these assays revealed that human cells infected with either producing or non-producing CDT *C. jejuni* bacteria displayed the same amount of alterations in DNA integrity and cell homeostasis (**Figure S9.4A,B**). Overall, this outcome corroborated previous studies that demonstrated that, under the tested conditions, CDT is not the sole factor responsible for disrupting chromatin and DNA integrity during human cell infection by *C. jejuni* (*438*, *439*). Again, we conclude that CjeCas9 contributes to the disruption of chromatin and DNA integrity in human cells.

To examine whether CjeCas9 nuclease activity specifically leads to changes in human cellular chromatin and DNA integrity, we quantified γ-H2AX stained nuclei in human cells transduced with plasmids encoding either eGFP-CjeCas9 or its catalytically inactive (dead) variant, eGFP-CjedCas9. At comparable transfection efficiencies (**Figure S9.5A,B**), γ-H2AX stained nuclei were significantly more often detected in human cells transfected with eGFP-CjeCas9 compared to cells transfected with eGFP-CjedCas9 (**Figure S9.5A,C**), a result that was validated by Western blotting (**Figure S9.5D**). The 53BP1, γ-H2AX immuno-histochemistry and γ-H2AX Western blotting results corroborated that CjeCas9 activity indeed alters human chromatin and DNA integrity. We realized that these findings could be of significant impact, but they did not reveal a clear insight in the chromosome-wide occurrence of DNA damage, let alone the mechanism behind these findings.

Therefore, at first, the BLESS (direct *in situ* breaks labelling, enrichment on streptavidin and next-generation sequencing) method (*440*) was used in infection experiments to obtain a genome-wide snapshot impression profile of DNA DSB induction that associates with CjeCas9 activity. The BLESS experiment (**Figure 9.3A**) revealed that human cells infected for 6 hours with WT or Δ*cas9*::*cas9* bacteria harboured a total of 678,910 or 942,104 CjeCas9-dependent DSBs, respectively (**Figure 9.3B**, **S9.6A,B**; **Table S9.1**).

These numbers of Cas9-dependent DSBs were obtained after background correction, *i.e.* after subtraction of DSBs in control infection experiments with Δ*cas9* bacteria and uninfected human cells (**Table S9.2**). Thus, the BLESS data revealed that native CjeCas9 nuclease activity is associated with increased numbers of DSBs in the nucleus (**Figure 9.3**, **S9.6B to D**), in agreement with our 53BP1 and γ-H2AX findings. Despite extensive *in-silico* analysis of raw and filtered CjeCas9-associated BLESS data, the three reported CjeCas9 PAM motifs (*431*, *441*, *442*) were not detected (**Figure 9.3C**; **Table S9.3**, **S9.4**), suggesting that the processing of human cellular DNA by native CjeCas9 during infection by *C. jejuni* is PAM-independent.

In all experiments described above, CjeCas9 activity was untargeted, since no exogenous RNA guides (gRNA) had been supplied. Our finding for guide-free CjeCas9 appears to resemble recent analyses described for Cas9 proteins from *Streptococcus pyogenes* (SpyCas9) and *F. novicida* (FnoCas9) that have been shown to cleave plasmid DNA in a guide- and RNA-independent manner *in vitro* (*443*). Indeed, SpyCas9 (type II-A) cleaves ssDNA, while FnoCas9 (type II-B) nicks plasmid dsDNA (*443*). This RNA-independent activity is dependent on Mg^{2+}/Mn^{2+} cations, and requires the RuvC domain of SpyCas9, and on the HNH domain of FnoCas9 (*443*). We considered that these metal ions may play an important role in the PAM- and guide- independent DNA processing activity by CjeCas9 as well. This was further validated after purification and verification of recombinant CjeCas9 and SpyCas9 proteins (**Figure S9.7A,B**) in a

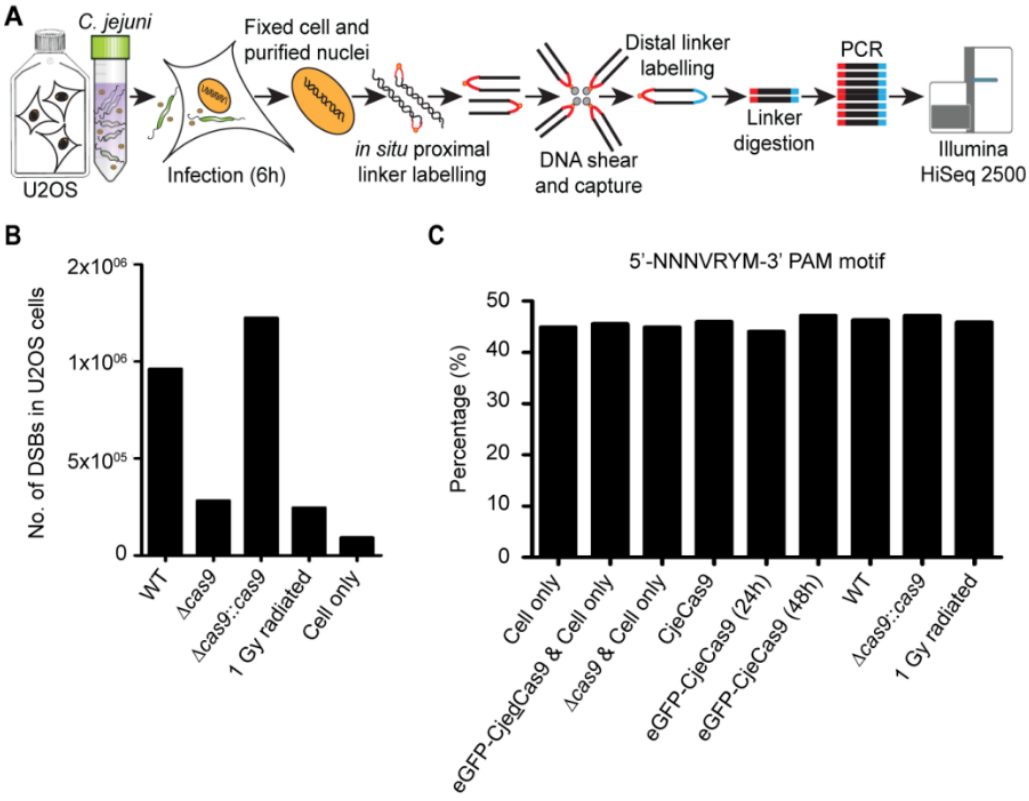


Figure 9.3 | Analysis of chromosomal DSBs induction using BLESS(A) Overview figure of the BLESS method after *C. jejuni* infection of U2OS cells. **(B)** Histogram showing numbers of CjeCas9-dependent DSBs accumulated during bacterial infection. **(C)** Example outcome of the PAM analyses. Y-axis shows the percentage of the 5'-NNNVRYM-3' PAM motif as observed in raw and filtered BLESS datasets. X-axis shows the sample names that were used in the BLESS analyses.

plasmid cleavage experiment in the presence of Mn^{2+} or Mg^{2+}. The obtained results revealed that CjeCas9 nicks plasmid dsDNA in the presence of at least 100 µM of Mn^{2+} and 500 µM of Mg^{2+} (**Figure 9.4A**). In contrast, under the same experimental conditions, SpyCas9 did not cleave dsDNA (**Figure S9.9A**), but ssDNA only (**Figure S9.9B**), corroborating a previous report (*443*). To identify if the cation dependent cleavage of plasmid dsDNA by CjeCas9 was catalysed by the RuvC or HNH domain, we carried out the same metal ion dependent cleavage assays with wild-type, RuvC and/or HNH active site mutants of CjeCas9. We found that the guide-independent activity of CjeCas9 to cleave plasmid dsDNA was completely abolished in both the HNH-inactive single mutant and the HNH/RuvC-inactive double mutant (**Figure 9.4A**, **S9.8A**). This shows that the guide-independent processing of double stranded plasmid DNA depends on the nicking activity of the CjeCas9 HNH domain in the presence of Mn^{2+} or Mg^{2+}. In addition, we incubated CjeCas9 and its active-site mutated variants with human chromosomal DNA isolated from U2OS cells. We found that CjeCas9 with a functional HNH domain and cations were sufficient to degrade human genomic DNA (**Figure S9.10**) confirming the results obtained with the plasmid DNA.

The observed RNA guide-independent processing of double stranded human DNA by CjeCas9

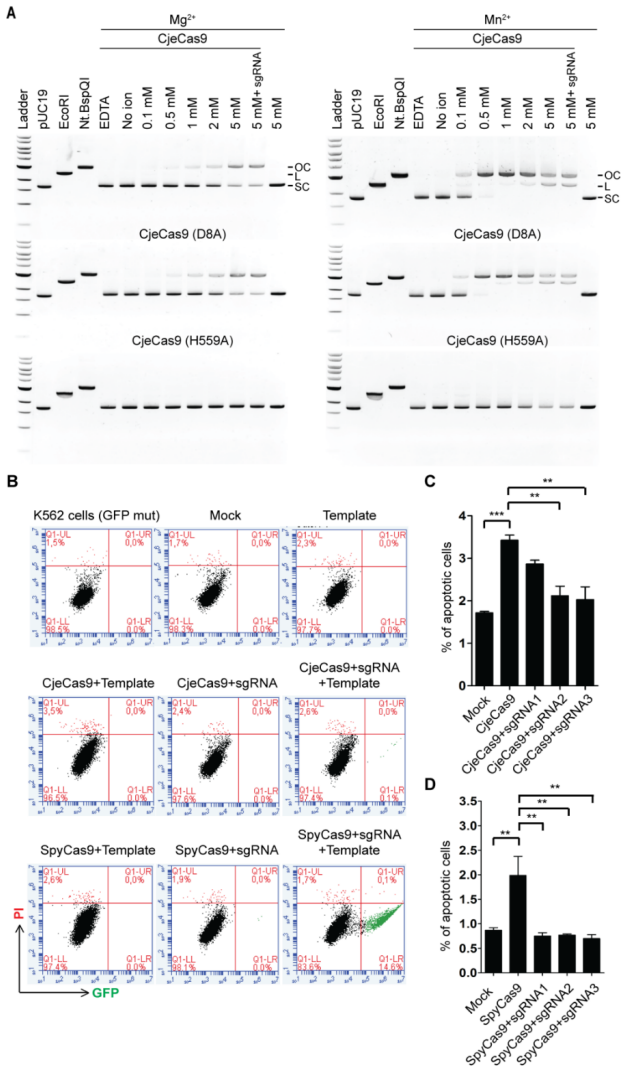


Figure 9.4 | In vitro cleavage assay and in vivo genome editing efficiencies of CjeCas9 and SpyCas9 in human cells. **(A)** CjeCas9 cleavage activity and the effect of divalent cations (Mn^{2+} and Mg^{2+}) concentration on plasmid DNA. **(B)** Restoration of GFP expression mediated by CjeCas9 or SpyCas9 editing in human cells (K562(GFPmut)) and quantified by FACS. (C and D) Percentages of apoptotic human cells (k562) obtained after six hours. Concentration of CjeCas9 or SpyCas9 is 30 pmol and for the sgRNA the concentrations are 30, 60 and 90 pmol (sgRNA1, sgRNA2 and sgRNA3, respectively). Results shown as the mean ± s.e.m., *** $p < 0.001$ and ** $p < 0.01$ by One-Way ANOVA (Bonferroni's Multiple Comparison test).

is very different from the reported guide-dependent targeting activity by SpyCas9, the most frequently used enzyme in genome editing applications (*254*). We tested to what extent the features observed for CjeCas9 also apply to SpyCas9. We first fused eGFP with full-length SpyCas9 or its arginine-rich BH motif using an eukaryotic expression vector, after which the vectors were transfected into human cells. Fluorescence microscopy analysis (**Figure S9.11A,B**) and Western blotting (**Figure S9.11C**) of the cytoplasmic and nuclear fractions revealed that like eGFP-CjeCas9, also eGFP-SpyCas9 does not require an exogenously introduced NLS, to reach the nucleus. However, γ-H2AX induction by guide-free SpyCas9 was significantly less, despite higher transfection efficiencies, compared to CjeCas9 (**Figure S9.12A to C**) Although *in vitro* analyses of plasmid and chromosomal DNA indicated that SpyCas9 does not cleave dsDNA as CjeCas9 does, *in vivo* analysis in human cells reveals that guide-free SpyCas9 does cause some DNA damage, most likely due to the guide-free ssDNA activity of transiently exposed regions of the chromosomes (**Figure S9.9B**).

Next, to compare genome editing efficiency of native SpyCas9 and CjeCas9 in human cells, we used an established experimental procedure where Cas9 editing restores a mutated GFP open reading frame in the AAVS1 locus, an event that can be detected by measuring restored GFP-expression by FACS (*58*). We found that genome editing by CjeCas9 is 100-fold less efficient as compared to SpyCas9 (**Figure 9.4B**, **S9.13A to D**), at least in part due to the reported poor helicase activity of this enzyme (*295*).

As a first approach to understand the mechanism behind this difference in editing efficiency, we used an unbiased proteomics approach to study the effect of Cas9 protein expression in human cells. We therefore transfected the human cells with expression vectors enabling the expression of active and inactive CjeCas9 and SpyCas9 proteins. The protein content and thus the cellular responses of human cells was analysed using an unbiased LC-MS based proteomics approach and showed that cell fate pathways appeared to be differentially regulated in cells exposed to guide-free versions of either CjeCas9 or SpyCas9. Of importance, pathways induced by SpyCas9 were more strongly associated with cell survival (**Table S9.5**, **S9.6**), including oxidative phosphorylation, DNA biosynthesis PI3K/AKT pathway that could have favoured the increased editing efficiency as observed for SpyCas9. In contrast, CjeCas9 transfection resulted in pathways that favoured apoptosis, including suppression of tRNA charging (**Table S9.5**, **S9.6**).

More detailed information was obtained after LC-MS based proteomics of transfected human cells in which Cas9 was expressed with or without a (non-targeting) sgRNA guide (**Table S9.7-S9.9**). This analysis suggested that eukaryotic cells exposed to SpyCas9 in the presence of sgRNAs accumulated DNA repair proteins associated with non-homologous end joining (XRCC5 and PARP1) (**Table S9.7-S9.9**), whereas cells exposed to CjeCas9 in presence of sgRNAs mainly accumulated DNA repair proteins associated with nucleotide excision pathway (*e.g.*, POLR2D and RPA1-3), both relative to exposure of cells with the corresponding dead Cas9 variants (**Table S9.7-S9.9**).

It was then demonstrated that saturation of native CjeCas9 with a sgRNA guide in case of CjeCas9 results in partially decrease of cell death (**Figure 9.4C**). Interestingly, in case of SpyCas9 cell death could be completely reduced to background level already with equimolar amounts of SpyCas9 protein and a sgRNA (**Figure 9.4D**). Compared to SpyCas9, binding of sgRNA guides appears to be less efficient in CjeCas9, possibly due to the smaller size of the latter protein, as suggested previously (*295*). Under the same conditions, also the effect of transfection of both nucleases was analysed by monitoring γ-H2AX activation in human cells (**Figure 9.4C,D**, **S9.13C,D**). Indeed, Western blot analysis of human cells exposed to CjeCas9 protein saturated with sgRNA showed considerable reductions of γ-H2AX activation, and even more so in case of SpyCas9 saturated with sgRNA (**Figure S9.14A to C**).

In conclusion, we reveal that Cas9 proteins can autonomously translocate into the nucleus of human cells. The consequence of this autonomous nuclear localization is that in the absence of guide RNAs, but in the presence of sufficient concentrations of intracellular Mn^{2+} and Mg^{2+} cations, DNA processing by Cas9 proteins occurs a-specifically. In this way the dsDNA nicking activity of CjeCas9 induces substantial genomic DNA damage that overwhelms human cells disabling their survival. Metal ion dependent and RNA-independent CjeCas9 nicking activity provides a clear mechanistic explanation for our previously observed cytotoxicity in *C. jejuni* infection experiments of human cells (*120*). Based on our data we therefore propose a working model to further study the role of Mg^{2+} and Mn^{2+} activated CjeCas9 in bacterial-induced

cytotoxicity (**Figure S9.15**). Further in-depth studies will then hopefully elucidate on which eukaryotic surface receptors CjeCas9 containing vesicles bind and how their content is released into the cytoplasm, a process that might be prone for the development of prophylactic measures. Although transduction of the SpyCas9 protein to human cells still may cause some DNA damage, most likely due to its ssDNA cleavage activity, this problem can be completely prevented by saturation with an appropriate guide, whereas this is only partly true for CjeCas9. Overall, we conclude that CjeCas9 has evolved towards a distinct natural function (attack of human cells rather than defence against viruses) and that its unique catalytic features may make this nuclease less suitable for genome editing applications.

Experimental Procedures

Bacterial strains and growth conditions

The used *C. jejuni* isolates are shown to harbour a type II-C CRISPR–Cas system, but GB11 lacks a CRISPR array (*120*, *429*) and the *C. jejuni* (cdt $^{-}$) strain only harbours *cas9*. All the used strains are listed together with their generated *cas9* variants in **Table S9.10**. The methods for generating Δ*cas9* and Δ*cas9::cas9* and growth conditions were described earlier (*120*). WT and the genetic variants strains were cultured on blood agar plates, containing 7% sheep blood (Becton Dickinson, Breda, The Netherlands) and the appropriate antibiotics (Sigma-Aldrich, Zwijndrecht, The Netherlands) for mutant selection (**Table S9.10**). Culturing under micro-aerophilic conditions was performed at 37 ℃, using anaerobic jars and an Anoxomat (Mart Microbiology B.V., Drachten, The Netherlands). Infection experiments occurred at a multiplicity of infection (M.O.I.) of 100 bacteria per 1 host cell unless stated otherwise.

Maintenance of human cell lines

Human U2OS cells (human osteosarcoma cells), K562 cells (human chronic myelogenous leukaemia) and HEK293T cells (human embryonic kidney cells) were maintained in Dulbecco's modified Eagle's medium (DMEM) (Thermo Fisher Scientific (TFS), Bleiswijk, The Netherlands) supplemented with 10% foetal bovine serum (FBS) (TFS, Bleiswijk, The Netherlands), 100 U/mL penicillin (TFS, Bleiswijk, The Netherlands), 100 µg/mL streptomycin (TFS, Bleiswijk, The Netherlands) and 1% non-essential amino acids (NEAA) (TFS, Bleiswijk, The Netherlands). The human cells were cultured in a 75-cm² flask (Greiner Bio-one, Alphen aan den Rijn, The Netherlands) at 37 ℃ and 5% CO2 in a humidified air incubator (Binder, Tuttlingen, Germany). All experiments were performed with the U2OS cell line unless stated otherwise. In example, HEK293T cells were used for technical reasons in experiments requiring a more loosely adhering cancer cell line for cytoplasmic and nuclear fraction separations. All cell lines were ordered from American type culture collection for human cell lines and were routinely tested for mycoplasma.

Isolation of *C. jejuni* outer membrane vesicles (OMVs)

C. jejuni OMVs were isolated as described previously (*428*, *438*). In brief, overnight *C. jejuni* cultures were centrifuged at 5,000 × g for 30 minutes, and the resulting supernatant was filtered across a 0.22-µm membrane (Corning, New York, USA). The filtrate was concentrated to 13 mL using an iCON™ Concentrator (TFS, Bleiswijk, The Netherlands) 20 mL/9K. The concentrated filtrate was ultra-centrifuged at 150,000 × g for three hours at 4 ℃ using a SW41 Ti Rotor (Beckman Coulter, Woerden, The Netherlands). All isolation steps were carried out at 4 ℃. The pellet was resuspended in PBS (TFS, Bleiswijk, The Netherlands) and stored at –80 ℃. OMV samples were pipetted onto blood agar plates containing 7% sheep blood (Becton Dickinson, Breda, The Netherlands) and incubated under both micro-aerophilic and aerobic conditions for

48 hours to confirm the absence of viable bacteria. OMVs were quantified using the EVQuant (*444*).

Generation of Cas9-mCherry fusion proteins

The *C. jejuni* Cas9-mCherry strain (**Table S9.10**) was generated by overlap extension PCR as described previously (*445*). In brief, the primers used to amplify *cas9* introduced overlapping ends for the mCherry gene (**Table S9.11**). By using overlap PCR, we generated a single PCR product combining the individual PCR fragments for Cas9 and mCherry and a kanamycin resistance gene to allow for antibiotic selection. The single Cas9-mCherry PCR product generated was then introduced into the native *cas9* locus of *C. jejuni* by electroporation to generate an mCherry tagged Cas9 strain by homologous recombination.

In silico Cas9 protein analyses

The Cas9 protein sequences from *Streptococcus pyogenes* M1-GAS (AAK33936.1), *Staphylococcus aureus* subsp. *aureus* (CCK74173.1), *Listeria monocytogenes* SLCC2482 (CBY05127.1), *Francisella tularensis* subsp. *novicida* U112 (ABK89648.1), *Neisseria meningitidis* 8013 (C9X1G5.1), *Helicobacter cinaedi* ATCCBAA-847 (YP_007601284.1), *Campylobacter lari* CF89-12 (BAK69486.1), *Campylobacter jejuni* NCTC11168 (YP_002344900.1) and *Campylobacter jejuni* GB11 (*120*) were uploaded into the online MyHits motif scan software tool (http://myhits.isb-sib.ch/cgi-bin/motif_scan). The NLS motifs obtained were aligned using MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/) and the alignment was visualized and evaluated for identity of amino acids with similar biochemical characteristics using MEGA6 software (*446*).

Cloning, DNA isolation and amplification of Cas9 proteins

The coding sequences of the Cas9 proteins of *C. jejuni* and *S. pyogenes* or their NLS motifs were cloned into the pEGFP-C1 (Clontech, Saint-Germain-en-Laye, France) vector for co-expression in eukaryotic cells. The pCDNA 3.1(+) vector (Invitrogen, Breda. The Netherlands) was used to express the Cas9 proteins of *C. sjejuni* or *S. pyogenes* in eukaryotic cells. The *SpyCas9* or its catalytically inactive (”dead”) variant, *spydcas9* gene were cloned from the plasmids pMJ806 and pMJ84, a kind gift from Prof. J. Doudna, UC Berkeley, USA (*51*). Primers and plasmids used in this study are listed in **Table S9.11** and **Table S9.12**, respectively. Presence of the CRISPR array, *cas* and *cdt* genes were analysed by PCR using previously described methods (*120*, *447*, *448*). Genomic DNA was isolated using the QIAamp® DNA tissue stool kit (QIAGEN, Venlo, The Netherlands) and PCR amplification was performed using a Biomed Thermal Cycler System 9700 (TFS, Bleiswijk, The Netherlands). DNA digestion with restriction enzymes (New England Biolabs, Leiden, The Netherlands), ligation with a T4 DNA ligase (TFS, Bleiswijk, The Netherlands), purification and agarose gel electrophoresis were performed according to the manufacturer’s protocols. Primer pairs and corresponding restriction enzymes used in this study are listed in **Table S9.11**. The resulting constructs were electroporated at 2.5 kV, 200 Ω, 25 µF into *Escherichia coli* Top10 cells (TFS, Bleiswijk, The Netherlands), resuspended in 37 ℃ pre-warmed SOC medium (TFS, Bleiswijk, The Netherlands), and allowed to recover by gentle shaking at 37 ℃. After recovery, 100 µL was plated onto a Lysogeny broth (LB) (Becton Dickinson, Breda, The Netherlands) agar plate containing the appropriate selection marker (**Table S9.12**). Positive colonies were cultured for plasmid DNA isolation (TFS, Bleiswijk, The Netherlands) and plasmids were subjected to sequencing using a Big Dye Terminator sequencing kit (TFS, Bleiswijk, The Netherlands) and an ABI Prism 3100 Genetic Analyser (TFS, Bleiswijk, The Netherlands), according to the manufacturer’s instructions.

Cas9 mutagenesis

A PCR based site directed mutagenesis method (*445*) was used to generate a catalytically inactive ("dead") Cas9 (D8A and H559A mutated) variant in *C. jejuni* Cas9 that was cloned into the pEGFP-C1 vector to obtain a GFP in-frame fusion protein at the C-terminal end of CjeCas9. Primers are listed in **Table S9.11**. Plasmids were transformed into chemically competent *E. coli* TOP10 cells (TFS, Bleiswijk, The Netherlands). After plasmid multiplication and isolation, presence of the correct genetic mutations was checked by DNA sequencing.

Plasmid transfection

Human cells were seeded onto a two well chamber slide (Sanbio, Uden, the Netherlands) at a density of 5.0 x 10^4 cells per millilitre. The next day, cells were transiently transfected with plasmid DNA using X-tremeGENE DNA Transfection Reagent (Roche Applied Science, Woerden, The Netherlands), according to the manufacturer's protocols. After the incubation periods at 37 ℃ and 5% CO2 in a humidified incubator (Binder, Tuttlingen, Germany), the human cells were washed with pre-warmed HBSS (TFS, Bleiswijk, The Netherlands) and fixed with 4% paraformaldehyde (Sigma-Aldrich, Zwijndrecht, The Netherlands) for immunocytochemistry and microscopy as described previously (*447*).

SDS-PAGE and Western Blot analysis

OMVs were analysed for the presence of CjeCas9 and CjeCas9(mCherry) using SDS-PAGE gel followed by Western blotting and staining with a *C. jejuni* anti-Cas9 antibody (*120*). Plasmid-transfected cells were washed with HBSS (TFS, Bleiswijk, The Netherlands) and cell lysates were prepared by adding Bond-Breaker® TCEP solution (TFS, Bleiswijk, The Netherlands) in 2X Laemmli sample buffer (Bio-Rad, Veenendaal, The Netherlands) (1:9) before loading on a SDS-PAGE gel. Cellular lysates were then denatured at 95 ℃ for 5 min and immediately run through 4 - 20% Mini-PROTEAN® TGX™ Precast Protein Gels (Bio-Rad, Veenendaal, The Netherlands). After electrophoresis the proteins were blotted onto a polyvinylidene difluoride (PVDF) nitrocellulose membrane (Millipore, Amsterdam, The Netherlands). The membranes were blocked with PBS (TFS, Bleiswijk, The Netherlands) containing 5% non-fat dry milk (Bio-Rad, Veenendaal, The Netherlands) for 2 - 3 hours at room temperature or overnight at 4 ℃. For detection of CjeCas9, SpyCas9 or their GFP-Cas9 fusion proteins, the membranes were incubated with primary antibodies against CjeCas9 (*120*), mouse-anti-SpyCas9 (Abcam, Cambridge, United Kingdom), or rabbit polyclonal anti-GFP (NB600-308; Novus Biologicals, Abingdon, United Kingdom) at a dilution of 1:1000 and incubated for 1.5 - 2 hours at room temperature. Appropriate secondary antibodies with alkaline phosphatase (Sigma-Aldrich, Zwijndrecht, The Netherlands) were used at a dilution of 1:1000 to visualize the primary antibody binding using NBT/BCIP solution (Sigma-Aldrich, Zwijndrecht, The Netherlands) according to the manufacturer's instructions. For cytoplasmic and nuclear protein fractions NE-PER™ Nuclear and Cytoplasmic Extraction Reagents (TFS, Bleiswijk, The Netherlands) were used and proteins from eukaryotic cells were extracted according to the manufacturer's instructions. Cytoplasmic and nuclear proteins were separated in a 4 - 20% Mini-PROTEAN® TGX™ Precast Protein Gel (Bio-Rad, Veenendaal, The Netherlands), Western blotted and incubated with the above mentioned primary Cas9 antibodies, or a mouse monoclonal anti-phospho-Histone H2A.X (Ser139) clone JBW301 (Millipore, Amsterdam, The Netherlands) at a 1 : 1000 dilution or a mouse monoclonal anti-GAPDH antibody (Abcam, Cambridge, United Kingdom) at a dilution of 1 : 5000 to detect the presence of Cas9, presence of DNA DSBs or control for fraction quality and equal loading. Visualization occurred with an NBT/BCIP solution (Sigma-Aldrich, Zwijndrecht, The

Netherlands) according to the manufacturer's instructions.

In vitro infection assays using glass chamber slides for microscopy

Cells were seeded onto two well chamber slides (Greiner Bio-one, Alphen aan den Rijn) at a density of 1.5 x 10^5 cells per well and incubated overnight at 37 ℃ with 5% CO_2 in a humidified incubator (Binder, Tuttlingen, Germany). Cells were exposed to GB11 or their derived genetic variants or *C. jejuni* bacteria lacking *cdt* (CDT^-). After overnight incubation at 37 ℃ with 5% CO_2 in a humidified incubator (Binder, Tuttlingen, Germany) human cells were washed three times with pre-warmed HBSS (TFS, Bleiswijk, The Netherlands) at 37 ℃ and fixed with 4% paraformaldehyde (Sigma-Aldrich, Zwijndrecht, The Netherlands). Fixed cells were prepared for γ-H2AX or 53BP1 immunocytochemistry (see below) and/or (fluorescence) microscopy analyses as described earlier (*447*).

Immunocytochemistry

For immunofluorescence detection, plasmid transfected human cells were washed and fixed with 4% paraformaldehyde before permeabilization with 0.1% Triton X-100 (Sigma-Aldrich, Zwijndrecht, The Netherlands). For γ-H2AX or 53BP1 antibody detection, human cells were incubated with a mouse anti-γ-H2AX antibody (Millipore, Amsterdam, The Netherlands) or rabbit polyclonal anti-53BP1 (NB100-304) (Novus Biologicals, Abingdon, United Kingdom) at a 1:1000 dilution. Secondary labelling and slide preparation occurred as described previously (*447*). To visualize *C. jejuni* in human cells, an anti-*C. jejuni* fluorescein isothiocyanate (FITC)-labelled antibody (Genway, San Diego, USA) was used. Labelling, detection of intracellular *C. jejuni* bacteria and slide preparation occurred as previously described (*447*). Images were taken using a XI51 phase-contrast fluorescence microscope (Olympus, Leiden, The Netherlands) and further analysed using Olympus CellSens or ImageJ/FIJI software (*449*).

Generation of Heterologous Plasmids for Cas9 protein expression

For heterologous expression, the *CjeCas9* gene from the pEC-K-CBP-NTAP plasmid (generous gift from Martin Jinek) and the *SpyCas9* gene from the pET-28b-Cas9-His plasmid [obtained from Alex Schier (Addgene plasmid # 47327)] were PCR amplified using Q5 DNA polymerase (New England Biolabs, Leiden, The Netherlands). The amplified genes were inserted into the plasmid pML-1B (obtained from the UC Berkeley MacroLab, Addgene #29653) or pET28b backbone by NEBuilder® HiFi DNA Assembly (New England Biolabs, Leiden, The Netherlands) using oligonucleotides shown in (**Table S9.11**) to generate a protein expression construct encoding the SpyCas9 or CjeCas9 polypeptide sequence fused with an C-terminal (for SpyCas9) and N-terminal tag (for CjeCas9) comprising a hexahistidine sequence and a Tobacco Etch Virus (TEV) protease cleavage site. The assembly mix was used to transform the competent *E. coli* DH5α strain (New England Biolabs, Leiden, The Netherlands). The plasmids were isolated and verified by Sanger sequencing (Macrogen, Amsterdam, The Netherlands) before transforming the *E. coli* Rosetta™ 2 (DE3) strain. The plasmids containing the wild type genes were used as PCR templates to generate active site mutants of SpyCas9 and CjeCas9 via site-directed mutagenesis. In SpyCas9, D10A, H840A, and D10A-H840A amino acid substitutions yielded a RuvC, HNH and double mutant, respectively. In CjeCas9, D8A, H559A and D8A-H558A amino acid substitutions yielded a RuvC, HNH and double mutant, respectively. All bacterial expression plasmids used in this study can be found in **Table S9.12**.

Cas9 expression and purification

The *CjeCas9* and *SpyCas9* genes were heterologous expressed in *E. coli* and purified using a combination of Ni^{2+} affinity, cation exchange and gel filtration chromatography steps. Three litres of LB growth medium with 100 µg mL^{-1} ampicillin was inoculated with 30 mL overnight culture of Rosetta (DE3) (EMD Millipore, Ontario, Cananda) cells containing the expression constructs. Cultures were grown to an OD_{600} nm 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 ℃ 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, Woerden, The Netherlands). Cells were lysed by sonication and the lysates were centrifuged for 45 min at 4 ℃ at 30,000x g to remove insoluble material. The clarified lysate was applied to a 5 mL HisTrap HP column (GE Healthcare, Eindhoven, The Netherlands). 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 ℃ 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, Eindhoven, The Netherlands). 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 analysed by SDS-PAGE and fractions containing the Cas9 protein were combined, and DTT (Sigma-Aldrich, Zwijndrecht, The Netherlands) 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, Eindhoven, The Netherlands) 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 ℃ for storage.

GFP expression and FACS-analysis

For genome editing and FACS analyses, we used the K562(GFPmut) cell line (*58*), which was seeded into a 24-well plate (Greiner Bio-one, Alphen aan den Rijn, The Netherlands). After overnight recovery at 37 ℃ with 5% CO_2 in a humidified incubator (Binder, Tuttlingen, Germany) we transfected the K562(GFPmut) cells using Lipofectamine 2000 (TFS, Bleiswijk, The Netherlands) and a RNP complex of the bacterial CjeCas9 and SpyCas9 (New England Biolabs, Leiden, The Netherlands) bound to their respective synthetic guide RNA (Biolegio, Nijmegen, The Netherlands) and a PCR product generated from the GFP gene. Production of active RNP complexes and lipofection were carried out according to the manufacturer's protocols. The synthetic guide RNA sequence for SpyCas9 (gRNA_AAVS1-T2) was described earlier (*58*) and CjeCas9 sgRNA listed in **Table S9.11**. Seventy two hours after transfection the K562(GFPmut) cells were harvested using ice cold HBSS (TFS, Bleiswijk, The Netherlands) and kept on ice; after the second wash the cells were resuspended in 0.4 mL HBSS (TFS) containing propidium iodide used from an apoptosis detection kit (GeneCopoeia, Rockville, USA). After 20 minutes the K562(GFPmut) cells were analyzed by FACS (BD Accuri C6) and BD Accuri C6 Software (version 8.8.6).

DNA double strand break (DSB) detection by BLESS

U2OS cells were seeded into 75 cm² flasks (± 1-5 million cells) (Greiner Bio-one, Bleiswijk, The Netherlands) and allowed to grow to confluence at 37 ℃ with 5% CO_2 in a humidified incubator (Binder, Tuttlingen, Germany). Cells were infected with GB11, GB11Δ*cas9* or GB11Δ*cas9*::*cas9* bacteria in a humidified incubator (Binder, Tuttlingen, Germany) at 37 ℃ with 5% CO_2 and after six hours, samples were fixed and prepared for sequencing according to the BLESS protocol (*440*) to obtain a snapshot of induced DSBs. This procedure was repeated for plasmid transfected cells which were harvested and processed for sequencing after 24 or 48 hours, respectively. As a positive control for DSB induction, a U2OS cell line harbouring a stably integrated genomic I-SceI restriction site was used according to the original BLESS protocol (*440*). Additionally, gamma (γ) irradiation was applied for random induction of DSBs. U2OS cells received 1 Gy of ^{137}Cs gamma radiation at a dose rate of 0.6 Gy min^{-1}. After irradiation, cells were allowed to recover for 30 minutes at 37 ℃ with 5% CO_2 in a humidified incubator (Binder, Tuttlingen, Germany) and further processed according to the original BLESS protocol. Paired-end sequencing of samples with Illumina HiSeq2500 v4 was performed according to the manufacturer's protocol (Illumina) by an accredited service provider (ServiceXS, Leiden, The Netherlands). The resulting FastQ files were demultiplexed, and 75 base pair genomic sequences including the BLESS primers (*440*) and sequencing barcodes were removed. The resulting FastQ files were mapped as paired reads per sample using BWA MEM (version 0.7.12.1) against the GRCh37/hg19 human genome assembly. Duplicate reads were flagged using Picard Mark Duplicates (version 1.136.0) and high-quality duplicate reads with a mapping quality score >30 were selected using SAM tools (version 1.19); the forward read from these sequences including genomic location were stored as ascii files. The frequency of reads at the same genomic location was determined for each sample, and only those positions for which 4 reads or more were mapped to DSBs were selected for further analysis. The frequency results from the experimental control sample were used to filter non-specific DSBs from all other test samples based on genomic location. Reads associated with positive DSBs were visualized using the Integrated Genome Viewer (version 2.3) and a summary of the breakpoint comparisons were visualized as Circos plots generated using the R (version 3.1.1) circlize package (version 0.3.8).

PAM motif identification

The identified DSB positions on the GRCh37/hg19 genome obtained by BLESS were used to extract nucleotides upstream and downstream of this position using the FETCH software tool to Extract Genomic DNA using coordinates from assembled/unassembled genomes (Galaxy Version 2.2.4) in Galaxy (http://usegalaxy.org). The extracted sequences, ten base pairs upstream or downstream of the cleavage site obtained from the raw and filtered BLESS datasets, were analysed from 5' - 3' and from 3' - 5' for the presence of a potential PAM motif using WebLogo (http://weblogo.berkeley. edu/logo.cgi) or manually in Excel (Microsoft) using frequency counting of the nucleotides A, T, C, G at each individual position upstream or downstream of the break position by text filter options. This was done since the BLESS technique can result in filling or removal of overhanging nucleotides at the DSB position that can change the potential PAM position (*440*). The extracted sequences were further analysed.

Proteomics

HEK293T cells were seeded onto a 6-well plate (Greiner Bio-One, Alphen aan den Rijn, The Netherlands) and allowed to grow to confluence. The next day, cells were transiently transfected with plasmid DNA using X-tremeGENE DNA Transfection Reagent (Roche Applied Sci-

ence, Woerden, The Netherlands), according to the manufacturer's protocols. After an incubation period of 24 hours at 37 ℃ and 5% CO2 in a humidified incubator (Binder, Tuttlingen, Germany). Total cell lysates were run in a 4-20% Mini-PROTEAN® TGX™ Precast Protein Gel (Bio-Rad, Veenendaal, The Netherlands). Sample preparation, nanoflow LC-MS analyses, protein identification and quantification, data management and statistical analyses were all carried out as described previously (*450*, *451*). Downstream analysis was performed by comparing protein abundance data for CjeCas9 over its inactivated mutant CjedCas9, SpyCas9 over SpydCas9, and each Cas9 protein bound to sgRNAs (**Table S9.11**) over the corresponding proteins without sgRNAs, using the Maxquant/Perseus software suite, and calculating which human cellular pathways were significantly induced or repressed by Ingenuity Pathway Analysis (QIAGEN Bioinformatics, 2018 release) software. To research if cellular pathways were differentially induced or repressed in cells upon exposure to Cas9 activity, in presence or absence of sgRNAs, protein abundances of CjeCas9 or SpyCas9 minus protein abundances of CjedCas9 or SpydCas9 were calculated for the first comparison, and CjeCas9+sgRNA or SpyCas9+sgRNAs minus CjeCas9 or SpyCas9 were calculated for the second comparison. Ingenuity Pathway Analysis (IPA; QIAGEN Bioinformatics, 2018 release) was used to identify significantly regulated pathways using default settings. Significantly modulated pathways, with –log(*p*-values) > 1.30 ($p<0.05$) are displayed in **Table S9.5** to **Table S9.9** together with an overview of differential abundances of proteins participating in nucleotide excision repair (NER) and non-homologous end-joining (NHEJ) pathways.

In vitro Cas9 U2OS genomic DNA digestion assays

Genomic DNA cleavage assays were performed in a final volume of 30 µL with 300 ng of DNA isolated from the U2OS cell line. DNA was isolated from cultured cells by DNeasy Blood & Tissue Kit (QIAGEN, Venlo, The Netherlands), then treated with RNase A (Promega Benelux B.V., Leiden, The Netherlands) and purified. Total 30 pmol of Cas9 proteins (final concentration of 1 µM) were used for cleavage reaction with or without different metal ions (Mg^{2+}/Mn^{2+}) in the cleavage buffer (20 mM HEPES (pH 7.5), 150 mM KCl, 0.5 mM DTT, 0.1 mM EDTA) and incubated at 37°C for one or six hours. Different restriction enzymes (New England Biolabs, Leiden, The Netherlands) were used with their appropriate buffer as a control. The reactions were stopped by adding proteinase K and 6X DNA loading dye Purple (New England Biolabs, Leiden, The Netherlands). The genomic DNA was resolved on an agarose gel (0.8%) or denaturing urea-agarose gel (1%) stained with SYBRTM Gold Nucleic Acid Gel Stain (TFS, Bleiswijk, The Netherlands) and visualized using an Amersham Typhoon Gel and Blot Imaging Systems (GE Healthcare, Eindhoven, The Netherlands). To denature genomic DNA the reactions were stopped at indicated time points by adding 40 µL denaturing sample buffer containing 8 M Urea and 0.3% SDS. Samples were denatured by incubation for ten minutes at 85 ℃ and loaded directly on a 1% urea (2 M) agarose gel in TAE buffer as described (*452*).

Plasmid cleavage assay

Plasmid cleavage assays were performed in a final volume of 20 µL with 6 nM of plasmid DNA (pUC19) and 100 nM protein in the presence of 10 mM EDTA or 5 mM $MgCl_2$ or $MnCl_2$ at 37°C for one hour. Other metal salts were tested but didn't show any activation of Cas9. As a control, the reaction was performed without any added metal or EDTA to account for any fortuitous metal associated with the protein preparation. The reactions were stopped by adding 10 mM EDTA and 1% SDS after which 6X DNA loading dye (New England Biolabs, Leiden, The Netherlands) was added. The DNA was resolved on an agarose gel (0.8%) stained

with SYBRTM Safe DNA Gel Stain (TFS, Bleiswijk, The Netherlands) and visualized using a G:BOX F3 gel imager (Syngene, Bangalore, India). To create DNA mobility standards, pUC19 was separately treated with a linearizing enzyme EcoRI (New England Biolabs, Leiden, The Netherlands) or a nicking enzyme Nt. BspQI (New England Biolabs, Leiden, The Netherlands). Both EcoRI and Nt. BspQI have a single recognition site in the pUC19 plasmid.

Acknowledgements

We would like to acknowledge the contributions of Prof. Anna Akhmanova, Prof. Jennifer Doudna, Dr. Tokameh Mahmoudi and Mir Mubashir Khalid for kindly providing plasmids and/or antibodies. We would like to thank Prof. Dr. Cynthia M. Sharma and Dr. David S. Weiss for helpful discussions. C.S. is a graduate student at Erasmus Postgraduate School of Molecular Medicine and is partially supported by I&I Fund (Erasmus Vrienden Fonds). S.B. is supported by the Ride for the Roses Cancer Research Grant obtained from the Dutch Cancer Society. J.V.D.O. is financially supported by the Netherlands Organization of Scientific Research (NOW, TOP grant 714.015.001), and R.L. by the Department of Medical Microbiology and Infectious Diseases and the Department of Bioinformatics, Erasmus MC; Author contributions: C.S., P.M., G.J.K., J.H.G.L., D.V.G., J.W.M., P.V.B., J.V.D.O., P.J.V.D.S., and R.L. contributed to the conception and design of the project; C.S., P.M., G.D., G.J.K., W.A.V.C., D.H.K., C.L. S.B., D.G., D.B., M.K., R.J., and R.L. performed the experiments and analysed the data; A.S. and Y.H. performed the sequencing analysis; J.H.G.L., D.V.G., J.W.M., P.J.V.D.S., J.V.D.O. and R.L. edited the manuscript; C.S., P.M., P.V.B., J.V.D.O. and R.L. wrote the manuscript, which was read and approved by all authors.

Author information

Correspondence should be addressed to john.vanderoost@wur.nl or p.vanderspek@erasmusmc.nl.

Competing interests

Related patent applications have been filed. All authors declare that there is no competing interest related to this work.

Data availability

Sequencing data are available at European Nucleotide Archive under accession number PRJEB34300.

Supplementary Figures and Tables

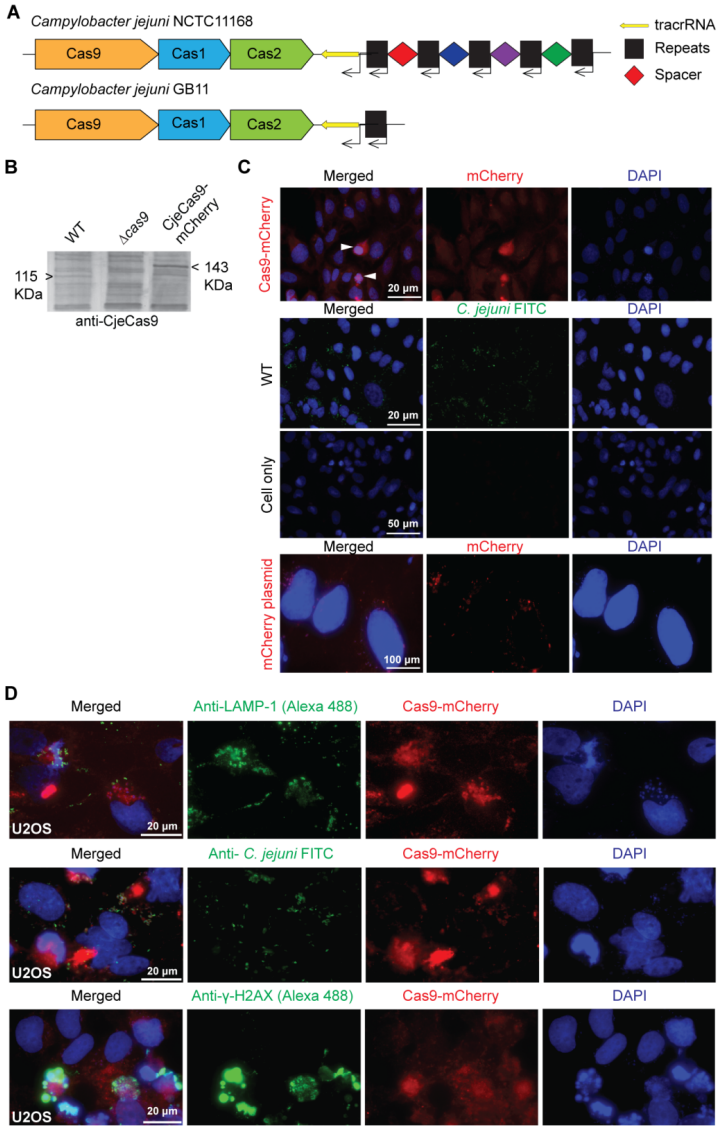


Figure S9.1|During human cell infections, CjeCas9 is detected in OMVs and accumulates in the host cell cytoplasm and nucleus. (A) Overview of the type II-C CRISPR–Cas system in our model strain GB11. Note the absence of the CRISPR array in GB11. **(B)** Western blot showing presence of CjeCas9 (WT) and CjeCas9-mCherry in *C. jejuni* produced OMVs. **(C)** *C. jejuni* bacteria CjeCas9-mCherry producing (red) or WT (green) infecting human cells. mCherry plasmid shows human cells expressing mCherry only (red). Nuclei were stained with DAPI (blue). Co-localization of CjeCas9-mCherry and nucleus indicated by white arrows. (D) *C. jejuni* bacteria expressing CjeCas9-mCherry (red) in human cells positive for lysosomal associated membrane protein 1 (LAMP-1) (green). (E) *C. jejuni* CjeCas9-mCherry producing bacteria (green) and their associated CjeCas9-mCherry expression (red). (F) Chromatin-accumulated γ-H2AX (green) is associated with CjeCas9-mCherry accumulation (red). Nuclei were stained with DAPI (blue).

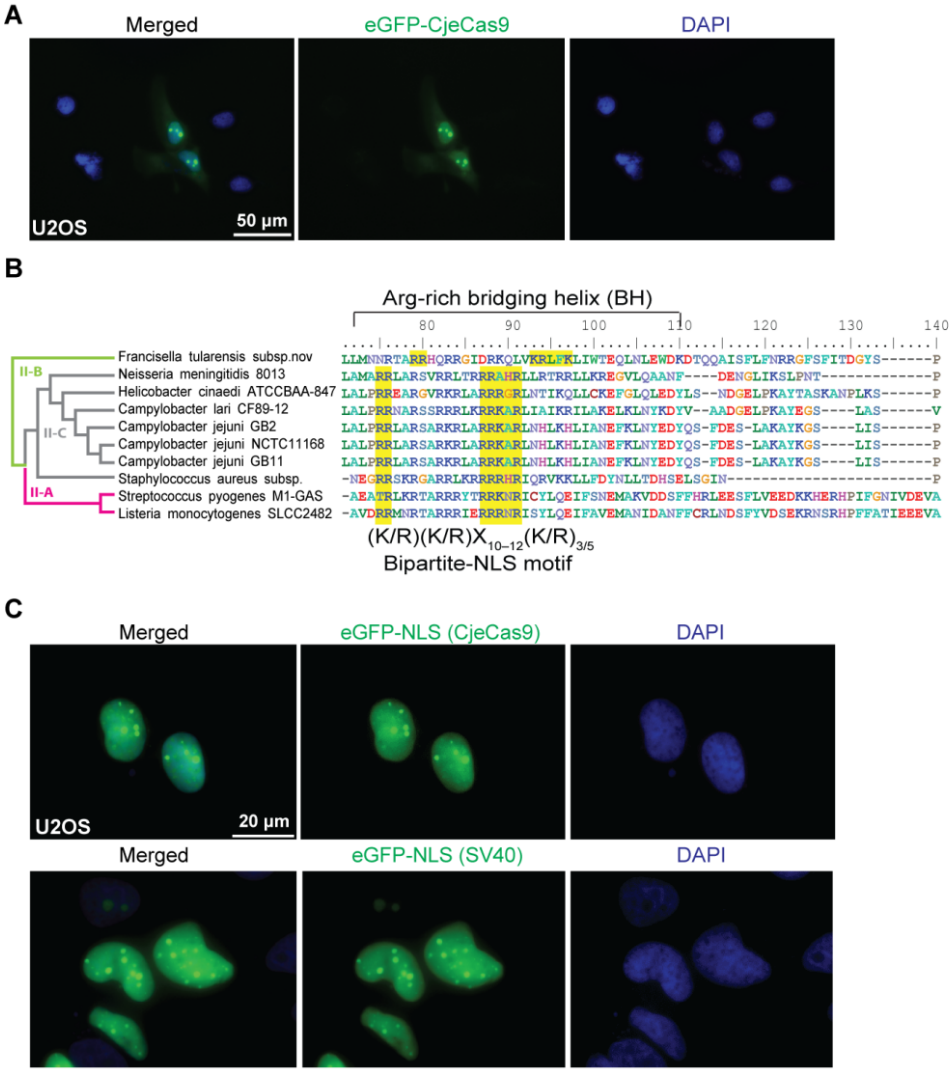


Figure S9.2 | GFP-CjeCas9 autonomously localizes into the eukaryotic nucleus. (A) Microscopic image of eGFP-CjeCas9 localization in the nuclei of human cells. **(B)** Multiple sequence alignment of orthologous Cas9 proteins produced by bacteria harbouring a type II CRISPR–Cas system; a conserved arginine-rich motif containing a bipartite NLS motif (highlighted in yellow) are annotated. The phylogenetic relatedness between the clusters of subtypes of type II [type II-A (purple); type II-B (green) and type II-C (grey)] CRISPR–Cas systems as determined by Maximum Likelihood is displayed as a dendrogram. **(C)** Microscopic image of eGFP localization (green) in the nucleus of human cells represented by eGFP fused to the NLS motif identified in CjeCas9 or eGFP fused to the NLS of the Simian Virus (SV)40 large T-antigen protein. NLS: nuclear localization signal. **(A)** and **(C)** Nuclei were stained with DAPI (blue).

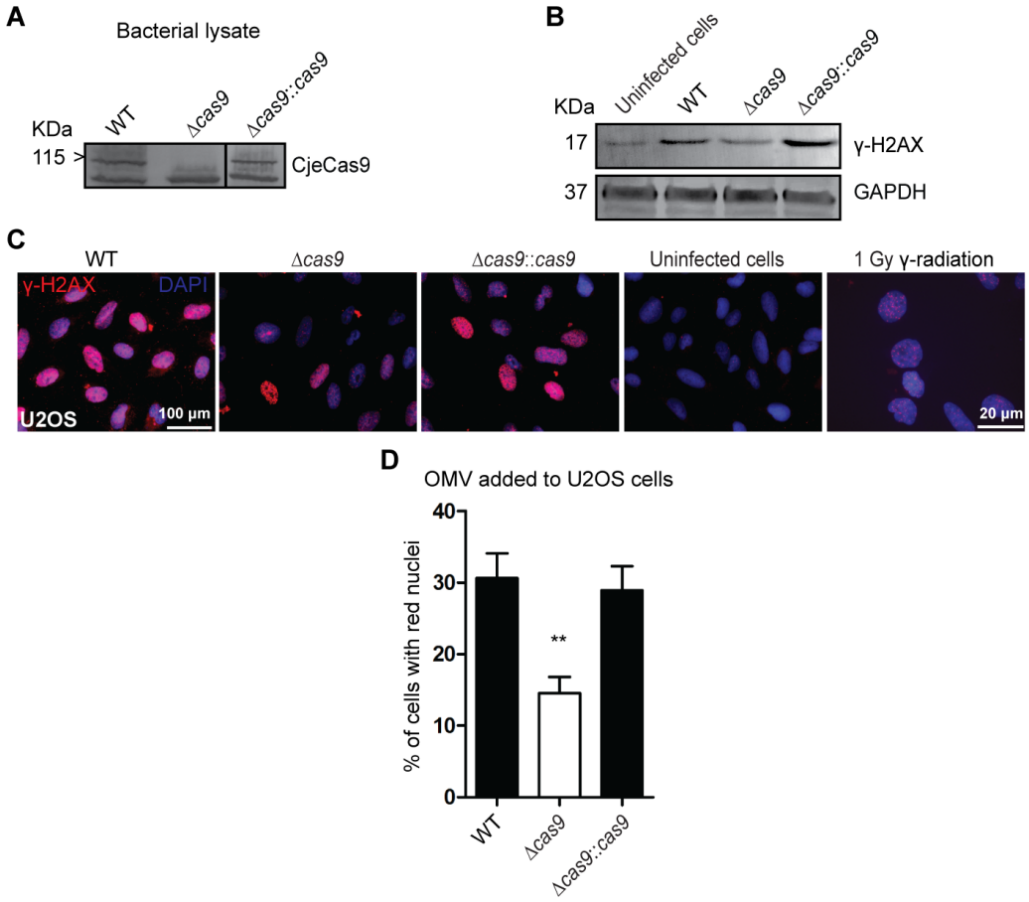


Figure S9.3|γ-H2AX activation is increased in human cells exposed to CjeCas9-producing bacteria. (A) Western blot analysis of *C. jejuni* bacterial lysates six hours post-infection of human cells reveals active CjeCas9 expression at 115kDa in the WT and Δ*cas9*::*cas9* bacteria. **(B)** Western blot shows six hours post-infection γ-H2AX activation in human cells exposed to WT or Δ*cas9*::*cas9* bacteria. GAPDH was used as a loading control. **(C)** γ-H2AX accumulation (red) in human cells exposed to OMVs produced by WT, Δ*cas9* or Δ*cas9*::*cas9* bacteria. 1 Gy dose of γ-radiation is used as a positive control. Nuclei were stained with DAPI (blue). **(D)** Quantification of γ-H2AX production in human cells (>100 analysed); Y-axis shows the percentage of human cells positive for bright red nuclei. The X-axis shows the corresponding *C. jejuni* bacteria from which the OMVs were isolated. Results shown as the mean ± s.e.m., ** $p < 0.01$ by One-Way ANOVA (Bonferroni's Multiple Comparison test).

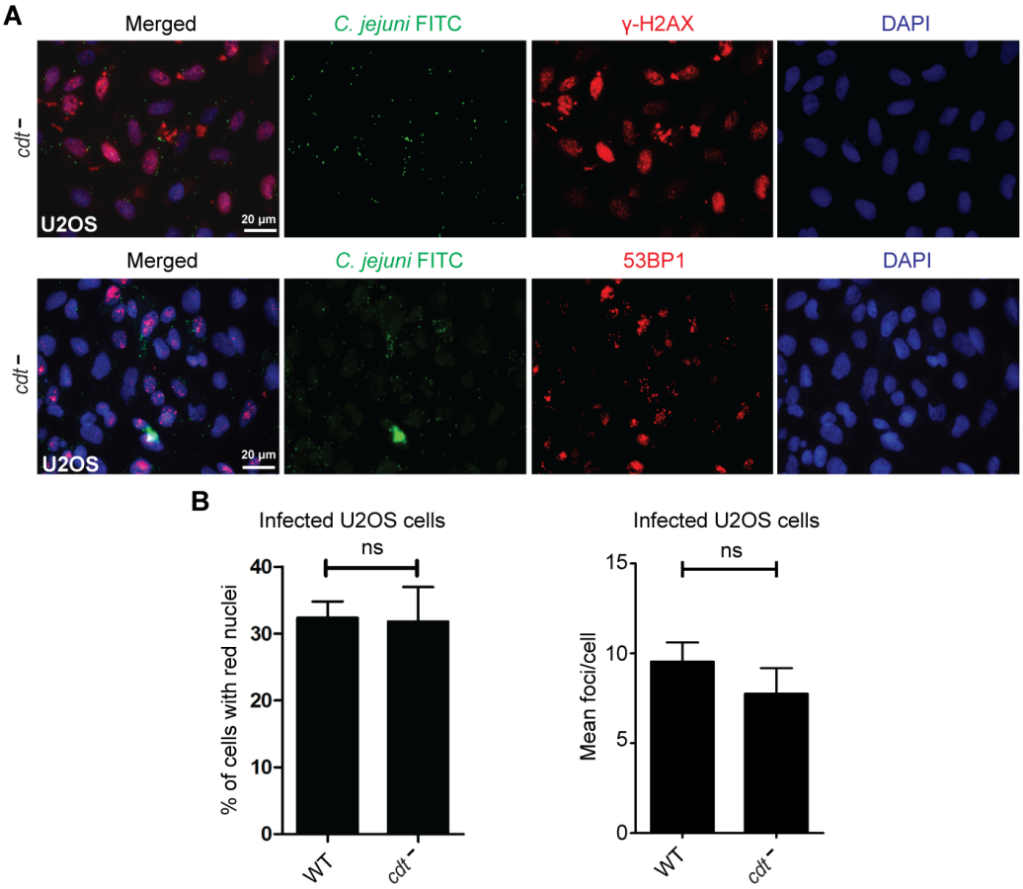


Figure S9.4 | CDT plays a minor role in the induction of DSBs in human cells. (A) γ-H2AX or 53BP1 accumulation in human cells (both in red after exposure to Cas9 positive *C. jejuni* bacteria (WT) or Cas9 positive bacteria that lack CDT (cdt $^{-}$); *C. jejuni* bacteria (WT) or (cdt $^{-}$) (green). Nuclei were stained with DAPI (blue). Of note (cdt $^{-}$) only harbors Cas9, but is lacking a CRISPR array, *cas2* and *cas1*. **(B)** Quantification of γ-H2AX or 53BP1 production in human cells (>100 analysed) exposed to bacteria (WT) or (cdt $^{-}$); Y-axis shows the percentage of human cells positive for bright red nuclei or mean number of foci per cell. The X-axis shows the corresponding *C. jejuni* strain used. Results shown as the mean ± s.e.m, ^{ns} $p > 0.05$ by two-tailed Student's t-test.

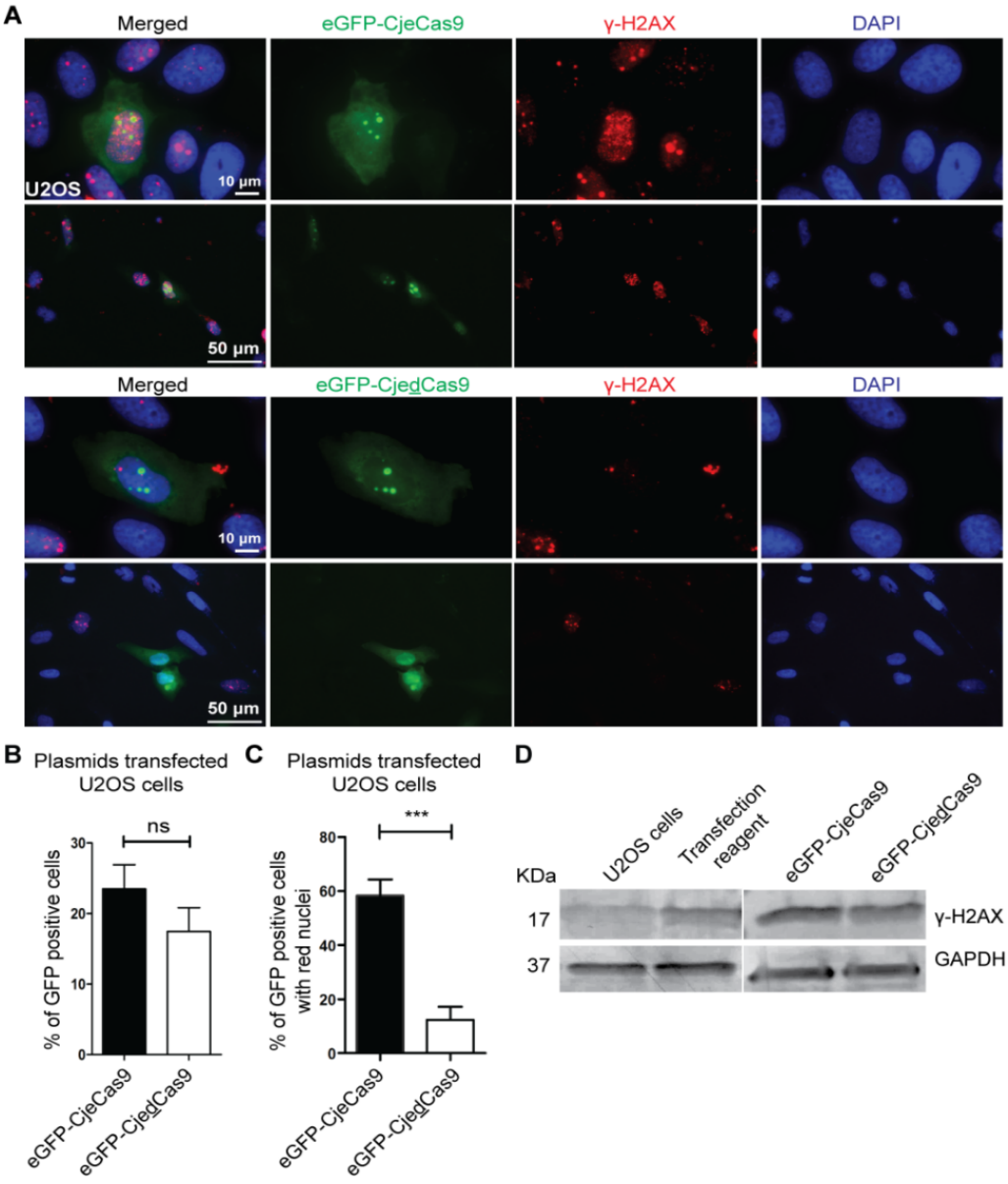


Figure S9.5 | GFP-CjeCas9 expression in human cells is associated with GFP and γ-H2AX accumulation in the nuclei. (A) γ-H2AX accumulation (bright red nuclei) in red in human cells expressing eGFP-CjeCas9 or eGFP-CjedCas9 in green. Nuclei were stained with DAPI (blue). **(B)** Y-axis shows the percentage of GFP positive cells transfected with eukaryotic expression plasmids encoding active (left) GFP-CjeCas9 or inactive ("dead"; right) GFP-CjedCas9 (X-axis); results shown as the mean ± s.e.m., ^{ns} $p > 0.05$ by two-tailed Student's *t*-test. **(C)** Y-axis shows the percentage of viable, GFP positive cells that displayed γ-H2AX activation as seen by bright red nuclei. X-axis shows the used plasmids encoding active (left) GFP-CjeCas9 or inactive ("dead"; right) GFP-CjedCas9; results shown as the mean ± s.e.m., ^{ns} $p > 0.05$ by two-tailed Student's *t*-test. **(D)** Western blot analysis of γ-H2AX production in human cells, left untreated (U2OS cells), exposed to the used transfection reagent or after transfection with plasmids expressing eGFP-CjeCas9 or eGFP-CjedCas9.

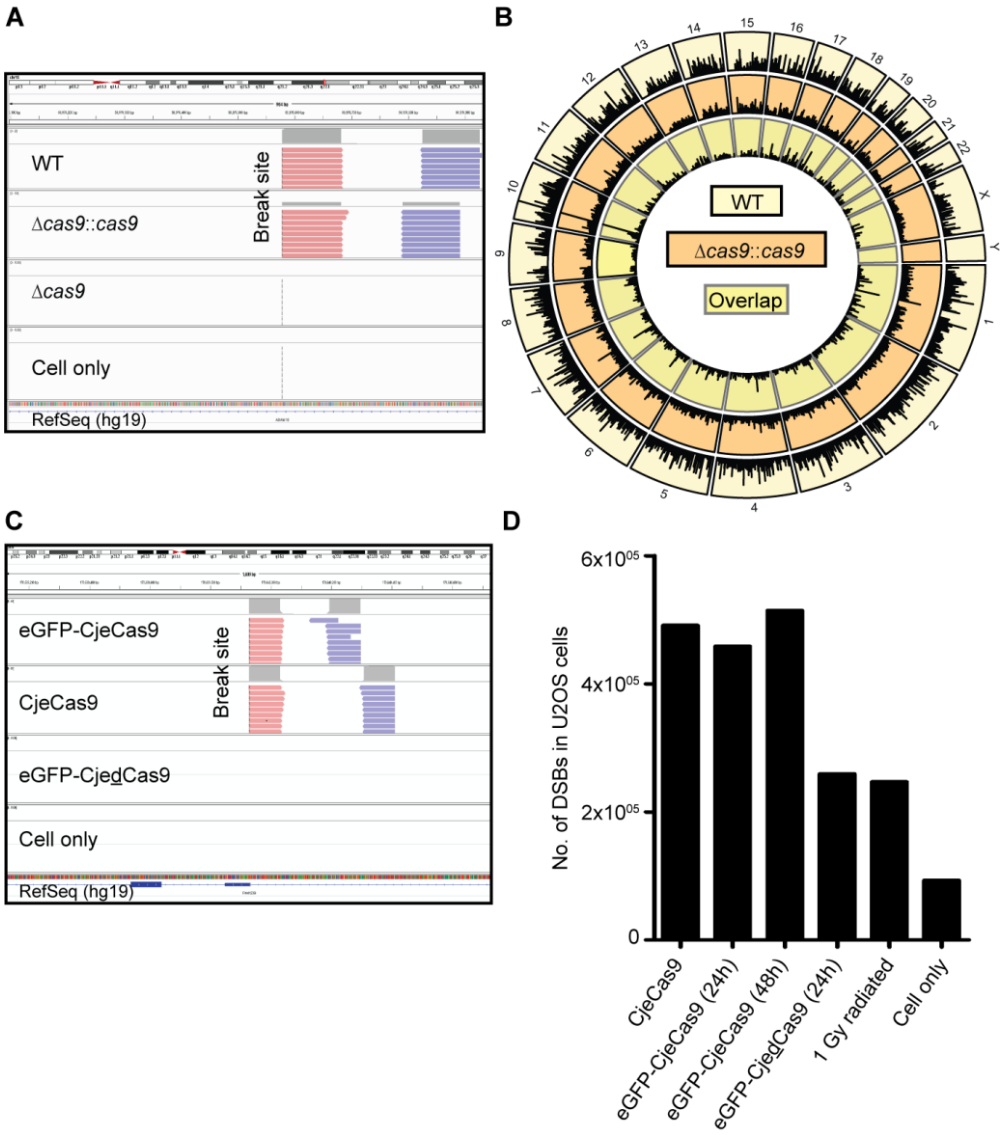


Figure S9.6 | BLESS analysis snapshot to detect CjeCas9-associated DSBs across the human genome. (A) An example of BLESS reads mapped (red, proximal linker used to detect the DSB break position and blue, distal linker) onto the reference genome (GRCh37/hg19) and was detected in human cells exposed to CjeCas9 Circus plot showing the genome wide distribution of CjeCas9-associated DSBs (related to **Table S9.1**). BLESS DNA sample of WT (yellow), Δ*cas9*::*cas9* (orange) and breaks that had occurred at the exact same genomic position (bright yellow) during infection of U2OS cells are shown. **(B)** An example of BLESS reads mapped (red, proximal linker used to detect the DSB break position and blue, distal linker) onto the reference genome (GRCh37/hg19) and was detected in human cells expressing GFP-CjeCas9 or CjeCas9 from an eukaryotic expression vector. **(C)** Bar graphs showing the number of DSBs detected in U2OS cells by a BLESS snapshot. Active GFP-CjeCas9 or CjeCas9 or inactive GFP-CjedCas9 were compared. 1 Gy radiated is used as a positive control, untreated U2OS cells (Cell only) as negative control.

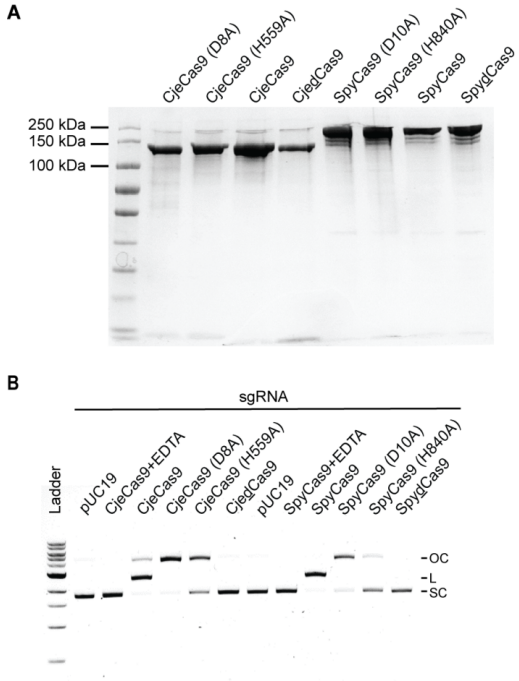


Figure S9.7 | SDS-PAGE of Cas9 purified proteins. (A) Black and white picture of a Coomassie blue stained SDS-PAGE gel in which the purified CjeCas9, SpyCas9 and their variants are visualized. **(B)** Agarose gel showing the cleavage of pUC19 target plasmid DNA by CjeCas9, SpyCas9 and its variants in the presence of a sgRNA, to verify the purified proteins. OC refers to the open circular, L to linear and SC to supercoiled confirmations of pUC19.

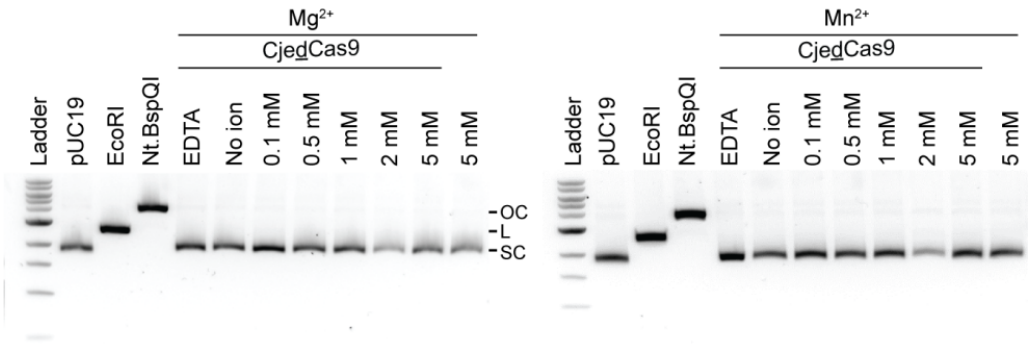


Figure S9.8 | Absence of DNA cleavage activity by CjedCas9 in the presence of Mn^{2+} or Mg^{2+} cations. Agarose gel showing the absence of plasmid DNA cleavage by CjedCas9 incubated with various concentrations of Mn^{2+} or Mg^{2+}. dCas9 refers to a catalytically inactive variant of CjeCas9. OC refers to the open circular, L to linear and SC to supercoiled confirmations of pUC19. EcoRI refers to pUC19 with a single double strand DNA break; Nt. BspQI refers to pUC19 with nicks. Relates to **Figure 9.4A**.

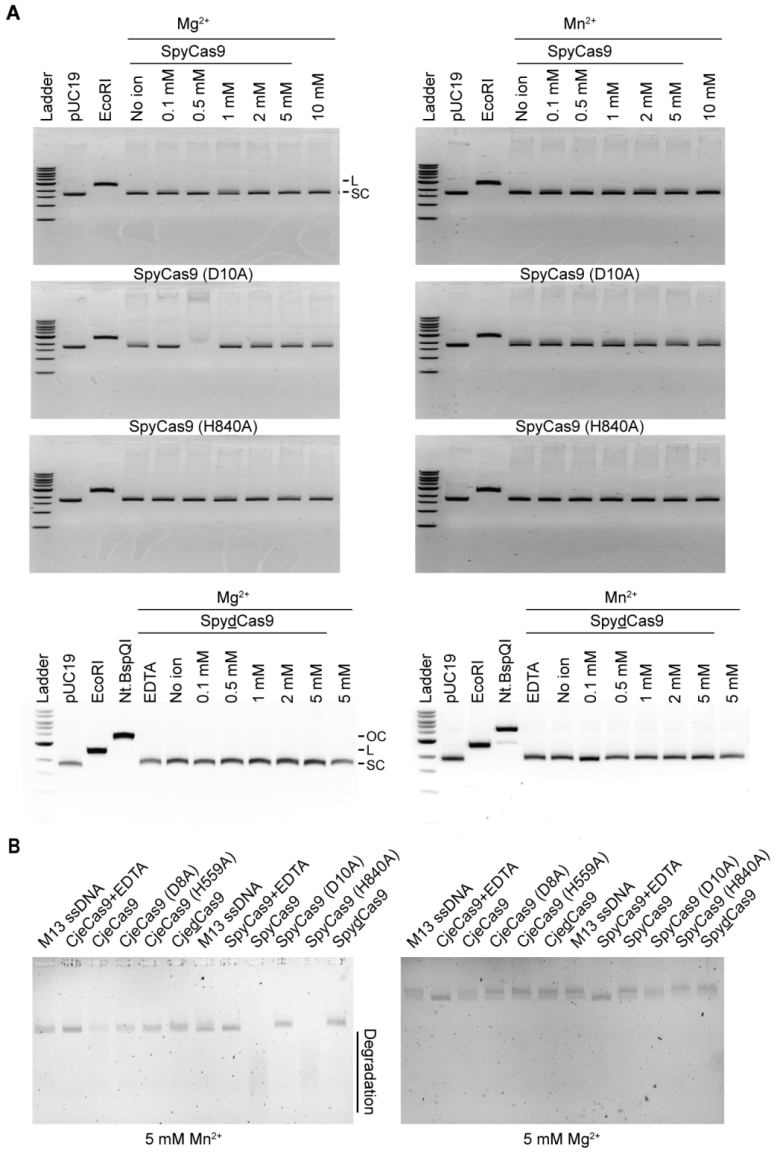


Figure S9.9 | The effect of Mg^{2+} and Mn^{2+} concentrations on plasmid DNA cleavage activity by SpyCas9. (A) Agarose gels showing the processing of plasmid DNA by SpyCas9 and its variants, incubated with various concentrations of Mn^{2+} or Mg^{2+}. D10A refers to a mutated SpyCas9 variant with inactive RuvC domain, H840A refers to a mutated SpyCas9 variant with inactive HNH domain and dCas9 refers to a mutated SpyCas9 variant with inactive RuvC and HNH domains. OC refers to the open circular, L to linear and SC to supercoiled confirmations of pUC19. EcoRI refers to pUC19 with a single double strand DNA break; Nt. BspQI refers to pUC19 with nicks. **(B)** Agarose gels showing the processing of circular ssDNA by SpyCas9, CjeCas9 variants at concentrations 0f 5 mM for both Mn^{2+} or Mg^{2+}.

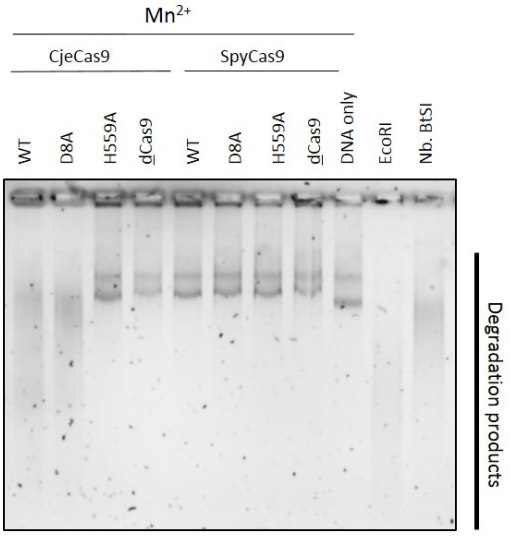


Figure S9.10 | CjeCas9 genomic DNA processing occurs mainly by the HNH domain. Denaturing urea-agarose gel analysis of human genomic DNA cleavage by CjeCas9 and SpyCas9 variants, incubated with 5 mM Mn^{2+}. D8A refers to a mutated CjeCas9 variant with inactive RuvC domain; H559A refers to a mutated CjeCas9 variant with inactive HNH domain. dCas9 refers to an inactive variant of CjeCas9 in which both the RuvC and HNH domain have been inactivated. EcoRI refers to genomic DNA with double strand DNA breaks; Nt. BspQI refers to genomic DNA with nicks.

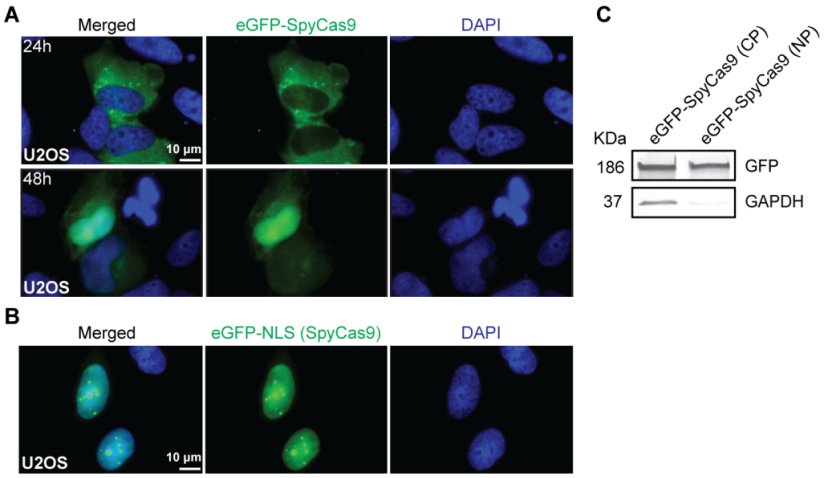


Figure S9.11 | SpyCas9 accumulates into the eukaryotic nucleus over time. (A) Microscopic image of eGFP-SpyCas9 expression in human cells (green). Time points shown, 24 and 48 hours after transfection with a eukaryotic expression vector that enables expression of eGFP-SpyCas9 **(B)** Microscopic image of eGFP localization (green) in the nucleus of human cells represented by eGFP fused to the NLS motif identified in SpyCas9. NLS: nuclear localization signal. (A and B) Nuclei were stained with DAPI (blue). **(C)** Western blot analysis of eGFP-SpyCas9 in nuclear (NP) and cytoplasmic (CP) protein fractions of human cells. GAPDH is used as a quality control for the separation of NP and CP.

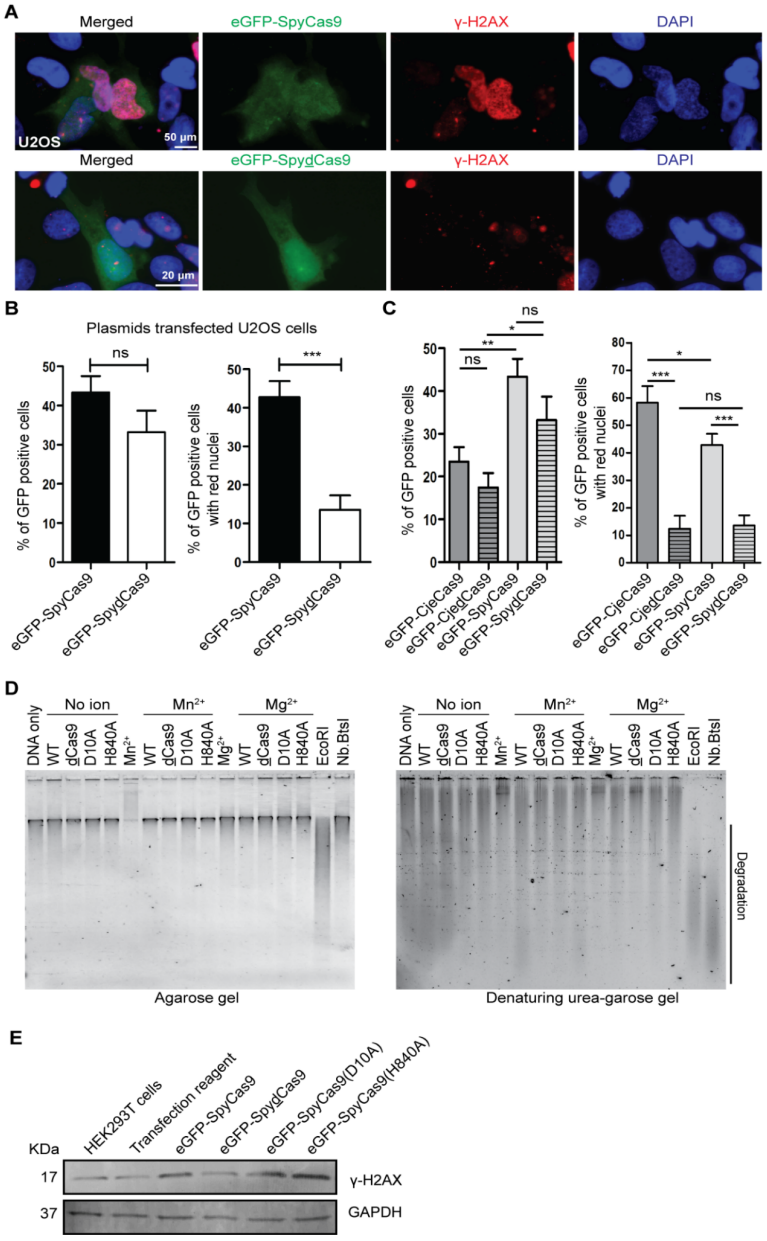


Figure S9.12 | SpyCas9 enters the nucleus and processes genomic DNA with its RuvC domain. (A) γ-H2AX accumulation (red) in the nuclei of transfected human cells expressing eGFP-SpyCas9 or eGFP-SpydCas9. Nuclei were stained with DAPI (blue). **(B)** Percentages of cells positive for eGFP only or eGFP positive cells with high numbers of γ-H2AX foci (detected by antibody, scored as bright red nuclei) detected in transfected human cells that efficiently expressed eGFP-SpyCas9 or eGFP-SpydCas9. Results shown as the mean ± s.e.m. *** $p < 0.001$, ns $p > 0.05$ by two-tailed Student's t-test. **(C)** Comparison between the transfection efficiencies of, and γ-H2AX activation by, eukaryotic Cas9 expression plasmids enabling the expression of eGFP-CjeCas9, eGFP-CjedCas9, eGFP-SpyCas9 and eGFP-SpydCas9. Transfection efficiencies are obtained by counting eGFP positive cells. Differences in chromatin alteration was obtained by counting eGFP and γ-H2AX (bright red nuclei positive cells. Y-axis shows the percentage of eGFP positive cells only or eGFP positive cells with high numbers of γ-H2AX foci (scored as bright red nuclei, see **(A)**). Results shown as the mean ± s.e.m. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns $p > 0.05$ by One-Way ANOVA (Bonferroni's Multiple Comparison test).

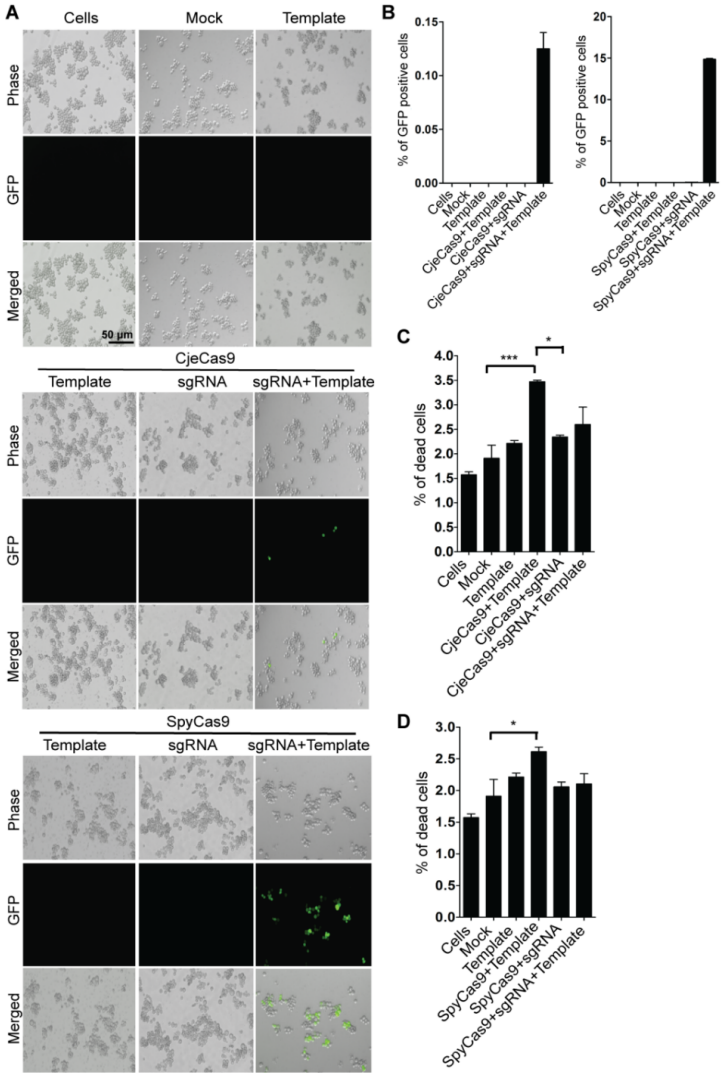


Figure S9.13|Restoration of the GFP expression in K562 (GFPmut) cells lipofected with purified CjeCas9 (upper panel) or SpyCas9 (lower panel) in presence/absence of sgRNAs. (A) Microscopic images of GFP edited cells. **(B)** Bar-graph showing the percentage of cells that produce GFP after transfection with purified CjeCas9 (left bar graph) or SpyCas9 (right bar graph) in presence/absence of sgRNAs. **(C & d)** FACS analysis of in vivo induction of dead cells (propidium iodide (PI) positive) during genome editing with CjeCas9 or SpyCas9. Mock is transfection agent. Results shown as the mean ± s.e.m, *** $p < 0.001$ and * $p < 0.05$ by One-Way ANOVA (Bonferroni's Multiple Comparison test).

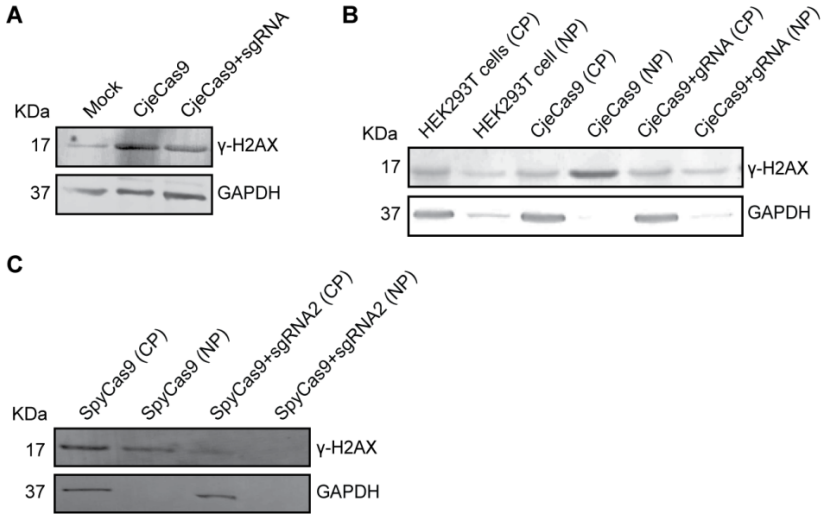


Figure S9.14 | Saturation of Cas9 with a guide RNA limits γ-H2AX accumulation *in vivo*. **(A)** Western blot analysis showing γ-H2AX abundance in human cells lipofected with CjeCas9 only or as RNP particles saturated with gRNAs. Mock is transfection agent only. (**(B)** and **(C)**) Western blot analysis of γ-H2AX abundance in nuclear (NP) and cytoplasmic (CP) protein fractions of human cells transfected with a plasmid expressing just (**(B)**) CjeCas9 or (**(C)**) SpyCas9 only or (**(B)**) CjeCas9 or (**(C)**) SpyCas9 with a guide RNA that lacks a target in the human genome.

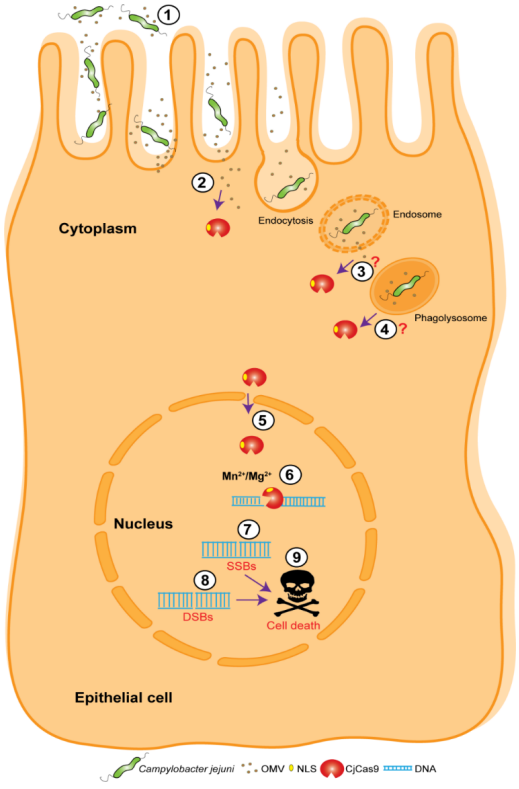


Figure S9.15 | Proposed model of CjeCas9 mediated DSBs induction and subsequent host cell death. During the *C. jejuni* infection process of intestinal epithelial cells, bacteria release CjeCas9 containing OMVs **(1)**, which are taken up by host cells **(2)** (*438*). During invasion of intestinal epithelial cells, *C. jejuni* bacteria generate *Campylobacter* containing vesicles (CCV) (*453*) that co-localize with Lamp-1 (*447*) in which *C. jejuni* can survive and cross the intestinal epithelial barrier. During endosomal maturation of early **(3)** and late **(4)** CCVs, *C. jejuni* may secrete CjeCas9. Upon release from the CCV into the cytoplasm, CjeCas9 will eventually translocate into the nucleus autonomously via the discovered NLS **(5)**. Our data support a model where nuclear CjeCas9 may associate with metal ions Mg^{2+} and/or Mn^{2+} that activate this nuclease **(6)**. From our work it follows that CjeCas9 induces SSB **(7)** and DSBs **(8)** forcing the human cells to activate the DNA damage, stress and cell death pathways, resulting in the death of the infected cells **(9)**. This implies that CjeCas9 is cytotoxic for human cells.

Table S9.1 | Total number of CjeCas9-dependent DSBs obtained by BLESS

Sample	Total break sites	Filter (>3 reads)	Not in reference	Cas9-dependent DSB	Overlap
WT	1,633,494	690,299	682,592	678,910*	
Δ*cas9*::*cas9*	2,150,945	956,998	947,022	942,104*	25,539***
CCas9 (24h)	795,106	236,446	232,812	231,912	
eGFP-CjeCas9 (24h)	722,654	203,758	200,373	199,326**	
eGFP-CjeCas9 (48h)	824,219	260,425	256,413	255,186**	

*CjeCas9-dependent DSBs that were not identified in cell-only or in Δ*cas9* infected human cells.
**CjeCas9-dependent DSBs that were not identified in cell-only or in eGFP-CjedCas9 transfected human cells.
***Shared CjeCas9-dependent DSBs present in WT and Δ*cas9*::*cas9* infected human cells; these DSBs overlap at the exact same nucleotide position in the human genome (reference genome GRCh37/hg19).

Table S9.2 | Total number of CjeCas9-independent DSBs obtained by BLESS

Sample	Total break sites	Filter (>3 reads)	Only in Chr1:22X
Δ*cas9*	727,672	191,271	190,111
eGFP-CjedCas9 (24h)	654,496	166,571	165,817
1 Gy irradiated	554,956	157,006	153,590
Untreated U2OS (cell only)	340,762	94,039	93,373

Table S9.3 | Nucleotide frequency counting upstream and downstream of the DSB position obtained by BLESS after infection with *C. jejuni* bacteria

	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11
A	74746	76874	75465	74137	75357	72481	78190	78856	78556	66661	49568	67227	62020	59822	76599	74603	81620	77448	77026	74775	76387	74686
T	73883	72845	72003	74788	73681	76958	66733	64752	66245	81311	67750	53813	66239	77283	92782	83785	74267	76839	78921	79677	76954	77215
C	51506	47851	51637	51378	47673	48007	44985	53244	50538	44826	47984	45400	90706	83935	47441	51229	48213	49895	48116	48818	48344	48955
G	50680	53245	51710	50512	54104	53369	60907	53963	55978	58017	85513	84376	31851	29776	33994	41169	46716	46634	46753	47546	49131	49960

	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11
A	75222	76664	76041	74343	75374	72840	78481	79927	78815	66625	47259	68162	62123	59124	75943	74341	81868	77583	77789	74509	76014	74944
T	74010	72909	71578	74687	73830	77165	66279	63336	65832	81779	68380	52602	66032	77854	93900	84233	73750	76560	78573	78885	76824	77065
C	51507	47686	51773	51739	47891	47922	44340	52971	50372	45143	49678	44755	91317	85001	47348	50917	48570	50389	47750	49060	48301	48562
G	50072	53552	51419	50042	53716	52884	61712	54579	55794	57266	85496	85295	31344	28837	33625	41325	46628	46284	46704	48362	49677	50245

	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11
A	71468	71856	71072	70448	71288	69595	67761	67972	68264	66196	62447	64829	66535	69097	76440	74434	75893	75687	75586	74274	74403	74005
T	71265	70287	69953	70857	71556	71559	69197	67255	67952	70767	68835	62327	68145	70951	77773	76715	75489	75488	76384	76457	75441	75632
C	53645	52800	54937	55321	52942	51743	51376	55781	57092	53390	49289	58341	74764	68206	51758	51634	50179	50475	49635	50730	50764	51246
G	54431	55866	54847	54184	55025	57914	62478	59805	57502	60461	70243	65318	41372	42562	44845	48033	49225	49166	49211	49355	50208	49933

Table S9.4 | Nucleotide Frequency counting upstream and downstream of the DSB position obtained by BLESS after plasmid-driven expression of CjeCas9

	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11
A	72062	72850	71478	71001	72238	69900	68890	69344	68717	65659	59895	64602	65306	66779	76564	74310	76285	75381	75000	74434	74780	74154
T	71404	70753	70660	71293	71410	72249	67979	66789	67425	72988	68722	61707	67992	71932	80313	77618	75346	75815	76631	76370	75592	76090
C	53987	51863	54668	54804	52347	51428	50178	55598	56818	52639	50407	55283	77383	71464	50923	51859	50207	50505	49981	50350	50604	50564
G	53358	55345	53985	53713	54816	57234	62766	59082	57854	59528	71790	69224	40135	40641	43016	47029	48978	49115	49204	49662	49890	50008

	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11
A	70251	71375	68935	68592	69195	66258	65658	66156	66503	60654	48979	57874	58479	62746	75866	72381	75427	73580	73471	71890	72386	71942
T	69749	68368	68303	69690	69919	69870	63996	62077	62897	68690	58212	49680	57872	66703	80038	75782	72987	74521	75331	75949	74707	74657
C	55738	53786	57860	57438	53749	52322	52080	59817	59374	51646	51661	63529	98371	84088	52084	53700	51308	51680	50498	51292	51046	52313
G	55078	57287	55718	55146	57953	62366	69082	62766	62042	69826	91964	79733	36094	37199	42828	48953	51094	51035	51516	51685	52677	51904

	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11
A	72626	74114	72430	71755	72417	70160	69147	69066	68463	66840	58366	65066	65646	66765	77226	75121	76517	75818	76187	74805	75243	75167
T	72842	71646	71165	72229	72703	73065	68313	66290	67676	72814	67812	59806	67339	70668	82080	79518	77126	77516	77956	78033	76690	76954
C	52800	50257	53846	53850	51344	50185	49296	55714	58018	52657	49051	53537	79599	74021	49728	50440	48888	48897	48315	49156	49057	49332
G	52538	54789	53365	52977	54343	57397	64053	59741	56656	58502	75584	72406	38232	39362	41782	45737	48285	48585	48358	48822	49826	49363

	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11
A	72307	73058	72059	71605	72259	69999	69838	69963	69332	65311	61366	64705	65558	67667	75747	74606	77012	76381	76351	74386	74806	74426
T	72069	71222	70920	72111	71795	72928	69232	67665	68402	72651	72176	64409	70720	74094	80528	78626	76684	76822	77911	77777	76481	76728
C	52965	51342	53540	53594	51893	51136	49897	54667	56036	54232	50870	57541	73960	67662	51096	51212	48929	49213	48848	49828	49976	50332
G	53467	55186	54298	53498	54861	56745	61842	58515	57040	58618	66401	64159	40578	41390	43445	46372	48191	48400	48206	48825	49553	49340

	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11
A	71180	71477	70592	69541	70714	69731	68212	68774	69201	67315	64101	65568	66610	69180	75411	74160	75392	75452	75386	74574	74150	73873
T	70638	69794	69805	71180	70883	71372	69072	68022	68125	69771	68712	62157	67644	70385	75995	75085	74394	74561	74743	75298	74577	74847
C	54475	53267	55231	55308	53727	52265	52420	55434	55603	53412	49224	58764	72886	66661	52799	52526	50967	50569	50706	50984	51137	51508
G	54516	56271	55181	54780	55485	57441	61106	59081	57882	60315	68776	64326	43676	44590	46611	49045	50063	50234	49981	50010	50946	50588

Table S9.5 | IPA canonical pathways significantly induced (blue bars) or repressed (red bars) in human cells exposed to CjeCas9 activity. Pathways in boldface (z-scores >2 or < -2) are significant after correction for participation of proteins in multiple pathways; only pathways with z-scores >1 or <1 are shown.

Ingenuity Canonical Pathways	-log(p-value)	z-score
Phospholipase C Signaling	2.75	**3**
Regulation of Actin-based Motility by Rho	6.74	**2.84**
Ephrin Receptor Signaling	5.08	**2.683**
CXCR4 Signaling	1.36	**2.53**
ILK Signaling	3.85	**2.357**
Rac Signaling	4.55	**2.324**
Cardiac Hypertrophy Signaling	1.83	**2.324**
fMLP Signaling in Neutrophils	4.29	**2.138**
Ephrin B Signaling	5.6	**2.121**
Macropinocytosis Signaling	1.51	**2**
Oxidative Phosphorylation	20.7	1.915
Actin Nucleation by ARP-WASP Complex	5.26	1.897
CD28 Signaling in T Helper Cells	1.6	1.897
PAK Signaling	1.89	1.89
Calcium Signaling	1.66	1.89
Huntington's Disease Signaling	10.2	1.732
Cdc42 Signaling	1.66	1.732
Protein Kinase A Signaling	2.36	1.706
Fcγ Receptor-mediated Phagocytosis in Macropha	7.31	1.698
Pyrimidine Deoxyribonucleotides De Novo Biosynt	7.19	1.667
NGF Signaling	1.44	1.667
Renal Cell Carcinoma Signaling	2.44	1.633
Antiproliferative Role of Somatostatin Receptor 2	1.65	1.633
Integrin Signaling	6.74	1.569
Pyrimidine Ribonucleotides Interconversion	6.26	1.508
Glycolysis I	5.53	1.414
Gluconeogenesis I	5.53	1.414
p70S6K Signaling	2.9	1.414
Telomerase Signaling	2.56	1.414
NRF2-mediated Oxidative Stress Response	7.25	1.387
EIF2 Signaling	60.3	1.372
14-3-3-mediated Signaling	7.81	1.155
Pyrimidine Ribonucleotides De Novo Biosynthesis	7.01	1.155
VEGF Signaling	5.23	1.134
PI3K/AKT Signaling	3.65	1.069
Aldosterone Signaling in Epithelial Cells	7.17	1
Pentose Phosphate Pathway	3.3	1
Apoptosis Signaling	2.21	1
Purine Nucleotides De Novo Biosynthesis II	7.56	-1.134
ATM Signaling	2.99	-1.414
tRNA Charging	6.96	**-2.714**

Table S9.6 | IPA canonical pathways significantly induced (blue bars) or repressed (red bars) in human cells exposed to SpyCas9 activity. Pathways in boldface (z-scores >2 or < -2) are significant after correction for participation of proteins in multiple pathways; only pathways with z-scores >1 or <1 are shown.

Ingenuity Canonical Pathways	-log(p-value)	z-score
Oxidative Phosphorylation	12.9	**5**
EIF2 Signaling	73.6	**4.529**
Ephrin Receptor Signaling	4.61	**3.441**
NRF2-mediated Oxidative Stress Response	6.15	**3.317**
Fcγ Receptor-mediated Phagocytosis in Macrophages and Monocytes	5.14	**3.207**
fMLP Signaling in Neutrophils	3.23	**3.162**
Pyrimidine Ribonucleotides Interconversion	4.51	**3**
Pyrimidine Ribonucleotides De Novo Biosynthesis	4.35	**3**
Remodeling of Epithelial Adherens Junctions	16.9	**2.828**
Actin Cytoskeleton Signaling	6.53	**2.746**
Pyrimidine Deoxyribonucleotides De Novo Biosynthesis I	4.89	**2.646**
Paxillin Signaling	2.06	**2.53**
Rac Signaling	1.96	**2.53**
CXCR4 Signaling	1.73	**2.333**
Signaling by Rho Family GTPases	3.79	**2.324**
Integrin Signaling	5.73	**2.294**
Antiproliferative Role of Somatostatin Receptor 2	2.18	**2.236**
Renal Cell Carcinoma Signaling	1.95	**2.236**
Gαs Signaling	1.31	**2.236**
RhoA Signaling	5.54	**2.183**
PI3K/AKT Signaling	3.7	**2.138**
Phospholipase C Signaling	2.43	**2.138**
Glutaryl-CoA Degradation	7.5	**2.121**
Tryptophan Degradation III (Eukaryotic)	5.71	**2.121**
Salvage Pathways of Pyrimidine Ribonucleotides	1.58	**2.121**
tRNA Charging	7	**2.111**
Aspartate Degradation II	4.24	**2**
Granzyme B Signaling	2.64	**2**
α-Adrenergic Signaling	2.34	**2**
Regulation of Actin-based Motility by Rho	5.31	1.941
Leukocyte Extravasation Signaling	1.32	1.732
Isoleucine Degradation I	5.25	1.633
Glutathione Redox Reactions I	3.74	1.633
TCA Cycle II (Eukaryotic)	3.74	1.633
Glutathione-mediated Detoxification	3.1	1.633
Fatty Acid β-oxidation I	3.02	1.633
Actin Nucleation by ARP-WASP Complex	1.79	1.633
ERK5 Signaling	1.47	1.633
Glycolysis I	5.56	1.414
Gluconeogenesis I	5.56	1.414
Mevalonate Pathway I	4.18	1.342
Superpathway of Geranylgeranyldiphosphate Biosynthesis I (via Mevalonate)	3.55	1.342
Superpathway of Cholesterol Biosynthesis	2.5	1.342
14-3-3-mediated Signaling	9.39	1.155
Purine Nucleotides De Novo Biosynthesis II	7.6	1.134
Telomerase Signaling	1.68	1.134
Aldosterone Signaling in Epithelial Cells	6.01	1
Ketogenesis	3.5	1
Ketolysis	3.5	1
Superpathway of Methionine Degradation	1.34	1
Pentose Phosphate Pathway	3.32	-1
Cardiac β-adrenergic Signaling	2.24	-1.134
eNOS Signaling	1.6	-1.667
Cyclins and Cell Cycle Regulation	2.02	**-2**
HIPPO signaling	4.11	**-2.121**

Table S9.7 | IPA canonical pathways significantly induced (blue bars) or repressed (red bars) in human cells exposed to CjeCas9 activity saturated with sgRNAs. Pathways in boldface (z-scores >2 or < -2) are significant after correction for participation of proteins in multiple pathways; only pathways with z-scores >1 or <1 are shown.

Ingenuity Canonical Pathways	-log(p-value)	z-score
Glutaryl-CoA Degradation	3.59	**2.236**
Tryptophan Degradation III (Eukaryotic)	2.63	**2.236**
Pentose Phosphate Pathway	3.23	**2**
TCA Cycle II (Eukaryotic)	4.62	1.89
eNOS Signaling	1.46	1.667
Glutathione-mediated Detoxification	2.98	1.633
Actin Cytoskeleton Signaling	6.43	1.347
Mevalonate Pathway I	4.08	1.342
Isoleucine Degradation I	3.90	1.342
Superpathway of Geranylgeranyldiphosphate Biosynthesis I (via Mevalonate)	3.45	1.342
Glutathione Redox Reactions I	2.71	1.342
Superpathway of Cholesterol Biosynthesis	2.40	1.342
Oxidative Phosphorylation	16.10	1.3
Integrin Signaling	4.86	1.279
RhoA Signaling	5.25	1.213
Actin Nucleation by ARP-WASP Complex	3.92	1
Androgen Signaling	3.07	1
Fatty Acid Î²-oxidation I	1.47	1
Glycolysis I	6.50	-1
Aldosterone Signaling in Epithelial Cells	5.10	-1
p70S6K Signaling	3.54	-1
Apoptosis Signaling	1.93	-1
Cell Cycle: G1/S Checkpoint Regulation	1.84	-1
IL-2 Signaling	1.74	-1
Non-Small Cell Lung Cancer Signaling	1.39	-1
Pyrimidine Deoxyribonucleotides De Novo Biosynthesis I	4.75	-1.134
Pyrimidine Ribonucleotides De Novo Biosynthesis	6.82	-1.155
Sirtuin Signaling Pathway	22.10	-1.219
Synaptic Long Term Potentiation	1.62	-1.265
PI3K/AKT Signaling	3.79	-1.291
CCR3 Signaling in Eosinophils	3.10	-1.342
Macropinocytosis Signaling	1.72	-1.342
Renal Cell Carcinoma Signaling	1.69	-1.342
Neuregulin Signaling	1.54	-1.342
Telomerase Signaling	2.25	-1.414
Pyrimidine Ribonucleotides Interconversion	6.10	-1.508
Paxillin Signaling	2.20	-1.508
Regulation of eIF4 and p70S6K Signaling	36.10	-1.633
Cardiac Î²-adrenergic Signaling	1.68	-1.89
PAK Signaling	1.65	-1.89
RhoGDI Signaling	5.30	**-2.183**
ERK/MAPK Signaling	2.61	**-2.309**

Table S9.8 | IPA canonical pathways significantly induced (blue bars) or repressed (red bars) in human cells exposed to SpyCas9 activity saturated with sgRNAs. Pathways in boldface (z-scores >2 or < -2) are significant after correction for participation of proteins in multiple pathways; only pathways with z-scores >1 or <1 are shown.

Ingenuity Canonical Pathways	-log(p-value)	z-score
Paxillin Signaling	2.66	**2.309**
Fatty Acid β-oxidation I	2.14	**2.236**
β-Adrenergic Signaling	2.01	**2.236**
Aspartate Degradation II	4.15	**2**
Pyrimidine Ribonucleotides Interconversion	5.19	1.897
Sirtuin Signaling Pathway	19	1.826
tRNA Charging	7.82	1.732
ERK/MAPK Signaling	2.61	1.732
Gluconeogenesis I	6.5	1.667
Telomerase Signaling	2.72	1.667
Granzyme B Signaling	4.73	1.633
CREB Signaling in Neurons	1.36	1.633
Pyrimidine Ribonucleotides De Novo Biosynthesis	5.89	1.508
RhoGDI Signaling	4.76	1.5
Isoleucine Degradation I	3.9	1.342
CCR3 Signaling in Eosinophils	3.1	1.342
Glioma Signaling	1.75	1.342
Macropinocytosis Signaling	1.72	1.342
Renal Cell Carcinoma Signaling	1.69	1.342
Melanoma Signaling	1.52	1.342
Sumoylation Pathway	8.41	1.291
Actin Cytoskeleton Signaling	5.89	1.177
NRF2-mediated Oxidative Stress Response	6.09	1.155
Rac Signaling	2.54	1.155
Cardiac Î²-adrenergic Signaling	1.68	1.134
Chemokine Signaling	1.54	1.134
Glycolysis I	6.5	1
RhoA Signaling	4.63	1
ATM Signaling	3.41	1
Cell Cycle: G1/S Checkpoint Regulation	2.38	1
Role of p14/p19ARF in Tumor Suppression	2.23	1
CDK5 Signaling	1.69	1
Non-Small Cell Lung Cancer Signaling	1.38	1
Glutathione Redox Reactions I	2.71	-1.342
Glutathione-mediated Detoxification	2.98	-1.633
eNOS Signaling	1.45	-1.667
Huntington's Disease Signaling	7.4	-1.897

Table S9.9 | Nuclear proteins involved in nucleotide excision repair (NER) or non-homologous end-joining (NHEJ) DNA repair pathways. The upper panel shows (log ratio) differential abundance of proteins for the comparison (Cas9 minus dCas9), the lower panel (log ratio) differential abundance of proteins for the comparison (Cas9+sgRNAs minus Cas9).

DNA DSB repair NHEJ		CjCas9-CjdeadCas9	SpyCas9-SpydeadCas9
DNA-PK	protein kinase, DNA-activated, catalytic polypeptide	-1.14	
Ku70	X-ray repair cross complementing 6		3.82
Ku80	X-ray repair cross complementing 5	-6.13	15.97
PARP1	poly(ADP-ribose) polymerase 1	-16.32	14.22
NER pathway			
POLR2D	RNA polymerase II subunit D	2.81	3.95
POLR2G	RNA polymerase II subunit G	-9.34	-2.19
POLR2H	RNA polymerase II subunit H		7.37
POLR2L	RNA polymerase II subunit L	2.84	
RPA1	replication protein A1	-1.58	-1.82
RPA2	replication protein A2	0.79	2.88
RPA3	replication protein A3	-1.49	6.26

DNA DSB repair NHEJ		CjCas9guideRNAs-CjCas9	SpyCas9guideRNAs-SpyCas9
DNA-PK	protein kinase, DNA-activated, catalytic polypeptide	-0.22	1.23
Ku70	X-ray repair cross complementing 6	-6.63	-0.79
Ku80	X-ray repair cross complementing 5	-14.98	20.78
PARP1	poly(ADP-ribose) polymerase 1	7.96	6.08
NER pathway			
POLR2D	RNA polymerase II subunit D	4.07	-2.11
POLR2G	RNA polymerase II subunit G	2.92	-5.03
POLR2H	RNA polymerase II subunit H	1.74	2.16
POLR2L	RNA polymerase II subunit L	-1.89	-6.17
RPA1	replication protein A1	1.86	-1.38
RPA2	replication protein A2	0.48	0.72
RPA3	replication protein A3	6.31	1.85

Table S9.10 | Bacterial strains used in the study.

Bacteria	Gene modified	Antibiotic marker	References
C. jejuni GB11	Wild type	Vancomycin	(*447*)
C. jejuni GB11	Δ*cas9*	Chloramphenicol	(*120*)
C. jejuni GB11	Δ*cas9*::*cas9*	Erythromycin	(*120*)
C. jejuni GB11Cas9-mCherry	Cas9-mCherry	Kanamycin	This study
C. jejuni (*cdt* -)		Vancomycin	This study

Table S9.11 | Primer list with related restriction sites and guide RNA used in this study.

Primer name	Direction	Restriction site	Primer sequence (5'-3')
CjeCas9_pEGFP-C1	Forward	SalI	TGCAGTCGACATGGCAAGAATTTTGGCATTTGATATA
CjeCas9_pEGFP-C1	Reverse	BamHI	CGGTGGATCCTTTTTTAAAATCTTCTCTTTGTCTAAA
pCjeCas9(GE)	Forward	BamHI	ATCGGGATCCATGGCAAGAATTTTGGCATTTGATATAGGAA
pCjeCas9(GE)	Reverse	XhoI	ATCGCTCGAGTCATTTTTTAAAATCTTCTCTTTGTCTAAA
NLS(Cje)	Forward	SalI	TGCAGTCGACGGAGAATCCTTAGCCTTGCCA
NLS(Cje)	Reverse	Bam HI	CGGTGGATCCTAAATGTTTTAAGTGATTTAG
Cje_D8A	Forward		TTTTGGCATTTGCTATAGGAATAAGCTCTATTGGC
Cje_D8A	Reverse		CTTATTCCTATAGCAAATGCCAAAATTCTTGCCAT
Cje_H559A	Forward		CTTGAAATTGATGCCATTTACCCTTATTCTAGAAGTT
Cje_H559A	Reverse		TAAGGGTAAATGGCATCAATTTCAAGCATTTTTTCATCT
SpyCas9_pEGFP-C1	Forward	SalI	TGCAGTCGACATGGATAAGAAATACTCAATAGGCTTA
SpyCas9_pEGFP-C1	Reverse	SacII	GGGCCCGCGGGTCACCTCCTAGCTGACT
SpydCas9_pEGFP-C1	Forward	SalI	TGCAGTCGACATGGATAAGAAATACTCAATAGGCTTA
SpydCas9_pEGFP-C1	Reverse	SacII	GGGCCCGCGGGTCACCTCCTAGCTGACT
NLS (Spy)	Forward	SalI	TGCAGTCGACGCGGAAGCGACTCGTCTCAAA
NLS (Spy)	Reverse	BamHI	CGGTGGATCCCTCCTGTAGATAACAAATACG
cas9-up (CSO 0374)	Forward		GTTTTTCTCGAGCCGCAAAGTAAATGGTAAAAT
cas9-up (CSO 0375)	Reverse		CCTCCTCGCCCTTGCTCACTTTTTTAAAATCTTCTCTTTGTCTAAAC
mCherry (CSO 2150)	Forward		GTGAGCAAGGGCGAGGA
mCherry (CSO 2151)	Reverse		TTACTTGTACAGCTCGTCCAT
Kan+*cas9*-dn (CSO2149)	Forward		GCATGGACGAGCTGTACAAGTAAGTACCCGGGTGACTAACTAG
Kan+*cas9*-dn (0377)	Reverse		GTTTTTTCTAGAAAAGCTCATCTCTTGAAAAAGAA
sgRNA-AAVS1-T2	Guide RNA	Synthetic	GGGGCCACUAGGGACAGGAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC
sgRNA-AAVS1_Cj	Guide RNA	Synthetic	AUCUUCUUCAAGGACGACGGCAGUUUUAGUCCCUGAAGGGACUAAAAUAAAGAGUUUGCGGGACUCUGCGGGGUUACAAUCCCCUAAAACCGCUUUUUU
Proximal linker		Synthetic	dTdAdCdTdAdCdCdTdCdGdAdGdAdGdTdTdAdCdGdCdTdAdGdGdGdAdTdAdAdCdAdGdGdGdTdAdAdTdAdTdAdGdTdTdTBiotindTdTdTdTdCdTdAdTdAdTdTdAdCdCdCdTdGdTdTdAdTdCdCdCdTdAdGdCdGdTdAdAdCdTdCdTdCdGdAdGdGdTdAdGdTdA
Distal linker		Synthetic	CGTCGTCTCGAGAGTTACGCTAGGGATAACAGGGTAATATAGTTTTTTTCTATATTACCCTGTTATCCCTAGCGTAACTCTCGAGACGACG
Primer proximal linker	Forward		CCCTAGCGTAACTCTCGAGGTAGTA
Primer distal linker	Reverse		CTAGCGTAACTCTCGAGACGACG
To create SpyCas9 variant expression plasmid	Forward		GCTCGATATCGGCACAAACAGC

Table S9.11 | Primer list with related restriction sites and guide RNA used in this study. (continued)

Primer name	Direction	Restriction site	Primer sequence (5'-3')
To create SpyCas9 variant expression plasmid	Reverse		GCTGTTTGTGCCGATATCGAGC
To create SpyCas9 variant expression plasmid	Forward		CCATTGGGCTCGCGATCGGCACAAACAGCG
To create SpyCas9 variant expression plasmid	Reverse		GCTGTTTGTGCCGATCGCGAGCCCAATGGAGTACTTC
To create SpyCas9 variant expression plasmid	Forward		CGACTACGACGTGGATCATATCG
To create SpyCas9 variant expression plasmid	Reverse		CGATATGATCCACGTCGTAGTC
To create SpyCas9 variant expression plasmid	Forward		CGACTACGACGTGGATGCGATCGTGCCCCAGTCTTTTC
To create SpyCas9 variant expression plasmid	Reverse		GGCACGATCGCATCCACGTCGTAGTCGG
To create SpyCas9 variant expression plasmid	Forward		CGTTGAATATGGCTCATAACACCC
To create SpyCas9 variant expression plasmid	Reverse		GGTGTTATGAGCCATATTCAACG
To create CjeCas9 variant expression plasmid	Reverse		GGTGTTATGAGCCATATTCAACG
To create CjeCas9 variant expression plasmid	Forward		GCAAGAATTTTGGCATTTGATATAGG
To create CjeCas9 variant expression plasmid	Reverse		CCTATATCAAATGCCAAAATTCTTGCTGC
To create CjeCas9 variant expression plasmid	Forward		GCAAGAATTTTGGCATTTGCGATAGGAATAAGCTCTATTGGCTGG
To create CjeCas9 variant expression plasmid	Reverse		GCTTATTCCTATCGCAAATGCCAAAATTCTTGCTGCATTGG
To create CjeCas9 variant expression plasmid	Forward		GCTTGAAATTGATCACATTTACCCTTATTCTAG
To create CjeCas9 variant expression plasmid	Reverse		GGGTAAATGTGATCAATTTCAAGC
To create CjeCas9 variant expression plasmid	Forward		GCTTGAAATTGATGCGATTTACCCTTATTCTAGAAGTTTTGATGATTCTTATATGAATAAAG
To create CjeCas9 variant expression plasmid	Reverse		GGGTAAATCGCATCAATTTCAAGCATTTTTTCATC
To create CjeCas9 variant expression plasmid	Forward		GCGAACGACCTACACCGAACTG
To create CjeCas9 variant expression plasmid	Reverse		CAGTTCGGTGTAGGTCGTTCGC
Sequencing primer-pET28b SpyCas9 - Insert - Seq 1	Forward		CTCACTATAGGGGAATTGTGAGCGG
Sequencing primer-pET28b SpyCas9 - Insert - Seq 2	Forward		CCTGTCACTCGGGCTGACC
Sequencing primer-pET28b SpyCas9 - Insert - Seq 3	Forward		CCTAAACACTCTCTGCTGTACGAG
Sequencing primer-pET28b SpyCas9 - Insert - Seq 4	Forward		GGGAAAGGATGAAGAGG
Sequencing primer-pET28b SpyCas9 - Insert - Seq 5	Forward		CAAGACCGAGATTACACTGGC
Sequencing primer-pET28b SpyCas9 - Insert - Seq 6	Forward		CATGAGCGGATACATATTTG
Sequencing primer-pET28b SpyCas9 - Vector - Seq 1	Forward		GGTTAAAAAATGAGCTGATTTAAC
Sequencing primer-pET28b SpyCas9 - Vector - Seq 2	Forward		CGCTACCTTTGCCATGTTTC
Sequencing primer-pET28b SpyCas9 - Vector - Seq 3	Forward		GGTATCTTTATAGTCCTGTCG
Sequencing primer-pET28b SpyCas9 - Vector - Seq 4	Forward		CATGCCCGGTTACTGGAAC
Sequencing primer-pET28b SpyCas9 - Vector - Seq 5	Forward		GGCGGTTTGCGTATTGGGCG
Sequencing primer-pET28b SpyCas9 - Vector - Seq 6	Forward		CAGACTGGAGGTGGCAACGC
Sequencing primer-pML1B CjeCas9 - Insert - Seq 1	Forward		GGGGAATTGTGAGCGG
Sequencing primer-pML1B CjeCas9 - Insert - Seq 2	Forward		GAAGAGGTGCTTAGTGTAGC

Table S9.11 | Primer list with related restriction sites and guide RNA used in this study. (continued)

Primer name	Direction	Restriction site	Primer sequence (5'-3')
Sequencing primer-pML1B CjeCas9 - Insert - Seq 3	Forward		GCCAAAGAGCTAAAATAGAAAAAG
Sequencing primer-pML1B CjeCas9 - Insert - Seq 4	Forward		GCGCCGCTACAGGGCGCG
Sequencing primer-pML1B CjeCas9 - Vector - Seq 1	Forward		GGAAGCTGAGTTGGCTGCTG
Sequencing primer-pML1B CjeCas9 - Vector - Seq 2	Forward		GATCCTGGTATCGGTCTGCG
Sequencing primer-pML1B CjeCas9 - Vector - Seq 3	Forward		GAGATCCTTTTTTTCTGCGCG
Sequencing primer-pML1B CjeCas9 - Vector - Seq 4	Forward		CTGTGCGGTATTTCACACCGC
Sequencing primer-pML1B CjeCas9 - Vector - Seq 5	Forward		GCTAACCAGTAAGGCAACCCCG
Sequencing primer-pML1B CjeCas9 - Vector - Seq 6	Forward		GCATGGTTTGTTGAAAACCG
Sequencing primer-pML1B CjeCas9 - Vector - Seq 7	Forward		CCTGCATTAGGAAGCAGCCC

Table S9.12 | Plasmids used in the study.

Plasmids	Gene Insert (origin)	Backbone	Antibiotic marker	References
pCjeCas9	Cas9 (*C. jejuni*)	pcDNA 3.1(+)	Ampicillin	This study
hCas9	Cas9 (*S. pyogenes*)	pcDNA3.3-TOPO	Ampicillin	Addgene plasmid #41815
pSPacdSRedE2ΔLac1 (p(mCherry))	Red fluorescent protein	pSP329	Kanamycin	This study
eGFP-CjeCas9	Cas9 (*C. jejuni*)	pEGFP-C1	Kanamycin	This study
eGFP-CjedCas9	Cas9 (*C. jejuni*) D8A and H559A mutated	pEGFP-C1	Kanamycin	This study
eGFP-NLS (CjeCas9)	NLS from ARM region (*C. jejuni cas9*)	pEGFP-C1	Kanamycin	This study
eGFP-SpyCas9	Cas9 (*S. pyogenes*) from pMJ806 (Addgene plasmid #39312)	pEGFP-C1	Kanamycin	This study
eGFP-SpydCas9	Cas9 (*S. pyogenes*) from pMJ841 (Addgene plasmid #39318)	pEGFP-C1	Kanamycin	This study
eGFP-NLS (SpyCas9)	NLS from ARM region (*S. pyogenes cas9*)	pEGFP-C1	Kanamycin	This study
pGFP-SV40	SV40 (NLS)	pEGFP-C1	Kanamycin	This study
gRNA_AAVS1-T2	gRNA_AAVS1-T2	pCR-BluntII-TOPO	Kanamycin	Addgene plasmid #41818
pU6-Cje-sgRNA	Cj-SgRNA	pUC57	Ampicillin	Addgene plasmid #89753
pET28b SpyCas9	Cas9 (*S. pyogenes*)	pET28b	Kanamycin	Addgene plasmid #47327
pML-1B CjeCas9	Cas9 (*C. jejuni*)	pML1-B Addgene plasmid #29653	Kanamycin	This study
pET28b SpyCas9-dRuVc	Cas9 (*S. pyogenes*) D8A mutated	pET28b	Kanamycin	This study
pET28b SpyCas9-dHNH	Cas9 (*S. pyogenes*) H559A mutated	pET28b	Kanamycin	This study

Table S9.12 | Plasmids used in the study. (continued)

Plasmids	Gene Insert (origin)	Backbone	Antibiotic marker	References
pET28b SpydCas9	Cas9 (*S. pyogenes*) D8A and H559A mutated	pET28b	Kanamycin	This study
pML-1B CjeCas9	Cas9 (*C. jejuni*)	pML1-B Addgene plasmid #29653	Kanamycin	This study
pML-1B CjeCas9 dRuvC	Cas9 (*C. jejuni*) D8A mutated	pML1-B Addgene plasmid #29653	Kanamycin	This study
pML-1B CjeCas9 dHNH	pML1-B Addgene plasmid #29653	pML1-B Addgene plasmid #29653	Kanamycin	This study
pML-1B CjedCas9	pML1-B Addgene plasmid #29653	pML1-B Addgene plasmid #29653	Kanamycin	This study

Chapter 10

Summary and general discussion

Summary

The diverse class 2 CRISPR nucleases have unique molecular features that contribute to an expansive toolbox for genome and transcriptome engineering. These nucleases differ greatly in their structure and mechanisms. These differences could be exploited as complementary applications creating numerous CRISPR-based technologies possibly with favourable specificity, efficiency and/or delivery. This thesis explores the diversity of Class 2 CRISPR–Cas systems and provides mechanistic insights into different class 2 nucleases. In addition, it describes potential applications to expand the current repertoire of CRISPR-based technologies.

As described in **Chapter 1**, prokaryotes and their viral predators are engaged in a never-ending combat. To evade viral invasion, prokaryotes rely on an impressive arsenal of diverse defence mechanisms. In addition to several innate defence systems, prokaryotes have evolved a sophisticated adaptive and heritable defence system, known as CRISPR–Cas. They are a relatively recent discovery (the first experimental evidence was uncovered 12 years ago) that provide protection against invasion all mobile genetic elements, including phages, plasmids, and conjugative elements. Adaptive immunity is conferred by the integration of DNA sequences from an invader into the CRISPR array (adaptation), which is transcribed into long pre-CRISPR RNAs (pre-crRNAs) and processed into short crRNAs (expression), that guide Cas proteins to specifically degrade the cognate DNA on subsequent exposures (interference). Moreover, the ability to of Cas proteins to be programmed for sequence-specific DNA targeting and cleavage has been adapted for the development of highly effective tools for genome engineering. In order to counter CRISPR–Cas systems, phages have a developed set of anti-CRISPR (Acr) proteins that provide them with direct and specific means to inhibit targeting by CRISPR–Cas systems. Furthermore, anti-CRISPRs have provided new biotechnological resources to precisely control the use of CRISPR–Cas systems for genome editing. Apart from their canonical role in host defence, CRISPR–Cas systems also appear to play a role in endogenous transcription control, and regulation of important lifestyle-based bacterial phenotypes such as pathogenicity.

Chapter 2 gives an overview of the basic biology, evolution, mechanisms, and applications of the CRISPR–Cas systems. Due to the everlasting arms race between prokaryotes and their viruses, the rapid evolution of CRISPR–Cas systems has resulted in extreme structural and functional diversity. As a result, a plethora of distinct CRISPR–Cas systems are represented in genomes of most archaea and almost half of the bacteria. The key players of this system are the crRNA binding effector complexes, and the associated nuclease domains. CRISPR–Cas systems are currently grouped into two classes each of which is subdivided into three types. Class 1 systems (consisting of types I, III, and IV) use a multi-subunit protein complex to achieve interference, and class 2 systems (consisting of types II, V, and VI) utilize a single multi-domain protein. The widespread occurrence of class 1 systems together with the proliferation of the ancient RNA recognition motif (RRM) domain in class 1 proteins strongly suggests that the ancestral CRISPR–Cas systems belonged to class 1. Most likely, class 2 variants then evolved via several independent replacements of the class 1 locus with nuclease genes that were derived from distinct mobile genetic elements. In particular, type V protein variants seem to have evolved from different families of the *tnpB* transposase genes that are widespread in transposons, whereas the type II nucleases may have evolved from IscB, a protein with two nuclease domains that belongs to a recently identified distinct transposon family.

Cas12a (previously Cpf1) nuclease of type V-A CRISPR–Cas systems has been successfully

repurposed for genome editing applications in a wide range of organisms. The high efficiency and specificity of Cas12a in human cells compared to the widely used type II nuclease Cas9, makes Cas12a a robust and reliable tool for genome editing. **Chapter 3** provides a step-by-step protocol for heterologous expression of *Francisella novicida* Cas12a in *Escherichia coli*. In addition, it includes a protocol for high-purity purification of recombinant FnCas12 protein that can be used for biochemical and genome editing experiments. In addition, instructions to perform *in vitro* cleavage assays that can be used to verify the activity of the purified proteins are provided. The protocols can also be used for the purification of other Cas12a orthologues. An adapted version of this protocol has been used for the purification of other Class 2 proteins throughout this thesis.

The mechanism of crRNA maturation in CRISPR–Cas12a systems is described in **Chapter 4**. Unlike the type II nuclease Cas9, which utilizes a tracrRNA as well as endogenous RNaseIII for maturation of its dual crRNA/tracrRNA guides, pre-crRNA processing in the Cas12a system proceeds in the absence of tracrRNA or other Cas proteins. It was demonstrated that Cas12a nucleases possess a previously unknown RNase domain that is responsible for cleaving the pre-crRNA to generate the mature crRNAs. The typical cleavage pattern revealed that Cas12a recognizes specific secondary structures and/or motifs on its direct repeat (DR). Furthermore, Cas12a recognizes specific nucleotides at the 5' flank of the DR stem loop (U22-A37). Substitution at positions A19 and U20 abolish RNA cleavage, substitution at position A18 and U21 weakens RNA cleavage, but substitutions at positions A16 and U17 did not affect pre-crRNA processing. Furthermore, the ability to autonomously process crRNA has significant implications from a genome editing standpoint, as it provides a simple route to editing multiple genomic loci at a time (multiplex editing). Using a single customized CRISPR array up to four genes (*DNMT1*, *EMX1*, *VEGFA*, and *GRIN2b*) in mammalian cells *ex vivo* and up to three genes (*Mecp2*, *Nlgn3*, and *Drd1)* in mouse brain cells *in vivo* were shown to be edited simultaneously.

The characterisation of a novel, diminutive type V-U1 Cas protein from *Mycolicibacterium mucogenicum* (MmuC2c4) is described in **Chapter 5**. Type V-U proteins are highly similar to the typical transposon-encoded TnpB-like proteins and each of them (type VU-1 to type VU-5) appear to have originated independently from distinct TnpB families. Akin to most type V proteins, MmuC2c4 was shown to recognize a 5'-TTN-3' PAM on a double-stranded target DNA. Unexpectedly, it was found that this enzyme does not cleave dsDNA and analysis in *E. coli* indicated a crRNA-guided MmuC2c4-mediated silencing of gene expression. In addition, MmuC2c4 was used for single- and multiplex- gene silencing in *E. coli*. Moreover, experiments that suggest that the RuvC-dependent ribonuclease activity of MmuC2c4 appears to enhance the silencing effect of MmuC2c4 have been performed.

The characterisation of a type II-C Cas9 orthologue of the thermophilic bacterium *Geobacillus thermodenitrificans* T12, ThermoCas9, is described in **Chapter 6**. This is one of the first reports that provides fundamental insights into a thermophilic CRISPR–Cas9 family member. It was demonstrated that ThermoCas9 is active *in vitro* between 20 and 70 ℃, that the structure of its sgRNA influences its activity at elevated temperatures, it has a more stringent PAM-preference at lower temperatures, it does not tolerate extensive spacer-protospacer mismatches, and it preferentially cleaves plasmid DNA compared to linear DNA. Furthermore, ThermoCas9 was employed for *pyrF* gene deletion and transcriptional silencing of *ldhL* gene at 55 ℃ in *Bacillus smithii* ET 138 and for *pyrF* gene deletion at 37 ℃ in *Pseudomonas putida*. This is the first time Cas9-based bacterial genome editing and silencing tools were used at temperatures above 42 ℃.

Four Cas12a orthologues were assessed for their salt tolerance as well as pH- and temperature-stability using biochemical assays as described in **Chapter 7**. Subsequently, *Francisella tularensis subsp. novicida* (FnCas12a) and *Eubacterium eligens* (EeCas12a) were applied for genome editing in a moderate thermophilic bacterium, *Bacillus smithii.* It is demonstrated that FnCas12a and EeCas12a are sub-optimally active *in vivo* at temperatures above 45 ℃. The wide growth temperature range of *B. smithii* ET 138 was employed for the controllable induction of Cas12a expression at temperatures below the 45 ℃ threshold. Plasmid-borne editing templates were employed for homologous recombination (HR)-based deletion of *pyrF* gene from the genome of *B. smithii* at 45-55 ℃. Transfer of the cultures at lower temperatures induces the expression of active Cas12a. Subsequently, Cas12a introduces lethal double-stranded DNA breaks to the non-edited cells, acting as a counter-selection mechanism. It was demonstrated that a mutant can be generated within a short span of 2-3 days. This process can be easily adapted for gene editing applications in a wide variety of both mesophilic and moderate thermophilic organisms.

The potential to harness the activity of anti-CRISPR (Acr) proteins for controllable bacterial genome engineering was investigated in **Chapter 8**. The Acr protein from *Neisseria meningitidis* (AcrIIC1$_{Nme}$) was employed as an "on/off-switch" to control the activity of thermostable Cas9 orthologues from *Geobacillus thermodenitrificans* T12 (ThermoCas9) and *Geobacillus stearothermophilus* (GeoCas9). Initially, it was proven that both ThermoCas9 and GeoCas9 can introduce lethal dsDNA breaks in *E. coli* at 37 ℃ in a tuneable manner. Next, it was demonstrated that AcrIIC1$_{Nme}$ traps both tested Cas9 orthologues in a DNA-bound, catalytically inactive state. The Cas9/AcrIIC1$_{Nme}$ complexes can promote a transcriptional silencing effect with efficiency comparable to the catalytically "dead" ThermodCas9 and GeodCas9 variants. Finally, a single-vector, tightly controllable and highly efficient dCas9/AcrIIC1$_{Nme}$-based tool for coupled silencing and targeting in *E. coli* was developed. This tool may serve as a basis for further developing a controllable genome editing and transcriptional regulation in model as well as non-model microorganisms.

In **Chapter 9**, a novel biological role and mechanism for the CRISPR–Cas9 system of the pathogen *Campylobacter jejuni* (CjeCas9) was uncovered. It was demonstrated that upon *C. jejuni* infection of human cells, CjeCas9 is secreted into the cytoplasm of the infected cells and it can autonomously enter the nucleus. Inside the nucleus, it catalyses metal-dependent and sequence-independent nicking of double stranded DNA, eventually leading to cell death. Genome editing using CjeCas9 was compared with the commonly used Cas9 from *Streptococcus pyogenes* (SpyCas9), and the latter was shown to be superior in creating indels. It was concluded that the unique catalytic features make CjeCas9 nuclease less suitable for genome editing applications.

In summary, the research described in this PhD thesis has uncovered novel molecular requirements and mechanisms of several unique Class 2 CRISPR–Cas nucleases. Besides gaining insights into their biochemical mechanism, the potential of Class 2 nucleases has been harnessed for biotechnological applications. Additionally, a unique role and mechanism of CRISPR–Cas in virulence has been elucidated. The characterisation of nucleases such as FnCas12a, EeCas12a, MmuC2c4 and ThermoCas9 opens up exciting possibilities of utilizing them as genome and transcriptome engineering tools.

"Learning is a process where knowledge is presented to us, then shaped through understanding, discussion and reflection"

– Paulo Freire

General discussion

Human stories behind ground-breaking discoveries and inventions are often considered as a series of lucky coincidences. However, most discoveries in the field of molecular biology (and perhaps other scientific disciplines) are seldom truly serendipitous. Extraordinary scientific advancements are often built on decades of collaborations and contributions of an inspiring ensemble of many accomplished scientists. Even the most serendipitous discoveries result from of extensive planning – of hypothesis-free and hypothesis-driven science, of a series of fresh perspectives and careful experiments to follow an interesting observation. This is how a group of researchers with creative minds and relevant expertise discovered the CRISPR–Cas system, unravelled its molecular mechanisms, and repurposed it as a powerful tool for fundamental biological research and applications in science, medicine, and biotechnology. Since 1987, CRISPR–Cas systems were discovered during sequencing projects. After their predicted function as a defence system, a series of seminal genetic and biochemical studies has eventually allowed for a myriad of applications, altogether resulting in a biotechnological revolution in the field of genome editing.

Ever since the first experimental evidence that CRISPR–Cas systems provide adaptive immunity in prokaryotes (*20*) scientists have endeavoured to understand the mechanistic basis of this fascinating immune system. By screening rapidly growing genomic and metagenomic databases along with different research groups using different model systems for elucidating the different systems, a wide diversity of CRISPR–Cas variants and mechanisms have been discovered and characterised. As outlined in **Chapter 2**, the latest CRISPR–Cas classification separates all systems into two main classes. The class 1 systems have multi-protein effector complex, whereas in class 2 systems, all interference activities are undertaken by a single, multi-domain protein complex. In 2015, the two classes were subdivided into six types and 19 subtypes which has very quickly expanded to more than 30 subtypes now (*31*, *284*). The classification of class 1 CRISPR–Cas systems that include types I, III and IV, has remained relatively stable over the years (*31*). They are the most commonly found CRISPR–Cas systems comprising ~90% of all identified CRISPR–Cas loci (*31*). On the other hand, class 2 CRISPR–Cas diversity has exploded in the recent years due to the dedicated efforts of comprehensive identification and characterisation of class 2 systems as a quest for potential new genome editing tools (*82*, *284*). Class 2 CRISPR–Cas systems include types II, V and VI. The most recent discoveries include diverse variants of type V as well as type VI, which is the first and so far, the only CRISPR–Cas type that exclusively cleaves RNA. Fundamental characterisation of different CRISPR–Cas (sub)types is vital to understand the molecular underpinnings of life in the microbial world.

Certain key mechanistic details of different class 2 effector nucleases have been uncovered. This includes the mechanism of crRNA maturation (**Chapter 4**), thermostability of the ribonucleoprotein complex (**Chapter 6** and **7**), RNA-independent cleavage of DNA (**Chapter 9**), and

the nature of the target nucleic acids (DNA or RNA) (**Chapter 4**, **5**, **6**). This last chapter aims to further reflect on our current knowledge of the different class 2 CRISPR–Cas types, and in particular the gaps in our current knowledge, as well as to discuss several additional experiments and place the data in a broader context. The different stages of CRISPR-based immunity (adaptation, expression, interference) of the different class 2 CRISPR–Cas systems will be discussed here, as well as their adoption for prokaryotic and eukaryotic genome editing.

Adaptation of the CRISPR memory

To keep up pace with a constantly changing pool of hostile invaders, CRISPR–Cas systems need to regularly update their stored genetic information, in order to facilitate efficient target recognition. The acquisition of new immunological memory occurs via a process termed adaptation. As briefly outlined in **Chapter 2**, adaptation of CRISPR arrays is crucial to ensure persistent CRISPR–Cas defence (*146*, *454*). Due to the complexity in studying CRISPR adaptation under laboratory conditions, its molecular details are one of the most enigmatic aspect of CRISPR immunity. Nevertheless, significant progress has been made towards understanding the molecular mechanisms governing the adaptation machinery.

Illustrative of their central roles in spacer integration, the *cas1* and *cas2* genes are associated with nearly all CRISPR–Cas systems (*82*). Cas1 and Cas2 proteins form a $Cas1_4$-$Cas2_2$ integrase complex (*152*, *455*) (hereafter Cas1-Cas2) which is the highly conserved "information processing" module of spacer integration. Spacer integration requires three main processes: i) substrate capture ii) recognition of the CRISPR locus and iii) integration of the spacer within the array. Cas1-Cas2-mediated spacer integration prefers dsDNA substrates and proceeds via a mechanism resembling retroviral integration (*158*, *161*). In addition to Cas1-Cas2, a single repeat, at least part of the leader sequence (*27*, *149*, *151*, *456*), and additional host factors for repair of the insertion sites (*e.g.* DNA polymerase) are essential for adaptation (*167*).

Type II-A systems are the best studied type II CRISPR–Cas systems. Two Cas1 dimers and a single Cas2 dimer form a heterohexameric integrase complex in type II-A CRISPR–Cas systems (*457*). Apart from Cas1-Cas2, adaptation in type II-A systems require Cas9, a tracrRNA and Csn2 protein (*150*, *168*). Cas9 aids in the generation of correct spacer substrates by enhancing the acquisition of spacers with compatible PAMs using its PAM-recognition domain (*168*). The deletion of *csn2* gene from the CRISPR locus impaired the acquisition of spacers from invading phages, although the exact role of *csn2* remains unclear (*150*, *168*). Recently, it was demonstrated that Csn2 directly interacts with Cas9 to form a stable, stoichiometric adaptation module with the Cas1-Cas2 integrase complex, suggesting that Csn2 mediates the involvement of Cas9 in adaptation (*458*).

The typical type V-A locus consists of *cas12a, cas4*, *cas1*, *cas2* genes. The potential role of Cas4 and/or Cas12a in adaptation of new spacers remains to be demonstrated. In type I systems, Cas4 together with Cas1 and Cas2 plays an important role in spacer acquisition. The Cas4 from *Bacillus halodurans* type I-C interacts tightly with Cas1 and processes pre-spacers (*459*). A similar role of two Cas4 proteins that act in coordination for spacer acquisition, has also been identified in the I-A CRISPR–Cas systems of *Pyrococcus furiosus* (*460*). Cas4 was shown to play a crucial role in the selection of functional PAM sequences and pre-spacer processing in type I-D CRISPR–Cas systems of *Synechocystis* sp. 6803 (*461*). It is anticipated that Cas4 in the type V-A locus could play a similar role in adaptation process of type V-A systems. In addition, akin to Cas9, Cas12a could also assist in the selection of suitable PAM sites on the pre-spacer substrates using the PAM-recognition domain.

After the discovery of type V-A systems two new V subtypes and one new VI subtypes were identified by using Cas1 as the seed for the search algorithm (*76*). In type V-B, Cas1 is fused to Cas4 to form a single polypeptide, indicating a stronger involvement of Cas4 in the spacer acquisition process. Recently, a gene fusion between *cas4* and *cas1* from the I-U CRISPR–Cas system of *Geobacter sulfurreducens* was shown to be capable of introducing functional spacers carrying interference proficient 5'-TTN -3' PAM sequences at much higher frequencies than unfused Cas4 adaptation modules (*462*). Mutations of the Cas4-domain catalytic residues resulted in dramatically decreased spacer acquisition, and a loss of PAM selectivity. Since the Cas4/1 proteins from I-U and V-B are phylogenetically closely related (*463*), it is possible that the adaptation modules have been shared between these two distinct systems by horizontal gene transfer and, therefore, it is very likely that the Cas4/1fusion protein plays a similar role in type V-B systems. Experimental characterisation of the adaptation machinery in these type V systems is pending.

Upon using only the CRISPR array as a search seed to detect non-autonomous class 2 systems, three new CRISPR–Cas subtypes were detected. To ensure precision in detecting the correct CRISPR arrays, only the predictions made using the Piler-CR (*464*) and CRISPRfinder (*30*, *465*) methods were taken for the analysis. The search yielded an additional heterogeneous subtype of putative type V systems and two subtypes of type VI systems (*82*). This reflects that many CRISPR–Cas loci lack the adaptation module and numerous such 'orphan' CRISPR arrays exist, some of which seem to be functional (*466*). Subtype V-(G-I) and all V-U systems, one of which is described in **Chapter 5**, do not contain an adaptation module. Notably, the effector modules of subtype III systems are known to utilize spacers from other systems, testifying to the modularity of CRISPR–Cas systems (*181*, *182*, *199*, *467*) Although most genomes encoding type V-U systems do not encode any other identifiable CRISPR–Cas systems, it is possible that these systems, use crRNAs transcribed from different CRISPR arrays, akin to III systems (*31*).

The lack of adaptation module in subtype V-U (*82*) suggests that the first stage of the evolution of new class 2 CRISPR–Cas systems involved the random insertion of a transposase TnpB-encoding element next to an orphan CRISPR array. Next, the association between CRISPR and the TnpB derivative would become fixed in the microbial population, conceivably owing to the emergence of a novel (unknown) function. Interestingly, two independent studies have demonstrated that a class of bacterial transposons encoding their own CRISPR–Cas systems can mobilize using RNA-guided DNA integration (*81*, *468*). Some of these TnpB-derived proteins could subsequently expand in protein size through internal duplications and/or the insertion of additional domains to grow into a functional effector protein, resulting in the archetypal bilobed structure and - in some cases - its association with an adaptation module through recombination with a different CRISPR–Cas locus. Compatible with this scenario, the Cas1 proteins of different subtypes of type II and of type V systems are most closely related to Cas1 proteins of different type I systems (*76*). It will be highly interesting to investigate the adaptation and interference mechanisms of the type V-U systems as it may provide exciting insights into the different stages of CRISPR–Cas evolution.

Expression of the guide RNAs that lead the Cas effector nucleases

The expression stage comprises the transcription of the CRISPR array and *cas* genes, followed by the processing of the precursor CRISPR RNA transcript (pre-crRNA) to generate mature crRNA guides (*40*, *469*). In class 1 CRISPR–Cas systems, three variants of Cas6, namely Cse3 (CasE, Cas6e), Cas6 and Csy4 (Cas6f), have been identified as endoribonucleases that cleave

the pre-crRNA within their repeat sequences to generate mature crRNAs (*40*, *179*, *469*).

Class 2 CRISPR–Cas type II systems employ a unique mechanism where Cas9 is the only Cas protein required for the steps of crRNA biogenesis as well as for the interference with invading DNA. Type II systems rely on a tracrRNA which binds to the complementary repeat sequences in pre-crRNA. Subsequently, the housekeeping endoribonuclease III catalyses tracrRNA-directed cleavage within the pre-crRNA repeats, as the RNA duplex is stabilized by Cas9 (*49*). Cas9 bound to the cleaved crRNA:tracrRNA hybrid forms the effector complex that mediates target DNA recognition and cleavage (*49*, *51*, *52*). The absence of genes encoding endoribonuclease III-like activities in archaeal genomes probably explains why type II-A systems are found exclusively in bacteria. Alternatively, endoribonuclease III is dispensable for the maturation of crRNAs in type II-C in *Neisseria meningitidis*. In this particular case, mature crRNAs are expressed directly from promoter sequences located within the repeats of the CRISPR arrays (*186*).

Investigation of pre-crRNA processing in type V-A

In type V-A systems, an associated *cas6* gene and a tracrRNA are absent (*83*). This suggested the existence of a novel, fundamentally different crRNA maturation process in the type V-A systems, involving (an)other Cas protein(s) and/or non-Cas ribonuclease(s).

We initially set out to check involvement of four *E. coli* RNases (RNase BN, Z, G and D), but the results were negative. To explore the plausible role of a Cas protein in pre-crRNA processing, it was tested if Cas12a autonomously processed the pre-crRNA. Together with an earlier report (*187*), this study described in **Chapter 4**, uncovered a novel pathway for the maturation of the crRNAs in which the effector protein Cas12a itself catalyses crRNA maturation. Apart from pre-crRNA processing by the AsCas12a and LbCas12a variants shown in **Chapter 4**, we additionally demonstrated the pre-crRNA processing activity by FnCas12a and Cas12a from *Moraxella bovoculi* 237 (MbCas12a) (**Figure 10.1**).

Structural studies based on different Cas12a variants bound to a crRNA revealed that the 5′-handle of the crRNA repeat adopts a pseudoknot structure, rather than a simple stem-loop structure predicted from its nucleotide sequence (*140*, *270*, *470*). Early biochemical studies suggested that pre-crRNA processing by Cas12a is catalysed by divalent cations upon (*187*). However, it was subsequently demonstrated that the two Mg^{2+} ions help stabilize Cas12a binding to the RNA but are not required for catalysis (*270*, *470*). Pre-crRNA processing by Cas12a proceeds via metal-independent acid-base catalysis involving a nucleophilic attack of the 2′-hydroxyl at the scissile phosphate of the ribonucleotide upstream of the processing site (*270*).

Investigation of pre-crRNA processing in type V-U1

While working on the characterisation of type V-U1 systems in **Chapter 5**, we also explored if, like type V-A systems, type V-U1 systems are capable of autonomous pre-crRNA processing. To this end, we codon-harmonized and synthesized the *c2c4* gene from *Clostridiales bacterium* DRI and cloned it into a pML-1B cloning vector containing an N-terminal hexahistidine tag and a TEV protease cleavage site. The plasmid was then transformed to *E. coli* Rosetta (DE3) for heterologous expression and the CbaC2c4 effector was purified using three chromatographic steps: Ni-NTA, ion exchange and gel filtration chromatography (**Figure 10.2A**). The pre-crRNA was designed based on the native array in the *Clostridiales bacterium* DRI locus and consists of

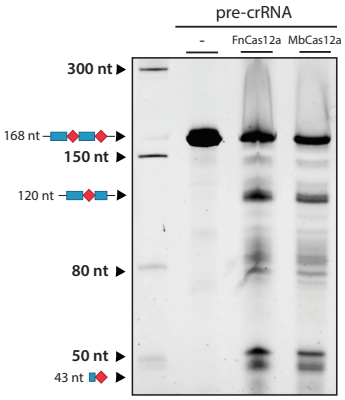


Figure 10.1|Pre-crRNA processing by recombinant FnCas12a and MbCas12a nucleases. 10% Urea-PAGE gel on which the processed pre-crRNA transcripts were resolved. RNA was visualized after staining with SYBR-gold. The blue rectangle represents a direct repeat, and the red diamond represents a spacer.

three repeats each of 37 nucleotides (nt) interspaced by synthetic spacers of 38 nt. Since the direction of transcription in the host organism was not clear, we generated artificial pre-crRNA transcripts from both the ends (Forward (FW) and Reverse (RV)) of the CRISPR array by *in vitro* transcription. The two artificial pre-crRNAs were subjected to the purified CbaC2c4 effector and incubated for 30 minutes at 37 ˚C. The CbaC2c4 effector was shown to possess crRNA maturation activity as it cleaved the FW pre-crRNA into smaller intermediates and mature crRNAs (**Figure 10.2B**). No cleavage of the control RV pre-crRNA was detected (**Figure 10.2B**). Sequencing of the generated mature crRNA revealed that CbaC2c4 cleaved at the phosphodiester bond four nucleotides upstream of the conserved (predicted) crRNA-repeat hairpin.

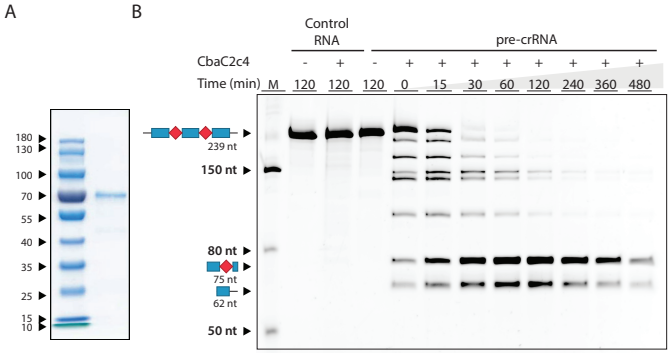


Figure 10.2|Pre-crRNA processing by CbaC2c4. (A) Coomassie Brilliant Blue stained 10% SDS-PAGE gel on which the purified CbaC2c4 was resolved. **(B)** 10% Urea-PAGE gel on which the processed pre-crRNA transcripts were resolved. RNA was visualized after staining with SYBR-gold. The control RNA is the pre-crRNA transcribed in the reverse orientation. The blue rectangle represents a direct repeat, and the red diamond represents a spacer.

Cleavage of the pre-crRNA by CbaC2c4 appears be divalent cation-dependent as the addition of the metal-chelating agent, EDTA abolished the pre-RNA processing activity of CbaC2c4 (**Figure 10.3**). Highest pre-crRNA cleavage activity was observed in the presence of Mg^{2+} and Mn^{2+} ions in the reaction (**Figure 10.3**). So far, no target (double-stranded DNA, dsDNA) binding or cleavage activity of CbaC2c4 has been detected. Next, it would be interesting to investigate its biological role (if any) in host defence. Therefore, crRNA-guided CbaC2c4 binding and/or cleavage of other potential target nucleic acid substrates such as ssDNA, ssRNA and dsRNA should be explored in future research.

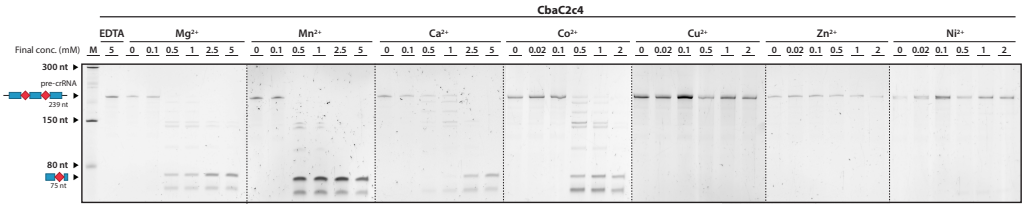


Figure 10.3 | Divalent cation dependent processing of the pre-crRNA by CbaC2c4. 10% Urea-PAGE gel on which the processed pre-crRNA transcripts were resolved. RNA was visualized after staining with SYBR-gold. The blue rectangle represents a direct repeat, and the red diamond represents a spacer.

Pre-crRNA processing in type VI

Type VI CRISPR–Cas systems are defined by the presence of the Cas13 effector (formerly C2c2). Cas13 nucleases have been demonstrated to exclusively cleave RNA targets (*76*, *82*, *224*, *471*, *472*). They possess two distinct active sites, both conferring RNase activity: one for pre-crRNA processing and the other is involved in target RNA degradation (*224*). Akin to Cas12a, crRNA maturation by Cas13 involves metal-independent cleavage upstream of the stem-loop, without the requirement of tracrRNA or other host factors (*76*). Typically, the direct repeats in type VI pre-crRNAs form a bulge-containing stem-looped structure, which appears to be a conserved feature of mature crRNAs in Cas13 nucleases. Disruption of this bulge did not affect pre-crRNA processing, whereas it impeded target RNA degradation (*85*, *473*). In subtype VI-A, the processing of pre-crRNA was shown to be dispensable for target RNA degradation; implying that pre-crRNA molecules were functional as guides for RNA targeting (*86*).

Concluding, pre-crRNA processing to generate mature crRNAs can be mediated by distinct components and mechanisms. In general, unique recognition of the pre-crRNA sequence and structure enable discrimination of cognate pre-crRNAs from other cytosolic RNAs, which results in specific maturation of crRNAs that associate with their respective interference complexes. It is interesting to note that CRISPR–Cas systems that harbour CRISPR-arrays with partial palindromic repeat sequences appear to correlate with tracrRNA-independent pre-crRNA processing systems. Future studies on some of the newly discovered novel systems (such as type V-D and V-H) is anticipated to shed more light on the molecular mechanisms of the crRNA maturation complexes of different CRISPR–Cas subtypes.

Interference of the invading nucleic acids

Together with the mature crRNAs, either a single Cas protein or a complex of Cas proteins seek recurring foreign nucleic acids and utilize the nuclease activity of the Cas proteins to cleave target DNA/RNA.

In type II systems, the binding of Cas9 to the crRNA:tracrRNA hybrid triggers a large conformational rearrangement, leading to its activation (*51*). Upon activation, the PAM recognition sites are prepositioned and the dual RNA-bound Cas9 complex scans for invading nucleic acids. Once a potential target with the appropriate PAM is identified, the dsDNA is unwound, and crRNA binds to the ssDNA Target Strand (TS) leading to the formation of an R-loop. Ultimately, complete crRNA-target strand base pairing allows the mobile HNH domain to reach a catalytic conformation for cleaving the target strand. This movement of HNH also allosterically regulates the cleavage of the Non-Target Strand (NTS) by the RuvC domain, resulting in a concerted blunt double-strand cleavage 3-nt upstream of the PAM site (*51*, *52*). Following

cleavage, Cas9 remains tightly bound to the cleaved target DNA until other cellular factors displace the enzyme for recycling (*188*, *218*).

In type V-A systems, crRNA-mediated DNA cleavage by Cas12a results in a 5′ overhang instead of a blunt end cut (*83*). Initial studies suggested that the Cas12a RuvC domain is responsible for the cleavage of the NTS, while the adjacent Nuc domain mediates TS cleavage. However, more recent biochemical insights imply that Cas12a cleaves both DNA strands using a single catalytic site located in the RuvC domain, while the Nuc domain plays a role in substrate DNA coordination (*270*). The current hypothesis is that after cleavage and dissociation of the NTS from the catalytic site, the PAM-distal part of the R-loop structure unwinds, allowing the TS DNA to "fold back" to fit into the RuvC catalytic centre. Hence, this sequential mechanism of DNA cleavage by a single active site would result in a staggered-end DNA break with the 5′ overhangs (*270*, *474*).

Cas12b contains a single RuvC domain and requires both crRNA and tracrRNAs for dsDNA cleavage (*76*). DNA cleavage by Cas12b generates 7-nt staggered ends distal to the protospacer adjacent motif (PAM) (*76*). Crystal structures of *Alicyclobacillus acidoterrestris* Cas12b (AaCas12b) have revealed that both the TS and NTS can be independently positioned for cleavage within a single active site cleft at the interface of the RuvC and Nuc domains in Cas12b (*293*). In view of this, it is highly interesting to elucidate the sequential order of cleavage of DNA strands leading to double-strand breaks of target DNA.

Besides canonical crRNA-target DNA targeting and cleavage, it was recently demonstrated - that in the presence of divalent manganese and/or magnesium cations - certain Cas nucleases cleave DNA in a crRNA- and sequence- independent manner. Recombinant SpyCas9 (type II-A) and FnCas12a (type V-A) were shown to degrade ssDNA while FnCas9 (type II-B) nicks dsDNA (*443*). This crRNA-independent activity is dependent on the RuvC domain for SpyCas9, the HNH domain for FnCas9, and both the RuvC and Nuc domains for FnCas12a (*443*). This ability of SpyCas9 was corroborated in **Chapter 9** where we tested metal independent DNA cleavage activities of SpyCas9 and CjeCas9 (*443*). Another interesting study demonstrated that target strand binding activates the RuvC catalytic site of Cas12a and Cas12b, which transforms the proteins into nonspecific ssDNases that do not only cut the targeted DNA in *cis* but also degrade non-target ssDNA in *trans* (*287*, *475*). Upon binding of the dsDNA target, Cas12a undergoes a series of checkpoints which exposes the RuvC domain, leading to the cleavage of the NTS followed by cleavage of the TS, and subsequently remains activated, thereby cleaving bystander ssDNA indiscriminately (*476*). This reveals a functional convergence of Cas12 enzymes with the type III CRISPR-Csm/Cmr systems which also exhibit target-activated, nonspecific ssDNase (*203*).

Several functionally diverse type V nucleases, including both DNA- and RNA- targeting nucleases, have been identified from extensive datamining of (meta)genomic databases. These include Cas12c (*76*), Cas12d (CasY) and Cas12e (CasX) (*74*), three subtypes of Cas12f (Cas14) (*77*), Cas12g, Cas12h, and Cas12i (*80*). Akin to Cas12b, the three characterised Cas12c nucleases utilize dual-RNA guides to bind and cleave dsDNA targets (*80*). While two of these Cas12c nucleases require a 5'-TG-3' PAM for dsDNA targeting, the third Cas12c requires a simple 5'-TN-3' PAM.

Cas12f is an RNA-guided nuclease of 400 to 700 amino acids that cleaves ssDNA targets without sequence restrictions. Akin to Cas12a and Cas12b, target recognition activates *trans*-acting, non-specific, ssDNase activity of Cas12f. Akin to Cas12f, Cas12g also utilizes a dual-RNA guide,

but targets RNA instead of ssDNA. This is the first type V effector to display RNA-guided RNA interference. In addition to cis-cleavage of target ssRNA, Cas12g displays non-specific trans-cleavage of both non-target ssDNA and ssRNA (*80*). Cas12i and -h are RNA-guided dsDNA cutters that recognize a 5'-TTN-3' PAM, however, Cas12i displays predominantly RNA-guided dsDNA nicking (*80*).

Cas13 contains two higher eukaryotes and prokaryotes nucleotide (HEPN)-binding EndoRNase domains (*76*, *82*). Cas13 nucleases display RNA-guided ssRNA targeting with a preference for targets with a protospacer flanking site (PFS) (*224*). Single and double mismatch analysis revealed the presence of a mismatch-sensitive 'seed region' in the centre of the spacer. This can be explained by the structural finding that the central part of the spacer is exposed to the solvent and recognizes the target RNA first (*477*). Target ssRNA binding to crRNA induces conformational changes in Cas13 that promotes further crRNA-ssRNA pairing. Upon full binding of the crRNA to the target RNA, Cas13 RNase activity is prompted by the approximation of the catalytic sites of both HEPN domains (*477*). This subsequently activates the Cas13 to exhibit collateral activity, leading to non-specific degradation of any nearby transcripts regardless of complementarity to the spacer, akin to the Csm6 enzyme of type III systems (*224*, *288*). Cleavage occurs primarily before uridine (U) residues. Additionally, Cas13 nucleases display a preference for ssRNA substrate without any higher-order secondary structures. This is most likely because they lack the helicase activity necessary for opening up double-stranded RNA regions for crRNA binding (*224*, *471*). The Cas nucleases of different class 2 CRISPR–Cas systems characterised during the course of this thesis are summarized in shown in **Table 10.1**.

Table 10.1 | Characteristics of Class 2 CRISPR–Cas nucleases studied during the course of this thesis.

Characteristic	SpyCas9	CjeCas9	ThermoCas9	Cas12a	MmuC2c4	CbaC2c4
CRISPR–Cas type	II-A	II-C	II-C	V-A	V-U1	V-U1
Size (amino acids)	1368	984	1082	1,200 - 1,500	596	568
Endonuclease domains[a]	RuvC (NTS) and HNH (TS)	RuvC (NTS) and HNH (TS)	RuvC (NTS) and HNH (TS)	RuvC (NTS and TS); uncharacterised domain for pre-crRNA processing	RuvC	RuvC
Source organism	*Streptococcus pyogenenes* SF370	*Campylobacter jejuni* NCTC 11168	*Geobacillus thermodenitrificans* T12	*Francisella novicida*, U112, *Moraxella bovoculi* 237, *Acidaminococcus* sp. BV3L6, *Eubacterium eligens*	*Mycolicibatcerium mucogenicum* CCH10 A2	*Clostridiales bacterium* DRI-13
Pre-crRNA processing	RNase III	RNase III	RNase III	autonomous processing	autonomous processing	autonomous processing
tracrRNA	yes	yes	yes	no	no	no
Canonical (cis) target cleavage	dsDNA	dsDNA	dsDNA	dsDNA and ssDNA	-	?
PAM[b]	5'-NGG-3'	5'-NNNNACA-3'	5′-CNAA-3′ 5′-CMCA-3′	5'-(T)TTV -3'	5'-TTN-3'	?
Collateral (trans) non-specific cleavage	x	x	x	ssDNA	*ssRNA*	?
Other Details	Target cleavage leads to blunt double-stranded DNA break, Guide-free cleavage of ssDNA in the presence of Mg^{2+} ions	Guide-free CjeCas9 nicks dsDNA using the HNH domain in the presence of Mg^{2+} ions	Thermostable enzyme, active from 20 °C to 70 °C, higher specificity observed at lower temperatures	Target cleavage results in staggered double-stranded DNA break	Alters nascent RNA in *trans*	?

a. TS and NTS indicate domain that cleaves target strand (TS) or non-target strand (NTS). b. N: A, C, G, or T; M: C or A; V: C or A or G

Genome editing of eukaryotic organisms

In the landmark paper released in 2012, Jinek *et al.* laid the foundation to what would ultimately become a revolution in genome editing and transcriptional control (*51*). It was demonstrated that the crRNA and the tracrRNA (see Interference) can be fused into a single-guide RNA (sgRNA) and that together with Cas9, this two-component Cas9-sgRNA endoribonuclease can be programmed to cleave plasmid DNA *in vitro*. Single-base mutations in the PAM and in the protospacer sequence targeted by the 3′ region of the sgRNA compromised Cas9 activity, whereas mutations in the protospacer sequence targeted by the PAM-distal 5′ region of the sgRNA do not always jeopardize Cas9 cleavage activity (*51*).

Subsequently, two orthologues of Cas9, *Streptococcus* pyogenes Cas9 (SpyCas9) and *Streptococcus thermophilus* Cas9 (StCas9) were programmed using their respective gRNAs to mediate genome editing in human and mouse cell lines (*57*). A combination of monopartite and a bipartite Nuclear Localization Signal (NLS) was added on C- and N- terminal of codon-optimised Cas9 to allow it to localize efficiently into the human cell nucleus without any aggregation in the nucleolus. However, as demonstrated in **Chapter 9**, the NLS is dispensable for nuclear compartmentalization, as autonomous nuclear entry activity of unmodified SpyCas9 was observed. Numerous additional studies reported gene editing using Cas9 in human and animal cells (*58*–*60*, *226*) (references not exhaustive) and the use of a catalytically inactivated variant of Cas9 to achieve targeted gene repression or activation (*61*, *67*, *71*, *73*, *228*, *229*, *238*, *478*, *479*) (references not exhaustive).

Case study: use of Cas9 to cure Crouzon syndrome and hereditary vitamin D-resistant Rickets

More than half of the known human pathological variants of genetic diseases are caused by point mutations. The Crouzon syndrome, recognized by Craniosynostosis, is an autosomal dominant genetic disease that causes premature fusion of the cranial sutures and malformations of the skull in infants (*480*–*482*). Crouzon syndrome is caused by a missense mutation, that is heterozygous and is present in the fibroblast growth factor receptor 2 (*FGFR2*) gene, chromosome 10q25-26, on exon 9 at cysteine 342, also known as FGFR2C^{342}Y/+ (*483*, *484*). Currently, the only possible treatment for the Crouzon syndrome is to perform surgery to accommodate the brain to expand. Hereditary vitamin D-Resistant Rickets (HVDRR) is an autosomal recessive disease that is recognized by the loss of function of the Vitamin D Receptor (VDR) resulting in no uptake of calcium, vitamin D (1,25(OH)2D3) and phosphate by the body (*485*, *486*). In VDR, a missense mutation is present, on chromosome 12 in the DNA-binding domain at lysine 45E, also known as VDRK^{45}E/K^{45}E (*485*, *486*). The disease can be treated by oral calcium therapy and vitamin D stimulation (1.25-(OH)2D3) therapy. Instead of treating these genetic diseases, gene therapy leading to repair of the mutation could result in curing of the disease and will remove the need for continuous treatment. Both Crouzon syndrome and HVDRR are genetic bone disorders, caused by a single nucleotide mutation, making them excellent targets for use in genome editing using Cas9 combined with homology direct repair (HDR).

During the course of this thesis, we explored the possibility of using SpyCas9 nuclease to repair craniofacial mice model system cells harbouring a single nucleotide mutation in the Fibroblast Growth Factor Receptor 2 (*FGFR2*), and the vitamin D receptor (*VDR*) gene in the fibroblasts of a rickets patient. The restoration of VDR function can be monitored by measuring expression and activity of the vitamin D target 24-hydroxylase (CYP24). Initially, to determine the opti-

mal sgRNA sequence for *in vivo* applications, we designed and tested four different guides each for *VDR* (VDR 1-4) and FGFR2 (*FGFR2* 1-4) and tested them using an *in vitro* assay using both the original and mutated genes as a target. All the tested sgRNAs allowed for Cas9-mediated target DNA cleavage, albeit with difference in cleavage efficiencies. The observed difference in cleavage efficiency for different sgRNAs is in agreement with previous observations demonstrating that the targeting efficiency of SpyCas9 system relies heavily on the selected sgRNA spacer segment sequence and that it can substantially differentiate amongst different targets, even within the same gene (*51*, *57*, *58*, *226*, *487*). No difference between the cleavage of the mutated and the original gene was observed, in line with the report that single base mismatches are generally tolerated by the Cas9-sgRNA complex (*51*). Therefore, the possibility of Cas9 recutting the repaired target gene or the HDR template must be circumvented. This can be done by either introducing synonymous mutations in either the PAM or the seed sequence, thereby blocking recognition by Cas9. It has been reported that such PAM or seed blocking mutations could increase the accuracy of editing by up to 10-fold per allele (*488*). To test the ability of the designed sgRNAs to function in human cells, they were transduced in fibroblasts harbouring the mutated *VDR* gene. Surveyor assay using genomic DNA isolated the transduced cells revealed that all the designed sgRNAs can target and induce a double stranded DNA break (DSB) by Cas9 in the *VDR* gene. Currently, the possibility to correct the *VDR* gene mutations using Cas9 and HDR are being assessed. Two methods for delivery of the SpyCas9 nuclease are being tested: i) to transfect the cells with *in vitro* obtained Cas9-ribonucleoproteins complexes (RNP) and a ssDNA repair template and, ii) to transfect cells with stable Cas9 expression using a plasmid containing the appropriate sgRNA together with a ssDNA repair template. The experiments are currently ongoing. Possibly, this method can offer a future treatment of the Crouzon syndrome and the Hereditary vitamin D-Resistant Rickets.

Potential application of ThermoCas9 for human genome editing

Despite the origin of ThermoCas9 from a thermophilic bacterium, it is also active at mesophilic temperatures (37 °C) in *E. coli and Pseudomonas putida,* as demonstrated in **Chapter 6** and **8**. Moreover, while exploring the mismatch tolerance ability of ThermoCas9 nuclease in **Chapter 6** we found that the specificity of the nuclease increases at lower temperatures. Therefore, it is highly interesting for future work to investigate if ThermoCas9 is a naturally 'high fidelity' Cas9 homolog, with lower off-target cleavage compared to other Cas9s. Such high fidelity Cas9 could be very valuable for the genome editing field, especially since the wild-type SpyCas9 has been shown to mediate unwanted off-target cleavage events in human cells (*281*, *487*). The lifetime of proteins in blood is often limiting for their use as a therapeutic (*489*, *490*). Another potential advantage of using ThermoCas9 for genome editing applications could be the extended lifetime of the nuclease in human plasma, as has been demonstrated for GeoCas9 (*419*). An extended lifetime may ensure better persistence of Cas9 RNPs upon delivery into bloodstreams. Additionally, ThermoCas9, originating from a soil microorganism, might be better suited for *in vivo* applications in terms of potential immunogenicity of the therapy, compared to SpyCas9 and *Staphylococcus aureus* Cas9, derived from pathogenic bacteria (*491*). However, it remains to be seen if the high fidelity and stability of ThermoCas9 can be exploited for ThermoCas9-mediated genome editing in eukaryotic cells.

Addition of type V nucleases to the genome editing toolkit

The discovery of Cas12a has expanded the current CRISPR-based genome editing toolbox. Perhaps, the most notable feature of Cas12a is that it is a single crRNA-guided endonuclease unlike Cas9 which requires a dual guide RNA. The shorter (44 nt) crRNAs employed by Cas12a are significantly easier and cheaper to chemically synthesize as RNA oligonucleotides compared to the long (100 nt) guide RNA used by Cas9. In addition, as described in **Chapter 4**, unlike Cas9, which utilizes host RNase III to process its CRISPR array, Cas12a itself has RNase activity and processes its cognate pre-crRNA array into individual crRNAs. This is ideally suited for multiplex genome editing because multiple crRNAs can be easily expressed as a single transcript and subsequently processed into individual crRNAs by Cas12a itself (*256*). This has a major impact for genetic engineering in biotechnology as well as in biomedicine applications. In contrast to G-rich PAMs required by most Cas9s, the T-rich PAMs by Cas12a orthologues expand the range of genomic targets that can be targeted by Cas nucleases (**Figure 10.4**) (*83*). On the contrary to cleavage by Cas9, cleavage by Cas12a results in DSB with 5′ overhangs instead of a blunt-ended DSB (*83*). It has been demonstrated that LbCas12a-mediated genome editing achieves higher HR rates than SpyCas9-mediated genome editing in Zebrafish (*266*). Finally, Cas12a is an intrinsically high-fidelity genome-editing enzyme (*246*, *269*) with fewer off-target cleavage compared to wild-type Cas9, which is of utmost importance for therapeutic applications.

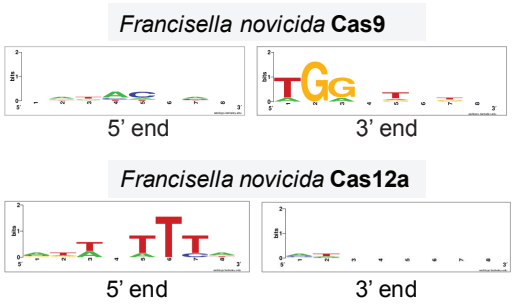


Figure 10.4 | Sequence logo of the results of *in silico* PAM prediction for two Cas nucleases, FnCas9 and FnCas12a, found in *Francisella novicida* U112 (NC_008601).

Cas12b-systems have also been repurposed for robust genome editing activity in human cells (*84*, *492*). Compared to wild-type SpyCas9, Cas12b has substantially reduced off-target activity when targeting the human genome, indicating it is inherently more specific than SpyCas9 (*81*, *84*). The recently characterised Cas12e orthologs, PlmCas12e and DpbCas12a have also been successfully used to achieve targeted gene knockout in human cells (*292*). The type V-U1 protein, MmuC2c4 described in **Chapter 5** holds great potential for genome editing and transcriptional silencing applications due to its small size. Currently, *in vivo* delivery modes for gene therapies typically rely on Adeno-associated viral (AAV), which have been approved by the U.S. Food and Drug Administration (FDA). Although promising, AAV vectors have relatively limited cargo capacity, making it challenging to deliver large Cas nucleases such as SpyCas9, AsCas12a and the guide RNAs effectively. Type V-U1 nucleases are less than half the size of the widely used nucleases, Cas9 and Cas12a, making them better suited for AAV delivery.

Genome editing of prokaryotic organisms

High-throughput genome editing methods for microbial strain engineering enable efficient development of next-generation microbial cell factories. CRISPR-based tools have advanced the use of metabolic engineering both for fundamental and systems biology studies as well as to produce green chemicals and biofuels. Shortly after the discovery of Cas9 in 2012, it was used to induce targeted mutations (insertions, deletions, and single-nucleotide substitutions) in the genomes of the bacteria *Streptococcus pneumoniae* and *Escherichia coli* (*392*). Catalytically inactivate dCas9 variants have been used for the regulation of expression of bacterial genes (either activating or repressing the transcription) (*73*). However, the application of CRISPR-based tools for genetic modification of industrially relevant non-model bacterial strains has progressed slowly (*387*, *493*, *494*), which stands in contrast to the rapid rise of CRISPR-based tools in model bacteria and eukaryotes. Nevertheless, CRISPR-based toolkits for genome editing, gene silencing, and genome-wide screening of essential genes are gradually being developed for diverse bacterial strains.

A major limitation of using CRISPR-based editing in prokaryotes is that most prokaryotes lack or only conditionally express an efficient NHEJ repair pathway (*371*). Nonetheless, some CRISPR–Cas nucleases have been combined with native or heterologously expressed (consisting of Ku and LigD) NHEJ systems (*356*, *396*, *495*). This has facilitated the construction of strain libraries with random indel mutations in the targeted genes. However, expression of the *ku* and *ligD* genes may not yield successful indel mutations in all bacterial strains, as demonstrated for *E. coli* (*368*). Overexpression of *ku* and *ligD* is cytotoxic in certain bacteria (*356*) and could lead to off-target mutations due to repair of spontaneous double-stranded DNA breaks. Therefore, more studies attempting this strategy are necessary to understand its full potential for genome editing.

The most commonly used approach for CRISPR-based genome editing in bacteria is based on crRNA-guided cleavage activity of SpyCas9 to generate DSBs combined with homologous recombination (HR). This strategy is easy to design, efficient, precise and has been used in a variety of organisms. However, the editing efficiencies of the SpyCas9-based tool can significantly and unpredictably vary amongst the same genomic targets (*496*), thereby justifying the need to develop different Cas9-based editing tools to improve the predictability of target selection. Furthermore, many other Cas nucleases, usually of comparable or smaller size than SpyCas9, have recently been characterised (*82*, *83*, *248*, *329*, *419*). The use of Cas nucleases other than CRISPR–Cas9 will be an important future research direction, as it will broaden the range of biotechnologically relevant microorganisms in which CRISPR–Cas can be used for genome editing. These novel variants can be used to: i) extend the number of available targets in a genome (via recognition of alternative protospacer-adjacent motifs (PAMs) as described in **Chapter 7**), ii) to increase targeting specificities, iii) to avoid cytotoxicity issues observed for SpyCas9 in certain bacteria iv) extend the range of species that can be edited using Cas9, for example by the thermo-tolerant Cas9 orthologue ThermoCas9 described in **Chapter 6**.

ThermoCas9- and Cas12a-based genome editing methods were developed for two non-model microorganisms: *Flavobacterium* sp. 98 and microalgae *Nannochloropsis oceanica* IMET1. The details of these tools will be discussed below.

Homologous Recombination-SpyCas9 mediated genome editing of *Flavobacterium* sp. 98

Disease suppressive soils rhizosphere bacteria act as first line of defence against pathogens (*497–500*). If the pathogen breaks through this first line of defence, it will encounter the basal and induced defence mechanisms of the plant (*501*). During the second stage of pathogen invasion of the roots, the endophytic microbiome can provide additional protection (Victor et al, submitted). Upon pathogen invasion, members of *Chitinophagaceae* and *Flavobacteriaceae* families are enriched in the endosphere and specific functions in these bacterial families becomes activated, in particular specific enzymatic activities associated with fungal cell wall degradation as well as different classes of secondary metabolites encoded by Nonribosomal Peptide Synthetases (NRPSs) and Polyketide synthases (PKSs) (Victor *et al*, 2019 submitted). *In vitro* assays have demonstrated that degradation products of chitin, in particular GlcNAc, triggers the expression of Biosynthetic Gene Cluster-298 (BGC-298) in *Flavobacterium* spp. BGC-298 expression coincides with enhanced gliding motility and containment of hyphal growth of the fungal pathogen *Rhizoctonia solanii*. Therefore, a *Flavobacterium* spp. BGC-298 mutant would help to identify and confirm its exact role in disease suppression. However, to generate mutant *Flavobacterium* strains through homologous recombination is very time-consuming and inefficient. We therefore set out to develop a CRISPR-SpyCas9 mediated genome-editing tool for generating mutants in *Flavobacterium* sp. 98.

Expression of SpyCas9 has been reported to be associated with toxicity in certain microbial species (*306*, *354*, *355*). To assess potential toxic effects of SpyCas9 expression in *Flavobacterium* sp. 98, we constructed the pSpyCas9Fb_NT control plasmid by cloning the *spyCas9* gene and a sgRNA expression locus with a non-targeting (NT) spacer in the *E. coli-Flavobacterium* shuttle vector pCP11. Since the GC-content and codon usage landscapes of *E. coli* and *Flavobacterium* are comparable, an *E. coli* codon-optimised *spyCas9* gene was employed to ensure proper expression in *Flavobacterium* sp. 98. The expression of the *spyCas9* gene and the sgRNA was put under the transcriptional control of the P_{ompA} promoter from *Flavobacterium johnsoniae* and the P_{HU} promoter from *Flavobacterium hibernum*, respectively. The pSpyCas9Fb_NT plasmids and empty pCP11 plasmids (control) were individually electroporated into the same batch of competent cells. The transformation efficiency for the cells harbouring the pSpyCas9Fb_NT plasmid was marginally lower compared to the cells harbouring the pCP11 plasmid (**Figure 10.5A**), indicating that the expression of SpyCas9 is not toxic for the *Flavobacterium* cells. The slight decline in the transformation efficiency is most likely due to almost 1.5 times the size of the former plasmid (9426 bp) compared to the latter (14005 bp).

Next, we developed a SpyCas9-mediated targeting system for efficient introduction of lethal DSBs in *Flavobacterium* sp. 98. Targeting plasmids (pSpyCas9Fb_Sp1-3), similar to the pSpyCas9Fb_NT plasmid were assembled. These plasmids encode sgRNAs with unique spacer sequences that corresponds to different target sequence within the type I PKS module in BGC-298 of the *Flavobacterium* sp. 98 genome. The pSpyCas9Fb series of plasmids were electroporated in *Flavobacterium* sp. 98 (**Figure 10.5B**). The transformation efficiency of the cells harbouring the targeting pSpyCas9Fb_Sp1-3 plasmids reduced drastically as compared to cells harbouring the pSpyCas9Fb_NT control plasmid. This shows that the designed SpyCas9-mediated targeting system is highly active in *Flavobacterium* sp. 98.

Previous studies have reported homologous recombination in *Flavobacterium columnare* C#2 *and Flavobacterium psychrophilum* THC02/90 (*502*, *503*), albeit with very low efficiency and

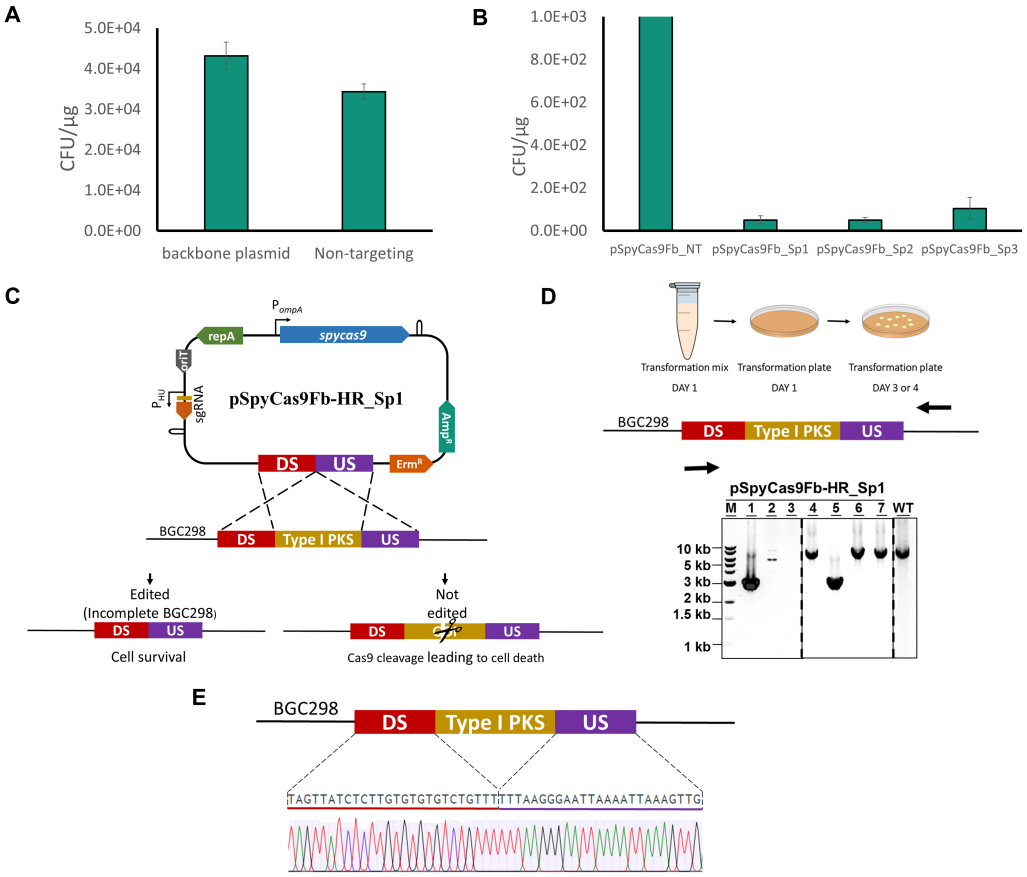


Figure 10.5 | Deletion of type I PKS module of BGC-298 in *Flavobacterium* sp. 98 by HR-SpyCas9 mediated genome editing(A) Results of the targeting experiment showing the transformation efficiency of the *Flavobacterium sp. 98* cells transformed with pSpyCas9Fb_NT compared to an empty pCP11 backbone plasmid. Average values of three replicates are shown, with error bars representing standard error. **(B)** Results of the targeting experiment showing the transformation efficiency of the *Flavobacterium* sp. 98 cells transformed with the targeting plasmids harbouring different spacers (pSpyCas9Fb_Sp1-3) compared to the non-targeting plasmid (pSpyCas9Fb_NT). Average values of three replicates are shown, with error bars representing standard error. **(C)** An overview of the editing plasmid and the Cas9 counter selection system. The editing plasmid contains *E. coli* codon optimised *SpyCas9* gene under the transcription regulation of constitutive promoter P_{ompA}, followed by the sgRNA module containing spacer 1 which targets the type I PKS module-BGC-298 of *Flavobacterium* sp. 98. The editing template includes 1.5 kb each of the upstream and downstream regions flanking the gene of type I PKS module (5 kb) of BGC-298. **(D)** Agarose gel electrophoresis showing the results from PCR on the colonies (1-7) of *Flavobacterium* sp. 98 cells transformed with pSpyCas9-HR_Sp1. A band at 3 kb indicates the presence of the correct mutant corresponding to the deleted 5 kb gene of type I PKS module. The last lane is the negative (wild type) control that corresponds to an 8 kb long DNA fragment. **(E)** Representative image of the sequence verification of the PCR product of the isolated genomic DNA from the type I PKS gene deletion mutant by Sanger sequencing.

here we developed a SpyCas9 targeting system in *Flavobacterium* sp. 98. Therefore, by coupling HR and SpyCas9 mediated counter selection, we set out to develop a HR-Cas9 system for the deletion of the type I PKS module in BGC-298 in *Flavobacterium* sp. 98. To this end, an editing plasmid (pSpyCas9Fb-HR_Sp1) was constructed and tested. It contains the previously described spacer 1 for efficient type I PKS module targeting and HR template composed

of 1500 bp each of the upstream and downstream genomic regions flanking the selected for deletion type I PKS fragment (**Figure 10.5C**). Two control plasmids (pSyCas9Fb-HR_NT and pCP11_w/o-Cas9_HR), one containing the same HR templates as the editing plasmid but a non-targeting spacer and another without the *SpyCas9* gene were also transformed to assess the contribution of SpyCas9 to the HR. Upon transformation of the above described constructs in *Flavobacterium* sp. 98, a drop was observed in the number of surviving colonies harbouring the editing plasmids compared to the control plasmids. We randomly picked and genotyped the colonies from the transformation plate containing cells harbouring the editing plasmid through colony PCR. Two out of the six tested colonies were confirmed to be the desired clean type I PKS deletion mutants (**Figure 10.5D**). The correct mutants were re-streaked on selection medium plates and the genomic DNA of single colonies was isolated. The gene deletion was further verified by Sanger sequencing using genome specific primers (**Figure 10.5E**). Meanwhile, none of the selected colonies from the transformation plates with control plasmids were deletion mutants, underlining the significant contribution of SpyCas9 targeting to the efficiency of the developed HR-SpyCas9 counter selection tool in *Flavobacterium* sp. 98.

By coupling the efficient SpyCas9 endonuclease targeting activity to the native homologous recombination mechanisms of *Flavobacterium* sp. 98 we have developed, to the best of our knowledge, the first Cas9-based genome editing method for *Flavobacterium.* Preliminary results of two independent BGC-298 mutants tested in a disease bioassay revealed that the mutants lost the ability to protect the plant as compared with the wild type. Furthermore, the consortia containing the mutants also showed an increase in the disease incidence at the same level as the control inoculated with the pathogen. Metabolomics analysis are currently being performed to investigate the compounds produced by BGC-298 that is responsible for defence against the pathogen. The here-developed method is potentially readily applicable to many other members of the genus *Flavobacterium* with a limited genetic toolbox and can be used to substantially accelerate fundamental and applied *Flavobacterium* studies.

Cas ribonucleoprotein complex mediated homology-directed repair in microalgae *Nannochloropsis oceanica* IMET1

Microalgae hold great potential as sustainable feedstock to produce biofuels and other value-added products. The *Nannochloropsis* spp. stands out from other microalgal species due to their capability to accumulate both triacylglycerol (TAG) and polyunsaturated fatty acids (PUFAs). However, the high production cost of TAGs and PUFAs in *Nannochloropsis spp.* is not competitive to the traditional fossil fuel-based industries and hinder the commercialization of microalgae-derived products. Efficient genome editing techniques would allow effective metabolic engineering which subsequently could reduce the production costs.

Efficient HR and antibiotic resistance-based selection to genetic mutants of *Nannochloropsis* spp have been reported (*504*). However, we were unable to generate any clean mutants by this approach. CRISPR-based genome editing techniques have been used in *Nannochloropsis* spp. (*505–508*). All reported CRISPR-based systems for this species rely on plasmid-borne constitutive expression of Cas9 and the sgRNA, combined with repair of DSB by NHEJ disruption of the target genes. We hypothesize that it would be advantageous to use Cas ribonucleoprotein (RNP) transfection methods. The advantages of using the Cas-RNP delivery approach include: (i) possibility to assay for the activity of RNP complex *in vitro* prior to tedious *in vivo* experiments, (ii) the dispensability of a codon harmonized Cas gene, as well as promoters and

terminators that are active in host of interest, and (iii) reduced off-target and cytotoxic effects as, during RNP transfection, the Cas nucleases are saturated with sgRNAs and is only transiently active before it is degraded by endogenous proteases in cells.

Therefore, we have developed an alternative approach for CRISPR-based genome editing by transfecting Cas RNPs and HR template via electroporation *in Nannochloropsis oceanica* IMET1. As a proof of concept, we targeted the *nitrate reductase* (*NR*) gene as NR mutants can be easily characterised based on their ability to grow on media containing nitrite or ammonia as the sole nitrogen source (*504*, *505*, *508*). We compared the efficiency of HR for four different Cas nucleases: SpyCas9, AsCas12a, FnCas12a and LbCas12a.

Upon transformation and screening of *N. oceanica* IMET1 cells with the SpyCas9 RNPs targeting the *NR* gene along with the HR editing template, up to 74% of the screened colonies were of the desired mutant genotype. For the Cas12a RNPs, we observed the highest genome editing efficiency upon using the FnCas12a RNP (up to 93% of screened cells were of desired mutant genotype), followed by LbCas12a (up to 78% of screened cells were of desired mutant genotype). For the cells transformed with AsCas12a RNPs, the genome editing was highly inefficient (only 4% of screened cells were of desired mutant genotype). The diminished activity of AsCas12a is likely due to the reduced *in vivo* activity of AsCas12a at lower temperatures such as 28 ℃ which is the growth temperature for microalgae. The absence of mutant colonies previously when only HR was used as compared to when HR in combination with Cas9 was used underscores the importance of DSB generated by the Cas nucleases for efficient HR. By assessing the efficiency of HR stimulated by different Cas nucleases, we substantially enhance the possibility to generate desired targeted mutants in *Nannochloropsis* sp. The same strategy could be used in the future to generate precise gene insertion and/or deletion of metabolically relevant genes. The future prospects include generation of marker-less mutants and multiplexing genome editing in *Nannochloropsis* spp.

Alternative CRISPR-based tools for bacterial genome editing

In most bacteria, cleavage in the genome by Cas nucleases results in cell death due to a lack of the NHEJ pathway (*368*, *371*). Therefore, CRISPR–Cas has been employed as a counter-selection tool to eliminate non-modified cells from a mixed population for enrichment of modified cells after homologous recombination (*392*, *397*). However, the efficiency of the CRISPR-HR system varies depending on the host cells. Consequently, other creative techniques to introduce changes in bacterial genomes have recently been developed. One such interesting alternative for bacterial genome editing is the development of modified Cas nucleases called base editors. Deaminase-mediated targeted base editing can be used to edit nucleotides at the target locus without DNA cleavage as well as the template donor DNA. As it employs DNA deamination in place of nuclease-mediated DNA cleavage, it may not induce cell death in bacteria. Base editors typically comprise fusions of dCas9 or Cas9 nickase variant and a cytidine deaminase domain. When fused to Cas9, the cytidine deaminase domain can convert cytidines to uracils on the non-target strand. Thereby, it can be used to introduce specific base mutations in a defined window adjacent to the PAM (*63*, *509*, *510*). Although dispensable, fusing an uracil DNA glycosylase inhibitor to the base editor has been shown to improve editing efficiencies by inhibiting uracil removal upon editing (*298*, *420*). To date, base editors have been mainly employed to generate point mutations or to insert premature stop codons, eliminating the need for homologous recombination templates carrying the desired mutations, as reported in *E. coli*

(*420*), *Klebsiella pneumoniae* (*511*), *Clostridium beijerinckii* (*512*) and *Pseudomonas aeruginosa* (*294*). More work is needed to assess cytotoxicity of these systems (if any), to advance the use of these base editors in other bacterial strains.

In another developed system, a nickase variant of Cas9 was fused to an error-prone polymerase to create EvolvR. This system allows for rapid diversification of the DNA sequence within a few hundred base-pairs window, which can be used for powerful functional screening platforms. A wide array of future applications for EvolvR can be envisaged, from evolving new protein activities to studying cellular diversification in response to exogenous stimuli.

Alternatively, endogenous Cas nucleases should be considered as a simple and powerful substitute to current strategies of CRISPR-based editing for strains harbouring native CRISPR–Cas systems (*513*). This may enable efficient selection without the need to heterologously express a Cas nuclease (*514*–*516*). Utilizing endogenous CRISPR–Cas systems, however, can be time-consuming, as it must be sufficiently characterised before use, including identifying PAMs and ensuring that the *cas* genes are actively expressed. Nonetheless, utilizing them for subsequent genome editing could potentially reduce cytotoxicity and simplify the editing process by excluding the expression of a heterologous Cas nuclease.

Genome editing of industrial bacteria is critical for strain improvement and for metabolic engineering. Several methods for CRISPR-based genome editing have been implemented in different bacterial species. The vast potential of CRISPR-based editing has accelerated development of genome and metabolic engineering studies on model as well as in more diverse, non-model bacteria. Finally, research that enhances CRISPR-based bacterial genome editing should be pursued. This will facilitate the application of such CRISPR-based tools to a wider range of non-model industrial bacteria to help the transition to green and sustainable bio-based economy.

Future perspectives

The rapidly expanding database of genomic sequences, which powered the initial and ongoing discovery of CRISPR–Cas systems, holds many hidden treasures. Exploration of the diversity of the prokaryotic universe will undoubtedly reveal new CRISPR–Cas systems as well as other defence systems that could facilitate subsequent development of novel molecular biology technologies. Our understanding of the fundamental mechanisms of CRISPR–Cas systems has increased tremendously over the past few years and will undoubtedly continue to do so. This is illustrated by the recent reports of not only novel Cas nucleases with unique mechanistic features (*74*, *76*, *77*, *80*), such as single-subunit RNA targeting systems, but also the discovery of non-canonical facets of CRISPR biology, such as CRISPR–Cas mediated offence and CRISPR–Cas mediated gene transposition (*81*, *468*). The discovery, functional and structural characterisation of several anti-CRISPR proteins against different type of CRISPR systems further underscore the diversity of the bacterium-phage defence arsenal. Further insights into the mechanistic variations of the novel 'ancestral' systems which likely evolved to functional Class 2 Cas nucleases, such as the type V-U or TnpB-like proteins, would shed light on the evolution CRISPR–Cas systems from mobile genetic elements to an adaptive immunity system.

In addition to the foundational research to unravel the molecular details of the elegant CRISPR–Cas systems, a huge stockpile of CRISPR-based tools have been adopted for powerful molecular technologies. Apart from exploration of natural Cas diversity, structure-guided engineering of Cas nuclease function is being used to expand and enhance CRISPR-based tools. RNA-guided

DNA and RNA cleavage through the nuclease activities of Cas proteins is only one-way CRISPR technology can be applied. The class 2 nucleases, such as Cas9 and Cas12a, can also be inactivated to turn them into RNA-guided DNA- or RNA-binding domains and repurposed for programmable transcriptional silencing as well as nucleic acid binding scaffolds for the recruitment of a variety of functional enzymes. Such scaffolds have been used for the development of programmable DNA and RNA base editors, which can achieve the precise conversion of one base to another. Base editors are particularly promising for therapeutics as they can be used to make targeted changes without the need for DNA cleavage or homologous recombination. Finally, it is anticipated that further unravelling of the CRISPR–Cas function, mechanism, and structures will uncover novel Cas nucleases with versatile functionalities that can be harnessed as efficient tools for nucleic acid manipulation.

Bibliography

1. C. A. Suttle, Viruses in the sea. *Nature* **437**, 356–361, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature04160 (Sept. 2005) (cit. on p. 3).
2. E. V. Koonin, V. V. Dolja, A virocentric perspective on the evolution of life. *Curr Opin Virol* **3**, 546–557, ISSN: 1879-6265 (Electronic) 1879-6257 (Linking), DOI 10.1016/j.coviro.2013.06.008 (Oct. 2013) (cit. on pp. 3, 19).
3. S. Gandon, Why Be Temperate: Lessons from Bacteriophage λ. *Trends Microbiol* **24**, 356–365, ISSN: 1878-4380 (Electronic) 0966-842X (Linking), DOI 10.1016/j.tim.2016.02.008 (May 2016) (cit. on p. 3).
4. J. A. Fuhrman, Marine viruses and their biogeochemical and ecological effects. *Nature* **399**, 541–548, ISSN: 0028-0836, DOI Doi10.1038/21119 (1999) (cit. on p. 3).
5. C. A. Suttle, Marine viruses — major players in the global ecosystem. *Nature Reviews Microbiology* **5**, 801–812, ISSN: 1740-1526, DOI 10.1038/nrmicro1750 (Oct. 2007) (cit. on p. 3).
6. H. Brussow, C. Canchaya, W.-D. Hardt, Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiology and Molecular Biology Reviews* **68**, 560–602, ISSN: 1092-2172, DOI 10.1128/mmbr.68.3.560-602.2004 (Sept. 2004) (cit. on p. 3).
7. A. M. Comeau, H. M. Krisch, War is peace — dispatches from the bacterial and phage killing fields. *Current Opinion in Microbiology* **8**, 488–494, ISSN: 1369-5274, DOI 10.1016/j.mib.2005.06.004 (Aug. 2005) (cit. on p. 3).
8. C. Venditti, A. Meade, M. Pagel, Phylogenies reveal new interpretation of speciation and the Red Queen. *Nature* **463**, 349–352, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature08630 (Dec. 2010) (cit. on p. 3).
9. A. F. Andersson, J. F. Banfield, Virus Population Dynamics and Acquired Virus Resistance in Natural Microbial Communities. *Science* **320**, 1047–1050, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.1157358 (May 2008) (cit. on p. 3).
10. N. L. Held, R. J. Whitaker, Viral biogeography revealed by signatures in Sulfolobus islandicus genomes. *Environ Microbiol* **11**, 457–466, ISSN: 1462-2920 (Electronic) 1462-2912 (Linking), DOI 10.1111/j.1462-2920.2008.01784.x (Feb. 2009) (cit. on p. 3).
11. J. F. Heidelberg, W. C. Nelson, T. Schoenfeld, D. Bhaya, Germ Warfare in a Microbial Mat Community: CRISPRs Provide Insights into the Co-Evolution of Host and Viral Genomes. *Plos One* **4**, ed. by N. Ahmed, e4169, ISSN: 1932-6203, DOI 10.1371/journal.pone.0004169 (Jan. 2009) (cit. on p. 3).
12. L. G. Guidotti, F. V. Chisari, Noncytolytic control of viral infections by the innate and adaptive immune response. *Annual Review of Immunology* **19**, 65–91, ISSN: 0732-0582, DOI DOI10.1146/annurev.immunol.19.1.65 (2001) (cit. on pp. 3, 6).
13. L. Carroll, *Through the looking-glass and what Alice found there* (Chicago : W.B. Conkey Co., [1900] ©1900, 1900) (cit. on p. 4).
14. M.-C. Chopin, A. Chopin, E. Bidnenko, Phage abortive infection in lactococci: variations on a theme. *Current Opinion in Microbiology* **8**, 473–479, ISSN: 1369-5274, DOI 10.1016/j.mib.2005.06.006 (Aug. 2005) (cit. on pp. 4, 19).
15. K. S. Makarova, Y. I. Wolf, E. V. Koonin, Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Research* **41**, 4360–4377, ISSN: 0305-1048, DOI 10.1093/nar/gkt157 (Mar. 2013) (cit. on pp. 4, 19).
16. J. E. Samson, A. H. Magadán, M. Sabri, S. Moineau, Revenge of the phages: defeating bacterial defences. *Nature Reviews Microbiology* **11**, 675–687, ISSN: 1740-1526, DOI 10.1038/nrmicro3096 (Aug. 2013) (cit. on pp. 4, 9, 10, 19, 33).
17. T. Goldfarb, H. Sberro, E. Weinstock, *et al.*, BREX is a novel phage resistance system widespread in microbial genomes. *Embo Journal* **34**, 169–183, ISSN: 0261-4189, DOI 10.15252/embj.201489455 (Dec. 2015) (cit. on pp. 4, 19).
18. D. C. Swarts, M. M. Jore, E. R. Westra, *et al.*, DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* **507**, 258–261, ISSN: 0028-0836, DOI 10.1038/nature12971 (Feb. 2014) (cit. on pp. 4, 19).
19. G. Ofir, S. Melamed, H. Sberro, *et al.*, DISARM is a widespread bacterial defence system with broad anti-phage activities. *Nature Microbiology* **3**, 90–98, ISSN: 2058-5276, DOI 10.1038/s41564-017-0051-0 (Oct. 2018) (cit. on p. 4).

20. R. Barrangou, C. Fremaux, H. Deveau, *et al.*, CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315**, 1709–1712, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.1138140 (Mar. 2007) (cit. on pp. 4, 6, 19, 23, 79, 111, 258).
21. R. Jansen, J. D. A. van Embden, W. Gaastra, L. M. Schouls, Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**, 1565–1575, ISSN: 0950-382X (Print) 0950-382X (Linking), DOI 10.1046/j.1365-2958.2002.02839.x (Mar. 2002) (cit. on pp. 4, 6, 19).
22. A. Stern, L. Keren, O. Wurtzel, G. Amitai, R. Sorek, Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet* **26**, 335–340, ISSN: 0168-9525 (Print) 0168-9525 (Linking), DOI 10.1016/j.tig.2010.05.008 (Aug. 2010) (cit. on pp. 5, 23).
23. C. Hale, S. Majumdar, J. Elmore, *et al.*, Essential Features and Rational Design of CRISPR RNAs that Function with the Cas RAMP Module Complex to Cleave RNAs. *Mol Cell* **45**, 292–302, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2011.10.023 (Feb. 2012) (cit. on p. 5).
24. R. K. Lillestøl, S. A. Shah, K. Brügger, *et al.*, CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. *Mol Microbiol* **72**, 259–272, ISSN: 1365-2958 (Electronic) 0950-382X (Linking), DOI 10.1111/j.1365-2958.2009.06641.x (Mar. 2009) (cit. on p. 5).
25. K. Pougach, E. Semenova, E. Bogdanova, *et al.*, Transcription, processing and function of CRISPR cassettes in Escherichia coli. *Mol Microbiol* **77**, 1367–1379, ISSN: 1365-2958 (Electronic) 0950-382X (Linking), DOI 10.1111/j.1365-2958.2010.07265.x (Sept. 2010) (cit. on p. 5).
26. Ü. Pul, R. Wurm, Z. Arslan, *et al.*, Identification and characterisation ofE. coliCRISPR-caspromoters and their silencing by H-NS. *Mol Microbiol* **75**, 1495–1512, ISSN: 1365-2958 (Electronic) 0950-382X (Linking), DOI 10.1111/j.1365-2958.2010.07073.x (Mar. 2010) (cit. on p. 5).
27. J. McGinn, L. Marraffini, CRISPR-Cas Systems Optimize Their Immune Response by Specifying the Site of Spacer Integration. *Molecular Cell* **64**, 616–623, ISSN: 1097-2765, DOI 10.1016/j.molcel.2016.08.038 (Nov. 2016) (cit. on pp. 5, 259).
28. J. van der Oost, M. M. Jore, E. R. Westra, M. Lundgren, S. J. Brouns, CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends in Biochemical Sciences* **34**, 401–407, ISSN: 0968-0004, DOI 10.1016/j.tibs.2009.05.002 (Aug. 2009) (cit. on pp. 5, 6, 19, 24, 26).
29. P. Mohanraju, K. S. Makarova, B. Zetsche, *et al.*, Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* **353**, aad5147, ISSN: 0036-8075, DOI 10.1126/science.aad5147 (Aug. 2016) (cit. on pp. 5, 39, 79, 111, 213).
30. I. Grissa, G. Vergnaud, C. Pourcel, The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**, 172, ISSN: 1471-2105 (Electronic) 1471-2105 (Linking), DOI 10.1186/1471-2105-8-172 (2007) (cit. on pp. 5, 260).
31. K. S. Makarova, Y. I. Wolf, O. S. Alkhnbashi, *et al.*, An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology* **13**, 722–736, ISSN: 1740-1526, DOI 10.1038/nrmicro3569 (Sept. 2015) (cit. on pp. 5, 20–22, 24, 27, 39, 79, 121, 258, 260).
32. D. Burstein, C. L. Sun, C. T. Brown, *et al.*, Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nature Communications* **7**, ISSN: 2041-1723, DOI Artn1061310.1038/Ncomms10613 (2016) (cit. on p. 5).
33. E. V. Koonin, K. S. Makarova, Y. I. Wolf, Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annu Rev Microbiol* **71**, 233–261, ISSN: 1545-3251 (Electronic) 0066-4227 (Linking), DOI 10.1146/annurev-micro-090816-093830 (Sept. 2017) (cit. on p. 5).
34. Y. Ishino, H. Shinagawa, K. Makino, M. Amemura, A. Nakata, Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of Bacteriology* **169**, 5429–5433, ISSN: 0021-9193, DOI 10.1128/jb.169.12.5429-5433.1987 (Dec. 1987) (cit. on p. 6).
35. K. S. Makarova, A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res* **30**, 482–496, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/30.2.482 (Jan. 2002) (cit. on p. 6).
36. F. J. Mojica, C. sar Díez-Villaseñor, J. García-Martínez, E. Soria, Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *J Mol Evol* **60**, 174–82, ISSN: 0022-2844 (Print) 0022-2844 (Linking), DOI 10.1007/s00239-004-0046-3 (Feb. 2005) (cit. on pp. 6, 19).
37. 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, ISSN: 1350-0872 (Print) 1350-0872 (Linking), DOI 10.1099/mic.0.27437-0 (Mar. 2005) (cit. on pp. 6, 8, 19).
38. A. Bolotin, B. Quinquis, A. Sorokin, S. D. Ehrlich, Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561, ISSN: 1350-0872 (Print) 1350-0872 (Linking), DOI 10.1099/mic.0.28048-0 (Aug. 2005) (cit. on pp. 6, 19).
39. 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 analo-

gies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* **1**, 7, ISSN: 1745-6150 (Electronic) 1745-6150 (Linking), DOI 10.1186/1745-6150-1-7 (2006) (cit. on pp. 6, 19, 26, 29).

40. S. J. J. Brouns, M. M. Jore, M. Lundgren, *et al.*, Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science* **321**, 960–964, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.1159689 (Aug. 2008) (cit. on pp. 6, 19, 26, 27, 29, 111, 260, 261).
41. T. Sinkunas, G. Gasiunas, C. Fremaux, *et al.*, Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *Embo Journal* **30**, 1335–1342, ISSN: 0261-4189, DOI 10.1038/emboj.2011.41 (Feb. 2011) (cit. on p. 6).
42. E. Westra, P. van Erp, T. Künne, *et al.*, CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. *Molecular Cell* **46**, 595–605, ISSN: 1097-2765, DOI 10.1016/j.molcel.2012.03.018 (June 2012) (cit. on pp. 6, 121).
43. L. A. Marraffini, E. J. Sontheimer, CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. *Science* **322**, 1843–1845, ISSN: 0036-8075, DOI 10.1126/science.1165771 (Dec. 2008) (cit. on pp. 7, 9, 19, 29).
44. C. R. Hale, P. Zhao, S. Olson, *et al.*, RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell* **139**, 945–956, ISSN: 0092-8674, DOI 10.1016/j.cell.2009.07.040 (Nov. 2009) (cit. on pp. 7, 19, 29).
45. F. J. M. Mojica, Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740, ISSN: 1350-0872, DOI 10.1099/mic.0.023960-0 (Mar. 2009) (cit. on pp. 7, 23, 82, 111, 147, 175).
46. H. Deveau, R. Barrangou, J. E. Garneau, *et al.*, Phage Response to CRISPR-Encoded Resistance in Streptococcus thermophilus. *Journal of Bacteriology* **190**, 1390–1400, ISSN: 0021-9193, DOI 10.1128/Jb.01412-07 (Dec. 2008) (cit. on pp. 7, 10, 23, 33, 111, 175).
47. D. Gleditzsch, P. Pausch, H. Müller-Esparza, *et al.*, PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *Rna Biology* **16**, 504–517, ISSN: 1547-6286, DOI 10.1080/15476286.2018.1504546 (Sept. 2019) (cit. on p. 7).
48. J. E. Garneau, M.-È. Dupuis, M. Villion, *et al.*, The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71, ISSN: 0028-0836, DOI 10.1038/nature09523 (Nov. 2010) (cit. on pp. 7, 9, 19, 23, 29, 31).
49. E. Deltcheva, K. Chylinski, C. M. Sharma, *et al.*, CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature09886 (Mar. 2011) (cit. on pp. 7, 21, 27, 30, 39, 147, 175, 261).
50. R. Sapranauskas, G. Gasiunas, C. Fremaux, *et al.*, The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Research* **39**, 9275–9282, ISSN: 0305-1048, DOI 10.1093/nar/gkr606 (Aug. 2011) (cit. on p. 7).
51. M. Jinek, K. Chylinski, I. Fonfara, *et al.*, A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821, ISSN: 0036-8075, DOI 10.1126/science.1225829 (June 2012) (cit. on pp. 7, 21, 27, 39, 79, 82, 111, 115, 122, 147, 175, 216, 221, 261, 263, 267, 268).
52. G. Gasiunas, R. Barrangou, P. Horvath, V. Siksnys, Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E2579–E2586, ISSN: 0027-8424, DOI 10.1073/pnas.1208507109 (Sept. 2012) (cit. on pp. 7, 27, 79, 111, 115, 261, 263).
53. P. Horvath, D. A. Romero, A.-C. Coute-Monvoisin, *et al.*, Diversity, Activity, and Evolution of CRISPR Loci in Streptococcus thermophilus. *Journal of Bacteriology* **190**, 1401–1412, ISSN: 0021-9193, DOI 10.1128/jb.01415-07 (Dec. 2008) (cit. on pp. 8, 147).
54. P. Horvath, A.-C. Coûté-Monvoisin, D. A. Romero, *et al.*, Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *International Journal of Food Microbiology* **131**, 62–70, ISSN: 0168-1605, DOI 10.1016/j.ijfoodmicro.2008.05.030 (Apr. 2009) (cit. on p. 8).
55. A. Quiberoni, S. Moineau, G. M. Rousseau, J. Reinheimer, H.-W. Ackermann, Streptococcus thermophilus bacteriophages. *International Dairy Journal* **20**, 657–664, ISSN: 0958-6946, DOI 10.1016/j.idairyj.2010.03.012 (Oct. 2010) (cit. on p. 9).
56. R. Sorek, V. Kunin, P. Hugenholtz, CRISPR — a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Reviews Microbiology* **6**, 181–186, ISSN: 1740-1526, DOI 10.1038/nrmicro1793 (Mar. 2008) (cit. on p. 9).
57. L. Cong, F. A. Ran, D. Cox, *et al.*, Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–823, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.1231143 (Jan. 2013) (cit. on pp. 9, 33, 267, 268).
58. P. Mali, L. Yang, K. M. Esvelt, *et al.*, RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823–826, ISSN: 0036-8075, DOI 10.1126/science.1232033 (Jan. 2013) (cit. on pp. 9, 33, 219, 224, 267, 268).

59. S. W. Cho, S. Kim, Y. Kim, *et al.*, Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* **24**, 132–141, ISSN: 1549-5469 (Electronic) 1088-9051 (Linking), DOI 10.1101/gr.162339.113 (Nov. 2014) (cit. on pp. 9, 267).
60. W. Y. Hwang, Y. Fu, D. Reyon, *et al.*, Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* **31**, 227–229, ISSN: 1546-1696 (Electronic) 1087-0156 (Linking), DOI 10.1038/nbt.2501 (Jan. 2013) (cit. on pp. 9, 267).
61. L. Qi, M. Larson, L. Gilbert, *et al.*, Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking), DOI 10.1016/j.cell.2013.02.022 (Feb. 2013) (cit. on pp. 9, 33, 111, 267).
62. F. A. Ran, P. D. Hsu, J. Wright, *et al.*, Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281–2308, ISSN: 1750-2799 (Electronic) 1750-2799 (Linking), DOI 10.1038/nprot.2013.143 (Oct. 2013) (cit. on p. 9).
63. N. M. Gaudelli, A. C. Komor, H. A. Rees, *et al.*, Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature24644 (Oct. 2017) (cit. on pp. 9, 274).
64. J. S. Gootenberg, O. O. Abudayyeh, J. W. Lee, *et al.*, Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* **356**, 438–442, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.aam9321 (Apr. 2017) (cit. on p. 9).
65. S. R. Choudhury, Y. Cui, K. Lubecka, B. Stefanska, J. Irudayaraj, CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. *Oncotarget* **7**, 46545–46556, ISSN: 1949-2553, DOI 10.18632/oncotarget.10234 (June 2016) (cit. on p. 9).
66. B. Chen, L. Gilbert, B. Cimini, *et al.*, Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell* **155**, 1479–1491, ISSN: 0092-8674, DOI 10.1016/j.cell.2013.12.001 (Dec. 2013) (cit. on pp. 9, 33).
67. I. B. Hilton, A. M. D'Ippolito, C. M. Vockley, *et al.*, Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature Biotechnology* **33**, 510–517, ISSN: 1087-0156, DOI 10.1038/nbt.3199 (Apr. 2015) (cit. on pp. 9, 33, 267).
68. K. Nishida, T. Arazoe, N. Yachie, *et al.*, Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729–aaf8729, ISSN: 0036-8075, DOI 10.1126/science.aaf8729 (Aug. 2016) (cit. on p. 9).
69. T. Wang, J. J. Wei, D. M. Sabatini, E. S. Lander, Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science* **343**, 80–84, ISSN: 0036-8075, DOI 10.1126/science.1246981 (Dec. 2014) (cit. on pp. 9, 33).
70. O. Shalem, N. E. Sanjana, E. Hartenian, *et al.*, Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* **343**, 84–87, ISSN: 0036-8075, DOI 10.1126/science.1247005 (Dec. 2014) (cit. on pp. 9, 33).
71. P. I. Thakore, A. M. D'Ippolito, L. Song, *et al.*, Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nature Methods* **12**, 1143–1149, ISSN: 1548-7091, DOI 10.1038/nmeth.3630 (Oct. 2015) (cit. on pp. 9, 267).
72. 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 in Biotechnology* **34**, 575–587, ISSN: 0167-7799, DOI 10.1016/j.tibtech.2016.02.004 (July 2016) (cit. on pp. 9, 111, 112, 118, 121, 150, 175, 185).
73. D. Bikard, W. Jiang, P. Samai, *et al.*, Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Research* **41**, 7429–7437, ISSN: 0305-1048, DOI 10.1093/nar/gkt520 (June 2013) (cit. on pp. 9, 33, 111, 120, 267, 270).
74. D. Burstein, L. B. Harrington, S. C. Strutt, *et al.*, New CRISPR–Cas systems from uncultivated microbes. *Nature* **542**, 237–241, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature21059 (Dec. 2017) (cit. on pp. 9, 79, 264, 275).
75. A. A. Gogleva, M. S. Gelfand, I. I. Artamonova, Comparative analysis of CRISPR cassettes from the human gut metagenomic contigs. *BMC Genomics* **15**, 202, ISSN: 1471-2164 (Electronic) 1471-2164 (Linking), DOI 10.1186/1471-2164-15-202 (2014) (cit. on p. 9).
76. S. Shmakov, O. Abudayyeh, K. Makarova, *et al.*, Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol Cell* **60**, 385–397, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2015.10.008 (Nov. 2015) (cit. on pp. 9, 20–23, 27, 31, 79, 260, 263–265, 275).
77. L. B. Harrington, D. Burstein, J. S. Chen, *et al.*, Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science* **362**, 839–842, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.aav4294 (Oct. 2018) (cit. on pp. 9, 79, 264, 275).
78. S. Konermann, P. Lotfy, N. J. Brideau, *et al.*, Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* **173**, 665–676.e14, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking), DOI 10.1016/j.cell.2018.02.033 (Apr. 2018) (cit. on pp. 9, 86).

79. S. A. Shmakov, K. S. Makarova, Y. I. Wolf, K. V. Severinov, E. V. Koonin, Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proc Natl Acad Sci U S A* **115**, E5307–E5316, ISSN: 1091-6490 (Electronic) 0027-8424 (Linking), DOI 10.1073/pnas.1803440115 (May 2018) (cit. on p. 9).
80. W. X. Yan, P. Hunnewell, L. E. Alfonse, *et al.*, Functionally diverse type V CRISPR-Cas systems. *Science* **363**, 88–91, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.aav7271 (Dec. 2019) (cit. on pp. 9, 79, 81–83, 86, 264, 265, 275).
81. J. Strecker, A. Ladha, Z. Gardner, *et al.*, RNA-guided DNA insertion with CRISPR-associated transposases. *Science* **365**, 48–53, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.aax9181 (June 2019) (cit. on pp. 9, 79, 81, 260, 269, 275).
82. S. Shmakov, A. Smargon, D. Scott, *et al.*, Diversity and evolution of class 2 CRISPR–Cas systems. *Nat Rev Microbiol* **15**, 169–182, ISSN: 1740-1534 (Electronic) 1740-1526 (Linking), DOI 10.1038/nrmicro.2016.184 (Jan. 2017) (cit. on pp. 9, 39, 79, 81, 258–260, 263, 265, 270).
83. B. Zetsche, J. Gootenberg, O. Abudayyeh, *et al.*, Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163**, 759–771, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking), DOI 10.1016/j.cell.2015.09.038 (Oct. 2015) (cit. on pp. 9, 21, 27, 29, 31, 33, 34, 39, 61, 79, 82, 147, 155, 157, 261, 264, 269, 270).
84. F. Teng, T. Cui, G. Feng, *et al.*, Repurposing CRISPR-Cas12b for mammalian genome engineering. *Cell Discov* **4**, 63, ISSN: 2056-5968 (Print) 2056-5968 (Linking), DOI 10.1038/s41421-018-0069-3 (Nov. 2018) (cit. on pp. 9, 269).
85. L. Liu, X. Li, J. Wang, *et al.*, Two Distant Catalytic Sites Are Responsible for C2c2 RNase Activities. *Cell* **168**, 121–134.e12, ISSN: 0092-8674, DOI 10.1016/j.cell.2016.12.031 (Jan. 2017) (cit. on pp. 9, 263).
86. A. East-Seletsky, M. R. O'Connell, D. Burstein, G. J. Knott, J. A. Doudna, RNA Targeting by Functionally Orthogonal Type VI-A CRISPR-Cas Enzymes. *Mol Cell* **66**, 373–383.e3, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2017.04.008 (May 2017) (cit. on pp. 9, 263).
87. S. J. Labrie, J. E. Samson, S. Moineau, Bacteriophage resistance mechanisms. *Nat Rev Microbiol* **8**, 317–327, ISSN: 1740-1534 (Electronic) 1740-1526 (Linking), DOI 10.1038/nrmicro2315 (Mar. 2010) (cit. on pp. 9, 10).
88. I. G. Bogdarina, M. Reiter, D. H. Kruger, I. Bur'ianov Ia, A. A. Baev, [DNA methylation in T3 and T7 phages by DNA-adenine methylases of various types and methylase EcoK ocR+ by protein]. *Dokl Akad Nauk SSSR* **273**, 234–7, ISSN: 0002-3264 (Print) 0002-3264 (Linking) (1983) (cit. on p. 9).
89. M. Drozdz, A. Piekarowicz, J. M. Bujnicki, M. Radlinska, Novel non-specific DNA adenine methyltransferases. *Nucleic Acids Res* **40**, 2119–2130, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gkr1039 (Nov. 2012) (cit. on p. 9).
90. R. Kahmann, in *Current Topics in Microbiology and Immunology* (Springer Berlin Heidelberg, 1984), vol. 108, pp. 29–47, DOI 10.1007/978-3-642-69370-0_4 (cit. on p. 9).
91. G. A. Roberts, A. S. Stephanou, N. Kanwar, *et al.*, Exploring the DNA mimicry of the Ocr protein of phage T7. *Nucleic Acids Research* **40**, 8129–8143, ISSN: 0305-1048, DOI 10.1093/nar/gks516 (June 2012) (cit. on p. 9).
92. F. W. Studier, N. R. Movva, Samase Gene of Bacteriophage-T3 Is Responsible for Overcoming Host Restriction. *Journal of Virology* **19**, 136–145, ISSN: 0022-538X (1976) (cit. on p. 10).
93. C. L. Sun, R. Barrangou, B. C. Thomas, *et al.*, Phage mutations in response to CRISPR diversification in a bacterial population. *Environ Microbiol* **15**, 463–470, ISSN: 1462-2920 (Electronic) 1462-2912 (Linking), DOI 10.1111/j.1462-2920.2012.02879.x (Oct. 2013) (cit. on p. 10).
94. J. Bondy-Denomy, B. Garcia, S. Strum, *et al.*, Multiple mechanisms for CRISPR–Cas inhibition by anti-CRISPR proteins. *Nature* **526**, 136–139, ISSN: 0028-0836 1476-4687, DOI 10.1038/nature15254 (Sept. 2015) (cit. on pp. 10, 33, 176).
95. E. Semenova, M. M. Jore, K. A. Datsenko, *et al.*, 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, ISSN: 1091-6490 (Electronic) 0027-8424 (Linking), DOI 10.1073/pnas.1104144108 (June 2011) (cit. on pp. 10, 29, 86).
96. K. D. Seed, D. W. Lazinski, S. B. Calderwood, A. Camilli, A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494**, 489–491, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature11927 (Feb. 2013) (cit. on pp. 10, 33).
97. J. Bondy-Denomy, A. Pawluk, K. L. Maxwell, A. R. Davidson, Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432, ISSN: 0028-0836, DOI 10.1038/nature11723 (Dec. 2013) (cit. on pp. 10, 176).
98. A. Pawluk, J. Bondy-Denomy, V. H. W. Cheung, K. L. Maxwell, A. R. Davidson, A New Group of Phage Anti-CRISPR Genes Inhibits the Type I-E CRISPR-Cas System of Pseudomonas aeruginosa. *MBio* **5**, ed. by R. Hendrix, e00896, ISSN: 2150-7511 (Electronic), DOI 10.1128/mbio.00896-14 (Apr. 2014) (cit. on pp. 10, 176).
99. K. L. Maxwell, B. Garcia, J. Bondy-Denomy, *et al.*, The solution structure of an anti-CRISPR protein. *Nat Commun* **7**, 13134, ISSN: 2041-1723 (Electronic) 2041-1723 (Linking), DOI 10.1038/ncomms13134 (Oct. 2016) (cit. on pp. 10, 176).

100. X. Wang, D. Yao, J.-G. Xu, *et al.*, Structural basis of Cas3 inhibition by the bacteriophage protein AcrF3. *Nature Structural & Molecular Biology* **23**, 868–870, ISSN: 1545-9993 1545-9985, DOI 10.1038/nsmb.3269 (July 2016) (cit. on pp. 10, 176).
101. D. Trasanidou, A. S. Gerós, P. Mohanraju, *et al.*, Keeping crispr in check: diverse mechanisms of phage-encoded anti-crisprs. *FEMS Microbiol Lett* **366**, ISSN: 1574-6968 (Electronic) 0378-1097 (Linking), DOI 10.1093/femsle/fnz098 (May 2019) (cit. on p. 10).
102. N. D. Marino, J. Y. Zhang, A. L. Borges, *et al.*, Discovery of widespread type I and type V CRISPR-Cas inhibitors. *Science* **362**, 240–242, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.aau5174 (Sept. 2018) (cit. on pp. 10, 176).
103. S. Chowdhury, J. Carter, M. F. Rollins, *et al.*, Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. *Cell* **169**, 47–57.e11, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking), DOI 10.1016/j.cell.2017.03.012 (Mar. 2017) (cit. on p. 10).
104. T. W. Guo, A. Bartesaghi, H. Yang, *et al.*, Cryo-EM Structures Reveal Mechanism and Inhibition of DNA Targeting by a CRISPR-Cas Surveillance Complex. *Cell* **171**, 414–426.e12, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking), DOI 10.1016/j.cell.2017.09.006 (Oct. 2017) (cit. on pp. 10, 176).
105. Y. Zhu, A. Gao, Q. Zhan, *et al.*, Diverse Mechanisms of CRISPR-Cas9 Inhibition by Type IIC Anti-CRISPR Proteins. *Molecular Cell* **74**, 296–309.e7, ISSN: 1097-2765, DOI 10.1016/j.molcel.2019.01.038 (Apr. 2019) (cit. on p. 10).
106. L. Dong, X. Guan, N. Li, *et al.*, An anti-CRISPR protein disables type V Cas12a by acetylation. *Nature Structural & Molecular Biology* **26**, 308–314, ISSN: 1545-9993, DOI 10.1038/s41594-019-0206-1 (Apr. 2019) (cit. on p. 10).
107. A. Pawluk, R. H. Staals, C. Taylor, *et al.*, Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nature Microbiology* **1**, ISSN: 2058-5276, DOI 10.1038/nmicrobiol.2016.85 (June 2016) (cit. on pp. 10, 176).
108. B. J. Rauch, M. R. Silvis, J. F. Hultquist, *et al.*, Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. *Cell* **168**, 150–158.e10, ISSN: 0092-8674, DOI 10.1016/j.cell.2016.12.009 (Jan. 2017) (cit. on pp. 10, 176).
109. Y. N. Srikhanta, K. L. Fox, M. P. Jennings, The phasevarion: phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes. *Nat Rev Microbiol* **8**, 196–206, ISSN: 1740-1534 (Electronic) 1740-1526 (Linking), DOI 10.1038/nrmicro2283 (Feb. 2010) (cit. on p. 10).
110. R. W. Carthew, E. J. Sontheimer, Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–55, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking), DOI 10.1016/j.cell.2009.01.035 (Feb. 2009) (cit. on p. 10).
111. L. Thöny-Meyer, D. Kaiser, devRS, an autoregulated and essential genetic locus for fruiting body development in Myxococcus xanthus. *J Bacteriol* **175**, 7450–7462, ISSN: 0021-9193 (Print) 0021-9193 (Linking), DOI 10.1128/jb.175.22.7450-7462.1993 (Nov. 1993) (cit. on p. 10).
112. K. C. Cady, G. A. O'Toole, Non-Identity-Mediated CRISPR-Bacteriophage Interaction Mediated via the Csy and Cas3 Proteins. *J Bacteriol* **193**, 3433–3445, ISSN: 1098-5530 (Electronic) 0021-9193 (Linking), DOI 10.1128/JB.01411-10 (Mar. 2011) (cit. on p. 11).
113. M. E. Zegans, J. C. Wagner, K. C. Cady, *et al.*, Interaction between Bacteriophage DMS3 and Host CRISPR Region Inhibits Group Behaviors of Pseudomonas aeruginosa. *Journal of Bacteriology* **191**, 210–219, ISSN: 0021-9193, DOI 10.1128/jb.00797-08 (Oct. 2009) (cit. on p. 11).
114. F. F. Gunderson, N. P. Cianciotto, The CRISPR-Associated Gene cas2 of Legionella pneumophila Is Required for Intracellular Infection of Amoebae. *Mbio* **4**, ISSN: 2150-7511, DOI ARTNe00074-1310.1128/mBio.00074-13 (2013) (cit. on p. 11).
115. P. Mandin, F. Repoila, M. Vergassola, T. Geissmann, P. Cossart, Identification of new noncoding RNAs in Listeria monocytogenes and prediction of mRNA targets. *Nucleic Acids Research* **35**, 962–974, ISSN: 0305-1048, DOI 10.1093/nar/gkl1096 (Jan. 2007) (cit. on p. 11).
116. K. L. Palmer, M. S. Gilmore, Multidrug-Resistant Enterococci Lack CRISPR-Cas. *MBio* **1**, ed. by R. Losick, ISSN: 2150-7511 (Electronic), DOI 10.1128/mbio.00227-10 (Oct. 2010) (cit. on p. 11).
117. M. Babu, N. Beloglazova, R. Flick, *et al.*, A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. *Mol Microbiol* **79**, 484–502, ISSN: 1365-2958 (Electronic) 0950-382X (Linking), DOI 10.1111/j.1365-2958.2010.07465.x (Dec. 2011) (cit. on p. 11).
118. E. Williams, T. M. Lowe, J. Savas, J. DiRuggiero, Microarray analysis of the hyperthermophilic archaeon Pyrococcus furiosus exposed to gamma irradiation. *Extremophiles* **11**, 19–29, ISSN: 1431-0651, DOI 10.1007/s00792-006-0002-9 (Aug. 2007) (cit. on p. 11).
119. D. H. Haft, J. Selengut, E. F. Mongodin, K. E. Nelson, A Guild of 45 CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes. *PLoS Comput Biol* **1**, e60, ISSN: 1553-7358 (Electronic) 1553-734X (Linking), DOI 10.1371/journal.pcbi.0010060 (2005) (cit. on pp. 11, 213).
120. R. Louwen, D. Horst-Kreft, A. G. Boer, *et al.*, A novel link between Campylobacter jejuni bacteriophage defence, virulence and Guillain–Barré syndrome. *European Journal of Clinical Microbiology & Infectious Diseases* **32**, 207–226, ISSN: 0934-9723, DOI 10.1007/s10096-012-1733-4 (Sept. 2013) (cit. on pp. 11, 213, 215, 219–222, 246).

121. T. R. Sampson, S. D. Saroj, A. C. Llewellyn, Y.-L. Tzeng, D. S. Weiss, A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature* **497**, 254–257, ISSN: 0028-0836, DOI 10.1038/nature12048 (Apr. 2013) (cit. on pp. 11, 213, 215).

122. C. L. Jones, T. R. Sampson, H. I. Nakaya, B. Pulendran, D. S. Weiss, Repression of bacterial lipoprotein production by Francisella novicida facilitates evasion of innate immune recognition. *Cellular Microbiology* **14**, 1531–1543, ISSN: 1462-5814, DOI 10.1111/j.1462-5822.2012.01816.x (June 2012) (cit. on p. 11).

123. F. Jacob, Evolution and tinkering. *Science* **196**, 1161–1166, ISSN: 0036-8075, DOI 10.1126/science.860134 (June 1977) (cit. on p. 11).

124. K. Makarova, Y. Wolf, E. Koonin, The basic building blocks and evolution of CRISPR–Cas systems. *Biochem Soc Trans* **41**, 1392–1400, ISSN: 1470-8752 (Electronic) 0300-5127 (Linking), DOI 10.1042/bst20130038 (Dec. 2013) (cit. on pp. 19, 21, 23).

125. M. R. Tock, D. T. Dryden, The biology of restriction and anti-restriction. *Curr Opin Microbiol* **8**, 466–472, ISSN: 1369-5274 (Print) 1369-5274 (Linking), DOI 10.1016/j.mib.2005.06.003 (Aug. 2005) (cit. on p. 19).

126. K. S. Makarova, L. Aravind, Y. I. Wolf, E. V. Koonin, Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol Direct* **6**, 38, ISSN: 1745-6150 (Electronic) 1745-6150 (Linking), DOI 10.1186/1745-6150-6-38 (2011) (cit. on p. 19).

127. M. P. Terns, R. M. Terns, CRISPR-based adaptive immune systems. *Curr Opin Microbiol* **14**, 321–327, ISSN: 1879-0364 (Electronic) 1369-5274 (Linking), DOI 10.1016/j.mib.2011.03.005 (June 2011) (cit. on p. 19).

128. B. Wiedenheft, S. H. Sternberg, J. A. Doudna, RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331–338, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature10886 (Feb. 2012) (cit. on p. 19).

129. R. Barrangou, L. Marraffini, CRISPR-Cas Systems: Prokaryotes Upgrade to Adaptive Immunity. *Mol Cell* **54**, 234–244, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2014.03.011 (Apr. 2014) (cit. on p. 19).

130. G. Gasiunas, T. Sinkunas, V. Siksnys, Molecular mechanisms of CRISPR-mediated microbial immunity. *Cell Mol Life Sci* **71**, 449–465, ISSN: 1420-9071 (Electronic) 1420-682X (Linking), DOI 10.1007/s00018-013-1438-6 (Aug. 2014) (cit. on p. 19).

131. E. R. Westra, A. Buckling, P. C. Fineran, CRISPR–Cas systems: beyond adaptive immunity. *Nat Rev Microbiol* **12**, 317–326, ISSN: 1740-1534 (Electronic) 1740-1526 (Linking), DOI 10.1038/nrmicro3241 (Apr. 2014) (cit. on pp. 19, 213).

132. T. R. Sampson, D. S. Weiss, CRISPR-Cas systems: new players in gene regulation and bacterial physiology. *Front Cell Infect Microbiol* **4**, 37, ISSN: 2235-2988 (Electronic) 2235-2988 (Linking), DOI 10.3389/fcimb.2014.00037 (Apr. 2014) (cit. on p. 19).

133. R. Louwen, R. H. J. Staals, H. P. Endtz, P. van Baarlen, J. van der Oost, The Role of CRISPR-Cas Systems in Virulence of Pathogenic Bacteria. *Microbiol Mol Biol Rev* **78**, 74–88, ISSN: 1098-5557 (Electronic) 1092-2172 (Linking), DOI 10.1128/mmbr.00039-13 (Mar. 2014) (cit. on pp. 19, 213).

134. N. Takeuchi, Y. I. Wolf, K. S. Makarova, E. V. Koonin, Nature and Intensity of Selection Pressure on CRISPR-Associated Genes. *J Bacteriol* **194**, 1216–1225, ISSN: 1098-5530 (Electronic) 0021-9193 (Linking), DOI 10.1128/JB.06521-11 (Dec. 2012) (cit. on p. 20).

135. E. V. Koonin, Y. I. Wolf, Evolution of the CRISPR-Cas adaptive immunity systems in prokaryotes: models and observations on virus–host coevolution. *Mol Biosyst* **11**, 20–27, ISSN: 1742-2051 (Electronic) 1742-2051 (Linking), DOI 10.1039/c4mb00438h (2015) (cit. on p. 20).

136. K. S. Makarova, Y. I. Wolf, S. Snir, E. V. Koonin, Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems. *J Bacteriol* **193**, 6039–6056, ISSN: 1098-5530 (Electronic) 0021-9193 (Linking), DOI 10.1128/jb.05535-11 (Sept. 2011) (cit. on pp. 21, 23).

137. C. Rouillon, M. Zhou, J. Zhang, *et al.*, Structure of the CRISPR Interference Complex CSM Reveals Key Similarities with Cascade. *Mol Cell* **52**, 124–134, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2013.08.020 (Oct. 2013) (cit. on pp. 21, 27, 29).

138. M. Spilman, A. Cocozaki, C. Hale, *et al.*, Structure of an RNA Silencing Complex of the CRISPR-Cas Immune System. *Mol Cell* **52**, 146–52, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2013.09.008 (Oct. 2013) (cit. on pp. 21, 29).

139. K. Chylinski, K. S. Makarova, E. Charpentier, E. V. Koonin, Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res* **42**, 6091–6105, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gku241 (Apr. 2014) (cit. on p. 21).

140. T. Yamano, H. Nishimasu, B. Zetsche, *et al.*, Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* **165**, 949–962, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking), DOI 10.1016/j.cell.2016.04.003 (May 2016) (cit. on pp. 21, 31, 61, 261).

141. M. Krupovic, K. S. Makarova, P. Forterre, D. Prangishvili, E. V. Koonin, Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity. *BMC Biol* **12**, 36, ISSN: 1741-7007 (Electronic) 1741-7007 (Linking), DOI 10.1186/1741-7007-12-36 (May 2014) (cit. on p. 21).
142. A. B. Hickman, F. Dyda, The casposon-encoded Cas1 protein from Aciduliprofundum booneiis a DNA integrase that generates target site duplications. *Nucleic Acids Res* **43**, 10576–10587, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gkv1180 (Nov. 2015) (cit. on p. 21).
143. M. Krupovic, S. Shmakov, K. S. Makarova, P. Forterre, E. V. Koonin, Recent Mobility of Casposons, Self-Synthesizing Transposons at the Origin of the CRISPR-Cas Immunity. *Genome Biol Evol* **8**, 375–386, ISSN: 1759-6653 (Electronic) 1759-6653 (Linking), DOI 10.1093/gbe/evw006 (Jan. 2016) (cit. on p. 21).
144. E. V. Koonin, M. Krupovic, Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nature Reviews Genetics* **16**, 184–192, DOI 10.1038/nrg3859 (Dec. 2014) (cit. on pp. 21, 23).
145. V. V. Kapitonov, K. S. Makarova, E. V. Koonin, ISC, a Novel Group of Bacterial and Archaeal DNA Transposons That Encode Cas9 Homologs. *J Bacteriol* **198**, ed. by I. B. Zhulin, 797–807, ISSN: 1098-5530 (Electronic) 0021-9193 (Linking), DOI 10.1128/JB.00783-15 (Dec. 2015) (cit. on p. 23).
146. G. Amitai, R. Sorek, CRISPR–Cas adaptation: insights into the mechanism of action. *Nat Rev Microbiol* **14**, 67–76, ISSN: 1740-1534 (Electronic) 1740-1526 (Linking), DOI 10.1038/nrmicro.2015.14 (Jan. 2016) (cit. on pp. 23, 24, 259).
147. D. C. Swarts, C. Mosterd, M. W. J. van Passel, S. J. J. Brouns, CRISPR Interference Directs Strand Specific Spacer Acquisition. *PLoS One* **7**, ed. by I. Mokrousov, e35888, ISSN: 1932-6203 (Electronic) 1932-6203 (Linking), DOI 10.1371/journal.pone.0035888 (Apr. 2012) (cit. on pp. 23, 24).
148. K. A. Datsenko, K. Pougach, A. Tikhonov, *et al.*, Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat Commun* **3**, 945, ISSN: 2041-1723 (Electronic) 2041-1723 (Linking), DOI 10.1038/ncomms1937 (Jan. 2012) (cit. on pp. 23, 24).
149. I. Yosef, M. G. Goren, U. Qimron, Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res* **40**, 5569–5576, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gks216 (Feb. 2012) (cit. on pp. 23, 24, 259).
150. Y. Wei, R. M. Terns, M. P. Terns, Cas9 function and host genome sampling in Type II-A CRISPR–Cas adaptation. *Genes Dev* **29**, 356–361, ISSN: 1549-5477 (Electronic) 0890-9369 (Linking), DOI 10.1101/gad.257550.114 (Feb. 2015) (cit. on pp. 23, 26, 259).
151. C. Díez-Villaseñor, N. M. Guzmán, C. Almendros, J. García-Martínez, F. J. Mojica, CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of Escherichia coli. *RNA Biol* **10**, 792–802, ISSN: 1555-8584 (Electronic) 1547-6286 (Linking), DOI 10.4161/rna.24023 (Feb. 2013) (cit. on pp. 23, 259).
152. J. K. Nuñez, P. J. Kranzusch, J. Noeske, *et al.*, Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nat Struct Mol Biol* **21**, 528–534, ISSN: 1545-9985 (Electronic) 1545-9985 (Linking), DOI 10.1038/nsmb.2820 (May 2014) (cit. on pp. 23, 24, 259).
153. A. Levy, M. G. Goren, I. Yosef, *et al.*, CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520**, 505–10, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature14302 (Apr. 2015) (cit. on pp. 23, 24).
154. S. Erdmann, S. L. M. Bauer, R. A. Garrett, Inter-viral conflicts that exploit host CRISPR immune systems of Sulfolobus. *Mol Microbiol* **91**, 900–17, ISSN: 1365-2958 (Electronic) 0950-382X (Linking), DOI 10.1111/mmi.12503 (Jan. 2014) (cit. on p. 23).
155. D. B. Wigley, RecBCD: the supercar of DNA repair. *Cell* **131**, 651–3, ISSN: 0092-8674 (Print) 0092-8674 (Linking), DOI S0092-8674(07)01408-0[pii]10.1016/j.cell.2007.11.004 (2007) (cit. on p. 23).
156. M.-È. Dupuis, M. Villion, A. H. Magadán, S. Moineau, CRISPR-Cas and restriction–modification systems are compatible and increase phage resistance. *Nat Commun* **4**, 2087, ISSN: 2041-1723 (Electronic) 2041-1723 (Linking), DOI 10.1038/ncomms3087 (July 2013) (cit. on p. 24).
157. Y. Wang, Z.-T. Zhang, S.-O. Seo, *et al.*, Markerless chromosomal gene deletion in Clostridium beijerinckii using CRISPR/Cas9 system. *Journal of Biotechnology* **200**, 1–5, ISSN: 0168-1656, DOI 10.1016/j.jbiotec.2015.02.005 (Apr. 2015) (cit. on pp. 24, 175).
158. J. K. Nuñez, A. S. Y. Lee, A. Engelman, J. A. Doudna, Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature* **519**, 193–198, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature14237 (Feb. 2015) (cit. on pp. 24, 259).
159. S. Silas, G. Mohr, D. J. Sidote, *et al.*, Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science* **351**, aad4234–aad4234, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.aad4234 (Feb. 2016) (cit. on p. 24).
160. P. C. Fineran, E. Charpentier, Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology* **434**, 202–9, ISSN: 1096-0341 (Electronic) 0042-6822 (Linking), DOI 10.1016/j.virol.2012.10.003 (2012) (cit. on p. 24).

161. Z. Arslan, V. Hermanns, R. Wurm, R. Wagner, Ü. Pul, Detection and characterisation of spacer integration intermediates in type I-E CRISPR–Cas system. *Nucleic Acids Res* **42**, 7884–7893, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gku510 (June 2014) (cit. on pp. 24, 259).

162. M. Li, R. Wang, D. Zhao, H. Xiang, Adaptation of the Haloarcula hispanica CRISPR-Cas system to a purified virus strictly requires a priming process. *Nucleic Acids Res* **42**, 2483–92, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gkt1154 (2014) (cit. on pp. 24, 26).

163. C. Richter, R. L. Dy, R. E. McKenzie, *et al.*, Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res* **42**, 8516–26, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gku527 (July 2014) (cit. on p. 24).

164. P. C. Fineran, M. J. H. Gerritzen, M. Suarez-Diez, *et al.*, Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc Natl Acad Sci U S A* **111**, E1629–E1638, ISSN: 1091-6490 (Electronic) 0027-8424 (Linking), DOI 10.1073/pnas.1400071111 (Apr. 2014) (cit. on pp. 24, 25).

165. C. Xue, A. S. Seetharam, O. Musharova, *et al.*, CRISPR interference and priming varies with individual spacer sequences. *Nucleic Acids Res* **43**, 10831–10847, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gkv1259 (Nov. 2015) (cit. on p. 26).

166. D. Vorontsova, K. A. Datsenko, S. Medvedeva, *et al.*, Foreign DNA acquisition by the I-F CRISPR–Cas system requires all components of the interference machinery. *Nucleic Acids Res* **43**, 10848–60, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gkv1261 (Nov. 2015) (cit. on p. 26).

167. I. Ivančić-Baće, S. D. Cass, S. J. Wearne, E. L. Bolt, Different genome stability proteins underpin primed and naïve adaptation inE. coliCRISPR-Cas immunity. *Nucleic Acids Res* **43**, 10821–10830, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gkv1213 (Nov. 2015) (cit. on pp. 26, 259).

168. R. Heler, P. Samai, J. W. Modell, *et al.*, Cas9 specifies functional viral targets during CRISPR–Cas adaptation. *Nature* **519**, 199–202, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature14245 (Feb. 2015) (cit. on pp. 26, 259).

169. S. P. T. Hooton, I. F. Connerton, Campylobacter jejuni acquire new host-derived CRISPR spacers when in association with bacteriophages harboring a CRISPR-like Cas4 protein. *Front Microbiol* **5**, 744, ISSN: 1664-302X (Electronic) 1664-302X (Linking), DOI 10.3389/fmicb.2014.00744 (Jan. 2014) (cit. on p. 26).

170. E. Charpentier, H. Richter, J. van der Oost, M. F. White, Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiology Reviews* **39**, 428–441, DOI 10.1093/femsre/fuv023 (May 2015) (cit. on pp. 26, 27).

171. 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, ISSN: 0890-9369 (Print) 0890-9369 (Linking), DOI 10.1101/gad.1742908 (Dec. 2008) (cit. on pp. 26, 27).

172. L. A. Marraffini, E. J. Sontheimer, Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**, 568–571, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature08703 (Jan. 2010) (cit. on p. 26).

173. M. M. Jore, M. Lundgren, E. van Duijn, *et al.*, Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* **18**, 529–536, ISSN: 1545-9985 (Electronic) 1545-9985 (Linking), DOI 10.1038/nsmb.2019 (Apr. 2011) (cit. on pp. 27, 29).

174. J. Reeks, J. Naismith, M. White, CRISPR interference: a structural perspective. *Biochem J* **453**, 155–166, ISSN: 1470-8728 (Electronic) 0264-6021 (Linking), DOI 10.1042/bj20130316 (July 2013) (cit. on p. 27).

175. O. Niewoehner, M. Jinek, J. A. Doudna, Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. *Nucleic Acids Res* **42**, 1341–1353, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gkt922 (Oct. 2014) (cit. on p. 27).

176. B. Wiedenheft, G. C. Lander, K. Zhou, *et al.*, Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **477**, 486–489, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature10402 (Sept. 2011) (cit. on pp. 27, 29).

177. M. F. Rollins, J. T. Schuman, K. Paulus, H. S. Bukhari, B. Wiedenheft, Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from Pseudomonas aeruginosa. *Nucleic Acids Res* **43**, 2216–2222, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gkv094 (Feb. 2015) (cit. on p. 27).

178. B. Wiedenheft, E. van Duijn, J. B. Bultema, *et al.*, RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci U S A* **108**, 10092–10097, ISSN: 1091-6490 (Electronic) 0027-8424 (Linking), DOI 10.1073/pnas.1102716108 (May 2011) (cit. on p. 27).

179. R. E. Haurwitz, M. Jinek, B. Wiedenheft, K. Zhou, J. A. Doudna, Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease. *Science* **329**, 1355–1358, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.1192272 (Sept. 2010) (cit. on pp. 27, 261).

180. K. Nam, C. Haitjema, X. Liu, *et al.*, Cas5d Protein Processes Pre-crRNA and Assembles into a Cascade-like Interference Complex in Subtype I-C/Dvulg CRISPR-Cas System. *Structure* **20**, 1574–1584, ISSN: 1878-4186 (Electronic) 0969-2126 (Linking), DOI 10.1016/j.str.2012.06.016 (Sept. 2012) (cit. on p. 27).

181. R. Staals, Y. Zhu, D. Taylor, *et al.*, RNA Targeting by the Type III-A CRISPR-Cas Csm Complex of Thermus thermophilus. *Mol Cell* **56**, 518–530, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2014.10.005 (Nov. 2014) (cit. on pp. 27, 29, 260).
182. S. Majumdar, P. Zhao, N. T. Pfister, *et al.*, Three CRISPR-Cas immune effector complexes coexist in Pyrococcus furiosus. *RNA* **21**, 1147–1158, ISSN: 1469-9001 (Electronic) 1355-8382 (Linking), DOI 10.1261/rna.049130.114 (Apr. 2015) (cit. on pp. 27, 260).
183. J. Zhang, C. Rouillon, M. Kerou, *et al.*, Structure and Mechanism of the CMR Complex for CRISPR-Mediated Antiviral Immunity. *Mol Cell* **45**, 303–313, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2011.12.013 (Feb. 2012) (cit. on p. 27).
184. A. Hatoum-Aslan, P. Samai, I. Maniv, W. Jiang, L. A. Marraffini, A Ruler Protein in a Complex for Antiviral Defense Determines the Length of Small Interfering CRISPR RNAs. *J Biol Chem* **288**, 27888–27897, ISSN: 1083-351X (Electronic) 0021-9258 (Linking), DOI 10.1074/jbc.m113.499244 (Aug. 2013) (cit. on p. 27).
185. R. Garrett, S. Shah, S. Erdmann, *et al.*, CRISPR-Cas Adaptive Immune Systems of the Sulfolobales: Unravelling Their Complexity and Diversity. *Life (Basel)* **5**, 783–817, ISSN: 2075-1729 (Electronic) 2075-1729 (Linking), DOI 10.3390/life5010783 (Mar. 2015) (cit. on p. 27).
186. Y. Zhang, N. Heidrich, B. Ampattu, *et al.*, Processing-Independent CRISPR RNAs Limit Natural Transformation in Neisseria meningitidis. *Mol Cell* **50**, 488–503, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2013.05.001 (May 2013) (cit. on pp. 27, 261).
187. I. Fonfara, H. Richter, M. Bratovič, A. L. Rhun, E. Charpentier, The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532**, 517–521, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature17945 (Apr. 2016) (cit. on pp. 27, 31, 61, 81, 147, 261).
188. S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, J. A. Doudna, DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature13011 (Jan. 2014) (cit. on pp. 29, 31, 264).
189. S. Redding, S. Sternberg, M. Marshall, *et al.*, Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. *Cell* **163**, 854–865, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking), DOI 10.1016/j.cell.2015.10.003 (Nov. 2015) (cit. on p. 29).
190. M. Rutkauskas, T. Sinkunas, I. Songailiene, *et al.*, Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. *Cell Rep* **10**, 1534–1543, ISSN: 2211-1247 (Electronic), DOI 10.1016/j.celrep.2015.01.067 (Mar. 2015) (cit. on p. 29).
191. T. Blosser, L. Loeff, E. Westra, *et al.*, Two Distinct DNA Binding Modes Guide Dual Roles of a CRISPR-Cas Protein Complex. *Mol Cell* **58**, 60–70, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2015.01.028 (Apr. 2015) (cit. on p. 29).
192. J. Wang, J. Li, H. Zhao, *et al.*, Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell* **163**, 840–853, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking), DOI 10.1016/j.cell.2015.10.008 (Nov. 2015) (cit. on p. 29).
193. R. N. Jackson, S. M. Golden, P. B. G. van Erp, *et al.*, Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. *Science* **345**, 1473–1479, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.1256328 (Aug. 2014) (cit. on p. 29).
194. R. Jackson, B. Wiedenheft, A Conserved Structural Chassis for Mounting Versatile CRISPR RNA-Guided Immune Responses. *Mol Cell* **58**, 722–728, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2015.05.023 (June 2015) (cit. on p. 29).
195. S. Mulepati, A. Heroux, S. Bailey, Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science* **345**, 1479–1484, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.1256996 (Aug. 2014) (cit. on p. 29).
196. D. W. Taylor, Y. Zhu, R. H. J. Staals, *et al.*, Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. *Science* **348**, 581–5, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.aaa4535 (Apr. 2015) (cit. on p. 29).
197. T. Osawa, H. Inanaga, C. Sato, T. Numata, Crystal Structure of the CRISPR-Cas RNA Silencing Cmr Complex Bound to a Target Analog. *Mol Cell* **58**, 418–430, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2015.03.018 (May 2015) (cit. on p. 29).
198. Z. Zebec, A. Manica, J. Zhang, M. F. White, C. Schleper, CRISPR-mediated targeted mRNA degradation in the archaeon Sulfolobus solfataricus. *Nucleic Acids Res* **42**, 5280–5288, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gku161 (Mar. 2014) (cit. on p. 29).
199. R. Staals, Y. Agari, S. Maki-Yonekura, *et al.*, Structure and Activity of the RNA-Targeting Type III-B CRISPR-Cas Complex of Thermus thermophilus. *Mol Cell* **52**, 135–145, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2013.09.013 (Oct. 2013) (cit. on pp. 29, 260).

200. G. Tamulaitis, M. Kazlauskiene, E. Manakova, *et al.*, Programmable RNA Shredding by the Type III-A CRISPR-Cas System of Streptococcus thermophilus. *Mol Cell* **56**, 506–517, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2014.09.027 (Nov. 2014) (cit. on pp. 29, 30).

201. W. Peng, M. Feng, X. Feng, Y. X. Liang, Q. She, An archaeal CRISPR type III-B system exhibiting distinctive RNA targeting features and mediating dual RNA and DNA interference. *Nucleic Acids Res* **43**, 406–417, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gku1302 (Dec. 2015) (cit. on p. 29).

202. M. A. Estrella, F.-T. Kuo, S. Bailey, RNA-activated DNA cleavage by the Type III-B CRISPR–Cas effector complex. *Genes Dev* **30**, 460–470, ISSN: 1549-5477 (Electronic) 0890-9369 (Linking), DOI 10.1101/gad.273722.115 (Feb. 2016) (cit. on pp. 29, 30).

203. J. R. Elmore, N. F. Sheppard, N. Ramia, *et al.*, Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR–Cas system. *Genes Dev* **30**, 447–459, ISSN: 1549-5477 (Electronic) 0890-9369 (Linking), DOI 10.1101/gad.272153.115 (Feb. 2016) (cit. on pp. 29, 30, 264).

204. L. Deng, R. A. Garrett, S. A. Shah, X. Peng, Q. She, A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus. *Mol Microbiol* **87**, 1088–1099, ISSN: 1365-2958 (Electronic) 0950-382X (Linking), DOI 10.1111/mmi.12152 (Feb. 2013) (cit. on p. 29).

205. P. Samai, N. Pyenson, W. Jiang, *et al.*, Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell* **161**, 1164–1174, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking), DOI 10.1016/j.cell.2015.04.027 (May 2015) (cit. on pp. 29, 30).

206. G. W. Goldberg, W. Jiang, D. Bikard, L. A. Marraffini, Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature* **514**, 633–637, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature13637 (Aug. 2014) (cit. on p. 29).

207. C. R. Hale, A. Cocozaki, H. Li, R. M. Terns, M. P. Terns, Target RNA capture and cleavage by the Cmr type III-B CRISPR–Cas effector complex. *Genes Dev* **28**, 2432–2443, ISSN: 1549-5477 (Electronic) 0890-9369 (Linking), DOI 10.1101/gad.250712.114 (Nov. 2014) (cit. on p. 29).

208. C. Benda, J. Ebert, R. Scheltema, *et al.*, Structural Model of a CRISPR RNA-Silencing Complex Reveals the RNA-Target Cleavage Activity in Cmr4. *Mol Cell* **56**, 43–54, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2014.09.002 (Oct. 2014) (cit. on p. 29).

209. N. Ramia, M. Spilman, L. Tang, *et al.*, Essential Structural and Functional Roles of the Cmr4 Subunit in RNA Cleavage by the Cmr CRISPR-Cas Complex. *Cell Rep* **9**, 1610–7, ISSN: 2211-1247 (Electronic), DOI 10.1016/j.celrep.2014.11.007 (Dec. 2014) (cit. on p. 29).

210. N. F. Ramia, L. Tang, A. I. Cocozaki, H. Li, Staphylococcus epidermidis Csm1 is a 3'-5' exonuclease. *Nucleic Acids Res* **42**, 1129–38, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gkt914 (2014) (cit. on p. 30).

211. N. F. Sheppard, C. V. Glover, R. M. Terns, M. P. Terns, The CRISPR-associated Csx1 protein of Pyrococcus furiosusis an adenosine-specific endoribonuclease. *RNA* **22**, 216–224, ISSN: 1469-9001 (Electronic) 1355-8382 (Linking), DOI 10.1261/rna.039842.113 (Dec. 2016) (cit. on p. 30).

212. W. Jiang, P. Samai, L. Marraffini, Degradation of Phage Transcripts by CRISPR-Associated RNases Enables Type III CRISPR-Cas Immunity. *Cell* **164**, 710–721, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking), DOI 10.1016/j.cell.2015.12.053 (Feb. 2016) (cit. on p. 30).

213. O. Niewoehner, M. Jinek, Structural basis for the endoribonuclease activity of the type III-A CRISPR-associated protein Csm6. *RNA* **22**, 318–329, ISSN: 1469-9001 (Electronic) 1355-8382 (Linking), DOI 10.1261/rna.054098.115 (Jan. 2016) (cit. on p. 30).

214. L. A. Marraffini, E. J. Sontheimer, CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* **11**, 181–190, ISSN: 1471-0064 (Electronic) 1471-0056 (Linking), DOI 10.1038/nrg2749 (Feb. 2010) (cit. on p. 30).

215. J. van der Oost, E. R. Westra, R. N. Jackson, B. Wiedenheft, Unravelling the structural and mechanistic basis of CRISPR–Cas systems. *Nat Rev Microbiol* **12**, 479–492, ISSN: 1740-1534 (Electronic) 1740-1526 (Linking), DOI 10.1038/nrmicro3279 (June 2014) (cit. on pp. 30, 39).

216. M. Jinek, F. Jiang, D. W. Taylor, *et al.*, Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science* **343**, 1247997–1247997, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.1247997 (Feb. 2014) (cit. on pp. 30, 31, 121, 129, 213).

217. H. Nishimasu, F. A. Ran, P. Hsu, *et al.*, Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **156**, 935–949, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking), DOI 10.1016/j.cell.2014.02.001 (Feb. 2014) (cit. on pp. 30, 31).

218. C. Anders, O. Niewoehner, A. Duerst, M. Jinek, Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature13579 (July 2014) (cit. on pp. 31, 264).

219. S. H. Sternberg, B. LaFrance, M. Kaplan, J. A. Doudna, Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature* **527**, 110–113, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature15544 (Oct. 2015) (cit. on p. 31).

220. A. H. Magadán, M.-È. Dupuis, M. Villion, S. Moineau, Cleavage of Phage DNA by the Streptococcus thermophilus CRISPR3-Cas System. *PLoS One* **7**, ed. by A. R. Poteete, e40913, ISSN: 1932-6203 (Electronic) 1932-6203 (Linking), DOI 10.1371/journal.pone.0040913 (July 2012) (cit. on p. 31).

221. Y. Zhang, R. Rajan, H. S. Seifert, A. Mondragón, E. Sontheimer, DNase H Activity of Neisseria meningitidis Cas9. *Mol Cell* **60**, 242–255, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2015.09.020 (Oct. 2015) (cit. on pp. 31, 82, 116).

222. P. Hsu, E. Lander, F. Zhang, Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **157**, 1262–1278, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking), DOI 10.1016/j.cell.2014.05.010 (June 2014) (cit. on pp. 31, 39).

223. V. Anantharaman, K. S. Makarova, A. M. Burroughs, E. V. Koonin, L. Aravind, Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing. *Biol Direct* **8**, 15, ISSN: 1745-6150 (Electronic) 1745-6150 (Linking), DOI 10.1186/1745-6150-8-15 (June 2013) (cit. on p. 31).

224. O. O. Abudayyeh, J. S. Gootenberg, S. Konermann, *et al.*, C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.aaf5573 (June 2016) (cit. on pp. 31, 79, 263, 265).

225. A. Levasseur, M. Bekliz, E. Chabrière, *et al.*, MIMIVIRE is a defence system in mimivirus that confers resistance to virophage. *Nature* **531**, 249–252, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature17146 (Feb. 2016) (cit. on p. 33).

226. M. Jinek, A. East, A. Cheng, *et al.*, RNA-programmed genome editing in human cells. *Elife* **2**, e00471, ISSN: 2050-084X (Print) 2050-084X (Linking), DOI 10.7554/eLife.00471 (Jan. 2013) (cit. on pp. 33, 267, 268).

227. S. W. Cho, S. Kim, J. M. Kim, J.-S. Kim, Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* **31**, 230–2, ISSN: 1546-1696 (Electronic) 1087-0156 (Linking), DOI 10.1038/nbt.2507 (Jan. 2013) (cit. on p. 33).

228. S. Konermann, M. D. Brigham, A. E. Trevino, *et al.*, Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472–476, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature12466 (Aug. 2013) (cit. on pp. 33, 63, 68, 267).

229. L. Gilbert, M. Larson, L. Morsut, *et al.*, CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **154**, 442–451, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking), DOI 10.1016/j.cell.2013.06.044 (July 2013) (cit. on pp. 33, 86, 267).

230. N. A. Kearns, H. Pham, B. Tabak, *et al.*, Functional annotation of native enhancers with a Cas9–histone demethylase fusion. *Nat Methods* **12**, 401–403, ISSN: 1548-7105 (Electronic) 1548-7091 (Linking), DOI 10.1038/nmeth.3325 (Mar. 2015) (cit. on p. 33).

231. S. Kiani, J. Beal, M. R. Ebrahimkhani, *et al.*, CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nat Methods* **11**, 723–6, ISSN: 1548-7105 (Electronic) 1548-7091 (Linking), DOI 10.1038/nmeth.2969 (May 2014) (cit. on p. 33).

232. Y. Liu, Y. Zeng, L. Liu, *et al.*, Synthesizing AND gate genetic circuits based on CRISPR-Cas9 for identification of bladder cancer cells. *Nat Commun* **5**, 5393, ISSN: 2041-1723 (Electronic) 2041-1723 (Linking), DOI 10.1038/ncomms6393 (Nov. 2014) (cit. on p. 33).

233. L. Nissim, S. Perli, A. Fridkin, P. Perez-Pinera, T. Lu, Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells. *Mol Cell* **54**, 698–710, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2014.04.022 (May 2014) (cit. on pp. 33, 61).

234. D. Bikard, C. W. Euler, W. Jiang, *et al.*, Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. *Nat Biotechnol* **32**, 1146–1150, ISSN: 1546-1696 (Electronic) 1087-0156 (Linking), DOI 10.1038/nbt.3043 (Oct. 2014) (cit. on p. 33).

235. H. Ebina, N. Misawa, Y. Kanemura, Y. Koyanagi, Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus. *Sci Rep* **3**, 2510, ISSN: 2045-2322 (Electronic) 2045-2322 (Linking), DOI 10.1038/srep02510 (Aug. 2013) (cit. on p. 33).

236. V. Ramanan, A. Shlomai, D. B. Cox, *et al.*, CRISPR/Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus. *Sci Rep* **5**, 10833, ISSN: 2045-2322 (Electronic) 2045-2322 (Linking), DOI 10.1038/srep10833 (June 2015) (cit. on p. 33).

237. W. Hu, R. Kaminski, F. Yang, *et al.*, RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. *Proc Natl Acad Sci U S A* **111**, 11461–11466, ISSN: 1091-6490 (Electronic) 0027-8424 (Linking), DOI 10.1073/pnas.1405186111 (July 2014) (cit. on p. 33).

238. S. Konermann, M. D. Brigham, A. E. Trevino, *et al.*, Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature14136 (Dec. 2015) (cit. on pp. 33, 267).

239. L. Gilbert, M. Horlbeck, B. Adamson, *et al.*, Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647–661, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking), DOI 10.1016/j.cell.2014.09.029 (Oct. 2014) (cit. on p. 33).

240. B. P. Kleinstiver, S. Q. Tsai, M. S. Prew, *et al.*, Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat Biotechnol* **34**, 869–874, ISSN: 1546-1696 (Electronic) 1087-0156 (Linking), DOI 10.1038/nbt.3620 (June 2016) (cit. on p. 33).

241. D. Kim, J. Kim, J. K. Hur, *et al.*, Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat Biotechnol* **34**, 863–868, ISSN: 1546-1696 (Electronic) 1087-0156 (Linking), DOI 10.1038/nbt.3609 (June 2016) (cit. on pp. 33, 39).

242. P. Mali, K. M. Esvelt, G. M. Church, Cas9 as a versatile tool for engineering biology. *Nat Methods* **10**, 957–63, ISSN: 1548-7105 (Electronic) 1548-7091 (Linking), DOI 10.1038/nmeth.2649 (Sept. 2013) (cit. on pp. 34, 39).

243. F. A. Ran, P. Hsu, C.-Y. Lin, *et al.*, Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154**, 1380–1389, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking), DOI 10.1016/j.cell.2013.08.021 (Sept. 2013) (cit. on p. 34).

244. 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, ISSN: 1552-4469 (Electronic) 1552-4450 (Linking), DOI 10.1038/nchembio.1793 (Apr. 2015) (cit. on p. 34).

245. Y. Fu, J. D. Sander, D. Reyon, V. M. Cascio, J. K. Joung, Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* **32**, 279–284, ISSN: 1546-1696 (Electronic) 1087-0156 (Linking), DOI 10.1038/nbt.2808 (Jan. 2014) (cit. on p. 34).

246. B. P. Kleinstiver, V. Pattanayak, M. S. Prew, *et al.*, High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature16526 (Jan. 2016) (cit. on pp. 34, 39, 86, 269).

247. I. M. Slaymaker, L. Gao, B. Zetsche, *et al.*, Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.aad5227 (Dec. 2016) (cit. on p. 34).

248. F. A. Ran, L. Cong, W. X. Yan, *et al.*, In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186–191, ISSN: 0028-0836, DOI 10.1038/nature14299 (Apr. 2015) (cit. on pp. 34, 81, 112, 270).

249. R. Platt, S. Chen, Y. Zhou, *et al.*, CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. *Cell* **159**, 440–455, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking), DOI 10.1016/j.cell.2014.09.014 (Oct. 2014) (cit. on p. 34).

250. W. Qin, S. L. Dion, P. M. Kutny, *et al.*, Efficient CRISPR/Cas9-Mediated Genome Editing in Mice by Zygote Electroporation of Nuclease. *Genetics* **200**, 423–30, ISSN: 1943-2631 (Electronic) 0016-6731 (Linking), DOI 10.1534/genetics.115.176594 (Mar. 2015) (cit. on p. 34).

251. D. D'Astolfo, R. Pagliero, A. Pras, *et al.*, Efficient Intracellular Delivery of Native Proteins. *Cell* **161**, 674–690, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking), DOI 10.1016/j.cell.2015.03.028 (Apr. 2015) (cit. on p. 34).

252. K. M. Esvelt, H. H. Wang, Genome-scale engineering for systems and synthetic biology. *Molecular Systems Biology* **9**, 641, ISSN: 1744-4292, DOI 10.1038/msb.2012.66 (Jan. 2014) (cit. on pp. 34, 175).

253. L. A. Marraffini, CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55–61, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature15386 (Oct. 2015) (cit. on p. 39).

254. J. A. Doudna, E. Charpentier, The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096–1258096, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.1258096 (Nov. 2014) (cit. on pp. 39, 218).

255. T. Karvelis, G. Gasiunas, A. Miksys, *et al.*, crRNA and tracrRNA guide Cas9-mediated DNA interference inStreptococcus thermophilus. *RNA Biol* **10**, 841–851, ISSN: 1555-8584 (Electronic) 1547-6286 (Linking), DOI 10.4161/rna.24203 (Mar. 2013) (cit. on pp. 39, 147).

256. B. Zetsche, M. Heidenreich, P. Mohanraju, *et al.*, Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nature Biotechnology* **35**, 31–34, ISSN: 1087-0156, DOI 10.1038/nbt.3737 (Dec. 2016) (cit. on pp. 39, 81, 121, 147, 159, 269).

257. 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, ISSN: 1752-9867 (Electronic) 1674-2052 (Linking), DOI 10.1016/j.molp.2017.03.001 (July 2017) (cit. on p. 39).

258. S.-Y. Li, G.-P. Zhao, J. Wang, C-Brick: A New Standard for Assembly of Biological Parts Using Cpf1. *ACS Synth Biol* **5**, 1383–1388, ISSN: 2161-5063 (Electronic) 2161-5063 (Linking), DOI 10.1021/acssynbio.6b00114 (June 2016) (cit. on p. 39).

259. C. Lei, S.-Y. Li, J.-K. Liu, *et al.*, The CCTL (Cpf1-assisted Cutting and Taq DNA ligase-assisted Ligation) method for efficient editing of large DNA constructsin vitro. *Nucleic Acids Res* **45**, e74, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gkx018 (Jan. 2017) (cit. on p. 39).
260. J. K. Hur, K. Kim, K. W. Been, *et al.*, Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins. *Nat Biotechnol* **34**, 807–808, ISSN: 1546-1696 (Electronic) 1087-0156 (Linking), DOI 10.1038/nbt.3596 (June 2016) (cit. on p. 39).
261. Y. Kim, S.-A. Cheong, J. G. Lee, *et al.*, Generation of knockout mice by Cpf1-mediated gene targeting. *Nat Biotechnol* **34**, 808–10, ISSN: 1546-1696 (Electronic) 1087-0156 (Linking), DOI 10.1038/nbt.3614 (June 2016) (cit. on p. 39).
262. A. Endo, M. Masafumi, H. Kaya, S. Toki, Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from Francisella novicida. *Sci Rep* **6**, 38169, ISSN: 2045-2322 (Electronic) 2045-2322 (Linking), DOI 10.1038/srep38169 (Dec. 2016) (cit. on p. 39).
263. R. Xu, R. Qin, H. Li, *et al.*, Generation of targeted mutant rice using a CRISPR-Cpf1 system. *Plant Biotechnol J* **15**, 713–717, ISSN: 1467-7652 (Electronic) 1467-7644 (Linking), DOI 10.1111/pbi.12669 (Feb. 2017) (cit. on p. 39).
264. 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, ISSN: 1097-0061 (Electronic) 0749-503X (Linking), DOI 10.1002/yea.3278 (Nov. 2018) (cit. on p. 39).
265. M. A. Świat, S. Dashko, M. den Ridder, *et al.*, FnCpf1: a novel and efficient genome editing tool for Saccharomyces cerevisiae. *Nucleic Acids Res* **45**, 12585–12598, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gkx1007 (Nov. 2017) (cit. on p. 39).
266. M. A. Moreno-Mateos, J. P. Fernandez, R. Rouet, *et al.*, CRISPR-Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing. *Nat Commun* **8**, 2024, ISSN: 2041-1723 (Electronic) 2041-1723 (Linking), DOI 10.1038/s41467-017-01836-2 (Dec. 2017) (cit. on pp. 39, 269).
267. A. Ferenczi, D. E. Pyott, A. Xipnitou, A. Molnar, Efficient targeted DNA editing and replacement in Chlamydomonas reinhardtii using Cpf1 ribonucleoproteins and single-stranded DNA. *Proc Natl Acad Sci U S A* **114**, 13567–13572, ISSN: 1091-6490 (Electronic) 0027-8424 (Linking), DOI 10.1073/pnas.1710597114 (Dec. 2017) (cit. on p. 39).
268. S. S.-e.-A. Zaidi, M. M. Mahfouz, S. Mansoor, CRISPR-Cpf1: A New Tool for Plant Genome Editing. *Trends Plant Sci* **22**, 550–553, ISSN: 1878-4372 (Electronic) 1360-1385 (Linking), DOI 10.1016/j.tplants.2017.05.001 (July 2017) (cit. on p. 39).
269. H. Kim, S.-T. Kim, J. Ryu, *et al.*, CRISPR/Cpf1-mediated DNA-free plant genome editing. *Nat Commun* **8**, 14406, ISSN: 2041-1723 (Electronic) 2041-1723 (Linking), DOI 10.1038/ncomms14406 (Feb. 2017) (cit. on pp. 39, 269).
270. 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, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2017.03.016 (Apr. 2017) (cit. on pp. 39, 56, 261, 264).
271. F. Studier, B. A. Moffatt, Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**, 113–130, ISSN: 0022-2836 (Print) 0022-2836 (Linking), DOI 10.1016/0022-2836(86)90385-2 (May 1986) (cit. on p. 39).
272. F. W. Studier, A. H. Rosenberg, J. J. Dunn, J. W. Dubendorff, in *Methods in Enzymology* (Elsevier, 1990), vol. 185, pp. 60–89, DOI 10.1016/0076-6879(90)85008-c (cit. on p. 39).
273. A. H. Rosenberg, B. N. Lade, C. Dao-shan, *et al.*, Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* **56**, 125–135, ISSN: 0378-1119 (Print) 0378-1119 (Linking), DOI 10.1016/0378-1119(87)90165-x (Jan. 1987) (cit. on p. 39).
274. A. M. Kabadi, D. G. Ousterout, I. B. Hilton, C. A. Gersbach, Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic Acids Res* **42**, e147, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gku749 (Aug. 2014) (cit. on p. 61).
275. T. Sakuma, A. Nishikawa, S. Kume, K. Chayama, T. Yamamoto, Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Sci Rep* **4**, 5400, ISSN: 2045-2322 (Electronic) 2045-2322 (Linking), DOI 10.1038/srep05400 (June 2014) (cit. on p. 61).
276. S. Q. Tsai, N. Wyvekens, C. Khayter, *et al.*, Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol* **32**, 569–576, ISSN: 1546-1696 (Electronic) 1087-0156 (Linking), DOI 10.1038/nbt.2908 (Apr. 2014) (cit. on p. 61).
277. K. Xie, B. Minkenberg, Y. Yang, Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc Natl Acad Sci U S A* **112**, 3570–3575, ISSN: 1091-6490 (Electronic) 0027-8424 (Linking), DOI 10.1073/pnas.1420294112 (Mar. 2015) (cit. on p. 61).
278. C. Ostlund, E. S. Folker, J. C. Choi, *et al.*, Dynamics and molecular interactions of linker of nucleoskeleton and cytoskeleton (LINC) complex proteins. *J Cell Sci* **122**, 4099–4108, ISSN: 1477-9137 (Electronic) 0021-9533 (Linking), DOI 10.1242/jcs.057075 (Oct. 2009) (cit. on p. 63).

279. L. Swiech, M. Heidenreich, A. Banerjee, *et al.*, In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. *Nature Biotechnology* **33**, 102–106, ISSN: 1087-0156, DOI 10.1038/nbt.3055 (Oct. 2015) (cit. on pp. 63, 68, 69).

280. N. Heidrich, G. Dugar, J. Vogel, C. M. Sharma, in *Methods in Molecular Biology* (Springer New York, 2015), vol. 1311, pp. 1–21, DOI 10.1007/978-1-4939-2687-9_1 (cit. on p. 67).

281. P. D. Hsu, D. A. Scott, J. A. Weinstein, *et al.*, DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827–832, ISSN: 1546-1696 (Electronic) 1087-0156 (Linking), DOI 10.1038/nbt.2647 (July 2013) (cit. on pp. 67, 268).

282. B. Giardine, Galaxy: A platform for interactive large-scale genome analysis. *Genome Res* **15**, 1451–1455, ISSN: 1088-9051 (Print) 1088-9051 (Linking), DOI 10.1101/gr.4086505 (Sept. 2005) (cit. on p. 67).

283. R. Barrangou, P. Horvath, A decade of discovery: CRISPR functions and applications. *Nat Microbiol* **2**, 17092, ISSN: 2058-5276 (Electronic) 2058-5276 (Linking), DOI 10.1038/nmicrobiol.2017.92 (June 2017) (cit. on pp. 79, 175).

284. E. V. Koonin, K. S. Makarova, F. Zhang, Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology* **37**, 67–78, ISSN: 1369-5274, DOI 10.1016/j.mib.2017.05.008 (June 2017) (cit. on pp. 79, 258).

285. G. J. Knott, J. A. Doudna, CRISPR-Cas guides the future of genetic engineering. *Science* **361**, 866–869, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.aat5011 (Aug. 2018) (cit. on p. 79).

286. 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, ISSN: 1552-4469 (Electronic) 1552-4450 (Linking), DOI 10.1038/s41589-018-0080-x (June 2018) (cit. on p. 79).

287. J. S. Chen, E. Ma, L. B. Harrington, *et al.*, CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436–439, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.aar6245 (Feb. 2018) (cit. on pp. 79, 264).

288. A. East-Seletsky, M. R. O'Connell, S. C. Knight, *et al.*, Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* **538**, 270–273, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature19802 (Sept. 2016) (cit. on pp. 79, 81, 86, 265).

289. Z. Wu, H. Yang, P. Colosi, Effect of Genome Size on AAV Vector Packaging. *Mol Ther* **18**, 80–86, ISSN: 1525-0024 (Electronic) 1525-0016 (Linking), DOI 10.1038/mt.2009.255 (Jan. 2010) (cit. on p. 79).

290. 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, ISSN: 1521-0464 (Electronic) 1071-7544 (Linking), DOI 10.1080/10717544.2018.1474964 (Jan. 2018) (cit. on p. 79).

291. R. Leenay, K. Maksimchuk, R. Slotkowski, *et al.*, Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Mol Cell* **62**, 137–47, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2016.02.031 (Apr. 2016) (cit. on pp. 82, 87, 159).

292. J.-J. Liu, N. Orlova, B. L. Oakes, *et al.*, CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* **566**, 218–223, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/s41586-019-0908-x (Feb. 2019) (cit. on pp. 82, 269).

293. 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, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking), DOI 10.1016/j.cell.2016.11.053 (Dec. 2016) (cit. on pp. 82, 264).

294. W. Chen, Y. Zhang, Y. Zhang, *et al.*, CRISPR/Cas9-based Genome Editing in Pseudomonas aeruginosa and Cytidine Deaminase-Mediated Base Editing in Pseudomonas Species. *iScience* **6**, 222–231, ISSN: 2589-0042, DOI 10.1016/j.isci.2018.07.024 (Aug. 2018) (cit. on pp. 82, 185, 275).

295. E. Ma, L. Harrington, M. O'Connell, K. Zhou, J. Doudna, Single-Stranded DNA Cleavage by Divergent CRISPR-Cas9 Enzymes. *Mol Cell* **60**, 398–407, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2015.10.030 (Nov. 2015) (cit. on pp. 82, 115, 116, 219).

296. C. Liao, F. Ttofali, R. A. Slotkowski, *et al.*, Modular one-pot assembly of CRISPR arrays enables library generation and reveals factors influencing crRNA biogenesis. *Nat Commun* **10**, 2948, ISSN: 2041-1723 (Electronic) 2041-1723 (Linking), DOI 10.1038/s41467-019-10747-3 (July 2019) (cit. on p. 86).

297. D. B. T. Cox, J. S. Gootenberg, O. O. Abudayyeh, *et al.*, RNA editing with CRISPR-Cas13. *Science* **358**, 1019–1027, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.aaq0180 (Oct. 2017) (cit. on p. 86).

298. 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, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature17946 (Apr. 2016) (cit. on pp. 86, 274).

299. P. Mohanraju, J. Oost, M. Jinek, D. Swarts, Heterologous Expression and Purification of the CRISPR-Cas12a/Cpf1 Protein. *Bio-protocol* **8**, e2842, ISSN: 2331-8325, DOI 10.21769/bioprotoc.2842 (2018) (cit. on p. 88).

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 Nucleic Acids* **25**, 171–189, ISSN: 1525-7770 (Print) 1525-7770 (Linking), DOI 10.1080/15257770500446915 (Jan. 2006) (cit. on p. 97).
301. K. S. Makarova, E. V. Koonin, Annotation and Classification of CRISPR-Cas Systems. *Methods Mol Biol* **1311**, 47–75, ISSN: 1940-6029 (Electronic) 1064-3745 (Linking), DOI 10.1007/978-1-4939-2687-9_4 (2015) (cit. on p. 111).
302. A. C. Komor, A. H. Badran, D. R. Liu, CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **168**, 20–36, ISSN: 0092-8674, DOI 10.1016/j.cell.2016.10.044 (Jan. 2017) (cit. on pp. 111, 121).
303. H. Puchta, Applying CRISPR/Cas for genome engineering in plants: the best is yet to come. *Current Opinion in Plant Biology* **36**, 1–8, ISSN: 1369-5266, DOI 10.1016/j.pbi.2016.11.011 (Apr. 2017) (cit. on pp. 111, 121).
304. X. Tang, L. G. Lowder, T. Zhang, *et al.*, A CRISPR–Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nat Plants* **3**, 17018, ISSN: 2055-0278 (Electronic) 2055-0278 (Linking), DOI 10.1038/nplants.2017.18 (Feb. 2017) (cit. on pp. 111, 121).
305. J. Xu, X. Ren, J. Sun, *et al.*, A Toolkit of CRISPR-Based Genome Editing Systems in Drosophila. *Journal of Genetics and Genomics* **42**, 141–149, ISSN: 1673-8527, DOI 10.1016/j.jgg.2015.02.007 (Apr. 2015) (cit. on pp. 111, 150).
306. Y. Jiang, F. Qian, J. Yang, *et al.*, CRISPR-Cpf1 assisted genome editing of Corynebacterium glutamicum. *Nature Communications* **8**, 15179, ISSN: 2041-1723, DOI 10.1038/ncomms15179 (May 2017) (cit. on pp. 111, 121, 147, 157, 159, 271).
307. M.-Y. Yan, H.-Q. Yan, G.-X. Ren, *et al.*, CRISPR-Cas12a-Assisted Recombineering in Bacteria. *Appl Environ Microbiol* **83**, ed. by V. Müller, ISSN: 1098-5336 (Electronic) 0099-2240 (Linking), DOI 10.1128/aem.00947-17 (June 2017) (cit. on pp. 111, 147, 157).
308. T. Karvelis, G. Gasiunas, J. Young, *et al.*, Rapid characterisation of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biology* **16**, 253, ISSN: 1474-760X, DOI 10.1186/s13059-015-0818-7 (Nov. 2015) (cit. on pp. 111, 113).
309. L. Lv, Y.-L. Ren, J.-C. Chen, Q. Wu, G.-Q. Chen, Application of CRISPRi for prokaryotic metabolic engineering involving multiple genes, a case study: Controllable P(3HB- co -4HB) biosynthesis. *Metabolic Engineering* **29**, 160–168, ISSN: 1096-7176, DOI 10.1016/j.ymben.2015.03.013 (May 2015) (cit. on p. 111).
310. E. Choudhary, P. Thakur, M. Pareek, N. Agarwal, Gene silencing by CRISPR interference in mycobacteria. *Nature Communications* **6**, 6267, ISSN: 2041-1723, DOI 10.1038/ncomms7267 (Feb. 2015) (cit. on p. 111).
311. E. F. Bosma, J. van der Oost, W. M. de Vos, R. van Kranenburg, Sustainable Production of Bio-Based Chemicals by Extremophiles. *Current Biotechnology* **2**, 360–379, ISSN: 2211-5501, DOI 10.2174/18722083113076660028 (Dec. 2013) (cit. on pp. 111, 118, 147, 175).
312. D. G. Olson, R. Sparling, L. R. Lynd, Ethanol production by engineered thermophiles. *Current Opinion in Biotechnology* **33**, 130–141, ISSN: 0958-1669, DOI 10.1016/j.copbio.2015.02.006 (June 2015) (cit. on pp. 111, 147).
313. M. P. Taylor, L. van Zyl, I. M. Tuffin, D. J. Leak, D. A. Cowan, Genetic tool development underpins recent advances in thermophilic whole-cell biocatalysts. *Microb. Biotechnol.* **4**, 438–448, ISSN: 1751-7915, DOI 10.1111/j.1751-7915.2010.00246.x (Feb. 2011) (cit. on pp. 111, 118).
314. B. M. Zeldes, M. W. Keller, A. J. Loder, *et al.*, Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals. *Frontiers in Microbiology* **6**, 1209, ISSN: 1664-302X, DOI 10.3389/fmicb.2015.01209 (Nov. 2015) (cit. on p. 111).
315. I. Mougiakos, E. F. Bosma, K. Weenink, *et al.*, Efficient Genome Editing of a Facultative Thermophile Using Mesophilic spCas9. *ACS Synthetic Biology* **6**, 849–861, ISSN: 2161-5063, DOI 10.1021/acssynbio.6b00339 (Feb. 2017) (cit. on pp. 111, 115, 118, 121, 122, 125, 141, 147, 148, 152, 155, 158, 159, 161, 162, 170, 176).
316. J. Wiktor, C. Lesterlin, D. J. Sherratt, C. Dekker, CRISPR-mediated control of the bacterial initiation of replication. *Nucleic acids research* **44**, 3801–3810, ISSN: 1362-4962, DOI 10.1093/nar/gkw214 (Apr. 2016) (cit. on pp. 111, 115, 121, 148).
317. M. J. A. Daas, A. H. P. van de Weijer, W. M. de Vos, J. van der Oost, R. van Kranenburg, Isolation of a genetically accessible thermophilic xylan degrading bacterium from compost. *Biotechnology for Biofuels* **9**, 210, ISSN: 1754-6834, DOI 10.1186/s13068-016-0618-7 (Oct. 2016) (cit. on pp. 111, 112, 121).
318. E. F. Bosma, A. H. P. van de Weijer, M. J. A. Daas, *et al.*, Isolation and Screening of Thermophilic Bacilli from Compost for Electrotransformation and Fermentation: Characterization of Bacillus smithii ET 138 as a New Biocatalyst. *Appl. Environ. Microbiol.* **81**, ed. by F. E. Löffler, 1874–1883, ISSN: 1098-5336, DOI 10.1128/aem.03640-14 (Jan. 2015) (cit. on pp. 111, 150).
319. Y. Li, S. Pan, Y. Zhang, *et al.*, Harnessing Type I and Type III CRISPR-Cas systems for genome editing. *Nucleic Acids Research* **44**, e34–e34, ISSN: 0305-1048, DOI 10.1093/nar/gkv1044 (Oct. 2016) (cit. on pp. 112, 121, 122).
320. M. J. A. Daas, B. Vriesendorp, A. H. P. van de Weijer, J. van der Oost, R. van Kranenburg, Complete Genome Sequence of Geobacillus thermodenitrificans T12, A Potential Host for Biotechnological Applications. *Current Microbiology* **75**, 49–56, DOI 10.1007/s00284-017-1349-0 (Sept. 2017) (cit. on p. 112).

321. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**, 845–858, ISSN: 1754-2189, DOI 10.1038/nprot.2015.053 (May 2015) (cit. on p. 113).

322. S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* **33**, 1870–1874, ISSN: 0737-4038, DOI 10.1093/molbev/msw054 (Mar. 2016) (cit. on pp. 113, 129).

323. A. Biswas, J. N. Gagnon, S. J. Brouns, P. C. Fineran, C. M. Brown, CRISPRTarget. *RNA Biology* **10**, 817–827, DOI 10.4161/rna.24046 (Mar. 2013) (cit. on pp. 112, 130).

324. S. Kim, D. Kim, S. W. Cho, J. Kim, J.-S. Kim, Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome research* **24**, 1012–9, ISSN: 1549-5469, DOI 10.1101/gr.171322.113 (Apr. 2014) (cit. on p. 113).

325. H. Chen, J. Choi, S. Bailey, Cut Site Selection by the Two Nuclease Domains of the Cas9 RNA-guided Endonuclease. *Journal of Biological Chemistry* **289**, 13284–13294, ISSN: 0021-9258, DOI 10.1074/jbc.m113.539726 (Mar. 2014) (cit. on p. 115).

326. B. M. A. Abdel-Banat, H. Hoshida, A. Ano, S. Nonklang, R. Akada, High-temperature fermentation: how can processes for ethanol production at high temperatures become superior to the traditional process using mesophilic yeast? *Appl. Microbiol. Biotechnol.* **85**, 861–867, ISSN: 1432-0614 (Electronic) 0175-7598 (Linking), DOI 10.1007/s00253-009-2248-5 (Oct. 2010) (cit. on pp. 118, 147, 150).

327. A. Bhalla, N. Bansal, S. Kumar, K. M. Bischoff, R. K. Sani, Improved lignocellulose conversion to biofuels with thermophilic bacteria and thermostable enzymes. *Bioresource Technology* **128**, 751–759, ISSN: 0960-8524, DOI 10.1016/j.biortech.2012.10.145 (Jan. 2013) (cit. on p. 118).

328. E. F. Bosma, A. H. P. van de Weijer, L. van der Vlist, *et al.*, Establishment of markerless gene deletion tools in thermophilic Bacillus smithii and construction of multiple mutant strains. *Microb Cell Fact* **14**, 99, ISSN: 1475-2859 (Electronic) 1475-2859 (Linking), DOI 10.1186/s12934-015-0286-5 (July 2015) (cit. on pp. 118, 120–122, 125, 150, 159, 161, 163).

329. I. Mougiakos, P. Mohanraju, E. F. Bosma, *et al.*, Characterizing a thermostable Cas9 for bacterial genome editing and silencing. *Nat Commun* **8**, 1647, ISSN: 2041-1723 (Electronic) 2041-1723 (Linking), DOI 10.1038/s41467-017-01591-4 (Nov. 2017) (cit. on pp. 118, 157, 158, 177, 270).

330. J.-H. Oh, J.-P. van Pijkeren, CRISPR–Cas9-assisted recombineering in Lactobacillus reuteri. *Nucleic acids research* **42**, e131–e131, ISSN: 1362-4962, DOI 10.1093/nar/gku623 (July 2014) (cit. on pp. 118, 150, 157).

331. P. I. Nikel, M. Chavarría, A. Danchin, V. de Lorenzo, From dirt to industrial applications: Pseudomonas putida as a Synthetic Biology chassis for hosting harsh biochemical reactions. *Current Opinion in Chemical Biology* **34**, 20–29, ISSN: 1367-5931, DOI 10.1016/j.cbpa.2016.05.011 (Oct. 2016) (cit. on p. 120).

332. I. Poblete-Castro, J. Becker, K. Dohnt, V. M. dos Santos, C. Wittmann, Industrial biotechnology of Pseudomonas putida and related species. *Applied Microbiology and Biotechnology* **93**, 2279–2290, ISSN: 0175-7598, DOI 10.1007/s00253-012-3928-0 (Feb. 2012) (cit. on p. 120).

333. T. Aparicio, S. I. Jensen, A. T. Nielsen, V. de Lorenzo, E. Martínez-García, The Ssr protein (T1E_1405) from Pseudomonas putida DOT-T1E enables oligonucleotide-based recombineering in platform strain P. putida EM42. *Biotechnology Journal* **11**, 1309–1319, ISSN: 1860-6768, DOI 10.1002/biot.201600317 (July 2016) (cit. on p. 120).

334. E. Martínez-García, V. de Lorenzo, The quest for the minimal bacterial genome. *Current Opinion in Biotechnology* **42**, 216–224, ISSN: 0958-1669, DOI 10.1016/j.copbio.2016.09.001 (Dec. 2016) (cit. on p. 120).

335. M. H. Larson, L. A. Gilbert, X. Wang, *et al.*, CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature Protocols* **8**, 2180–2196, ISSN: 1754-2189, DOI 10.1038/nprot.2013.132 (Oct. 2013) (cit. on p. 120).

336. T. Xu, Y. Li, Z. Shi, *et al.*, Efficient Genome Editing in Clostridium cellulolyticum via CRISPR-Cas9 Nickase. *Applied and Environmental Microbiology* **81**, ed. by A. M. Spormann, 4423–4431, ISSN: 0099-2240, DOI 10.1128/aem.00873-15 (Apr. 2015) (cit. on p. 121).

337. H. Yang, D. J. Patel, Inhibition Mechanism of an Anti-CRISPR Suppressor AcrIIA4 Targeting SpyCas9. *Molecular Cell* **67**, 117–127.e5, ISSN: 1097-2765, DOI 10.1016/j.molcel.2017.05.024 (July 2017) (cit. on pp. 121, 176).

338. A. D. Weinberger, Y. I. Wolf, A. E. Lobkovsky, M. S. Gilmore, E. V. Koonin, Viral Diversity Threshold for Adaptive Immunity in Prokaryotes. *mBio* **3**, ed. by S. J. Giovannoni, e00456–12, ISSN: 2150-7511, DOI 10.1128/mbio.00456-12 (Dec. 2012) (cit. on p. 121).

339. J. Sambrook, E. Fritsch, T. Maniatis, *Molecular cloning : a laboratory manual* (Cold Spring Harbor Laboratory Pr, 1989), ISBN: 0879693096 (cit. on p. 122).

340. G. E. Crooks, WebLogo: A Sequence Logo Generator. *Genome Research* **14**, 1188–1190, ISSN: 1088-9051, DOI 10.1101/gr.849004 (May 2004) (cit. on pp. 124, 130).

341. K.-H. Choi, A. Kumar, H. P. Schweizer, A 10-min method for preparation of highly electrocompetent Pseudomonas aeruginosa cells: Application for DNA fragment transfer between chromosomes and plasmid trans-

formation. *Journal of Microbiological Methods* **64**, 391–397, ISSN: 0167-7012, DOI 10.1016/j.mimet.2005.06.001 (Mar. 2006) (cit. on p. 126).

342. S. A. van Hijum, A. de Jong, R. J. Baerends, *et al.*, A generally applicable validation scheme for the assessment of factors involved in reproducibility and quality of DNA-microarray data. *BMC genomics* **6**, 77, ISSN: 1471-2164, DOI 10.1186/1471-2164-6-77 (2005) (cit. on p. 126).
343. M. Larkin, G. Blackshields, N. Brown, *et al.*, Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–8, ISSN: 1367-4811 (Electronic) 1367-4803 (Linking), DOI 10.1093/bioinformatics/btm404 (Sept. 2007) (cit. on p. 129).
344. P. Gouet, E. Courcelle, D. Stuart, F. Metoz, ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308, ISSN: 1367-4803 (Print) 1367-4803 (Linking), DOI 10.1093/bioinformatics/15.4.305 (Apr. 1999) (cit. on p. 129).
345. E. Martínez-García, V. de Lorenzo, Engineering multiple genomic deletions in Gram-negative bacteria: analysis of the multi-resistant antibiotic profile of Pseudomonas putida KT2440. *Environ Microbiol* **13**, 2702–16, ISSN: 1462-2920 (Electronic) 1462-2912 (Linking), DOI 10.1111/j.1462-2920.2011.02538.x (Aug. 2011) (cit. on p. 141).
346. U. Bonnet, M. Banger, F. M. Leweke, *et al.*, Treatment of Acute Alcohol Withdrawal With Gabapentin: Results From a Controlled Two-Center Trial. *J Clin Psychopharmacol* **23**, 514–519, ISSN: 0271-0749 (Print) 0271-0749 (Linking), DOI 10.1097/01.jcp.0000088905.24613.ad (Oct. 2003) (cit. on p. 147).
347. M. P. Taylor, K. L. Eley, S. Martin, *et al.*, Thermophilic ethanologenesis: future prospects for second-generation bioethanol production. *Trends Biotechnol.* **27**, 398–405, ISSN: 1879-3096 (Electronic) 0167-7799 (Linking), DOI 10.1016/j.tibtech.2009.03.006 (July 2009) (cit. on p. 147).
348. P. K. R. Kambam, M. A. Henson, Engineering bacterial processes for cellulosic ethanol production. *Biofuels* **1**, 729–743, ISSN: 1759-7269, DOI 10.4155/bfs.10.46 (Sept. 2010) (cit. on p. 147).
349. M. Beckner, M. Ivey, T. Phister, Microbial contamination of fuel ethanol fermentations. *Lett. Appl. Microbiol.* **53**, 387–394, ISSN: 1472-765X (Electronic) 0266-8254 (Linking), DOI 10.1111/j.1472-765X.2011.03124.x (Aug. 2011) (cit. on p. 147).
350. J. Qin, B. Zhao, X. Wang, *et al.*, Non-Sterilized Fermentative Production of Polymer-Grade L-Lactic Acid by a Newly Isolated Thermophilic Strain Bacillus sp. 2–6. *PLoS One* **4**, ed. by M. Isalan, e4359, ISSN: 1932-6203 (Electronic) 1932-6203 (Linking), DOI 10.1371/journal.pone.0004359 (Feb. 2009) (cit. on p. 147).
351. K. Ma, T. Maeda, H. You, Y. Shirai, Open fermentative production of l-lactic acid with high optical purity by thermophilic Bacillus coagulans using excess sludge as nutrient. *Bioresour. Technol.* **151**, 28–35, ISSN: 0960-8524, DOI 10.1016/j.biortech.2013.10.022 (Jan. 2014) (cit. on p. 147).
352. M. S. Ou, N. Mohammed, L. O. Ingram, K. T. Shanmugam, Thermophilic Bacillus coagulans Requires Less Cellulases for Simultaneous Saccharification and Fermentation of Cellulose to Products than Mesophilic Microbial Biocatalysts. *Applied Biochemistry and Biotechnology* **155**, 76–82, ISSN: 0273-2289, DOI 10.1007/s12010-008-8509-4 (Jan. 2009) (cit. on p. 147).
353. C. Weber, A. Farwick, F. Benisch, *et al.*, Trends and challenges in the microbial production of lignocellulosic bioalcohol fuels. *Applied Microbiology and Biotechnology* **87**, 1303–1315, ISSN: 0175-7598, DOI 10.1007/s00253-010-2707-z (June 2010) (cit. on p. 147).
354. H. Li, C. R. Shen, C.-H. Huang, *et al.*, CRISPR-Cas9 for the genome engineering of cyanobacteria and succinate production. *Metabolic Engineering* **38**, 293–302, DOI 10.1016/j.ymben.2016.09.006 (Nov. 2016) (cit. on pp. 147, 271).
355. K. E. Wendt, J. Ungerer, R. E. Cobb, H. Zhao, H. B. Pakrasi, CRISPR/Cas9 mediated targeted mutagenesis of the fast growing cyanobacterium Synechococcus elongatus UTEX 2973. *Microb Cell Fact* **15**, 115, ISSN: 1475-2859 (Electronic) 1475-2859 (Linking), DOI 10.1186/s12934-016-0514-7 (June 2016) (cit. on pp. 147, 271).
356. L. Li, K. Wei, G. Zheng, *et al.*, CRISPR-Cpf1-Assisted Multiplex Genome Editing and Transcriptional Repression in Streptomyces. *Appl Environ Microbiol* **84**, ed. by R. M. Kelly, ISSN: 1098-5336 (Electronic) 0099-2240 (Linking), DOI 10.1128/AEM.00827-18 (July 2018) (cit. on pp. 147, 157, 159, 270).
357. J. Ungerer, H. B. Pakrasi, Cpf1 Is A Versatile Tool for CRISPR Genome Editing Across Diverse Species of Cyanobacteria. *Scientific Reports* **6**, ISSN: 2045-2322, DOI Artn3968110.1038/Srep39681 (2016) (cit. on p. 147).
358. W. Hong, J. Zhang, G. Cui, L. Wang, Y. Wang, Multiplexed CRISPR-Cpf1-Mediated Genome Editing in Clostridium difficile toward the Understanding of Pathogenesis of C. difficile Infection. *Acs Synthetic Biology* **7**, 1588–1600, ISSN: 2161-5063, DOI 10.1021/acssynbio.8b00087 (June 2018) (cit. on p. 147).
359. W. Strutz, Exploring Protein Stability by NanoDSF. *Biophysical Journal* **110**, 393a–393a, ISSN: 0006-3495, DOI DOI10.1016/j.bpj.2015.11.2126 (2016) (cit. on p. 149).
360. A. T. Kovacs, M. van Hartskamp, O. P. Kuipers, R. van Kranenburg, Genetic Tool Development for a New Host for Biotechnology, the Thermotolerant Bacterium Bacillus coagulans. *Appl. Environ. Microbiol.* **76**, 4085–4088, DOI 10.1128/aem.03060-09 (Apr. 2010) (cit. on p. 150).
361. Y. Wang, Z.-T. Zhang, S.-O. Seo, *et al.*, Bacterial Genome Editing with CRISPR-Cas9: Deletion, Integration, Single Nucleotide Modification, and Desirable "Clean" Mutant Selection in Clostridium beijerinckii as an Example.

ACS Synthetic Biology **5**, 721–732, ISSN: 2161-5063, DOI 10.1021/acssynbio.6b00060 (Apr. 2016) (cit. on pp. 150, 176).

362. H. Suzuki, Genetic Transformation of Geobacillus kaustophilus HTA426 by Conjugative Transfer of Host-Mimicking Plasmids. *J Microbiol Biotechnol* **22**, 1279–1287, ISSN: 1738-8872 (Electronic) 1017-7825 (Linking), DOI 10.4014/jmb.1203.03023 (Sept. 2012) (cit. on p. 151).
363. S. A. Tripathi, D. G. Olson, D. A. Argyros, *et al.*, Development of pyrF-Based Genetic System for Targeted Gene Deletion in Clostridium thermocellum and Creation of a pta Mutant. *Appl. Environ. Microbiol.* **76**, 6591–6599, DOI 10.1128/aem.01484-10 (Aug. 2010) (cit. on p. 151).
364. A. KITA, Y. IWASAKI, S. YANO, *et al.*, Isolation of Thermophilic Acetogens and Transformation of Them with the pyrF and kan(r) Genes. *Biosci., Biotechnol., Biochem.* **77**, 301–306, DOI 10.1271/bbb.120720 (2013) (cit. on p. 151).
365. H. K. Kim, M. Song, J. Lee, *et al.*, In vivo high-throughput profiling of CRISPR–Cpf1 activity. *Nat Methods* **14**, 153–159, ISSN: 1548-7105 (Electronic) 1548-7091 (Linking), DOI 10.1038/nmeth.4104 (Dec. 2017) (cit. on pp. 152, 159).
366. H. K. Kim, S. Min, M. Song, *et al.*, Deep learning improves prediction of CRISPR–Cpf1 guide RNA activity. *Nat Biotechnol* **36**, 239–241, ISSN: 1546-1696 (Electronic) 1087-0156 (Linking), DOI 10.1038/nbt.4061 (Jan. 2018) (cit. on p. 152).
367. J. G. Doench, N. Fusi, M. Sullender, *et al.*, Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184–191, ISSN: 1546-1696 (Electronic) 1087-0156 (Linking), DOI 10.1038/nbt.3437 (Jan. 2016) (cit. on p. 152).
368. L. Cui, D. Bikard, Consequences of Cas9 cleavage in the chromosome ofEscherichia coli. *Nucleic Acids Res* **44**, 4243–4251, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gkw223 (Apr. 2016) (cit. on pp. 152, 175, 270, 274).
369. R. Barrangou, J.-P. van Pijkeren, Exploiting CRISPR–Cas immune systems for genome editing in bacteria. *Current Opinion in Biotechnology* **37**, 61–68, ISSN: 0958-1669, DOI 10.1016/j.copbio.2015.10.003 (Feb. 2016) (cit. on pp. 157, 175).
370. T. Jakočiūnas, M. K. Jensen, J. D. Keasling, CRISPR/Cas9 advances engineering of microbial cell factories. *Metab Eng* **34**, 44–59, ISSN: 1096-7184 (Electronic) 1096-7176 (Linking), DOI 10.1016/j.ymben.2015.12.003 (Mar. 2016) (cit. on p. 157).
371. R. Bowater, A. J. Doherty, Making Ends Meet: Repairing Breaks in Bacterial DNA by Non-Homologous End-Joining. *PLoS Genet* **2**, e8, ISSN: 1553-7404 (Electronic) 1553-7390 (Linking), DOI 10.1371/journal.pgen.0020008 (2006) (cit. on pp. 159, 175, 270, 274).
372. E. F. Bosma, J. J. Koehorst, S. A. F. T. van Hijum, *et al.*, Complete genome sequence of thermophilic Bacillus smithii type strain DSM 4216T. *Stand Genomic Sci* **11**, 52, ISSN: 1944-3277 (Print) 1944-3277 (Linking), DOI 10.1186/s40793-016-0172-8 (Aug. 2016) (cit. on p. 159).
373. S. Shuman, M. S. Glickman, Bacterial DNA repair by non-homologous end joining. *Nat Rev Microbiol* **5**, 852–861, ISSN: 1740-1534 (Electronic) 1740-1526 (Linking), DOI 10.1038/nrmicro1768 (Nov. 2007) (cit. on p. 159).
374. S. K. Kim, H. Kim, W.-C. Ahn, *et al.*, Efficient Transcriptional Gene Repression by Type V-A CRISPR-Cpf1 fromEubacterium eligens. *ACS Synthetic Biology* **6**, 1273–1282, DOI 10.1021/acssynbio.6b00368 (Apr. 2017) (cit. on p. 159).
375. C. Merryman, D. G. Gibson, Methods and applications for assembling large DNA constructs. *Metab Eng* **14**, 196–204, ISSN: 1096-7184 (Electronic) 1096-7176 (Linking), DOI 10.1016/j.ymben.2012.02.005 (May 2012) (cit. on p. 159).
376. X. Ao, Y. Yao, T. Li, *et al.*, A Multiplex Genome Editing Method for Escherichia coli Based on CRISPR-Cas12a. *Front Microbiol* **9**, 2307, ISSN: 1664-302X (Print) 1664-302X (Linking), DOI 10.3389/fmicb.2018.02307 (Oct. 2018) (cit. on p. 159).
377. B. Adiego-Pérez, P. Randazzo, J. M. Daran, *et al.*, Multiplex genome editing of microorganisms using CRISPR-Cas. *FEMS Microbiology Letters* **366**, ISSN: 0378-1097, DOI 10.1093/femsle/fnz086 (Apr. 2019) (cit. on p. 159).
378. B. X. H. Fu, J. D. Smith, R. T. Fuchs, *et al.*, Target-dependent nickase activities of the CRISPR–Cas nucleases Cpf1 and Cas9. *Nat Microbiol* **4**, 888–897, ISSN: 2058-5276 (Electronic) 2058-5276 (Linking), DOI 10.1038/s41564-019-0382-0 (Mar. 2019) (cit. on p. 160).
379. L. Greenough, K. M. Schermerhorn, L. Mazzola, *et al.*, Adapting capillary gel electrophoresis as a sensitive, high-throughput method to accelerate characterisation of nucleic acid metabolic enzymes. *Nucleic Acids Research* **44**, ISSN: 0305-1048, DOI ARTNe1510.1093/nar/gkv899 (2016) (cit. on p. 160).
380. A. D. Haimovich, P. Muir, F. J. Isaacs, Genomes by design. *Nature Reviews Genetics* **16**, 501–516, ISSN: 1471-0056 1471-0064, DOI 10.1038/nrg3956 (Aug. 2015) (cit. on p. 175).
381. J. M. Reyrat, V. Pelicic, B. Gicquel, R. Rappuoli, Counterselectable markers: untapped tools for bacterial genetics and pathogenesis. *Infect Immun* **66**, 4011–7, ISSN: 0019-9567 (Print) 0019-9567 (Linking) (1998) (cit. on p. 175).

382. H. P. Schweizer, Bacterial genetics: past achievements, present state of the field, and future challenges. *BioTechniques* **44**, 633–641, ISSN: 0736-6205 1940-9818, DOI 10.2144/000112807 (Apr. 2008) (cit. on p. 175).
383. G. Pines, E. F. Freed, J. D. Winkler, R. T. Gill, Bacterial Recombineering: Genome Engineering via Phage-Based Homologous Recombination. *ACS Synthetic Biology* **4**, 1176–1185, ISSN: 2161-5063, DOI 10.1021/acssynbio.5b00009 (Apr. 2015) (cit. on p. 175).
384. C. L. Frazier, J. S. Filippo, A. M. Lambowitz, D. A. Mills, Genetic Manipulation of Lactococcus lactis by Using Targeted Group II Introns: Generation of Stable Insertions without Selection. *Applied and Environmental Microbiology* **69**, 1121–1128, ISSN: 0099-2240, DOI 10.1128/aem.69.2.1121-1128.2003 (Feb. 2003) (cit. on p. 175).
385. J. Yao, A. M. Lambowitz, Gene Targeting in Gram-Negative Bacteria by Use of a Mobile Group II Intron ("Targetron") Expressed from a Broad-Host-Range Vector. *Applied and Environmental Microbiology* **73**, 2735–2743, ISSN: 0099-2240, DOI 10.1128/aem.02829-06 (Feb. 2007) (cit. on p. 175).
386. Y. Wang, X. Li, C. B. Milne, *et al.*, Development of a Gene Knockout System Using Mobile Group II Introns (Targetron) and Genetic Disruption of Acid Production Pathways in Clostridium beijerinckii. *Applied and Environmental Microbiology* **79**, 5853–5863, ISSN: 0099-2240 1098-5336, DOI 10.1128/aem.00971-13 (July 2013) (cit. on p. 175).
387. P. D. Donohoue, R. Barrangou, A. P. May, Advances in Industrial Biotechnology Using CRISPR-Cas Systems. *Trends in Biotechnology* **36**, 134–146, ISSN: 0167-7799, DOI 10.1016/j.tibtech.2017.07.007 (Feb. 2018) (cit. on pp. 175, 185, 270).
388. 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. *Current Opinion in Biotechnology* **50**, 146–157, ISSN: 0958-1669, DOI 10.1016/j.copbio.2018.01.002 (Apr. 2018) (cit. on pp. 175, 185).
389. E. Stout, T. Klaenhammer, R. Barrangou, CRISPR-Cas Technologies and Applications in Food Bacteria. *Annual Review of Food Science and Technology* **8**, 413–437, ISSN: 1941-1413 1941-1421, DOI 10.1146/annurev-food-072816-024723 (Feb. 2017) (cit. on p. 175).
390. L. Aravind, Prokaryotic Homologs of the Eukaryotic DNA-End-Binding Protein Ku, Novel Domains in the Ku Protein and Prediction of a Prokaryotic Double-Strand Break Repair System. *Genome Research* **11**, 1365–1374, ISSN: 1088-9051, DOI 10.1101/gr.181001 (Aug. 2001) (cit. on p. 175).
391. G. Iliakis, H. Wang, A. Perrault, *et al.*, Mechanisms of DNA double strand break repair and chromosome aberration formation. *Cytogenetic and Genome Research* **104**, 14–20, ISSN: 1424-8581 1424-859X, DOI 10.1159/000077461 (2004) (cit. on p. 175).
392. W. Jiang, D. Bikard, D. Cox, F. Zhang, L. A. Marraffini, RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology* **31**, 233–239, ISSN: 1087-0156, DOI 10.1038/nbt.2508 (Jan. 2013) (cit. on pp. 175, 270, 274).
393. Q. Li, J. Chen, N. P. Minton, *et al.*, CRISPR-based genome editing and expression control systems in Clostridium acetobutylicum and Clostridium beijerinckii. *Biotechnology Journal* **11**, 961–972, ISSN: 1860-6768, DOI 10.1002/biot.201600053 (June 2016) (cit. on p. 175).
394. H. Huang, C. Chai, N. Li, *et al.*, CRISPR/Cas9-Based Efficient Genome Editing in Clostridium ljungdahlii, an Autotrophic Gas-Fermenting Bacterium. *ACS Synthetic Biology* **5**, 1355–1361, ISSN: 2161-5063, DOI 10.1021/acssynbio.6b00044 (June 2016) (cit. on p. 175).
395. Y. Li, Z. Lin, C. Huang, *et al.*, Metabolic engineering of Escherichia coli using CRISPR–Cas9 meditated genome editing. *Metabolic Engineering* **31**, 13–21, ISSN: 1096-7176, DOI 10.1016/j.ymben.2015.06.006 (Sept. 2015) (cit. on p. 176).
396. Y. Tong, P. Charusanti, L. Zhang, T. Weber, S. Y. Lee, CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. *Acs Synthetic Biology* **4**, 1020–1029, ISSN: 2161-5063, DOI 10.1021/acssynbio.5b00038 (Apr. 2015) (cit. on pp. 176, 270).
397. Y. Jiang, B. Chen, C. Duan, *et al.*, Multigene Editing in the Escherichia coli Genome via the CRISPR-Cas9 System. *Applied and Environmental Microbiology* **81**, ed. by R. M. Kelly, 2506–2514, ISSN: 0099-2240 1098-5336, DOI 10.1128/aem.04023-14 (Jan. 2015) (cit. on pp. 176, 274).
398. Q. Yan, S. S. Fong, Challenges and Advances for Genetic Engineering of Non-model Bacteria and Uses in Consolidated Bioprocessing. *Frontiers in Microbiology* **8**, ISSN: 1664-302X, DOI 10.3389/fmicb.2017.02060 (Oct. 2017) (cit. on p. 176).
399. F. Richter, I. Fonfara, B. Bouazza, *et al.*, Engineering of temperature- and light-switchable Cas9 variants. *Nucleic Acids Research*, gkw930, ISSN: 0305-1048 1362-4962, DOI 10.1093/nar/gkw930 (Oct. 2016) (cit. on p. 176).
400. L. E. Dow, J. Fisher, K. P. O'Rourke, *et al.*, Inducible in vivo genome editing with CRISPR-Cas9. *Nature Biotechnology* **33**, 390–394, ISSN: 1087-0156 1546-1696, DOI 10.1038/nbt.3155 (Feb. 2015) (cit. on p. 176).
401. L. R. Polstein, C. A. Gersbach, A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nature Chemical Biology* **11**, 198–200, ISSN: 1552-4450 1552-4469, DOI 10.1038/nchembio.1753 (Feb. 2015) (cit. on p. 176).

402. Y. Nihongaki, F. Kawano, T. Nakajima, M. Sato, Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nature Biotechnology* **33**, 755–760, ISSN: 1087-0156 1546-1696, DOI 10.1038/nbt.3245 (June 2015) (cit. on p. 176).

403. A. V. Wright, S. H. Sternberg, D. W. Taylor, *et al.*, Rational design of a split-Cas9 enzyme complex. *Proceedings of the National Academy of Sciences* **112**, 2984–2989, ISSN: 0027-8424 1091-6490, DOI 10.1073/pnas.1501698112 (Feb. 2015) (cit. on p. 176).

404. B. Zetsche, S. E. Volz, F. Zhang, A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat Biotechnol* **33**, 139–142, ISSN: 1546-1696 (Electronic) 1087-0156 (Linking), DOI 10.1038/nbt.3149 (Feb. 2015) (cit. on p. 176).

405. J. K. Nuñez, L. B. Harrington, J. A. Doudna, Chemical and Biophysical Modulation of Cas9 for Tunable Genome Engineering. *ACS Chemical Biology* **11**, 681–688, ISSN: 1554-8929 1554-8937, DOI 10.1021/acschembio.5b01019 (Feb. 2016) (cit. on p. 176).

406. J. Cao, L. Wu, S.-M. Zhang, *et al.*, An easy and efficient inducible CRISPR/Cas9 platform with improved specificity for multiple gene targeting. *Nucleic Acids Research*, gkw660, ISSN: 0305-1048 1362-4962, DOI 10.1093/nar/gkw660 (July 2016) (cit. on p. 176).

407. K. I. Liu, M. N. B. Ramli, C. W. A. Woo, *et al.*, A chemical-inducible CRISPR–Cas9 system for rapid control of genome editing. *Nature Chemical Biology* **12**, 980–987, ISSN: 1552-4450 1552-4469, DOI 10.1038/nchembio.2179 (Sept. 2016) (cit. on p. 176).

408. A. Pawluk, N. Amrani, Y. Zhang, *et al.*, Naturally Occurring Off-Switches for CRISPR-Cas9. *Cell* **167**, 1829–1838.e9, ISSN: 0092-8674, DOI 10.1016/j.cell.2016.11.017 (Dec. 2016) (cit. on p. 176).

409. A. P. Hynes, G. M. Rousseau, M.-L. Lemay, *et al.*, An anti-CRISPR from a virulent streptococcal phage inhibits Streptococcus pyogenes Cas9. *Nature Microbiology* **2**, 1374–1380, ISSN: 2058-5276, DOI 10.1038/s41564-017-0004-7 (Aug. 2017) (cit. on p. 176).

410. F. He, Y. Bhoobalan-Chitty, L. B. Van, *et al.*, Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype I-D immunity. *Nature Microbiology* **3**, 461–469, ISSN: 2058-5276, DOI 10.1038/s41564-018-0120-z (Mar. 2018) (cit. on p. 176).

411. K. E. Watters, C. Fellmann, H. B. Bai, S. M. Ren, J. A. Doudna, Systematic discovery of natural CRISPR-Cas12a inhibitors. *Science* **362**, 236–239, ISSN: 0036-8075 1095-9203, DOI 10.1126/science.aau5138 (Sept. 2018) (cit. on p. 176).

412. S. Chowdhury, J. Carter, M. Rollins, *et al.*, *Cryo EM structure of anti-CRISPRs, AcrF1 and AcrF2, bound to type I-F crRNA-guided CRISPR surveillance complex*, Figure, Apr. 2017, DOI 10.2210/pdb5uz9/pdb (cit. on p. 176).

413. R. Peng, Y. Xu, T. Zhu, *et al.*, Alternate binding modes of anti-CRISPR viral suppressors AcrF1/2 to Csy surveillance complex revealed by cryo-EM structures. *Cell Research* **27**, 853–864, ISSN: 1001-0602 1748-7838, DOI 10.1038/cr.2017.79 (June 2017) (cit. on p. 176).

414. D. Dong, M. Guo, S. Wang, *et al.*, Structural basis of CRISPR–SpyCas9 inhibition by an anti-CRISPR protein. *Nature* **546**, 436–439, ISSN: 0028-0836 1476-4687, DOI 10.1038/nature22377 (Apr. 2017) (cit. on p. 176).

415. J. Shin, F. Jiang, J.-J. Liu, *et al.*, Disabling Cas9 by an anti-CRISPR DNA mimic. *Science Advances* **3**, e1701620, ISSN: 2375-2548, DOI 10.1126/sciadv.1701620 (July 2017) (cit. on p. 176).

416. I. Kim, M. Jeong, D. Ka, *et al.*, Solution structure and dynamics of anti-CRISPR AcrIIA4, the Cas9 inhibitor. *Scientific Reports* **8**, ISSN: 2045-2322, DOI 10.1038/s41598-018-22177-0 (Mar. 2018) (cit. on p. 176).

417. L. B. Harrington, K. W. Doxzen, E. Ma, *et al.*, A Broad-Spectrum Inhibitor of CRISPR-Cas9. *Cell* **170**, 1224–1233.e15, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking), DOI 10.1016/j.cell.2017.07.037 (Sept. 2017) (cit. on pp. 176–178).

418. E. M. Basgall, S. C. Goetting, M. E. Goeckel, *et al.*, Gene drive inhibition by the anti-CRISPR proteins AcrIIA2 and AcrIIA4 in Saccharomyces cerevisiae. *Microbiology* **164**, 464–474, ISSN: 1350-0872 1465-2080, DOI 10.1099/mic.0.000635 (Apr. 2018) (cit. on p. 176).

419. L. B. Harrington, D. Paez-Espino, B. T. Staahl, *et al.*, A thermostable Cas9 with increased lifetime in human plasma. *Nature Communications* **8**, ISSN: 2041-1723, DOI 10.1038/s41467-017-01408-4 (Nov. 2017) (cit. on pp. 177, 268, 270).

420. S. Banno, K. Nishida, T. Arazoe, H. Mitsunobu, A. Kondo, Deaminase-mediated multiplex genome editing in Escherichia coli. *Nature Microbiology* **3**, 423–429, ISSN: 2058-5276, DOI 10.1038/s41564-017-0102-6 (Feb. 2018) (cit. on pp. 185, 274, 275).

421. K. Zheng, Y. Wang, N. Li, *et al.*, Highly efficient base editing in bacteria using a Cas9-cytidine deaminase fusion. *Communications Biology* **1**, ISSN: 2399-3642, DOI 10.1038/s42003-018-0035-5 (Apr. 2018) (cit. on p. 185).

422. T. Gu, S. Zhao, Y. Pi, *et al.*, Highly efficient base editing in Staphylococcus aureus using an engineered CRISPR RNA-guided cytidine deaminase. *Chemical Science* **9**, 3248–3253, ISSN: 2041-6520 2041-6539, DOI 10.1039/c8sc00637g (2018) (cit. on p. 185).

423. Y. Wang, Y. Liu, J. Liu, *et al.*, MACBETH: Multiplex automated Corynebacterium glutamicum base editing method. *Metabolic Engineering* **47**, 200–210, ISSN: 1096-7176, DOI 10.1016/j.ymben.2018.02.016 (May 2018) (cit. on p. 185).

424. R. Green, E. J. Rogers, in *Methods in Enzymology* (Elsevier, 2013), pp. 329–336, ISBN: 0076-6879 9780124186873, DOI 10.1016/b978-0-12-418687-3.00028-8 (cit. on pp. 186–188).
425. C. Chumduri, R. K. Gurumurthy, R. Zietlow, T. F. Meyer, Subversion of host genome integrity by bacterial pathogens. *Nature Reviews Molecular Cell Biology* **17**, 659–673, DOI 10.1038/nrm.2016.100 (Aug. 2016) (cit. on pp. 213, 215).
426. C. Schwechheimer, M. J. Kuehn, Outer-membrane vesicles from Gram-negative bacteria: biogenesis and functions. *Nature reviews microbiology* **13**, 605–619, DOI 10.1038/nrmicro3525 (Sept. 2015) (cit. on p. 213).
427. J. Parkhill, B. W. Wren, K. Mungall, *et al.*, The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences. *Nature* **403**, 665–668, DOI 10.1038/35001088 (Feb. 2000) (cit. on p. 213).
428. K.-S. Jang, M. J. Sweredoski, R. L. Graham, S. Hess, W. M. Clemons, Comprehensive proteomic profiling of outer membrane vesicles from Campylobacter jejuni. *Journal of proteomics* **98**, 90–98, ISSN: 1874-3919, DOI 10.1016/j.jprot.2013.12.014 (Feb. 2014) (cit. on pp. 213, 220).
429. M. Gilbert, P. C. R. Godschalk, M.-F. Karwaski, *et al.*, Evidence for Acquisition of the Lipooligosaccharide Biosynthesis Locus in Campylobacter jejuni GB11, a Strain Isolated from a Patient with Guillain-Barre Syndrome, by Horizontal Exchange. *Infect Immun* **72**, 1162–1165, ISSN: 0019-9567 (Print) 0019-9567 (Linking), DOI 10.1128/iai.72.2.1162-1165.2004 (Jan. 2004) (cit. on pp. 213, 220).
430. K. T. Young, L. M. Davis, V. J. Dirita, Campylobacter jejuni: molecular biology and pathogenesis. *Nat Rev Microbiol* **5**, 665–679, ISSN: 1740-1534 (Electronic) 1740-1526 (Linking), DOI nrmicro1718[pii]10.1038/nrmicro1718 (2007) (cit. on p. 213).
431. I. Fonfara, A. L. Rhun, K. Chylinski, *et al.*, Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic acids research* **42**, 2577–2590, ISSN: 1362-4962, DOI 10.1093/nar/gkt1074 (Nov. 2013) (cit. on pp. 213, 214, 216).
432. E. C. LaCasse, Y. A. Lefebvre, Nuclear localization signals overlap DNA-or RNA-binding domains in nucleic acid-binding proteins. *Nucleic Acids Research* **23**, 1647–1656, DOI 10.1093/nar/23.10.1647 (May 1995) (cit. on p. 214).
433. M. K. Hibbard, R. M. Sandri-Goldin, Arginine-rich regions succeeding the nuclear localization region of the herpes simplex virus type 1 regulatory protein ICP27 are required for efficient nuclear localization and late gene expression. *Journal of Virology* **69**, 4656–4667, ISSN: 0022-538X (Print) (1995) (cit. on p. 214).
434. S. Panier, S. J. Boulton, Double-strand break repair: 53BP1 comes into focus. *Nature reviews Molecular cell biology* **15**, 7–18, ISSN: 1471-0080, DOI 10.1038/nrm3719 (Dec. 2014) (cit. on p. 214).
435. E. P. Rogakou, C. Boon, C. Redon, W. M. Bonner, Megabase Chromatin Domains Involved in DNA Double-Strand Breaks in Vivo. *The Journal of cell biology* **146**, 905–916, ISSN: 0021-9525 (Print)0021-9525 (Linking), DOI 10.1083/jcb.146.5.905 (Sept. 1999) (cit. on p. 214).
436. C. P. A. Skarp, O. Akinrinade, A. J. E. Nilsson, *et al.*, Comparative genomics and genome biology of invasive Campylobacter jejuni. *Scientific reports* **5**, 17300–17300, ISSN: 2045-2322, DOI 10.1038/srep17300 (Nov. 2015) (cit. on p. 215).
437. G. Iraola, R. Pérez, H. Naya, *et al.*, Genomic Evidence for the Emergence and Evolution of Pathogenicity and Niche Preferences in the Genus Campylobacter. *Genome biology and evolution* **6**, 2392–2405, ISSN: 1759-6653, DOI 10.1093/gbe/evu195 (Sept. 2014) (cit. on p. 215).
438. A. Elmi, E. Watson, P. Sandu, *et al.*, Campylobacter jejuni Outer Membrane Vesicles Play an Important Role in Bacterial Interactions with Human Intestinal Epithelial Cells. *Infection and immunity* **80**, ed. by J. L. Flynn, 4089–4098, ISSN: 0019-9567, DOI 10.1128/IAI.00161-12 (Sept. 2012) (cit. on pp. 216, 220, 239).
439. L. D. Kalischuk, G. D. Inglis, A. G. Buret, Strain-dependent induction of epithelial cell oncosis by Campylobacter jejuni is correlated with invasion ability and is independent of cytolethal distending toxin. *Microbiology* **153**, 2952–2963, ISSN: 1350-0872 (Print) 1350-0872 (Linking), DOI 10.1099/mic.0.2006/003962-0 (Sept. 2007) (cit. on p. 216).
440. N. Crosetto, A. Mitra, M. J. Silva, *et al.*, Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nature methods* **10**, 361–365, ISSN: 1548-7091, DOI 10.1038/nmeth.2408 (Mar. 2013) (cit. on pp. 216, 225).
441. M. Yamada, Y. Watanabe, J. S. Gootenberg, *et al.*, Crystal Structure of the Minimal Cas9 from Campylobacter jejuni Reveals the Molecular Diversity in the CRISPR-Cas9 Systems. *Molecular Cell* **65**, 1109–1121.e3, DOI 10.1016/j.molcel.2017.02.007 (Mar. 2017) (cit. on p. 216).
442. E. Kim, T. Koo, S. W. Park, *et al.*, In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nature Communications* **8**, 14500–14500, DOI 10.1038/ncomms14500 (Feb. 2017) (cit. on p. 216).
443. R. Sundaresan, H. P. Parameshwaran, S. Yogesha, M. W. Keilbarth, R. Rajan, RNA-Independent DNA Cleavage Activities of Cas9 and Cas12a. *Cell Rep* **21**, 3728–3739, ISSN: 2211-1247 (Electronic), DOI 10.1016/j.celrep.2017.11.100 (Dec. 2017) (cit. on pp. 216, 217, 264).

444. T. A. Hartjes, J. Leivo, M. Vredenbregt-van den Berg, *et al.*, EV analysis of clinical urine samples from prostate cancer patients. *Journal of Extracellular Vesicles* **7**, 14–14, DOI 10.1080/20013078.2018.1461450 (Apr. 2018) (cit. on p. 221).

445. G. Dugar, S. L. Svensson, T. Bischler, *et al.*, The CsrA-FliW network controls polar localization of the dual-function flagellin mRNA in Campylobacter jejuni. *Nature Communications* **7**, 11667–11667, DOI 10.1038/ncomms11667 (May 2016) (cit. on pp. 221, 222).

446. K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30**, 2725–2729, ISSN: 1537-1719, DOI 10.1093/molbev/mst197 (Oct. 2013) (cit. on p. 221).

447. R. Louwen, E. E. S. Nieuwenhuis, L. van Marrewijk, *et al.*, Campylobacter jejuni Translocation across Intestinal Epithelial Cells Is Facilitated by Ganglioside-Like Lipooligosaccharide Structures. *Infection and Immunity* **80**, ed. by B. A. McCormick, 3307–3318, DOI 10.1128/iai.06270-11 (July 2012) (cit. on pp. 221–223, 239, 246).

448. K. A. Talukder, M. Aslam, Z. Islam, *et al.*, Prevalence of Virulence Genes and Cytolethal Distending Toxin Production in Campylobacter jejuni Isolates from Diarrheal Patients in Bangladesh. *Journal of Clinical Microbiology* **46**, 1485–1488, DOI 10.1128/jcm.01912-07 (Feb. 2008) (cit. on p. 221).

449. J. Schindelin, I. Arganda-Carreras, E. Frise, *et al.*, Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676–682, ISSN: 1548-7105 (Electronic)1548-7091 (Linking), DOI 10.1038/nmeth.2019 (June 2012) (cit. on p. 223).

450. K. A. Sap, K. Bezstarosti, D. H. W. Dekkers, *et al.*, Global quantitative proteomics reveals novel factors in the ecdysone signaling pathway inDrosophila melanogaster. *Proteomics* **15**, 725–738, DOI 10.1002/pmic.201400308 (Jan. 2015) (cit. on p. 226).

451. A. Kuzniar, C. Laffeber, B. Eppink, *et al.*, Semi-quantitative proteomics of mammalian cells upon short-term exposure to non-ionizing electromagnetic fields. *PLoS ONE* **12**, ed. by M. R. Scarfi, e0170762, DOI 10.1371/journal.pone.0170762 (Feb. 2017) (cit. on p. 226).

452. É. Hegedüs, E. Kókai, A. Kotlyar, V. Dombrádi, G. Szabó, Separation of 1–23-kb complementary DNA strands by urea–agarose gel electrophoresis. *Nucleic acids research* **37**, e112–e112, DOI 10.1093/nar/gkp539 (June 2009) (cit. on p. 226).

453. R. O. Watson, J. E. Galán, Campylobacter jejuni Survives within Epithelial Cells by Avoiding Delivery to Lysosomes. *PLoS Pathogens* **4**, e14, ISSN: 1553-7374 (Electronic), DOI 10.1371/journal.ppat.0040014 (2008) (cit. on p. 239).

454. S. Sternberg, H. Richter, E. Charpentier, U. Qimron, Adaptation in CRISPR-Cas Systems. *Molecular Cell* **61**, 797–808, ISSN: 1097-2765, DOI 10.1016/j.molcel.2016.01.030 (Mar. 2016) (cit. on p. 259).

455. C. Richter, T. Gristwood, J. S. Clulow, P. C. Fineran, In Vivo Protein Interactions and Complex Formation in the Pectobacterium atrosepticum Subtype I-F CRISPR/Cas System. *PLoS One* **7**, ed. by D. Arnold, e49549, ISSN: 1932-6203 (Electronic) 1932-6203 (Linking), DOI 10.1371/journal.pone.0049549 (Dec. 2012) (cit. on p. 259).

456. 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, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gku1407 (Jan. 2015) (cit. on p. 259).

457. Y. Xiao, S. Ng, K. H. Nam, A. Ke, How type II CRISPR–Cas establish immunity through Cas1–Cas2-mediated spacer integration. *Nature* **550**, 137–141, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature24020 (Sept. 2017) (cit. on p. 259).

458. D. Ka, D. Jang, B. Han, E. Bae, Molecular organization of the type II-A CRISPR adaptation module and its interaction with Cas9 via Csn2. *Nucleic Acids Research* **46**, 9805–9815, DOI 10.1093/nar/gky702 (Aug. 2018) (cit. on p. 259).

459. H. Lee, Y. Zhou, D. W. Taylor, D. G. Sashital, Cas4-Dependent Prespacer Processing Ensures High-Fidelity Programming of CRISPR Arrays. *Mol Cell* **70**, 48–59.e5, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2018.03.003 (Apr. 2018) (cit. on p. 259).

460. M. Shiimori, S. C. Garrett, B. R. Graveley, M. P. Terns, Cas4 Nucleases Define the PAM, Length, and Orientation of DNA Fragments Integrated at CRISPR Loci. *Mol Cell* **70**, 814–824.e6, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2018.05.002 (June 2018) (cit. on p. 259).

461. S. N. Kieper, C. Almendros, J. Behler, *et al.*, Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation. *Cell Rep* **22**, 3377–3384, ISSN: 2211-1247 (Electronic), DOI 10.1016/j.celrep.2018.02.103 (Mar. 2018) (cit. on p. 259).

462. C. Almendros, F. L. Nobrega, R. E. McKenzie, S. J. Brouns, Cas4–Cas1 fusions drive efficient PAM selection and control CRISPR adaptation. *Nucleic Acids Res* **47**, 5223–5230, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gkz217 (Apr. 2019) (cit. on p. 260).

463. S. Hudaiberdiev, S. Shmakov, Y. I. Wolf, *et al.*, Phylogenomics of Cas4 family nucleases. *BMC Evol Biol* **17**, 232, ISSN: 1471-2148 (Electronic) 1471-2148 (Linking), DOI 10.1186/s12862-017-1081-1 (Nov. 2017) (cit. on p. 260).

464. R. C. Edgar, PILER-CR: Fast and accurate identification of CRISPR repeats. *BMC Bioinformatics* **8**, 18, ISSN: 1471-2105 (Electronic) 1471-2105 (Linking), DOI 10.1186/1471-2105-8-18 (Jan. 2007) (cit. on p. 260).

465. I. Grissa, G. Vergnaud, C. Pourcel, CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* **35**, W52–W57, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gkm360 (May 2007) (cit. on p. 260).

466. C. Almendros, N. M. Guzmán, J. García-Martínez, F. J. M. Mojica, Anti-cas spacers in orphan CRISPR4 arrays prevent uptake of active CRISPR–Cas I-F systems. *Nature Microbiology* **1**, ISSN: 2058-5276, DOI 10.1038/nmicrobiol.2016.81 (June 2016) (cit. on p. 260).

467. S. Silas, P. Lucas-Elio, S. A. Jackson, *et al.*, Type III CRISPR-Cas systems can provide redundancy to counteract viral escape from type I systems. *Elife* **6**, ISSN: 2050-084X, DOI ARTNe2760110.7554/eLife.27601 (2017) (cit. on p. 260).

468. 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, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/s41586-019-1323-z (June 2019) (cit. on pp. 260, 275).

469. J. Carte, N. T. Pfister, M. M. Compton, R. M. Terns, M. P. Terns, Binding and cleavage of CRISPR RNA by Cas6. *RNA* **16**, 2181–2188, ISSN: 1469-9001 (Electronic) 1355-8382 (Linking), DOI 10.1261/rna.2230110 (Sept. 2010) (cit. on pp. 260, 261).

470. D. Dong, K. Ren, X. Qiu, *et al.*, The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* **532**, 522–526, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature17944 (Apr. 2016) (cit. on p. 261).

471. A. A. Smargon, D. B. Cox, N. K. Pyzocha, *et al.*, Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Mol Cell* **65**, 618–630.e7, ISSN: 1097-4164 (Electronic) 1097-2765 (Linking), DOI 10.1016/j.molcel.2016.12.023 (Feb. 2017) (cit. on pp. 263, 265).

472. G. Tamulaitis, Č. Venclovas, V. Siksnys, Type III CRISPR-Cas Immunity: Major Differences Brushed Aside. *Trends in Microbiology* **25**, 49–61, ISSN: 0966-842X, DOI 10.1016/j.tim.2016.09.012 (Jan. 2017) (cit. on p. 263).

473. G. J. Knott, A. East-Seletsky, J. C. Cofsky, *et al.*, Guide-bound structures of an RNA-targeting A-cleaving CRISPR–Cas13a enzyme. *Nature Structural & Molecular Biology* **24**, 825–833, ISSN: 1545-9993, DOI 10.1038/nsmb.3466 (Sept. 2017) (cit. on p. 263).

474. D. C. Swarts, M. Jinek, Cas9 versus Cas12a/Cpf1: Structure-function comparisons and implications for genome editing. *Wiley Interdiscip Rev RNA* **9**, e1481, ISSN: 1757-7012 (Electronic) 1757-7004 (Linking), DOI 10.1002/wrna.1481 (May 2018) (cit. on p. 264).

475. S.-Y. Li, Q.-X. Cheng, J.-K. Liu, *et al.*, CRISPR-Cas12a has both cis- and trans-cleavage activities on single-stranded DNA. *Cell Research* **28**, 491–493, ISSN: 1001-0602, DOI 10.1038/s41422-018-0022-x (Mar. 2018) (cit. on p. 264).

476. D. C. Swarts, M. Jinek, Mechanistic Insights into the cis- and trans-Acting DNase Activities of Cas12a. *Molecular Cell* **73**, 589–600.e4, ISSN: 1097-2765, DOI 10.1016/j.molcel.2018.11.021 (Feb. 2019) (cit. on p. 264).

477. L. Liu, X. Y. Li, J. Ma, *et al.*, The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a. *Cell* **170**, 714–+, ISSN: 0092-8674, DOI 10.1016/j.cell.2017.06.050 (2017) (cit. on p. 265).

478. M. L. Maeder, S. J. Linder, V. M. Cascio, *et al.*, CRISPR RNA–guided activation of endogenous human genes. *Nat Methods* **10**, 977–979, ISSN: 1548-7105 (Electronic) 1548-7091 (Linking), DOI 10.1038/nmeth.2598 (July 2013) (cit. on p. 267).

479. P. Perez-Pinera, D. D. Kocak, C. M. Vockley, *et al.*, RNA-guided gene activation by CRISPR-Cas9–based transcription factors. *Nat Methods* **10**, 973–976, ISSN: 1548-7105 (Electronic) 1548-7091 (Linking), DOI 10.1038/nmeth.2600 (July 2013) (cit. on p. 267).

480. C. Sinha, A. Kumar, N. Goel, A. Singh, Anesthetic implications in a child with crouzon syndrome. *Anesth Essays Res* **11**, 246–247, ISSN: 0259-1162 (Print) 2229-7685 (Linking), DOI 10.4103/0259-1162.200234 (2017) (cit. on p. 267).

481. Y. K. Tandon, Bilateral squamosal suture synostosis: A rare form of isolated craniosynostosis in Crouzon syndrome. *World J Radiol* **6**, 507, ISSN: 1949-8470 (Print) 1949-8470 (Linking), DOI 10.4329/wjr.v6.i7.507 (2014) (cit. on p. 267).

482. S. Twigg, A. Wilkie, A Genetic-Pathophysiological Framework for Craniosynostosis. *Am J Hum Genet* **97**, 359–377, ISSN: 1537-6605 (Electronic) 0002-9297 (Linking), DOI 10.1016/j.ajhg.2015.07.006 (Sept. 2015) (cit. on p. 267).

483. K. Mangasarian, Y. Li, A. Mansukhani, C. Basilico, Mutation associated with Crouzon syndrome causes ligand-independent dimerization and activation of FGF receptor-2. *J Cell Physiol* **172**, 117–125, ISSN: 0021-9541 (Print) 0021-9541 (Linking), DOI 10.1002/(sici)1097-4652(199707)172:1<117::aid-jcp13>3.0.co;2-9 (July 1997) (cit. on p. 267).

484. A. K. Snyder-Warwick, C. A. Perlyn, J. Pan, *et al.*, Analysis of a gain-of-function FGFR2 Crouzon mutation provides evidence of loss of function activity in the etiology of cleft palate. *Proc Natl Acad Sci U S A* **107**, 2515–2520, ISSN: 1091-6490 (Electronic) 0027-8424 (Linking), DOI 10.1073/pnas.0913985107 (Feb. 2010) (cit. on p. 267).

485. P. J. Malloy, V. Tasic, D. Taha, *et al.*, Vitamin D receptor mutations in patients with hereditary 1,25-dihydroxyvitamin D-resistant rickets. *Mol Genet Metab* **111**, 33–40, ISSN: 1096-7206 (Electronic) 1096-7192 (Linking), DOI 10.1016/j.ymgme.2013.10.014 (Jan. 2014) (cit. on p. 267).

486. S. B. Ameur, C. Silve, I. Chabchoub, *et al.*, Clinical and Genetic Characterization of Tunisian Children with Hereditary 1,25-Dihydroxyvitamin D-Resistant Rickets. *Horm Res Paediatr* **87**, 23–29, ISSN: 1663-2826 (Electronic) 1663-2818 (Linking), DOI 10.1159/000452886 (Dec. 2017) (cit. on p. 267).

487. Y. Fu, J. A. Foden, C. Khayter, *et al.*, High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**, 822–826, ISSN: 1546-1696 (Electronic) 1087-0156 (Linking), DOI 10.1038/nbt.2623 (June 2013) (cit. on p. 268).

488. D. Paquet, D. Kwart, A. Chen, *et al.*, Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125–129, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature17664 (Apr. 2016) (cit. on p. 268).

489. R. E. Kontermann, Strategies for extended serum half-life of protein therapeutics. *Curr Opin Biotechnol* **22**, 868–876, ISSN: 1879-0429 (Electronic) 0958-1669 (Linking), DOI 10.1016/j.copbio.2011.06.012 (Dec. 2011) (cit. on p. 268).

490. V. Schellenberger, C.-w. Wang, N. C. Geething, *et al.*, A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nat Biotechnol* **27**, 1186–1190, ISSN: 1546-1696 (Electronic) 1087-0156 (Linking), DOI 10.1038/nbt.1588 (Nov. 2009) (cit. on p. 268).

491. C. T. Charlesworth, P. S. Deshpande, D. P. Dever, *et al.*, Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nature Medicine* **25**, 249–254, ISSN: 1078-8956, DOI 10.1038/s41591-018-0326-x (Jan. 2019) (cit. on p. 268).

492. J. Strecker, S. Jones, B. Koopal, *et al.*, Engineering of CRISPR-Cas12b for human genome editing. *Nat Commun* **10**, 212, ISSN: 2041-1723 (Electronic) 2041-1723 (Linking), DOI 10.1038/s41467-018-08224-4 (Jan. 2019) (cit. on p. 269).

493. S. Kaboli, H. Babazada, CRISPR Mediated Genome Engineering and its Application in Industry. *Current Issues in Molecular Biology* **26**, 81–92, ISSN: 1467-3037, DOI 10.21775/cimb.026.081 (2018) (cit. on p. 270).

494. J. M. Vento, N. Crook, C. L. Beisel, Barriers to genome editing with CRISPR in bacteria. *J Ind Microbiol Biotechnol*, ISSN: 1476-5535 (Electronic) 1367-5435 (Linking), DOI 10.1007/s10295-019-02195-1 (June 2019) (cit. on p. 270).

495. B. B. Sun, J. J. Yang, S. Yang, *et al.*, A CRISPR-Cpf1-Assisted Non-Homologous End Joining Genome Editing System of Mycobacterium smegmatis. *Biotechnology Journal* **13**, ISSN: 1860-6768, DOI Artn170058810.1002/Biot.201700588 (2018) (cit. on p. 270).

496. R. T. Leenay, J. M. Vento, M. Shah, *et al.*, Genome Editing with CRISPR-Cas9 in Lactobacillus plantarum Revealed That Editing Outcomes Can Vary Across Strains and Between Methods. *Biotechnol J* **14**, 1700583, ISSN: 1860-7314 (Electronic) 1860-6768 (Linking), DOI 10.1002/biot.201700583 (Sept. 2019) (cit. on p. 270).

497. R. Mendes, M. Kruijt, I. de Bruijn, *et al.*, Deciphering the Rhizosphere Microbiome for Disease-Suppressive Bacteria. *Science* **332**, 1097–100, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking), DOI 10.1126/science.1203980 (May 2011) (cit. on p. 271).

498. E. Chapelle, R. Mendes, P. A. H. Bakker, J. M. Raaijmakers, Fungal invasion of the rhizosphere microbiome. *ISME J* **10**, 265–268, ISSN: 1751-7370 (Electronic) 1751-7362 (Linking), DOI 10.1038/ismej.2015.82 (May 2016) (cit. on p. 271).

499. V. J. Carrión, V. Cordovez, O. Tyc, *et al.*, Involvement of Burkholderiaceae and sulfurous volatiles in disease-suppressive soils. *ISME J* **12**, 2307–2321, ISSN: 1751-7370 (Electronic) 1751-7362 (Linking), DOI 10.1038/s41396-018-0186-x (June 2018) (cit. on p. 271).

500. V. Cordovez, V. J. Carrion, D. W. Etalo, *et al.*, Diversity and functions of volatile organic compounds produced by Streptomyces from a disease-suppressive soil. *Front Microbiol* **6**, 1081, ISSN: 1664-302X (Print) 1664-302X (Linking), DOI 10.3389/fmicb.2015.01081 (Oct. 2015) (cit. on p. 271).

501. J. D. G. Jones, J. L. Dangl, The plant immune system. *Nature* **444**, 323–329, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking), DOI 10.1038/nature05286 (Nov. 2006) (cit. on p. 271).

502. A. M. Staroscik, D. W. Hunnicutt, K. E. Archibald, D. R. Nelson, Development of methods for the genetic manipulation of Flavobacterium columnare. *BMC Microbiol* **8**, 115, ISSN: 1471-2180 (Electronic) 1471-2180 (Linking), DOI 10.1186/1471-2180-8-115 (2008) (cit. on p. 271).

503. E. Gómez, B. Álvarez, E. Duchaud, J. A. Guijarro, Development of a Markerless Deletion System for the Fish-Pathogenic Bacterium Flavobacterium psychrophilum. *PLoS One* **10**, ed. by U. G. Munderloh, e0117969, ISSN: 1932-6203 (Electronic) 1932-6203 (Linking), DOI 10.1371/journal.pone.0117969 (Feb. 2015) (cit. on p. 271).

504. O. Kilian, C. S. E. Benemann, K. K. Niyogi, B. Vick, High-efficiency homologous recombination in the oil-producing alga Nannochloropsis sp. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 21265–21269, ISSN: 0027-8424, DOI 10.1073/pnas.1105861108 (Nov. 2011) (cit. on pp. 273, 274).

505. E. Poliner, T. Takeuchi, Z.-Y. Du, C. Benning, E. M. Farré, Nontransgenic Marker-Free Gene Disruption by an Episomal CRISPR System in the Oleaginous Microalga, Nannochloropsis oceanica CCMP1779. *ACS Synth Biol* **7**, 962–968, ISSN: 2161-5063 (Electronic) 2161-5063 (Linking), DOI 10.1021/acssynbio.7b00362 (Mar. 2018) (cit. on pp. 273, 274).

506. J. Verruto, K. Francis, Y. Wang, *et al.*, Unrestrained markerless trait stacking in Nannochloropsis gaditana through combined genome editing and marker recycling technologies. *Proc Natl Acad Sci U S A* **115**, E7015–E7022, ISSN: 1091-6490 (Electronic) 0027-8424 (Linking), DOI 10.1073/pnas.1718193115 (July 2018) (cit. on p. 273).

507. I. Ajjawi, J. Verruto, M. Aqui, *et al.*, Lipid production in Nannochloropsis gaditana is doubled by decreasing expression of a single transcriptional regulator. *Nature Biotechnology* **35**, 647–652, ISSN: 1087-0156, DOI 10.1038/nbt.3865 (June 2017) (cit. on p. 273).

508. Q. Wang, Y. Lu, Y. Xin, *et al.*, Genome editing of model oleaginous microalgae Nannochloropsis spp. by CRISPR/Cas9. *Plant Journal* **88**, 1071–1081, ISSN: 0960-7412, DOI 10.1111/tpj.13307 (Oct. 2016) (cit. on pp. 273, 274).

509. Y. B. Kim, A. C. Komor, J. M. Levy, *et al.*, Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol* **35**, 371–376, ISSN: 1546-1696 (Electronic) 1087-0156 (Linking), DOI 10.1038/nbt.3803 (Feb. 2017) (cit. on p. 274).

510. A. Eid, S. Alshareef, M. M. Mahfouz, CRISPR base editors: genome editing without double-stranded breaks. *Biochem J* **475**, 1955–1964, ISSN: 1470-8728 (Electronic) 0264-6021 (Linking), DOI 10.1042/bcj20170793 (June 2018) (cit. on p. 274).

511. Y. Wang, S. Wang, W. Chen, *et al.*, CRISPR-Cas9 and CRISPR-Assisted Cytidine Deaminase Enable Precise and Efficient Genome Editing in Klebsiella pneumoniae. *Appl Environ Microbiol* **84**, ed. by H. L. Drake, ISSN: 1098-5336 (Electronic) 0099-2240 (Linking), DOI 10.1128/aem.01834-18 (Sept. 2018) (cit. on p. 275).

512. Q. Li, F. M. Seys, N. P. Minton, *et al.*, CRISPR–Cas9 D10A nickase-assisted base editing in the solvent producer Clostridium beijerinckii. *Biotechnol Bioeng* **116**, 1475–1483, ISSN: 1097-0290 (Electronic) 0006-3592 (Linking), DOI 10.1002/bit.26949 (Feb. 2019) (cit. on p. 275).

513. M. L. Luo, A. S. Mullis, R. T. Leenay, C. L. Beisel, Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. *Nucleic Acids Res* **43**, 674–681, ISSN: 1362-4962 (Electronic) 0305-1048 (Linking), DOI 10.1093/nar/gku971 (Oct. 2015) (cit. on p. 275).

514. C. Hidalgo-Cantabrana, Y. J. Goh, R. Barrangou, Characterization and Repurposing of Type I and Type II CRISPR–Cas Systems in Bacteria. *Journal of Molecular Biology* **431**, 21–33, ISSN: 0022-2836, DOI 10.1016/j.jmb.2018.09.013 (Jan. 2019) (cit. on p. 275).

515. M. E. Pyne, M. R. Bruder, M. Moo-Young, D. A. Chung, C. P. Chou, Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in Clostridium. *Scientific Reports* **6**, ISSN: 2045-2322, DOI 10.1038/srep25666 (May 2016) (cit. on p. 275).

516. J. Zhang, W. Zong, W. Hong, Z.-T. Zhang, Y. Wang, Exploiting endogenous CRISPR-Cas system for multiplex genome editing in Clostridium tyrobutyricum and engineer the strain for high-level butanol production. *Metabolic Engineering* **47**, 49–59, ISSN: 1096-7176, DOI 10.1016/j.ymben.2018.03.007 (May 2018) (cit. on p. 275).

About the author

Prarthana Mohanraju was born on the 20th of June, 1991 in Chennai, India. In 2008 she started her studies in Biotechnology at SRM University in India. For her BSc thesis, she performed research on optimising the parameters for the production of an industrially relevant enzyme, naringinase in *Staphylococcus aureus*. After obtaining her Bachelor's degree in 2012, she moved to The Netherlands to pursue a Master's degree in Biotechnology (specialising in cellular and molecular biology) at Wageningen University. For her MSc thesis, she used synthetic biology approaches to increase the production of itaconic acid in *Aspergillus niger* at the Systems and Synthetic biology group, Wageningen University. In 2014, she spent six months at the Genetics R&D group at DSM B.V, Delft, The Netherlands, where she characterised the expression of hexose transporters in DSM advanced yeast strains.

In March 2015, Prarthana commenced her PhD research in the Laboratory of Microbiology at Wageningen University under the supervision of Prof. John van der Oost and Dr. Daan Swarts. Her work focused on providing biological and mechanistic insights into novel Class 2 CRISPR-Cas systems and harnessing their potential to design and develop innovative tools for genome engineering. Most of her work has been described in this thesis.

Co-author affiliations

Laboratory of Microbiology, Department of Agrotechnology and Food Sciences
Wageningen University, 6703 HB Wageningen, Netherlands
John van der Oost, Wen Y. Wu, Jeroen Kneppers, Sjoerd C. A. Creutzburg, Karlijn Keessen, Tahseen S. Khan, Stijn Prinsen, Ioannis Mougiakos, Elleke F. Bosma[a], Valentijn Vrouwe, Max Finger Bou, Mihris I. S. Naduthodi, Alex Gussak, Richard van Kranenburg, Justin Albers, Despoina Trasanidou, Maarten Klunder, Rob Joosten

National Center for Biotechnology Information, National Library of Medicine, National Institute
Bethesda, MD 20894, USA
Kira S. Makarova, Eugene V. Koonin

Broad Institute of MIT and Harvard
Cambridge, MA 02142, USA
Bernd Zetsche, Feng Zhang, Matthias Heidenreich, Jeroen Kneppers, Ellen M. DeGennaro, Nerges Winblad, Sourav R. Choudhury, Omar O. Abudayyeh, Jonathan S. Gootenberg, David A. Scott

Department of Biochemistry, University of Zurich
Zurich, Switzerland
Martin Jinek, Daan C. Swarts[b]

McGovern Institute for Brain Research, MIT
Cambridge, MA 02139
Bernd Zetsche, Matthias Heidenreich, Iana Fedorova, Nerges Winblad, Sourav R. Choudhury, Omar O. Abudayyeh, Jonathan S. Gootenberg, David A. Scott

Department of Brain and Cognitive Sciences, MIT
Cambridge, MA 02139
Bernd Zetsche, Matthias Heidenreich, Iana Fedorova, Nerges Winblad, Sourav R. Choudhury, Omar O. Abudayyeh, Jonathan S. Gootenberg

Department of Biological Engineering, MIT
Cambridge, MA 02139
Bernd Zetsche, Matthias Heidenreich, Iana Fedorova, Nerges Winblad, Sourav R. Choudhury, Omar O. Abudayyeh, Jonathan S. Gootenberg

Skolkovo Institute of Science and Technology
Skolkovo, 143025, Russia
Iana Fedorova, Konstantin Severinov

Harvard-MIT Division of Health Sciences and Technology, MIT
Cambridge, MA 02139
Ellen M. DeGennaro, Omar O. Abudayyeh

Waksman Institute for Microbiology, Rutgers, The State University of New Jersey
Piscataway, NJ 08854, USA
Konstantin Severinov

Institute of Molecular Genetics, Russian Academy of Sciences
Moscow, 123182, Russia
Konstantin Severinov

Arbor Biotechnologies
Cambridge, MA 02139, USA
Winston X. Yan, David A. Scott

Corbion
Arkelsedijk 46, 4206 AC Gorinchem, The Netherlands
Rudolf B. L. Brinkman, Richard van Kranenburg,

New England Biolabs
240 County Road, Ipswich, MA, USA, 01938
Megumu Mabuchi, Ryan T. Fuchs, G. Brett Robb

Department Medical Microbiology and Infectious Diseases, Erasmus MC University Medical Center Rotterdam
Rotterdam, The Netherlands
Chinmoy Saha, Deborah Horst-Kreft, Dior Beerens[c], Rogier Louwen

Department of Bioinformatics, Erasmus MC University Medical Center Rotterdam
Rotterdam, The Netherlands
Andrew Stubbs, Youri Hoogstrate, Peter J. van der Spek

Institute of Molecular Infection Biology (IMIB)/Research Center for Infectious Diseases (ZINF), University of Würzburg
Würzburg, Germany
Gaurav Dugar[d]

Optical Imaging Center, Erasmus MC University Medical Center Rotterdam
Rotterdam, The Netherlands
Gert-Jan Kremers, Wiggert A. van Cappellen

Department of Genetics, Erasmus MC University Medical Center Rotterdam
Rotterdam, The Netherlands
Joyce H.G. Lebbink, Serena Bruens, Dik van Gent

Institute of Food Research, Gut Health and Food Safety Programme
Norwich Research Park, Norwich, United Kingdom
Duncan Gaskin

Proteomics Center, Erasmus MC University Medical Center Rotterdam
Rotterdam, The Netherlands
Jeroen A. A. Demmers

Host–Microbe Interactomics Group, University of Wageningen
Wageningen, The Netherlands
Peter van Baarlen

Current Addresses

[a] **The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark**, *Kemitorvet B220, 2800 Kgs. Lyngby, Denmark*

[b] **Laboratory of Biochemistry, Department of Agrotechnology and Food Sciences**, *Wageningen University, 6703 HB Wageningen, Netherlands*

[c] **Max Planck Institute for Infection Biology**, *Berlin, Germany*

[d] **Swammerdam Institute for Life Sciences, University of Amsterdam**, *Amsterdam, The Netherlands*

List of publications

13. C. Saha*, **P. Mohanraju***, A. Stubbs, G. Dugar, Y. Hoogstrate, G. Y. Kremers, W. A. van Cappellen, D. Horst-Kreft, J. H. G. Lebbink, S. Bruens, D. Gaskin, D. Beerens, M. Klunder, R. Joosten, J. A. A. Demmers, D. van Gent, J. W. Mouton, P. J. van der Spek, P. van Baarlen, J. van der Oost, R. Louwen. *Guide-free Cas9 from pathogenic Campylobacter jejuni bacteria causes severe DNA damage in cultured human cells. Manuscript submitted.*

12. **P. Mohanraju***, I. Mougiakos*, J. Albers, M. Mabuchi, R. Fuchs, R. van Kranenburg, B. Robb, J. van der Oost. *Development of a CRISPR-Cas12a based genome editing tool for facultative thermophiles. Manuscript in preparation.*

11. **P. Mohanraju***, W. Y. Wu*, S. C. A. Creutzburg, Y. Winston, D. A. Scott, K. S. Makarova, E. V. Koonin, J. van der Oost. *Characterising the CRISPR–Cas type V-U1 C2c4 effector enzyme. Manuscript in preparation.*

10. J. C. Carrión, J. Perez-Jaramillo*, V. Cordovez*, V. Tracanna*, M. de Hollander, D. Ruiz-Buck, L. W Mendes, W. F. J van Ijcken, R. Gomez-Exposito, S. S Elsayed, **P. Mohanraju**, A. Arifah, J. van der Oost, R. Mendes, G. P. van Wezel, M. H. Medema, J. M. Raaijmakers (2019). *Metagenome-guided reconstruction and functional characterisation of a disease-suppressive endophytic bacterial consortium.* **Science**, *accepted for publication.*

9. S. 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. Manuscript under revision.*

8. D. Trasanidou*, **P. Mohanraju***, R. van Kranenburg, I. Mougiakos, J. van der Oost. *Developing an anti-CRISPR-based toolbox for bacterial genome engineering. Manuscript in preparation.*

7. D. Trasanidou, A. S. Geros, **P. Mohanraju**, A. C. Nieuwenweg, F. L Nobrega, R. H. J. Staals (2019). *Keeping CRISPR in check: diverse mechanisms of phage-encoded Anti-CRISPRs.* **FEMS Microbiology Letters**, 366(9). DOI 10.1093/femsle/fnz098.

6. M. I. S Naduthodi, **P. Mohanraju**, C. Sudfeld, S. D'Adamo, M. J. Barbosa, J. van der Oost (2019). *CRISPR-Cas ribonucleoprotein mediated homology-directed repair for efficient targeted genome editing in microalgae* Nannochloropsis oceanica *IMET1.* **Biotechnology for Biofuels**, 12:66. DOI 10.1186/s13068-019-1401-3.

5. M. Vlot, J. Houkes, S. J. A Lochs, D. C. Swarts, P. Zheng, T. Kunne, **P. Mohanraju**, C. Anders, M. Jinek, J. van der Oost, S. J. J. Brouns (2018) *Bacteriophage DNA glucosylation impairs target DNA binding by type I and II but not by type V CRISPR-Cas effector complexes.* **Nucleic Acids Research**, 46(2), 873-885. DOI 10.1093/nar/gkx1264.

4. **P. Mohanraju**, J. van der Oost, M. Jinek, D. C. Swarts (2018). *Heterologous Expression and Purification of CRISPR-Cas12a/Cpf1.* **Bio-Protocol**, 8(9). DOI 10.21769/BioProtoc.2842.

3. I. Mougiakos*, **P. Mohanraju***, E.F. Bosma*, V. Vrouwe, M. F. Bou, M. I. S. Naduthodi, A. Gussak, R. B. L. Brinkman, R. van Kranenburg, J. van der Oost (2017). *Characterising a thermostable Cas9 for bacterial genome editing and silencing..* **Nature Communications**, 8(1):1647. DOI 10.1038/S41467-017-01591-4.

2. 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, van der Oost, F. Zhang (2017). *Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array*. **Nature Biotechnology**, 35(1), 31-34. DOI 10.1038/nbt.3737.

1. **P. Mohanraju**, K. S Makarova, B. Zetsche, F. Zhang, E. V. Koonin, J. van der Oost (2016). *Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems*. **Science**, 353(6299). DOI 10.1126/science.aad5147.

*denotes equal contribution

Patent applications

4. J. van der Oost, **P. Mohanraju**, W. Y. Wu, S. Creutzburg. *Type V-U CRISPR system*. UK1909597.5 (filed on July 3, 2019).

3. J. van der Oost, R. van Kranenburg, E.F. Bosma, I. Mougiakos, **P. Mohanraju**. *Thermostable Cas9 nucleases with reduced off-target activity*. 18 Apr 2019, Patent No. *WO2019072596*.

2. J. van der Oost, R. van Kranenburg, E.F. Bosma, I. Mougiakos, **P. Mohanraju**. *Thermostable Cas9 Nucleases*. 21 Jun 2018, Patent No. *WO2018109101*.

1. J. van der Oost, R. van Kranenburg, E.F. Bosma, I. Mougiakos, **P. Mohanraju**. *Thermostable Cas9 Nucleases*. 21 Jun 2018, Patent No. *WO2018108339*.

Overview of completed training activities

Discipline specific activities

Meetings & conferences

- Annual Molecular Genetics meeting, Lunteren, The Netherlands (2015)
- Genome Engineering and Synthetic Biology: Tools and Technologies, Ghent, Belgium (2016)
- CRISPR Conference, Rehovat, Israel (2016)
- Annual Molecular Genetics meeting, Lunteren, The Netherlands (2016)
- 4th Annual Wageningen PhD Symposium, Science: from Local to Global, Wageningen, The Netherlands (2017)*
- CRISPR: From Biology to Technology and Novel Therapeutics, Sitges, Spain (2017)*
- Microbiology Centennial, Wageningen, The Netherlands (2017)**
- NWO Chains, Veldhoven, The Netherlands (2017)**
- 5th Annual Wageningen PhD Symposium, Bridging Science and Society: Unifying Knowledge (2018)**
- CRISPR Conference, Vilnius, Lithuania (2018)*
- CRISPRCon, Wageningen, The Netherlands (2019)**

*poster presentation **oral presentation

Courses

- Biocatalysis and Protein Engineering, Delft, The Netherlands (2018)

General courses

- Project and Time Management, The Netherlands (2017)
- Career Assessment, The Netherlands (2017)
- Scientific Publishing, The Netherlands (2017)
- Scientific Writing, The Netherlands (2017)
- Presenting with Impact, The Netherlands (2017)
- Brain Friendly Working and Writing, The Netherlands (2017)
- Career Perspectives, The Netherlands (2017)

Optionals

- Preparation of research proposal

- Organizing committee, PhD study trip to Germany, Sweden, Denmark (2017)
- Bacterial Genetics group meetings, Wageningen, The Netherlands
- Microbiology PhD meetings, Wageningen, The Netherlands
- Microbiology seminars

Acknowledgments

"Appreciation is for now, gratitude is forever"

–unknown

Four point five years. Summarised in 320 pages, the PhD thesis is here! The past years have truly been an amazing learning experience that I will cherish for the rest of my life. While the new title (Dr.) does sound cool, the biggest result is this little book. Getting to this point has been a rollercoaster ride consisting of long hours of purifying proteins, running what-feels-like-a-million gels, culturing litres of *E. coli*, frustrations of irreproducible experiments, getting scooped more often than not, glimmers of hope in a seemingly simple experiment, the unparalleled joy of an accepted manuscript and the disappointments accompanying manuscript rejections. Writing this final and special section of this thesis feels a bit strange, but at the same time very rewarding. The more I introspect, the more I realise the modesty of my contribution to my own achievements, as compared to the people who have supported me emotionally and intellectually. A lot of people have been pulling and pushing me along the path that has led me here and I am thankful to each one of you.

First and foremost, my promotor and mentor **John**, I am forever indebted to you for carving out the researcher in me. Words cannot do justice to the central role you have played in my scientific career. You've taught me patience, optimism, how to define a problem and find a solution for it. I vividly remember my first interview with you for the position of a research assistant, I wasn't too sure of the outcome, especially given my limited CRISPR knowledge at the time. I was overjoyed when you offered me an opportunity to express my scientific acumen. Since then, you have provided me with a lot of freedom as well as support to explore my ideas and hypotheses. You have always provided me the necessary impetus when the morale seemed low, giving me the needed push in the right direction. I thoroughly enjoyed all our brainstorming sessions and discussions, after which I always walked away feeling motivated. I will always remember your very first advice to me – "*Always do what your heart says, that is what has gotten me where I am*". Thanks for always taking the time to critically assess my work and manuscripts. I appreciate your humility that allowed me to just drop by your office and discuss whatever was on my mind. It has been a pleasure to work with you, deliberate unlimited new (wild!) ideas and learn from you at every step of the way as a student. Your endless passion for science is admirable. I hope I can continue to count on your advice even after I have left BacGen. I wish you happiness and a great life ahead with lovely **Paulien**, your sons and **Charlie**.

Daan, thank you for accepting to be my co-promotor and for your continuous support during my PhD. A house is as good as it's foundation. You were my direct supervisor during my first three months at BacGen and helped me pick up many techniques and tricks, particularly about protein purification, that formed the basis for my next four years of research. I have learnt a lot from you when it comes to approaching science. I am amazed at the clarity of thought I have been left with after the numerous (nice) discussions with you. I appreciate the critical and constructive feedback that led to the improvement of some of my chapters in my thesis. It's been a real pleasure collaborating with you over the past years and I certainly

hope we will continue doing so.

Willem, it's been an honour to work in a chair group headed by a great scientist and a nice person like you. Thank you for setting such a high benchmark. **Richard**, I have thoroughly enjoyed our brief discussions and thank you for the fruitful collaborations. **Raymond**, I am still amazed at how you manage to do all the things you do. Thanks for being so chill, down-to-earth and always willing to help. I appreciate your advice during the last months of my PhD, reassuring me that everything will work out fine. **Stan**, thanks for organizing the CRISPR meetings and outings that I truly enjoyed and for giving me a timely valuable tip for the Cas12a pre-processing experiments. Many thanks to my defence committee, **Michiel Kleerebezem**, **Niels Geijsen**, **Edze Westra** and **Pascale Daran-Lapujade**, who have sacrificed their precious time to evaluate my PhD thesis. **Edze**, special thanks for coming all the way from the UK for the defence.

I had the honour to work with various experts from diverse research fields. **Kira** and **Eugene**, thank you for providing me bioinformatic insights on CRISPR systems right from the first year of my PhD. It was my pleasure and a great honour to write a review with CRISPR pioneers like you. I am often amazed at the speed at which I would get email replies from both of you. **Bernd** and **Feng**, it was a great experience to collaborate with you on the CRISPR review, and the Cas12a pre-crRNA processing project. **Hans**, **Jeroen** and the gracious family, thanks for involving me in the nice project about using CRISPR nucleases for treating craniosynothosis. I really enjoyed our meetings and discussions. **David** and **Winston**, thanks for performing the NGS to determine the PAM for MmuC2c4 as well as the insights to determine the function of the type VU proteins. I learnt a lot from our discussions and efforts trying to understand the elusive type VU proteins. **Brett**, **Ryan** and **Megumu**, it was nice that after our interesting discussion at the CRISPR onference, we got together to work on the thermo-Cas12a project. **Chinmoy**, **Rogier**, **Peter**, thank you for the long-standing collaborative efforts to unravel a very interesting aspect of CRISPR biology. **Victor**, **Viviane** and **Jos**, I am happy that together with **Adini**, we were able to contribute to the massive flavo project. Thank you for including us.

I feel very fortunate to be surrounded by so many kind-hearted people, that each helped me in their own way. **Teunke**, you're really the nicest person I have ever met in my life. I am so glad we connected at how "quiet" our office was back in Dreijn, and the rest as they say is history. I'm always amazed at how well you understand me, sometimes even without saying anything (bit freaky!). Right from the very beginning we got along well, thanks to our mutual love for the sun, fresh air, ice-cream and tea. I have thoroughly enjoyed all our talks, walks, lunches, outings, game nights, pub quizzes, and escape rooms together. I am extremely proud of how well you handled the challenges during your PhD and defended it so well last year! I also really enjoyed the tradition of bringing back gifts from the new places we visit each year. This allowed me to experience a part of different countries without ever being there. I would also like to acknowledge your failed attempts to get me to the gym and leaving work early. Thanks for organising together with **Joyshree** the surprise evening out after I handed in my thesis. I hope you're never followed by a wolf in real life! I wish you and **Mark**, a very happy future in your new house.

Joyshree, I am sure there is no one else at work I have spent more chatting with than you. Thank you for providing tons of support, laughs, coffees, lunches, and most importantly a listening ear for my vent sessions over the past years. Whining helps me calm down sometimes—I am glad I have you to whine to. I really appreciate that you were always there to help and provide company during late lab evenings. It's amazing how we share a lot of commonalities in our thoughts and practices. Thanks for inspiring me to eat freshly made salads for lunch and for the impromptu lets-go-out-for-dinner-tonight outings. I am grateful for all the serious life conversations as well as for the ones we have laughed till our tummies hurt. I would also like to thank you for putting me in my place when I get too adamant about my opinion. It is reassuring to know that I can count on your support, in good but also in bad times. I appreciate you checking on me constantly while I was writing and encouraging me to look at the brighter side. I hope I can return the favour to you. All the very best with the last leg of the PhD, I am sure you'll do a great job at it! Please don't forget the pact we made about Kenya, but first Luxembourg, maybe?

One of my strongest pillars of support in my research has been you, **Yannis**. You've been a wonderful friend, colleague, science-partner, travel companion, all wrapped in one! I owe you big time for your help and support during some of the most stressful times of my PhD. We have shared together moments of anxiety, but also of big excitement. Our research collaborations have yielded many fruitful results and many more on the way. Thank you for the lovely dinners and vibrant parties. I thoroughly enjoyed all our trips and conference visits. It was also a lot of fun organizing and going on the PhD trip together. Thank you for teaching me to chill, introducing me to smartbars, being so approachable, caring and an amazing comrade. It's great that I can discuss anything with you, whether it is science, personal matters or just retelling old jokes. I apologize for moments I have let you down, as what we have is special and I only hope for many better moments to come. I wish you all the best for the next phase of your life.

Wen, I would like to steal your words and say them right back to you - "you're the *in vivo* to my *in vitro*" and "you're the day to my night". When you joined the lab, I was immediately struck by your infectious enthusiasm and positive attitude. I think by now you have managed to instil quite a level of optimism in me. It has been a lot of fun working together and sharing the office together with you. To be very honest, I did not think we would get along THIS well when we had just started to work together. I believe our working ethics, habits and passion for curiosity-driven science has helped a lot to cement our bond. You were instrumental in helping me finish my last chapter in time for my deadline and I appreciate your efforts. Thanks for being a great teammate, fellow CRISPRer and an absolutely fun conference buddy. I can't wait to see you as a PI making tons of money with the cool patents you're going to file.

Thanks to all the colleagues that shared an office with me, for making my time at work highly enjoyable. **Sjoerd**, it has been a lot of fun working together on the spacer efficiency (except the part at the 37 °C room) and the type VU projects. I am eternally grateful for all your help, the amazing .jsx file, guidance and troubleshooting in the lab. Also, thanks for the very many random facts, geography trivia, Nederlands lessons, etymology discussions... okay I can go on and on (as our conversations), but I think I should stop here. **Yifan**, your hardworking and always eager-to-learn attitude is inspiring. Thank you for lending a helping hand with the radiolabelling experiments. Also, thanks for the tasty dinners! **Jorrit**, thanks for being my protein-buddy and for always being ready to help with the FPLC throughout the years. I am also glad to have met **Maria** through you. I am happy you found a great job and I am sure you'll make an amazing team manager. **Marnix**, it was great to have random chats in the office with you and to learn a bit about wood-working every now and then. **Despoina**, we got along well right from the first day of your master thesis, and I feel very happy to share an office with you as colleagues. That smile of yours is truly infections, always keep it on! **Thomas**, I must say 6 out of the 10 puns you try do work out, so never give up on them. Thanks for constantly checking on me and motivating me during the last bits of my thesis and for teaching me how to make *pretty* protein structures in Illustrator.

Tom, thank you for always cheering for me. You're one of the nicest, fun-loving, no-nonsense people I met at BacGen. I wish you, **Juliette**, **Sophie** and **Coen** the best in your lives. **Jeroen**, I have thoroughly enjoyed our discussions (especially the ones about the future) together. Thanks for the dry fruits that you would share with **Wen** who in turn would share them with me. I wish you a great post-doc life at BacGen on a very interesting project. **Costas**, my dear ~~Greek~~ (oops) Cypriot friend! Thanks for the best music, the best carrot cake and the nice chats at work. You've inspired me with your love for entrepreneurship, I am sure you're going to go a long way. **Mihris**, working with you on the Algae-Cas12a project was a real pleasure and I am glad that it led to a nice paper. I was very impressed with how quickly you learnt new things and how independent you are. Also, your choice of Indian music is great, not like you needed any validation! **Rob**, thank you for joining and assisting me on the never-ending-quest-for-the-perfect-gel-images, I am glad in the end we succeeded. Many thanks for taking care of the organisational aspects of running the BacGen lab. I appreciate your disciplined and hands-on mentality.

Alex, I am glad that I got to know you better. Thanks for making me feel so comfortable around you (almost like meeting a long-lost friend, sniff) and for your sense of humour. I owe you for flying over for my defence! **Nico**, thanks for the nice conversations and accompanying me in the lab during late evening

experiments. **Elleke**, it was great fun working together on Yannis's video and getting to know each other a bit better. Thanks for the nice collaboration on the ThermoCas9 project. Your positivity and undeniable efficiency is inspiring. It was my pleasure to get to know **Zeno** and your lovely family. **Serve**, though our interaction has been limited, the communication we had over the years has been joyful. **Belen**, it's always been fun to talk to you. Thanks for helping with the CjeCas9 protein experiments in yeast. I am happy to hear you enjoyed your trip to India, you must visit me in the South next time! **Stijn**, thanks for assisting with the MmuC2c4 EMSA and DNA cleaving experiments. **Thijs**, **Joep**, **Max** and **Lorenzo**, it's been great to see you guys grow professionally through the years from a master student to a colleague (ok somehow that makes me feel very old!), thanks for the good times and nice conversations. I wish you all the best!

I would also like to extend my gratitude to all the colleagues at BacGen, in the past, present and future. Many thanks to **Melvin**, **Tijn**, **Tim**, **Mark Y.**, **Mark L.**, **Brenda**, **Emiel**, **Lucia**, **Shreyans**, **Mamou**, **Enrico Janneke**, **Catarina**, **Hanne**, **Ismael**, **Jie**, **Carina**, **Guus**, **James**, **Lione**, **Eric**, and **Jurre** for all the fun coversations during the Christmas dinners, bbqs, working in the lab, corridors, cakes, coffee, games and lunch breaks.

I would also like to thank **Maria**, **Linde**, **Enrique**, **Christos Erika** and **Lyon** from SSB for the interesting lunch-table conversations and fun at parties.

Bich, thank you for bringing a fresh perpective and letting me explore my work beyond the lab. It was a pleasure and a lot of fun collaborating with you for your (sci-fi) art project.

Franklin, **Seb**, **Patrick**, thanks for the good times! **Jochem**, thank you for the awesome Israel trip you planned for us and for the nice discussions. **Becca**, from the moment we met, I knew that we were going to be friends. I'm so glad I was right. You are kind, understanding and great fun to hang out with. We share many similarities – from thoughts to likes to dislikes. I have immensely enjoyed our times together at Wageningen, Israel, Delft and more so at our bi/tri-monthly dinners where we would hardly realize the time pass by. We need to keep them on!

I would also like to thank the other members of the ice-cream group for the fun activities and outings over the past years. **Emmy**, thanks for enlightening me with how we can help towards sustainable living. Thanks **Mark**, for the interesting conversations mostly revolving around life, post-doc, how-not-to-kill-plants and food. **Dorett**, thanks for being so pleasant and approachable. We should someday complete the *deep* discussion we started before your trip to China! **Ruben**, thanks for all the fun times, especially at the PhD trip. I think it's about time we get that ice-cream!

I have been very fortunate that I got to work with many talented BSc and MSc students. **Jara** and **Bas**, you two started almost at the same time as me, so I considered you my colleagues rather than students. It was great to work together as a trio to attempt to crack the Cpf1 nuclease. Thank you for your hard work and determination, that led to successfully setting the project up and running it from scratch. **Tom**, it was a pleasure working with you. You were very motivated and persistent on getting the CjeCas9 to cleave DNA, and in the end it did! Thanks for being thoughtful, I have been using the pair of scissors and the screwdriver regularly for the past 3 years. **Tess**, you juggled two projects at once well, and laid the basis for the Cpf1-*B. smithii* work. You demonstrated great accuracy, dedication and eagerness to learn, which made working with you very smooth. **Valentijn**, I am really proud of your growth (as a researcher, but also personally) during the 6-month period of your thesis. Your self-confidence, cheerfulness and never-give-up attitude really inspired me! **Maarten**, you were undeniably efficient, with a good eye for detail and independent. Even though we didn't really manage to work on co-cultures, thanks for joining me on a year-long journey, figuring out new techniques, coming up with new ideas, executing them, collaborating with different groups, and having a lot of fun at the end of it all! **Max**, I really admired your reasoning, analytical and critical thinking. I think these qualities will take you far as a researcher. **Patrick**, I am so glad we managed to get on the Skype call from different continents and start working together on the type VU project. You're a great guy with the right combination of intelligence, hard

work and dedication. **Yuxin**, you are extremely hard-working and philomathic. Your independence and commitment to research made working together very enjoyable. **Despoina**, from the first day at the lab to your last, I did not observe any drop in your motivation, dedication or commitment. Your enthusiasm towards learning and mastering new things is inspiring. Always keep that with you! **Mauricio**, I had a great time working with you, your enthusiasm and kindness was infectious. **Adini**, you're too kind, sweet and a very curiousity-driven researcher. I loved answering every relevant question from you, which also made me learn a lot. Thanks for shuttling your time between the NIOO-MIB collaboration so competently. **Justin**, your "jokes" did crack me up from time to time. It was so easy and smooth working with you, and your commitment to your work lifted up a zombie project and brought it back to life. **Jasper**, you were pure joy to work with. You're smart, creative, critical, and possess great people's skills. This deadly combination is going to take you places. I am happy you've decided to go for a PhD as I think you're going to be great at it! **Tahseen**, I am glad you joined me as my last student during my PhD. Your motivation and inquisitiveness inspired me to come up with new ideas. I am very happy to see you evolving into an independent researcher. I would also like to specially thank you for being so understanding during the final leg of writing this thesis. Even though you were not directly under my supervision, many thanks to **Jorik**, **Finn**, **Ana** and **Eugene** for all the interesting and fun-filled discussions. I am truly grateful to each one for your commendable effort. Working with you has truly enriched my PhD, and I hope you learned as much from me as I did from you. I am happy to see that many of you chose to continue your careers in (academic) science, and I can't wait to hear all about the great work you'll do in future!

I thoroughly enjoyed being part of the supervision team of iGEM 2017, together with **Raymond**, **Christian**, **Rob**, **Emma**, **Marta**, **Wen**, and **Rik**. I learnt so much within a short period of time. All thanks to the talented team – **Niels**, **Jurre**, **Bart**, **Natalia**, **Linda**, **Stijn**, **Tom**, **Sabine**, **Jose**, and **Mark** – that came back from the Jamboree with many laurels! Thanks all of you for a great iGEM experience.

It was good fun organizing the PhD trip to Germany, Sweden and Denmark with **Johanna**, **Yannis**, **Wen**, **Joyshree**, **Erica**, **Hugo**, **Aleks**, and **Jeroen**. I would also like to thank **Jie**, **Emma**, **Martijn**, **Ran**, **Sudarshan**, **Hikmah**, **Tika**, **Indra**, **Daan**, **Emmy**, **Giannis**, **Caifang**, **Ran**, **Erika**, **Ruben**, and **Nong** for the wonderful discussions and times during the trip. **Irene** and **Richard**, thank you for being great PhD trip supervisors.

Thanks, **Detmer** for the coffee corner conversations and career advices. **Anja** and **Heidi**, thank you for always being ready to help with the administrative and bureaucratic queries. The technicians team, **Sjon**, **Philippe**, **Steven**, **Tom**, **Ton**, **Ineke** and **Merlijn**, thank you for the organization, practical help and support to ensure everything runs smoothly in the lab.

I am lucky enough to be surrounded by a great bunch of friends, that have each in their own way contributed to my personal development and ensured I remain sane. Well **Shreyans**, as they say, friends are the family you choose, and you for sure fall in that category. One of the reasons I am currently writing this section in my PhD thesis at BacGen is you and I will be eternally grateful for that. You've been extremely supportive and supported me since our Masters. I am so happy for you and **Jose**, thank you two for making me a part of your wonderful wedding earlier this year. I have really enjoyed the dinners, games, king's days together with **Kal** and **Yannis**. **Kal**, I am glad to have gotten to know you over the past 2-3 years, thanks for the nice conversations, food and ~~*jokes*~~ *mokka kaddis*. I wish you luck with your new venture! **Christina**, our travels, catch-ups and discussions have been truly enriching and have often provided me with the much-needed break from work. **Ju**, you're really a very sweet and kind person, and I feel lucky to have gotten to know you. I thoroughly enjoyed our discussions and giggling sessions. **Bas**, you're one of the most cheerful person I have met, thank you for all the smiles and nice (travel) memories. **Chetna**, thanks for the help and fun times during my MSc days. I cannot wait to meet **Onkar** and **Manu** as soon as I visit India. **Surya**, it was a lot of fun hanging out during the initial years of my PhD. We must plan our next meet up. **Juliana**, **Durk**, and **Ruben**, it's been fun getting together for a game's night occasionally – we must keep it going. **Majic**, **Andrea**, **Iza**, and **Damien**, thank you for the entertaining conversations and the awesome trip to Poland. **Carmen**, It was my pleasure to have you

as my housemate. Thank you for the long discussions and nice dinners together. **Ernst** and **Tineke**, I would like to express my appreciation for making my initial years back at Wageningen very smooth.

I have been away from home for about seven years now, I consider it my good fortune that I have some extremely special friends, that have stood by me through the years. **Shruti**, I know I can count on you for anything and everything. Thanks for being there even after months of absence and always reconnecting just where we left off la st. I am so happy for you and **Fahim**, and I wish you guys the best life has to offer! **Pari**, thanks for being so understanding, for constantly checking on me and encouraging m e. I thoroughly enjoyed your visit, though we couldn't visit as many places as we had originally planned to. Come visit me again, this time, with **Nandini** and let's go exploring more places together. **Rawoof**, thanks for being in touch and supporting me the last years, I cannot wait to meet you soon.

Robin, I remember calling you the "glue" once, and I feel that is still very apt, even after all these years. I was very glad that I could attend your wedding and now I am looking forward to becoming an aunt. It was great fun to have you and **Shahenaaz** visiting me in The Netherlands. **Shravan**, it was really nice to meet you few months ago and catch up on all thing's adult life. To all my dear friends: **Pavi**, **Poo**, **Suresh**, **Ninja**, **Riju**, and **Motu**, thanks for the continous love and support. **Motu**, thanks a lot for your patience and for designing my thesis cover.

To my complete family, thank you for encouraging me in all my pursuits. **Amoma** and **Babu thatha**, you two have always been an inspiration for me and I am grateful for your constant encouragement. **Manju atha**, thank you for standing by me and cheering for me. **Narayan mama**, **Ani** and **Ashwin**, thanks for your good wishes. It was great fun exploring eastern Europe together! **Logu ava** and **Munu thatha**, thank you for always showering me with your blessings and good wishes. I'd also like to extend my gratitude to **Jyanthi atha, Janikiraman mama, Suganthi atha** and **Santhanam mama**. Thanks to all my dear cousins and their respective partners who have always been supportive and fun to discuss all things life with: **Kousi, Sugi, Siddhu, Vandana** and **Archana**.

Ma and **dad**, my deepest gratitude goes to you for your unconditional love and support throughout my life; this thesis wouldn't have been there without you. I shall be indebted to as long as the space–time continuum exists. **Ma**, thank you for being my rock. **Dad**, your belief in me has been my constant source of encouragement. **Nikki**, thanks for being my personal cheerleader at every step of my PhD. I will always strive to make you proud!

Ashwin, I feel extremely privileged to have you by my side. You've been so patient, understanding and supported me through thick and thin. You make every day so exciting and challenge me to be a better person. I am grateful for the love and happiness you've brought to my life! Thank you for being you.

Prarthana Mohanraju
Wageningen, Sep 2019

The research described in this thesis was financially supported by a Netherlands Organisation for Scientific Research (NWO) TOP grant (714.015.001) to Prof. Dr John van der Oost.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by Koushik Gopal.
Lay-out aided by Ashwin A. Candade.
Printed by GVO drukkers & vormgevers B.V Ede.