A COLLECTION OF STORIES ON HOT'N'SHARP DNA ENGINEERING

IOANNIS MOUGIAKOS

2019

IOANNIS MOUGIAKOS

FEEL THE BURN

A COLLECTION OF STORIES ON HOT'N'SHARP DNA ENGINEERING

INVITATION

You are kindly invited to attend the public defense of my PhD thesis

FEEL THE BURN

A COLLECTION OF STORIES ON HOT'N'SHARP DNA ENGINEERING

Friday 15 March 2019
1.30 pm

Aula of Wageningen University

Generaal Foulkesweg 1A
Wageningen

Ioannis Mougiakos

ioannis.mougiakos@wur.nl

Paranymphs:

Elleke F. Bosma
elleke.b@gmail.com

Prarthana Mohanraju
prarthana.mohanraju@wur.nl

Propositions

1. Many breakthroughs in genetic engineering and molecular biology were based on bacterial-phage interaction mechanisms.
   (this thesis)

2. The use of temperature for regulating molecular mechanisms and microbial production profiles holds great potential.
   (this thesis)

3. Enzymatic promiscuity hinders the development of accurate metabolic models.

4. Laboratory cultured meat is not a realistic replacement of real meat from a social, financial and scientific perspective.

5. Vaccine refusal has revived previously eradicated diseases, mostly affecting the offspring of those who refuse to vaccinate.

6. The EU regulatory policy on genetically modified organisms (GMOs) is an act of reality denial.

7. Hard work increases one's chances to get lucky.

Propositions belonging to the thesis entitled
"Feel the burn: a collection of stories on hot'n'sharp DNA engineering"

Ioannis Mougiakos
Wageningen, 15 March 2018

Feel the burn:

A collection of stories on hot'n'sharp DNA engineering

Ioannis Mougiakos

Thesis committee

Promotors
Prof. Dr John van der Oost
Personal chair at the Laboratory of Microbiology
Wageningen University & Research

Prof. Dr Richard van Kranenburg
Special Professor Bacterial Cell Factories
Wageningen University & Research
Corporate Scientist Cell Factories and Team Leader Strain Development
Corbion, Gorinchem

O ther members
Prof. Dr M. Kleerebezem , Wageningen University& Research
Prof. Dr A.T. Nielsen, DTU Center for Biosustainability, Lyngby, Denmark
Prof. Dr C. Beisel, Helmh oltz Institute for RNA-based Infection Research, Würzburg, Germany
Dr H.P. Goorissen, Corbion, Gorinchem

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

Feel the burn:

A collection of stories on hot'n'sharp DNA engineering

Ioannis Mougiakos

Thesis
submitted in fulfillment 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 15 March 2019
at 1.30 p.m . in the Aula.

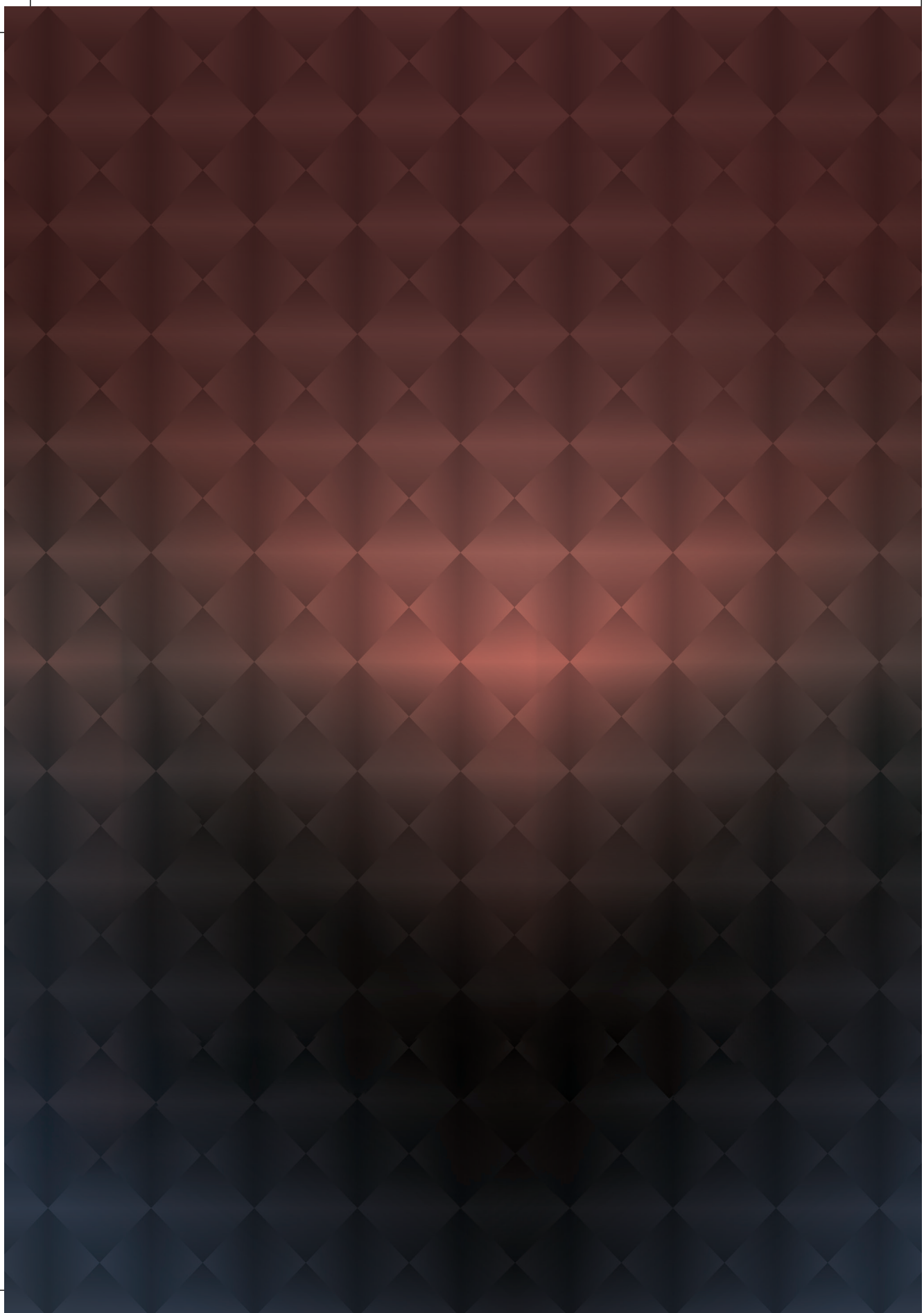
Ioannis Mougiakos
Feel the burn: A collection of stories on hot'n'sharp DNA engineering, 302 pages.

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

ISBN 978-94-6343-409-6
DOI 10.18174/468570

Για τη Γιαγιά, την Ελίζα,
τη Μαμά και τον Μπαμπά

TABLE O F CO NTENTS



Chapter 1

GENERAL INTRODUCTION AND THESIS OUTLINE

GENERAL INTRODUCTION

SHORT HISTORY OF MICROBIOLOGY & MICROBIAL BIOTECHNOLOGY

The founding stone of microbiology as a scientific discipline was set in the 1670s (Fig. 1). This was the decade that Antonie van Leeuwenhoek and Robert Hooke, the first acknowledged microscopists and the "Fathers of Microbiology", started corresponding their microscopic observations on "animalcules" ("tiny animals") to the Royal Society[1]. Two centuries later, Louis Pasteur experimentally connected the spoilage of milk and wine with bacterial activity, also postulating that bacteria could be the causative agents of human diseases[2]. This hypothesis was confirmed by Robert Koch, the founder of modern bacteriology, towards the end of the 19th century; Koch identified the bacteria that cause anthrax (*Bacillus anthracis*), tuberculosis (*Mycobacterium tuberculosis*) and cholera (*Vibrio cholerae*), introduced the concept of infectious diseases, and established the first pure bacterial culturing methods, which are still used today[2]. Except for infection agents it was later shown that bacteria are an important contributor to the human microbiome and numerous studies have demonstrated the close relation between the human microbiome composition and human health[3].

Microbes have been extensively used throughout the thousands of years of human history for basic biotechnological applications (Fig. 1), such as the production of foods (e.g. bread and cheese) and beverages (e.g. wine and beer)[2,4]. Nonetheless, the conscious and controlled biotechnological exploitation of microbes began only upon the identification of bacteria, yeasts and fungi as entities and of microbial fermentation as a process. The use of pure yeast cultures as a practice for beer brewing, aiming to keep the quality and the taste of a beer constant, started only towards the end of the 19th century. Additionally, during World War I chemicals, such as acetone and glycerol that were used for the production of explosives[2], respectively, were overconsumed and found to be in short supply. The replenishment of these products was achieved via bacterial fermentations, and it was then that the "biotechnology" definition was coined as the conversion of raw material to valuable products by biological systems. Since then, many additional useful products, such as chemicals and health related molecules, have been produced via microbial based biotechnological processes. Moreover, many ancient biotechnological production processes, such as food and beverage fermentations, have been improved and protocols have been established for performing them in a more controlled manner[2,4].

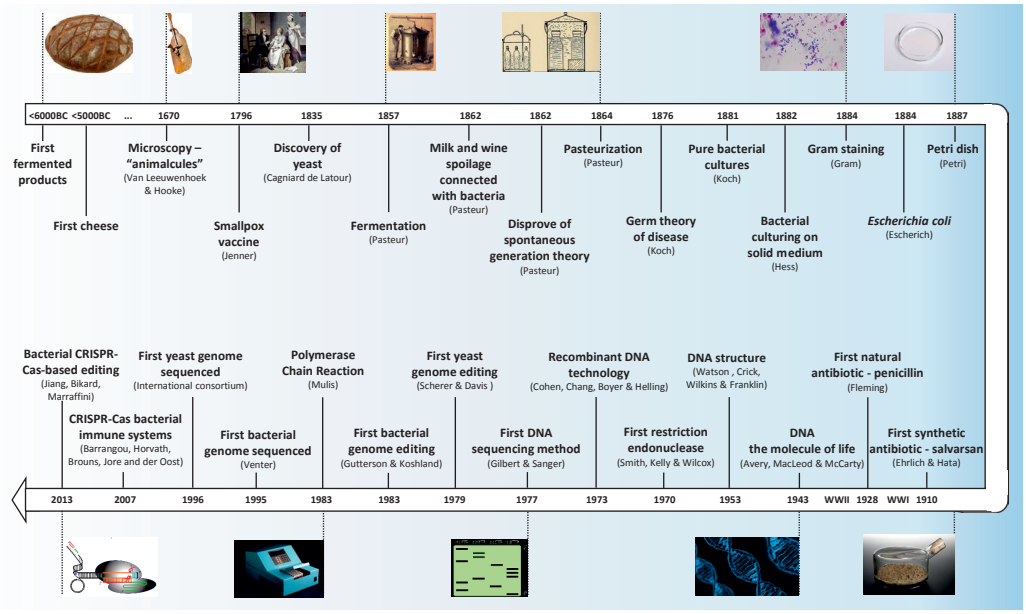


Figure 1. Timeline of the summarised history of microbial biotechnology, molecular biology and genome engineering.

MOLECULAR BIOLOGY AND GENOME ENGINEERING

In the forty-year long period right after World War II, a series of milestones in the field of molecular biology extended the abilities to modify the production pattern and capacities of microbes in a targeted manner, providing the foundation of the microbial genome engineering field (Fig. 1). These milestones include the discovery of DNA as the molecule of life[5], the elucidation of the DNA structure[6], the discovery of the restriction enzymes[7], the transformation and propagation of DNA plasmid material into *E. coli*[8] and the development of the polymerase chain reaction[9] (Fig. 1). In the middle of the twentieth century it was shown that radiation[10] and certain chemical agents[11] can increase the natural rate with which the genome of an organism is mutated, while a later it was demonstrated that transposition genetic elements can be randomly inserted in genomes, disrupting genes and hence their function[12]. These approaches, as well as laboratory evolution, were extensively used for decades in industry and academia, aiming to develop randomly mutated microbial strains with desired traits and enhanced production capacities[13]. The random nature of these tools resulted in strain development processes with low efficiencies and high screening requirements, making clear the need for the construction of efficient and targeted genome engineering tools.

The targeted microbial genome engineering era opened up in the late 1970s with the targeted introduction of heterologous DNA fragments into the *Saccharomyces cerevisiae* genome via transformation and homologous recombination (HR)[14]. As for the first bacterial targeted genome editing applications, these came slightly later and were also relying on the native bacterial HR systems and provided HR templates carrying the desired modifications[15]. In the cases of tools that employed replicating DNA vectors, an extensive screening process was required upon transformation of the vector to the cells, aiming at the identification of the clones where a single cross over event (SCO) had occurred. Subsequently, the use of a plasmid-borne counter-selection system, such as the *pyrF*-system, the *lacZ*-system and the *tdk-hpt* system[16], was used for the selection of clones that lost the editing plasmid. This kind of tools commonly resulted in wild type revertant clones, especially for genes that code for important functions. For the tools that were based on suicide DNA vectors, an antibiotic resistance marker gene was introduced to the HR template for selection purposes, restricting the execution of sequential editing rounds for a species to the number of available marker genes[16]. Moreover, an extensive screening process was required for identification of the clones where a double cross-over (DCO) event occurred resulting in the elimination of the suicide plasmid from the genome together the wild type targeted gene. Further employment of recombinase-based systems (Cre-lox system, FLP-FRT system)[16] allowed for marker gene recycling. Nonetheless, genome editing using these methods was proved to be time consuming, laborious, frequently resulted in chromosomal scars and could induce undesirable chromosomal rearrangements. As an alternative, group II intron retrotransposition systems (ClosTron system, TargeTron system, ThermoTargeTron system)[17-20] have been developed and successfully employed for a limited number of bacterial species, nonetheless they often resulted in unstable insertions and polar effects, making their usefulness even more restricted. Additionally, the developed recombineering-based methods (phage λ-Red system, Rac prophage-RecET system)[21,22] that rely on the heterologous expression of viral recombination systems and immediate transformation of the targeted strain with ssDNA or dsDNA linear fragments, are easy-to-apply and time-efficient; however, this system only provides relatively low engineering efficiencies for just a handful of model bacteria[23-26]. In sum, all the previously described tools were proved to be generally inefficient for editing the genome of non-model bacteria with restricted genetic toolbox, low transformation and recombination efficiencies.

Worth mentioning is that, starting from the last couple of years of the 20th century, the eukaryotic genome engineering field met a new era with the development of the ZFNs (zi-

nc-finger nucleases) and the transcription activator-like effector nucleases (TALENs)[27,28], giving great hope that these tools could be adapted for prokaryotic genome engineering too. These chimeric nucleases are constructed as a "chain" of ZF and TALE DNA-binding domains fused to the non-specific DNA cleavage domain of the FokI restriction enzyme. The ability of FokI-dimers to introduce double stranded DNA breaks (DSDBs) in a sequence specific and programmable manner was exploited for stimulation of the error-prone non-homologous end-joining (NHEJ) or homology directed repair mechanisms mostly in eukaryotic cells, resulting in targeted mutations in their genomes[27,28]. Nonetheless, the complexity in the designing and construction process of these tools discouraged their extensive use for bacterial genome engineering purposes. Hence, the need for user-friendly, efficient and generally applicable bacterial genome editing tools remained.

MICROBIAL METABOLIC ENGINEERING: THE CRISPR-CAS ERA

Microbial metabolic engineering is a practice dedicated to the exploration of the metabolism and the exploitation of the production capacities of industrially relevant microorganisms. Over the last decades, extensive microbial metabolic engineering studies have been conducted both for fundamental purposes and for biotechnological applications. These studies have resulted in the impressive gain of insight in the metabolism of -mostly- model microorganisms, such as *E. coli*, *B. subtilis* and *S. cerevisiae*, and in the development of strains that produce a wide range of compounds at high titers. Hence, it seems like a safe bet to hypothesize that thorough investigation of the metabolism of additional, non-model microorganisms holds great promise on the extension of the products range and the further improvement of the production processes.

All the past and current metabolic engineering studies rely on at least one of the following approaches: 1) episomal- or chromosomal-based overexpression of the native rate-limiting enzyme(s) of the desired production pathway, 2) episomal- or chromosomal-based overexpression of the heterologous enzyme(s) of the desired production pathway, 3) downregulation of competing pathways either via deletion or via down regulation of the corresponding pathway gene(s), 4) enzyme engineering. All these approaches depend on synthetic biology and genome engineering applications, which are organism-specific and rarely applicable to a wide range of microorganisms. Hence, the development of highly effective general genome engineering tools, is very important for further metabolic engineering studies.

In 2007 it was discovered that bacterial and archaeal genomes carry adaptive immune systems against viruses and mobile genetic elements in the form of clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins[29]. The importance of these systems was underlined by the observation that approximately half of the currently sequenced bacterial and archaeal genomes contain at least one CRISPR-Cas system[30]. Moreover, the specificity of the CRISPR-Cas targeting mechanisms was proved to be adaptable in a very simple way, especially when compared to ZFNs and TALENs; small RNA molecules could lead the single or multi-subunit effectors of the CRISPR-Cas systems to their DNA or RNA cleavage targets. Not surprisingly, at the beginning of the 2010s, the first Cas endonuclease was employed for the development of eukaryotic and prokaryotic genome editing and silencing tools[31-34]; it was the Cas9 RNA-guided DNA endonuclease from S*treptococcus pyogenes*, denoted as SpCas9. The first generation of the CRISPR-Cas-based tools for bacterial genome editing, which are reviewed in chapter 2 of this thesis, almost exclusively relied on the introduction of desired modifications to a bacterial genome via HR templates while the SpCas9-targeted DSDBs act as moderate inducers of the HR cellular mechanism and as strict counter selection system in favor of the modified genomes. Moreover, the first CRISPR-Cas-based tool for bacterial transcriptional silencing was based on the targeting of the catalytically inactive ("dead") SpdCas9 variant to the promoter region of the gene of interest, or within the gene sequence for lower silencing effect. Chapter 8 presents the metabolic engineering studies on model and non-model bacteria that relied on these tools for improvement of their production profiles in green chemicals. The remarkable simplicity and efficiency of the SpCas9-based engineering tools lead to the further development of a number of genome engineering related applications, such as base editing[35], EvolvR systems[36] and the use of antiCRISPR systems for improving the control over Cas9-based tools and their editing efficiencies, as described in chapter 6 of this thesis. Additionally, it is anticipated that the evaluation of a rapidly increasing number of Cas effectors with improved characteristics compared to SpCas9, as described in chapter 5 of this thesis for a thermoactive Cas9 homologue, will result in a collection of CRISPR-Cas tools that will further facilitate metabolic engineering studies.

TRANSITIONING TO THE GREEN-CHEMICALS AND BIOFUELS ERA

Fossil resources are currently extensively exploited for covering the exponentially growing global demand for fuels and chemicals. The environmental impacts caused by this exploitation are already visible and impose serious threats on the inhabitability of our planet in the near future. The use of renewable resources for the microbial-based production of chemicals and fuels in a biorefinery is an attractive and environmentally friendly ("green") approach.The complete transition from the fossil-based to the microbial-based production of chemicals and fuels requires the development of financially competitive biotechnological processes. The total cost of each process is determined by the cost of:

1. the fermentation process, including:
    a. the cost of the employed carbon and electron sources, as well as additional nutritional requirements of the microbial workhorse
    b. the cost of the required oxygen, carbon dioxide and energy inputs etc.
    c. the cost of the equipment and related operational costs
2. The products separation and purification process, i.e. the downstream processing

Both cost categories heavily rely on the characteristics, requirements and possible byproducts of the employed microbial production platform, especially when the selling price of the targeted product is low, as is the case for bulk chemicals and fuels. Hence, the careful selection of an appropriate microbial workhorse, according to the requirements of an *a priori* designed production process, is mandatory in order to secure the financial success of a green chemicals or biofuels production process. Additionally, it is required that the selected microbes can be subjected to targeted genetic modifications for metabolic engineering purposes, aiming to achieve the production of heterologous products or to enhance the production of native products, at least up to financially viable levels. Nonetheless, not all the commonly employed model microbial platforms, such as *E. coli*, *S. cerevisiae* and *B. subtilis*, can withstand production conditions (such as high temperatures, low pH, etc.) that in some cases would enable the financial viability of a production process. Moreover, the vast majority of the remaining, non-model microbes are usually not genetically accessible, while if they are, they have limited -if any- genetic and genome engineering toolboxes, a fact that hinders the efficient engineering of their metabolism towards the desired products. Hence, the genetic "taming" of biotechnologically relevant microbes is gaining increasing attention the last years, in parallel with the efforts to transition to a sustainable economic and environmental model.

NON-MODEL BACTERIA WITH BIOTECHNOLOGICAL PERSPECTIVE: *BACILLUS SMITHII, PSEUDOMONAS PUTIDA* AND *RHODOBACTER SPHAEROIDES*

The main focus of this thesis, as presented in Chapters 3, 4 and 5, was the development of Cas9-based efficient genome engineering tools for three biotechnologically interesting non-model bacteria: the moderately thermophilic bacterium *Bacillus smithii* and the mesophilic bacteria *Rhodobacter sphaeroides* and *Pseudomonas putida*.

The biotechnological relevance of *B. smithii* resides in its abilities to utilize both C5 and C6 sugars, to grow in a wide temperature range withstanding wide temperature fluctuations, to tolerate low pH conditions for an extended period of time and to be genetically accessible[37]. These characteristics render *B. smithii* a strain with great potential to become a production platform in a Simultaneous Saccharification and Fermentation (SSF) set-up that employs lignocellulosic biomass as carbon source. Chapter 8 presents further work on the employment of the developed Cas9-based tools towards the exploration and exploitation of its metabolism.

P. putida has attracted a lot of attention for its innate biotechnological potential towards bioremediation[38,39] and biocontrol[40,41] that emanates from its adaptability to harsh physicochemical conditions (such as in the presence of solvents) and its diverse metabolism. Moreover, its ability to consume a wide range of substrates only enhanced this interest[42]. Metabolically engineered *P. putida* strains have already been constructed and used towards the production of bioplastics[39] as well as agricultural and pharmaceutical compounds[42]. Hence, the development of tools that enhance the genome engineering throughput for *P. putida*, like the tool described in this thesis in Chapter 5, will only accelerate the biotechnological exploitation of this bacterium.

The highly adaptable metabolism of *R. sphaeroides* has been extensively investigated in diverse studies with fundamental and biotechnological interests[43]. *R. sphaeroides* is a facultative phototrophic microorganism, grows on a wide range of carbon sources and can use many different electron acceptors for respiration purposes[44,45]. The studies that employed *R. sphaeroides* for bioremediation, photoheterotrophic hydrogen production and heterotrophic terpene biosynthesis have underlined its biotechnological versatility and relevance[46,47]. It is anticipated that the development of efficient engineering tools for *R. sphaeroides*, like the one described in Chapter 4, will contribute to the elucidation of its complex metabolism as well as to the construction of readily applicable hydrogen or terpene production strains.

THESIS OUTLINE

As briefly introduced above in **Chapter 1**, microbes have been extensively exploited throughout the reported human history for biotechnological applications, especially in the field of food biotechnology. Nowadays, the microbial production of green chemicals and fuels gains increasing attention, especially due to the ease in the construction of metabolically engineered microorganisms with high production capacities. This ease was achieved due to the development of efficient genome engineering tools. This thesis describes the development of genetic tools as well as novel CRISPR-Cas9-based and antiCRISPR-based prokaryotic genome engineering tools, for mesophilic and thermophilic bacteria, and the use of these tools for the metabolic exploration and exploitation of the moderate thermophilic bacterium *B. smithii*.

Chapter 2 discusses the currently established and extensively used CRISPR-Cas-based tools for engineering the genomes of model and non-model prokaryotes. Additionally, it analyses the possibilities to improve and expand these tools for further enhancement of their throughput.

Chapter 3 describes the first application of *Streptococcus pyogenes* Cas9 (spCas9)-based genome editing on a moder ate thermophilic bacterium, *B. smithii* . Plasmid-borne editing templates introduce the desired modifications to the *B. smithii* genome at higher culturing temperatures via homologous recombination (HR), while consequent culturing at lower temperatures allows for spCas9-based introduction of lethal double-stranded DNA breaks to the genomes of the non-edited cells, acting as a counter-selection mechanism.

Chapter 4 describes the development of the first SpCas9-based genome editing tool for the α-proteobacterium *Rhodobacter sphaeroides*. Plasmid-borne HR editing templates introduce the desired modifications to the *R. sphaeroides* genome and the SpCas9-based introduction of lethal genomic DNA breaks allows the survival and easy selection of the edited cells. Additionally, this chapter shows that the native *R. sphaeroides* non-homologous end joining (NHEJ) mechanism can repair the SpCas9-induced genomic DNA breaks and is based on the LigD and not on the predicted Ku homologs, strongly suggesting the existence of an unprecedented bacterial NHEJ system.

Chapter 5 focuses on the characterization of ThermoCas9, a Cas9 orthologue from the type II-C CRISPR-Cas system of the thermophilic bacterium *Geobacillus thermodenitrificans*. Amongst others, it is demonstrated that ThermoCas9 is *in vitro* active up to 70 °C, motivating its further use for the development of uniquely versatile and robust Cas9-based genome engineering tools, applied on *B. smithii* at 55 °C and on *P. putida* at 37 °C.

Chapter 6 studies the potential to harness the activity of an antiCRISPR protein for genome engineering purposes. The AcrIIC1$_{Nme}$ antiCRISPR protein from *Neisseria meningitidis* is proven to capture two thermo-active Cas9 orthologues (ThermoCas9 and GeoCas9) in a DNA-bound, but catalytically inactive form *in vivo* in *E. coli* at 37°C. The Cas9/AcrIIC1$_{Nme}$ complexes have a transcriptional silencing effect with efficiency comparable to the catalytically "dead" Thermo-dCas9 and Geo-dCas9 variants. Finally, this chapter describes the controllable and efficient Cas9/AcrIIC1$_{Nme}$-based tool for coupled silencing and targeting in bacteria.

Chapter 7 reviews recent metabolic engineering studies that aimed to improve the production capacities of model and non-model prokaryotes and that were executed employing CRISPR-Cas-based DNA/RNA engineering technologies. Additionally, this chapter analyses how novel CRISPR-Cas systems from diverse environments will amplify the number of CRISPR-Cas-based engineering applications and extend the variety of organisms on which these applications can be employed for metabolic engineering purposes.

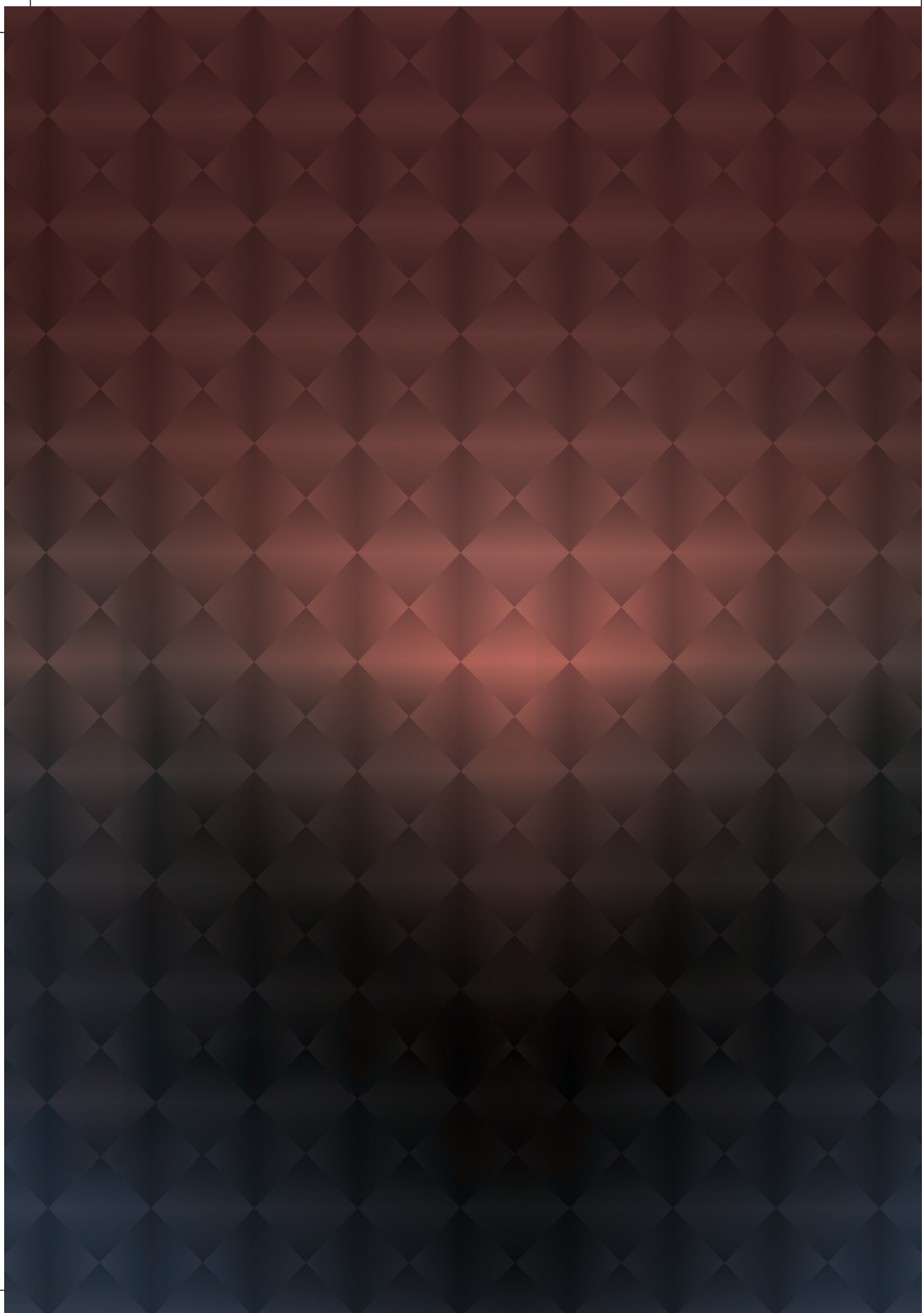
Chapter 8 presents the work performed on studying the metabolism of *B. smithii* via the employment of the developed Cas9-based genome editing tools. The described work aimed to either relieve the NAD^+ deficiency of the *B. smithii* Δ*ldhL* strain under oxygen limited conditions or to study the non-canonical acetate production pathway of *B. smithii*. Moreover, *B. smithii* was cultured under at different temperatures aiming to discover differences in growth, production and protein content. Finally, a *B. smithii* strain was engineered to produce higher amounts of dicarboxylic acids and lower amounts of acetate.

Chapter 9 summarises the work presented in this thesis, discusses its general scope and puts the presented results in perspective. Moreover, further work towards the development of additional genetic tools for *B. smithii* and genome editing tools for thermophilic and mesophilic bacteria is also discussed together with the encountered bottlenecks and future perspectives. Finally, additional work aiming the identification of the *B. smithii* acetate production pathway and the enhancement of its dicarboxylic acids production levels is presented and future engineering avenues are discussed.

REFERENCES

1. Gest, H., The discovery of microorganisms by Robert Hooke and Antoni van Leeuwenhoek, Fellows of The Royal Society. Notes and Records of the Royal Society of London, 2004. 58(2): p. 187-201.
2. Kirk, R.E. and D.F. Othmer, Food and feed technology. 2008: Wiley Interscience.
3. Gilbert, J.A., et al., Current understanding of the human microbiome. Nature medicine, 2018.24(4):p.392.
4. Bourdichon, F., et al., Food fermentations: microorganisms with technological beneficial use. International journal of food microbiology, 2012. 154(3): p. 87-97.
5. Avery, O.T., C.M. MacLeod, and M. McCarty, Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. Journal of experimental medicine, 1944. 79(2): p. 137-158.
6. Watson, J.D. and F.H. Crick, Molecular structure of nucleic acids. Nature, 1953. 171(4356): p. 737-738.
7. Meselson, M. and R. Yuan, DNA restriction enzyme from *E. coli*. Nature, 1968. 217(5134): p. 1110.
8. Cohen, S.N., et al., Construction of Biologically Functional Bacterial Plasmids *In Vitro*. Proceedings of the National Academy of Sciences, 1973. 70(11): p. 3240-3244.
9. Mullis, K.B., et al., Process for amplifying, detecting, and/or-cloning nucleic acid sequences. 1987, Google Patents.
10. Muller, H.J., Artificial transmutation of th e gene. Science, 1927. 66(1699): p. 84-87.
11. Auerbach, C., J.M. Robson, and J.G. Carr, The Chemical Production of Mutations. Science, 1947. 105(2723): p. 243-247.
12. McClintock, B., The origin and behavior of mutable loci in maize. Proceedings of the National Academy of Sciences, 1950. 36(6): p. 344-355.
13. Derkx, P.M., et al. The art of strain improvement of industrial lactic acid bacteria without the use of recombinant DNA technology. in Microbial cell factories. 2014. BioMed Central.
14. Scherer, S. and R.W. Davis, Replacemen t of chromosome segments with altered DNA sequences constructed *in vitro*. Proceedings of the National Academy of Sciences, 1979. 76(10): p. 4951-4955.
15. Gutterson, N.I. and D.E. Koshland, Replacement and amplification of bacterial genes with sequences altered *in vitro*. Proceedings of the National Academy of Sciences, 1983. 80(16): p. 4894-4898.
16. Schweizer, H.P., Bacterial genetics: past achievements, present state of the field, and future challenges. Biotechniques, 2008. 44(5): p. 633-641.
17. Wang, Y., 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 micro biology, 2013: p. AEM. 00971-13.
18. Esvelt, K.M. and H.H. Wang, Genome‐scale engineering for systems and synthetic biology. Molecular systems bio logy, 2013. 9(1): p. 641.
19. Frazier, C.L., et al., Genetic manipulation of Lactococcus lactis by using targeted group II introns: generation of stable insertions without selection. Applied and environmental microbiology, 2003. 69(2): p. 1121-1128.
20. Mohr, G., et al., A targetron system for gene targeting in thermophiles and its application in. *Clostridium thermocellum*. PloS one, 2013. 8(7): p. e69032.
21. Datsenko, K.A. and B.L. Wanner, One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proceedings of the National Academy of Sciences, 2000. 97(12): p. 6640-6645.
22. Pines, G., et al., Bacterial recombineering: genome engineering via phage-based. homologous recombination. ACS synthetic biology, 2015. 4(11): p. 1176-1185.
23. Ellis, H.M., D. Yu, and T. DiTizio, High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. Proceedings of the National Academy of Sciences, 2001. 98(12): p. 6742-6746.
24. Oh, J.-H. and J.-P. van Pijkeren, CRISPR–Cas9-assisted recombineering in Lactobacillus reuteri. Nucleic acids research, 2014. 42(17): p. e131-e131.
25. Swingle, B., et al., Recombineering Using RecTE from Pseudomonas syringae. Applied and Environmental Microbio logy, 2010. 76(15): p. 4960-4968.
26. Binder, S., et al., Recombineering in *Corynebacterium glutamicum* combined with optical nanosensors: a general strategy for fast producer strain generation. Nucleic acids research, 2013. 41(12): p. 6360-6369.
27. Carroll, D., Genome engineering with targetable nucleases. Annual review of biochemistry, 2014. 83: p. 409-439.

28. Gaj, T., C.A. Gersbach, and C.F. Barbas, ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in Biotechnology, 2013. 31(7): p. 397-405.
29. Barrangou, R., et al., CRISPR provides acquired resistance against viruses in prokaryotes. Science, 2007. 315(5819): p. 1709-1712.
30. Makarova, K.S., et al., An updated evolutionary classification of CRISPR–Cas systems. Nature Reviews Microbiology, 2015. 13(11): p. 722.
31. Mali, P., et al., RNA-guided human genome engineering via Cas9. Science, 2013. 339(6121):p.823-826.
32. Jinek, M., et al., RNA-programmed genome editing in human cells. elife, 2013. 2: p. e00471.
33. Jiang, W., et al., RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature biotechnology, 2013. 31(3): p. 233.
34. Larson, M.H., et al., CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nature protocols, 2013. 8(11): p. 2180.
35. Komor, A.C., et al., Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature, 2016. 533(7603): p. 420.
36. Halperin, S.O., et al., CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. Nature, 2018. 560(7717): p. 248-252.
37. Bosma, E.F., Isolation, characterization and engineering of *Bacillus smithii* : a novel thermophilic platform organism for green chemical production. 2015, Wageningen University: Wageningen.
38. Gomes, N.C., et al., Effects of the inoculant strain Pseudomonas putida KT2442 (pNF142) and of naphthalene contamination on the soil bacterial community. FEMS microbiology ecology, 2005. 54(1): p.21-33.
39. Ward, P.G., et al., A two step chemo-biotechnological conversion of polystyrene to a biodegradable thermoplastic. Environmental science & technology, 2006. 40(7): p. 2433-2437.
40. Amer, G. and R. Utkhede, Development of formulations of biological agents for management of root rot of lettuce and cucumber. Canadian journal of microbiology, 2000. 46(9): p. 809-816.
41. Validov, S., et al., Selection of bacteria able to control Fusarium oxysporum f. sp. radicis‐lycopersici in stonewool substrate. Journal of applied microbiology, 2007. 102(2): p. 461-471.
42. Poblete-Castro, I., et al., Industrial biotechnology of Pseudomonas putida and related species. Applied microbiology and biotechnology, 2012. 93(6): p. 2279-2290.
43. Imam, S., D.R. Noguera, and T.J. Donohue, Global insights into energetic and metabolic networks in Rhodobacter sphaeroides. BMC systems biology, 2013. 7(1): p. 89.
44. Tabita, F.R., The biochemistry and metabolic regulation of carbon metabolism and CO_2 fixation in purple bacteria, in Anoxygenic photosynthetic bacteria. 1995, Springer. p. 885-914.
45. Zannoni, D., B. Schoepp-Cothenet, and J. Hosler, Respiration and respiratory complexes, in the purple phototrophic bacteria. 2009, Springer. p. 537-561.
46. Mackenzie, C., et al., Postgenomic adventures with Rhodobacter sphaeroides. Annu. Rev. Microbiol., 2007. 61: p. 283-307.
47. Imam, S., D.R. Noguera, and T.J. Donohue, Global analysis of photosynthesis transcriptional regulatory networks. PLoS genetics, 2014. 10(12): p. e1004837



CHAPTER 2

NEXT GENERATION BACTERIAL ENGINEERING : THE CRISPR-CAS TOOLKIT

Ioannis Mougiakos[1,#], Elleke F. Bosma[1,#],Willem M. de Vos[1], Richard van Kranenburg[1,2], John van der Oost[1*]

[1]Laboratory of Microbiology, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, The Netherlands.

[2]Corbion, Arkelsedijk 46, 4206 AC Gorinchem , The Netherlands.

[#]Contributed equally

*Corresponding author

Chapter adapted from publication:

Trends Biotechnol. 2016 Jul;34(7):575-587. doi: 10.1016/j.tibtech.2016.02.004

ABSTRACT

The increasing demand for environmentally friendly production processes of green chemicals and fuels has stimulated research in microbial metabolic engineering. CRISPR-Cas-based tools for genome editing and expression control have enabled fast, easy and accurate strain development for established production platform organisms, such as *Escherichia coli* and *Saccharomyces cerevisiae*. However, the growing interest in alternative production hosts for which genome editing options are generally limited, requires further development of such engineering tools. In this review we discuss established and emerging CRISPR-Cas-based tools for genome editing and transcription control of model and non-model bacteria, and we analyse the possibilities for further improvement and expansion of these tools for next generation bacterial engineering.

- What is the efficiency of bacterial genome editing by CRISPR-Cas variants such as Cpf1 or C2C1, as compared to Cas9? In particular, could such systems enhance recombination-mediated genome editing?
- Can endogenous CRISPR-Cas systems substitute SpCas9 for genome editing purposes in non-model bacteria that do not support SpCas9 activity?
- How to rapidly characterize novel CRISPR-Cas-systems, including PAM-identification, so they can be applied for genome editing?
- Can CRISPR-Cas-systems be exploited for purposes beyond genome editing?
- Whereas gene silencing seems highly effective in prokaryotes, gene activation needs further improvement – how can this be achieved?
- Can NHEJ in bacteria be established heterologously, so it can be used as general means for CRISPR-Cas-based gene disruption?

KEYWORDS

CRISPR-Cas, bacteria, recombineering, genome editing, Cas9

G LO SSARY

- **Clustered regularly interspaced short palindromic repeats (CRISPR)**: a bacterial or archaeal DNA array constituted of small (30-45 nt long) sequences, usually of foreign origin, which are separated by (almost) identical repeat sequences of similar size.
- **CRISPR-associated (Cas) enzymes**: enzymes encoded by *cas* genes that generally reside in close proximity to a CRISPR array, taking part in any of the three stages of the CRISPR–Cas based immunity.
- **CRISPR locus**: a bacterial or archaeal DNA locus constituted of a CRISPR array, its corresponding Cas genes and possible ancillary modules.
- **CRISPR RNA (crRNA)**: an RNA molecule that guides the targeting of CRISPR-Cas systems. It originates from the CRISPR locus and is comprised of the processed transcript of a spacer and a CRISPR array repeat.
- **Double-Stranded DNA Break (DSDB)**: damage of the DNA such that the backbones of both strands are cleaved simultaneously.
- **Non-Homologous End Joining (NHEJ)**: An error-prone DSDB-repair mechanism, based on the concerted activities of DNA-binding protein Ku and ligase LigD.
- **Protospacer**: a DNA-sequence, targeted for cleavage by a CRISPR-Cas system, that is identical to a spacer sequence of the corresponding CRISPR-array.
- **Protospacer adjacent motif (PAM)**: the short (2-8 nt) conserved sequence adjacent to a protospacer, mandatory for recognition and targeting by CRISPR-C as systems (except Type III), as a means to discriminate self (CRISPR spacer) from non-self (invader protospacer).
- **Recombineering (recombination-mediated genetic engineering)**: genetic engineering tool mediated by phage-derived recombination systems, such as λ-Red or RecET, that facilitate the homologous recombination between ssDNA or dsDNA fragments with complementary genome sequences.
- **Seed**: the first 8-12 bp of a protospacer immediately adjacent to the PAM-sequence; required together with the PAM for successful targeting.
- **Single-Stranded DNA Break (SSDB)**: cleavage of the DNA such that the backbone of one of the strands is nicked while the other one remains intact.
- **Single guide RNA (sgRNA)**: a synthetic chimera combining the crRNA and tracrRNA into a single CRISPR guide.
- **Spacer**: short (30-45 nt) DNA sequences derived from target genetic elements; located in the CRISPR array and flanked by repeat sequences.
- **Trans-activating CRISPR RNA (tracrRNA)**: an RNA molecule encoded by Type II CRISPR loci, that hybridizes to the repeat parts of crRNAs. The crRNA:tracrRNA hybrid is required for Cas9 activation.

FROM EXPLORATION TO EXPLOITATION

The circular economy requires sustainable and biobased alternatives for generating fuels and chemicals that are produced via microbial fermentation. High-throughput genome editing tools are essential for the development of economically viable production organisms. Therefore, ample research focuses on adapting the efficient use of CRISPR -Cas9-based genome editing in eukaryotes for applications in prokaryotes. The main bottleneck of this adaptation process is the limitations of the bacterial DNA -repair systems.

Less than 10 years ago it was discovered that the bacterial and archaeal Clustered regularly interspaced short palindromic repeats (CRISPR)-**CRISPR-associated (Cas)** (see glossary) systems are in fact RNA-guided adaptive immune systems, conferring sequence-specific resistance against invasive mobile genetic elements (MGEs)[1,2]. Since then, considerable insight in the molecular events has been acquired and the CRISPR-Cas based microbial immunity is now divided into the 3 distinct molecular stages of acquisition, expression and interference[3] (Table 1). During the acquisition stage the Cas1 and Cas2 proteins recognize and digest invasive MGEs to sample short DNA sequences from them and introduce these into the **CRISPR** array as **spacers**. The spacers constitute the memory of the system and efficiently vaccinate the microbe against the corresponding MGEs. The transcription of the array as one large premature transcript occurs during the expression stage and starts from the leader sequence. **Cas enzymes** recognize and bind to the premature transcript, which is then processed into smaller mature **CRISPR RNAs (crRNAs)** either by specific Cas nucleases or by the cellular RNase III. These mature crRNAs are comprised of two parts: the transcript of a spacer followed or preceded by the transcript of a repeat. During the interference stage the crRNAs guide the Cas nucleases to target and cleave **protospacer** sequences in invading MGEs.

To date, 6 (I-VI) types of CRISPR-Cas systems have been identified and divided into two major classes, class 1 and class 2, according to the complexity of their interference machinery[4-6] (Table 1). The CRISPR-Cas class 1 comprises types I, III and the putative type IV, all of which confer immunity to their hosts, based on multi-subunit, Cascade-like effector-complexes (Table 1). The CRISPR-Cas class 2 encompasses types II, V and VI, which base their functionality on a large, single protein, multi-domain nuclease (Cas9, Cpf1, C2C1, C2C3 and C2C2) (Table 1). The type II CRISPR-Cas systems confer immunity to their hosts through the Cas9-mediated introduction of sequence-specific, blunt, **Double-Stranded DNA Breaks (DSDBs)** to the targeted MGE protospacers. The Cas9 endonuclease has two nickase domains, an HNH and a RuvC, and is active when loaded with a small guide CRISPR-RNA:**trans-acti-**

vating CRISPR RNA (crRNA:**tracrRNA**) duplex. The spacer part of the crRNA molecule is responsible for the specificity of the Cas9 activity, due to its complementarity to one of the strands of the targeted protospacer (Fig. box 1). A short and conserved **protospacer adjacent motif (PAM)**, flanking the 3'-end of the protospacer, is required for Cas9-mediated cleavage activity[7, 8]. This is crucial for avoiding autoimmunity: the absence of a PAM motif flanking the spacers of the CRISPR array on the host's chromosome prevents lethal self-targeting events.

Class	Type	Subtypes	Acquisition	Expression	Interference		
				Processing of the pre-crRNA transcript	Target binding[a,b]	Target cleavage	Type of nuclease activity
1	I	A-F,U	Cas1, Cas2, Cas4	Cas6	Cas7, Cas5, Cas8 (LS), SS	Cas3	3'-5' exonuclease (SSDB) [55]
	III	A-D	(Cas1, Cas2)	(Cas6)	[c,d] Cas7, Cas5, Cas10 (LS), SS	Cas7 Cas10	RNase/DNase [56, 57]
	IV	-	Unknown	unknown	Cas7, Cas5, Csf1, (SS)	unknown	unknown
2	II	A-C	Cas1, Cas2, (Cas4), Cas9	RNase III (Cas9)	Cas9	Cas9	Blunt-end DSDB inside seed [58]
	V	A-C	(Cas1), (Cas2), (Cas4)	unknown	Cas12a, C2C1, C2C3	Cpf1, C2C1, C2C3	Cpf1: sticky-end DSDB outside seed [59]
	VI	-	Cas1, Cas2	unknown	C2C2	C2C2	unknown (RNase?)

Table 1. Classification of Cas proteins in CRISPR-Cas systems according to their mode of action. Based on[5]. The components that are not present in all the subtypes of the corresponding type are indicated by brackets. [a] Abbreviations: LS: large subunit protein; SS: small subunit protein. In some type I systems SS is fused to LS (Cas8). [b] The Cas proteins that are responsible for the target binding in Type I, III, and probably type IV systems form Cascade-like complexes. [c] The Cas proteins that are responsible for the target binding in Type III-A and D systems form the Csm-complex. [d] The Cas proteins that are responsible for the target binding in Type III-B and C systems form the Cmr-complex.

BOX 1. *IN SILICO* PREDICTION OF CRISPR-CAS ELEMENTS AND PAM

In order to perform *in vitro* or *in vivo* PAM-determination studies (Box 2) for Type II systems, it is necessary to *in silico* predict the CRISPR array of the system, the tracrRNA-expressing module and the Cas9 gene. The CRISPR array is used for the identification of the crRNA module. The tracrRNA-expressing sequence is located either within a 500 bp-window flanking Cas9 or between the Cas genes and the **CRISPR locus**[14]. The tracrRNA should consist of a 5'-sequence with high level of complementarity to the direct repeats of the CRISPR array, followed by a predicted structure of no less than two stem-loop structures and a Rho-independent transcriptional termination signal[15] (Fig. box 1). The crRNA and tracrRNA molecule can then be used to design a chimeric sgRNA module. The 5'-end of the sgRNA consists of a truncated 20 nt long spacer followed by the 16-20 nt long truncated repeat of the CRISPR array. The repeat is followed by the corresponding truncated anti-repeat and the stem loop of the tracrRNA module. The repeat and anti-repeat parts of the sgRNA are generally connected by a GAAA linker[16] (Fig. box 1).

***IN SILICO* PREDICTION OF THE PAM**

In some cases, it is possible to predict the PAM *in silico* by bioinformatic s analyses. The *in silico* predictions are only possible if enough protospacer sequences are available in the databases. The *in silico* process starts with finding hits of spacers from the CRISPR array in the genome of the organism under study against databases such as GenBank. A useful tool to identify the CRISPR loci is "CRISPR finder" (http://crispr.u-psud.fr/Server/). The identified CRISPR loci output can then be loaded into "CRISPR target" (http://bioanalysis.otago.ac.nz/CRISPRTarget/crispr_analysis.html), which will search selected databases and provide an output with matching protospacers. These protospacer sequences should then be checked for unique hits and for complementarity to spacers – for example, mismatches in the seed sequence are likely to be false positive hits and might be excluded from further analysis. Especially hits with prophages and (integrated) plasmids provide confidence that the obtained hits are true positives. Subsequently, the flanking regions (3' for Type II and 5'for Type I) of the remaining, unique protospacer hits can be aligned a nd compared for consensus sequences using tool such as Web Logo (http://weblogo.berkeley.edu/logo.cgi). Resulting consensus sequences can then be verified via *in vivo* or *in vitro* processes (Box 2).

BOX 1. *IN SILICO* PREDICTION OF CRISPR-CAS ELEMENTS AND PAM (continuation)

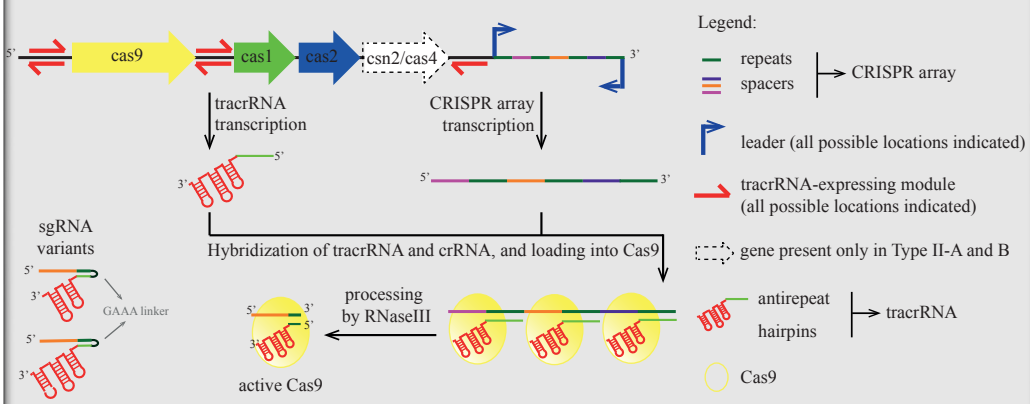


Fig. box 1. Overview of CRISPR-Cas9 locus components and their structures. The figure depicts all the required and optional genomic components for Type II CRISPR-Cas systems, with all the possible locations and orientations of the tracrRNA-expressing module and leader sequence[14]. The CRISPR array consists of interspaced repeats and leads to the expression of the crRNA-modules of the system. The transcription of the CRISPR array into a pre-crRNA transcript and the transcription of the tracrRNA expressing module are followed by the formation of RNA duplexes by these two elements. The tracrRNA consists of a 3'-end 3D structure of two or three loops and a 5'-part which is complementary to the repeats of the CRISPR array. The 5'-half of the crRNA in the duplex is comprised of the transcript of a spacer, while the 3'-half of the crRNA is formed by the transcript of the CRISPR array repeat. The pre-crRNA:tracrRNA duplexes are loaded into Cas9s, leading to the drastic conformational change of the enzymes and their activation. Finally, cellular RNase III enzymes processes the pre-mature complexes into mature crRNA:tracrRNA:Cas9-complexes. The designing and construction of an sgRNA molecule involves the rational truncation of the crRNA and tracrRNA modules, the addition of a linker sequence (mostly GAAA) between these two modules and possibly the truncation of the last tracrRNA hairpin.

The first programmable genome editing tool for bacteria dates from 2013 and it was based on the Cas9 endonuclease from the *Streptococcus pyogenes* type II-A CRISPR-Cas system, denoted as SpCas9[17-20]. Around the same time, SpCas9 has also been extensively exploited for genome editing purposes of a wide range of eukaryotes[18-21]. In bacteria, the CRISPR-Cas9-mediated genome editing tools are based on the heterologous co-

expression of two components: the SpCas9 and a crRNA:tracrRNA duplex that directs the nuclease to the targeted protospacer. Rationally designed chimeric **single guide RNA (sgRNA)** molecules form a convenient and effective alternative to the crRNA:tracrRNA duplex[12]. The SpCas9-crRNA:tracrRNA (or SpCas9-sgRNA) complex introduces DSDBs to the target sites. Cells that carry mutations at these sites can avoid the Cas9-induced DSDBs and survive (Figure 1, Key Figure). In eukaryotes, the **Non-Homologous End Joining (NHEJ)** mechanism can repair DSDBs in an error-prone manner, introducing insertion/deletion (indel) mutations in the targeted site that frequently lead to frame shifts and gene disruption. However, in contrast to the eukaryotic genomes that all encode NHEJ-like systems, the enzymes that are responsible for the bacterial NHEJ mechanism (Ku/LigD) are not encoded by all the bacterial genomes[22]. This fact explains the small number of CRISPR-Cas9-based cases of gene knockouts in bacteria compared to the large number of eukaryotic knockouts using Cas9. In bacterial CRISPR-Cas9-based genome editing, the DSDBs are repaired either by the cellular homologous recombination (HR) system in combination with chromosome/plasmid-borne templates, or by an heterologous **recombineering** system and linear single or double stranded DNA templates. Moreover, it is possible that, at least to some extent, the introduction of DSDBs in the bacterial genomes triggers the recombination mechanisms[17]. In case of genetic engineering, the rescuing templates are designed to introduce chromosomal deletions varying from a single to thousands of nucleotides or deliver genetic sequences of variable size. Recombination occurs at a specific chromosomal site, depending on the Cas9-associated guide; the only limitation with respect to the selected target sequence relates to the requirement for a PAM sequence at the right position. The clones with the desired mutations are easily selected through a negative selection process, due to the CRISPR-Cas9-based elimination of all the wild type clones. This tandem Recombination-CRISPR-Cas9-based counter-selection process has been applied to almost all the CRISPR-Cas9-based genome editing examples published to date.

In this review, we summarize the advances in the CRISPR-Cas-based bacterial genome editing and gene expression-modulation technologies, with an emphasis on the Cas9-based technologies and their application in biotechnologically-relevant bacteria. In addition, we discuss the limitations of bacterial genome editing using these technologies and propose solutions to overcome them for application in next generation bacterial engineering.

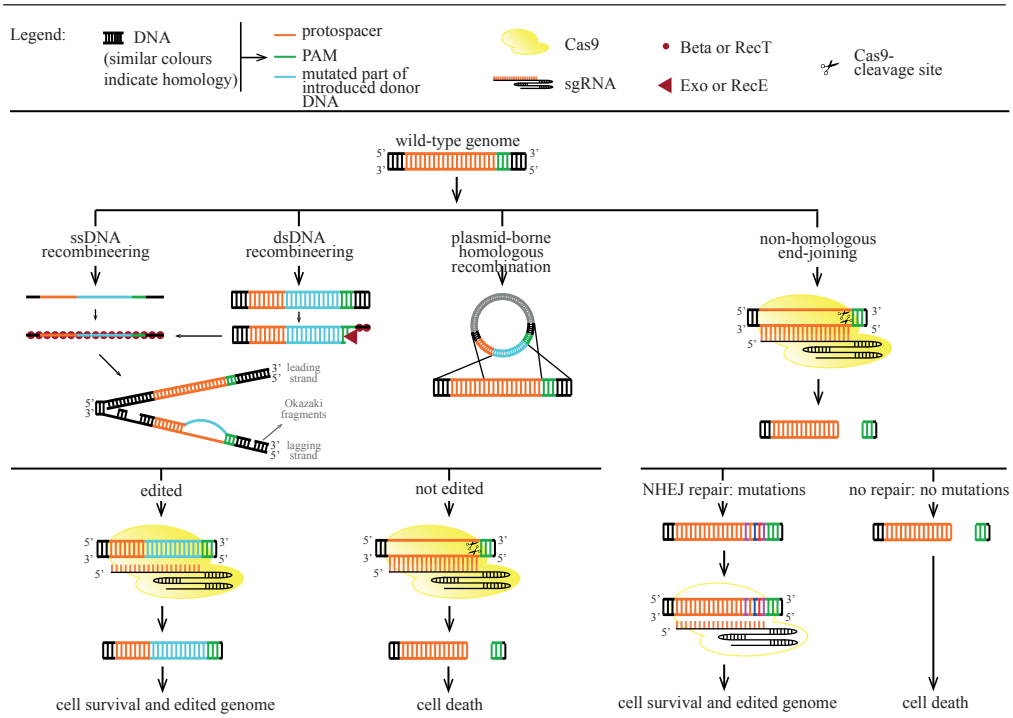


Figure 1. (Key figure) Overview of different systems for Cas9-mediated genome editing and counter-selection or repair mechanisms. Note that in all cases, Cas9 might be replaced by any endogenous system as well. Also, the nickase variant of Cas9 might be used to create single strand DNA breaks instead of double strand DNA breaks. In the figure, only the introduction of mutations in the seed region of the protospacer is shown, but also mutations in the PAM-sequence will prevent Cas9-targeting.

BOX 2. PAM-DETERMINATION METHODS

IN VITRO

This approach has been mainly used for PAM-characterization of type II CRISPR-Cas systems [15,16]. Initially, the *in silico* identification of three elements is required: the *cas9* gene, the CRISPR array and the tracrRNA (Box 1). The process proceeds with *in vitro* transcription and purification of either the crRNA and tracrRNA modules or the sgRNA module, and the design and construction of a targeted PAM-containing vector library16. Each library member contains a constant 20 bp target protospacer, flanked by a 3' 5-7 bp degenerate sequence, which serves as the PAM. The members of the library that contain the right PAM are cleaved by the crRNA: tracrRNA-loaded (or the sgRNA-loaded) Cas9 during an *in vitro* cleavage assay. The cleaved vectors are purified and specific sequencing adaptors are added to the blunt end created approximately 3 bp away from the PAM in the protospacer seed sequence. The PAM of the cleaved vectors compared to the uncleaved library as negative control.

BOX 2. PAM DETERMINATION METHODS

IN VIVO

The *in vivo* PAM characterization of a type II CRISPR-Cas system is based on the heterologous expression of the Cas9 endonuclease and a negative selection process[17,23]. The first step of the *in vivo* process is the same as for the *in vitro*, i.e. the identification of the different elements (Box 1). The second step is the plasmid-based expression of the Cas9 and tracrRNA module in a selected host. A second plasmid encoding the crRNA module is transformed into the same host. The crRNA module guides Cas9 to target the members of a PAM-containing vector library. The members of the library contain a constant 20 bp target protospacer followed by a 3' 5-7 bp degenerate PAM-sequence. After transformation, the vectors from surviving colonies on selective medium are sequenced. The PAM is then extracted as a consensus from the sequences that are not present in the sequencing results as compared to the library prior to transformation (or in a control transformed with a non-targeted plasmid). In case that the organism harboring the studied CRISPR-Cas system is transformable, it is possible to immediately transform the crRNA-expressing vector and the PAM-containing vector-library into that strain. This method eliminates the requirement for heterologous expression of Cas9 and tracrRNA. Similarly, the *in vivo* approach can be followed for the PAM characterization of native type I CRISPR-Cas systems present in the genome of transformable strains.

CRISPR-CAS MEDIATED ENGINEERING OF BACTERIAL GENOMES

SINGLE-STRANDED DNA RECOMBINEERING (SSDR)

Jiang *et al.* were the first to exploit the heterologous expression of SpCas9 to introduce defined mutations or to delete genes in bacterial genomes[17]. During this study, the recombineering strain *E. coli* HME63 and the highly recombinogenic *Streptococcuspneumoniae* crR6Rk were transformed with a pCas9 vector. The vector was responsible for the heterologous expression of the SpCas9 and the tracrRNA molecule in the cells. The next step in both processes was the simultaneous transformation with one or two linear single stranded DNA oligonucleotides for single-stranded DNA recombineering (SSDR) and with the pCRISPR vector. The SSDR templates were designed either to introduce point mutations to the targeted genes, or to delete them. The pCRISPR vector carried and expressed a minimal CRISPR array with one or two spacers that targeted the gene or genes of interes leading to a CRISPR-Cas9 ba-

sed counter-selection against the wild type genes. This tandem use of an SSDR system followed by CRISPR-Cas9-based counter-selection led to a highly efficient genome editing tool (Fig. 1). The tool was successfully applied for additional gene disruptions and large chromosomal insertions and deletions in *E. coli*, relying on co-expression either of a crRNA:tracrRNA pair[24] or of a chimeric sgRNA[25,26] (Table S1). Remarkably, high levels of editing efficiency were achieved when the editing ssDNA oligonucleotides were designed to be identical to the lagging strand of the protospacers; this suggests that the editing oligonucleotides may be recognized as Okazaki fragments by the cellular replication mechanism or may prime the Okazaki fragments synthesis[27].

Soon after the first SSDR- CRISPR-Cas9-based genome editing tool for *E. coli* was published, a similar tool was developed for *Lactobacillus reuteri*, a lactic acid bacterium used as a probiotic[28]. The tool was based on a SpCas9 and tracrRNA expressing vector, a single or dual step procedure and the RecT expressing *L. reuteri* 6475. During the single step procedure the SpCas9 and tracrRNA expressing *L. reuteri* 6475 was simultaneously transformed with a CRISPR-array expressing vector and ssDNA recombineering fragments, while during the dual step procedure *L. reuteri* 6475 was transformed with the same ssDNA recombineering fragments prior to the transformation with the CRISPR-array expressing vector. Whereas both processes yielded almost only recombinant clones, the dual step approach allowed for more replication cycles before the CRISPR-Cas9-based counter selection, leading to more efficient incorporation of the recombineering fragments and elevated number of surviving recombined clones. The same dual step SSDR-CRISPR-Cas9 counter-selection-based editing tool was used for time-efficient *L. reuteri* 6475 chromosomal deletions of up to 1 kb and codon saturation mutagenesis[28].

DOUBLE-STRANDED DNA RECOMBINEERING (DSDR)

A double-stranded DNA recombineering (DSDR)-CRISPR-Cas9-based counter-selection editing system (Fig. 1) was recently developed for *E. coli*[29]. The editing templates for this system were either cloned into an sgRNA-expressing plasmid or introduced into the cells as dsDNA linear fragments. When dsDNA fragments were used as HR templates, the single gene deletion efficiency was high and comparable to the efficiency observed before[17], but the gene insertion efficiency was low and immediately proportional to the length of the homologous regions in the HR template fragments.

The two-plasmid based DSDR-CRISPR-Cas9 counter-selection editing tool was employed for *E. coli* genome editing in many occasions. It was used as an editing tool to introduce and express the *Shewanella frigidimarina* fatty acids production pathway into the *E. coli* MG1655 strain[30], to perform targeted codon substitution in the chromosome of *E. coli* BW25113 Δ*mutS*[31], and to delete genes from the chromosome of another member of the *Enterobacteriaceae* , the biotechnologically interesting *Tatumella citrea* DSM 13699 strain[29], suggesting the broader applicability of the system.

In a recent study,a highly efficient DSDR-CRISPR-Cas9 counter-selection editing tool was used to quickly and efficiently engineer the *E. coli* metabolism[25] (Table S1). The promoters and the RBSs of several genes encoding enzymes involved in the beta-carotene pathway and in the methyl-erythritol-phosphate (MEP) pathway were sequentially substituted by variant expression-controlling elements. The genetic variants with the highest enhancement in beta-carotene production were selected after each editing cycle for further editing, rendering this study an example of Cas9-based *E. coli* metabolic engineering. Interestingly, it was noticed that the mismatch repair mechanism of *E. coli* could efficiently correct point mutations introduced by the SSDR- CRISPR-Cas9 counter-selection editing tool. This revertant issue was solved either by introducing a ssDNA template carrying 9 consecutive mismatch mutations or by using dsDNA editing templates with long homology arms and the DSDR- CRISPR-Cas9 counter-selection editing tool[25].

HOMOLOGOUS RECOMBINATION (HR)

Very recently, plasmid-borne homologous recombination (HR) templates and CRISPR-Cas9-based counter-selection were used a for the simultaneous insertion of 2 or deletion of 3 genes with high efficiency[29]. A non-recombineering HR- CRISPR-Cas9-based counter selection editing tool (Fig. 1), which targets insertion sequences (IS) naturally present in the genome of *E. coli*, was used for large chromosomal deletions of up to 133 kb[32]. In the latter study, a Cas9 nickase variant (SpCas9D10A that encompasses a non-functional RuvC nuclease domain) was employed to introduce targeted Single Stranded DNA Breaks (SSDB)[12,33]. These breaks, when introduced in pairs, in the correct orientation and position around the IS elements, led to 5’ overhangs that induce direct recombination between the neighboring and identical IS elements. This editing approach solves the problem of CRISPR-Cas9-induced lethal DSDBs, and may be applied for genome editing of prokaryotes that lack well-developed recombineering systems.

A CRISPR-Cas9-based genome editing tool was recently developed for the solvent-producing *Clostridium beijerinkii* NCIMB 8052 (Table S1)[34]. The low natural recombination efficiency of the latter species together with the lack of a functional recombineering system led to a single-step HR- CRISPR-Cas9-based editing approach in which the *spcas9* gene was cloned into the same vector as the sgRNA expressing module and the editing template. The CRISPR-Cas9-based counter selection of cells with the desired markerless deletion was highly efficient and the multiplex genome editing potential of this tool is currently under examination[34]. The same one-step tool was employed for the editing of the *Clostridium cellulolyticum* H10 (ATCC 35319) chromosome[35], recently renamed as *Ruminiclostridium cellulolyticum*, an important cellulose-degrading anaerobe used for solvent production[36]. However, the approach contained an important variation: the wild-type SpCas9 was replaced by the SpCas9^{D10A} mutant, which introduced SSDBs instead of DSDBs in the targeted regions. The editing efficiency of the system was affected by the length of the editing template's homologous arms, and the profusion of edited genomes was enhanced with serial, post-transformation recovery transfers[35]. This single-nick-induced HR-CRISPR-Cas9-based counter-selection approach allowed for efficient genome editing of a prokaryote sensitive to DSDBs, paving the way for the editing of other prokaryotes with inefficient NHEJ and HR mechanisms. It is striking though that the native *C. cellulolyticum* type II CRISPR-Cas system was not employed for the development of the editing tool, due to limited understanding of this system[35]. This facts underlines the requirement for solid *in silico* sgRNA designing tools and *in vitro* or *in vivo* high-throughput PAM determination tools (Box 1 & 2).

A dedicated system has recently been developed for efficient genome editing of *Streptomyces lividans*, *Streptomyces viridochromogenes* DSM 40736 and *Streptomyces albus* J1074 (Table S1)[37], all relevant antibiotic producers, by using an approach that resembles the aforementioned approach for *Clostridia*. The system was based on two types of vectors, both with a temperature-sensitive replicon, a codon-optimized *spcas9* gene and an HR template. The first vector type additionally carried a crRNA:tracrRNA-expressing module and the other an sgRNA-expressing module[37]. This one-plasmid based system was employed for efficient single and double gene deletions and for large-size chromosomal deletions of up to 30 kb. Interestingly, the sgRNA expressing vectors showed higher editing efficiencies compared to the corresponding crRNA:tracrRNA expressing vectors. This could be due either to reduced

pre-crRNA:tracrRNA processing by the host's RNase III[37], or to deletion of the spacer parts by spontaneous homologous recombination of the CRISPR repeats. Three independent studies describe the development of HR-CRISPR-Cas9-based counter-selection editing systems for the antibiotic-producing model organism *Streptomyces coelicolor*: two systems were developed for *S. coelicolor* M145[38,39] and one for *S. coelicolor* A3(2)[40] (Table S1). Zeng *et al*. and Tong et al. based the editing systems on the above described one-step HR-CRISPR-Cas9-based counter-selection genome editing approach. Both systems yielded efficient single[40] and double gene deletions, large-size chromosomal deletions and the simultaneous introduction of three point mutations in one gene[38]. Zeng *et al.* based their editing system on the introduction of the *codA*(sm) mutant cytosine deamidase gene into the vector of the one-step HR-CRISPR-Cas9-based counter-selection genome editing system. The CodA(sm) serves as a counter-selection marking by converting 5-fluorocytocine into toxic 5-fluorouracil. This CRISPR-Cas/CodA(sm)-based editing approach resulted in a double counter-selection system that selects for edited and plasmid-free clones in one step and it was successfully used for rapid and efficient in-frame gene deletion[39].

NON-HOMOLOGOUS END JOINING (NHEJ)

The only existing example of targeted DSDB introduced by Cas9 combined with repair via the NHEJ mechanism in a bacterium (Fig. 1) is presented in one of the mentioned S. coelicolor studies[40]. The incomplete NHEJ mechanism of the S. coelicolor A3(2) repaired the Cas9-targeted sites, introducing deletions of variable sizes. The efficiency of the CRISPR-Cas9-induced DSDBs– NHEJ editing tool was increased by the heterologous expression of a ligase D into the *S. coelicolor* A3(2) cells.

CRISPR-CAS-BASED EXPRESSION CONTROL IN BACTERIA

GENE REPRESSION

In some cases, it may be of interest to silence genes rather than knock them out. Whereas RNAi is available for this purpose in eukaryotic systems, these knock down tools are not yet available for prokaryotes. In 2013, however, two studies reported gene silencing in bacteria using catalytically inactive *S. pyogenes* $Cas9^{D10A, H840D}$, so-called dead Cas9 (dCas9), in which the two nuclease domains are mutated (D10A and H840D) and inactivated (Table S2, Fig. 2)[41,42]. Contrary to RNAi, dCas9 does not act on the mRNA but rather on the DNA level, by directly blocking transcription initiation or elongation (CRISPRi) depending on whether the promoter region or open reading frame is targeted (Box 3).

An advantage of dCas9-based silencing over Cas9-based editing is that no repair pathway is required. CRISPRi can be readily applied in cells lacking the NHEJ pathway orfor which no fast and efficient HR method is available, making it a promising method especially for hard to engineer prokaryotes. Additionally, CRISPRi can be an important tool for studying the phenotypic effect of essential genes silencing and for targeting specific genes or genotypes in mixed populations or undefined cultures. In all cases so far, repression by CRISPRi was shown to be highly specific, reversible after removal of the inducer compound and finely tuneable (Box 3).

The principle of using dCas9 was shown in *E. coli* by sgRNA-guided silencing of reporter genes such as *gfp*[41] or *gfp* and *rfp* simultaneously[42]. The possibility of using the system for hard-to-engineer organisms in a mixed culture was shown shortly after[40,43,44] (Table S2). Mimee *et al.* showed rapid and efficient dCas9-based gene silencing in the gut commensal *Bacteroides thetaiotaomicron* during *in vivo* colonization of the mouse gut. In *E. coli*, conjugation was used for horizontal transfer of a plasmid containing dCas9 and an sgRNA directed against an *rfp* reporter gene from a

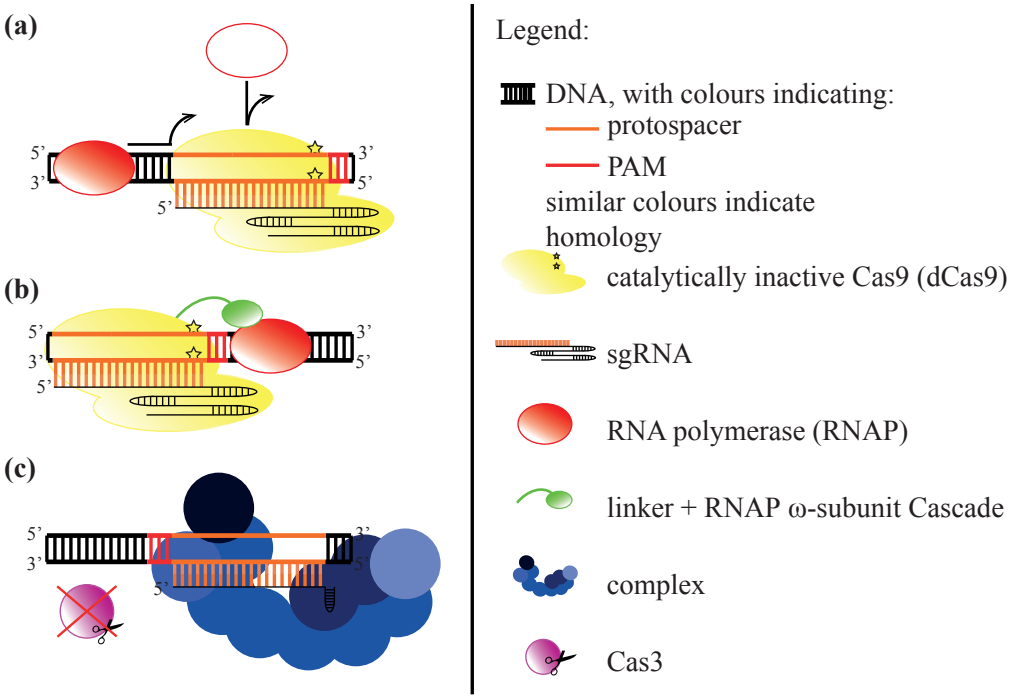


Figure 2. Overview of different systems for CRISPR-Cas-mediated gene expression modulation. a) Gene silencing using catalytically inactive Cas9 (dCas9, indicated with two stars in the nuclease domains), showing transcription block by directly binding the promoter, or blocking transcription elongation when the ORF is targeted. **b)** Gene activation via the RNA Polymerase ω-subunit fused to dCas9 in a strain that lacks the ω-subunit gene (ΔrpoZ). **C.** Gene silencing using Cascade in a *cas3*-knockout background.

donor strain into an *rfp*-expressing *E. coli* acceptor strain. The acceptor strain was part of a mixed *E. coli* culture and its RFP expression was successfully silenced[43]. A similar application of using CRISPR-Cas-system as highly specific antimicrobial in a complex environment was also shown for the wild-type Cas9 variant which was used as highly specific antimicrobial[45].

Another important application of CRISPRi is the targeting of essential genes[46,47] (Table S2). Previously, laborious methods were required in order to study or silence essential genes, for example by establishing conditional knockouts. Using CRISPRi, the amount of silencing can be appropriately fine-tuned (Box 3), and repression can be induced in a simple and fast way, also in hard-to-engineer organisms. This has allowed for the identification of essential genes in the pathogen *Mycobacterium* tuberculosis, which could be possible new drug targets[46]. A first successful example of using CRISPRi for metabolic engineering aimed at the production of PHA in *E. coli* and used dCas9-based downregulation of several essential TCA cycle genes[47].

GENE ACTIVATION

The properties of dCas9 as a non-cleaving DNA-binding protein can be further exploited by attaching other functional molecules to it. This was shown in bacteria in which dCas9 was fused to the RNA Polymerase ω-subunit (dCas9::ω) in an *E. coli* Δ*rpoZ* strain that lacks the ω-subunit gene[41] (Fig. 2b). The dCas9::ω chimera was programmed to target a promoter driving *lacZ* expression. A maximum of 23-fold induction in *lacZ* expression was observed, which might be further optimized by adjusting the linker or by using a different activator domain[41].

BOX 3. TUNING CRISPRi

Whereas in some cases maximal repression via CRISPRi is desired, in other cases (e.g. when genes are essential) it is necessary to fine-tune repression levels. Below, several determinants of the strength of CRISPRi are listed that can be used when tuning repression and designing targets and sp spacers.

TARGET REGION

Repression strength changes when targeting different regions of the promoter or open reading frame (ORF) relative to the transcription start site (TSS). The most efficient repression using dCas9 can be obtained by targeting the -35 box[41,42,47], which was found to be similar when using Type I-E Δ*cas3* systems[48]. The further away the target is from the TSS, then the less

BO X3. TUNING CRISPRi (continuation)

efficient the silencing – this holds true for targets in both the promoter and the ORF[41,42,46,47]. Moreover, it should be taken into account that many bacterial genes are in an operon and therefore silencing might have an undesired downstream effect[49].

STRAND BIAS

When targeting the promoter, no difference was observed between targeting the template or non-template strand. However, when dCas9 was used to target a region within the ORF, targeting the non-template (or coding) strand was consistently shown to be highly efficient, while targeting the template strand gave less or no repression[40-42,46]. It was hypothesized that this is due to the ability of the RNAP to 'unzip' dCas9 from the template strand via its helicase activity[42,46]. Remarkably, when using Type I-E Δ*cas3* systems this strand bias for ORF-targets was found to be inversed[48,50].

SPACER MISMATCHES

dCas9 binds steadily to the DNA target only if the 12 nt long protospacer seed sequence is fully complimentary to the corresponding part of the spacer sequence[41,42]. The more mismatches after the 12 nt seed sequence, the less efficient the binding and repression. Interestingly, when there are only 12 complementarynucleotides when using wild-type Cas9, the active protein can bind but not cleave, hence functioning as a dCas9 without the active site mutations[41]. This hold promise for future engineering purposes especially for those using endogenous systems, which can in this way also be used as repressor without making genomic mutations in *cas9*.

MULTIPLEXING

The use of multiple sgRNAs targeting different regions of the same gene (either promoter and/or ORF region) can further enhance repression[42,46].

INDUCTION

If an inducible promoter is used for dCas9, Cascade or guide RNA expression, in most cases the amount of repression was shown to increase with increasing inducer concentration or time[42,48]. On the other hand, in *Mycobacteria* using dCas9 it was found that longer induction time did not increase left-over protein levels, but increasing the length of hybridizing spacer part did, indicating differences per organism, and the need for optimization of the system for each new host[46].

GENOME EDITING AND REPRESSION USING ENDOGENOUS CRISPR-CAS SYSTEMS

The dCas9 system is easily transferable to different hosts as it requires only a single plasmid containing dCas9 and the targeting spacer(s). However, it might not be readily applicable in thermophilic hosts, as SpCas9 is derived from a mesophilic organism, or in poorly genetically accessible organisms due to the large size of *cas9*. In such cases, an alternative may be to use the host's endogenous CRISPR system for silencing. As the number of available plasmids and inducible promoters for non-model organisms is often limited, a major advantage of using an endogenous CRISPR-Cas system for editing or silencing is that only one, small plasmid containing a spacer (crRNA) and -if necessary- an HR-template, is required and no inducible promoters are needed. The only prerequisites are that the CRISPR-Cas system is functionally expressed, and that (in case of Type I, II and V systems) the PAM sequence is known (Box 2). The genome of the thermophilic archaeon *Sulfolobus islandicus* has recently been engineered using its endogenous Type I-A and Type III-B CRISPR-Cas systems, including gene deletion, insertion and site-directed mutagenesis[51]. More specifically, for gene deletions using the Type I system, the authors used a single plasmid which included a template for HR and a module expressing a crRNA against the target gene[51]. As the Type I system is most wide-spread in bacteria[5], promising proofs of principle were provided in two parallel studies in which after knockout of the nuclease cas3 in the *E. coli* Type I-E CRISPR-Cas system, this system functioned as a guide-directed endogenous silencing system (Fig. 2c)[48,50]. The Type I-E Cascade complex consisting of Cas proteins and crRNA is able to bind target DNA without cleaving it (due to the absence of Cas3), and as such Cascade can be used as a transcriptional road block. Furthermore it was demonstrated that the *E. coli* Type I-E Cascade complex can be successfully used for gene silencing when introduced into *Salmonella typhimurium*[48]. In *E. coli*, expression of the Cascade operon is normally repressed under laboratory conditions, requiring promoter replacement or induction of the Cascade-genes even for endogenous use in this species[48,50]. However, in many other organisms the CRISPR-Cas-system appears to be constitutively expressed and can readily be used, such as shown for *Sulfolobus*[51].

Next to genome editing and silencing, endogenous CRISPR-Cas-systems have also been shown to be effective for self-targeting or targeted killing of specific strains. As described earlier, the introduction of DSDBs in bacterial genomes is lethal in the absence of either a template for HR or a functional NHEJ system, which is the case in most bacteria. This feature can be exploited to specifically kill bacteria with certain genotypes in a mixed popula-

tion. Cas9-mediated killing of cells in the human gut was successfully shown[45], but also native systems have been evaluated for this purpose using the *E. coli* Type I-E[52] and the *Pectobacterium atrosepticum* Type I-F system[53] (Table S2). The endogenous *E. coli* CRISPR Type I-E system was used for highly selective targeting and removal of *E. coli* strains from a mixed culture[52]. In *P. atrosepticum*, targeting a genomic pathogenicity island using its endogenous CRISPR-Cas-system led to excision of the island[53].

CONCLUDING REMARKS AND FUTURE DIRECTIONS

Whereas CRISPR-Cas9-mediated engineering has moved extremely fast in eukaryotes, prokaryotic genome editing using this system is taking off more slowly. This is mainly due to the lack or the conditional expression of an efficient NHEJ repair pathway in most prokaryotes[54,55]. The Ku and LigD proteins constitute a minimal, error-prone bacterial NHEJ system[56] and their heterologous expression in the host of interest may lead to the construction of a CRISPR-Cas-based high-throughput tool for gene knock-outs in the host.

The absence of an effective repair system in prokaryotes might be highly useful for applications other than genome editing. Examples of this are the CRISPR-Cas-mediated vaccination of biotechnologically interesting strains against bacteriophages by natural spacer acquisition[1] or by CRISPR design[2] as well as the targeted killing of specific bacterial strains in complex environments[45,52]. At the moment, the delivery of a CRISPR-Cas-mediated targeted killing or delivery system in undefined mixed cultures is still a challenge, but conjugation systems[43] or virus-based systems[45,57] are possible strategies to allow for lateral transfer. In addition, targeting a genome either randomly or via IS elements[32,53] and selecting for survivors might be a useful tool when creating and studying minimal genomes. For other functional genomic studies, the possibility of adding functional molecules to dCas9 holds promise for applications such as the attachment of fluorescent groups to direct these to specific genomic loci and study subcellular localization linked to gene expression.

It is clear that high transformation and recombination efficiencies of the target organism are required for high-throughput genome editing. As already discussed, the nickase-Cas9 holds potential for enhancing editing efficiencies. Additional Class 2 effectors, such as Cpf1 and the recently discovered C2C1 and C2C3[6], are currently under evaluation for their editing potential. Cpf1 is an endonuclease that serves as the interference machine of the recently identified type V CRISPR-Cas system[13]. It was shown that Cpf1 introduces staggered DSDBs at the PAM distal end of a protospacer, which might enhance the efficiency of both HR and directional insertion of dsDNA fragments into eukaryotic and prokaryotic ge-

nomes[13]. Moreover, the fact that the seed sequence of a protospacer is not affected by the Cpf1 activity leaves also space for additional rounds of editing using the same protospacer. Additionally, Cpf1 does not require a tracrRNA, simplifying the design process of a Cpf1-based counter-selection tool.

The use of native CRISPR-Cas systems for editing purposes could overcome the challenges in efficient heterologous expression of a Cas9/Cpf1 system in non-model organisms (see Outstanding Questions). This underlines the requirement for rapid characterization methods of the host's CRISPR systems and especially identification of the PAM sequenc that these systems recognise. The use of CRISPR-Cas-systems other than Cas9 will be an important future research direction, as it will broaden the range of biotechnologically relevant micro-organisms in which it can be used.

AUTHOR CONTRIBUTIONS

I.M, E.F.B. designed and coordinated the manuscriptpreparation. I.M., E.F.B., W. d. V., R.v.K., and J.v.d.O. wrote the manuscript.

CO MPETING INTERESTS

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

ACK NO WLEDG EMENTS

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.

SUPPLEMENTAL INFORMATION

Table S1. Overview of CRISPR-Cas-based bacterial genome editing

Species	Cas type	HR method[a]	Nr of plasmids used	sgRNA or crRNA:tracrRNA used	Multiplex yes/no	Ref
Escherichia coli	SpyCas9	SSDR	2	crRNA:tracrRNA	no	[17]
	SpyCas9	SSDR and DSDR	2	sgRNA	yes	[25]
	SpyCas9	SSDR and DSDR	3	crRNA:tracrRNA	yes	[24]
	SpyCas9	λ Red recombineering-assisted Plasmid-HR	2	sgRNA	yes	[29]
	SpyCas9	DSDR	2	sgRNA	no	[31]
	SpyCas9	SSDR and DSDR	2	sgRNA	yes	[26]
	SpyCas9	DSDR	1	sgRNA	yes	[30]
	SpyCas9	None (aimed at cell killing)	1	crRNA:tracrRNA	yes	[45]
	SpyCas9 nickase	Chromosomal IS5 direct repeats and 2-4 target CRISPR system	2	sgRNA	yes	[32]
Tatumella citrea	SpyCas9	λ Red recombineering-assisted Plasmid-HR	2	sgRNA	yes	[29]
Streptococcus pneumoniae	SpyCas9	SSDR	2	crRNA:tracrRNA	no	[17]
Clostridium beijerinckii	SpyCas9	Plasmid-HR	1	sgRNA	no	[34]
Streptomyces coelicolor	SpyCas9	Plasmid-HR	1	sgRNA	no	[38]
Streptomyces coelicolor	SpyCas9	NHEJ and Plasmid-HR	1	sgRNA	no	[40]
Streptomyces coelicolor	SpyCas9	Plasmid-HR, combined with codA-counter-selection	1	sgRNA	no	[39]
Streptomyces lividans	SpyCas9	Plasmid-HR	1	sgRNA	yes	[37]
Streptomyces viridochromogenes	SpyCas9	Plasmid-HR	1	sgRNA	yes	[37]
Streptomyces albus	SpyCas9	Plasmid-HR	1	sgRNA	yes	[37]
Lactobacillus reuteri	SpyCas9	SSDR	3	crRNA:tracrRNA	no	[28]

Table S1. Overview of CRISPR-Cas-based bacterial genome editing (continuation)

Species	Cas type	HR method[a]	Nr of plasmids used	sgRNA or crRNA:tracrRNA used	Multiplex yes/no	Ref
Clostridium cellulolyticum	SpyCas9-nickase	Plasmid-HR	1	sgRNA	no	[35]
Sulfolobus islandicus	Native Type I-A	Plasmid-HR	1	NA	no	[51]
Sulfolobus islandicus Δcas3	Native Type III-B Cmr-α	Plasmid-HR	1	NA	no	[51]
Pectobacterium atrosepticum	Native Type I-F	Genomic island reshuffling (AEJ)	1	NA	yes[b]	[53]
Escherichia coli	Native Type I-E	None (aimed at cell killing)	3	NA	no	[52]

[a] Abbreviations: HR: homologous recombination; SSDR: single-stranded DNA recombineering (SSDR); DSDR: double-stranded DNA recombineering; plasmid-HR: plasmid-borne homologous recombination; NHEJ: non-homologous end-joining; AEJ: alternative end-joining

[b] Multiplexing was possible but not very effective, probably due to rearrangements of the repeats in the plasmid.

Table S2. Overview of CRISPR-Cas-based bacterial gene repression.

Species	CRISPR-Cas type	Max. fold expression reduction	Nr of plasmids used[a]	Required/modified genomic components[a]	sgRNA or crRNA:tracrRNA used	Multiplex yes/no	Ref
Escherichia coli	Sp-dCas9	100[b]	1	NA	tracrRNA:crRNA	no	[41]
	Sp-dCas9	300[b] (multiplex 1000)	2	NA	sgRNA	yes	[42]
	Sp-dCas9	330[b]	1	conjugation machinery	sgRNA	no	[43]
	Sp-dCas9	100 (multiplex on different genes)	1	NA	sgRNA	yes	[47]
Escherichia coli Δcas3	Native Type I-E	Plasmid target: 900[b] Genome target: 2200 (multiplex on different genes)	1	promoter replacement to activate Cascade	NA	yes	[50]
	Native Type I-E	Strain K12: 50[b] Strain BL21: 15[b]	2	NA	NA	yes	[48]
Streptococcus pneumoniae	Sp-dCas9	14[b]	1	dCas9 and tracrRNA integrated in genome	tracrRNA:crRNA	no	[41]
Mycobacteria	Sp-dCas9	Max. 94% Multiplex test: 4-fold each vs. 16-fold multiplex	1	dCas9 integrated in genome	sgRNA	yes	[46]
Bacteroides thetaiotaomicron	Sp-dCas9	45	1	NA	sgRNA	no	[44]
Streptomyces coelicolor	Sp-dCas9	100[b]	1	NA	sgRNA	no	[40]
Salmonella typhymurium	Type I-E (no Cas3)	12[c]	2	NA	NA	no	[48]

[a] These are only the elements required for the CRISPR-Cas-based system and excludes reporter plasmids etc.

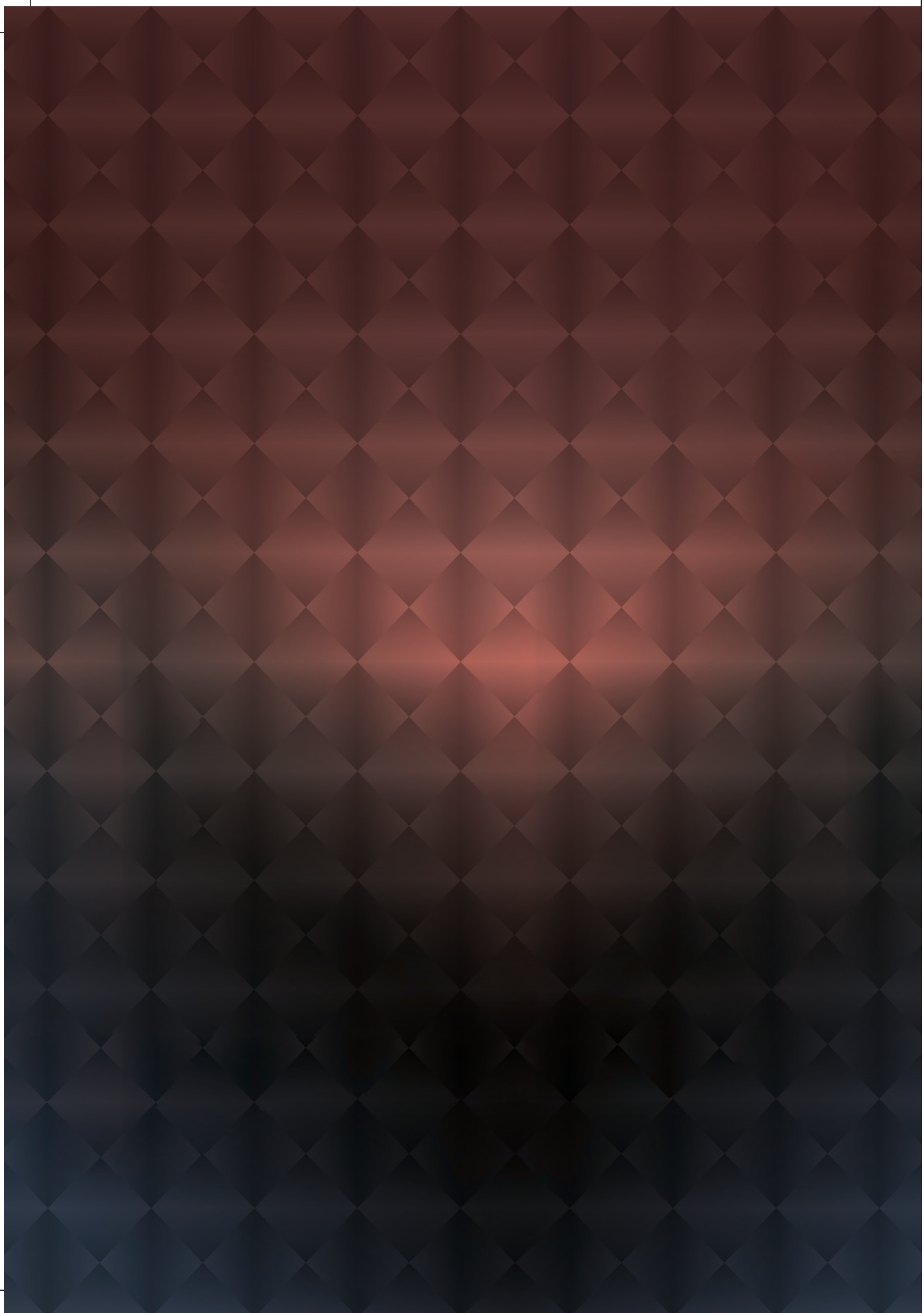
[b] This represents reduction of the measured fluorescent, coloured or luminescent product rather than gene expression.

[c] Longer induction times resulted in stronger repression, but these data were not shown.

REFERENCES

1. Barrangou, R., et al., CRISPR provides acquired resistance against viruses in prokaryotes. Science, 2007. 315(5819): p. 1709-1712.
2. Brouns, S.J.J., et al., Small CRISPR RNAs guide antiviral defense in prokaryotes. Science, 2008. 321(5891) p. 960-964.
3. van der Oost, J., et al., CRISPR-based adaptive and heritable immunity in prokaryotes. Trends in Biochemical Sciences, 2009. 34(8): p. 401-407.
4. Lundgren, M., et al., Annotation and classification of CRISPR-Cas systems, in CRISPR. 2015, Springer New York. p. 47-75.
5. Makarova, K.S., et al., An updated evolutionary classification of CRISPR-Cas systems. Nature Reviews Microbiology, 2015. advance online publication.
6. Shmakov, S., et al., Discovery and functional characterization of diverse Class 2 CRISPR-Cas systems. Molecular Cell, 2015. 60(3): p. 385-397.
7. Deveau, H., et al., Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. Journal of Bacteriology, 2008. 190(4): p. 1390-1400.
8. Mojica, F.J.M., et al., Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology, 2009. 155(3): p. 733-740.
9. Sinkunas, T., et al., Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. EMBO Journal, 2011. 30(7): p. 1335-1342.
10. Samai, P., et al., Co-transcriptional DNA and RNA cleavage during Type III CRISPR-Cas immunity. Cell, 2015. 161(5): p. 1164-74.
11. Tamu laitis, G., et al., Programmable RNA shredding by the type III-A CRISPR-Cas system of *Streptococcus thermophilus*. Molecular cell, 2014. 56(4): p. 506-17.
12. Jinek, M., et al., A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science, 2012. 337(6096): p. 816-821.
13. Zetsche, B., et al., Cpf1 Is a single RNA-guided endonuclease of a Class 2 CRISPR-Cas system. Cell, 2015. 163(3): p. 759-771.
14. Chylinski, K., et al., Classification and evolution of type II CRISPR-Cas systems. Nucleic Acids Research, 2014. 42(10): p. 6091-6105.
15. Bosma, E.F., et al., Isolation and Screening of Thermophilic Bacilli from Compost for Electrotransformation and Fermentation: Characterization of *Bacillus smithii* ET 138 as a New Biocatalyst. Applied and Environmental Microbiology, 2015. 81(5): p. 1874-1883.
16. Karvelis, T., et al., Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. Genome Biology, 2015. 16(1): p. 253.
17. Jiang, W., et al., RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature Biotechnology, 2013. 31(3): p. 233-239.
18. Jinek, M., et al., RNA-programmed genome editing in human cells. Elife, 2013. 2: p. e00471.
19. Kim, S., et al., Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Research, 2014. 24(6): p. 1012-1019.
20. Mali, P., K.M. Esvelt, and G.M. Church, Cas9 as a versatile tool for engineering biology. Nature Methods, 2013. 10(10): p. 957-963.
21. Cong, L., et al., Multiplex genome engineering using CRISPR/Cas systems. Science, 2013. 339(6121): p. 819-23.
22. Bowater, R. and A.J. Doherty, Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. PLoS genetics, 2006. 2(2).
23. Esvelt, K.M., et al., Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nature Methods, 2013. 10(11): p. 1116-1121.
24. Pyne, M.E., et al., Coupling the CRISPR/Cas9 system with lambda Red recombineering enables simplified chromosomal gene replacement in *Escherichia coli*. Applied and Environmental Microbiology, 2015. 81(15): p. 5103-5114.
25. Li, Y., et al., Metabolic engineering of *Escherichia coli* using CRISPR–Cas9 meditated genome editing. Metabolic Engineering, 2015. 31: p. 13-21.
26. Reisch, C.R. and K.L.J. Prather, The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in *Escherichia coli*. Scientific Reports, 2015. 5: p. 15096.
27. van Pijkeren, J.-P. and R.A. Britton, High efficiency recombineering in lactic acid bacteria. Nucleic Acids Research, 2012.
28. Oh, J.-H. and J.-P. van Pijkeren, CRISPR–Cas9-assisted recombineering in Lactobacillus reuteri. Nucleic acids research, 2014. 42: p. e131.
29. Jiang, Y., et al., Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system. Applied and Environmen talMicrobiology, 2015. 81(7): p. 2506-2514.
30. Xia, J., et al., Expression of Shewanella frigidimarina fatty acid metabolic genes in *E. coli* by CRISPR/ Cas9-coupled lambda Red recombineering. Biotechnology Letters, 2015: p. 1-6.

31. Pines, G., et al., Codon compression algorithms for saturation mutagenesis. ACS synthetic biology, 2015. 4(5): p. 604-14.
32. Standage-Beier, K., Q. Zhang, and X. Wang, Targeted large-scale deletion of bacterial genomes using CRISPR-nickases. ACS Synthetic Biology, 2015. 4(11): p. 1217-1225.
33. Gasiunas, G., et al., Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proceedings of the National Academy of Sciences, 2012. 109(39): p. E2579–E2586.
34. Wang, Y., et al., Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system. Journal of Biotechnology, 2015. 200: p. 1-5.
35. Xu, T., et al., Efficient genome editing in *Clostridium cellulolyticum* via CRISPR-Cas9 nickase. Applied and Environmen talMicrobiology, 2015. 81(13): p. 4423-4431.
36. Ravachol, J., et al., Combining free and aggregated cellulolytic systems in the cellulosome-producing bacterium Ruminiclostridium cellulolyticum. Biotechnology for Biofuels, 2015. 8(1): p. 1-14.
37. Cobb, R.E., Y. Wang, and H. Zhao, High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. ACS synthetic biology, 2014. 4(6): p. 723-728.
38. Huang, H., et al., One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. Acta Biochimica et Biophysica Sinica, 2015. 47(4): p. 231-243.
39. Zeng, H., et al., Highly efficient editing of the actinorhodin polyketide chain length factor gene in *Streptomyces coelicolor* M145 using CRISPR/Cas9-CodA(sm) combined system. Applied Microbiology and Biotechnology, 2015. 99(24): p. 10575-10585.
40. Tong, Y., et al., CRISPR-Cas9 based engineering of actinomycetal genomes. ACS synthetic biology, 2015. 4(9): p. 1020-1029.
41. Bikard, D., et al., Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Research, 2013. 41(15): p. 7429-7437.
42. Qi, Lei S., et al., Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell, 2013. 152(5): p. 1173-1183.
43. Ji, W., et al., Specific gene repression by CRISPRi system transferred through bacterial conjugation. ACS synthetic biology, 2014. 3(12): p. 929-931.
44. Mimee, M., et al., Programming a human commensal bacterium, *Bacteroides thetaiotaomicron*, to sense and respond to stimuli in the murine gut microbiota. Cell Systems, 2015. 1(1): p. 62-71.
45. Citorik, R.J., M. Mimee, and T.K. Lu, Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. Nature Biotechnology, 2014. 32(11): p. 1141-1145.
46. Choudhary, E., et al., Gene silencing by CRISPR interference in mycobacteria. Nature Communications, 2015. 6.
47. Lv, L., et al., Application of CRISPRi for prokaryotic metabolic engineering involving multiple genes, a case study: Controllable P(3HB-co-4HB) biosynthesis. Metabolic Engineering, 2015. 29(0): p. 160-168.
48. Rath, D., et al., Efficient programmable gene silencing by Cascade. Nucleic Acids Research, 2015. 43(1): p. 237-246.
49. Qi, L.S. and A.P. Arkin, A versatile framework for microbial engineering using synthetic non-coding RNAs. Nature Reviews Microbiology, 2014. 12(5): p. 341-354.
50. Luo, M.L., et al., Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. Nucleic Acids Research, 2015. 43(1): p. 674-681.
51. Litsanov, B., et al., Efficient aerobic succinate production from glucose in minimal medium with *Corynebacterium glutamicum*. Microbial biotechnology, 2012. 5(1): p. 116-128.
52. Gomaa, A.A., et al., Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. mBio, 2 014.5(1).
53. Vercoe, R.B., et al., Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacteria l genomes and expel or remodel pathogenicity islands. PLoS Genetics, 2013. 9(4): p. e1003454.
54. Aravind, L. and E.V. Koonin, 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, 2001. 11(8): p. 1365-1374.
55. Wang, S.T., et al., The forespore line of gene expression in Bacillus subtilis. Journal of Molecular Biology, 2006. 358(1): p. 16-37.
56. Shuman, S. and M.S. Glickman, Bacterial DNA repair by non-homologous end joining. Nature Reviews Microbiology, 2007. 5(11): p. 852-861.
57. Bikard, D., et al., Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. Nature Biotechnology, 2014. 32(11): p. 1146-1150.



CHAPTER 3

EFFICIENT GENOME EDITING OF A FACULTATIVE THERMOPHILE USING THE MESOPHILIC SpCAS9

Ioannis Mougiakos[1,#], Elleke F. Bosma[1,#], Koen Weenink[1], Eric Vossen[1], Kirsten Goijvaerts[1], John van der Oost[1], Richard van Kranenburg[1,2,*]

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

[2] Corbion, Arkelsedijk 46, 4206 AC Gorinchem , The Netherlands.

#Contributed equally

*Corresponding author

Chapter adapted from publication:

ACS Synth Biol. 2017 May 19;6(5):849-861. doi: 10.1021/acssynbio.6b00339

ABSTRACT

Well-developed genetic tools for thermophilic microorganisms are scarce, despite their industrial and scientific relevance. Whereas highly efficient CRISPR/Cas9-based genome editing is on the rise in prokaryotes, it has never been employed in a thermophile. Here, we apply *Streptococcus pyogenes* Cas9 (spCas9)-based genome editingto a moderate thermophile, i.e. *Bacillus smithii*, including a gene deletion, gene knock-out via insertion of premature stop codons, and gene insertion. We show that spCas9 is inactive *in vivo* above 42°C and we employ the wide temperature growth range of *B. smithii* as an induction system for spCas9 expression. Homologous recombination with plasmid-born editing templates is performed at 45-55°C, when spCas9 is inactive. Subsequent transfer to 37°C allows for counter-selection through production of active spCas9 that introduces lethal double-stranded DNA breaks to the non-edited cells. The developed method takes 4 days with 90, 100 and 20% efficiency for gene deletion, knockout and insertion, respectively. The major advantage of our system is the limited requirement for genetic parts: only one plasmid, one selectable marker and a promoter, which does not need to be inducible or well-characterized. Hence, it can be easily applied for genome editing purposes in both mesophilic and thermophilic non-model organisms with limited genetic toolbox and ability to grow at temperatures between 37 and above 42°C.

KEYWORDS

CRISPR/Cas9, bacteria, *Bacillus smithii*, thermophiles, genome editing, homologous recombination

INTRODUCTION

Microbial fermentation of renewable resources into green fuels and chemicals is playing a major part in the development of the bio-based economy. The production costs of these environmentally friendly processes have to be reduced before they become competitive with the traditional fossil fuel-based industries. To this end, microbes other than the widely-used model organisms, such as *Escherichia coli* and *Saccharomyces cerevisiae*, are being evaluated for their perspective abilities to act as production hosts. Thermophilic organisms are of particular interest due to their multiple advantages over mesophilic organisms when being used as production hosts[1-3]. For example, their ability to grow and ferment at thermophilic temperatures reduces the contamination risk with mesophiles[4-6] and there are examples of using thermophiles for non-sterilized fermentations, which would reduce sterilization costs[7,8], it reduces the cooling costs[9,10], increases substrate and product solubility[11], and the fermentation process runs at the optimum temperature for enzymatic lignocellulose degradation, allowing for efficient simultaneous saccharification and fermentation[12,13]. However, the use of non-model thermophiles as production hosts is generally hampered by the lack of well-developed genome editing tools compared to the currently used mesophilic model organisms[1,14].

Previous work has established basic genome editing tools for the thermophilic strain *Bacillus smithii* ET 138 (referred to as ET 138 herein), allowing for the introduction of scar-free markerless gene deletions using homologous recombination[15]. *B. smithii* grows between 37°C and 65°C and efficiently utilizes both C_5 and C_6 sugars[16,17]. Its main product is L-lactate and in order to use ET 138 as a versatile platform host for the production of other chemicals and fuels, its ldhL gene was deleted. No counter-selection tool was used in this process, resulting in a very laborious screening process to obtain clean mutants[15]. Subsequently, a lacZ-based counter-selection was developed relying on the toxicity of high concentrations of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal)[15,18]. Using this system, the sporulation gene sigF and the pyruvate dehydrogenase complex subunit gene pdhA were consecutively deleted in the ldhL mutant. The resulting triple mutant strain was sporulation-deficient, which is desired in industrial fermentations for safety reasons, and did not produce L-lactate and acetate. Since the pdhA mutant was also acetate -auxotroph, the double mutant ET 138 ΔldhL ΔsigF was selected as a basic platform strain for future studies[15]. Although the lacZ-based counter-selection system significantly decreased the time needed for mutant selection, the developed process is still time-consuming, with the fastest possible route

to gene deletion taking approximately 2 to 3 weeks from transformation to generation ofa scar-free markerless knockout[15]. Moreover, the counter-selection step is not stringent enough for removal of genes that are essential for the fitness and the metabolism of the strain. For the successfully engineered genes, only 12.5-33% of the colonies had the mutant genotype, whereas for other genes only wild-type revertants and false positives were obtained, generating mostly (66-88%) or in some cases only wild-type revertants or false positives[15]. The large number of wild-type revertants due to the absence or inefficiency of counter-selection is a general issue in several non-model organisms[19,20]. This creates the need for laborious PCR-based screening and strongly decreases throughput of the engineering process, limiting the study of these organisms and their development into industrial production hosts.

One of the fastest and most efficient methods currently available for genome editing is a system based on the *Streptococcus pyogenes* Cas9 (spCas9) RNA-guided DNA endonuclease of the type II CRISPR-Cas defence system[21]. Jinek et al. showed that a short single guide RNA (sgRNA) molecule can direct the spCas9 endonuclease to a desired complementary target , called protospacer[22]. In the presence of the short 5'-NGG-3' DNA motif right downstream of the 3'-end of the protospacer, called protospacer adjacent motif (PAM), spCas9 introduces a lethal chromosomal double stranded DNA break (DSDB)[22]. The system was applied for genome editing of human[23-26] and mice cells[25], paving the road for genome editing of a wide range of eukaryotic cells[27-29]. In these eukaryotes, the Non-Homologous End Joining (NHEJ) system repairs spCas9-induced DSDBs in an error-prone manner, creating insertion/deletion (indel) mutations[23-26]. These mutations usually render the gene inactive through frameshifting and simultaneously prevent further spCas9-recognition and subsequent cleavage due to the alteration of the target site. Whilst for most prokaryotes the NHEJ system is generally not present or not active[30], the ET 138 genome contains the genes for the basic prokaryotic NHEJ-like system, consisting of DNA ligase LigD and DNA-end-binding protein Ku[31,32]. However, the functionality of the NHEJ-like system in ET 138 is unknown.

The combination of the spCas9 activity with editing templates, such as recombineering oligonucleotides or plasmid-borne sequences for homologous recombination, has been recently exploited for prokaryotic genome editing[33-35]. spCas9 was employed to introduce DSDBs in prokaryotic genomes. These breaks modestly induced the recombination of a provided rescuing/editing template into the targeted chromosome, resulting in genetically modified cells[33,36,37].The edited cells avoided subsequent spCas9 targeting events, but in many studies the number of surviving/edited colonies was low with high percentage of mixed

(both wild type and mutant) or escape mutant genotypes[38-40]. The number of surviving colonies as well as the percentage of successfully edited cells was higher in studies that allowed homologous recombination of the editing templates to take place prior to the spCas9 targeting. This way spCas9 was employed for stringent counter-selection of the unedited genomes. For this approach either homologous recombination was faster than spCas9-targeting or cas9sp expression was induced after homologous recombination[41-43]. Moreover, the vast majority of the studies required either a multiple-plasmid system or very tightly controlled promoters[42,44]. Currently, only one plasmid, one selection marker and no inducible promoters are available for ET 138, limiting the options for such systems. Many of the well-known and widely applied genome editing tools, including CRISPR-Cas9 editing, are not amenable to thermophiles. The enzymatic machineries of these tools have not proven stable at temperatures higher than 37°C. Whereas the native CRISPR-Cas type I system of a thermophilic archaeon has been employed for genome editing[45] as well as chromosome-based genetic manipulations have been reported for a few naturally competent thermophiles[46], no reports are available on using Cas9-based editing in thermophilic organisms.

In this study, we show that spCas9 is inactive in ET 138 at temperatures above 42°C and we tightly control its activity by altering the cultivation temperature rather than using an inducible promoter. We create a clean gene deletion, a gene disruption and a gene insertion by employing a plasmid-borne homologous recombination template for introduction of the desired modifications to the genome at 45°C, while the non-edited cells are subsequently eliminated by spCas9 counter-selection at 37°C. To the best of our knowledge, this is the first time that a temperature-controlled recombination/counter-selection tool is employed for genome editing purposes, as well as the first time that a Cas9-based editing tool is used for engineering the genome and the metabolism of a moderate thermophile.

RESULTS AND DISCUSSION

IN VIVO EXPRESSION VALIDATION OF SpCAS9 AT DIFFERENT TEMPERATURES

In a parallel study, we demonstrated that the *in vitro* sgRNA-loading to spCas9 requires temperatures below 42°C and is the limiting step towards the *in vitro* formation of the active gRNA-spCas9-complex (to be published). This result motivated us to evaluate the *in vivo* activity of spCas9 in ET 138 at different temperatures. We designed and constructed the modular pWUR_Cas9nt construct, that encompasses the *cas9* gene of *S. pyogenes* (referred to as *cas9sp* herein) and an sgRNA-expressing module for which the

is predicted not to target any site of the *B. smithii* genome, (i.e. nt, for non-targeting). The backbone of the pWUR_Cas9nt is the pNW33n vector, which was the only available vector for *B. smithii* (Figure 1A)[15,16]. The first basic requirement for the design of the pWUR_Cas9nt was the development of promoters that will drive the expression of the two components of the system: the cas9sp effector and its sgRNA guide module. For many non-model organisms, the number of available promoters and plasmids is very limited. ET 138 is no exception: only two promoters have been evaluated for expression in *B. smithii*: the heterologous P_{pta} from *B. coagulans*, and the native P_{ldhL}[15]. The latter is undesired, as integration of pNW33n-based HR-plasmids into the ET 138 genome is possible[15] and we want to prevent crossover events between the pWUR_Cas9nt construct and the *B. smithii* genome over the promoter region. Additionally, an inducible system would be desirable. For these purposes we tested the *xynA* promoter (P_{xynA}) from *Thermoanaerobacterium saccharolyticum*[47]. On the genome of its native host this promoter is induced by xylose and repressed by glucose. To test expression in ET 138, we constructed the pWUR_lacZ vector; we used pNW33n as cloning and expression vector and we introduced P_{xynA} in front of the *B. coagulans*-derived *lacZ* gene, previously shown to be functional in ET 138[15]. Low expression was observed after overnight culturing in LB2 liquid medium without added carbon source, and strong expression was observed with both xylose and glucose (Figure S1). It is probable that the catabolite repression mechanism of ET 138 does not recognize the cre (catabolite repression element) sequence present on P_{xynA}, resulting in lack of transcriptional repression by glucose in *B. smithii* ET 138. Although P_{xynA} is not tightly controllable in *B. smithii* we still selected it to drive the expression of cas9 from the pWUR_Cas9nt, maintaining the possibility to explore the effects of different cas9sp expression levels. During all further experiments in this study, xylose and glucose were both added to the media at all the culturing temperatures, to ensure constant induction of SpCas9 expression at all times from P_{xynA}. Additionally, we placed the sgRNA-module under the control of the *B. coagulans* phosphotransacetylase (*pta*) promoter P_{pta}[5,49] without its RBS. It is generally stressed that the transcriptional start site (TSS) of the promoter driving the sgRNA expression must be characterized. However, the TSS of the P_{pta} we use for the same purpose in our study is not characterized and since we remove only the predicted RBS we expect that a few nucleotides are added to the 5'-end of the sgRNA molecu-

les. Even so, we hypothesized hat this extended sgRNA module will be efficiently loaded to the spCas9 and it will not influence the Cas9 targeting efficiency and specificity, since it is located at the PAM distal end. By practically verifying our hypothesis we would facilitate the application of our system to other non-model organisms with promoters that do not have their transcription start determined.

We evaluated spCas9-expression and toxicity levels in ET 138 by transforming a single batch of competent cells, with the non-targeting pWUR_Cas9nt vector and with the empty pNW33n control vector. The transformed cells were plated on LB2 plates supplemented with chloramphenicol. We incubated the plates overnight only at 55°C,

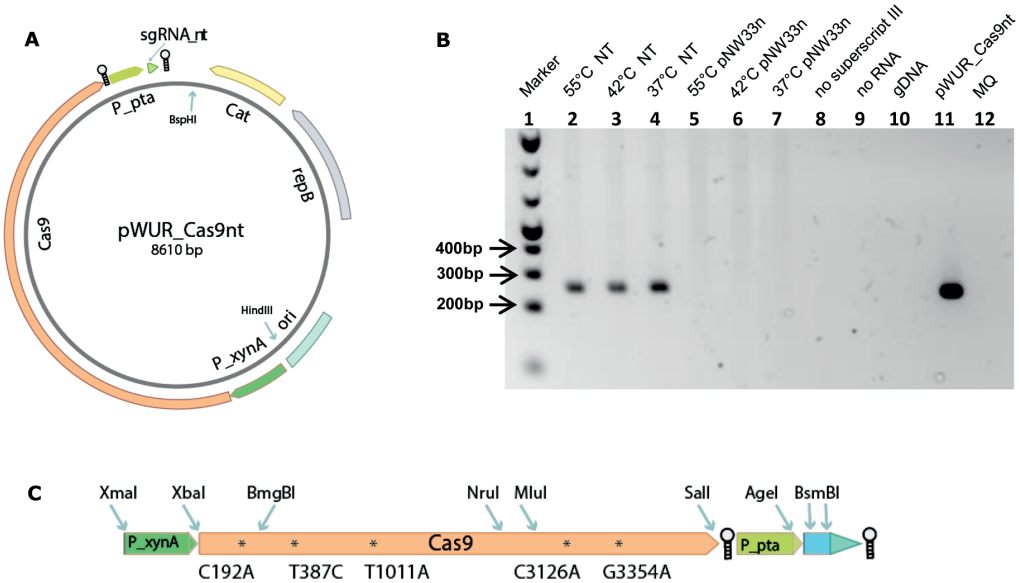


Figure 1. Schematic overview of the basic pWUR_Cas9nt construct. **A.** The non-codon optimized $cas9_{sp}$ gene was employed for the construction of the pWUR_Cas9nt vector, since the *S. pyogenes* and *B. smithii* GC-content and codon usage are highly similar. In the pNW33n-based basic construct the spCas9 was placed under the control of P_{xynA}. A Rho-independent terminator from *B. subtilis*[59] was introduced after the stop codon of the gene. The spCas9 module is followed by an sgRNA-expressing module that encompasses a spacer which does not target the genome of ET 138. The sgRNA module was placed under the transcriptional control of P_{pta} from *B. coagulans* -without its RBS-[15,49] and it is followed by a second Rho-independent terminator from *B. subtilis*. The spCas9 and sgRNA modules were synthesized as one fragment, which was subsequently cloned into the pNW33n through the BspHI and HindIII restriction sites. **B.** To prevent double restriction sites and create a modular system, 5 point silent mutations (C192A, T387C, T1011A, C3126A, G354A) were introduced to the gene (depicted as *). The depicted restriction sites are unique in the construct and introduced to facilitate the exchange of genetic parts. The spacer was easily exchanged to targeting spacers via BsmBI restriction digestion or Gibson assembly. The basic construct did not contain any HR-templates, but in the cases that these were added they were always inserted right upstream the spCas9-module and downstream the origin of replication. **C.** Total RNA was isolated from ET 138 wild-type cells transformed with pWUR_Cas9nt or pNW33n and grown at 55, 45 and 37°C. 6 cDNA libraries were produced with rt-PCR and used as templates for PCR with $cas9_{sp}$-specific primers that amplify a 255 bp long region. The PCR results are depicted in this agarose gel electrophoresis: Lane 1 corresponds to the marker (1kb+ DNA ladder, ThermoFisher), Lanes 2 to 4 correspond to ET 138 wild type cultures transformed with the pWUR_Cas9nt and grown at 55, 42, 37°C respectively, Lanes 5 to 7 correspond to ET 138 wild type cultures transformed with the pNW33n and grown at 55, 42 and 37°C respectively, Lanes 7, 8, 9, 11, 12 correspond to different negative controls and Lane 10 corresponds to the positive control, for which pWUR_Cas9nt was used as the PCR template.

as previous incubation attempts at 37 and 42°C were not successful (data not shown). One single colony per transformation was used for sequential transfers in LB2xg broth, transferring the cultures from 55°C to 37°C, with two intermediate steps at 45°C and 42°C. We isolated total RNA from each pWUR_Cas9nt culture after 18 hours of incubation at 55°C, 42°C and 37°C and performed semi-quantitative reverse transcription (rt)-PCR using specific primers for the *cas9sp* gene. Transcription of cas9sp was observed for all tested temperatures (Figure 1C). The growth of the cas9sp-expressing cultures was similar to the pNW33n control cultures at all temperatures (Figure 2B), indicating that the expression of *cas9sp* is not toxic for the ET 138 cells at any of the tested temperatures.

IN VIVO VALIDATION OF SPCAS9 ACTIVITY AT DIFFERENT TEMPERATURES

As *B. smithii* genome encompasses the genes for the basic prokaryotic NHEJ-like system[31,32] our first approach to construct a spCas9-based genome editing tool focused on determining the *in vivo* temperature limits of spCas9 and the capacity of the ET 138 NHEJ-like mechanism to repair the spCas9-induced DSDBs. We chose to target *pyrF* as a first proof of principle. The *pyrF* gene encodes the orotidine 5′-phosphate decarboxylase and is part of the operon for pyrimidine biosynthesis. Removal of the gene causes uracil-auxotrophy and resistance to the toxic uracil-analogue 5-fluorootic acid (5-FOA). It is a frequently used selection and counter-selection system in many organisms including thermophiles[50-52], and to this end we had initially made a clean *pyrF* deletion mutant ET 138 Δ*ldhL* Δ*sigF* Δ*pyrF* by adding the fused 1-kb up- and downstream *pyrF* flanks to pNW33n and applying 5-FOA pressure to select for double cross-over mutants. A total of 9 rounds of subculturing on selective media containing uracil and 5-FOA was required before pure knockouts were obtained with a knockout efficiency of around 50% (data not shown), making the process rather time-consuming. Whereas integration of pNW33-based HR-plasmids in ET 138 strains occurs[15], likely due to the rolling circle-based replication of pNW33n[53,54], subsequent excision of the plasmid and identification of clean mutant strains is difficult[15]. The speed and ease of this process could be significantly increased by using a strong counter-selection system, such as spCas9.

We continued our study by constructing 3 vectors based on the pWUR_Cas9nt sequence, designated pWUR_Cas9sp1, pWUR_Cas9sp2, and pWUR_Cas9sp3, each containing a different pyrF-targeting spacer (sp1-3). We transformed a single batch of wild type ET 138 competent cells with the 3 targeting vectors, the non-targeting control pWUR_Cas9nt and the empty vector control pNW33n. After initial outgrowth at 55°C on

LB2 without sugar, one confirmed transformant for each construct was subjected to a sequential plating and culturing process in which the temperature was stepwise lowered from 55°C to 37°C to induce spCas9-expression (Figure 2A). Normal growth was observed for all cultures at 55°C, 45°C and 42°C, as well as for the control cultures at 37°C. No growth was observed for any of the cultures with *pyrF*-targeting sgRNA-modules at 37°C (Figure 2B). We repeated the transformation and culturing process using the double mutant *B. smithii* ET138 Δ*ldhL* Δ*sigF* strain[15]. This strain is sporulation-deficient, it cannot produce L-lactate and it was constructed as the basic platform strain for further metabolic engineering work We obtained similar results as for the wild type *B. smithii* ET138 strain (Figure S3) and we continued our study using the double mutant strain.

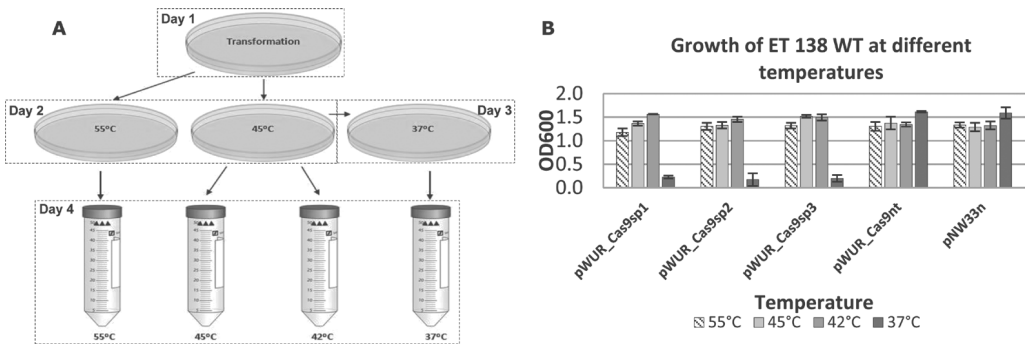


Figure 2. A. Sequential transfer scheme of wild-type ET 138 cultures to evaluate spCas9 expression and targeting efficiency at different temperatures. ET 138 cells were transformed with the pWUR_Cas9sp1, pWUR_Cas9sp2, pWUR_Cas9sp3, pWUR_Cas9nt and pNW33n vectors and plated on LB2 agar plates with chloramphenicol (Day 1). After overnight (ON) incubation at 55°C, single colonies were restreaked on LB2 agar plates with chloramphenicol and incubated ON at 55°C and 45°C (Day 2). Colonies from the 45°C plates were transferred to LB2 agar plates with chloramphenicol for ON incubation at 37°C (Day 3). The plates from days 2 and 3 were then used for inoculation of liquid medium (Day 4); 1 colony per 55°C, 45°C and 37° plate was transferred to LB2 medium with xylose, glucose and chloramphenicol for ON incubation at 55°C, 45°C and 37°C respectively, while 1 extra colony per 45°C plate was transferred to the same medium for overnight incubation at 42°C. **B.** Results of the targeting experiment showing OD_{600} measurements from cultures of wild-type ET 138 transformed with the 3 different pyrF targeting $cas9_{sp}$ constructs, the non-targeting $cas9_{sp}$ construct and pNW33n. The growth of the cells with the pyrF targeting constructs is greatly affected at 37°C, which is not observed for cells containing the non-targeting constructs.

In line with the *in vitro* assay data, the aforementioned results indicate that the designed spCas9 system is active at 37°C, but is inactive and does not introduce lethal SDBs to the ET 138 genome at temperatures above 42°C. It also indicates that the NHEJ system in ET 138 is inactive at 37°C or not active enough to counteract the spCas9 activity. In addition, the sequencing results of the pWUR_Cas9sp2 construct revealed the deletion of 7 nucleotides near the 3' end of the Ppta (Figure S2). Nonetheless, the results of the targeting ex-

periment support our hypothesis that the uncharacterized nature of the P_{pta}, and possibly the extra nucleotides at the 5' end of the sgRNA module, do not hinder the targeting efficiency of spCas9. Moreover, the truncated P_{pta} of the pWUR_Cas9sp2 construct clearly expresses enough sgRNA molecules and the observed 7-nucleotides long deletion has no impact to the targeting efficiency of the corresponding spCas9/sgRNA complex. The latter two observations reveal that our system tolerates a certain level of flexibility in the 5`-extension of the sgRNA molecule.

EFFICIENT GENE DELETION USING A HR-CRISPR-CAS9 COUNTER-SELECTION SYSTEM

Taking into account the results described so far, we decided to develop a Cas9-based editing system for ET 138 exploiting its efficient homologous recombination mechanism[15] and the temperature-induced spCas9 activity at 37°C. The new experimental setup consists of a single plasmid that combines the recombination template and the spCas9- and sgRNA-expressing modules. Providing the cells with the appropriate plasmid-borne editing template at 55°C and then inducing the expression of active spCas9 at 37°C through a sequential culturing process, is expected to form a powerful homologous recombination and counter-selection system.

To generate a pyrF deletion mutant, pWUR_Cas9sp1 was selected as the pyrF-targeting vector for further experiments, which was always compared to the non-targeting control pWUR_Cas9nt. To both vectors, we added a fusion of the two *pyrF*-flanks (each 1kb), creating the pWUR_Cas9nt_hr and pWUR_Cas9sp1_hr vectors (i.e. hr, for homologous recombination) (Figure 3A). After transforming the 4 constructs (2 with flanks and 2 withou flanks) into ET 138 Δ*ldhL* Δ*sigF* at 55°C, one verified transformant per construct was inoculated into TVMY selection medium containing xylose, glucose and supplemented with uracil ($TVMY_{xgu}$) to complement the auxotrophy in case of successful *pyrF* deletion. After growth at 55°C for 24 hours, cells were sequentially transferred every 24 hours to fresh media while gradually lowering the culturing temperature from 55°C to 37°C, with an intermediate transfer at 45°C. After 3 more transfers at 37°C to allow for efficient spCas9-counter-selection, cells were transferred back to 55°C with an intermediate transfer at 45°C, allowing for gradual adjustment of their metabolism. As expected, PCR on genomic DNA from the pWUR_Cas9nt and pWURCas9_sp1 cultures showed no *pyrF* knockout bands at any culturing temperature due to the lack of homologous recombination template in the constructs. In line with the initial *in vivo* tests, pWURCas9_sp1 cultures at 37°C showed almost no growth, while the pWUR_Cas9nt cultures at all the temperatures showed the expected growth

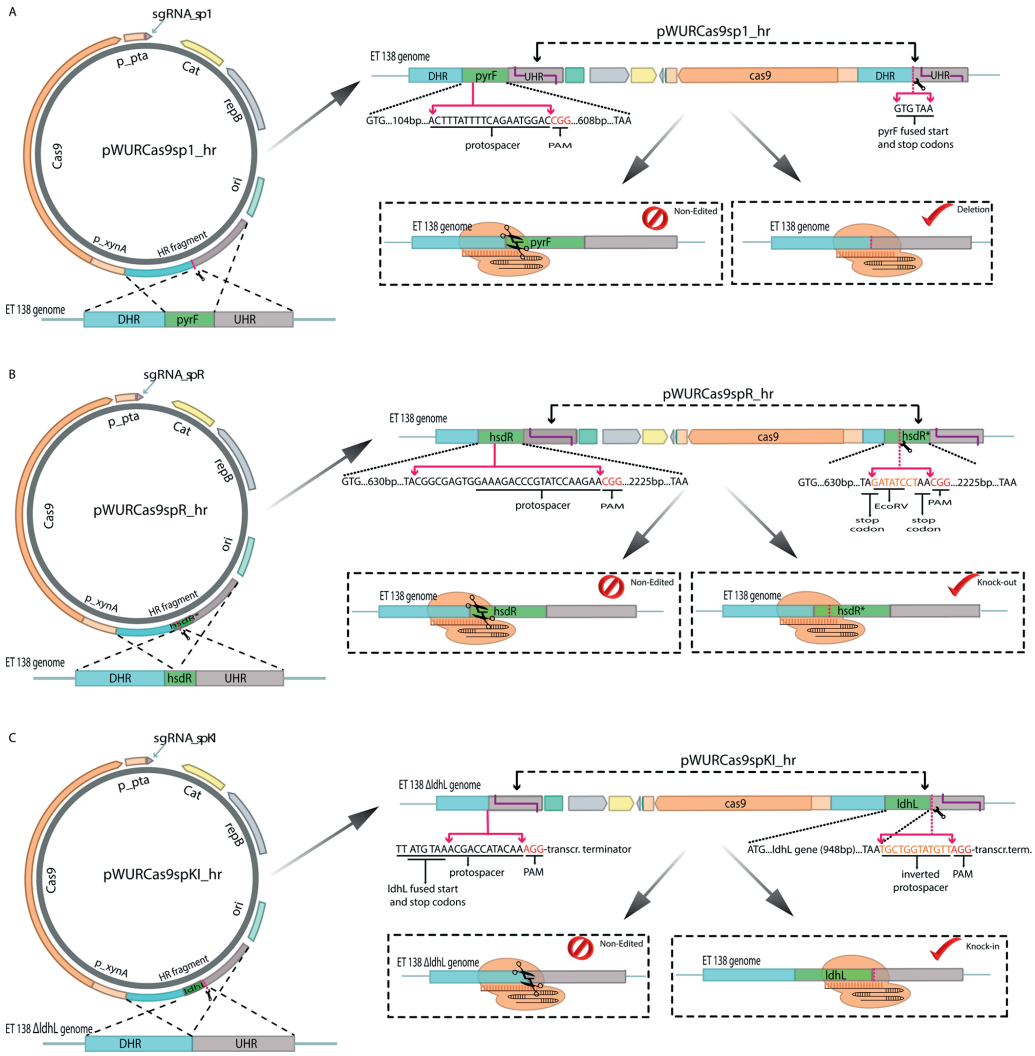


Figure 3. Schematic representation of the different hom logous recombination andspCas9-mediated mutation s described in this study. The first single cross over event (SCO) can occur by insertion of the editing plasmid into the chromosome either through the upstream homologous region (UHR) -as depicted here- or through the downstream homologous region (DHR). A. Gene deletion. A scarless, markerless *pyrF* gene deletion was established after insertion of the editing vector into the chromosome via homologous recombination with the plasmid-borne editing template (2 × 1 kb flanks, immediately flanking the *pyrF* gene and thus removing it from start to stop codon), after which a second SCO event results in either wild-type revertants or edited cells. The spCas9 targeting of the wild-type cells acts as counter-selection for the *pyrF* mutants. (B) Gene knockout via insertion of stop codons and a restriction site. The followed process was similar to the gene deletion described above. The *hsdR* restriction gene was inactivated by inserting stop codons and a restriction site between codons 212 and 221 that were contained in a 2 kb HR fragment that expands 289 bp upstream and 1.65 kb downstream from the start codon of the *hsdR* gene on the genome of the ET 138. Between the two stop codons, an EcoRV restriction site was added, generating a frame shift and facilitating the screening process. The spacer was designed to target the original sequence without stop codons and restriction site. (C) Gene knock-in. The followed process was similar to the gene deletion and gene knockout processes described above. The ldhL gene was reinserted into mutant strain ET 138 Δ*ldhL* Δ*sigF*. This was achieved by adding the original *ldhL* gene sequence between 2 × 1 kb HR flanks. The spacer was designed to target the area between the *ldhL* stop codon and the beginning of the adjacent rho-independent transcriptional terminator. On the HR flanks, the region between the *ldhL* stop codon and its rho-independent transcriptional terminator was inverted, avoiding the spCas9 targeting of edited cells.

for ET 138 culture. PCR on genomic DNA from the liquid cultures containing the vectors with *pyrF* flanks showed the absence of k, but very strong *pyrF* knockout bands for the pWUR_Cas9sp1_hr cultures for the same culturing steps mentioned above (Figure 4A). The striking difference in the density of the knockout bands between the targeting pWUR_Cas9sp1_hr and the non-targeting pWUR_Cas9nt_hr cultures indicates successful spCas9 activity and *pyrF*-targeting by the pWUR_Cas9sp1_hr construct. It furthermore indicates that the counter-selection activity of spCas9 is already efficient from the first culturing step at 37°C. Growth of the pWUR_Cas9nt_hr cultures was similar to the control at all temperatures, while the pWUR_Cas9sp1_hr cultures showed poor growth in the first 2 culturing steps at 37°C, but the growth was reconstituted from the 3rd culturing step at 37° C onwards to control levels (Figure S4). After temperature up shift, colony PCR of the 55°C colonies showed that 5 out of 10 tested pWUR_Cas9sp1_hr colonies were pure *pyrF* deletion mutants, i.e. an editing efficiency of 50% for our system (data not shown). The correct mutations were also verified by sequencing. None of the 10 tested pWUR_Cas9nt_hr colonies (non-targeting control) were *pyrF* deletion mutants demonstrating functional spCas9-targeting.

To improve the efficiency and speed of the system, we repeated the process for the pWUR_Cas9sp1_hr-containing strain and reduced the number of culturing steps at 37°C from 4 to 1 while keeping the culturing time of each step in a window between 8 and 16 hours. Moreover, we used 3 different media in order to observe possible medium-dependent variations in the efficiency of the system: TVMY selection medium supplemented with xylose, glucose and uracil (TVMYxgu), TVMY selection medium supplemented with xylose, glucose but not with uracil (TVMYxg) and LB2 medium supplemented with xylose and glucose (LB2xg). After the final culturing step at 55°C, cells were plated on selective agar of the corresponding medium, supplemented with uracil. Colony PCR on 10 randomly selected colonies for each medium and construct showed that from the culturing process on TVMYxgu medium 9 out of 10 colonies were pure *pyrF* deletion mutants colonies (Figure 4B). From the process with TVMYxg medium, 4 of the checked colonies were pure *pyrF* deletion mutants, proving the efficiency of the counter-selection tool even in the presence of an auxotrophy barrier (Figure 4B). The process with the LB2xg medium was repeated twice, with 2 different pWUR_Cas9sp1_hr containing ET 138 clones. Surprisingly, all the checked colonies from the LB2 medium process (40 in total) contained the wild type *pyrF* gene (data

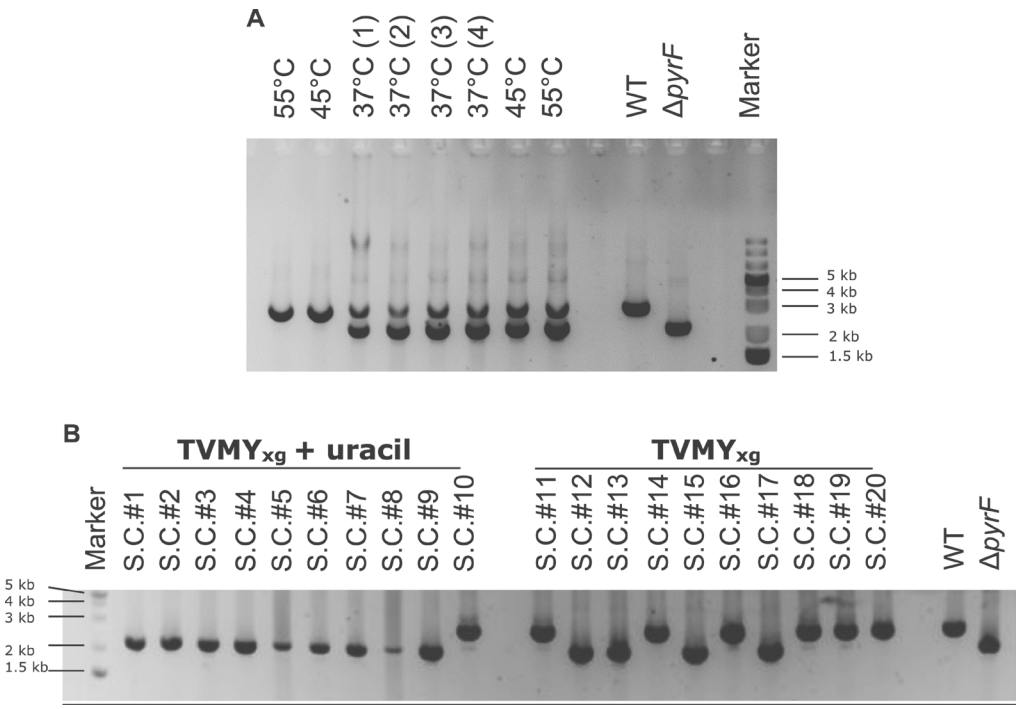


Figure 4. **A.** Agarose gel electrophoresis showing the results from PCR on the genomic DNA of a ET 138 Δ*ldhL* Δ*sigF* culture transformed with pWUR_Cas9sp1_hr and sequentially transferred to different temperatures (following the depicted temperature sequence) for detection of *pyrF* deletion mutants in the culture mixture. The *pyrF* deletion mutant band appears at the 1 st 37°C culturing step (Lane 3) onwards. The last 2 lanes are the negative (wild-type) and positive (Δ*pyrF*) controls, that correspond to 2.9 kb and 2.1 kb long DNA fragments respectively. **B.** Agarose gel electrophoresis showing the resulting products from colony PCR on colonies transformed with pWUR_Cas9sp1_hr for the detection of deletion mutants. 9 out of the 10 tested colonies (S.C.#1 to S.C.#10) that resulted from the 3-day long editing process in $TVMY_{xgu}$ (TVMY supplemented with xylose, glucose and uracil) medium were deletion mutants. 4 out of the 10 tested colonies (S.C.#11 to S.C.#20) that resulted from the 3-day long editing process in $TVMY_{xg}$ (TVMY supplemented with xylose and glucose) medium were deletion mutants. The last 2 lanes are the negative (wild-type) and positive (Δ*pyrF*) controls, respectively.

not shown). At this moment it is unclear what causes this difference between the LB2xg medium and the 2 types of TVMY media and this will be addressed in future research. To retain spCas9-activity, antibiotics were added in all steps. To allow for subsequent metabolic engineering steps, however, plasmid curing is required. After transferring a sequence-verified *pyrF* deletion mutant twice in TVMY medium without antibiotics cells were plated on TVMY plates without antibiotics. Colony PCR with plasmid-specific primers showed that all tested colonies had lost the plasmid. Finally, we verified the 5-FOA sensitivity and the uracil auxotrophy of 2 ET 138 Δ*ldhL* Δ*sigF* Δ*pyrF* cultures that originated from 2 of the tested colonies ET 138 (Figure S5).

EXPANDING THE TOOLBOX: KNOCKING OUT THE TYPE-I RM SYSTEM

To increase the potential of ET 138 as platform organism for the production of green chemicals we aimed to improve its transformation efficiency. ET 138 has a type I Restriction-Modification (R-M) system. Methylation analysis of the PacBio genome sequencing data showed the existence of the single motif "Cm6AGNNNNNNTGT/ACm6ANNNNNNCTG" with N6-methyladenine (m6A) modifications (unpublished data). Our attempts to transform different pNW33n-based constructs containing one or multiple copies of this motif were unsuccessful or gave 3-orders of magnitude lower transformation efficiencies compared to the constructs of the same size that do not contain the aforementioned motif. We decided to knock out the hsdR restriction gene of the *B. smithii* ET138 type I R-M system, expecting to overcome the transformation obstacle and further expand its genetic accessibility.

For knocking out the hsdR gene, we followed a different approach compared to the *pyrF* deletion process in order to be able to evaluate the efficiency of the Cas9-based counter-selection editing method in the introduction of limited-sized modifications to the genome. Between the origin of replication (ori) and the $cas9_{sp}$ gene of the pWUR_Cas9nt vector, we introduced a 1,920 bp long HR-fragment that resulted from the fusion of 2 separate fragments; one that is comprised of the 924 bp upstream of codon 212 of the *hsdR* gene (including the 2 first nucleotides of the codon) and one that is comprise d of the 986 bp downstream of codon 221 of the *hsdR* gene. In this HR-fragment we replaced the 25 nt long part between codons 212 and 221 of the *hsdR* gene, including the last nucleotide of codon 212, with an 8 nt long sequence comprised of 2 stop codons and the EcoRV restriction site, generating a frame shift and facilitating the screening process (Figure 3B). Since the *hsdR* gene is 2952 nucleotides (984 codons) long, only a fifth of it will be translated due the introduction of the stop codons. We also introduced a spacer in the sgRNA module for spCas9 targeting of the unmodified genomes, completing the *B. smithii* ET138 Δ*ldhL* Δ*sigF* cells were transformed with the new vector and sequentially cultured as before, gradually lowering the temperature from 55°C to 37°C, with an intermediate transfer at 45°C, and then back up to 55°C. The duration of each culturing step was within a window of 8 to 16 hours. Moreover, we used 2 types of selection media, LB2xg and TVMYxg. Five transformants per medium were subjected to colony PCR, after which the PCR fragments were digested with EcoRV. All colonies from the LB2-culturing process were successfully modified (Figure 5A) giving 100% editing efficiency, whereas only 2 of the colonies from the TVMY process were

modified giving 40% editing efficiency (Figure 5B). This is in contrast with the result from the *pyrF* deletion process, where there were no modified colonies resulting from the LB2-culturing process. This suggests that the selection of culturing medium influences the efficiency of our editing system in a variable and gene-specific manner. Plasmid curing was performed as before and the correct mutations were verified by sequencing.

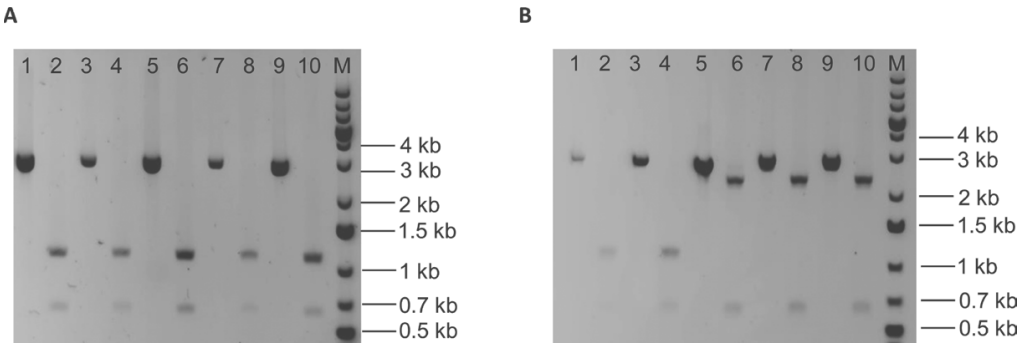


Figure 5. A. Agarose gel electrophoresis showing the resulting products from colony PCR on ET 138 colonies transformed with pWUR_Cas9spR_hr for the detection of hsdR knock-out mutants from the 3-day long editing process in LB2xg medium. The 2.75 kb PCR fragments resulting after using genome-specific primers (Lanes 1, 3, 5, 7, 9) were digested with the EcoRV restriction enzyme. Each digestion mixture was loaded next to its corresponding undigested PCR-fragment (Lanes 2, 4, 6, 8, 10). All the checked colonies were shown to be knock-out mutant cells as the restriction digestion gave the expected bands of 1.1, 1.05 and 0.6 kb. **B.** Agarose gel electrophoresis showing the resulting products from colony PCR on ET 138 colonies transformed with pWUR_Cas9spR_hr for the detection of hsdR knock-out mutants from the 3-d ay long editing process in TVMYxg mediu.m. The 2.75 kb PCR-fragments resulting after using genome-specific primers (Lanes 1, 3, 5, 7, 9) were digested with the EcoRV restriction enzyme. Each digestion mixture was loaded next to its corresponding undigested PCR-fragment (Lanes 2, 4, 6, 8, 10). For the 2 colonies comprised of knock-o ut mutant cells (Lanes 1-2 and 3-4) the restriction digestion gives the expected bands of 1.1, 1.05 and 0.6 kb. For the 3 non-edited colonies comprised of wild-type cells (Lanes 5-6, 7-8, 9-10), the restriction digestion gives the expected bands of 2.15 kb and 0.6 kb.

We confirmed the lack of a functional R-M system in the newly developed ET 138 Δ*ldhL* Δ*sigF* Δ*hsdR* strain by successfully transforming the plasmid-cured cells with vector pG2K (Table S2)[51]. In previous attempts we did not succeed in transforming this vector into other (*hsdR*+) ET 138 strains as it contains the aforementioned methylation motif in its antibiotic resistance marker gene, the kanamycin nucleotidyltransferase (*aadA*) gene derived from *Geobacillus stearothemophilus*. This way we added a new antibiotic resistance marker to the toolbox of ET 138, and we confirmed that the ET 138 Δ*ldhL* Δ*sigF* Δ*hsdR* strain can be utilized for the expansion of the genetic parts toolbox.

METABOLIC ENGINEERING USING SPCAS9: KNOCK-IN OF THE *LDHL* GENE

Next, we evaluated the applicability of our Cas9-based system in markerless gene chromosomal integrations by knocking into the genome of ET 138 Δ*ldhL* Δ*sigF* Δ*hsdR* the 942 bp long genomic fragment between the start and the stop codons of the lactate dehydrogenase (*ldhL*) gene. The reconstitution of the lactate production in the resulting ET 138 Δ*sigF* Δ*hsdR* strain would allow for efficient growth under anaerobic conditions, while retaining the advantages of a sporulation- and R-M-deficient strain.

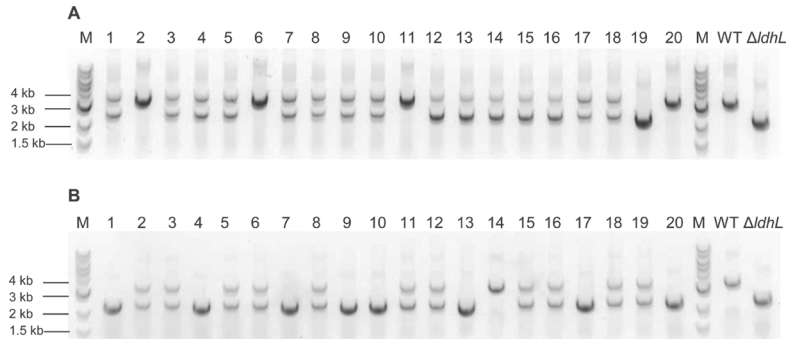


Figure 6. Agarose gel electrophoresis showing the resulting products from colony PCR on the colonies from the 3-day long *ldhL* knock-in culturing processes in LB2 medium, using ET 138 Δ*ldhL* Δ*sigF* Δ*hsdR* cells transformed with the pWUR_Cas9spKI_hr1 vector. For the colony PCR genome-specific primers BG8145 and BG8146 were used with the expected size of the PCR fragment for the knock-in mutations being 3.2 kb (equal to the size of the PCR fragment for the wild type cells). The expected size of the PCR fragment for the knock-out (non-edited) mutants was 2.3 kb. **A.** When the 45°C culturing step was performed, the process resulted in 4 out of the 20 tested colonies having the knock-in genotype (20% editing efficiency), 15 colonies with mixed knock-in and wild-type genotype and only 1 colony with wild-type genotype. **B.** When the 45°C culturing step was omitted, the process resulted in only 1 out of the 20 tested colonies having the knock-in genotype, 11 colonies with mixed genotype and 8 colonies with the wild-type genotype.

Two versions were constructed of a pWUR_Cas9-based vector that target the ET 138 Δ*ldhL* Δ*sigF* Δ*hsdR* genome at the same position between the *ldhL* stop codon and the beginning of the adjacent rho-independent transcriptional terminator. HR was facilitated with 1kb flanks (pWUR_Cas9spKI_hr1) or 0.75kb flanks (pWUR_Cas9spKI_hr2). For both versions, the region between the *ldhL* stop codon and its rho-independent transcriptional terminator was inverted, avoiding spCas9 targeting (Figure 3C). The region between the start and stop codon was provided with the wild-type *ldhL* sequence to allow its knocking in. ET138 Δ*ldhL* Δ*sigF* Δ*hsdR* was transformed with the 2 pWUR_Cas9spKI_hr versions and the transformants were sequentially cultured as described before, gradually lowering the temperature from 55 to 37°C, with or without an intermediate transfer at 45°C, an then back up to 55°C. Each culturing step was within a window of 8 to 16 hours. Again, we used 2 types of selection media, LB2xg and TVMYxg. The colony PCR results of the TVMY culturing pro

cesses showed that none out of the tested colonies had the knock-in genotype. The colony PCR results of the LB2 culturing processes with the pWUR_Cas9spKI_hr1 transformant showed that with the additional culturing step at 45°C, 4 out of the 20 tested colonies had the knock-in genotype (20% editing efficiency), 15 colonies had a mixed knock-in/wild type genotype and only 1 colony had the wild-type genotype (Figure 6A). When the culturing step at 45°C was omitted, only 1 out of the 20 tested colonies had the knock-in genotype, 11 colonies had the mixed genotype and 8 colonies had the wild-type genotype (Figure 6B). The colony PCR results of the LB2 culturing processes with the pWUR_Cas9spKI_hr2 transformant showed that with the additional culturing step at 45°C, 1 out of the 20 tested colonies had the knock-in genotype (5% editing efficiency), 7 colonies had the mixed knock-in and wild-type genotype, and only 1 colony had the wild-type genotype (Figure 6C). When the culturing step at 45°C was omitted none of the 20 tested colonies had the knock-in genotype, only 6 colonies had the mixed genotype while the remaining 14 colonies had the wild-type genotype (Figure 6D). This was the first time that we observed colonies with mixed genotypes using our editing approach. The appearance of such colonies could be explained by relatively inefficient Cas9 targeting when the enzyme is loaded with a suboptimal sgRNA module, as has been described for *E. coli*[53] although the molecular basis for this phenomenon remains elusive. Alternative sgRNAs may lead to a more stringent counter-selection, thereby improving the editing efficiency. Additionally, the difference in recombination efficiency at specific chromosomal sites might influence the editing efficiency of the tool[34]. In a recent study, that combined dsDNA recombineering with Cas9-counter-selection, the efficiency of the employed system for insertions was lower than the efficiency for deletions, while longer homologous regions lead to higher editing efficiencies[41]. Our results confirm the influence of the HR-template length, as well as the importance of the culturing period before the induction of the counter-selection. The editing efficiency of the tool was higher when we employed the editing construct with the 1 kb HR flanks compared to the editing construct with the 0.75 kb HR flanks (20% vs. 5% efficiency, respectively, Figure 6). Furthermore, we observed that a culturing period with an additional intermediate step at 45°C allows for efficient homologous recombination and double cross over events to occur, leading to the appearance of the mutants that Cas9 will select for. This is in line with the observations in *L. reuterii*[54] and supports that the efficiency of our approach to use spCas9 as a counter-selection tool is higher compared to using the spCas9 as a tool for induction of the cellular HR mechanism after introduction of targeted DS-

DBs. In addition, it may be that the stress of the temperature drop increases the efficiency of the homologous recombination mechanism.

Next, we attempted to cure the constructed *B. smithii* ET138 Δ*sigF* Δ*hsdR* strain from the pWUR_Cas9spKI_hr1 plasmid using the sequential transferring approach in LB2 medium without antibiotic at 55°C. However, after the usual 2 transfers none of the tested colonies had lost the plasmid. We repeated the same sequential transferring process but raised the culturing temperature to 65°C, as the pNW33n replicon might be less stable at elevated temperatures. After 2 transfers at 65°C, 1 out of 8 tested colonies was confirmed to be plasmid-free by PCR and antibiotic sensitivity and the correct mutations were verified by sequencing. Plasmid curing might be simplified in the future by adding to the system an sgRNA expressing module with a spacer against the editing plasmid[41,55]; the module will have to either be under the control of a tightly inducible promoter or cloned into a second expression vector and transformed into the edited cells in a second transformation round.

Finally, evaluation of lactate production in the resulting *B. smithii* ET138 Δ*sigF* Δ*hsdR* strain under aerobic, microaerobic and anaerobic conditions showed a complete restoration of lactate production to wild-type levels (Table S4).

CONCLUSIONS

In this study, spCas9-based genome editing was applied for the first time to a moderate thermophile, establishing a gene deletion, a gene knockout and a gene insertion. Major advantage of this system is the requirement of only one plasmid and no inducible or highly characterized promoters to drive the spCas9 and sgRNA expression. Additionally, the speed and efficiency of the genome editing process of ET 138 has been substantially improved compared to the previous *lacZ*-based counter-selection system. For the three cases presented in the present study, it took on average 1 week from transformation to clean deletion, knock out or knock in –including the plasmid curing step- with an editing efficiency of 90% for the gene deletion, 100% for the gene knock out and 20% for the gene insertion.

In the course of our study we showed that spCas9 is not active *in vivo* above 42°C. This observation allowed us to develop an editing system where mutants are constructed via homologous recombination events at higher temperatures (>42°C) before Cas9-induced counter-selection takes place at 37°C. Crucial factors for obtaining high editing efficiencies is to allow for sufficient time for HR at elevated temperatures before starting the spCas9-based counter-selection, and the length of the HR flanks. Moreover, we hypothesize that testing different sgRNA modules may improve the editing efficiency of the tool. During our study we

we observed gene-dependent differences in numbers of obtained mutants when repeating the same process using different media. There is no obvious link between the editing of a specific gene and the medium used for the editing process and this will be subject of further studies.

The results of the editing approach we developed make our system potentially applicable for many interesting model or non-model organisms with an active HR mechanism and a growth temperature range covering 37°C to over 42°C. It is anticipated that the approaches reported here will expand the range of organisms for which the powerful Cas9-counter-selection tool can be used, thereby greatly increasing engineering throughput for these organisms and allowing for both their fundamental study and biotechnological exploitation.

MATERIALS AND METHODS

BACTERIAL STRAINS AND GROWTH CONDITIONS

Strains used in this study are listed in Table S1. All *B. smithii* strains were routinely cultured at 55°C unless stated otherwise. TVMY medium and LB2 medium were used as described previously. $TVMY_{xgu}$ is TVMY supplemented with 0.5 g/l xylose, 0.5 g/l glucose and 50 mg/L uracil. $TVMY_{xg}$ is TVMY supplemented with 0.5 g/l xylose and 0.5 g/l glucose. $LB2_{xg}$ is LB2 supplemented with 0.5 g/l xylose and 0.5 g/l glucose. Substrates were added separately as 50% autoclaved solutions after autoclavation of the medium. Uracil was added as 50 mg/ml filter sterilized solution in 1M NaOH after autoclavation of the medium and addition of the substrates. *E. coli* strains were grown in LB medium at 37°C. For plates, 30 g of agar (Difco) per liter of medium was used for *B. smithii* in all experiments; 15 g of agar (Difco) per liter of LB was used for *E. coli*. If required, chloramphenicol was added in concentrations of 25 µg/ml for *E. coli* DH5α, 15 µg/ml for *E. coli* TG90, and 7 µg/ml for *B. smithii.*

GENOMIC DNA ISOLATION, TRANSFORMATIONS, COLONY PCR, SEQUENCE AND PHENOTYPIC VERIFICATION

Genomic DNA from *B. smithii* strains was isolated using the MasterPure™ Gram Positive DNA Purification Kit (Epicentre). Heat shock transformation of *E. coli* strains was performed according to the NEB supplier's protocol. Transformation of *B. smithii* strains was performed as described previously[15]. Plasmids for transforming *B. smithii* were extracted from *E. coli* via maxiprep isolation (Genomed Jetstar 2.0). For transformation of *B. smithii* strains, 1 µg DNA was used unless stated otherwise in the plasmid construction sections.

Potential *B. smithii* ET138 Δ*ldhL* Δ*sigF* Δ*pyrF* colonies were randomly selected and subjected to colony PCR using the InstaGene™ Matrix (BIORAD), Taq DNA Polymerase (NEB) and the genome specific primers BG6420 and BG6421. Potential *B. smithii* ET138 Δ*ldhL* Δ*sigF* Δ*hsdR* and *B. smithii* ET138 Δ*sigF* Δ*hsdR ldhL* knock-in colonies were randomly selected and subjected to colony PCR using the Phire Plant Direct PCR kit (ThermoFisher Scientific) and the genome specific primers BG7881, BG7882 and BG8142, BG8143 respectively. Purification of PCR products was performed using the Zymoclean™ Gel DNA Recovery Kit, after running them on 0.8% agarose gels. The DNA fragments were subsequently sent for sequencing to GATC Biotech. The DNA fragments from the potential *B. smithii* ET138 Δ*ldhL* Δ*sigF* Δ*hsdR* colonies were subjected to EcoRV (NEB) restriction digestion. To evaluate the 5-FOA sensitivity and uracil auxotrophy of *B. smithii* ET138 Δ*ldhL* Δ*sigF* Δ*pyrF* sequence confirmed strains, cells were plated on TVMY medium with 30 g/l agar and the following additions: a) 2 g/l 5-FOA and 50 mg/l uracil, b) 2 g/l 5-FOA and no uracil, d) no 5-FOA and no uracil. To evaluate lactate production from *B. smithii* ET138 Δ*sigF* Δ*hsdR ldhL* knock-in cultures, sequence verified cells were grown overnight in TVMY medium containing 10 g/l glucose and subsequently transferred to the same medium and grown for 24 h, after which L-lactate specific measurements were performed using MegaZyme kit K-LATE.

PLASMID CONSTRUCTION

Plasmids and primers used in this study are shown in Table S2 and S3. Q5 polymerase (NEB) was used for all PCR reactions for cloning purposes. NEB T4 ligase was used for assembling vectors pWUR_lacZ, pWUR_Cas9nt, pWUR_Cas9nt_hr, pWUR_Cas9sp1, pWUR_Cas9sp1_hr, pWUR_Cas9sp2 and pWUR_Cas9sp3. The NEBuilder® HiFi DNA Assembly Master Mix was used for assembling the pWUR_Cas9spR_hr, pWUR_Cas9spKI_hr1 and pWUR_Cas9spKI_hr2 constructs. All the used restriction enzymes were obtained from NEB. Purification of PCR products was performed after running them on a 0.8% agarose gel using the Zymoclean™ Gel DNA Recovery Kit. To test the PxynA promoter, a DNA fragment comprised of the PxynA and the lacZ gene was synthesized by GeneArt and inserted into pNW33n using digestion with BspHI and KpnI and subsequent ligation and cloning into *E. coli* DH5α, creating plasmid pWUR_lacZ. The P_{xynA} sequence was used exactly as originally described, using the sequence until the start codon of the corresponding gene in the original host.

For the construction of the basic, modular pWUR_Cas9nt construct, a synthetic gene string was synthesized by GeneArt containing the elements depicted in Figure 1B except the P_{xynA} promoter. P_{xynA} was amplified from pWUR_lacZ using primers BG6538 and BG6541. Primer BG6541 replaces the final 6 bp of P_{xynA} by an XbaI site, changing the final -1 to -6 sequence from GTAAGA to TCTAGA and keeping the total length the same as in the original promoter. Primer BG6538 adds a BspHI site on the start of the PxynA. The entire synthesized spCas9 module without promoter for spCas9 was amplified using primers BG6542 and BG6543, keeping the XbaI and HindIII sites already present in the module. Subsequently, vector pNW33n was digested with BspHI and HindIII, the P_{xynA} PCR-product with XbaI and BspHI and the spCas9 module PCR-product with XbaI and HindIII. The three elements were ligated in a 3-point ligation and cloned into *E. coli* TG90. Plasmid was extracted and the correct sequence was verified by sequencing, creating plasmid pWUR_Cas9nt (Figure 2B). For transformation of this construct to *B. smithii* strains, 0.1 µg DNA was used rather than the standard 1 µg in order to more precisely determine the transformation efficiencies together with the targeting constructs described in the next section. The resulting CFUs were around 3,000 and 200 per µg DNA for the Δ*ldhL* Δ*sigF* and wild-type strain, respectively. Control transformations with empty vector pNW33n yielded CFUs of 10,000 and 1,800, respectively.

To insert the 3 different targeting spacers into pWUR_Cas9nt (which contains a non-targeting spacer), 3 sets of oligos were annealed to create the 3 spacers, after which the annealed spacers were inserted into the construct as follows. Oligo sets were BG6017 and BG6021 for spacer 1, BG6018 and BG6022 for spacer 2, BG6019 and BG6023 for spacer 3. Each set was annealed by adding 5 µl 10 mM oligo sets together with 10 µl NEB buffer 2.1 and 74 µL MQ water. Mixtures were heated to 94°C for 5 min and gradually cooled down to 37°C at 0.03°C/sec using a PCR machine. Annealed oligos and plasmid pWUR_Cas9nt were digested with BspEI and BsmBI (NEB). First, BspEI digestion was performed at 37°C for 15 min, after which BsmBI was added and the mixture was further incubated at 55°C for 15min. After gel purification of the digested products, ligation was performed using NEB T4 ligase and mixtures were transformed to *E. coli* TG90. All constructs were verified by sequencing spacer 2, which is missing 7 nt from the Ppta that is driving spacer expression (Figure S2).

Constructs were named pWUR_Cas9sp1 until pWUR_Cas9sp3 according to their corresponding spacer. For transformation of this construct to *B. smithii* strains, 0.1 µg DNA was used rather than the standard 1 µg in order to more precisely determine the transformation efficiencies. The resulting CFUs were 700-3,300 and 9-300 per µg DNA for

the Δ*ldhL* Δ*sigF* and wild-type strain, respectively. Control transformations with empty vector pNW33n yielded CFUs of 10,000 and 1,800, respectively.

To insert the *pyrF*-flanks into the pWUR_Cas9nt and pWUR_Cas9sp1 constructs, the already fused *pyrF*-flanks were amplified from a previous plasmid in which the flanks were added as follows: flanks were cloned from genomic DNA of ET 138 using primers BG 5798, BG5799 (upstream, 958 bp) and BG580, BG5801 (downstream, 979 bp), introducing SalI and XbaI restriction sites. The flanks were fused in an overlap extension PCR using primers BG5798 and BG5801 making use of the complementary overhangs in primers BG5799 and BG5800. Subsequently, the flanks and pNW33n were digested with SalI and XbaI, ligated and transformed into *E. coli* DH5α. To amplify the flanks for insertion into spCas9-editing plasmids, primers BG6850 and BG6849 were used, which both introduce a BspHI site. The pWUR_Cas9nt and pWUR_Cas9sp1 plasmids, and the amplified *pyrF*-flanks were digested with BspHI, followed by alkaline phosphatase treatment of the vectors (Thermo Scientific), ligated and transformed in *E. coli* TG90. Since only one restriction site was used, the flanks could have been inserted in both orientations. For both constructs multiple colonies were verified by sequencing and for all constructs the same flank-orientation was selected and used for future experiments, namely with the downstream flank on the P_{xynA} side. The resulting plasmids were named pWUR_Cas9nt_hr and pWUR_Cas9sp1_hr. Transformation of these constructs to *B. smithii* Δ*ldhL* Δ*sigF* using 0.1 µg DNA like for the non-flanked versions did not yield any colonies, most likely because there is a RM-recognition site present in the flanking regions (data not shown). Transformations performed using 5-8 µg DNA resulted in CFUs of around 10 per µg DNA (compared to around 700-3,300 for the non-flanked versions and 10,000 for control transformations with empty vector pNW33n).

A 4-fragment NEBuilder® HiFi DNA Assembly was designed and executed for the construction of the *hsdR*–modifying plasmid pWUR_Cas9spR_hr. The backbone of the vector was PCR amplified from pWUR_Cas9sp1 using primers BG7836 and BG7837. The HR fragment upstream of the targeted site in the *hsdR* gene was PCR amplified from the *B. smithii* ET 138 genome using primers BG7838 and BG7839. The HR fragment downstream of the targeted site in the hsdR gene was PCR amplified from the ET 138 genome using primers BG7840 and BG7841. The *cas9sp* and sgRNA containing fragment was PCR amplified from the pWUR_Cas9sp1 vector using primers BG7842 and BG7843.

Two 4-fragment NEBuilder® HiFi DNA Assemblies were designed and executed for the construction of the *ldhL*-restoration plasmids pWUR_Cas9spKI_hr1 and pWUR_Cas9

spKI_hr2. The backbone of both vectors was PCR amplified from the pWUR_Cas9sp1 vector using primers BG8134 and BG7837. The HR fragment upstream of and including the *ldhL* gene was PCR amplified from the *B. smithii* ET 138 genome using primers BG8135 and BG8137 for the pWUR_Cas9spKI_hr1 vector and primers BG8135 and BG8136 for the pWUR_Cas9spKI_hr2 vector. The HR fragment downstream of the *ldhL* gene was PCR amplified from the *B. smithii* ET 138 genome using primers BG8138 and BG8139 for the pWUR_Cas9spKI_hr1 vector and primers BG8138 and BG8140 for the pWUR_Cas9spKI_hr2 vector. The cas9sp and sgRNA containing fragment of both vectors was PCR amplified from the pWUR_Cas9sp1 vector using primers BG8141 and BG7842.

RNA ISOLATION AND RT-PCR

RNA isolation was performed by the phenol extraction based on Van Hijum *et al.*. Overnight 10 ml cultures were centrifuged at 4°C and 4816 x g for 15 min and immediately used for RNA isolation. After removal of the medium, cells were resuspended in 0.5 ml 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 zirconium beads, 30 µl 10% SDS, 30 µl 3 M sodium acetate (pH 5.2), and 500 µl Roti-Phenol (pH 4.5-5.0, Carl Roth GmbH). Cells were disrupted using a FastPrep-24 apparatus (MP Biomedicals) at speed 5500 rpm for 45 sec and centrifuged at 4°C and 10,000 rpm for 5 min. 400 µl of the water phase from each tube was transferred to a new tube, to which 400 µl chloroform-isoamyl alcohol (Carl Roth GmbH) was added, after which samples were centrifuged at 4°C and 18,400 x 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. Integrity and concentration of the isolated RNA was checked on Nanodrop-1000.

Reverse Transcriptase PCR was performed using SuperScript III Reverse Transcriptase kit (Invitrogen) according to the manufacturer´s protocol. For synthesis of the first-strand cDNA, 2 µg of RNA and 200 ng of random primers were used. After cDNA synhesis, the products were used as a template for PCR using spCas9-specific forward and reverse primers BG6237 and BG6232, resulting in a 255 bp product. Products were visualized on a 2% agarose gel ran for 20 min.

CONFLICT OF INTEREST

The authors declare no competing financial interest. RvK is employed by the commercial company Corbion (Gorinchem, The Netherlands).

ACKNOWLEDGEMENTS

The authors thank Dr. Ben Reeves for the kind donation of the pG2K vector and Prof. Dr. Oscar Kuipers for kindly providing *E. coli* TG90. This work was 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.

REFERENCES

1. Bosma, E. F., van der Oost, J., de Vos, W. M., and van Kranenburg, R. (2013) Sustainable production of bio-based chemicals by extremophiles, Curr. Biotechnol. 2, 360-379.
2. Lin, L., and Xu, J. (2013) Dissecting and engineering metabolic and regulatory networks of thermophilic bacteria for biofuel production, Biotechnol. Adv. 31, 827-837.
3. Olson, D. G., Sparling, R., and Lynd, L. R. (2015) Ethanol production by engineered thermophiles, Curr. Opin. Biotechnol. 33, 130-141.
4. Beckner, M., Ivey, M. L., and Phister, T. G. (2011) Microbial contamination of fuel ethanol fermentatio ns, Lett.Appl. Microbiol. 53, 387-394.
5. Qin, J., Zhao, B., Wang, X., Wang, L., Yu, B., Ma, Y., Ma, C., Tang, H., Sun, J., and Xu, P. (2009) Non-sterilized fermentativ eproduction of polymer-grade L-lactic acid by a newly isolated thermophilic strain Bacillus sp. 2-6, PLoS One 4, e4359.
6. Abdel-Banat, B. M., Hoshida, H., Ano, A., Nonklang, S., and Akada, R. (2010) High-temperature fermentatio n:how can processes for ethanol production at high temp eraturesbecome superior to the traditional process using mesophilic yeast?, Appl. Microbiol. Biotechnol. 85, 861-867.
7. Ouyang, J., Ma, R., Zheng, Z., Cai, C., Zhang, M., and Jiang, T. (2013) Open fermentativ eproduction of L-lactic acid by Bacillus sp. strain NL01 using lignocellulosic hydrolyzates as low-cost raw material, Bioresour. Technol. 135, 475-480.
8. Ye, L., Zhou, X., Hudari, M. S., Li, Z., and Wu, J. C. (2013) Highly efficient production of L-lactic acid from xylose by newly isolated Bacillus coagulans C106, Bioresour. Technol. 132, 38-44.
9. Taylor, M. P., Eley, K. L., Martin, S., Tuffin, M. I., Burton, S. G., and Cowan, D. A. (2009) Thermophilic ethanologenesis: future prospects for second-generation bioethanol production, Trends Biotechnol. 27, 398-405.
10. Kambam, P. K. R., and Henson, M. A. (2010) Engineering bacterial processes for cellulosic ethanol production, Biofuels 1, 729-743.
11. Ma, K., Maeda, T., You, H., and Shirai, Y. (2014) 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
12. Ou, M., Mohammed, N., Ingram, L., and Shanmugam, K. (2009) Thermophilic Bacillus coagulans requires less cellulases for simultan eous saccharification and fermentation of cellulose to products than mesophilic microbial biocatalysts, Appl. Biochem. Biotechnol. 1, 76-82.

13. Bhalla, A., Bansal, N., Kumar, S., Bischoff, K. M., and Sani, R. K. (2013) Improved lignocellulose conversion to biofuels with thermophilic b acteria and thermostableenzymes, Bioresour. Technol. 128,751-759.
14. Taylor, M. P., van Zyl, L., Tuffin, I. M., Leak, D. J., and Cowan, D. A. (2011) Genetic tool development underp ins recent advancesin thermophilic whole-cell biocatalysts,Microb. Biotechnol4,438-448.
15. Bosma, E. F., van de Weijer, A. H. P., van der Vlist, L., de Vos, W. M., van der Oost, J., and van Kranenburg, R. (2015) Establishment of markerless gene deletion tools in thermophilic *Bacillus smithii* and construction of multiple mutant strain s, Microb. Cell Fact. 14, Art.nr.99.
16. Bosma, E. F., van de Weijer, A. H. P., Daas, M. J. A., van der Oost, J., de Vos, W. M., and van Kranenburg, R. (2015) 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, 1874-1883.
17. Nakamura, L. K., Blumen stock, I., and Claus, D. (1988) Taxonomic studyof Bacillus coagulans Hammer 1915 with a proposal for *Bacillus smithii* sp. nov, Int. J. Syst. Bacteriol. 38, 63-73.
18. Van Spanning, R. J. M., Wansell, C. W., Reijnders, W. N. M., Harms, N., Ras, J., Oltmann, L. F., and Stouthamer, A. H. (1991) A method for introduction of unmark ed mutations in thegenome of Paracoccus denitrificans: Construction of strains with multiple mutations in the genes encoding periplasmic cytochromes c550, c(551i), and c(553i), J. Bacteriol. 173, 6962-6970.
19. Wang, Q., Ingram, L. O., and Shanmugam, K. T. (2011) Evolution of D-lactate dehydrogenase activity from glycerol dehydrogenase and its utility for D-lactate production from lignocellulose, Proc. Natl. Acad. Sci. U. S. A. 108, 18920-18925.
20. Wang, Q., Chen, T., Zhao, X., and Chamu, J. (2012) Metabolic engineering of thermophilic Bacillus licheniformis for chiral pure D-2,3 -butanediolproduction, Biotechnol. Bioeng. 109, 1610-1621.
21. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes, Science 315, 1709-1712.
22. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity, Science 337, 816-821.
23. Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013) RNA-programmed genome editing in human cells, eLife 2, e0 0471.
24. Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M.(2013) RNA-guided human genome engineering via Cas9, Science 339, 823-826.
25. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F. (2013) Multiplex genome engineering using CRISPR/Cas systems, Science 339, 819-823.
26. Kim, S., Kim, D., Cho, S. W., Kim,J., and Kim, J.-S. (2014) Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins, Genome Res. 24, 1012-1019.
27. Komor, A. C.,Badran, A. H., and Liu, D. R. (2016) CRISPR-based technologies for the manipulation of eukaryotic genomes, Cell.
28. Puchta, H. (2017) Applying CRISPR/Cas for genome engineering in plants: the best is yet to come, Curr. Opin. Plant Biol. 36, 1-8.
29. Xu, J., Ren, X., Sun, J., Wang, X., Qiao, H. H., Xu, B. W., Liu, L. P., and Ni, J. Q. (2015) A Toolkit of CRISPR-Based Genome Editing Systems in Drosophila, Journal of genetics and genomics = Yi chuan xuebao 42, 141-149.
30. Bowater, R., and Doherty, A. J. (2006) Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining, PLoS Genet. 2, 9 3-99.
31. Bosma, E. F.,Koehorst, J. J., van Hijum, S. A., Renckens, B., Vriesendorp, B., van de Weijer, A. H., Schaap, P. J., de Vos, W. M., van der Oost, J., and van Kranenburg, R. (2016) Complete genome sequence of thermophilic *Bacillus smithii* type strainDSM 4216(T), Stand. Genomic Sci. 11, 52.
32. Shuman, S., and Glickman, M. S. (2007) Bacterial DNA repair by non-homologous end joining, Nature Reviews Microbiology 5, 852-861.
33. Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L. A. (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems, Nat. Biotechnol . 31, 233-239.
34. Barrangou, R., and van Pijkeren, J. P. (2016) Exploiting CRISPR-Cas immune systems for genome editing in bacteria, Curr. Opin. Biotechnol. 37, 61-68.
35. Mougiakos, I., Bosma, E. F., de Vos, W. M., van Kranenburg, R., and van der Oost, J. (2016) Next generation prokaryotic engineering: The CRISPR-Cas toolkit, Trends Biotechnol. 34, 575-587.
36. Xu, T., Li, Y., Shi, Z., Hemme, C. L., Li, Y., Zhu, Y.,Van Nostrand, J. D., He, Z., and Zhou, J. (2015) Efficient genome editing in *Clostridium cellulolyticum* via CRISPR-Cas9 nickase, Appl. Environ. Microbiol. 81, 4423-4431.

37. Huang, H., Zheng, G., Jiang, W., Hu, H., and Lu, Y. (2015) One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*, Acta Biochim Biophys. Sin. 47, 231-243.
38. Huang, H., Chai, C., Li, N., Rowe, P., Minton, N. P., Yang, S., Jiang, W., and Gu, Y. (2016) CRISPR/Cas9-based efficient genome editing in *Clostridium ljungdahlii,* an autotrophic gas-fermenting bacterium, ACS Synth. Biol. 5, 1355-1361.
39. Li, Q., Chen, J., Minton, N. P., Zhang, Y., Wen, Z., Liu, J., Yang, H., Zeng, Z., Ren, X., Yang, J., Gu, Y., Jiang, W., Jiang, Y., and Yang, S. (2016) CRISPR-based genome editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii,* Biotechnol. J. 11, 961-972.
40. Wang, Y., Zhang, Z. T., Seo, S. O., Choi, K., Lu, T., Jin, Y. S., and Blaschek, H. P. (2015) Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRI SPR/Cas9 system, J. Biotechnol. 200, 1-5.
41. Li, Y., Lin, Z., Huang, C., Zhang, Y., Wang, Z., Tang, Y.-j., Chen, T., and Zhao, X. (2015) Metabolic engineering of *Escherichia coli* using CRISPR–Cas9 meditated genome editing, Metab. Eng. 31, 13-21.
42. Wang, Y., Zhang, Z. T., Seo, S. O., Lynn, P., Lu, T., Jin, Y. S., and Blaschek, H. P. (2016) Bacterial genome editing with CRISPR-Cas9: deletion, integration, single nucleotide modification, and desirable "clean" mutant selection in *Clostridium beijerin* ckii as an example, ACS Synth. Biol. 5, 721-732.
43. Tong, Y., Charusanti, P., Zhang, L., Weber, T., and Lee, S. Y. (2015) CRISPR-Cas9 based engineering of Actinomycetal genomes, ACS Sy nth. Biol. 4, 1020-1029.
44. Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J., and Yang, S. (2015) Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 sy stem, Appl. Environ. Microbiol. 81, 2506-2514.
45. Li, Y., Pan, S., Zhang, Y., Ren, M., Feng, M., Peng, N., Chen, L., Liang, Y. X., and She, Q. (2016) Harnessing Type I and Type III CRISPR-Cas systems for genome e diting, Nucleic Acids Res. 44, e34.
46. Zeldes, B. M., Keller, M. W., Loder, A. J., Straub, C. T., Adams, M. W., and Kelly, R. M. (2015) Extremely thermophilic microorganisms as metab olic engineering platforms for production of fuels and industrial chemicals, Front. Microbiol. 6, 1209.
47. Currie, D., Herring, C., Guss, A., Olson, D., Hogsett, D., and Lynd, L. (2013) Functional heterologous expression of an engineered full length CipA from *Clostridium thermocellum* in Thermoanaerobacteriu m saccharolyticum, Biotechnol. Biofu els 6, 32.
48. van Hijum, S. A., de Jong, A., Baerends, R. J., Karsens, H. A., Kramer, N . E., Larsen, R., den Hengst, C. D., Albers, C. J., Kok, J., and Kuipers, O. P. (2005) A generally applicable validation scheme for the assessment of factors involved in reproducibility and quality of DNA-microarray data, BMC Genomics 6 ,77.
49. Kovács, Á. T., van Hartskamp, M., Kuipers, O. P., and van Kranenburg, R. (2010) Genetic tool development for a new host for biotechnology, the thermotoleran t bacterium Bacillus coagulans, Appl. Environ. Microbiol. 76, 4085-4088.
50. Tripathi, S. A., Olson, D. G., Argyros, D. A., Miller, B. B., Barrett, T. F., Murphy, D. M., McCool, J. D., Warner, A. K., Rajgarhia, V. B., Lynd, L. R., Hogsett, D. A., and Caiazza, N. C. (2010) 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.
51. Kita, A., Iwasaki, Y., Yano, S., Nakashimada, Y., Hoshino, T., and Murakami, K. (2013) Isolation of thermophilic acetogens and transformation of them with the pyrF and kanr genes, Biosci., Biotechnol., Biochem. 77, 301-306.
52. Chung, D., Farkas, J., Huddleston, J. R., Olivar, E., and Westpheling, J. (2012) Methylation by a unique alpha-class N4-cytosine methyltransferase is required for DNA transformation of Caldicellulosiruptor bescii DSM6725, PLoS One 7, e43844.
53. Biswas, I., Gruss, A., Ehrlich, S. D., and Maguin, E. (1993) High-efficiency gene inactivation and replacement system for gram-positive bacteria, J. Bacteriol. 175, 3628-3635.
54. Morel-Deville, F., and Ehrlich, S. D. (1996) Theta-type DNA replication stimulates homolog ous recombination in the Bacillus subtilis chromosome, Mol. Microbiol. 19, 587-598.
55. Reeve, B., Martinez-Klimova, E., de Jonghe, J., Leak, D. J., and Ellis, T. (2016) The Geobacillus plasmid set: a modular toolkit for thermophile engineering ,ACS Synth. Biol. 5, 1342-1347.
56. Cui, L., and Bikard, D. (2016) Consequences of Cas9 cleavage in the chromoso me of *Escherichia coli*, Nucleic Acids Res. 44, 4243-4251.
57. Oh, J.-H., and van Pijkeren, J.-P. (2014) CRISPR–Cas9-assisted recombineering in Lactobacillus reuteri, Nucleic Acids Res. 42, e131.
58. Ronda, C., Pedersen, L. E., Sommer, M. O., and Nielsen, A. T. (2016) CRMAGE: CRISPR optimized MAGE recombineering, Sci. Rep. 6, 19452.
59. Kingsford, C. L., Ayanbule, K., and Salzberg, S. L. (2007) Rapid, accurate, computational discovery of Rho independent transcription termin ators illuminates their relationship to DNA uptake, Genome Biol. 8, R22.

SUPPLEMENTARY INFO RMATIO N

SUPPLEMENTARY TABLES

Table S1. Strains used in this study

Strain	Description	Reference/origin
E. coli DH5α	Cloning host	NEB
E. coli TG90	Cloning host	[1]
B. smithii		
strains: ET 138	Wild-type, natural isolate	[2]
ET 138 Δ*ldhL* Δ*sigF*	ET 138 Δ*ldhL* Δ*sigF* with *pyrF*-deletion	[3]
ET 138 Δ*ldhL* Δ*sigF* Δ*pyrF*	ET 138 with *ldhL* and *sigF*-deletions	This study
ET 138 Δ*ldhL* Δ*sigF* Δ*hsdR*	ET 138 Δ*ldhL* Δ*sigF* with stop codons and frame shift inserted in the *hsdR* gene	This study
ET 138 Δ*sigF* Δ*hsdR*	ET 138 ΔiΔsigF Δ*hsdR* with the l*dhL* gene re-inserted to restore its original sequence	This study

Table S2. Plasmids used in this study

Plasmid	Description	Reference/origin
pNW33n	*E. coli-Bacillus* shuttle vector, cloning vector, Cm^R.	BGSC
pWUR_lacZ	pNW33n with *B. coagulans* lacZ gene under *T. saccharolyticum* P_{xynA} promoter.	This study; [4,5]
pWUR_Cas9nt	pNW33n with spCas9-module[1] containing non-targeting spacer.	This study
pWUR_Cas9sp1	pNW33n with spCas9-module[1] containing spacer targeting the *pyrF* gene at bp nr 124	This study
pWUR_Cas9sp2	pNW33n with spCas9-module[1] containing spacer targeting the *pyrF* gene at bp nr 267	This study
pWUR_Cas9sp3	pNW33n with spCas9-module[1] containing spacer targeting the *pyrF* gene at bp nr 583	This study
pWUR_Cas9nt_hr	pNW33n with spCas9-module[1] containing non-targeting spacer and the fused us+ds *pyrF*-flanks.	This study
pWUR_Cas9sp1_hr	pNW33n with spCas9-module[1] containing spacer targeting the *pyrF* gene and the fused us+ds *pyrF*-flanks.	This study
pWUR_Cas9spR_hr	pNW33n with spCas9-module[1] containing spacer targeting the *hsdR* gene and the fused us+ds *hsdR*-flanks containing mutations as described in M&M.	This study
pG2K	Modular shuttle vector for *Geobacillus* and *E. coli*	Reeve *et al,* 2016[6]
pWUR_Cas9spKI_hr1	pNW33n with spCas9-module[1] containing spacer targeting directly downstream of the *ldhL* gene and the fused us+ds *ldhL*-flanks (~1000 bp each), in which the area between the *ldhL* stop codon and its terminator is inverted.	This study
pWUR_Cas9spKI_hr2	pNW33n with spCas9-module[1] containing spacer targeting directly downstream of the *ldhL* gene and the fused us+ds *ldhL*-flanks (~750 bp each), in which the area between the *ldhL* stop codon and its terminator is inverted.	This study

[1] The spCas9 module contains *S. pyogenes cas9* under the *T. saccharolyticum* P_{xynA} promoter followed by a *B. subtilis*-derived Rho-independent terminator, followed by a spacer under the *B. coagulans* P_{pta} promoter followed by another *B. subtilis*-derived Rho-independent terminator (Figure 1).

Abbreviations: Cm^R: chloramphenicol resistance gene (chloramphenicol acetyltransferase); BGSC: *Bacillus* Genetic Stock Centre, USA; us: upstream ;ds: downstream ; bp: base pairs.

Table S3. Primers used in this study

BG nr	Sequence 5'-3'[1]	Target/purpose
5798	GCCTCTAGATAGGGATAACAGGGTAATATGCAGCGAT GGTCCGGTGTTC	*pyrF*-us-Fw, XbaI
5799	TTTCAGATCTGCTGGTTTACACGCACTTTCCAGCTCCTT C	*pyrF*-us-Rv, overhang with BG5800
5800	GAAGGAGCTGGAAAGTGCGTGTAAACCAGCAGATCTG AAA	*pyrF*-ds-Fw, overhang with BG5799
5801	GCCGTCGACTAGGGATAACAGGGTAATGTTCCCATGTT GTGATTC	*pyrF*-ds-Rv, SalI
6017	CCGGAAAAATGGATTATTCTTAAAATGGATACCGTTAT ACACTTTATTTTCAGAATGGAC	Fw oligo spacer 1
6018	CCGGAAAAATGGATTATTCTTAAAATGGATACCGTTAT ACTCAACGTGCATGCAGCAGGC	Fw oligo spacer 2
6019	CCGGAAAAATGGATTATTCTTAAAATGGATACCGTTAT ACGCAAAAGCGGCTGGTTGCAA	Fw oligo spacer 3
6021	AAACGTCCATTCTGAAAATAAAGTGTATAACGGTATCC ATTTTAAGAATAATCCATTTTT	Rv oligo spacer 1
6022	AAACGCCTGCTGCATGCACGTTGAGTATAACGGTATCC ATTTTAAGAATAATCCATTTTT	Rv oligo spacer 2
6023	AAACTTGCAACCAGCCGCTTTTGCGTATAACGGTATCC ATTTTAAGAATAATCCATTTTT	Rv oligo spacer 3
6850	GCCTCATGAATGCAGCGATGGTCCGGTGTTC	*pyrF*-us-Fw, BspHI site
6849	GCCTCATGAGTTCCCATGTTGTGATTC	*pyrF*-ds-Rv, BspHI site
6237	ATTGTGATCCCTAGTAACTC	spCas9-specific-Fw, to evaluate expression
6232	GCGCAAAGTATTGTCCATGCC	spCas9-specific-Rv, to evaluate expression
6420	TCGGGGGTTCGTTTCCCTTG	*pyrF* KO check Fw
6421	CTTACACAGCCAGTGACGGAAC	*pyrF* KO check Rv
7836	gaaagacccgtatccaagaaGTTTTAGAGCTAGAAATAGCAAG	FW_pWUR_ori_SpacerR
7837	TCATGACCAAAATCCCTTAAC	RV_pWUR_ori
7838	tcacgttaagggattttggtcatgaGATATAAGCAATACTGTTTTTCTA ATC	FW_*hsdR*_USflank1_ori
7839	ggaaccgttaggatatcTAATGTTGTACTTTTGCACC	RV_ *hsdR* _USflank1
7840	aagtacaacattagatatcctaaCGGTTCCTGAACTCGGTGAC	FW_ *hsdR* _Dsflank1
7841	attatcctcagctcactagcgccatGGCCGCTGCTTCACGAGT	RV_ *hsdR* _DSflank1_Cas9
7842	ATGGCGCTAGTGAGCTG	FW_pWUR_cas9
7843	ctaaaacttcttggatacgggtctttcGTATAACGGTATCCATTTTAAG AATAATCC	RV_pWUR_cas9_SpacerR
7881	AATACCAAACACGCCCTATAGC	FW_ *hsdR* _check
7882	CTTGCCCTGTTGAAATAATTAATGC	RV_ *hsdR* _check
8134	ttatgtaaacgaccatacaaGTTTTAGAGCTAGAAATAGCAAG	FW_pNW_ori_SpacerL
8135	tcacgttaagggattttggtcatgaTAAAGCTGGAAGGATGCC	FW_LDH_DS1_ori
8136	tcacgttaagggattttggtcatgaACATCAAAAGGATAGAGACAAA ATC	FW_LDH_DS2_ori
8137	ctttcttaatgctggtatgAAAGGGAGCGTTTTAAGTTAAATG	RV_LDH_DS
8138	aaacgctccctttcataccagcaTTAAGAAAGTACTTTATTCATCGT TTC	FW_LDH+US
8139	attatcctcagctcactagcgccatATGATGAAATACGCTGGC	RV_LDH+US1
8140	attatcctcagctcactagcgccatTTTGGATTTTTGGCAGTATTC	RV_LDH+US2
8141	ctaaaacttgtatggtcgtttacataaGTATAACGGTATCCATTTTAAG AATAATCC	RV_pNW_spCas9_SpacerL
8142	GCCAACGCTTCATTGTTTCC	FW_LDH_Check
8143	TCAGCCCATTGACTAGGAAG	RV_LDH_Check

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

Table S4. Results of the enzymatic assays for L-lactate production, averages of technical duplos.

	DKO 10ml	DKO 25ml	DKO 50ml	WT 10ml	WT 25ml	WT 50ml	DKO_ki 10ml	DKO_ki 25ml	DKO_ki 50ml
Lactate g/L	0.01	0.05	0.02	1.08	1.48	1.46	1.06	1.39	1.54
Lactate mM	0.09	0.60	0.24	12.04	16.47	16.21	11.76	15.40	17.07
O D600	1.28	0.66	0.16	1.10	0.74	0.64	1.18	0.86	0.74
Prod.	0.07	0.91	1.51	10.94	22.26	25.34	9.97	17.91	23.06

Abbreviations: DKO: double knockout ET 138 Δ*ldhL* Δ*sigF*; WT: wild-type ET 138; DKO_ki: *ldhL* knock-in strain ET 138 Δ*ldhL* Δ*sigF* +*ldhL*. The results prove the insertion and expression of the *ldhL* gene in the DKO_ki strain; the lactate production and the growth under anaerobic conditions are reconstituted to the wild-type levels.

SUPPLEMENTARY FIGURES

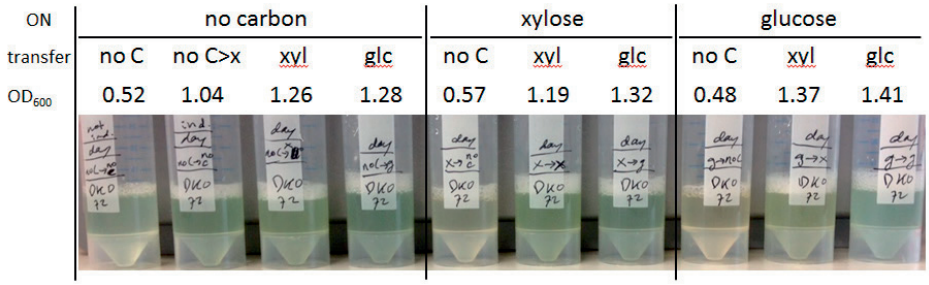


Figure S1. *lacZ*-expression controlled by P_{xynA} in ET 138 Δ*ldhL* Δ*sigF* grown in TVMY medium containing different carbon sources. Cells were grown overnight ('ON', top row) in TVMY medium either without carbon source, with xylose or with glucose. Each preculture was used to inoculate 3 different TVMY cultures either without carbon source, with xylose or with glucose. The cultures were grown for 8 h (2nd row and tubes shown). Higher LacZ expression leads to darker blue culture colour.

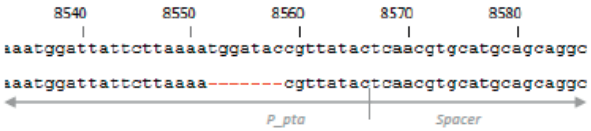


Figure S2. Sequencing results of the P_{pta} from pWUR_Cas9sp2. A 7 bp-long deletion is located right downstream of the predicted -10 region of the prom oter and expands until 8bp upstream of the spacer 2 sequence.

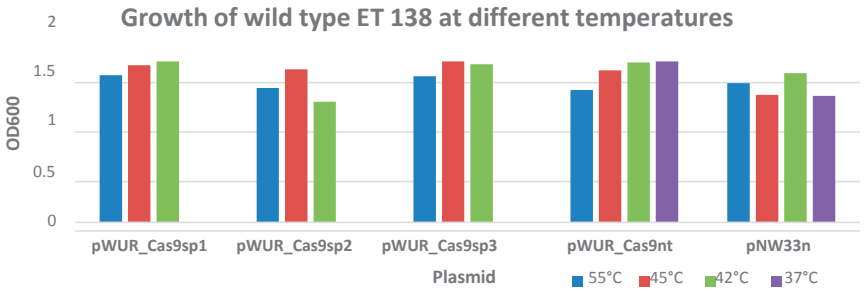


Figure S3. OD_{600} measurements from the wild type ET 138 spCas9 targeting experiment. The growth of the cells with the spCas9 targeting constructs is greatly affected at 37°C, which is not the case for the cells with the negative control constructs.

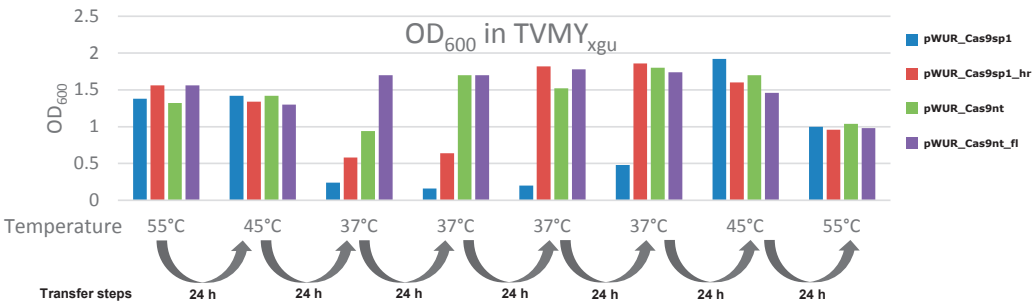


Figure S4. OD600 measurements from the 7-days long *pyrF* deletion culturing process. ET 138 Δ*ldhL* Δ*sigF* cells transformed with pWUR_Cas9sp1, pWUR_Cas9sp1_hr, pWUR_Cas9nt or pWUR_Cas9nt_hr were cultured in $TVMY_{xgu}$ medium. After growth at 55°C for 24 hours, the cultures were sequentially transferred every 24 hours to fresh media while gradually lowering the culturing temperature from 55°C to 37°C, with an intermediate transfer at 45°C. After 3 more transfers at 37°C cells were transferred back to 55°C with an intermediate transfer at 45°C. The pWURCas9_sp1 cultures at 37°C showed almost no growth, indicating efficient spCas9 targeting, while the pWUR_Cas9nt and pWUR_Cas9nt_hr control cultures grew at all the temperatures as expected for ET 138 cultures. The pWUR_Cas9_sp1hr cultures showed poor growth in the first 2 culturing steps at 37°C, but the growth was reconstituted from the 3rd culturing step at 37°C onwards to control levels, indicating the development of either sgRNA and spCas9 escape mutants, or Δ*pyrF* deletion mutants that can avoid the spCas9 targeting.

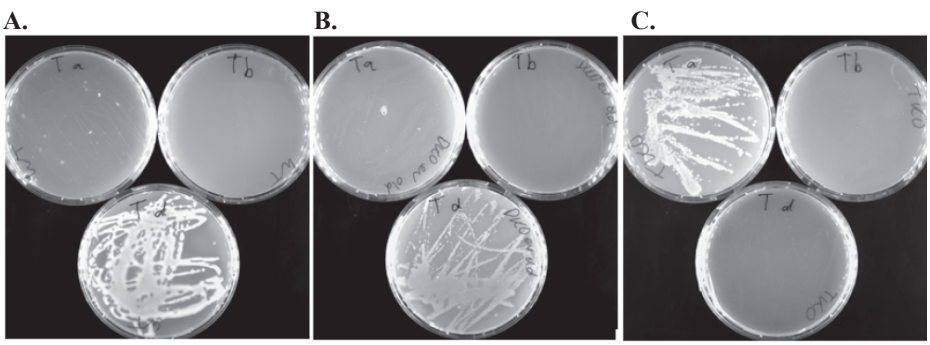


Figure S5. Phenotypic evaluation of 5-FOA sensitivity and uracil auxotrophy of ET 138 *pyrF* wild-types and mutant. Cells were grown overnight on TVMY medium with the following additions: Plates annotated as "Ta" contained: 2 g/L 5-FOA and 50 mg/L uracil; Plates annotated as "Tb" contained: 2 g/L 5-FOA and no uracil; Plates annotated as "Td" contained: no 5-FOA and no uracil. **A.** ET 138 wild-type. **B.** ET 138 Δ*ldhL* Δ*sigF*. **C.** ET 138 Δ*ldhL* Δ*sigF* Δ*pyrF*.

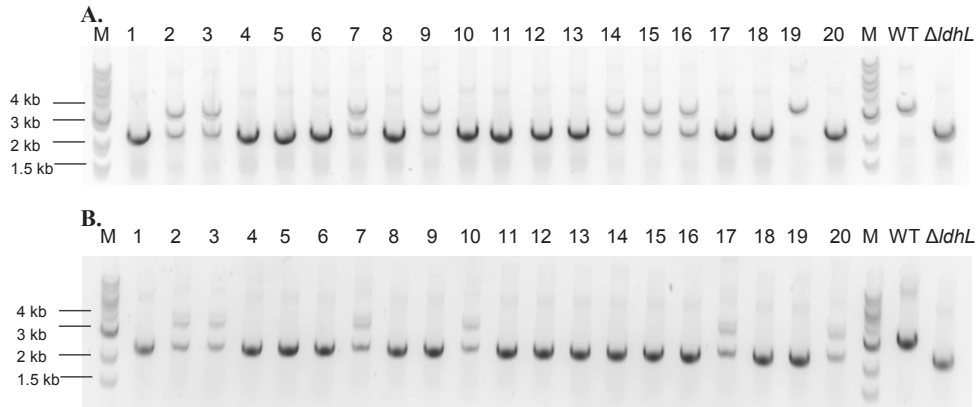


Figure S6. Agarose gel electrophoresis showing the resulting products from colony PCR on the colonies from the 3-day long *ldhL* knock-in culturing processes in LB2 medium , using ET 138 Δ*ldhL* Δ*sigF* Δ*hsdR* cells transform ed with the pWUR_Cas9spKI_hr2 vector. For the colony PCR genome specific primers BG8145 and BG8146 were used with the expected size of the PCR fragment for the knock-in mutations being 3.2 kb (equal to the size of the PCR fragment for the wild type cells). The expected size of the PCR fragment for the knock-out (non-edited) mutants was 2.3 kb. **A.** When the 45°C culturing step was performed, the process resulted in 1 out of the 20 tested colonies having the knock-in genotype, 7 colonies with mixed knock-in and wild-type genotype, and 12 colonies with the wild-type genotype. **B.** When the culturing step at 45°C was omitted none of the 20 tested colonies had the knock-in genotype, only 6 colonies had the mixed genotype while the remaining 14 colonies had the wild-type genotype.

SUPPLEMENTAL REFERENCES

1. Lopilato, J., Bortner, S., and Beckwith, J. (1986) Mutations in a new chromosomal gene of *Escherichia coli* K-12, pcnB, reduce plasmid copy number of pBR322 and its derivatives, Mol. Gen. Genet. 205, 285-290.

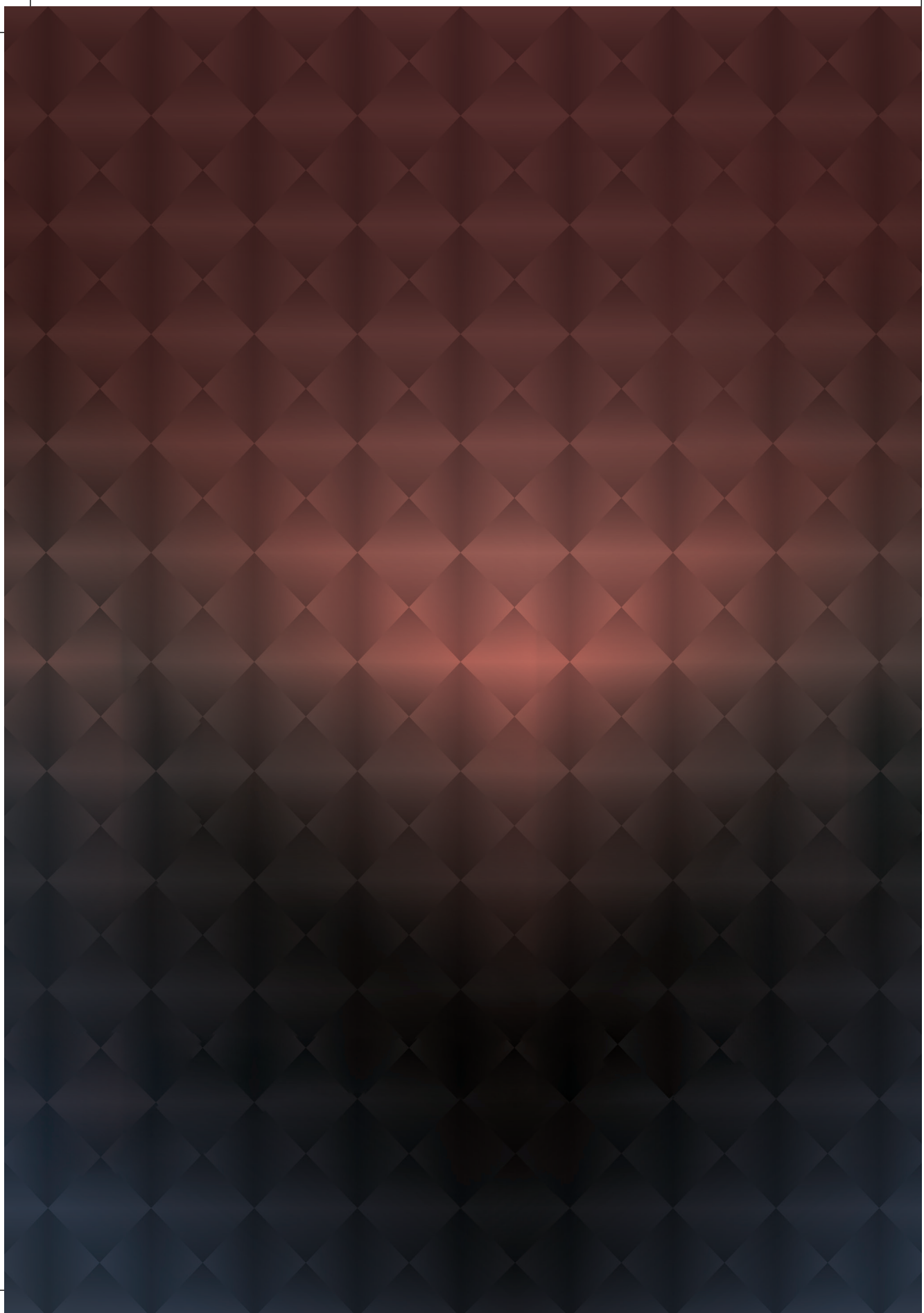
2. Bosma, E. F., van de Weijer, A. H. P., Daas, M. J. A., van der Oost, J., de Vos, W. M., and van Kranenburg, R. (2015) Isolation and screening of thermophilic Bacilli from compost for electrotransformation and fermentation: characterization of *Bacillus smithii* ET 138 as a new biocatalyst, Appl. Environ. Microbiol. 81, 1874-1883.

3. Bosma, E. F., van de Weijer, A. H. P., van der Vlist, L., de Vos, W. M., van der Oost, J., and van Kranenburg, R. (2015) Establishment of markerless gene deletion tools in thermophilic *Bacillus smithii* and construction of multiple mutant strains, Microb. Cell Fact . 14, Art.nr.99.

4. Currie, D., Herring, C., Guss, A., Olson, D., Hogsett, D., and Lynd, L. (2013) Functional heterologous expression of an engineered full length CipA from *Clostridium* thermocellum in Thermoanaerobacterium saccharolyticum, Biotechnol. Biofuels 6, 32.

5. Kovács, Á. T., van Hartskamp, M., Kuipers, O. P., and van Kranenburg, R. (2010) Genetic tool development for a new host for biotechnology, the thermotolerant bacterium Bacillus coagulans, Appl. Environ. Microbiol. 76, 4085-4088.

6. Reeve, B., Martinez-Klimova, E., de Jonghe, J., Leak, D. J., and Ellis, T. (2016) The Geobacillus plasmid set: a modular toolkit for thermophile engineering, ACS Synth. Biol. 5, 1342-1347.



CHAPTER 4

EFFICIENT CAS9-BASED GENOME EDITING OF *RHODOBACTER SPHAEROIDES* BY NATIVE HOMOLOGOUS RECOMBINATION AND NON-HOMOLOGOUS REPAIR SYSTEMS

Ioannis Mougiakos[1,#], Enrico Orsi[2,#], Mohammad Rifqi Ghiffary[1,2], Alberto de Maria[1,2], Belén Adiego Perez[1,2], Frank Fluitman[3], Servé W.M. Kengen[1], Ruud A. Weusthuis[2]*, John van der Oost[1]*

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

[2]Bioprocess Engineering, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

[3]Isobionics, Urmonderbaan 22-B 45, 6167 RD Geleen, TheNetherlands

#Contributed equally

*Corresponding authors

Chapter submitted for publication

ABSTRACT

Rhodobacter sphaeroides is a metabolically versatile bacterium that serves as a model for analysis of photosynthesis, hydrogen production and terpene biosynthesis. However, the lack of efficient tools for *R. sphaeroides* genome editing is a bottleneck for both fundamental studies and biotechnological exploitation. Here we describe the development of a highly efficient spCas9-based targeting system for *R. sphaeroides*. We combined the targeting system with plasmid-borne homologous recombination (HR) templates for knocking-out and knocking-in different genes, and for single nucleotide substitution with efficiencies of up to 100%. The *R. sphaeroides* genome encompasses genes encoding LigD and Ku, the canonical components of the non-homologous end joining (NHEJ) system. A unique feature of *R. sphaeroides* strains appears to be the presence of a cluster of two predicted Ku homologs that share low sequence similarity. We succeeded in combining spCas9 targeting with the native bacterial NHEJ system of *R. sphaeroides*. The tool resulted in survival through indel formation, and subsequent gene inactivation with efficiencies of up to 20%. Surprisingly, the NHEJ activity of *R. sphaeroides* is demonstrated to be based on the LigD but not on the predicted Ku homologs, strongly suggesting the existence of an unprecedented variant of the NHEJ system. Overall, we anticipate that the presented work will facilitate molecular research on *R. sphaeroides* and will be the starting point for further characterization of a non-canonical NHEJ system.

SIGNIFICANCE

Bacterial and archaeal genomes encompass arrays of clustered regularly interspaced short palindromic repeats (CRISPR) along with CRISPR-associated (Cas) proteins and it was recently shown that these systems constitute adaptive immune systems against viruses and mobile genetic elements. The Cas9 RNA-guided DNA-endonuclease from the type II CRISPR-Cas system of *Streptococcus pyogenes* (spCas9) has been extensively employed for the development of effective and time-efficient genome engineering tools for prokaryotes and eukaryotes. In this study, we combine the spCas9 targeting activity with the native homologous recombination (HR) and non-homologous end joining (NHEJ) mechanisms of the biotechnologically interesting bacterium *Rhodobacter sphaeroides* for the development of genome editing tools. Moreover, we used these tools to delete the NHEJ genes of *R. sphaeroides* and reveal an unprecedented NHEJ mechanism.

INTRODUCTION

The purple non-sulphur bacterium *Rhodobacter sphaeroides* is a microorganism with an extremely adaptable metabolism[1]: it is a facultative phototroph that can grow on many different carbon substrates and can respire aerobically and anaerobically with different electron acceptors[2,3]. This versatility is reflected in a number of different studies, which range from fundamental understanding of photosynthesis and quorum sensing to more applied fields like bioremediation and photoheterotrophic hydrogen or heterotrophic terpene biosynthesis[4–8]. Still, a lot of work is required to further increase our knowledge on *R. sphaeroides* metabolism and exploit its biotechnological potential.pot

The high-throughput exploration of *R. sphaeroides* metabolism requires the development of highly effective and time-efficient genome engineering tools[9]. Currently, the introduction of genomic mutations in *R. sphaeroides* is based on suicide-plasmid driven homologous recombination (HR)[10,11]. This approach allows for markerless modifications, but it is laborious and inefficient[12]; the lack of a strict counter-selection method usually leads to high rates of wild-type revertants during the curation of the suicide plasmid, especially for essential genes. Therefore, screening of mutants can be time consuming and is frequently unsuccessful[12].

A wide range of CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR associated proteins) bacterial and archaeal adaptive immune systems have been repurposed as tools for eukaryotic, prokaryotic and archaeal genome engineering[13–21]. Most of these tools exploit the RNA-guided DNA endonuclease from the type II CRISPR-Cas system of *Streptococcus pyogenes*, denoted as spCas9. As any Cas9 orthologue, spCas9 can be easily programmed to precisely introduce double stranded DNA breaks (DSB) to a selected DNA sequence, denoted as protospacer. A single, customizable guide RNA (sgRNA) molecule directs Cas9 to the protospacer via complementarity between its exchangeable 5'-end sequence, denoted as spacer, and the protospacer sequence. The only additional requirement for Cas9-based targeting is the presence of a specific, 3-8nt long motif right after the 3'-end of the selected protospacer, denoted as protospacer adjacent motif (PAM)[14,22]. All in all, the simplicity in the design and construction of a Cas9-based DNA targeting system has made it popular as the basis for numerous genome manipulation applications. A plethora of Cas9-based tools have been developed the last 5 years for prokaryotes. Cas9 can serve as an efficient counter selection

tool, when combined with plasmid-based ho-mologous recombination (HR) or recombineering; Alternative DSB repairing mechanisms (like the template independent non-homologous end joining, NHEJ) are lacking in most prokaryotes[23]. Therefore, Cas9-based DSBs after HR are lethal if introduced to the wild type genomes: unmodified cells will be eliminated from the treated population, allowing survival only for the recombined ones[23]. Additionally, substituting catalytic residues from either one or both of the Cas9 active sites leads to the construction of nickase (nCas9) and catalytically inactive (dCas9) variants which have been extensively employed for bacterial genome editing and silencing experiments[19,24]. Finally, the construction of dCas9 fusions with transcriptional activation domains[25,26], or with cytidine deaminase -for single nucleotide base editing-[27,28] have further expanded the Cas9-based bacterial genome engineering toolbox[25].

An alternative approach for introduction of knock-out mutations in bacterial genomes with Cas9 targeting relies on the exploitation of the template-independent non-homologous end joining (NHEJ) repairing mechanism; however, as indicated before, only few bacteria have been reported to encompass a functional NHEJ system[29,30]. Upon the formation of a DSB within a genome, an active NHEJ system would allow for repairing the break in an error-prone manner, randomly introducing small insertions or deletions (indel) that can inactivate a gene via open reading frame shift or premature stop codon formation[31–33]. A Cas9-NHEJ knock-out system has several advantages when compared to other systems, including reduced cloning complexity and the potential to create a library of knock-out mutants in a high throughput and cost effective manner.

In this study, we develop a highly efficient HR-Cas9 counter selection tool for *R. sphaeroides* and we implement it for generating gene deletions (knock-outs), insertions (knock-ins) and single nucleotide substitutions. Moreover, we demonstrate that *R. sphaeroides* has a functional, non-canonical NHEJ mechanism that can repair spCas9-induced DSBs.

RESULTS

SpCAS9 TARGETING IN *R. SPHAEROIDES*

The first aim of this study was to assess potential toxic effects of spCas9 expression in *R. sphaeroides* cells, as previously reported for other microbial species[37–39]. For this purpose we constructed the pBBR_Cas9_NT control vector by cloning the *spCas9* gene, codon optimized for *R. sphaeroides*, and a sgRNA expressing module with a non-targeting (NT) spacer in the *E. coli-Rhodobacter* shuttle vector pBBR1MCS2. The expression of the *spCas9* gene was under control of the P_{lac} promoter that due to the absence of the l*acIq* repressor gene in the *R. sphaeroides* genome, has constitutive transcription activity. Moreover the sgRNA expressing module was under the control of the synthetic constitutive P_{BBa_J95023} promoter. The pBBR_Cas9_NT vector in parallel with the pBBR1MCS2 control vector were conjugated in *R. sphaeroides* cells (Figure 1A). Even though the conjugation efficiency for the pBBR_Cas9_NT vector was reduced compared to the pBBR1MC2 (Figure 1B), it remained high and the drop in efficiency can most likely be attributed to the almost double size of the former vector (9566 bp) compared to the latter (5144bp). Moreover, we confirmed the transcription of the *spCas9* gene in *R. sphaeroides* by RT-PCR (Figure S1).

We further developed a spCas9-based system for efficient introduction of lethal double stranded DNA breaks (DSBs) in *R. sphaeroides*. The pBBR_Cas9_NT vector was employed for the construction of three targeting plasmids (pBBR_Cas9_sp1-3) each containing a unique spacer that corresponds to a different target sequence (protospacer) within the *upp* (uracil phosphoribosyl transferase) gene of the *R. sphaeroides* genome. It is expected that if the constructed spCas9 system is efficiently expressed in *R. sphaeroides*, the number of obtained colonies upon conjugation with the targeting plasmids is going to be substantially lower compared to the number of obtained colonies upon conjugation with the control plasmid.

The pBBR_Cas9 series of plasmids was conjugated in *R. sphaeroides* (Figure 1C). Remarkably, the conjugation efficiency for the pBBR_Cas9_sp1 and pBBR_Cas9_sp2 constructs dropped more than 3 orders of magnitude compared to the pBBR_Cas9_NT control, while for the pBBR_Cas9_sp3 the conjugation efficiency dropped only by 50% (Figure 1C). This result shows that the constructed spCas9 system is highly active in *R. sphaeroides*.

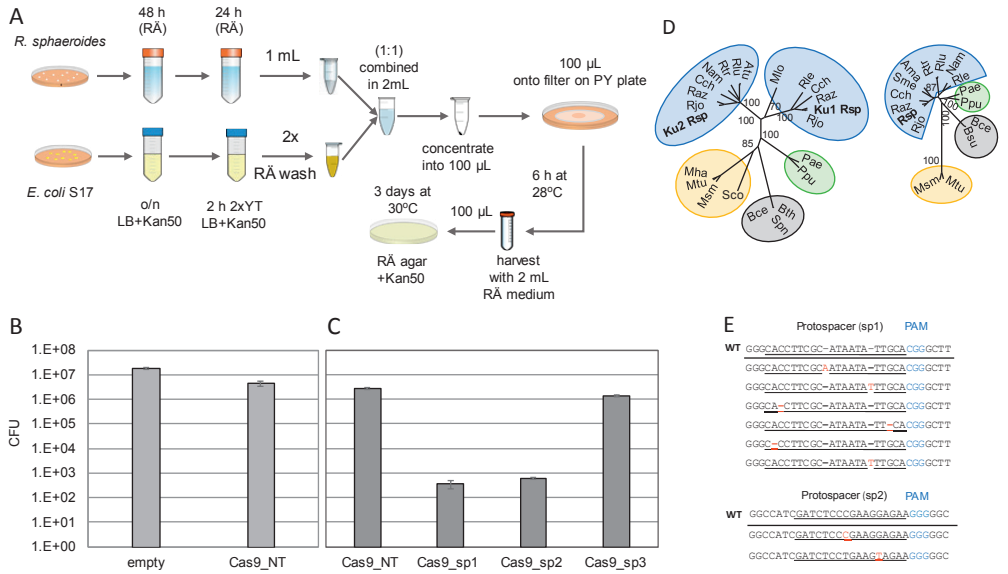


Fig. 1. Conjugation of *upp* targeting plasmids. (A) Schematic overview of the di-parental conjugation protocol used: detailed description of the protocol can be found in the Material and Methods section. (B) Colony forming units (CFUs) obtained after conjugation of pBBR_Cas9_NT in *R. sphaeroides*. The efficiency was compared to an empty pBBR1MCS2 plasmid. (C) CFU obtained after conjugation of the plasmid pBBR_Cas9 harbouring different spacers for *upp* (pBBR_Cas9_sp1-sp3); the plasmid with the non-targeting spacer (pBBR_Cas9_NT) was used as control. (D) Phylogenetic comparison of Ku (left) and LigD (right) orthologues. Colours in the background are used to cluster taxa from the same phylogenetic class: Alphaproteobacteria (blue), Gammaproteobacteria (green), Bacilli (grey) and Actinobacteria (red). The evolutionary history was inferred using the Neighbor-Joining method[51]. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the most important nodes. Evolutionary analyses were conducted in MEGA7[35]. Abbreviations: Rsp – *Rhodobacter sphaeroides,* Rjo – *Rhodobacter jorhii,* Raz – *Rhodobacter azotoformans,* Cch – *Cereibacter changlensis,* Sme - *Sinorhizobium meliloti,* Ama - *Aurantimonas manganoxydans,* Rtr - *Rhizobium tropici,* Rlu - *Rhizobium lusitanum,* Nam - *Nitrospirillum amazonense,* Rle - *Rhizobium leguminosarum,* Pae - *Pseudomonas aeruginosa,* Ppu - *Pseudomonas putida,* Bce - *Bacillus cereus,* Bsu - *Bacillus subtilis,* Mtu - *Mycobacterium tuberculosis,* Msm - *Mycobacterium smegmatis,* Mha – *Mycobacterium haemophilum,* Sco – *Streptomyces coelicolor,* Bth – *Bacillus thuringensis,* Spn – *Streptococcus pneumoniae,* Mlo – *Mesorhizobium loti,* Atu – *Agrobacterium tumefaciens*. (E) Examples of insertions or deletions (indel, red) in the protospacer loci detected by the spacers sp1 and sp2 (underlined) in the *upp* gene. The protospacer adjacent motifs (PAM) are shown in light blue.

Moreover, the observed fluctuation of targeting efficiency of different spacers is in agreement with previous reports that the targeting efficiency of spCas9 system relies heavily on the selected spacer and can substantially differentiate amongst different targets, even within the same gene[40].

REPAIR OF SpCAS9 INDUCED DSBS BY THE NATIVE *R. SPHAEROIDES* NHEJ

The majority of the prokaryotic NHEJ mechanisms consist of one ATP-dependent DNA ligase *ligD* gene and one *ku* gene, which are typically adjacent within a genome[41,42]. Surprisingly, the genome of all sequenced *R. sphaeroides* strains, including *R. sphaeroides* 2.6.5 (data not shown), encompasses a NHEJ system consisting of 1 *ligD* and 2 *ku* homologs. The latter are located next to each other within the 1st chromosome of *R. sphaeroides*, in opposing directions with their stop codons being in proximity, while the *ligD* is located within the same chromosome but approximately 1Mb upstream of the 2 *ku* homologs. Additionally, phylogenetic analysis of the *R. sphaeroides* NHEJ genes not only shows that they only cluster together with other α-proteobacteria NHEJ genes but also that the 2 *ku* homologs are phylogenetically distant (Figure 1D). The latter is also apparent by the great sequence dissimilarity of the 2 *ku* homologs (no apparent similarity on nucleotide level when aligned with BLASTn), which suggests that they cannot be the result of a duplication event.

The existence of NHEJ genes within the genome of *R. sphaeroides* motivated to further study the genomic content of the colonies that escaped the spCas9 targeting of the *upp* gene. Previous studies reported that the toxicity of 5-fluorouracil (5-FU) in bacterial species can be circumvented by the deactivation of the uracyl-phosphoribosytransferase (*upp*) gene[43]. 5-Fluorouracil (5-FU) toxicity on *R. sphaeroides* was studied by growing the bacteria on RÄ agar media with and without 5-FU (100 μg/mL). No colonies were observed on the 5-FU containing plates, indicating that, under the used conditions, the spontaneous formation of 5-FU resistant *R. sphaeroides* Δ*upp* mutants occurs in frequency less than 1 per 10^5 CFU.

The pBBR_Cas9_sp1 and pBBR_Cas9_sp2 vectors showed the highest targeting efficiency in our previous experiment and were selected for further conjugation experiments. Surviving colonies were randomly selected and streaked on RÄ_5-FU agar plates and only 7% of them were able to grow on the selection plates (Table S5). We genotyped all the colonies that survived spCas9 targeting and 5-FU selection, through colony PCR and sequencing. The results showed that all the 5-FU resistant colonies included mutations within the corresponding protospacer regions (Figure 1E). We additionally genotyped all the colonies that survived spCas9 targeting but not 5-FU selection, through colony PCR and sequencing. The results showed that all the non 5-FU resistant colonies did not encompass mutations within he upp gene and promoter regions. We further cultured and isolated plasmid from 25% of the non-mutant escape colonies; surprisingly the sequencing re-

sults showed intact spCas9 and sgRNA sequences for all of the examined plasmids, suggesting that still a small number of *R. sphaeroides* cells can survive spCas9 targeting. As bacteria often have more than a single copy of their chromosome, these survivors can result from the RecA-dependent HR repair of the formed DSB, with the use of an intact chromosomal copy of the targeted gene as template[40].

The obtained results suggested that the *R. sphaeroides* native non-homologous end joining (NHEJ) mechanism is active, repairing spCas9-induced DSBs through error prone correction. Alternatively, the spCas9 targeting mechanism acts as a selection tool in favor of cells containing random-naturally occurring mutations within the targeted protospacer regions. The second option though is considered unlikely, given the low natural rate of spontaneous Δ*upp* mutants formation under 5-FU pressure as well as the relatively high rate of Δ*upp* mutants formed upon spCas9 targeting, with mutations within the targeted protospacer regions. Nonetheless, we decided to proceed with the construction of several NHEJ mutant strains by repeating the targeting assays.

HOMOLOGOUS RECOMBINATION-SpCAS9 COUNTER SELECTION FOR GENE DELETIONS

There are multiple studies on bacterial genome editing employing i) homologous recombination (HR) via plasmid-borne templates that carry the desired modifications, and ii) Cas9-induced DSBs for the induction of the cellular HR machinery and as a counter-selection system for the unedited cells[19,23]. Previous studies reported HR activity in *R. sphaeroides*[12] and here we developed a spCas9 targeting system. Hence, we set out to develop a HR-spCas9 counter selection system for efficient genome editing in *R. sphaeroides*. As a proof of principle, we programmed the designed HR-spCas9 system to knock out the upp gene of *R. sphaeroides*, excluding the start codon and the last 12 nucleotides of the gene in order to avoid potential polar effects for neighboring genes. Two editing plasmids were constructed and tested, both containing the previously described spacer 2 for efficient upp targeting and HR templates consisted of either 500bp (pBBR_Cas9_Δupp500HR_sp2) or 1kb (pBBR_Cas9_Δupp1000HR_sp2) upstream and downstream genomic regions flanking the selected for deletion upp fragment. Two control plasmids (pBBR_Cas9_Δupp500HR_NT and pBBR_Cas9_Δupp1000HR_NT), containing the same HR templates as the editing plasmids but a non-targeting spacer, were also taken along for assessing the contribution of spCas9 to the efficiency of the tool.

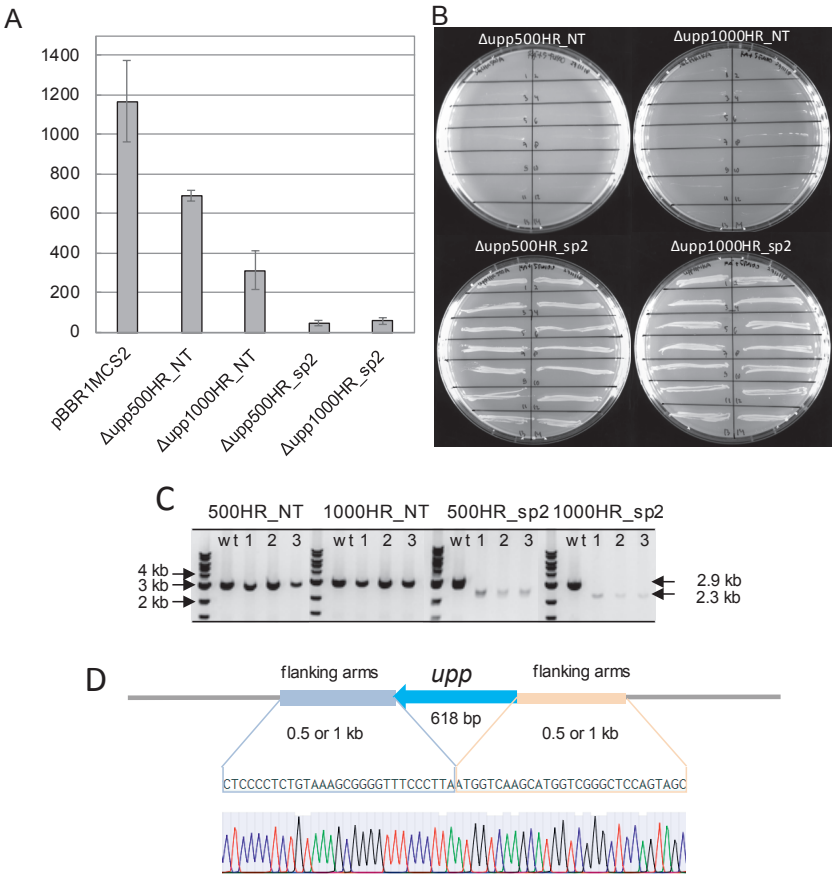


Fig. 2. Deletion of *upp* gene from *Rhodobacter sphaeroides* genome. (A) Number of colonies obtained on RÄ agar plates plated with 10^{-3} dilutions of *R. sphaeroides* conjugation mixtures with the HR control vectors pBBR_Cas9_Δupp500HR_NT and pBBR_Cas9_Δupp1000HR_NT, and HR editing vectors pBBR_Cas9_Δupp500HR_sp2 and pBBR_Cas9_Δupp1000HR_sp2. The error bars represent standard deviations from three replicate experiments. (B) Restreaks of randomly selected colonies from the above mentioned conjugations on RÄ agar plates supplemented with 5-FU. Only Δ*upp* mutants can grow on 5-FU plates. (C) Genome specific colony PCR amplification of the *upp* locus in cells conjugated with the pBBR_Cas9_Δupp500HR_NT, pBBR_Cas9_Δupp1000HR_NT, pBBR_Cas9_Δupp500HR_sp2 and pBBR_Cas9_Δupp1000HR_sp2 vectors. Amplification yields a 2992 bp product for the wild type *upp* gene and a 2374 bp product for the deleted *upp* gene. (D) Sequence verification of the desired *upp* deletion by sanger sequencing.

Upon conjugation of the above described constructs in *R. sphaeroides*, a drop was observed of at least 3 orders of magnitude in the number of surviving colonies upon conjugation with the editing constructs compared to the non-targeting control construct (Figure 2A). 14 colonies were randomly picked from each plate and streaked on RÄ agar plates supplemented with 5-FU. All selected colonies from the conjugation plates with editing constructs -regardless the size of the employed HR template- grew on the 5-FU selection plates (Figure 2B). We genotyped all the selected co-

lonies through colony PCR and sequencing, and we confirmed the desired clean deletions (Figure 2C, 2D). Meanwhile none of the selected colonies from the conjugation plates with control constructs grew on 5-FU selection plates (Figure 2B), underlining the significant contribution of spCas9 targeting to the high efficiency of the developed HR-spCas9 counter selection tool in *R. sphaeroides*.

Prerequisite for extensive metabolic engineering studies with the HR-spCas9 tool in *R. sphaeroides* is the performance of rapid iterative genome editing cycles. This demands the curing of constructed mutants from the spCas9-harbouring editing plasmids via an easy and time-effective way. For this purpose, one of the previously constructed Δ*upp* mutants was grown in LB broth medium without antibiotic for two iterative inoculation cycles of 24 hours each. After plating and incubating the final culture on LB agar plates without antibiotic for two days, 5 colonies were selected and re-streaked on LB agar plates with and without antibiotic (kanamycin). Colonies that grew only on the agar plate without antibiotic were subjected to colony PCR with plasmid specific primers and the plasmid loss was verified (Figure S2). The cured *R. sphaeroides* Δ*upp* strain was further employed for the expansion of the spCas9-based toolbox.

HOMOLOGOUS RECOMBINATION- SpCAS9 COUNTER-SELECTION FOR GENE INSERTIONS AND SINGLE NUCLEOTIDE SUBSTITUTIONS

To further expand the genome editing toolbox of *R. sphaeroides*, we proceeded with the development of a plasmid-based HR-spCas9 counter selection knock-in (KI) system. As a proof of principle, we designed a system for reinsertion of the upp gene in the genome of the previously constructed Δ*upp* mutant strain. We selected two spacers (sp4 and sp5, Table S3) for the editing plasmids, aiming to target the remaining fraction of the upp gene in the genome of the Δ*upp* mutant (Figure 3A). Moreover, we constructed HR templates containing the *upp* gene - with the "ATG" start codon substituted by the "GTG" as sequence verification marker - flanked by either the 500bp (pBBR_Cas9_KIupp500HR_sp4 and pBBR_Cas9_KIupp500HR_sp5) or the 1kb (pBBR_Cas9_KIupp1000HR_sp4 and pBBR_Cas9_KIupp1000HR_sp5) upstream and downstream genomic regions. Upon reinsertion of the *upp* gene into the Δupp genome the corresponding protospacers would be disrupted, providing resistance from the spCas9 targeting.

The four editing plasmids and the pBBR_Cas9_NT control plasmid were conjugated in the plasmid-cured *R. sphaeroides* Δ*upp* mutant. As expected, a great decrease in the number of surviving conjugants was observed for all the editing plasmids compared to the

non-targeting control plasmid (Figure 3B). The *upp* genomic region from several colonies was amplified by colony PCR with genome specific primers. The editing constructs with 500bp HR-flanking sites resulted in editing efficiencies of 5% and 10% for spacers 4 and 5 respectively (Figure S3A, S3B). Increasing the size of the HR-flanking sites increased the editing efficiency to 15% for spacer 4 and to 95% for spacer 5 (Figure S3C, S3D). All the mutants were subsequently sequence verified for the existence of the "GTG" start codon (Figure 3C). This result shows that, unlike the *upp* deletion process, the efficiency of the upp knock-in process is influenced by the size of the HR-template; HR-templates with longer flanking regions lead to higher editing efficiencies.

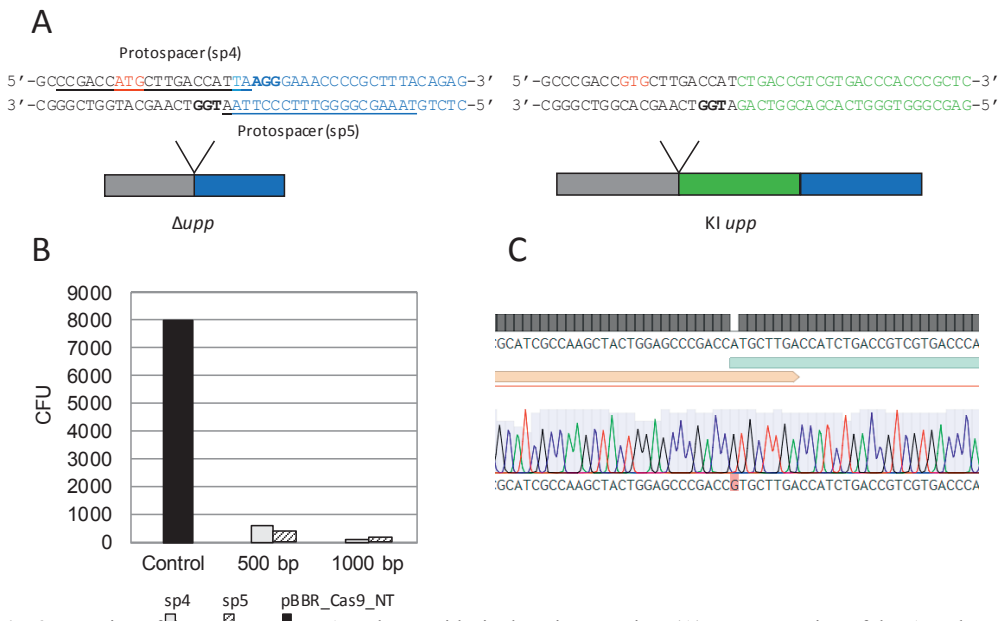


Fig. 3. Insertion of the gene *upp* in a *Δupp* locus with single point mutation. (A) Representation of the *Δupp* locus before (left) and after (right) knock-in of the *upp* gene. The protospacers recognized by spacers sp4 and sp5 are underlined and their PAM is in bold; spacers sp4 and sp5 were used for counter selection against *Δupp*. The flanking sites surrounding *Δupp* are in grey and blue, while the inserted *upp* sequence is in green; after introduction of the *upp* sequence, the protospacers sequences are disrupted, and the starting codon ATG is replaced by GTG via a 1 bp substitution (red). (B) Colony forming units (CFU) obtained after conjugation with pBBR_Cas9 plasmids (pBBR_Cas9_KIupp500HR_sp4, pBBR_Cas9_KIupp500HR_sp5, pBBR_Cas9_KIupp1000HR_sp4 and pBBR_Cas9_KIupp1000HR) harbouring different Δ*upp* targeting spacers (sp4 and sp5) and different size HR templates (500 bp and 1000 bp) for the knock-in of the *upp* gene accompanied by single nucleotide substitution; the plasmid with the non-targeting spacer (pBBR_Cas9_NT) was used as control. (C) Sequence verification of the desired *upp* insertion and single nucleotide substitution, by sanger sequencing.

NHEJ GENES DELETIONS VIA HR-CAS9 COUNTER SELECTION TOOL

The majority of the prokaryotic NHEJ mechanisms consist of an ATP-dependent DNA ligase (LigD) and a DNA-end-binding protein (Ku), encoded by the *ligD* and *ku* genes that are typically clustered[41,42]. Surprisingly, the *R. sphaeroides* genome encompasses a NHEJ system consisted of one ligD and two ku homologs. The latter are adjacent on the large chromosome of *R. sphaeroides*, in opposing directions with their stop codons being in proximity. The sequence difference of the two ku homologs renders it unlikely that they are the result of a recent duplication event. The *ligD* gene, on the other hand, is located separately at a distance of approximately 1Mb of the *ku* genes.

The aforementioned spCas9 targeting assays results suggested that the native NHEJ system of *R. sphaeroides* can repair spCas9-induced DSBs. To analyze the involvement of LigD and the Ku proteins in the NHEJ process, single and combinatorial deletion of the corresponding genes were generated and the resulting mutants were subjected and to the spCas9 targeting assays. The knockout of the *ligD*, *ku1* and *ku2* genes was carried out using the here established HR-spCas9 counter selection tool. The pBBR_Cas9_Δku1, pBBR_Cas9_Δku2 and pBBR_Cas9_ΔligD vectors were employed for the substitution of 150bp long DNA fragments, roughly located 100bp downstream the start codon of each gene, with a "stop-sequence" containing three stop codons in a row and a BamHI restriction site (5'-TGATGATGAGGATCC-3'). The double Δ*ku1*_Δ*ku2* mutant strain was constructed via a sequential process, that started with the knock out of the *ku1* gene followed by the knock out of the *ku2* gene. The employed spacers are listed in Table S3. The efficiencies of the knock out processes were lower than the efficiency of the upp deletion process, although still 11 to 30% of the screened colonies were sequence verified as clean mutants (Table S6). At least two verified mutant colonies per strain were cultured, then cured from the corresponding editing plasmids and finally employed for the following targeting experiment.

TARGETING ASSAYS WITH NHEJ MUTANTS

We aimed to evaluate the efficiency of the native NHEJ mechanism of *R. sphaeroides* to repair spCas9 induced DSBs. For this purpose we conjugated each constructed NHEJ mutant strain (Δ*ku1*, Δ*ku2*, Δ*ku1*_Δ*ku2*, Δ*ligD*) and the wild type *R. sphaeroides* with the non-targeting vector pBBR_Cas9_NT and the previously tested *upp* targeting vectors pBBR_Cas9_sp1 and pBBR_Cas9_sp2. As expected, there was a drop of at least 3 orders of magnitude in the number of surviving colonies upon conjugation with the targeting constructs compared to the control construct, for all the strains (Figure 4a). Surviving were fu

rther streaked on solid RÄ_5-FU agar plates. None of the Δ*ligD* streaked colonies survived the 5-FU selection, confirming the significance of this gene for the NHEJ activity in *R. sphaeroides* (Table S7). On the contrary, 5% to 20% of the wild type, Δ*ku1*, Δ*ku2* and Δ*ku1*_ Δ*ku2* streaked colonies survived the 5-FU selection and their upp genomic region was PCR amplified and sequenced (Table S7). The wild type colonies were shown to contain stop codon forming SNPs within the selected protospacer regions. Surprisingly, SNPs and indel mutations were observed in the protospacer regions within the genomes of most Δ*ku1*, Δ*ku2* and Δ*ku1*_ Δ*ku2* mutant strains (Table S8). Hence, the vast majority of the screened colonies that survived the 5-FU selection and contained an intact *ligD* gene copy in their genome were proven to be Δ*upp* mutants with mutations within the targeted protospacer region. Meanwhile, the spCas9 targeting system based on spacer 2 exhibited the highest efficiency (Table S7). This finding, in combination with the corresponding result from the Δ*ligD* mutants, suggests either that the Ku activity is not essential for the *R. sphaeroides* NHEJ system, or that *R. sphaeroides* uses an alternative NHEJ mechanism in which the LigD activity requires assistance of one or more elusive proteins.

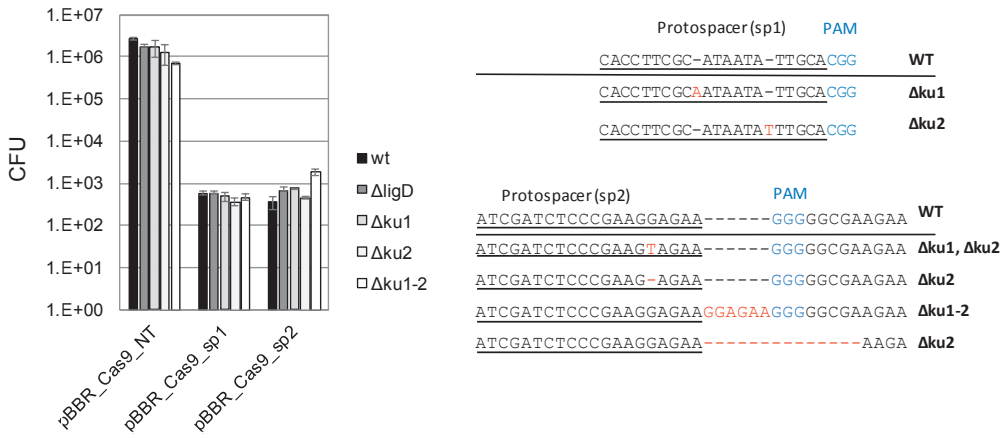


Fig. 4. Assessment of NHEJ activity in *R. sphaeroides* Δ*ligD*, Δ*ku1*, Δ*ku2* and Δ*ku1-2* mutant strains. (A) CFUs obtained after conjugation of the mutant strains with the pBBR_Cas9_sp1 and pBBR_Cas9_sp2 *upp* targeting plasmids. The plasmid with the non-targeting spacer (pBBR_Cas9_NT) was used as control. Average values of three replicates are shown, with error bars representing standard deviations. (B) Examples of Sanger sequencing results on the *upp* loci corresponding to the protospacers (underlined) recognized by spacers sp1 ans sp2; the prosotspacer adjacent motifs (PAM) are in light blue. Above the line is shown the protospacer in the wild-type genome while, below the continuous dark line, are shown examples of mutations detected in the mutant strains Δ*ku1*, Δ*ku2* and Δ*ku1-2*, in bold. The insertions or deletions (indels) detected in the protospacer regions are shown in red. The selected colonies had previously survived the 5FU selection (no survival detected in Δ*ligD* strain).

DISCUSSION

A number of studies have reported the efficient manipulation of native CRISPR-Cas systems in bacteria and archaea for genome engineering purposes[44,45]. *R. sphaeroides* encompasses a type I-E CRISPR-Cas system within its genome. We reasoned not to characterize and employ it for genome editing purposes, as we wanted to create a tool adaptable to other *Rhodobacter* species and *R. sphaeroides* strains that do not encompass the same CRISPR-Cas system as the here employed strain. Moreover, recent studies have reported the existence of a wide range of anti-CRISPR systems within bacterial genomes that deactivate native CRISPR-Cas systems[46]. The existence of such a system in the genome of *R. sphaeroides* has not been identified and reported yet, but possible existence could render the development of a native CRISPR-Cas based tool difficult.

In this study we developed a spCas9-based genome editing toolbox for *R. sphaeroides*. We demonstrated that it is possible to construct knock-out and knock-in strains, as well as strains with single nucleotide substitutions, with high efficiency and in a reduced time-frame, by employing an HR-spCas9 counter selection approach. This potential will accelerate fundamental studies, amongst others, on membrane and photosynthetic reaction centre proteins, as well as metabolic engineering studies for improved hydrogen production and terpene synthesis.

The use of NHEJ-based error prone repair of Cas9-induced DSBs is a common approach for gene knock-outs in eukaryotes[47]. In prokaryotes this is a rather rarely employed approach[40,48,49], since most bacteria appear to lack a native NHEJ system. That explains why the few available examples of successful knock-outs through the Cas9-NHEJ combination rely on the expression of heterologous NHEJ systems[40,49]. Here, we demonstrated that the native NHEJ system of *R. sphaeroides* can repair spCas9-induced breaks at easy-to-screen rates. This tool is based on a low complexity cloning approach, allowing for the simultaneous and cost-efficient construction of multiple spCas9-based targeting constructs at once. Hence, the development of an efficient Cas9-NHEJ tool in *R. sphaeroides* allows for rapid construction of an extensive library of knock-out strains. Potentially, the efficiency of the here developed tool could be further improved by the deletion of the recA gene from the *R. sphaeroides* genome. This approach was previously reported to enhance the spCas9 targeting efficiency in *E. coli*, regardless the employed spacer, and to eliminate non-mutated surviving colonies rescued by unedited genome templates via RecA-dependent HR[40].

The α-proteobacteria are unique amongst bacteria because they possess two genes encoding distantly related Ku variants. Since Ku is a component of the canonical NHEJ system(Figure 1D), we decided to analyse the potential role of these Ku homologs in the NHEJ system of *R. sphaeroides*. After performing multiple spCas9 targeting experiments with a series of the NHEJ mutants, we unexpectedly demonstrated that deletion of the two Ku homologs did not affect the efficiency of the *R. sphaeroides* NHEJ. A possible explanation could recite in the expression of a protein with Ku characteristics (DNA-end binding), such as the host-nuclease inhibitor protein Gam from the Mu phage, but a gene encoding such protein was not identified in the *R. sphaeroides* genome[50]. On the other hand, the LigD ligase was proven to be indispensable for NHEJ activity. This conclusion comes in complete contrast to what it was previously demonstrated when the only other Cas9-native NHEJ tool was reported for bacteria[48]. In that study, S*treptomyces coelicolor* was demonstrated to contain an active but incomplete NHEJ system, lacking a *ligD* homolog. It was hypothesized that the LigD function was substituted by other ligases. This explanation was supported by the random size of the created indel mutations from the employed spCas9-targeting system, that could extent up to 30kb deletions. In comparison, the targeted indel mutations generated by our system within the *R. sphaeroides* genome were restricted in length and the longest observed deletion was only 55bp long. Heterologous expression of the *Mycobacterium tuberculosis ligD* in S. *coelicolor* raised the efficiency of the system and resulted in shorter indel mutations of 1 to 3 nucleotides. All in all, these results indicate the existence of an unprecedented prokaryotic NHEJ in *R. sphaeroides*, and probably in all α-proteobacteria, that contains either an additional not identified distantly related *ku* homolog, or one or more elusive enzymes with an analogous function, i.e. to protect loose DNA ends and recruit the LigD.

In our work we demonstrated that the spCas9 endonuclease is active in the biotechnologically interesting bacterium *R. sphaeroides*. We further combined the spCas9 targeting activity with the native homologous recombination (HR) and non-homologous end joining (NHEJ) mechanisms of *R. sphaeroides* for the development of efficient genome editing tools. These tools can substantially accelerate fundamental and metabolic engineering studies on *R. sphaeroides*, as already have been achieved for other microorganisms with developed spCas9 toolbox. Finally, we used the tools for the deletion of the NHEJ genes of *R. sphaeroides* and we demonstrated the existence of an unprecedented NHEJ mechanism.

MATERIALS AND METHODS

BACTERIAL STRAINS, MEDIA AND GROWTH CONDITIONS

The strains used in this study are listed in Table 1. *E. coli* DH5a was used for cloning and routine amplification. *E. coli* S17-1 was used as vector donor for *R. sphaeroides* in diparental conjugation.

Culturing of *R. sphaeroides* was performed in RÄ minimal or LB (Luria-Bertani) medium at 250 rpm, 30ºC. Growth on solid medium was performed on Ä R supplemented with 15% w/v agar. Culturing of *E. coli* strains was performed in LB medium at 250 rpm, 37ºC. Growth on solid medium was performed on LB supplemented with 15% w/v agar. When required, kanamycin (50 µg/mL) was added for all the mentioned growing conditions.

Strain	**Description**	**Reference**
E. coli		
DH5α	. *fhuA2* Δ(*argF-lacZ*)U169 *phoA glnV44* Φ80 Δ(*lacZ*)M15 *gyrA96 recA1 relA1 endA1 thi-1 hsdR*[17] . Strain for assembly and plasmid amplification	Laboratory stock
S17-1	Host strain for transconjugation, thi pro *recA hsdR* [RP4-2Tc::Mu-Km ::Tn7] Tp^r Sm^r	Laboratory stock
R. sphaeroides		
265-9c	Wild type	ATCC 35053 derivative (52, 53)
Δ*upp*	265-9cΔ*upp*	This study
Δ*ku1*	265-9cΔ*ku1*	This study
Δ*ku2*	265-9cΔ*ku2*	This study
Δ*ligD*	265-9cΔ*ligD*	This study
Δ*ku1-2*	265-9cΔ*ku1*_Δ*ku2*	This study
*upp*KI	265-9cΔ*upp*, KI(ATG->GTG)*upp*	This study
Δ*ku1*Δ*upp*	265-9cΔ*ku1*Δ*upp*	This study
Δ*ku2*Δ*upp*	265-9cΔ*ku2*Δ*upp*	This study
Δ*ku1-2*Δ*upp*	265-9cΔ*ku1*-2Δ*upp*	This study

Table 1. List of strains used in this study

PLASMIDS CONSTRUCTION

The plasmids constructed and the primers used for cloning and sequencing are listed in Supplementary Tables S1 and S2, respectively. All primers for PCR were prepared by Integrated DNA Technologies (IDT) and all the PCR amplifications were performed using Q5 High-Fidelity DNA polymerase Master Mix 2x from New England Biolabs (NEB) according to the manufacturer's protocol. Due to the high GC content of *R. sphaeroides* genome, all PCR reactions using its genomic DNA as template were supplemented with DMSO 3% v/v. PCR fragment assemblies were performed employing the HiFi Assembly kit (NEB). Prior to plasmids assemblies, all PCR products were purified via gel extraction using the Zymoclean Gel DNA Recovery Kit (Zymo Research). Plasmid extractions were performed using the GeneJET Plasmid Miniprep kit (Thermo Fisher Scientific) and *R. sphaeroides* genome extractions were performed using the GeneJET Genomic DNA Purification Kit (Thermo Fisher Scientific).

GENERATION OF THE NON-TARGETING CAS9 PLASMID

A synthetic construct containing the sgRNA construct (promoter BBa_J95023, sgRNA scaffold and BBa_J95029 terminator) was ordered as gBlock from Integrated DNA Technologies (Leuven, Belgium). After restriction digestion with SmaI and SalI (Thermo Fisher Scientific), the sgRNA construct was ligated into the pUC19 vector and transformed into *E. coli* TOP10 cells for storage and amplification. Subsequently, the pUC19 containing the sgRNA construct and the pBBR1MCS2 *E. coli-R. sphaeroides* shuttle vector were digested with the same restriction enzymes. After digestion, the sgRNA construct and the pBBR1MCS2 backbone were ligated with T4 ligase (Thermo Fisher Scientific) yielding the pBBR1MCS2-sgRNA. The spCas9 gene, was codon-harmonized according to *R. sphaeroides* codon-usage using the Galaxy/Codon harmonizer on-line tool[34]. The codon-harmonized spCas9 was synthetically constructed by Baseclear (the Netherlands) with a 6xHis-tag fused at its C-terminus, and delivered inserted in a pUC57 vector backbone. The spCas9 gene was PCR amplified using the pUC57hspCas9 vector as template and primers BG10937/BG10938. The pBBR-sgRNA vector was linearized using primers BG10939/BG10941. The construction of the non-targeting plasmid pBBR_Cas9_NT vector was done via HiFi assembly of the two aforementioned PCR amplicons and the spCas9 amplicon was inserted downstream of the lac promoter site of the pBBR-sgRNA. The lack of *lacIq* repressor or its homolog in *R. sphaeroides* genome allowed constitutive transcription of the harmonized spCas9 gene.

GENERATION OF THE UPP TARGETING PLASMIDS

The spacers introduced in the sgRNA construct of the pBBR_Cas9 vectors are listed in the Supplementary Table S3. Using the plasmid pBBR_Cas9_NT as template, it was possible to substitute the non-targeting spacer contained in the sgRNA construct with three different targeting spacers for the gene uracyl-phosphoribosytransferase *upp* (sp1-sp3). The pBBR_Cas9_NT was used as template for PCR amplification of the backbone, the start codon and part of the spCas9 gene, employing primers BG11415/BG11416. The remaining spCas9 sequence and the promoter of the sgRNA construct were PCR amplified employing primer BG11412 in combination with primers BG11486, BG11487 and BG11411, which also contained the spacer 1, 2 and 3 sequences respectively as overhangs. The remaining part of the backbone and the constant part of the sgRNA construct were PCR amplified employing primer BG11413 in combination with primers BG11488, BG11489 and BG11414 which also contained the spacer 1, 2 and 3 sequences respectively as overhangs. 3 fragments assemblies were performed and the reaction mixtures were transformed in DH5α *E. coli* competent cells (NEB) via heat shock, for storage and plasmid amplification. Upon sequence verification, the pBBR_Cas9_NT plasmid was transformed to *E. coli* S17-1 cells.

GENERATION OF VECTORS FOR GENE KNOCK -OUTs (KO) AND GENE KNOCK -INs (KI)

Homologous recombination (HR) followed by Cas9-based counter selection was used for the KO of the *upp, ligD, ku1* and *ku2* genes, as well as the KI of the upp gene back to its original place in the *R. sphaeroides* genome combined with single nucleotide substitution.

Plasmids for *upp* K O

We started by constructing the HR control, non-targeting plasmids pBBR_Cas9_Δupp500HR_NT and pBBR_Cas9_Δupp1000HR_NT. The pBBR_Cas9_NT backbone was PCR linearized into two fragments via two PCR reactions using the primers sets BG11886/BG11182 and BG11887/BG11888. For HR purposes, the 500bp or 1kb long upstream and downstream flanking sites of the *upp* gene from the *R. sphaeroides* genome were PCR amplified employing the primer sets BG11866/BG11867 and BG11869/BG11871 for the 500bp sizes and the primer sets BG11866/BG11868 and BG11870/BG11871 for the 1kb sizes. All 4 fragments contained overhangs suitable for HiFi assembly reactions. The two pBBR_Cas9_NT fragments and the two combinations of the HR fragments were assembled generating the non-targeting, HR control plasmidspBBR_Cas9_Δupp500HR_NT and

pBBR_Cas9_Δupp1000HR_NT, which were employed as controls for the HR efficiency in *R. sphaeroides*. Upon sequence verification in DH5α *E. coli*, both plasmids were transformed to *E. coli* S17-1 cells.

We proceeded with the construction of the HR-editing, targeting plasmids pBBR_Cas9_Δupp500HR_sp2 and pBBR_Cas9_Δupp1000HR_sp2. The previously mentioned couples of primer sets were used for the generation of the 500bp and 1kb HR flanking sites. The pBBR_Cas9_NT vector was employed as template for the PCR amplification of the conserved backbone region included within the amplification with the BG11887/BG11888 primer set. In addition, the spacer sp2 was included in the corresponding position employing the primer sets BG11182/BG11487 and BG11489/BG11886 for the amplification of the spCas9 gene and the sgRNA module respectively. Then, 5 fragments were assembled to generate pBBR_Cas9_Δupp500HR_sp2 and pBBR_Cas9_Δupp1000HR_sp2. Upon sequence verification in *E. coli* DH5α, both plasmids were transformed to *E. coli* S17-1 cells.

Plasmids for NHEJ genes KO

For the selective deletion of *ku1, ku2* and *ligD*, targeting plasmids were generated with a similar approach used for assembling the *upp* targeting plasmids. The primers set BG11887/ BG11888 was used for amplifying the conserved backbone sequence of pBBR_Cas9_NT. The 1000 bp flanking sites (with overhangs) for *ligD, ku1* and *ku2* were obtained with the primers sets BG12363/BG12366 and BG12365/BG12364 (*ligD*), BG12360/BG12359 and BG12361/ BG12362 (*ku1*), BG12357/BG12356 and BG12355/BG12358 (*ku2*). The backbone interspace included between the sgRNA (listed in Supplementary Table 3) and the flanking sites of the target genes was amplified from pBBR_Cas9_NT using the primer sets BG11886/BG12340 (*ligD*), BG11886/BG12336 (*ku1*) and BG11886/BG12339 (*ku2*). Similarly, from the same pBBR_Cas9_NT template, the spCas9 sequences including the sgRNA with the spacer of interest were obtained combining primers BG11182/BG12325 (*ligD*), BG11182/BG12333 (*ku1*) and BG11182/BG12330 (*ku2*). The 5 fragments assembly in *E. coli* DH5α, generated the plasmids pBBR_Cas9_Δku1, pBBR_Cas9_Δku2 and pBBR_Cas9_ΔligD. Upon sequence verification, the plasmids were transformed in *E. coli* S17.

Plasmids for *upp* K I

Also here we employed the aforementioned strategy as for generating the *upp* KO plasmids. The pBBR_Cas9_NT plasmid was used as template for the amplification of the fragments containing the backbone (primer set BG11887/BG11888), spCas9 and the targeting sgRNAs.

For the spacer sp4, the fragments including spCas9 and the sgRNA were generated via the amplification with primers sets BG11182/BG12907 and BG11186/BG12908, respectively. The spacer sp5 was included in the fragments for spCas9 and sgRNA amplifications using primers set BG11182/BG12909 and BG11186/BG12910, respectively. Additionally, the templates used for upp insertion included within two flanking sites of 500 bp or 1000 bp around upp, both amplified from *R. sphaeroides* genome. For the 500 bp flanking sites, genomic amplification was performed using the primers pairs BG12347/ BG11867 and BG11869/BG12348; for the 1000 bp ones, primers sets BG12347/BG11868 and BG11870/BG12348 were employed. The primers BG12347 and BG12348 introduced the single point mutation in the start codon, which changed from ATG to GTG. Assembly with 5 fragments per plasmid allowed the generation and replication in *E. coli* DH5α of four plasmids:pBBR_Cas9_KIupp500HR_sp4,pBBR_Cas9_KIupp500HR_sp5,pBBR_Cas9_KIupp1000HR _sp4 and pBBR_Cas9_KIupp1000HR_sp5. Once their sequence was verified, the plasmids were transformed to *E.coli* S17.

Diparental conjugation and plasmid curation

R. sphaeroides was inoculated from glycerol stock in 10 mL RÄ medium and incubated for 2 days at 30°C and 200 rpm prior to conjugation. After 48 hours, 50 µL of culture were transferred to fresh 10 mL RÄ medium, incubated for 24 h under the same conditions, and harvested when the OD at 600nm was approximately 3. *E. coli* S17-1 harbouring the desired plasmid was grown overnight in LB supplemented with kanamycin at 37°C and 200 RPM. Then, the overnight culture was transferred to fresh 2xYT media with antibiotic and grown for 2 h until the OD at 600nm was approximately 1. The cell suspension was harvested and washed twice with 1 mL of RÄ medium, mixed with *R. sphaeroides* culture at a 1:1 ratio (1 mL each) and centrifuged for 1 minute (maximum speed). The pellet was concentrated by suspension in 100 µL of RÄ medium, which was transferred to a sterile 0.22 µm 47 mm diameter nitrocellulose filter disc (Sigma-Aldrich) on a PY agar plate and incubated for 6 hours at 28°C. The conjugation mixture was harvested and re-suspended in 2 mL of RÄ medium. 100 µL of diluted culture was spread on RÄ agar plates containing kanamycin and incubated at 30°C for 3 days until colonies appeared. Purification of *R. sphaeroides* conjugants from *E. coli* S17-1 donors was performed by subsequent transfer of *R. sphaeroides* colonies to LB agar medium supplemented with kanamycin until no *E. coli* colonies appeared. Finally, a colony PCR was performed using primer set P3/P4 to verify that *E. coli* was not present in the culture.

Plasmid curing from positive *R. sphaeroides* mutants was carried out by cultivating the mutants in 5 mL of RÄ medium without antibiotic for 24 h, twice. The cells from the second cultivation were then diluted and spread on RÄ agar plates. The individual colonies were streaked onto LB agar with and without kanamycin. Those cells that grew on LB without antibiotic but were unable to grow in the presence of antibiotic were considered to have lost the plasmid. Final verification of plasmid curing was conducted by performing PCR with primers BG10937/BG10938, which are specific for the Cas9 gene.

RNA EXTRACTIO N AND REVERSETRANSCRIPTASE PCR (RT-PCR)

To check expression of the spCas9 gene under the lac promoter, total RNA of *R. sphaeroides* conjugated with the pBBR_Cas9_NT plasmid was extracted employing the RNAeasy mini kit (Qiagen, Germany) according to the manufacturer's protocol and treated with DNAseI (NEB) to remove genomic DNA contamination in the sample. The SuperScript III Reverse Transcriptase kit (Invitrogen) was used for RT-PCR. When the first strand of cDNA was synthesized, the primers BG11112 and BG11115 were used to verify transcription activity of Cas9 gene using standard PCR.

5-FU SCREENING

For the 5-fluorouracil (5-FU) screening process we prepared RÄ agar plates supplemented with 5-FU up to 100 μg/ml final concentration, using a 50 mg/mL stock solution prepared in dimethyl sulfoxide (DMSO). Colonies obtained from conjugation were picked and streaked on RÄ_5-FU plates with kanamycin. Survival colonies were isolated and their upp locus was then PCR amplified and sequenced.

PHYLOGENETC ANALYSIS OF THE AMINO ACID SEQUENCES OF KU1, KU2 AND LIGD

From the sequenced genome of *R. sphaeroides* 265-9c, the amino acid sequences of the NHEJ proteins Ku1, Ku2 and LigD were compared to the protein sequences deposited on NCBI using the protein BLAST tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The top hits from each alignment were used to build phylogenetic trees for the two Ku homologs and LigD, respectively. For each tree, the sequences were aligned with already characterized prokaryotic orthologous belonging to other bacterial classes. The trees were built using MEGA7[35] employing the multiple sequence alignment program CulstalW[36] and confirmed for their robustness using a bootstrapping value of 1000 repetitions. Table S4 contains the species names and the deposited accession numbers of the proteins used to build the trees.

REFERENCES

1. Imam S, Noguera DR, Donohue TJ (2013) Global insights into energetic and metabolic networks in *Rhodobacter sphaeroides* . BMC Syst Biol 7(1):89.
2. Tabita RF (1995) The biochemistry and metabdic regulation of carbon metabolism and CO2 fixation in purple bacteria. ed Blankenship R.E., Madigan M.T. BCE (Springer, Dordrecht) doi:10.1007/0-306-479540_41.
3. Zannoni D, Schoepp-Cothenet B, Hosler J (2009) Respiration and Respiratory Complexes. The Purple Phototrophic Bacteria (Springer), pp 537–561.
4. Mackenzie C, et al. (2007) Postgenomic adventures with *Rhodobacter sphaeroides*. Annu Rev Microbiol 61:283–307.
5. Imam S, Noguera DR, Donohue TJ (2014) Global Analysis of Photosynthesis Transcriptional Regulatory Networks. PLoS Genet 10(12). doi:10.1371/journal.pgen.1004837.
6. Koku H, Eroğlu I, Gündüz U, Yücel M (2002) Aspects of the meta bolism of hydrogen production by *Rhodobacter sphaeroides* . Int J Hydrogen Energy 27(11–12):1315–1329.
7. Lu W, et al. (2015) Identification and elimination of metabolic bottlenecks int he quinone modification pathway for enhanced coenzyme Q10 production in *Rhodobacter sphaeroides* . Metab Eng 29:208–216.
8. Bai HJ, Zhang ZM, Yang GE, Li BZ (2008) Bioremediation of cadmium by growing *Rhodobacter sphaeroides*: Kinetic characteristic and mechanism studies. Bioresour Technol 99(16):7716–7722.
9. Nybo SE, Khan NE, Woolston BM, Curtis WR (2015) Metabolic engineering in chemolitho autotrophic hosts for the production of fuels and chemicals. Metab Eng 30:105–120.
10. Porter, S. L., Wadhams, G. H., & Armitage JP (2007) *In Vivo* and *In Vitro* Analysis of the *Rhodobacter sphaeroides* Chemotaxis Signaling Complexes. Methods Enzymol 423(07):39 2–413.
11. Jaschke PR, Saer RG, Noll S, Beatty JT (2011) Modification of the genome of R. sphaeroides and construction of synthetic operons (Elsevier Inc.). 1st Ed. doi:10.1016/B978-0-12-385075-100023-8.
12. Jaschke PR, Saer RG, Noll S, Beatty JT (2011) Chapter twenty-three – Modification of the Genome of *Rhodobacter sphaeroides* and Construction of Synthetic Operons. Methods in Enzymology, pp 519–538.
13. Mougiakos I, et al. (2017) Efficient genome editing of a facultative thermophile using the mesophilic spCas9. ACS Synth Biol:acssynbio.6b00339.
14. Mougiakos I, et al. (2017) Characterizing a thermostable Cas9 for bacterial genome editing and silencing. Nat Commun 8(1). doi:10.1038/s41467-017-01591-4.
15. Jinek M, et al. (2012) A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptative Bacterial Immunity. Science (80-) 337(August):816–822.
16. Hsu PD, Lander ES, Zhang F (2014) Development and applications of CRISPR-Cas9 for genome engineering. Cell 157(6):1262–1278.
17. Selle K, Barrangou R (2015) Harnessing CRISPR-Cas systems for bacterial genome editing. Trends Microbiol 23(4):225–232.
18. Doudna JA, Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. Science (80-) 346(6213). doi:10.1126/science.1258096.
19. Mougiakos I, et al. (2018) Hijacking CRISPR-Cas for high-throughput bacterial metabolic engineering: advances and prospects. Curr Opin Biotechnol 50:146–157.
20. Gophna U, Allers T, Marchfelder A (2017) Finally, Archaea Get Their CRISPR-Cas Toolbox. Trends Microbiol 25(6):430–432.
21. Stachler AE, Marchfelder A (2016) Gene repression in haloarchaea using the CRISPR (Clustered regularly interspaced short palindromic repeats)-Cas I-B system. J Biol Chem 291(29):15226–15242.
22. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol 31(3):233–239.
23. Mougiakos I, Bosma EF, de Vos WM, van Kranenburg R, van der Oost J (2016) Next Generation Prokaryotic Engineering: The CRISPR-Cas Toolkit. Trends Biotechnol 34(7):575–587.
24. Arazoe T, Kondo A, Nishida K (2018) Targeted Nucleotide Editing Technologies for Microbial Metabolic Engineering. Biotechnol J 1700596:1–12.
25. Bikard D, et al. (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res 41(15):7429–7437.
26. Dong C, Fontana J, Patel A, Carothers JM, Zalatan JG (2018) Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. Nat Commun 9(2489). doi:10.1038/s41467-018-04901-6.
27. Zheng K, et al. (2018) Highly efficient base editing in bacteria using a Cas9-cytidine deaminase fusion. Commun Biol 1(1):32.
28. Banno S, Nishida K, Arazoe T, Mitsunobu H, Kondo A (2018) Deaminase-mediated multiplex genome editing in *Escherichia coli*. Nat Microbiol 3(4):423–429.
29. Bowater R, Doherty AJ (2006) Making ends meet: Repairing breaks in bacterial DNA by non-homologous end-joining. PLoS Genet 2(2):93–99.

30. Pitcher RS, Brissett NC, Doherty AJ (2007) Nonhomologous End-Joining in Bacteria: A Microbial Perspective. Annu Rev Microbiol 61(1):259–282.
31. Cho SW, Kim S, Kim JM, Kim J (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol 31(3):230–232.
32. Sakuma T, Nishikawa A, Kume S, Chayama K, Yamamoto T (2014) Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. Sci Rep 4:4–9.
33. Cong L, et al. (2013) Multiplex Genome Engineering Using CRISPR/Cas Systems. Science (80-) 339(February):819–824.
34. Claassens NJ, et al. (2017) Improving heterologous membrane protein production in *Escherichia coli* by combining transcriptional tuning and codon usage algorithms. PLoS One 12(9). doi:10.1371/journal.pone.0184355.
35. Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33(7):1870–1874.
36. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22(22):4673–4680.
37. Li H, et al. (2016) CRISPR-Cas9 for the genome engineering of cyanobacteria and succinate production. Metab Eng 38(June):293–302.
38. Wendt KE, Ungerer J, Cobb RE, Zhao H, Pakrasi HB (2016) CRISPR/Cas9 mediated targeted mutagenesis of the fast growing cyanobacterium Synechococcus elongatus UTEX 2973. Microb Cell Fact 15(1):1–8.
39. Jiang Y, et al. (2017) CRISPR-Cpf1 assisted genome editing of *Corynebacterium glutamicum*. Nat Commun 8(May):1–11.
40. Cui L, Bikard D (2016) Consequences of Cas9 cleavage in the chromosome of *Escherichia coli*. Nucleic Acids Res 44(9):4243–4251.
41. Y AJD, Jackson SP (2001) Identification of bacterial homologues of the Ku DNA repair proteins. FEBS Lett 500:186–188.
42. Aravind L, Koonin E V. (2001) 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 Res 11(8):1365–1374.
43. Graf N, Altenbuchner J (2011) Development of a method for markerless gene deletion in Pseudomonas putida. Appl Environ Microbiol 77(15):5549–5552.
44. Li Y, et al. (2015) Harnessing Type I and Type III CRISPR-Cas systems for genome editing. Nucleic Acids Res 44(4). doi:10.1093/nar/gkv1044.
45. Pyne ME, Bruder MR, Moo-Young M, Chung DA, Chou CP (2016) Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *Clostridium*. Sci Rep 6(May):1–15.
46. Stanley SY, Maxwell KL (2018) Phage-Encoded Anti-CRISPR Defenses. Annu Rev Genet 52(14):1–20.
47. Komor AC, Badran AH, Liu DR (2017) Erratum: CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes (Cell (2017) 168(2) (20–36)(S0092867416314659)(10.1016/j.cell.2016.10.044)). Cell 169(3):559.
48. Tong Y, Charusanti P, Zhang L, Weber T, Lee SY (2015) CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. ACS Synth Biol 4(9):1020–1029.
49. Su T, et al. (2016) A CRISPR-Cas9 assisted non-homologous end-joining strategy for one-step engineering of bacterial genome. Sci Rep 6(November):1–11.
50. Aniukwu J, Glickman MS, Shuman S (2008) The pathways and outcomes of mycobacterial NHEJ depend on the structure of the broken DNA ends. Genes Dev 22(4):512–527.
51. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4):406–425.
52. Markus Huembelin, Matrinus Julius Beekwilder JGTK (2014) Rhodobacter for preparing terpenoids.
53. Beekwilder J, et al. (2014) Valencene synthase from the heartwood of Nootka cypress (Callitropsis nootkatensis) for biotechnological production of valencene. Plant Biotechnol J 12(2):174–182

SUPPLEMENTARY INFO RMATIO N

SUPPLEMENTARY FIG URES

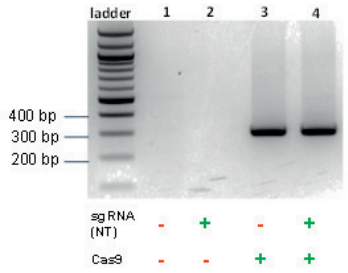


Fig. S1. Reverse Transcriptase PCR (RT-PCR) to assess transcription of Cas9 under the constitutive lac promoter. The expected size of the cDNA amplicon for Cas9 is of 307bp, and it was obtained b y sonication of *R. sphaeroides* cells suspension conjugated with pBBR1MCS2 harbouring: no additional features (well 1), non-targeting sgRNA (pBBR_NT, well 2), harmonized Cas9 (pBBR_Cas9, well 3) and harmonized Cas9 plus NT-sgRNA (pBBR_Cas9_NT, well 4). The primers set used was BG11112/BG11115.

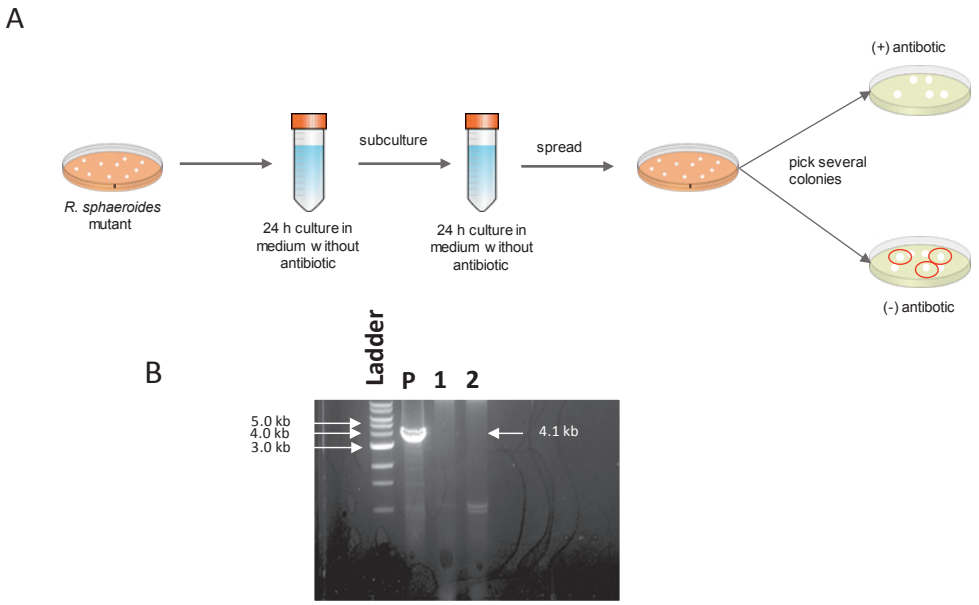


Figure S2. Curing from the plasm id after successful conjugation. A) Schematic overview of the workflow: mutants from *R. sphaeroides* were transferred twice to liquid medium without antibiotic and incubated over-night. Subsequently, the liquid culture was spread on LB plates. Once colonies form ed, they were transferred to LB plates with and without antibiotic. B) The colonies that grew only on the plate without antibiotic were PCR amplified with primers annealing to the Cas9 gene in the vector backbone using primers set BG10937/BG10938 (P: plasmid control, 1 and 2: curated *R. sphaeroides* mutants).

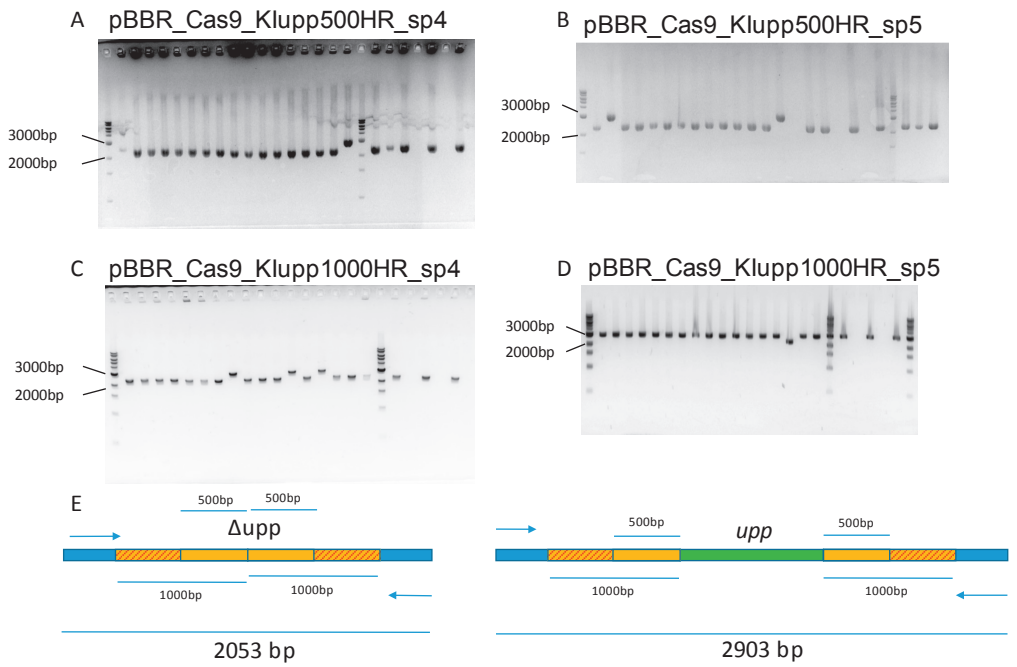


Figure S3. Figure S4. 1% agarose electrophoresis showing the colony PCRs after the conjugation of the *Δupp* strain of *R. sphaeroides* with the plasmids for upp integration. A) pBBR_Cas9_KIupp500HR_sp4. B) pBBR_Cas9_KIupp500HR_sp5. C) pBBR_Cas9_KIupp1000HR_sp4. D) pBBR_Cas9_KIupp1000HR_sp5. E) Schematic representation of the amplicons generated via colony PCR: the Δ*upp* gene leads to an amplicon size of 2053bp (left), while a restored *upp* gene (green) generates an amplicon of the size of 2903bp. The homology regions (500bp and 1000bp) are highlighted in yellow and in yellow with diagonal red bars, respectively.

SUPPLEMENTARY TABLES

Table S1. List of plasmids used in this study

Plasmid	Description	Source
pUC57hCas9	pUC57 vector containing DNA encoding the Cas9 gene harmonized for expression in *Rhodobacter sphaeroides (*hCas9)	Baseclear
pBBR1MCS2-sgRNA	Includes a synthetic construction between the SmaI and the SalI restriction sites. This synthetic construction contains the promoter BBa_J95023, two BsaI restriction sites, the sgRNA the terminator BBa_J95029	Baseclear, Huo *et. al*, 2011, this study
pBBR1MCS2_Cas9_NT	Derivative of pBBR1MCS2-sgRNA. Broad host range plasmid (KanR) containing hCas9 under constitutive lac promoter and non-targeting sgRNA under the promoter BBa_J95023 and the terminator BBa_J95029	this study
pBBR1MCS2_Cas9_sp1	derivative of pBBR1MCS2_Cas9_NT containing sgRNA with spacer 1 for targeting the gene *upp*	this study
pBBR1MCS2_Cas9_sp2	derivative of pBBR1MCS2_Cas9_NT containing sgRNA with spacer 2 for targeting the gene *upp*	this study
pBBR1MCS2_Cas9_sp3	derivative of pBBR1MCS2_Cas9_NT containing sgRNA with spacer 3 for targeting the gene *upp*	this study
pBBR_Cas9_Δupp500HR_NT	derivative of pBBR1MCS2_Cas9_NT containing non-targeting sgRNA and 2 flanking sites of 500 bp each for homologous recombination of the *upp* gene	this study
pBBR_Cas9_Δupp1000HR_NT	derivative of pBBR1MCS2_Cas9_NT containing non-targeting sgRNA and 2 flanking sites of 1000 bp each for homologous recombination of the *upp* gene	this study
pBBR_Cas9_Δupp500HR_sp2	derivative of pBBR1MCS2_Cas9_NT containing sgRNA with spacer 2 and 2 flanking sites of 500 bp each for counter selection of homologous recombination of the *upp* gene	this study
pBBR_Cas9_Δupp1000HR_sp2	derivative of pBBR1MCS2_Cas9_NT containing sgRNA with spacer 2 and 2 flanking sites of 1000 bp	this study

	each for conter selection of homologous recombination in the *upp* gene	
pBBR_Cas9_KIupp500HR_sp4	derivative of pBBR1MCS2_Cas9_NT with 500 bp flanking sites for knock-in of *upp* in a *Δupp* strain. It also contains sgRNA with spacer 4 for targeting the genomic locus of the *Δupp* previously obtained via precise HR-CRISPR-Cas.	this study
pBBR_Cas9_KIupp500HR_sp5	derivative of pBBR1MCS2_Cas9_NT with 500 bp flanking sites for knock-in of *upp* in a *Δupp* strain. It also contains sgRNA with spacer 5 for targeting the genomic locus of the *Δupp* previously obtained via precise HR-CRISPR-Cas.	this study
pBBR_Cas9_KIupp1000HR_sp4	derivative of pBBR1MCS2_Cas9_NT with 1000 bp flanking sites for knock-in of *upp* in a *Δupp* strain. It also contains sgRNA with spacer 4 for targeting the genomic locus of the *Δupp* previously obtained via precise HR-CRISPR-Cas.	this study
Plasmid	**Description**	**Source**
pBBR_Cas9_KIupp1000HR_sp5	derivative of pBBR1MCS2_Cas9_NT with 1000 bp flanking sites for knock-in of *upp* in a *Δupp* strain. It also contains sgRNA with spacer 5 for targeting the genomic locus of the *Δupp* previously obtained via precise HR-CRISPR-Cas.	this study
pBBR_Cas9_Δku1	derivative of pBBR1MCS2_Cas9_NT containing sgRNA with spacer targeting *ku1* and 2 flanking sites of 1000 bp each for counter selection of homologous recombination of the *ku1* gene	this study
pBBR_Cas9_Δku2	derivative of pBBR1MCS2_Cas9_NT containing sgRNA with spacer targeting *ku2* and 2 flanking sites of 1000 bp each for counter selection of homologous recombination of the *ku2* gene	this study
pBBR_Cas9_ΔligD	derivative of pBBR1MCS2_Cas9_NT containing sgRNA with spacer targeting *ligD* and 2 flanking sites of 1000 bp each for counter selection of homologous recombination of the *ligD* gene	this study

Table S2. List of primers used in this study

	Oligo	Sequence	Description
Ubiquitin PCR	P3-Ec_ubi1	GATAGCGTCCTAACGGCGTT	FW for ubiquitin PCR to check post-conjugational *E. coli* contamination. Anneals to *E .coli* genome
	P4-Ec_ubi2	GGGTTTGTCGAGCAGAATGA	RV for ubiquitin PCR to check post-conjugational *E. coli* contamination. Anneals to *E. coli* genome
	P5-Rs_ubi105F	GAACCCCTGCCGGCTGGATTACA	FW for ubiquitin PCR to check post-conjugational *E. coli* contamination. Anneals to *R. sphaeroides* genome
	P6-Rs_ubi501R	CATGGCGAAGCTCTTGGGATTG	RV for ubiquitin PCR to check post-conjugational *E. coli* contamination. Anneals to *R. sphaeroides* genome
RT-PCR Cas9	BG11112	GCCCGTTCTCGAGTTCG	FW to check Cas9 transcription in pBBR_Cas9_NT
	BG11115	GAATCGATCCTCCCCAAGAG	RV to check Cas9 transcription in pBBR_Cas9_NT
Non-targeting: pBBR_Cas9_NT	BG10937	GACAAAAAGTACTCGATTGGGC	FW for amplifying hCas9 from pUC57
	BG10938	CTAATGGTGATGGTGATGGTGATC	RV for amplifying hCas9 from pUC57
	BG10939	GCCCAATCGAGTACTTTTTGTCCATAGCTGTTTCCTGTGTGAAATTG	FW for linearizing pBBR1MCS2-sgRNA with homologous overhang for Gibson assembly with hCas9
	BG10941	TCACCATCACCATCACCATTAGGAGGTCGACGGTGATTGATTGAG	RV for linearizing pBBR1MCS2-sgRNA with homologous overhang for Gibson assembly with hCas9
Targeting *upp:* pBBR_Cas9_sp1-3	BG11415	ACCTTAACAAGCTCGTCGACGAC	FW for linearizing the conserved parts of *hCas9* and backbone of pBBR_Cas9_NT
	BG11416	GCTATCTGGACAAGGGAAAACGC	RV for linearizing the conserved parts of *hCas9* and backbone of pBBR_Cas9_NT
	BG11412	GTCGTCGACGAGCTTGTTAAGGT	RV conserved for amplifying part of the *hCas9* gene and include the spacer of interest for *upp* in the sgRNA (provided by FW)
	BG11411	CGGCGTCTGGATCGTCGTCGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	FW to combine with BG11412 for amplifying the sgRNA

			including the spacer 3 for *upp* and part of *hCas9* gene
	BG11486	CACCTTCGCATAATATTGCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	FW to combine with BG11412 for amplifying the sgRNA including the spacer 1 for *upp* and part of *hCas9* gene
	BG11487	ATCGATCTCCTGAAGGAGAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	FW to combine with BG11412 for amplifying the sgRNA including the spacer 2 for *upp* and part of *hCas9* gene
	BG11413	GCGTTTTCCCTTGTCCAGATAGC	FW conserved for amplifying the conserved part of the backbone and include the spacer of interest for *upp* in the sgRNA (provided by RV)
	BG11414	CGACGACGATCCAGACGCCGAACCAGCGATCCCGTCCG	RV to combine with BG11413 for amplifying the sgRNA including the spacer 3 for *upp* and part of *pBBR1MCS2-sgRNA backbone*
	BG11488	TGCAATATTATGCGAAGGTGAACCAGCGATCCCGTCCG	RV to combine with BG11413 for amplifying the sgRNA including the spacer 1 for *upp* and part of *pBBR1MCS2-sgRNA backbone*
	BG11489	TTCTCCTTCAGGAGATCGATAACCAGCGATCCCGTCCG	RV to combine with BG11413 for amplifying the sgRNA including the spacer 2 for *upp* and part of *pBBR1MCS2-sgRNA backbone*
HR plasmids for *upp* knock-out	BG11886	gtgcgggcctcttcgctatta	FW to amplify pBBR_Cas9NT including the sgRNA with non-targeting spacer and the *hCas9* gene
	BG11182	CCATGTCGGCAGAATGCTTAATG	RV to amplify pBBR_Cas9NT including the sgRNA with non-targeting spacer and the *hCas9* gene
	BG11887	CATTAAGCATTCTGCCGACATGG	FW to amplify pBBR_Cas9NT including the backbone
	BG11888	gcctgaatggcgaatggaaattgtaa	RV to amplify pBBR_Cas9NT including the backbone
	BG11866	CTCTGTAAAGCGGGGTTTCCCTTAATGGTCAAGCATGGTCGGGC	FW for 500/1000 bp amplification of the first homology region of *upp*. It contains overhangs with BG11871
	BG11867	taatagcgaagaggcccgcacCGACGAGACGGTGGGCAA	RV for 500 bp amplification of the homology region of *upp*.

	BG11869	ttacaatttccattcgccattcaggcTTTCCACGTCTTCGCCGG	FW for 500 bp amplification of the second homology region of *upp*.
	BG11871	TAAGGGAAACCCCGCTTTACAGAG	RV for 500/1000 bp amplification of the second homology region of *upp*. It contains overhangs with BG11866
	BG11868	taatagcgaagaggcccgcacCGCATCTGGGCCTCTCCAA	RV for 1000 bp amplification of the first homology region of *upp*.
	BG11870	ttacaatttccattcgccattcaggcTGTCTCTAGCGGAGGAATGCGT	FW for 1000 bp amplification of the second homology region of *upp*.
HR plasmids for *ku1*, *ku2* and *ligD* knock-outs	BG12363	tgatgatgaGGATCCATCCGCTGGATTATGCCGAG	FW for *ligD* flanking site 2 amplification, including overhangs with flanking site 1
	BG12366	taatagcgaagaggcccgcacGCCTCGATATCCTCGACGAG	RV for *ligD* flankins site 2 amplification, with overhang for the backbone
	BG12365	ttacaatttccattcgccattcaggcCCGGAGACAAGGACAGTCTCC	FW for *ligD* flanking site 2 amplification, including overhangs with the plasmid backbone
	BG12364	**CGGATGGATCC**tcatcatcaTTGCACGACGAAGGACAGGC	RV for *ligD* flanking site 2 amplification, including overhangs with the flanking site 1
	BG12340	**GGAGGCTGCATGTTTTTGCA**AACCAGCGATCCCGTCCG	RV to combine with BG11886 for amplifying the sgRNA including the spacer for *ligD* and part of *pBBR1MCS2-sgRNA backbone*
	BG12325	**TGCAAAAACATGCAGCCTCC**GTTTTAGAGCTAGAAATAGCAAGTT A AAATAAGGCTAGTC	FW to combine with BG11182 for amplifying the sgRNA including the spacer for *ligD* and part of *hCas9* gene
	BG12360	tgatgatgaGGATCCCGGACCATCGACATCGACTG	FW for *ku1* flanking site 2 amplification, including overhangs with flanking site 1
	BG12359	taatagcgaagaggcccgcacAGCCCTTGCCCTCGATGC	RV for *ku1* flankins site 2 amplification, with overhang for the backbone
	BG12361	ttacaatttccattcgccattcaggcCCCACGAGGTCGAGGATTATG	FW for *ku1* flanking site 2 amplification, including overhangs with the plasmid backbone
	BG12362	**GTCCGGGATCC**tcatcatcaGTTCAGCGTACGGAAGCGCAC	RV for *ku1* flanking site 2 amplification, including overhangs with the flanking site 1

	BG12336	**GGTGAAGGGCTATCCCCGCA**AACCAGCGATCCCGTCCG	RV to combine with BG11886 for amplifying the sgRNA including the spacer for ku1 and part of *pBBR1MCS2-sgRNA backbone*
	BG12333	**TGCGGGGATAGCCCTTCACC**GTTTTAGAGCTAGAAATAGCAAGTT A AAATAAGGCTAGTC	FW to combine with BG11182 for amplifying the sgRNA including the spacer for *ku1* and part of *hCas9* gene
	BG12357	tgatgatgaGGATCCGTTCCCGACAGCGACAAGTG	FW for *ku2* flanking site 2 amplification, including overhangs with flanking site 1
	BG12356	taatagcgaagaggcccgcacATGAGATCGGACGGCTCCCG	RV for *ku2* flankins site 2 amplification, with overhang for the backbone
	BG12355	ttacaatttccattcgccattcaggcGATGGCGGAATGATCGAGGC	FW for *ku2* flanking site 2 amplification, including overhangs with the plasmid backbone
	BG12358	**GGAACGGATCC**tcatcatcaGAAACTCACCCGGTCCGAGG	RV for *ku2* flanking site 2 amplification, including overhangs with the flanking site 1
	BG12339	**CTCACCGCATCGACGAATTC**AACCAGCGATCCCGTCCG	RV to combine with BG11886 for amplifying the sgRNA including the spacer for *ku2* and part of *pBBR1MCS2-sgRNA backbone*
	BG12330	**GAATTCGTCGATGCGGTGAG**GTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTC	FW to combine with BG11182 for amplifying the sgRNA including the spacer for *ku2* and part of *hCas9* gene
Knock-in via HR and counter selection	BG12347	GATGGTCAAGCA**C**GGTCGGG	FW for *upp* flanking site 1 amplification, including overhang for BG12348 and point mutation for the start ATG codon
	BG12348	CCCGACCGTGCTTGACCATC	RV for *upp* flanking site 1 amplification, including overhang for BG12347 and point mutation for the start ATG codon
	BG12908	**TAATGGTCAAGCATGGTCGG**AACCAGCGATCCCGTCCG	RV to combine with BG11886 for amplifying the sgRNA including the spacer 4 for *Δupp* and part of *pBBR1MCS2-sgRNA backbone*
	BG12907	**CCGACCATGCTTGACCATTA**GTTTTAGAGCTAGAAATAGCAAGTT A AAATAAGGCTAGTC	FW to combine with BG11182 for amplifying the sgRNA

			including the spacer 4 for *Δupp* and part of *hCas9* gene
	BG12910	**TTAAGGGAAACCCCGCTTTA**AACCAGCGATCCCGTCCG	RV to combine with BG11886 for amplifying the sgRNA including the spacer 5 for *Δupp* and part of *pBBR1MCS2-sgRNA backbone*
	BG12909	**TAAAGCGGGGTTTCCCTTAA**GTTTTAGAGCTAGAAATAGCAAGTTA A AATAAGGCTAGTC	FW to combine with BG11182 for amplifying the sgRNA including the spacer 5 for *Δupp* and part of *hCas9* gene
sequencing	BG12037	GGCCGAAGGGCTGCATC	for Δupp sequencing
	BG12038	CATGGCAACGATCCCACCTT	for Δupp sequencing
	BG12367	TTCTTGTCTTGTGACTGTCG	for ΔligD sequencing
	BG12368	TAGACCAGACGGCTCTTTTC	for ΔligD sequencing
	BG12369	CGGTATTTCTCGGGCGAGTG	for Δku2 sequencing
	BG12370	AGGAGACCCTGCTCGAACTG	for Δku2 sequencing
	BG12371	TCATGGATCTGGGCGTGCATC	for Δku1 sequencing
	BG12372	CCTCGGACCGGGTGAGTTTC	for Δku1 sequencing

Table S3. List of spacers used for targeting *Rhodobacter sphaeroides'* genomic DNA

Spacer name	Sequence	PAM	Targeted gene
NT	AGGAGGGACGTTGCGACAAG	---	
sp1	CACCTTCGCATAATATTGCA	CGG	*upp*
sp2	ATCGATCTCCTGAAGGAGAA	GGG	*upp*
sp3	CGGCGTCTGGATCGTCGTCG	TGG	*upp*
sp4	CCGACCATGCTTGACCATTA	AGG	*Δupp*
sp5	TAAAGCGGGGTTTCCCTTAA	TGG	*Δupp*
ligD	TGCAAAAACATGCAGCCTCC	CGG	*ligD*
ku1	TGCGGGGATAGCCCTTCACC	TGG	*ku1*
ku2	GAATTCGTCGATGCGGTGAG	CGG	*ku2*

Table S4. List of bacteria and NCBI accession numbers of the proteins used for generating Ku and LigD phylogenetic trees

LigD	
Species	**Accession #**
Rhodobacter sphaeroides	Sequenced genome (Isobionics)
Rhodobacter azotoformans	WP_101340727.1
Cereibacter changlensis	WP_107663713.1
Sinorhizobium meliloti	WP_088200687.1
Aurantimonas manganoxydans	WP_009208938.1
Mycobacterium tuberculosis	ALB18079.1
Pseudomonas putida	ADR60141.1
Mycolicibacterium smegmatis	AFP41860.1
Rhizobium leguminosarum	AUW40742.1
Pseudomonas aeruginosa	BAR68071.1
Nitrospirillum amazonense	ASG25270.1
Rhodobacter johrii	RAZ83231.1
Rhizobium lusitanum	SCB45385.1
Mycobacterium tuberculosis	AJF02293.1
Bacillus cereus	AUZ26128.1
Rhizobium tropici	RAX41073.1
Bacillus subtilis	AQZ90234.1
Ku 1	
Species	**Accession #**
Rhodobacter sphaeroides Ku1	Sequenced genome (Isobionics)
Rhodobacter sphaeroides Ku2	Sequenced genome (Isobionics)
Rhodobacter azotoformans	WP_101341724
Cereibacter changlensis	PTE23121
Cereibacter changlensis	WP_107662491
Rhizobium leguminosarum	WP_047555545
Rhizobium leguminosarum	WP_028734704.1
Pseudomonas putida	KZO53158.1
Mycobacterium tuberculosis	KXN95658.1
Rhodobacter johrii	WP_108223209.1
Rhodobacter azotoformans	WP_101340822.1
Rhizobium lusitanum	WP_092574817.1
Nitrospirillum amazonense	WP_004274142.1
Agrobacterium tumefaciens	WP_025591969.1
Mycobacterium haemophilum	WP_047314295.1
Cereibacter changlensis	WP_107662492.1

Rhizobium tropici	RAX38214.1
Bacillus thuringiensis	AKR33785.1
Bacillus cereus	PGV94828.1

	Ku 2
Species	**Accession #**
Rhodobacter johrii	WP_069332344.1
Streptococcus pneumoniae	CRF99327.1
Mesorhizobium loti	OBQ66756.1
Pseudomonas aeruginosa	BAR68056.1
Mycolicibacterium smegmatis	YP_889815.1

Table S5. NHEJ in *R. sphaeroides* conjugated with pBBR_Cas9_sp1 and pBBR_Cas9_sp2.

Plasmid	**CFUs (100x dilution)**	**5FU-screening: edited colonies/streaked colonies**	**% NHEJ efficiency ***
pBBR_Cas9_sp1	113	3/40	7.5
pBBR_Cas9_sp2	89	6/85	7.1

* The percentage of NHEJ efficiency has been determined by restreak of conjugants strains on 5FU and subsequent sequencing of the protospacer locus. All the survival colonies on the 5FU plates showed indels in the protospacer region.

Table S6. Efficiency of the knock-outs generated via HR-CRISPR/Cas9 counter-selection for the NHEJ genes *ligD, ku1* and *ku2*

Plasmid used	**Mutated colonies/colonies screened**	**Mutation efficiency (%)**
pBBR_Cas9_ΔligD (spacer1)	1/7	14%
pBBR_Cas9_Δku1 (spacer3)	1/9	11%
pBBR_Cas9_Δku2 (spacer3)	6/20	30%

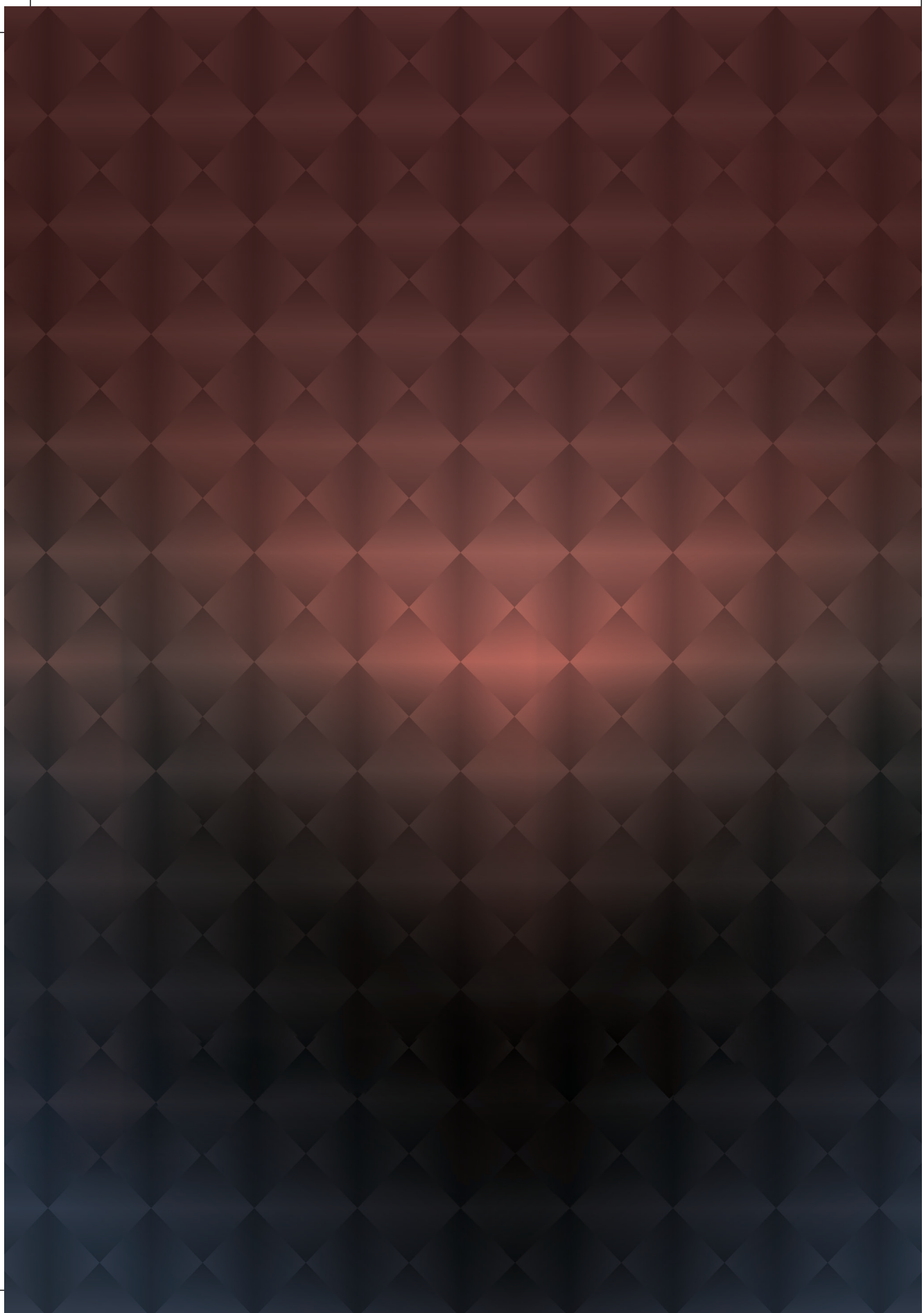
Table S7. Comparison of the NHEJ efficiencies in the different NHEJ KO strains conjugated with pBBR_cas9_sp1/2 targeting *upp*.

strain and spacers	dilution	# conjugants	# screened colonies	# survivals (5FU)	# NHEJ mutants	% NHEJ efficiency *
wt-sp1	100x	17	15	0	0	0
wt-sp2	100x	60	60	3	3	5
ΔligD-sp1	1x	1726	179	0	0	0
ΔligD-sp2	1x	2004	195	0	0	0
Δku1-sp1	100x	81	65	3	1	1.54
Δku1-sp2	1000x	658	81	5	4	4.94
Δku2-sp1	10x	155	65	4	2	3.08
Δku2-sp2	100x	839	100	21	20	20
Δku1-2-sp1	10x	120	106	3	0	0
Δku1-2-sp2	100x	568	130	8	7	5.38

* The efficiencies were determ ined as ratio of the num ber of sequenced NHEJ m utants to the num ber of colonies screened for 5FU toxicity.

Table S8. Description of the type of indels registered for the m utants obtained from the pBBR_cas9_sp1/2 in the NHEJ KO strains (Δ*ligD*, Δ*ku1*, Δ*ku2*, Δk*u1-2*).

strain	colony	mutation	strain	colony	mutation
wt-sp2	wt-sp2c24	1bp substitution	Δku2-sp2	k2'-sp2c2	1bp substitution
	wt-sp2c32	1bp substitution		k2'-sp2c9	9bp deletion
	wt-sp2c49	1bp substitution		k2'-sp2c11	1bp substitution
Δku1-sp1	k1-sp1c23	1bp insertion		k2'-sp2c12	1bp substitution
Δku1-sp2	k1-sp2c19	1bp substitution		k2'-sp2c15	6bp repeat insertion
	k1-sp2c49	6bp repeat insertion		k2'-sp2c34	9bp deletion
	k1-sp2c55	1bp substitution		k2'-sp2c37	1bp substitution
	k1-sp2c77	1bp substitution		k2'-sp2c39	1bp substitution
Δku1-2-sp2	k12-sp2c50	6bp repeat insertion		k2'-sp2c40	1bp substitution
	k12-sp2c53	1bp substitution		k2'-sp2c43	6bp repeat insertion
	k12-sp2c57	1bp substitution		k2'-sp2c49	1bp substitution
	k12-sp2c76	1bp substitution		k2'-sp2c51	6bp repeat insertion
	k12-sp2c81	1bp substitution		k2'-sp2c66	1bp substitution
	k12-sp2c106	6bp repeat insertion		k2'-sp2c67	1bp substitution
	k12-sp2c108	1bp deletion		k2'-sp2c69	6bp repeat insertion
Δku2-sp1	k2'-sp1c6	1bp insertion		k2'-sp2c74	55bp deletion
	k2'-sp1c36	1bp insertion		k2'-sp2c79	1bp substitution
				k2'-sp2c95	6bp repeat insertion
				k2'-sp2c96	1bp substitution
				k2'-sp2c98	9bp deletion



CHAPTER 5

CHARACTERIZING A THERMOSTABLE CAS9 FOR BACTERIAL GENOME EDITING AND SILENCING

Ioannis Mougiakos[1#], Prarthana Mohanraju[1#], Elleke F. Bosma[1#], Valentijn Vrouwe[1], Max Finger Bou[1], Mihris I. S. Naduthodi[1], Alex Gussak[1], Rudolf B. L. Brinkman[2], Richard van Kranenburg[1, 2], John van der Oost[1*]

[1]Laboratory of Microbiology, Wageningen University, Stippeneng 4, 6708 WE

Wageningen, The Netherlands

[2] Corbion, Arkelsedijk 46, 4206 AC Gorinchem, The Netherlands

#Contributed equally

*Corresponding authors

Chapter adapted from publication:

Nat Commun. 2017 Nov 21;8(1):1647. doi: 10.1038/s41467-017-01591-4

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 characterize ThermoCas9 from the thermophilic bacterium *Geobacillus thermodenitrificans* T12. We show that ThermoCas9 is active *in vitro* between 20°C and 70°C, 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°C in *Bacillus smithii* and for gene deletion at 37°C 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.

KEYWORDS

CRISPR/Cas9, thermoCas9, *Bacillus smithii*, *Pseudomonas putida*, thermophiles, genome editing, genome silencing

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[1–4]. CRISPR-Cas systems are subdivided into two classes (1 and 2) and six types (I-VI), depending on their complexity and signature proteins[5]. 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[6–10] and prokaryotes[11–13]. 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[14]. 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[15–17]. 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[18-20]. The inactive or dead Cas9 (dCas9) has been instrumental as an efficient gene silencing system and for modulating the expression of essential genes[11,21,22].

To date, *Streptococcus pyogenes* Cas9 (SpCas9) is the best characterized 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[23]. 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[24–27], 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°C, 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[28]. However, a Cas9-based editing and silencing tool for obligate thermophiles is not yet available a s SpCas9 is not active at elevatedtemperatures[28,29], and to date no thermophilic Cas9 has been adapted for such purpose. Here, we describe the characterization of ThermoCas9, an RNA-guided DNA-endonuclease from the CRISPR-Cas type-IIc system of the thermophilic bacterium *Geobacillus thermodenitrificans* T12[30]. We show that ThermoCas9 is active *in vitro* between 20 and 70°C 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[31] at 55°C, 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[32,33], 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°C[30]. Contrary to previous claims that type II CRISPR-Cas systems are not present in thermophilic bacteria[34], the sequencing results revealed the existence of a type-IIC CRISPR-Cas system in the genome of *G. thermodenitrificans* T12 (Figure 1A)[35]. 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[36]. Furthermore, ThermoCas9 was expected to be active at least around the temperature optimum of *G. thermodenitrificans* T12[30]. 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, Supplementary Table 1), mostly within the *Geobacillus* genus, supporting the idea that ThermoCas9 is part of a highly conserved defense system of thermophilic bacteria (Figure 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[11,36]. 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 a 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°C and 65°C, for 15 and 30 min. SDS-PAGE analysis showed that the purified ThermoCas9 denatures at 65°C but not at 60°C, while the denaturation temperature of ThermoCas9-sgRNA complex is above 65°C (Figure 1C). The demonstrated thermostability of ThermoCas9 implied its potential as a thermo-tolerant CRISPR-Cas9 genome editing tool, and encouraged us to analyze some relevant molecular features in more detail.

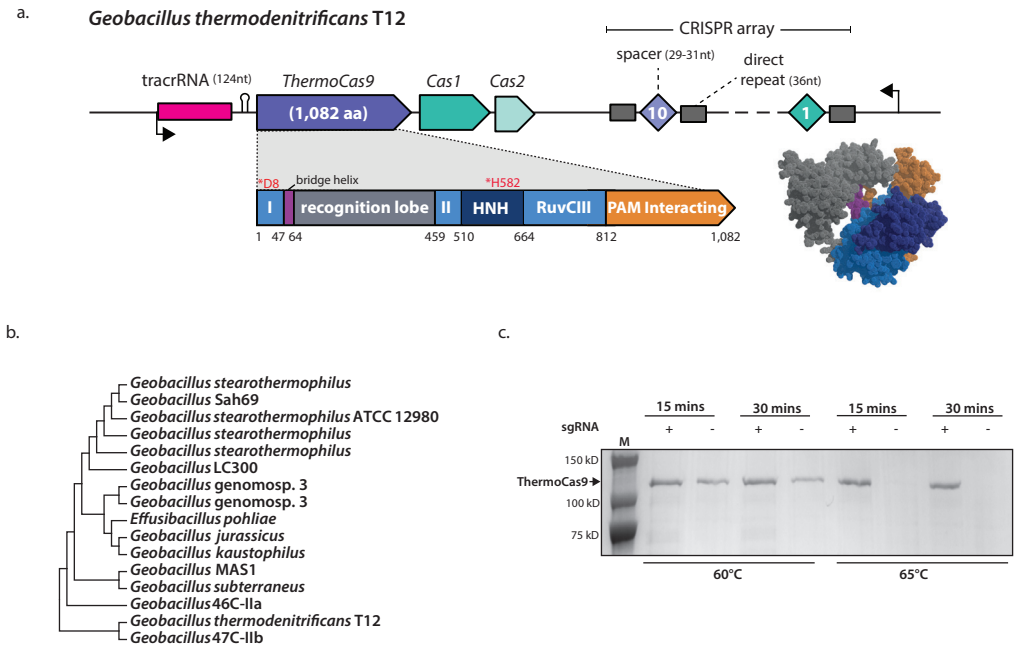


Figure 1. The *Geobacillus thermodenitrificans* T12 type-IIC CRISPR-Cas locus encoding ThermoCas9.
a) Schematic representation of the genomic locus encoding ThermoCas9. The domain architecture of ThermoCas9 based on sequence comparison, with predicted active sites residues highlighted in magenta. A homology model of ThermoCas9 generated using Phyre 2 [58] is shown, with different colours for the domains. b) Phylogenetic tree of Cas9 orthologues highly identical to ThermoCas9. Evolutionary analysis was conducted in MEGA7[59]. 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 apo-ThermoCas9 molecular weight of 126 kDa.

THERMOCAS9 PAM DETERMINATION

The first step towards the characterization 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[37]. As only two hits were obtained with phage genomes (Supplementary Fig. 1A), 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°C, the cleaved sequences of the library (together with a non-targeted library sample as control) were separated from uncleaved sequences, by gel electrophoresis, and analyzed by deep-sequencing in order to identify the ThermoCas9 PAM preference (Figure 2A). The sequencing results revealed that ThermoCas9 introduces double stranded DNA breaks that, in analogy with the mesophilic Cas9 variants, 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 2B). Recent studies have revealed the importance of the 8th PAM position for target recognition of some Type IIC Cas9 orthologues[17,38]. For this purpose, and taking into account the results from the *in silico* ThermoCas9 PAM prediction (Supplementary Fig. 1), we performed additional PAM determination assays. This revealed optimal targeting efficiency in the presence of an adenine at the 8th PAM position (Figure 2C). Interestingly, despite the limited number of hits, the aforementioned *in silico* PAM prediction (Supplementary Fig. 1B) 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 6 th 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°C) for 10 min before combining and incubating them for 1 hour at the corresponding assay temperature.

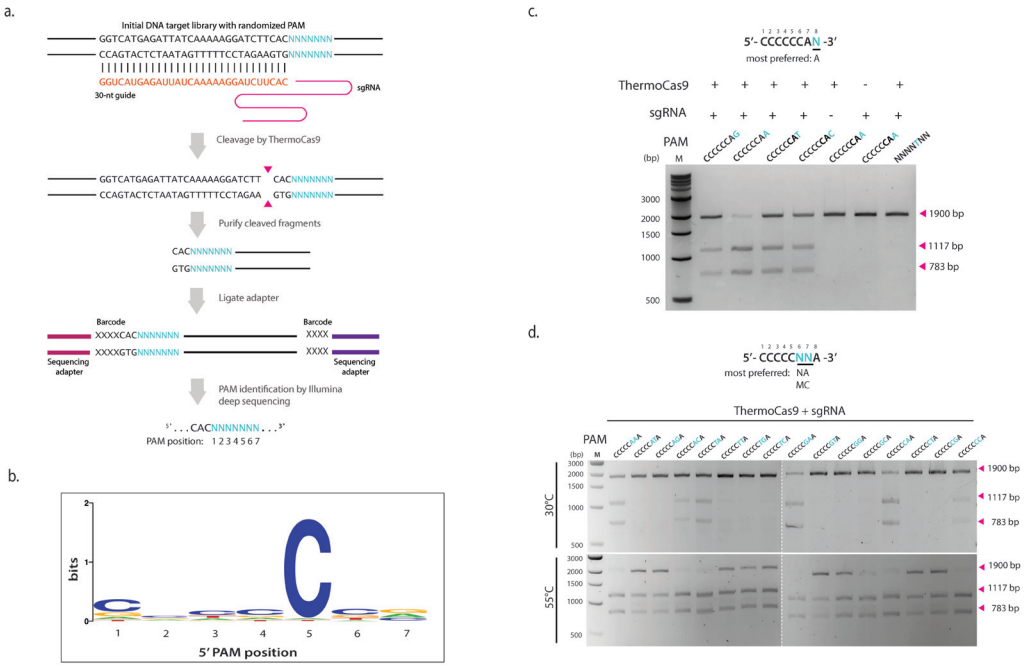


Figure 2. ThermoCas9 PAM analysis. a) Schematic illustrating the *in vitro* cleavage assay for discovering the position and identity (5′-NNNNNNN-3') of the protospacer adjacent motif (PAM). Magenta triangles indicate the cleavage position. b) Sequence logo of the consensus 7nt long PAM of ThermoCas9, 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 Supplementary Fig. 2. Supplementary Fig. 9 shows the uncropped gel images.

When the assays were performed at temperatures between 37°C and 60°C, all the different DNA substrates were cleaved (Figure 2D, S3). However, the most digested target fragments 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°C, 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 2D). Lastly, at 20°C only the DNA substrates with (5th to 8th PAM positions) 5'-CVAA-3' and 5'-CCCA-3' PAM sequences were targeted (Supplementary Fig. 2), 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.

METAL ION DEPENDENCY

Previously characterized, mesophilic Cas9 endonucleases employ divalent cations to catalyze the generation of DSBs in target DNA[14,39]. 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 (Supplementary Fig. 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[14,20,40]. Addition of Fe^{2+}, Co^{2+}, Ni^{2+}, or Zn^{2+} ions also mediated cleavage. Ca^{2+} only supported plasmid nicking, suggesting that with this cation only one of the endonuclease domains is functional.

THERMOSTABILITY AND TRUNCATIONS

The predicted tracrRNA consists of the anti-repeat region followed by three hairpin structures (Figure 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 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 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°C) for 1, 2 and 5 min before combining and incubating them for 1 hour at various assay temperatures (37-65°C). 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 3B). Moreover, the cleavage efficiency drastically dropped upon removal of both the middle and the 3' hairpins (Supplementary Fig. 4). Whereas pre-heating ThermoCas9 at 65°C 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°C, 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

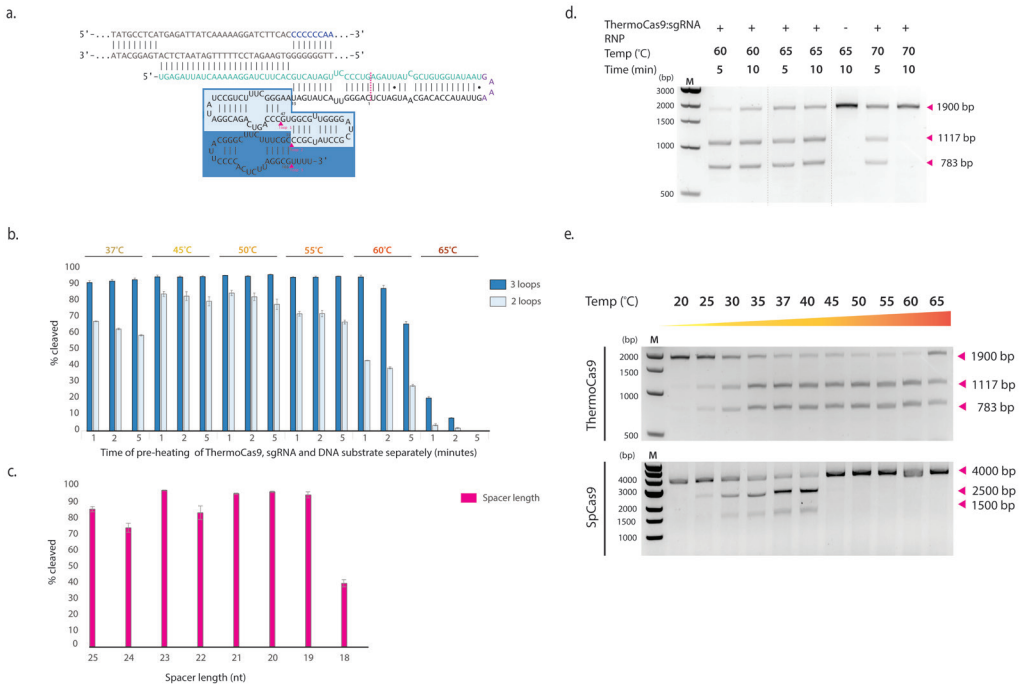


Figure 3. ThermoCas9 temperature range and effect of sgRNA-binding. a) Schematic representation of the sgRNA and a matching target DNA. The target DNA, the PAM and the crRNA are shown in grey, blue and green, respectively. The site where the crRNA is linked with the tracrRNA is shown in purple. The dark 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.. The blots of one of the replicates are shown in Supplementary Fig. 10. 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°C. Average values of three replicates are shown, with error bars representing S.D.. The blots of one of the replicates are shown in Supplementary Fig. 11. d) To identify the maximum temperature, endonuclease activity of ThermoCas9:sgRNA RNPcomplex 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 preheated DNA substrate was added and the reaction was incubated for 1 hour at the same temperature.

cleavage efficiency drops significantly when a spacer of 18 nt is used (Figure 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°C for 5 and 10 min before adding pre-heated DNA and subsequent incubation for 1 hour at 60, 65 and 70°C. Strikingly, we observed that the ThermoCas9 RNP was active up to 70°C, in spite of its pre-heating for 5 min at 70°C (Figure 3D). This finding confirmed our assumption that the ThermoCas9 stabili-

ty strongly correlates with the association of an appropriate sgRNA guide.

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°C. 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 previousanalysis[28,29], the mesophilic SpCas9 was active only between 25 and 44°C (Figure 3E); above this temperature SpCas9 activity rapidly decreased to undetectable levels. In contrast, ThermoCas9 cleavage activity could be detected between 25 and 65°C (Figure 3). This indicates the potential to use ThermoCas9 as a genome editing tool for both thermophilic and mesophilic organisms.

Based on previous reports that certain type-IIC systems were efficient single stranded DNA cutters[41,40], we tested the activity of ThermoCas9 on ssDNA substrates. However, no cleavage was observed, indicating that ThermoCas9 is a dsDNA nuclease (Supplementary Fig. 5B).

SPACER-PROTOSPACER MISMATCH TOLERANCE

The predicted tracrRNA consists 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 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°C.

The ThermoCas9:sgRNA activity on linear dsDNA targets was abolished at 37°C for most of the single-mismatch targets (Figure 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 4b). At 55°C, the cleavage efficiency for single-mismatch linear targets was higher than at 37°C, 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 4c). On the contrary, single-mismatches at positions 1, 2 and 20 were the most tolerated for cleavage (Figure 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°C, re-

gardless 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 4b). At 55°C, a similar trend was observed, with elevated cleavage rates (Figure 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 4b, 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° C (Figure 4b, 4c).

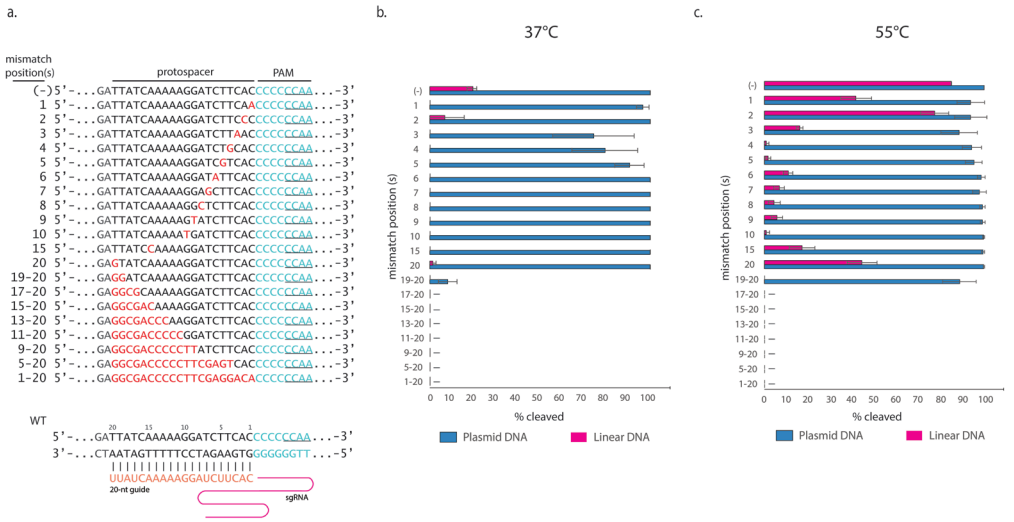


Figure 4. Protospacer targeting Specificity of ThermoCas9. a) Scheme of the generated mismatch protospacers library, employed for evaluating the ThermoCas9:sgRNA targeting specificity *in vitro*. The mismatch spacer-protospacer positions are shown in red, the PAM in light blue with the 5th to 8th positions underlined. b) Graphical representation of the ThermoCas9:sgRNA cleavage efficiency over linear or plasmid targets with different mismatches at 37°C. 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 blots of one of the replicates are shown in Supplementary Fig. 12. c) Graphical representation of the ThermoCas9:sgRNA cleavage efficiency over linear or plasmid targets with different mismatches at 55°C. 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.

THERMO CAS9-BASED GENE DELETION IN THE THERMOPHILE *B. 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*[24,25,42,43]. 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°C. 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[24,28,31,44]. 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 xylA promoter (P_{xylA}), 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 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_bsΔpyrF1 and pThermoCas9_bsΔpyrF2 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°C 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[28,44]. In contrast, transformation with the pThermoCas9_bsΔpyrF1 and pThermoCas9_bsΔpyrF2 plasmids resulted in 20 and 0 colonies respectively. Out of the ten pThermoCas9_ΔpyrF1 colonies screened, one was a clean Δ*pyrF* mutant whereas the rest had a mixed wild type/Δ*pyrF* genotype (Figure 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[11,28]. Hence, Cas9 functions as stringent counter-selection system to kill cells that have not performed the desired HR prior to Cas9 cleavage[11,28,45]. 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[11,28,45]. As the promoter we use here for thermocas9-expression is not sufficiently controllable and HR is inefficient in *B. smithii*[28,44], the low number (pyrF1) or even complete

lack (pyrF2) of colonies we observed here in the presence of an HR template confirms the high *in vivo* activity of ThermoCas9 at 55°C. In the SpCas9-based counter-selection system we previously developed for *B. smithii*, the activity of SpCas9 was 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[28]. We anticipate that the use of a tightly controlled promoter will increase efficiencies of the ThermoCas9-system.

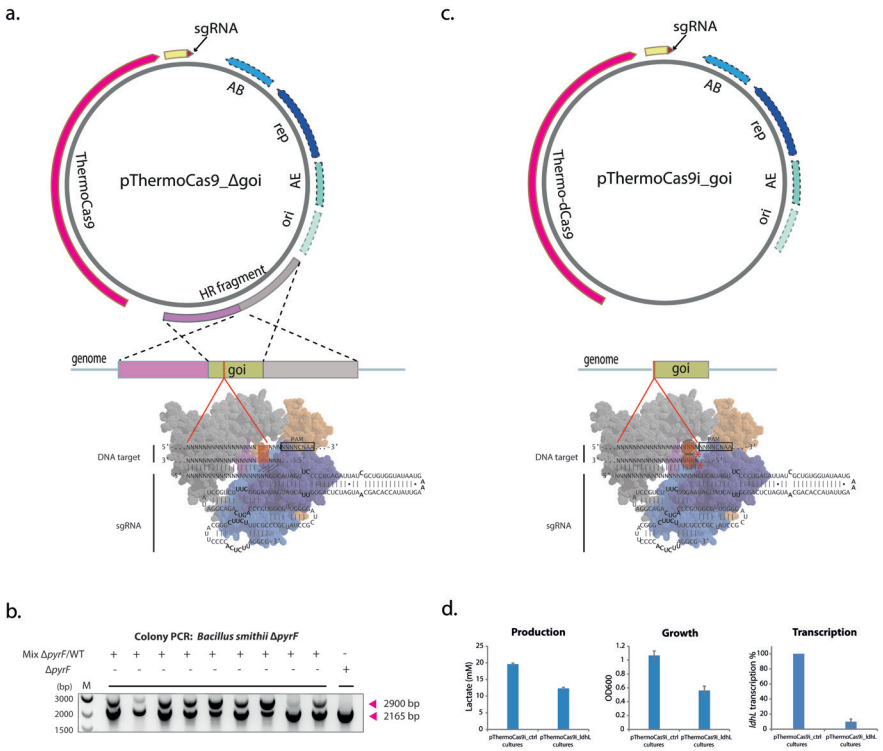


Figure 5. ThermoCas9-based genome engineering in thermophiles. a) Schematic overview of the basic pThermoCas9_Δgene-of-interest (goi) construct. The *thermocas9* gene was introduced either to the pNW33n (*B. smithii*) or to the pEMG (*P. putida*) vector. Homologous recombination flanks were introduced upstream thermocas9 and encompassed the 1kb (*B.smithii*) or 0.5kb (*P. putida*) upstream and and downstream regions 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 accesory 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. Supplementary Fig. 13 shows the uncropped gel image. 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*. d) Graphical representation of the production, growth and RT-qPCR results from the ldhL silencing experimen t 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 *P. 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[46,47]. 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[32,33]. Once more, we followed a single plasmid approach and combined homologous recombination and ThermoCas9-based counter-selection. We constructed the pEMG-based pThermoCas9_ppΔpyrF plasmid containing the thermocas9 gene under the control of the 3-methylbenzoate-inducible P_m-promoter, a homologous recombination template for deletion of the *pyrF* gene and a sgRNA expressing module under the control of the constitutive P_3 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° C. 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 (Supplementary Fig. 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°C. **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 hete-

rologously 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 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 (Supplementary Fig. 7). The position and targeted strand selection were based on previous studies[18,48], 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°C 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[44]. The growth of the pThermoCas9i_ldhL cultures was 50% less than the growth of the pThermoCas9i_ctrl cultures (Figure 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 LdhL-based NAD+-regenerating capacity under micro-aerobic conditions[44]. 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 5D).

DISCUSSION

Most CRISPR-Cas applications are based on RNA-guided DNA interference by Class 2 CRISPR-Cas proteins, such as Cas9 and Cas12a[6–13]. Prior to this work, there were only a few examples of Class 1 CRISPR-Cas systems present in thermophilic bacteria and archaea[5,49], which have been used for genome editing of thermophiles[34]. As a result, the application of CRISPR-Cas technologies was mainly restricted to temperatures below 42°C, due to the mesophilic nature of the employed Cas-endonucleases[28,29]. 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 characterized ThermoCas9, a Cas9 orthologue from the thermophilic bacterium *G. thermodenitrificans* T12 that has been isolated from compost[30]. 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°C, which is much broader than the 25–44°C 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°C, 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°C), whereas more variety in the PAM is allowed for activity at the moderate to optimal temperatures (37-60°C). 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[50]. Additionally, we showed that the ThermoCas9 activity on linear DNA targets is very sensitive to spacer-protospacer mismatches. At 55°C, 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[51]. 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°C. 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 tolerancei of ThermoCas9 compared to SpCas9[52]and highlight its potential as a genome editing tool for eukaryotic cells with enhanced target specificity. Elucidation of the ThermoCas9 crystal structure is required

togain insight on the molecular basis of the ThermoCas9 cleavage mechanism.

Based on the here described characterization of the novel ThermoCas9, we successfully developed genome engineering tools for strictly thermophilic prokaryotes. We showed that ThermoCas9 is active *in vivo* at 55°C and 37°C, 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°C up to 70°C. 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[34]. 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.

METHODS

BACTERIAL STRAINS AND GROWTH CONDITIONS

The moderate thermophile *B. smithii* ET 138 ΔsigF Δ*hsdR*[28] was used for the gene editing and silencing experiments using ThermoCas9. It was grown at 55°C in LB2 medium[44] . 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[44] in a 50 mL Greiner tube for 24 h at 55°C and 120 rpm after transfer from an overnight culture.[44]. For protein expression, *E. coli* Rosetta (DE3) was grown in LB medium in flasks at 37°C in a shaker incubator at 120 rpm until an $OD_{600\ nm}$ of 0.5 was reached after the temperature was switched to 16°C. 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, a fter which incubation was co

ntinued at 16°C. For cloning PAM constructs for 6th and 7th, and 8th 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°C. For cloning degenerate 7-nt long PAM library, electro-competent DH10B *E. coli* cells were transformed according to standard procedures[53] and grown on LB agar plates at 37°C overnight. *E. coli* DH5α λpir (Invitrogen) was used for *P. putida* plasmid construction using the transformation procedure described by Ausubel *et al*.[54]. For all *E. coli* strains, if required chloramphenicol was used in concentrations of 25 m g L-1 and kanamycin in 50 m*g*L^{-1}. *P. putida* KT2440 (DSM 6125) strains were cultured at 37°C 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 Ni2+-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 (Supplementary Table 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 OD600nm 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°C 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 16,000xg for 1 hour at 4°C 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°C 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 Supplementary Table 2) were incubated separately (unless otherwise indicated) at the stated temperature for 10 min, followed by combining the components together and incu bating 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 the n 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* transcribe d 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 reacti on volume. The PAM containing cleavage fragments from the 55°C 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 [55].

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 Supplementary Table 2 and 3 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 electro-competent cells[44]. 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_bsΔpyrF1 and pThermoCas9_bsΔpyrF2 vectors, the pNW33n backbone together with the Δ*pyrF* homologous recombination flanks were PCR amplified from the pWUR_Cas9sp1_hr vector[28] (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 [28] (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 and the

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_ppΔpyrF. 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 P_3 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 P_m-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*.[56]. 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°C 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 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.

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_ppΔpyrF. 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 P_3 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.[56]. 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°C 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.

DATA AVAILABILITY

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

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.

COMPETING INTERESTS

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

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.

REFERENCES

1. Brouns, S. J. J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science (80-.).. 321,(2008).
2. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science (80-.). 315, (2007).
3. Wright, A., Nunez, J. & Doudna,J. Biology and applications of CRISPR systems: Harnessing nature's toolbox for genome engineering.Cell 164, 29–44 (2016).
4. Mohanraju, P. et al. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. Science (80-.). 353, aad5147 (2016).
5. Makarova, K. S. et al. An updated evolutionary classification of CRISPR–Cas systems. Nat. Rev. Microbiol. 13, 722–736 (2015).
6. Komor, A. C. et al. CRISPR-based technologies for the manipulation of eukaryotic genomes. Cell 168, 20–36 (2017).
7. Puchta, H. Applying CRISPR/Cas for genome engineering in plants: the best is yet to come. Curr. Opin. Plant Biol. 36, 1–8 (2017).
8. Xu, J. et al. A toolkit of CRISPR-based genome editing systems in Drosophila. J. Genet.Genomics 42, 141–149 (2015).
9. Tang, X. et al. A CRISPR–Cpf1 system for efficient genome editing and transcriptional repression in plants. Nat. Plants 3, 17018 (2017).
10. Zetsche, B. et al. Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. Nat. Biotechnol. 35, 31–34 (2016).
11. Mougiakos, I., Bosma, E. F., de Vos, W. M., van Kranenburg, R. & van der Oost, J. Next generation prokaryotic engineering: The CRISPR-Cas toolkit. Trends Biotechnol. 34, 575–587 (2016).
12. Yan, M.-Y. et al. CRISPR-Cas12a-assisted recombineering in bacteria. Appl. Environ. Microbiol. AEM.00947–17 (2017). doi:10.1128/AEM.00947-17.
13. Jiang, Y. et al. CRISPR-Cpf1 assisted genome editing of *Corynebacterium glutamicum*. Nat. Commun. 8, 15179 (2017).

14. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial. immunity. Science (80-.). 337, 816–821 (2012).
15. Mojica, F. J. M. Short motif sequences determin e the targetsof the prokaryotic CRISPR defence system. Microbiology 155, 733–740 (2009).
16. Deveau, H. et al. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J. Bacteriol. 190, 1390–400 (2008).
17. Karvelis, T. et al. Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. Genome Biol. 16, 253 (2015).
18. Bikard, D. et al. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res. 41, 7429–7437 (2013).
19. Qi, L. S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152, 1173–1183 (2013).
20. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. U. S. A. 109, E2579–86 (2012).
21. Lv, L., Ren, Y.-L., Chen, J.-C., Wu, Q. & Chen, G.-Q. Application of CRISPRi for prokaryotic metabo lic engineering involving multiple genes, a case study: Controllable P(3HB-co-4HB) biosynthesis. Metab. Eng. 29, 160–168 (2015).
22. Choudhary, E., Thakur, P., Pareek, M., Agarwal, N. & Rajagopalan, M. Gene silencing by CRISPR interference in mycobacteria. Nat. Commun. 6, 6267 (2015).
23. Nakade, S., Yamamoto, T. & Sakuma, T. Cas9, Cpf1 and C2c1/2/3—What's next? Bioengineered 1–9 (2017). doi:10.1080/21655979.2017.1282018.
24. Bosma, E. F., van der Oost, J., de Vos, W. M. & van Kranenburg, R. Sustainable production of bio-based chemicals by extremophiles. Curr. Biotechnol. 2, 360–379 (2013).
25. Taylor, M. P., van Zyl, L., Tuffin, I. M., Leak, D. J. & Cowan, D. A. Genetic tool development underpins recent advances in thermophilic whole-cell biocatalysts. Microb. Biotechnol. 4, 438–448 (2011).
26. Olson, D. G., Sparling, R. & Lynd, L. R. Ethanol production by engineered thermophiles. Curr. Opin. Biotechnol. 33, 130–141 (2015).
27. Zeldes, B. M. et al. Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals. Front. Microbiol. 6, 1209 (2015).
28. Mougiakos, I. et al. Efficient genome editing of a facultative thermophile using mesophilic spCas9. ACS Synth. Biol. 6, 849–861 (2017).
29. Wiktor, J., Lesterlin, C., Sherratt, D. J. & Dekker, C. CRISPR-mediated control of the bacterial initiation of replication. Nucleic Acids Res. 44, 3801–10 (2016).
30. Daas, M. J. A., van de Weijer, A. H. P., de Vos, W. M., van der Oost, J. & van Kranenburg, R. Isolation of a genetically accessible thermophilic xylan degrading bacterium from compost. Biotechnol. Biofuels 9, 210 (2016).
31. Bosma, E. F. 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, 1874–1883 (2015).
32. Aparicio, T., Jensen, S. I., Nielsen, A. T., de Lorenzo, V. & Martínez-García, E. The Ssr protein (T1E_1405) from Pseudomonas putida DOT-T1E enables oligonucleotide-based recombineering in platform strain P. putida EM42. Biotechnol. J. 11, 1309–1319 (2016).
33. Martínez-García, E. & de Lorenzo, V. The quest for the minimal bacterial genome. Curr. Opin. Biotechnol. 42, 216–224 (2016).
34. Li, Y. et al. Harnessing Type I and Type III CRISPR-Cas systems for genome editing. Nucleic Acids Res. 44, e34–e34 (2016).
35. Daas, M. J. A., Vriesendorp, B., Van de Weijer, A. H. P., Van der Oost, J. & Van Kranenburg, R. Complete Geno me Sequence of *Geobacillus thermodenitrificans* T12, A Potential Host for Biotechnological Applications. doi:10.1007/s00284-017-1349-0.
36. Ran, F. A. et al. *In vivo* genome editing using Staphylococcus aureus Cas9. Nature 520, 186–191 (2015).
37. Biswas, A., Gagnon, J. N., Brouns, S. J. J., Fineran, P. C. & Brown, C. M. CRISPRTarget: bioinformatic prediction and analysis of crRNA targets. RNA Biol. 10, 817–827 (2013).
38. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J.-S. Highly efficient RNA-guidedgenome editing in human cells via delivery ofpurified Cas9 ribonucleoproteins. Genome Res. 24, 1012–9 (2014).
39. Chen, H., Choi, J. & Bailey, S. Cut site selection by the two nuclease domains of the Cas9 RNA-guided endonuclease. J. Biol. Chem. 289, 13284–13294 (2014).

40. Zhang, Y., Rajan, R., Seifert, H. S., Mondragon, A. & Sontheimer, E. J. DNase H Activity of Neisseria meningitidis Cas9. Mol. Cell 60, 242–255 (2015).
41. Ma, E., Harrington, L. B., O'Connell, M. R., Zhou, K. & Doudna, J. A. Single-Stranded DNA Cleavage by Divergent CRISPR-Cas9 Enzymes. Mol. Cell 60, 398–407 (2015).
42. Bhalla, A., Bansal, N., Kumar, S., Bischoff, K. M. & Sani, R. K. Improved lignocellulose conversion to biofuels with thermophilic bacteria and thermostable enzymes. Bioresour. Technol.128, 751–759 (2013).
43. Abdel-Banat, B. M. A., Hoshida, H., Ano, A., Nonklang, S. & Akada, R. 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 (2010).
44. Bosma, E. F. et al. Establishment of markerless gene deletion tools in thermophilic*Bacillus smithii* and construction of multiple mutant strainsMicrob. Cell Fact. 14, 99 (2015).
45. Oh, J.-H. & van Pijkeren, J.-P. CRISPR-Cas9-assisted recombineering in Lactobacillus reuteri. Nucleic Acids Res. 42, e131 (2014).
46. Nikel, P. I., Chavarría, M., Danchin, A. & de Lorenzo, V. From dirt to industrial applications: Pseudomonas putida as a Synthetic Biology chassis for hosting harsh biochemical reactions. Curr. Opin. Chem. Biol. 34, 20–29 (2016).
47. Poblete-Castro, I., Becker, J., Dohnt, K., dos Santos, V. M. & Wittmann, C. Industrial biotechnology of Pseudomonas putida and related species. Appl. Microbiol. Biotechnol. 93, 2279–2290 (2012).
48. Larson, M. H. et al. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat. Protoc. 8, 2180–2196 (2013).
49. Weinberger, A. D., Wolf, Y. I., Lobkovsky, A. E., Gilmore, M. S. & Koonin, E. V. Viral diversity threshold for adaptive immunity in prokaryotes. MBio3, e00456–12 (2012).
50. Jinek, M. et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science (80). 343, 1247997–1247997 (2014).
51. Westra, E. R. et al. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Mol. Cell 46, 595–605 (2012).
52. Jinek, M. et al. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science (80-.). 337, 816–821 (2012).
53. Sambrook, J., Fritsch, E. F. & Maniatis, T. Molecular cloning : a laboratory manual. (Cold Spring Harbor Laboratory, 1989).
54. Current Protocols in Molecular Biology. (John Wiley & Sons, Inc., 2001). doi:10.1002/0471142727
55. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A sequence logo generator. Genome Res. 14, 1188–1190 (2004).
56. Choi, K.-H., Kumar, A. & Schweizer, H. P. A 10-min method for preparation of highly electrocompetent Pseudomonas aeruginosa cells: Application for DNA fragment transfer between chromosomes and plasmid transformation. J. Microbiol. Methods 64, 391–397 (2006).
57. van Hijum, S. A. F. T. 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 (2005).
58. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10, 845–858 (2015).
59. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874 (2016).

SUPPLEMENTARY INFORMATION

SUPPLEMENTARY TABLE 1 | PBLAST RESULTS OF CAS9 PROTEIN SEQUENCES FROM FIGURE 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 genom osp. 3	87
Geobacillus genom osp. 3	87
Geobacillus subterraneus	87
Effusibacillus pohliae	86

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

SUPPLEMENTARY TABLE 2 | OLIGONUCLEOTIDES USED IN THIS STUDY, RELATED TO FIGURES 1 TO 4.

	Oligo	Sequence	Description
PAM Library construction	BG6494	TATGCC***TCATGA***GATTATCAAAAAGGATCTTCACNNNNNN NCTAGA TCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with 7-nt long random PAM sequence
	BG6495	TATGCC***GGATCC***TCAGACCAAGTTTACTCATATATACTTTAG ATTGAT TTAAAACTTCATTTTTAATTTAAAAGGATCTAG	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	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACCATG ATTACG CCAAGC	RV sequencing adapter for PCR amplification of the ThermoCas9 cleaved fragments
	BG7616	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTCATGAG ATTAT CAAAAAGGATCTTC	RV sequencing adaptor for PCR amplification of the control fragments
	BG8157	TATGCC***TCATGA***GATTATCAAAAAGGATCTTCACCCCCCC AGCTAG ATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM "CCCCCCAG"
	BG8158	TATGCC**TCATGA**GATTATCAAAAAGGATCTTCACCCCCC CAACTA GATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM "CCCCCCAA"
	BG8159	**TATGCC*TCATGA*GATTATCAAAAAGGATCTTCACCCCCC CATCTA** GATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM "CCCCCCAT"
	BG8160	TATGCC***TCATGA***GATTATCAAAAAGGATCTTCACCCCCCC ACCTAG ATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM "CCCCCCAC"
	BG8161	TATGCC***TCATGA***GATTATCAAAAAGGATCTTCACNNNNTN NCTAGA TCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	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	TATGCC***TCATGA***GATTATCAAAAAGGATCTTCACCCCCCA AACTAG ATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "AA"
	BG8764	**TATGCC*TCATGA*GATTATCAAAAAGGATCTTCACCCCCC ATACTA** GATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "AT"
	BG8765	TATGCC***TCATGA***GATTATCAAAAAGGATCTTCACCCCCCA GACTAG ATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of in *vitro* target DNA with PAM position 6&7 "AG"
	BG8766	TATGCC***TCATGA***GATTATCAAAAAGGATCTTCACCCCCCA CACTAG ATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "AC"

	Oligo	Sequence	Description
PAM Library construction	BG8767	TATGCC***TCATGA***GATTATCAAAAAGGATCTTCACCCCCCT AACTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "TA"
	BG8768	TATGCC***TCATGA***GATTATCAAAAAGGATCTTCACCCCCCT TACTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "TT"
	BG8769	TATGCC***TCATGA***GATTATCAAAAAGGATCTTCACCCCCCT GACTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "TG"
	BG8770	TATGCC***TCATGA***GATTATCAAAAAGGATCTTCACCCCCCT CACTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "TC"
	BG8771	TATGCC***TCATGA***GATTATCAAAAAGGATCTTCACCCCCCG AACTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "GA"
	BG8772	TATGCC***TCAT***GAGATTATCAAAAAGGATCTTCACCCCCCG TACTAG ATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "GT"
	BG8773	TATGCC***TCATGA***GATTATCAAAAAGGATCTTCACCCCCCG GACTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "GG"
	BG8774	TATGCC***TCATGA***GATTATCAAAAAGGATCTTCACCCCCCG CACTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "GC"
	BG8775	TATGCC***TCATGA***GATTATCAAAAAGGATCTTCACCCCCCC AACTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "CA"
	BG8776	TATGCC***TCATGA***GATTATCAAAAAGGATCTTCACCCCCCC TACTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "CT"
	BG8777	TATGCC***TCATGA***GATTATCAAAAAGGATCTTCACCCCCCC GACTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	FW for construction of *in vitro* target DNA with PAM position 6&7 "CG"
	BG8778	TATGCC***TCATGA***GATTATCAAAAAGGATCTTCACCCCCCC CACTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC	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	AAGCTTGAAATAATACGACTCACTATAGGTGAGATTATCAAA AAGG ATCTTCACGTC	RV for PCR amplification of the sgRNA template for all the PAM identification processes except the first one (25nt long spacer)
	BG9309	AAAACGCCTAAGAGTGGGGAATG	RV for PCR amplification of the 3-hairpins long sgRNA template for all the PAM identification processes except for the first one
	BG9310	AAAAGGCGATAGGCGATCC	RV for PCR amplification of the 2-hairpins long sgRNA template for all the PAM identification processes except for the first one
	BG9311	AAAACGGGTCAGTCTGCCTATAG	RV for PCR amplification of the 1-hairpin long sgRNA template for all the PAM identification processes except for the first one
	BG9308	AAGCTTGAAATAATACGACTCACTATAGGTGAGATTATCAAA AAGG ATCTTCACGTC	pT7 and 25nt spacer sgRNA Fw

	Oligo	Sequence	Description
sgRNA module for *in vitro* transcription	BG10118	AAGCTTGAAATAATACGACTCACTATAGGAGATTATCAAAAA GGAT CTTCACGTCA	pT7 and 24nt spacer sgRNA Fw
	BG10119	AAGCTTGAAATAATACGACTCACTATAGGAAGATTATCAA AAAGGA TCTTCACGTCATAG	pT7 and 23nt spacer sgRNA Fw
	BG10120	AAGCTTGAAATAATACGACTCACTATAGGATTATCAAAAAGG ATCTT CACGTCATAGT	pT7 and 22nt spacer sgRNA Fw
	BG10121	AAGCTTGAAATAATACGACTCACTATAGGAATTATCAAAAA GGATCT TCACGTCATAGTT	pT7 and 21nt spacer sgRNA Fw
	BG10122	AAGCTTGAAATAATACGACTCACTATAGGTTATCAAAAAGGA TCTTC ACGTCATAGTT	pT7 and 20nt spacer sgRNA Fw
	BG10123	AAGCTTGAAATAATACGACTCACTATAGGTATCAAAAAGGA TCTTCA CGTCATAGTTC	pT7 and 19nt spacer sgRNA Fw
	BG10124	AAGCTTGAAATAATACGACTCACTATAGGATCAAAAAGGATC TTCAC GTCATAGTTC	pT7 and 18nt spacer sgRNA Fw
Editing and silencing constructs	BG9312	AAAACGCCTAAGAGTGGGGAATGCCCGAAGAAAGCGGGCGA TAGGCGATCC	3 loops sgRNA OH Rv
	BG8191	AAGCTTGGCGTAATCATGGTC	For the construction of the pThermoCas9_ctrl plasmid & pThermoCas9_bsΔpyrF1/2
	BG8192	TCATGAGTTCCCATGTTGTG	For the construction of the pThermoCas9_ctrl plasmid & pThermoCas9_bsΔpyrF1/2
	BG8194	tatggcgaatcacaacatgggaactcatgaGAACATCCTCTTTCTTAG	For the construction of the pThermoCas9_ctrl plasmid & pThermoCas9_bsΔpyrF1/2
	BG8195	gccgatatcaagaccgattttatacttcatTTAAGTTACCTCCTCGATTG	For the construction of the pThermoCas9_ctrl plasmid & pThermoCas9_bsΔpyrF1/2
	BG8196	ATGAAGTATAAAATCGGTCTTG	For the construction of the pThermoCas9_ctrl plasmid & pThermoCas9_bsΔpyrF1/2
	BG8197	TAACGGACGGATAGTTTC	For the construction of the pThermoCas9_ctrl plasmid & pThermoCas9_bsΔpyrF1/2
	BG8198	gaaagccggggaaactatccgtccgttataAATCAGACAAAATGGCCTGCTTA TG	For the construction of the pThermoCas9_ctrl plasmid & pThermoCas9_bsΔpyrF1/2
	BG8263	gaactatgacactttatttcagaatggacGTATAACGGTATCCATTTTAAGAAT AATCC	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_bsΔpyrF1/2
	BG8261	gaactatgacatcatggagttttaaatccaGTATAACGGTATCCATTTTAAGAAT AATCC	For the construction of the pThermoCas9_bsΔpyrF1
	BG8266	accgttatactggatttaaaactccatgatGTCATAGTTCCCCTGAGAT	For the construction of the pThermoCas9_bsΔpyrF2
	BG8317	gaactatgaccacccagcttacatcaacaaGTATAACGGTATCCATTTTAAGAA TAATCC	For the construction of the pThermoCas9_ΔbspyrF2
	BG8320	accgttatacttgttgatgtaagctgggtgGTCATAGTTCCCCTGAGAT	For the construction of the pThermoCas9_bsΔpyrF2

	Oligo	Sequence	Description
Editing and silencing constructs	BG9075	CTATCGGCATTACGTCTATC	For the construction of the pTherm oCas9i_ctrl
	BG9076	GCGTCGACTTCTGTATAGC	For the construction of the pTherm oCas9i_ctrl
	BG9091	TGAAGTATAAAATCGGTCTTGCTATCGGCATTACGTCTATC	For the construction of the pTherm oCas9i_ctrl
	BG9092	CAAGCTTCGGCTGTATGGAATCACAGCGTCGACTTCTGTATAGC	For the construction of the pTherm oCas9i_ctrl
	BG9077	GCTGTGATTCCATACAG	For the construction of the pTherm oCas9i_ctrl
	BG9267	GGTGCAGTAGGTTGCAGCTATGCTTGTATAACGGTATCCAT	For the construction of the pTherm oCas9i_ctrl
	BG9263	AAGCATAGCTGCAACCTACTGCACCGTCATAGTTCCCCTGAGATTAT CG	For the construction of the pTherm oCas9i_ctrl
	BG9088	TCATGACCAAAATCCCTTAACG	For the construction of the pTherm oCas9i_ctrl
	BG9089	TTAAGGGATTTTGGTCATGAGAACATCCTCTTTCTTAG	For the construction of the pTherm oCas9i_ctrl
	BG9090	GCAAGACCGATTTTATACTTCATTTAAG	For the construction of the pTherm oCas9i_ctrl
	BG9548	GGATCCCATGACGCTAGTATCCAGCTGGGTCATAGTTCCCCTGAGAT TATCG	For the construction of the pTherm oCas9i_ldhL
	BG9601	TTCAATATTTTTTTTGAATAAAAAATACGATACAATAAAAATGTCTA GAAAAAGATAAAAATG	For the construction of the pTherm oCas9i_ldhL
	BG9600	TTTTTTATTCAAAAAAAATATTGAATTTTAAAAATGATGGTGCTAGT ATGAAG	For the construction of the pTherm oCas9i_ldhL
	BG9549	CCAGCTGGATACTAGCGTCATGGGATCCGTATAACGGTATCCATTTT AAGAATAATCC	For the construction of the pTherm oCas9i_ldhL
	BG8552	TCGGGGGTTCGTTTCCCTTG	FW to check genom icpyrF deletion KO check
	BG8553	CTTACACAGCCAGTGACGGAAC	RV to check genom icpyrF deletion KO check
	BG2365	GCCGGCGTCCCGGAAAACGA	For the construction of the pTherm oCas9_ppΔpyrF
	BG2366	GCAGGTCGGGTTCCTCGCATCCATGCCCCCGAACT	For the construction of the pTherm oCas9_ppΔpyrF
	BG2367	ggcttcggaatcgttttccgggacgccggcACGGCATTGGCAAGGCCAAG	For the construction of the pTherm oCas9_ppΔpyrF
	BG2368	gacacaggcatcggtGCAGGGTCTCTTGGCAAGTC	For the construction of the pTherm oCas9_ppΔpyrF
	BG2369	gccaagagaccctgCACCGATGCCTGTGTCGAACC	For the construction of the pTherm oCas9_ppΔpyrF
	BG2370	cttggcggaaaacgtcaaggtctttttttacACGCGCATCAACTTCAAGGC	For the construction of the pTherm oCas9_ppΔpyrF
	BG2371	atgac**gagctgttcaccagcagcgc**TATTATTGAAGCATTTATCAGGG	For the construction of the pTherm oCas9_ppΔpyrF
	BG2372	GTAAAAAAGACCTTGACGTTTTC	For the construction of the pTherm oCas9_ppΔpyrF

	Oligo	Sequence	Description
Editing and silencing constructs	BG2373	tatgaagcgggccatTTGAAGACGAAAGGGCCTC	For the construction of the pThermoCas9_ppΔpyrF
	BG2374	taata**gcgctgctggtgaacagctc**GTCATAGTTCCCCTGAGATTATCG	For the construction of the pThermoCas9_ppΔpyrF
	BG2375	tggagtcatgaacatATGAAGTATAAAATCGGTCTTG	For the construction of the pThermoCas9_ppΔpyrF
	BG2376	ccctttcgtcttcAAATGGCCCGCTTCATAAGCAG	For the construction of the pThermoCas9_ppΔpyrF
	BG2377	gattttatacTTCATATGTTCATGACTCCATTATTATTG	For the construction of the pThermoCas9_ppΔpyrF
	BG2378	gggggcatggatgCGAGGAACCCGACCTGCATTGG	For the construction of the pThermoCas9_ppΔpyrF
	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	TGGACGTGTACTTCGACAAC	*thermocas9* seq. 1
	BG8197	TAACGGACGGATAGTTTC	*thermocas9* seq. 2
	BG6850	GCCTCATGAATGCAGCGATGGTCCGGTGTTC	*pyr*F 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
Cleavage assays	BG8363	ACGGTTATCCACAGAATCAG	FW for PCR linearization of DNA target
	BG8364	CGGGATTGACTTTTAAAAAAGG	RV for PCR linearization of DNA target
	BG9302	AAACTTCATTTTTAATTTAAAAGGATCTAGAACCCCCCGTGA AGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGA GTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAA	Non-template strand oligonucleotide for ssDNA cleavage assays
	BG9303	TTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGT TAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCAC CCCCCCAACTAG ATCCTTTTAAATTAAAAATGAAGTTT	Template strand oligonucleotide for ssDNA cleavage assays

	Oligo	Sequence	Description
	BG9304	TTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTT AAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACGG GGGGTTCTAG ATCCTTTTAAATTAAAAATGAAGTTT	Template strand oligonucleotide for ssDNA cleavage assays
ThermoCas9 expression	BG7886	TACTTCCAATCCAATGCAAAGTATAAAATCGGTCTTGATATCG	FW LIC_*thermocas9*
	BG7887	TTATCCACTTCCAATGTTATTATAACGGACGGATAGTTTCCCC GGCTTTC	RV LIC_*thermocas9*
RT-qPCR	BG9665	ATGACGAAAGGAGTTTCTTATTATG	RV qPCR check *ldhl*
	BG9666	AACGGTATTCCGTGATTAAG	FW qPCR check *ldhl*
***In vitro* ThermoCas9 Mismatch assay**	BG10561	CATGAGATTATCAAAAAGGATCTTCAACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatch at PAM proximal position 1
	BG10562	GATCCCTAGTTGGGGGGTTGAAGATCCTTTTTGATAATCT	RV for construction of *in vitro* target DNA with mismatch at PAM proximal position 1
	BG10563	CATGAGATTATCAAAAAGGATCTTCCCCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatch at PAM proximal position 2
	BG10564	GATCCCTAGTTGGGGGGGGGAAGATCCTTTTTGATAATCT	RV for construction of *in vitro* target DNA with mismatch at PAM proximal position 2
	BG10565	CATGAGATTATCAAAAAGGATCTTAACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatch at PAM proximal position 3
	BG10566	GATCCCTAGTTGGGGGGGTTAAGATCCTTTTTGATAATCT	RV for construction of *in vitro* target DNA with mismatch at PAM proximal position 3
	BG10567	CATGAGATTATCAAAAAGGATCTGCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatch at PAM proximal position 4
	BG10568	GATCCCTAGTTGGGGGGGTGCAGATCCTTTTTGATAATCT	RV for construction of *in vitro* target DNA with mismatch at PAM proximal position 4
	BG10569	CATGAGATTATCAAAAAGGATCGTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatch at PAM proximal position 5
	BG10570	GATCCCTAGTTGGGGGGGTGACGATCCTTTTTGATAATCT	RV for construction of *in vitro* target DNA with mismatch at PAM proximal position 5
	BG10571	CATGAGATTATCAAAAAGGATATTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatch at PAM proximal position 6

Oligo	Sequence	Description
BG10572	GATCCCTAGTTGGGGGGGTGAATATCCTTTTTGATAATCT	RV for construction of *in vitro* target DNA with mismatch at PAM proximal position 6
BG10573	CATGAGATTATCAAAAAGGAGCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatch at PAM proximal position 7
BG10574	GATCCCTAGTTGGGGGGGTGAAGCTCCTTTTTGATAATCT	RV for construction of *in vitro* target DNA with mismatch at PAM proximal position 7
BG10575	CATGAGATTATCAAAAAGGCTCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatch at PAM proximal position 8
BG10576	GATCCCTAGTTGGGGGGGTGAAGAGCCTTTTTGATAATCT	RV for construction of *in vitro* target DNA with mismatch at PAM proximal position 8
BG10577	CATGAGATTATCAAAAAGTATCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatch at PAM proximal position 9
BG10578	GATCCCTAGTTGGGGGGGTGAAGATACTTTTTGATAATCT	RV for construction of *in vitro* target DNA with mismatch at PAM proximal position 9
BG10579	CATGAGATTATCAAAAATGATCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatch at PAM proximal position 10
BG10580	GATCCCTAGTTGGGGGGGTGAAGATCATTTTTGATAATCT	RV for construction of *in vitro* target DNA with mismatch at PAM proximal position10
BG10581	CATGAGATTATCCAAAAGGATCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatch at PAM proximal position 15
BG10582	GATCCCTAGTTGGGGGGGTGAAGATCCTTTTGGATAATCT	RV for construction of *in vitro* target DNA with mismatch at PAM proximal position 15
BG10583	CATGAGAGTATCAAAAAGGATCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatch at PAM proximal position 20
BG10584	GATCCCTAGTTGGGGGGGTGAAGATCCTTTTTGATACTCT	RV for construction of *in vitro* target DNA with mismatch at PAM proximal position 20
BG10585	CATGAGAGGATCAAAAAGGATCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatch at PAM distal positions 19-20
BG10586	GATCCCTAGTTGGGGGGGTGAAGATCCTTTTTGATCCTCT	RV for construction of *in vitro* target DNA with mismatch at PAM distal position 19-20
BG10587	CATGAGAGGCGCAAAAAGGATCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatch at PAM distal positions 17-20
BG10588	GATCCCTAGTTGGGGGGGTGAAGATCCTTTTTGCGCCTCT	RV for construction of *in vitro* target DNA with mismatch at PAM distal position 17-20
BG10589	CATGAGAGGCGACAAAAGGATCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatch at PAM distal positions 15-20

	Oligo	Sequence	Description
***In vitro* ThermoCas9 Mismatch assay**	BG10590	GATCCCTAGTTGGGGGGGTGAAGATCCTTTTGTCGCCTCT	RV for construction of *in vitro* target DNA with mismatch at PAM distal position 15-20
	BG10591	CATGAGAGGCGACCCAAGGATCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatch at PAM distal positions 13-20
	BG10592	GATCCCTAGTTGGGGGGGTGAAGATCCTTGGGTCGCCTCT	RV for construction of *in vitro* target DNA with mismatch at PAM distal position 13-20
	BG10593	CATGAGAGGCGACCCCCGGATCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatch at PAM distal positions 11-20
	BG10594	GATCCCTAGTTGGGGGGGTGAAGATCCGGGGGTCGCCTCT	RV for construction of *in vitro* target DNA with mismatch at PAM distal position 11-20
	BG10595	CATGAGAGGCGACCCCCTTATCTTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatch at PAM distal position 9-20
	BG10596	GATCCCTAGTTGGGGGGGTGAAGATAAGGGGGTCGCCTCT	RV for construction of *in vitro* target DNA with mismatch at PAM distal position 9-20
	BG10597	CATGAGAGGCGACCCCCTTCGAGTCACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatch at PAM distal position 5-20
	BG10598	GATCCCTAGTTGGGGGGGTGACTCGAAGGGGGTCGCCTCT	RV for construction of *in vitro* target DNA with mismatch at PAM distal position 5-20
	BG10599	CATGAGAGGCGACCCCCTTCGAGGACACCCCCCCAACTAGG	FW for construction of *in vitro* target DNA with mismatch at PAM distal position 0-20
	BG10600	GATCCCTAGTTGGGGGGTGTCCTCGAAGGGGGTCGCCTCT	FW for construction of *in vitro* target DNA with mismatch at PAM distal position 0-20

Restriction sites are shown in italics. The PAMs are underlined. Spacer regions are shown in bold. Nucleotides in lowercase letters correspond to primer overhangs for HiFi DNA Assembly. LIC: Ligase Independent cloning; FW: Forward primer; RV: Reverse primer.

SUPPLEMENTARY TABLE 3 | PLASMIDS USED IN THIS STUDY, RELATED TO FIGURES 1 TO 4.

Plasmid	Description	Restriction sites used	Primers	Source
pNW33n	*E. coli-Bacillus* shuttle vector, cloning vector, Cam^R	-	-	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
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, Kan^R	-	-	EMD Millipore
pML-1B	*E. coli* Rosetta™ (DE3) plasmid, encodes rare tRNAs, Cam^R	-	-	Macrolab, Addgene
pEMG	*P. putida* suicide vector, used as template for replicon and Kan^R		See table S1	1
pSW_I-SceI	*P. putida* vector containing *I-SceI*, used as template for *xylS* and P_{Pm}		See table S1	1
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	-	-	2
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	cas9*dthermo*cas9 with N-term. His-tag and TEV cleavage site in pML-1B. Expression vector for catalytically inactive (dead) dThermoCas9	SspI and Ligase Independent Cloning	BG7886 and BG7888	This study

Plasmid	Description	Restriction sites used	Primers	Source
pNW-PAM7nt	Target sequence in pNW33n vector containing a 7-nt degenerate PAM for *in vitro* PAM determination assay	BamHI and BspHI	See table S1	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 S1	This study
pThermoCas9_ctrl	pNW33n with ThermoCas9-module[1] containing a non-targeting spacer. Used as a negative control	-	See table S1	This study
pThermoCas9_bsΔpyrF1	pNW33n with ThermoCas9-module[1] containing spacer 1 targeting the *pyrF* gene and the fused us+ds *pyrF*-flanks	-	See table S1	This study
pThermoCas9_bsΔpyrF2	pNW33n with ThermoCas9-module[1] containing spacer 2 targeting the *pyrF* gene and the fused us+ds *pyrF*-flanks	-	See table S1	This study
pThermoCas9i_ctrl	pNW33n with Thermo-dCas9-module[2] containing a non-targeting spacer. Used as a wild-type control	-	See table S1	This study
pThermoCas9i_ldhL	pNW33n with Thermo-dCas9-module[2] containing spacer 2 targeting the *ldhL* gene	-	See table S1	This study
pThermoCas9_ppΔpyrF	pEMG with ThermoCas9-module[3] for *Pseudomonas putida* containing a spacer targeting the a spacer targeting the *pyrF* gene and the fused us+ds *pyrF*-flanks	-	See table S1	This study

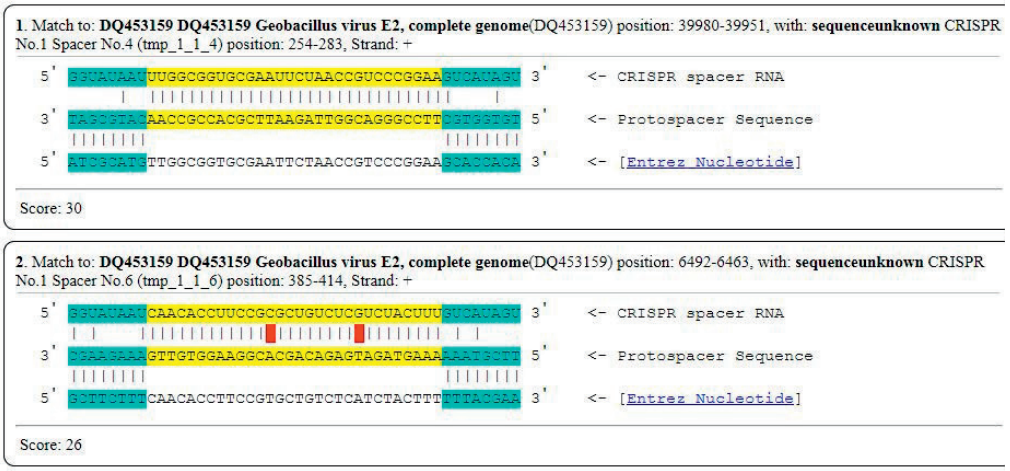
[1] The ThermoCas9 module contains *thermocas9* under the native P*xylL* promoter followed by the sgRNA under the *B. coagulans* P*pta* promoter (Figure 4).

[2] Like the ThermoCas9 module, but with the *thermo-dCas9* instead of *thermocas9* (Figure 4).

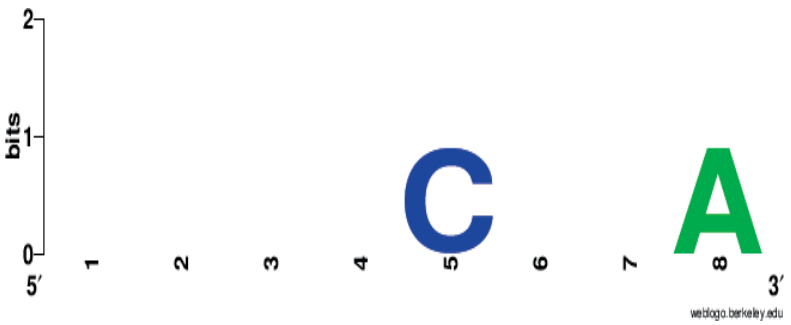
[3] The ThermoCas9 module for *Pseudomonas putida* contains *thermocas9* under the transcriptional control of the inducible Pm-XylS system followed by the sgRNA under the constitutive P3 promoter.

SUPPLEMENTARY FIG URES

A.

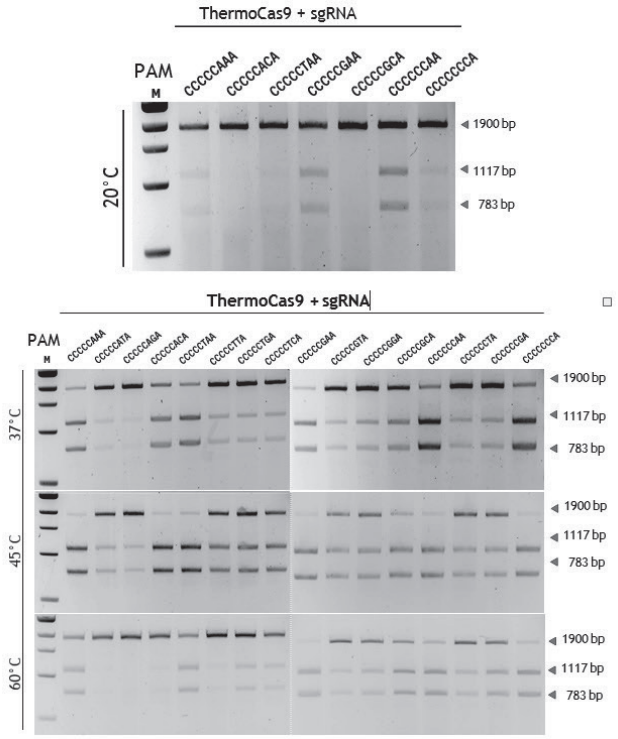


B.

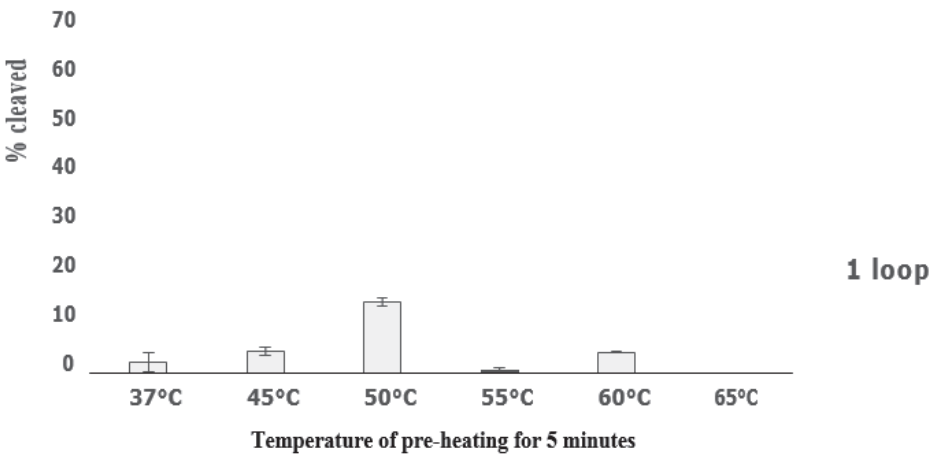


Supplementary Fig. 1 | *In silico* PAM determination results.

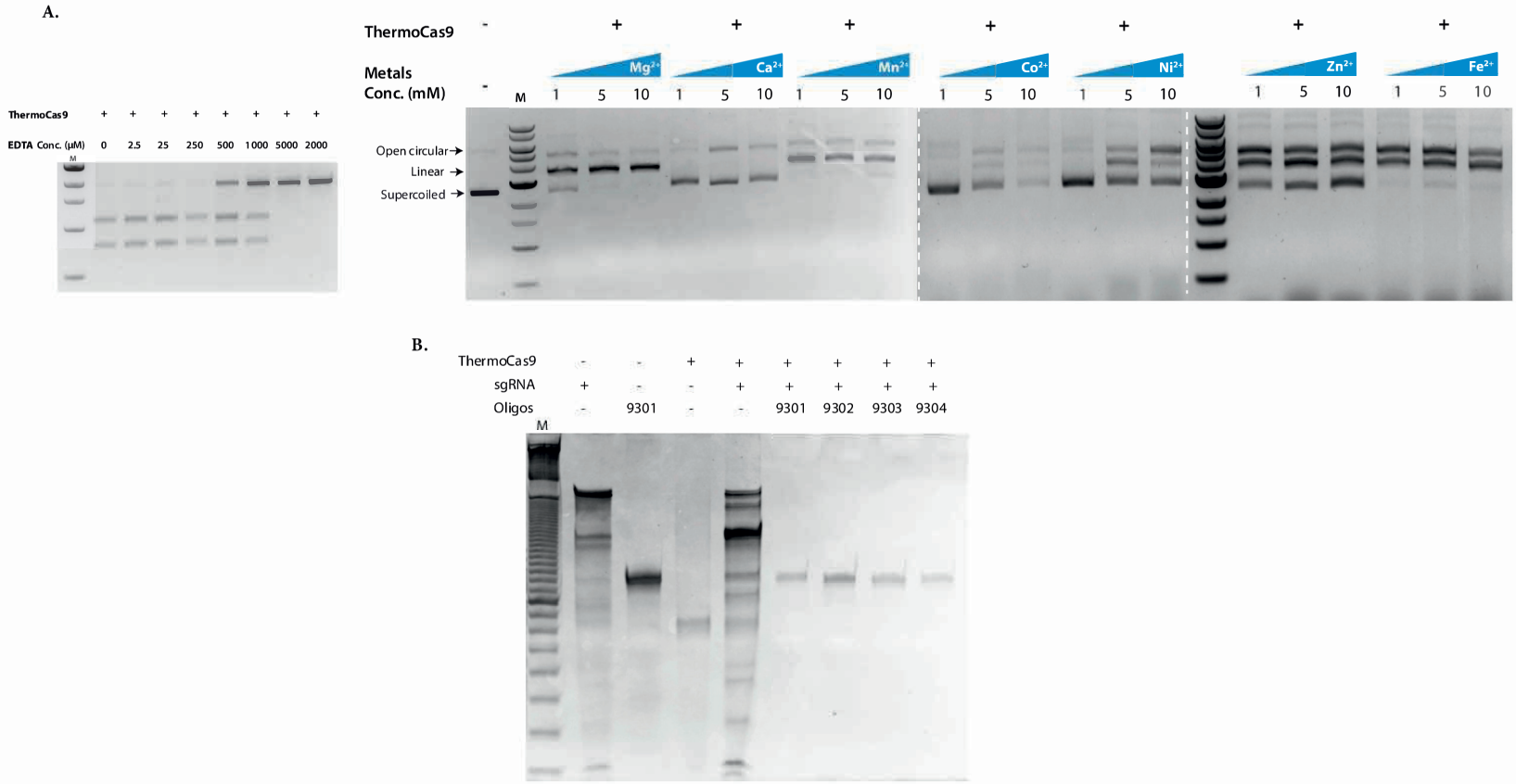
(A) The two hits obtained with phage genomes using CRISPRtarget[7].
(B) Sequence logo of the consensus 7-nt long PAM of ThermoCas9, obtained by *in silico* PAM analysis. Letter height at each position is measured by information content[8].



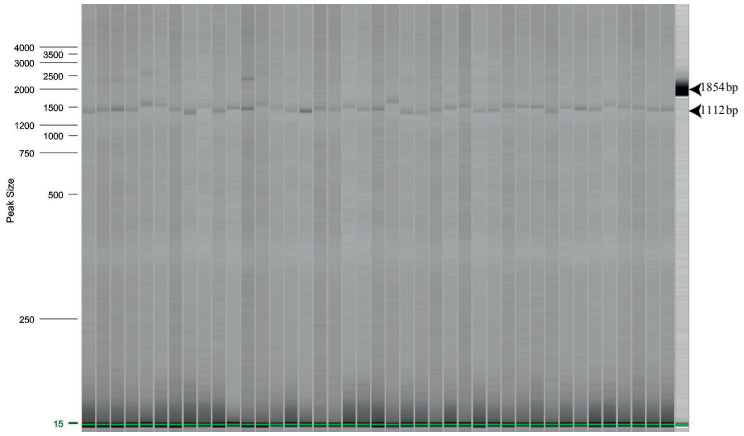
Supplementary Fig. 2 | ThermoCas9 PAM discovery. I*n vitro* cleavage assays for DNA targets with different PAMs at 20˚C, 37˚C, 45˚C and 60˚C. Seven (20˚C) or sixteen (37˚C, 45˚C, 60˚C) linearized plasmid targets, each containing a distinct 5'-CCCCCNNA-3' PAM, were incubated with ThermoCas9 and sgRNA, then analysed by agarose gel electrophoresis.



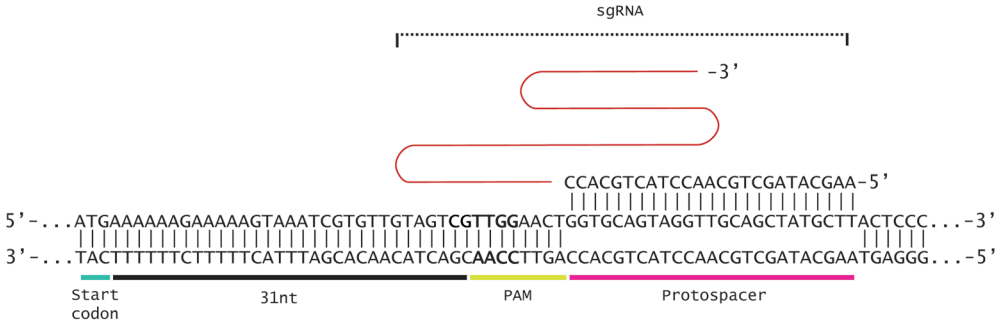
Supplementary Fig. 3 | Activity of ThermoCas9 at a wide temperature range using sgRNA containing one stem-loop. The importance of the predicted three stem loops of the tracrRNA scaffold was tested by transcribing truncated variations of the sgRNA and evaluating their ability to guide ThermoCas9 to cleave 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.



Supplementary Fig. 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.



Supplementary Fig. 5 | 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.



Supplementary Fig. 6 | 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.

SUPPLEMENTARY REFERENCES

1. Martinez-Garcia, E. & de Lorenzo, V. Engineering multiple genomic deletions in Gram-negative bacteria: analysis of the multi-resistant antibiotic profile of Pseudomonas putida KT2440. Environ. Microbiol. 13, 2702–2716 (2011).

2. Mougiakos, I. et al. Efficient genome editing of a facultative thermophile using mesophilic spCas9. ACS Synth. Biol. 6, 849–861 (2017).

3. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947– 2948 (2007).

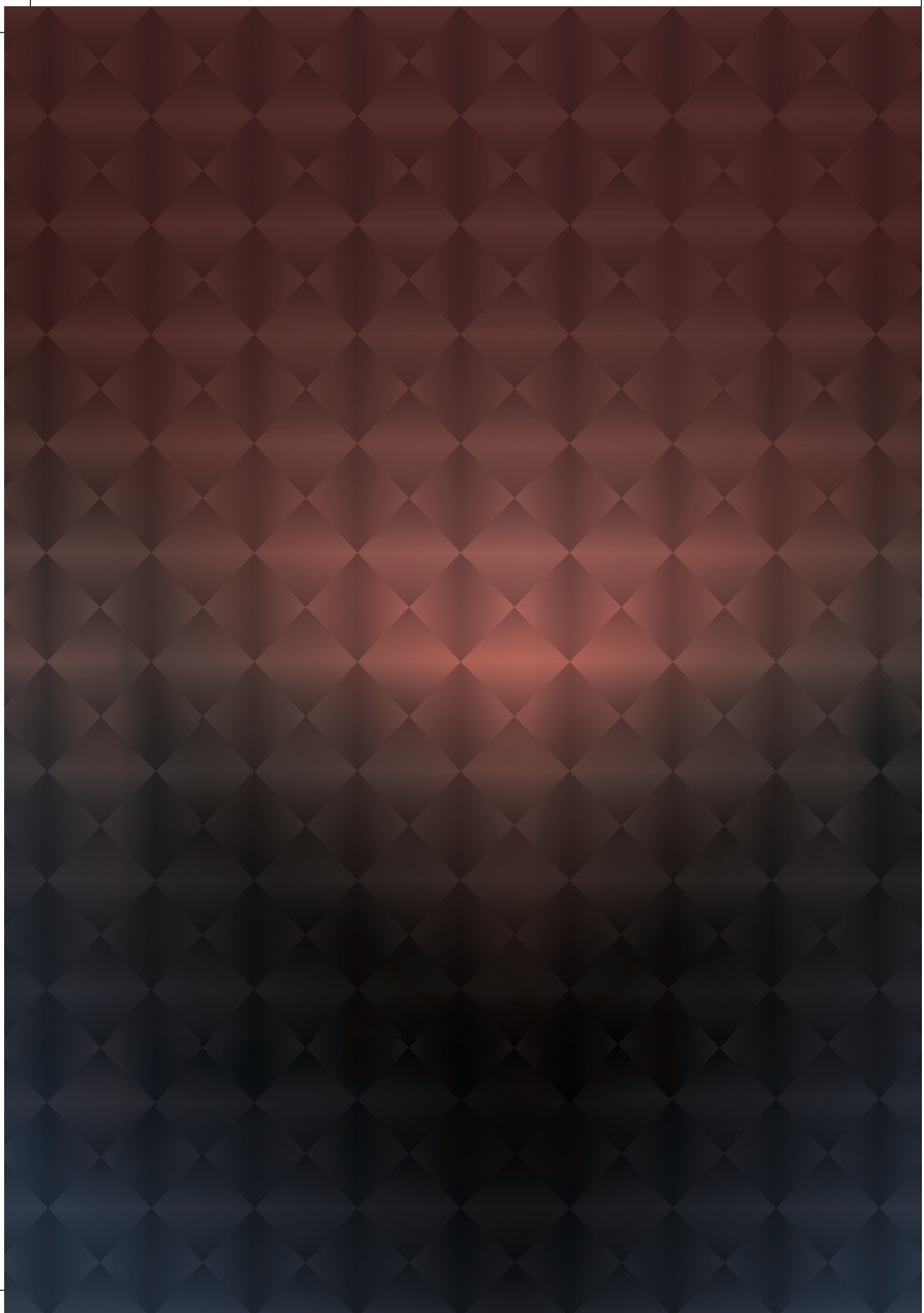
4. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874 (2016).

5. Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics 15, 305–8 (1999).

6. Jinek, M. et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science (80-.). 343, 1247997–1247997 (2014).

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

8. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A sequence logo generator. Genome Res. 14, 1188–1190(2004).



CHAPTER 6

CHARACTERIZING AN ANTI-CRISPR-BASED ON/OFF SWITCH FOR BACTERIAL GENOME ENGINEERING

Despoina Trasanidou[1#], Prarthana Mohanraju[1#], Richard van Kranenburg[1,2], Ioannis Mougiakos[1*]& John van der Oost[1*]

[1]Laboratory of Microbiology, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, The Netherlands.

[2]Corbion, Arkelsedijk 46, 4206 AC Gorinchem, The Netherlands.

[#]Contributed equally

*Corresponding authors

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* ($AcrIIC1_{Nme}$) 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°C and can be used for introducing dsDNA breaks in *E. coli*, in a tunable and spacer-dependent manner. In addition, we show that $AcrIIC1_{Nme}$ 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 (Thermo-dCas9 and Geo-dCas9). Moreover, we describe a single-vector, tightly controllable and highly efficient Cas9/$AcrIIC1_{Nme}$-based tool for silencing in bacteria. Altogether, an anti-CRISPR protein has been used to control Cas9-based genome editing and transcriptional regulation.

KEYWORDS

ThermoCas9, GeoCas9, antiCRISPR, bacteria, genome engineering

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[1]. 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)[2-5]. 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[5,6]. 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[5,7-9]. 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[10-12]. 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[11,13,14]. 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, SpyCas9)[10-12], and (ii) a programmable single-guide RNA (sgRNA) molecule[15]. The sgRNA guides the Cas9 nuclease to introduce double-stranded DNA breaks (DSDBs) to a desired DNA sequence (protospacer), which is complementary to the exchangeable 5'-end of the sgRNA (spacer)[16] and flanked downstream by a 3-8nt long conserved motif (protospacer adjacent motif, PAM)[17,18]. Unlike eukaryotes, Cas9-mediated DSDBs 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[19-22]. 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[10,23,24]. The Cas9 endonuclease either triggers cellular repair mechanisms, usually resulting in escape mutants and mixed wild type-mutant genotypes[25-27], 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[28-31]. Due to the absence of well-characterized inducible promoters in non-model organisms[32], 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 SpyCas9 in a moderate thermophile[33], and through the development of synthetic Cas9 variants regulated by drugs, light or other external stimuli[34-42].

Intriguingly, several small (50 - 150 amino acids) proteins naturally neutralizing CRISPR-Cas systems were recently identified in prokaryotic viruses and mobile genetic elements[43-50], launching the now rapidly expanding anti-CRISPR field. Out of the currently thirty-seven identified anti-CRISPR proteins, ten have been characterized 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)[48,51-55], by masking of the PAM-interacting domain of Cas9 (AcrIIA2, AcrIIA4)[56-59] or by forcing dimerization of the Cas9 (AcrIIC3)[46,60]. Such proteins have been successfully applied to limit off-target editing[46,57,61], control leaky CRISPR interference (CRISPRi)[46], and optimize CRISPR-based imaging[46] in human cells as well as to inactivate a gene drive in budding yeast[62], and to hamper CRISPRi in *E. coli*[61]. 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 Cas 3 endonuclease to the DNA-associated Cascade[51,63] 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[60]. 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[46,60] 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 G*eobacillus stearothermophilus* (GeoCas9), which are functional at wide-temperature range, both efficiently introduce lethal DSDBs 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 (Geo-dCas9, Thermo-dCas9). 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 characterized *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[64]. We further employed ThermoCas9 to develop the first Cas9-based genome editing and silencing tools for thermophilic bacteria, readily applicable to mesophilic bacteria too[64]. ThermoCas9 and its catalytically inactive mutant Thermo-dCas9 were applied for gene deletion and transcriptional silencing, respectively, in *Bacillus smithii* ET138 (55°C) as well as for gene deletion in *Pseudomonas putida* KT2440 (37°C)[64]. 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 characterized *in vitro* and employed for *in vitro* DNA targeting in human plasma[65], for gene disruption in mammalian cells[65] and for *in vivo* targeting of viral DNA during *E. coli* infection[60]. The wide-temperature activity range (20-70° C *in vitro*), the recalcitrance and the relatively small sizes of ThermoCas9 and GeoCas9 (Sup. Fig. S1a) 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°C. 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, $P_{tet\text{-}lac}$). 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) (Fig. 1; Sup. Fig. S1b). 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 (Fig. 1a), and the fourth spacer was scrambled, not targeting any region in the genome of the *E. coli* DH10B*_gfp* strain (Sup. Table S1). 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 i*n vitro*[64], 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 (Fig. 1; Sup. Table S2). 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 (Fig. 1b; Sup. Fig. 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 (Fig. 1c; Sup. Fig. 2b). However, the GeoCas9 plasmid with the spacer 3 cause d 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.

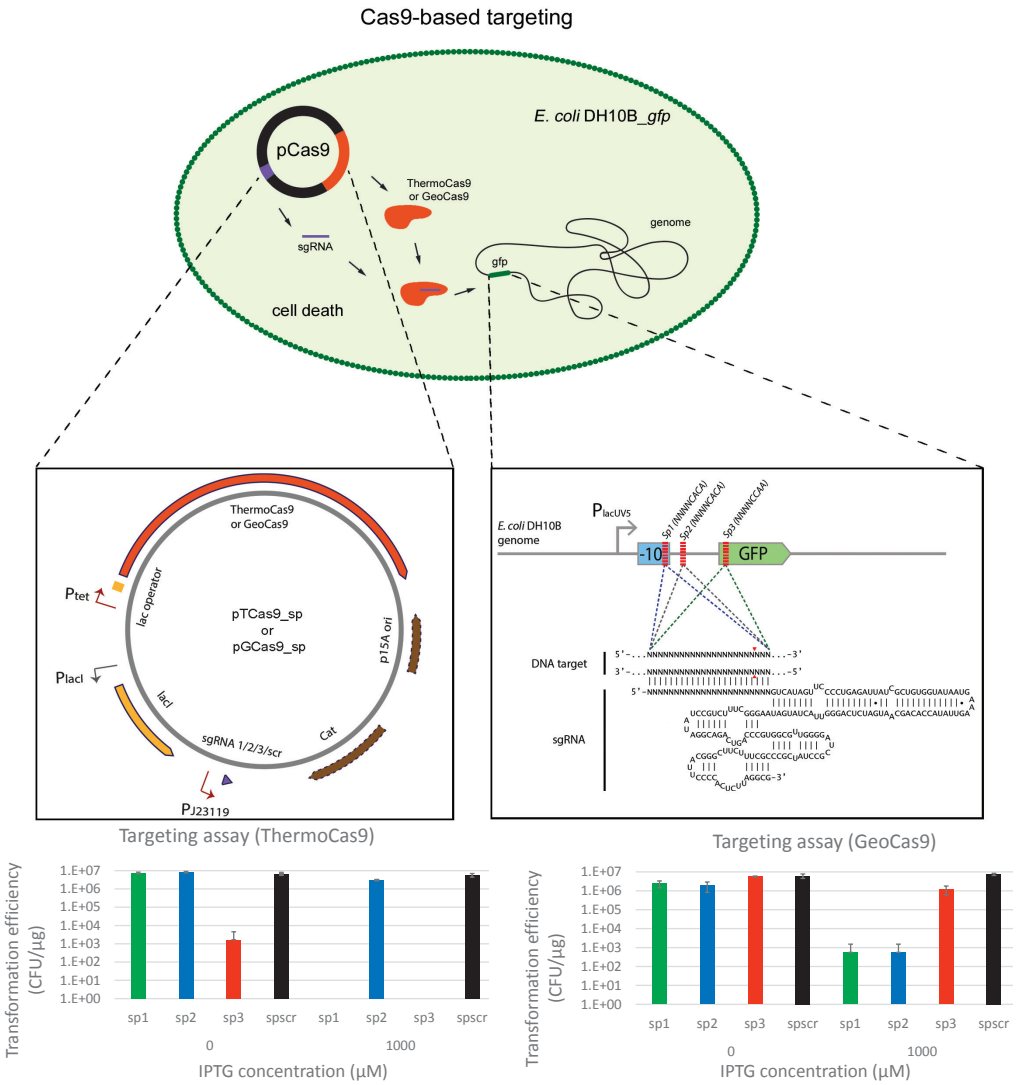


Figure 1 | **ThermoCas9- and G eoCas9-based targeting in *E. coli* DH 10B_*gfp*. a)** Schematic illustration of the strategy for ThermoCas9- and GeoCas9-based targeting in *E. coli* DH10B_gfp. The pCas9 plasmid was used for the expression of the ThermoCas9/GeoCas9 nuclease and its respective sgRNA to 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.

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[60], 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[60]. 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[60]. 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 (Sup. Fig. S3). 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 (Fig. 2a). The pAcr was constructed by cloning the *E. coli* codon optimized *acrIIC1$_{Nme}$* gene in the high copy number pUC19 backbone plasmid under the transcriptional control of the L-rhamnose-inducible promoter (P$_{rha}$) (Fig. 2a). This tactic was expected 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 (Fig. 2a), as well as the pTCas9L551H_sp3 plasmid that is similar to the pTCas9_sp3 plasmid but encompasses the *thermoCas9$_{L551H}$* gene that codes for the ThermoCas9$_{L551H}$ variant, which harbors the L551H substitution (Sup. -

Fig. S4a). Assays were conducted with and without L-rhamnose supplementation (Sup. Fig. S4a). 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 (Fig. 2b,c; Sup. Fig. S4a,b; Sup. Fig. S5a,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-tunable manner.

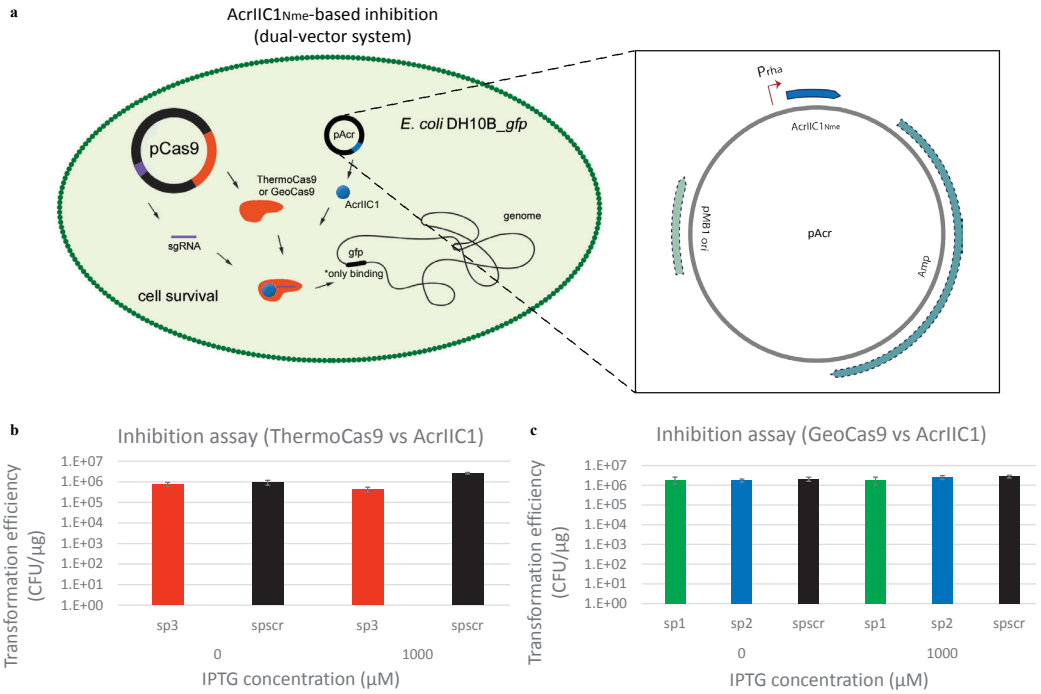


Figure 2 | AcrIIC1$_{Nme}$-mediated inhibition of ThermoCas9- and GeoCas9-based targeting in *E. coli* DH10B*_gfp* (dual-vector system). a) Schematic illustration of the dual-vector strategy for AcrIIC1$_{Nme}$-mediated inhibition of ThermoCas9- and GeoCas9-based targeting in *E. coli* DH10B*_gfp*. The 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, P_{lacUV5}). 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.

THE 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$_{D8A,H582A}$ (Thermo-dCas9) and GeoCas9$_{D8A,H582A}$ (Geo-dCas9) 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 *thermo-dcas9* and *geo-dCas9* 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, Thermo-dCas9 and Geo-dCas9 expression (Fig. 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_i(pACYC184), *E. coli* DH10B_*gfp*(pAcr, pTCas9_spscr), *E. coli* DH10B_*gfp*(pAcr, pTCas9L551H_spscr), *E. coli* DH10B_*gfp*(pAcr, pGCas9_spscr), *E. coli* DH10B_*gfp*(pT-dCas9_spscr), and *E. coli* DH10B_*gfp*(pG-dCas9_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-AcrIIC1$_{Nme}$ 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 (Fig. 3a; Sup. Fig. 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 pT-dCas9_sp3 plasmid under non-induced conditions, absolute fluorescence loss was observed (Fig. 3b; Sup. Fig. 8a). Hence, the ThermoCas9-AcrIIC1$_{Nme}$, and the ThermoCas9$_{L551H}$-AcrIIC1$_{Nme}$ complexes mediated sile-

ncing of the *gfp* gene with the same efficiency as Thermo-dCas9 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 (Fig. 3c; Sup. Fig. 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 pG-dCas9_sp1 and 2, separately and in absence of induction of Geo-dCas9 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*(pG-dCas9_spscr) control strain (Fig. 3d; Sup. Fig. 8b). When the Geo-dCas9 expression was 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*(pG-dCas9_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 Geo-dCas9 for the spacer 1. However, regarding spacer 2, Geo-dCas9-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 Geo-dCas9 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 AcrIIC1$_{Nme}$ via L-rhamnose supplementation. Notably, no substantial difference was observed in the results of these assays compared to the assays without L-rhamnose supplementation (Sup. Fig. S6b; Sup. Fig. S7b). Moreover, single cell populations were always observed for all the previously described

flow cytometry-based fluorescence loss assays (Sup. Fig. S9), 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 (Sup. Fig. S10; S11; S12).

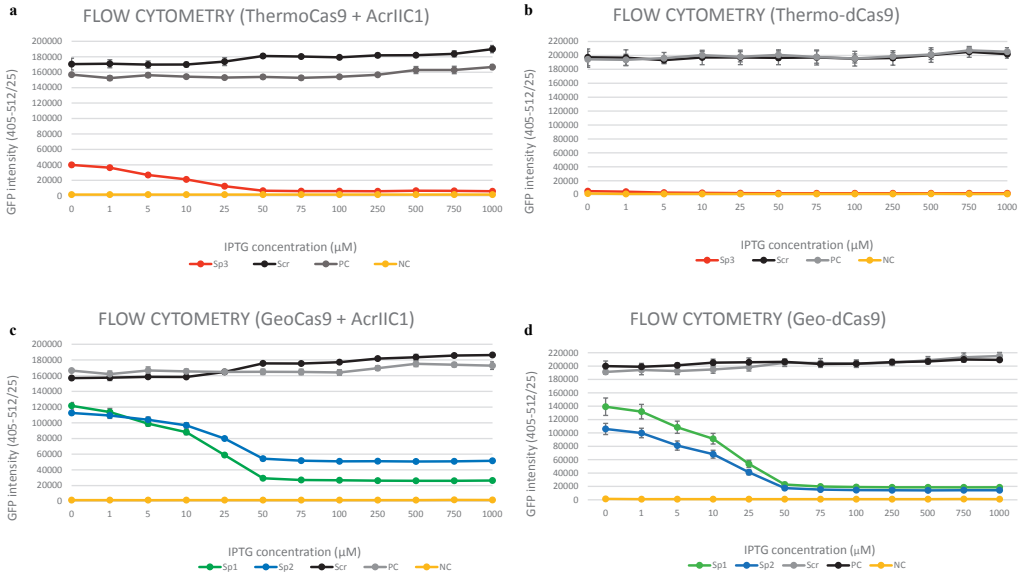


Figure 3 | Comparison of the ThermoCas9/GeoCas9:AcrIIC1$_{Nme}$-based and the Thermo-dCas9/Geo-dCas9-based silencing in *E. coli* DH10B_*gfp* (flow cytometry-based fluorescence loss assays). a) ThermoCas9:AcrIIC1$_{Nme}$-mediated silencing/fluorescence downregulation assays in *E. coli* DH10B_*gfp* cells 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 pT-dCas9_sp3 or the pT-dCas9_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. **b)** GeodCas9-mediated silencing/fluorescence downregulation assays in *E. coli* DH10B_*gfp* cells harbouring either the pG-dCas9_sp1 or the pG-dCas9_sp2 or the pT-dCas9_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.

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-AcrCII1$_{Nme}$-mediated targeting and targeting inhibition in *E. coli*. To this end, we added the *E. coli* codon-optimized *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 (Fig. 4a).

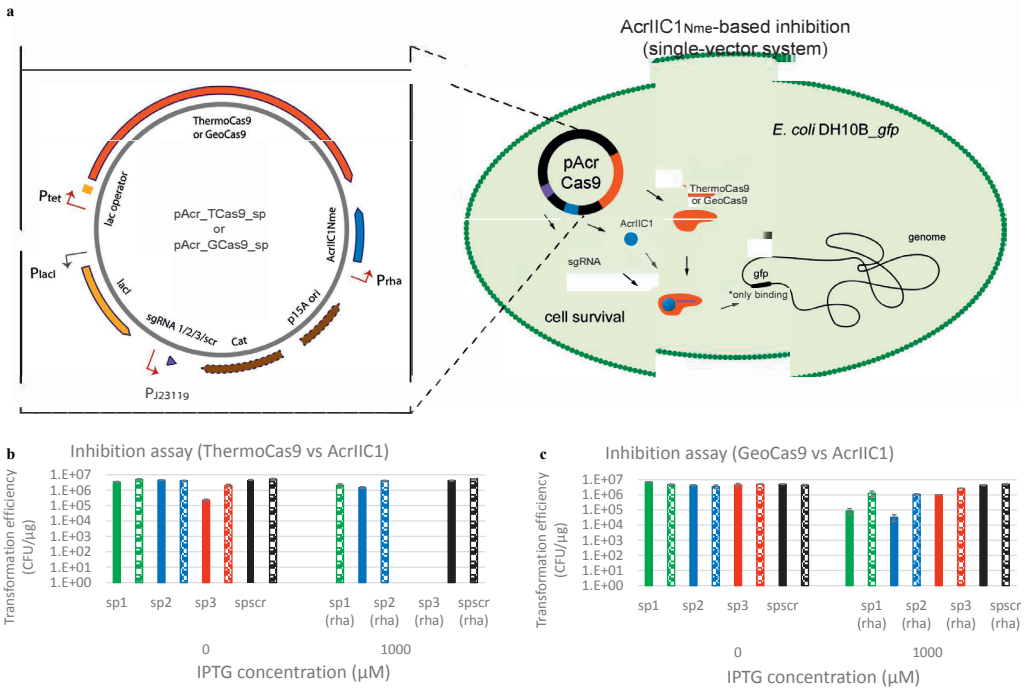


Figure 4 | AcrIIC1$_{Nme}$-mediated inhibition of ThermoCas9- and GeoCas9-based targeting in *E. coli* DH 10B_*gfp* (single-vector system). a) Schematic illustration of the single-vector strategy for AcrIIC1$_{Nme}$-mediated inhibition ofThermoCas9- and GeoCas9-based targeting in *E. coli* DH10B_*gfp*. 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 rhamnose supplementation are depicted with dotted bars.

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 (Fig. 4b; Sup. Fig. 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 suppleme-

ntation 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 (Fig. 4b; Sup. Fig. 13a). The transformation efficiency of the cells transformed with the pAcr_TCas9_sp3 plasmi d 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 (Fig. 5c). Interestingly, there was no loss of 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 (Fig. 4c; Sup. Fig. 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 (Fig. 5d). 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 (Fig. 4c ; Sup. Fig. 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.

DISCUSSION

In this study, we demonstrate that the thermostable ThermoCas9 and GeoCas9 endonucleases are highly active *in vivo* at 37°C and introduce lethal DSDBs 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 the levels *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* and the silencing effect was proved to be spacer-dependent, tuneable and

comparable to the silencing effect of the catalytically inactive ("dead") variants Thermo-dCas9 and Geo-dCas9, 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[10-12]. The limitation of this approach, for the majority of microorganisms, is that the Cas9-based introduction of lethal DSDBs often precedents the incorporation of the desired modifications in a genome, resulting in the dramatic reduction of the surviving mutants 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*$_{Nme}$ 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 DSDBs, 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 mRNAs, leading to ThermoCas9/GeoCas9 expression levels that prevail over the antiCRISPR-mediated inhibition, allowing the endonuclease to act as a stringent cou-

nter-selection system. We plan to establish the proof of principle approach in *E. coli*.

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*[66,67], *Brucella melitensis*[67], *Staphylococcus aureus*[68], *Corynebacterium glutamicum*[69] and several *Pseudomonas* species[70], employing a fusion of catalytically inactivated or nickase SpyCas9 and cytidine deaminase (PmCDA1; APOBEC1) that converts cytosine (C) to uracil (U) at spacer-targeted sites without DSDBs, active Thermo-Cas9/ GeoCas9-AcrIIC1$_{Nme}$ may be applied instead of the catalytically deficient SpyCas9 variants to broaden the spectrum of DNA targets (different PAMs) and reduce the off-target effects (stringent PAM recognition at 37°C). As such, a ThermoCas9/GeoCas9-cytidine deaminase fusion can be generated and trapped by AcrIIC1Nme 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 favor 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 for 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 characterized, 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 relevant bacteria, is now closer than ever before.

METHODS

BACTERIAL STRAINS AND GROWTH CONDITIONS

Bacterial strains used in this study are listed in Sup. Table S3. All *E. coli* strains were cultured in Luria-Bertani broth (LB) or on LB agar plates, supplemented with chloramphenicol (15 μg/ml), and/or ampicillin (100 μg/ml). 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, primers/oligonucleotides, and Cas9/anti-CRISPR genes used in this study are presented in Supplementary Table S2, S4, and S5, 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[71] and plated on LB agar containing chloramphenicol (15 μg/ml) or ampicillin (100 μg/ml) 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 Supplementary Table S2. 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°C for 5 min, and slowly cooled to room temperature.

TARGETING ASSAYS

To target bacterial DNA, chemically competent *E. coli* DH10B cells[71] harbouring a genome-integrated *gfp* gene under the transcriptional control of P_{lacUV5} (*E. coli* DH10B_*gfp*) were transformed[71] 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_{lacUV5} (pTCas9_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) and different IPTG concentrations (0 - 1000 μM) for 17 h at 37°C. Colony forming units (CFUs) were counted after plating undilutedbiological 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[71] 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), chloramphenicol (15 μg/mL), different IPTG concentrations (0 - 1000 μM), and L-rhamnose (0 or 0.2%) for 17 h at 37°C. 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[71] 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) and chloramphenicol (15 µg/mL) in a Masterblock® 96-well deep microplate (Greiner Bio-One) for 22 h at 37°C 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 aMasterblock® 96-well deep microplate (Greiner Bio-One), and incubated for 22 h at 37°C with vigorous shaking (900 rpm). The third day, 40 µl of overnight cultures were diluted i n 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:$AcrII_{Nme}$ to that of the dead variants, *E. coli* DH10B_gfp cells were heat-transformed[71] 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} (pT-dCas9_sp1), the region immediately after the -10 sequence (pT-dCas9_sp2) or the start of the gfp gene (pT-dCas9_sp3). As a positive control, 3 ng of empty vector (pACYC184) was used while as negative controls 3 ng of the non-targeting plasmid (pT-dCas9_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 containi ng chloramphenicol (15 µg/mL) in a Masterblock® 96-well deep microplate (Greiner Bio-One) for 22 h at 37°C 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°C 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 SynergyM

x Monochromator-Based Multi-Mode Micro-plate reader (BioTek). The plasmid pGeodCas9_sp1, pGeodCas9_sp2, pGeodCas9_sp3, pGeodCas9_spscr were used for assessing the fluorescence loss by Geo_dCas9 following the previously described approach.

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[71] with equal amounts (5 ng) of plasmids expressing AcrIIC1$_{Nme}$, ThermoCas9, 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), different IPTG concentrations (0 - 1000 µM), and L-rhamnose (0 or 0.2%) for 17 h at 37°C. 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.

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.

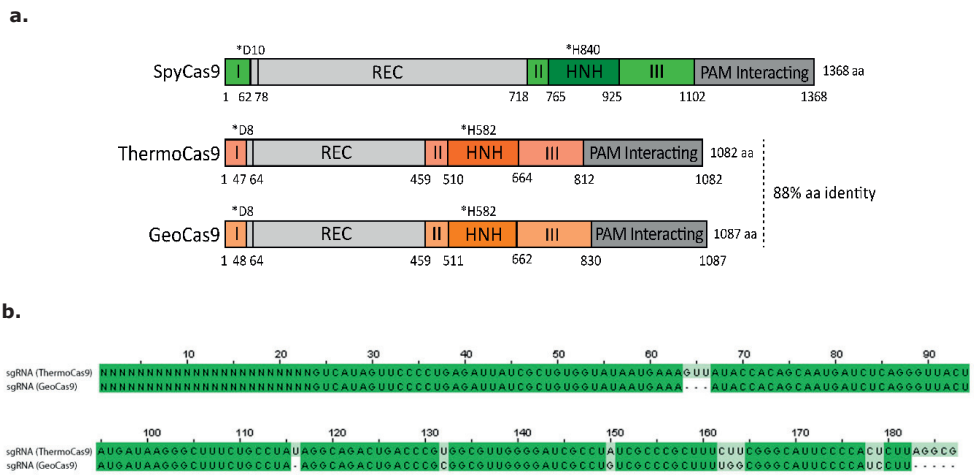
COMPETING INTERESTS

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

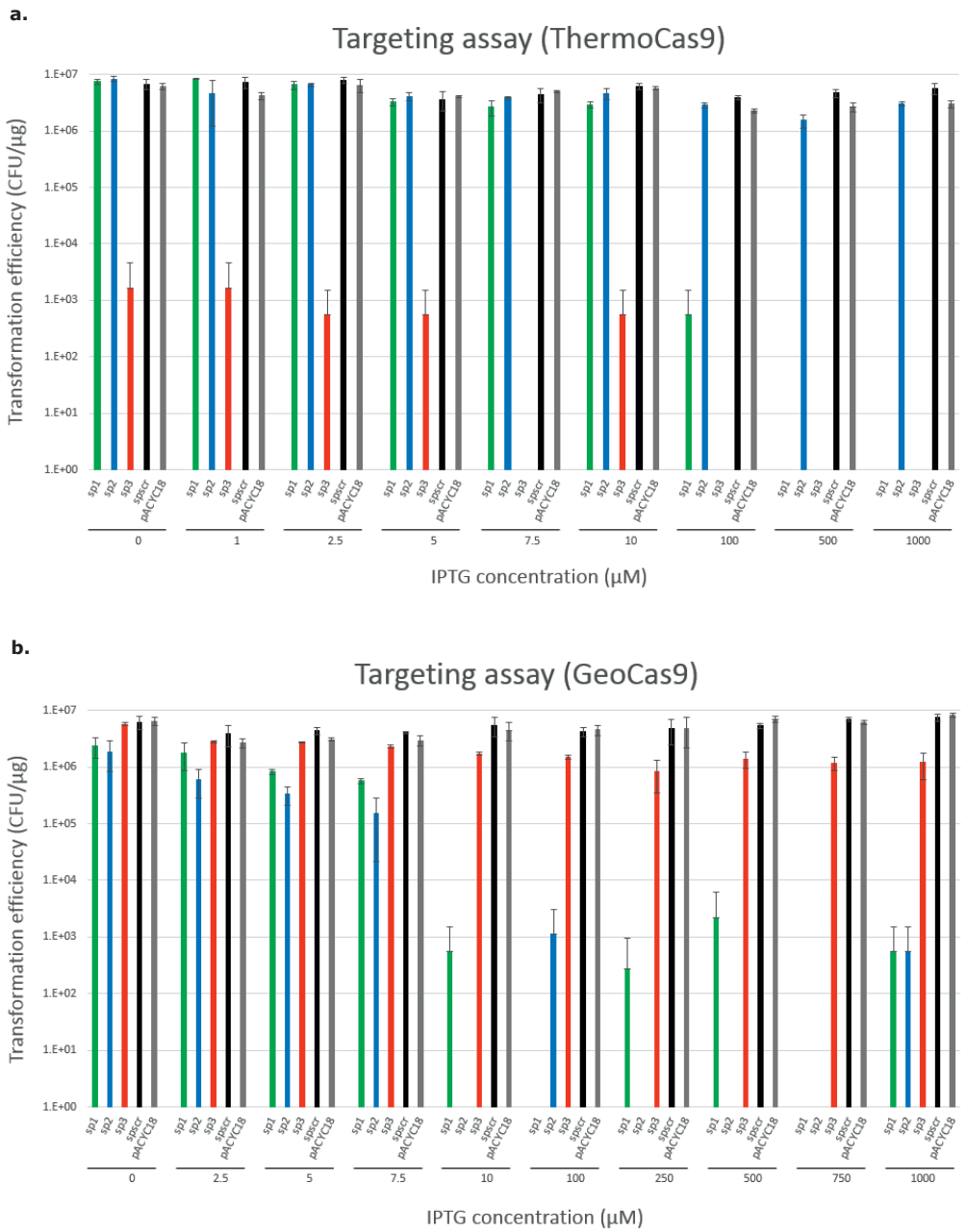
ACKNOWLEDGEMENTS

We would like to thank Nieuwkoop, Thijs 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.

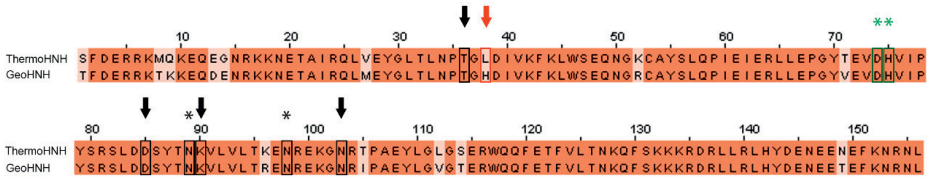
Supplementary Figures S1-S13



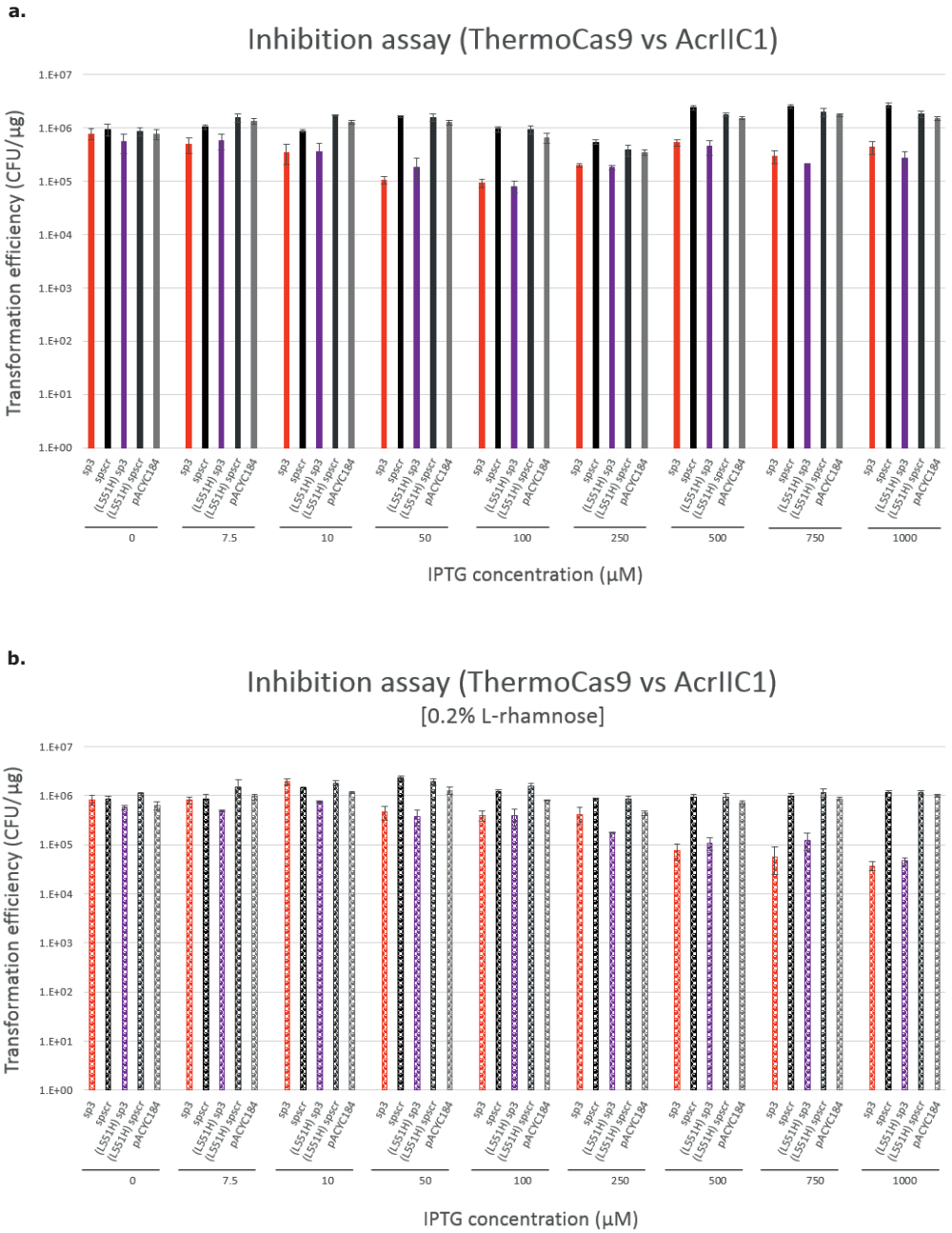
Sup. Fig. S1 | Comparison of the domain architecture and the sgRNA modules of ThermoCas9 and GeoCas9. a. Schematic illustration of the domain architecture of SpyCas9 (green), ThermoCas9 (dark orange), and GeoCas9 (light orange). The active site residues are indicated with asterisks. **b.** Sequence alignment of the sgRNA modules of ThermoCas9 and GeoCas9, used in previous studies for *in vivo* experiments[62,63]. Dark green colour indicates identical nucleotides, while light green colour represents non-identical regions.



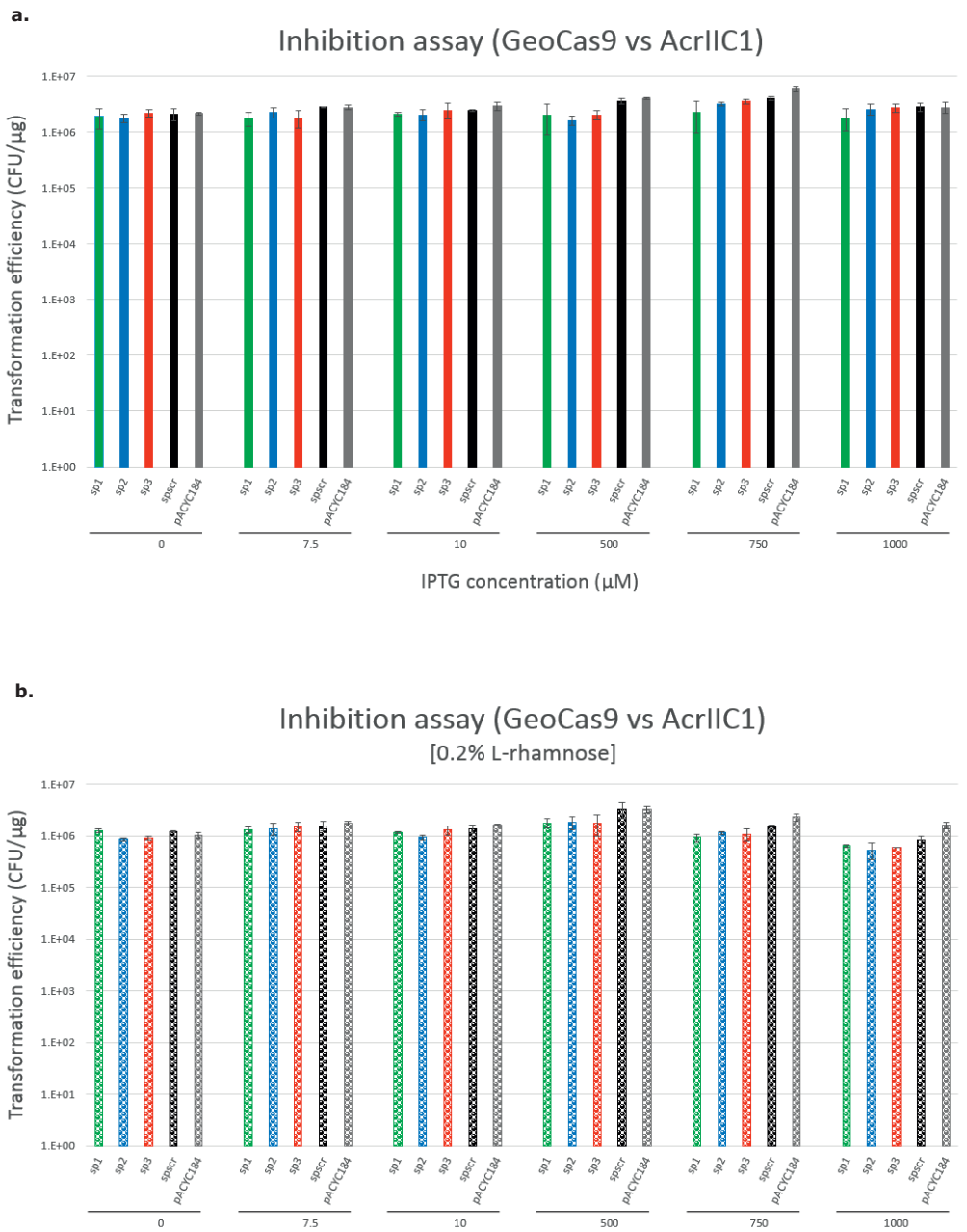
Sup. Fig. S2 | Comparison of ThermoCas9- and GeoCas9-based targeting in *E. coli* DH 10B_*gfp*. a. Complete data for ThermoCas9-based targeting in *E. coli* DH10B_*gfp* including the whole range of IPTG concentrations used for the induction of the ThermoCas9 expression. **b.** Complete data for GeoCas9-based targeting in *E. coli* DH10B_gfp including the whole range of IPTG concentrations used for the induction of the GeoCas9 expression.



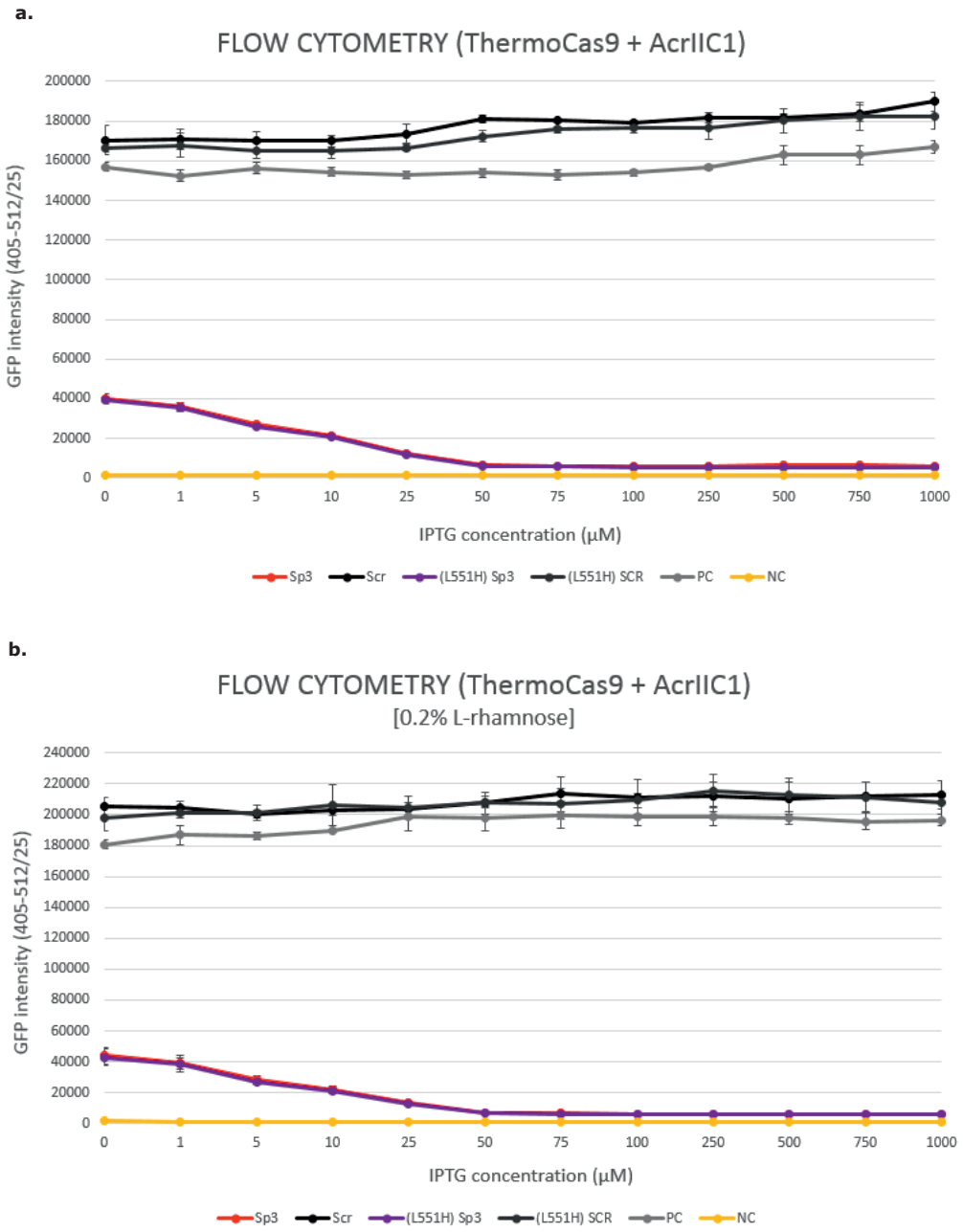
Sup. Fig. S3 | Amino acid sequence alignment of the HNH domain of ThermoCas9 and GeoCas9. Dark orange colour indicates identical amino acid, while light orange colour represents non-identical residues. Asterisks indicate catalytic residues, and, specifically, green asterisks represent catalytic residues 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.



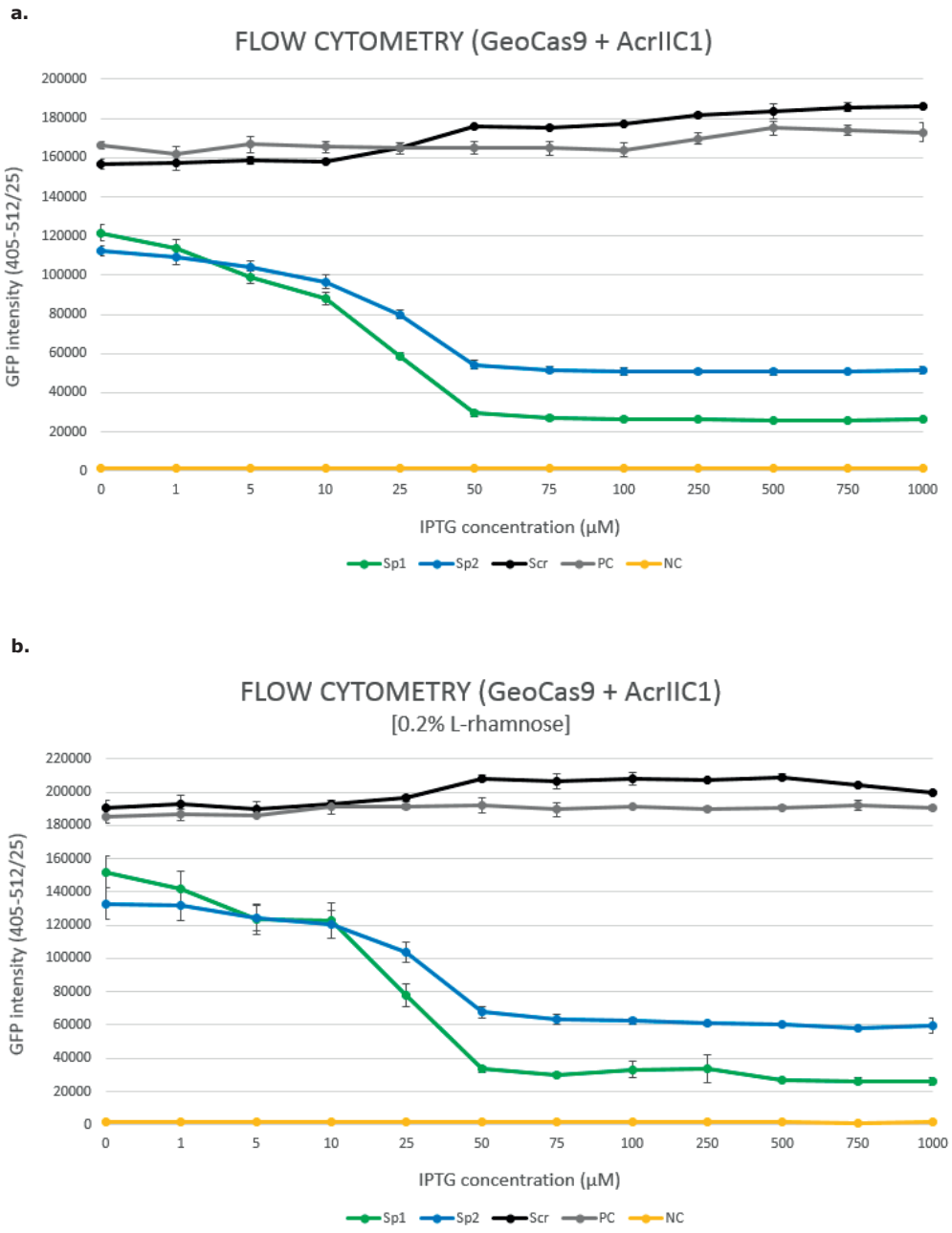
Sup. Fig. S4 | **AcrIIC1$_{Nme}$-mediated inhibition of ThermoCas9 and ThermoCas9$_{L551H}$ in *E. coli* DH 10B_*gfp* (dual-vector system). a.** Complete data for inhibition of ThermoCas9 and ThermoCas9$_{L551H}$ cleavage activity, when the expression of AcrIIC1$_{Nme}$ is not induced (no supplementation of L-rhamnose). **b.** Complete data for inhibition of ThermoCas9 and ThermoCas9$_{L551H}$ cleavage activity, when the expression of AcrIIC1$_{Nme}$ is induced (supplementation of 0.2% L-rhamnose).



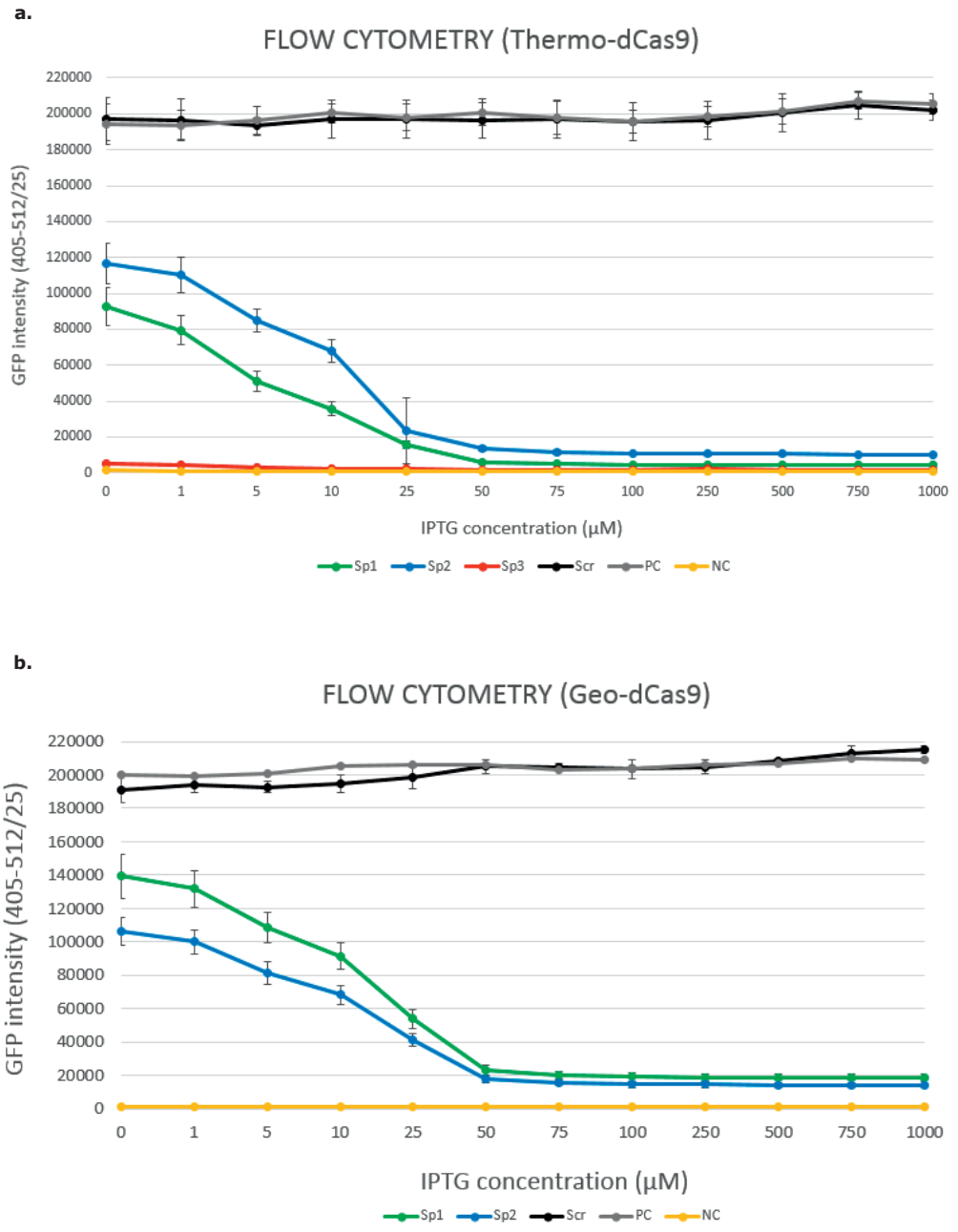
Sup. Fig. S5 | AcrIIC1$_{Nme}$-mediated inhibition of GeoCas9 in *E. coli* DH10B *gfp* (dual-vector system). **a.** Complete data for inhibition of GeoCas9 cleavage activity, when the expression of AcrIIC1$_{Nme}$ is not induced (no supplementation of L-rhamnose). **b.** Complete data for inhibition of GeoCas9 cleavage activity, when the expression of AcrIIC1$_{Nme}$ is induced (supplementation of 0.2% L-rhamnose).



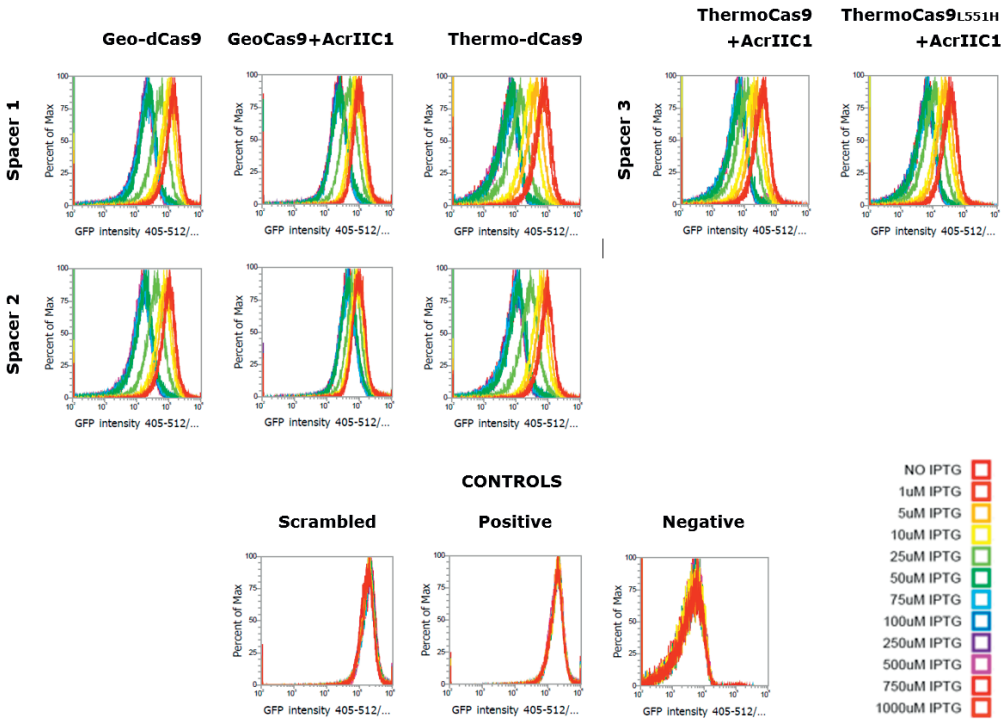
Sup. Fig. S6 | ThermoCas9/ThermoCas9$_{L551H}$:AcrIIC1$_{Nme}$-based silencing in *E. coli* DH10B_*gfp* (flow cytometry-based fluorescence loss assays). a. Complete data for Thermo Cas9/ThermoCas9$_{L551H}$–mediated silencing, when the expression of AcrIIC1$_{Nme}$ is not induced (no supplementation of L-rhamnose). **b.** Complete data for ThermoCas9/Th ermoCas9$_{L551H}$–mediated silencing, when the expression of AcrIIC1$_{Nme}$ is induced (supplementation of 0.2% L-rhamnose).



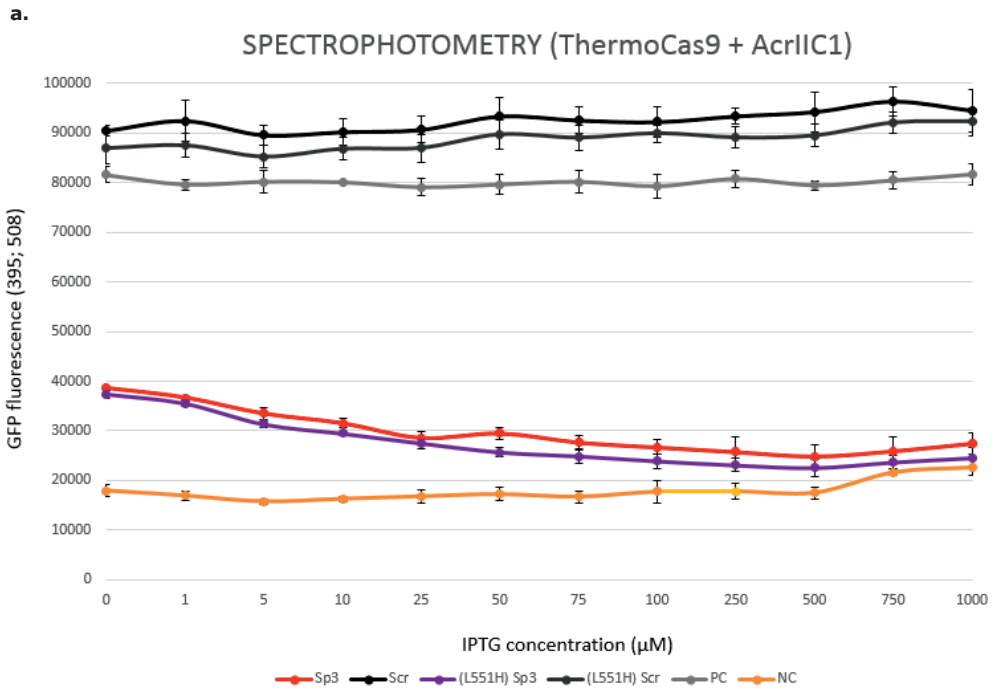
Sup. Fig. S7 | GeoCas9:AcrIIC1$_{Nme}$-based silencing in *E. coli* DH10B_*gfp* (flow cytometry-based fluorescence loss assays). a. Complete data for GeoCas9-m ediated silencing, when the expression of AcrIIC1$_{Nme}$ is not induced (no supplementation of L-rhamnose). **b.** Complete data for GeoCas9-m ediated silencing, when the expression of AcrIIC1$_{Nme}$ is induced (supplementation of 0.2% L-rhamnose).



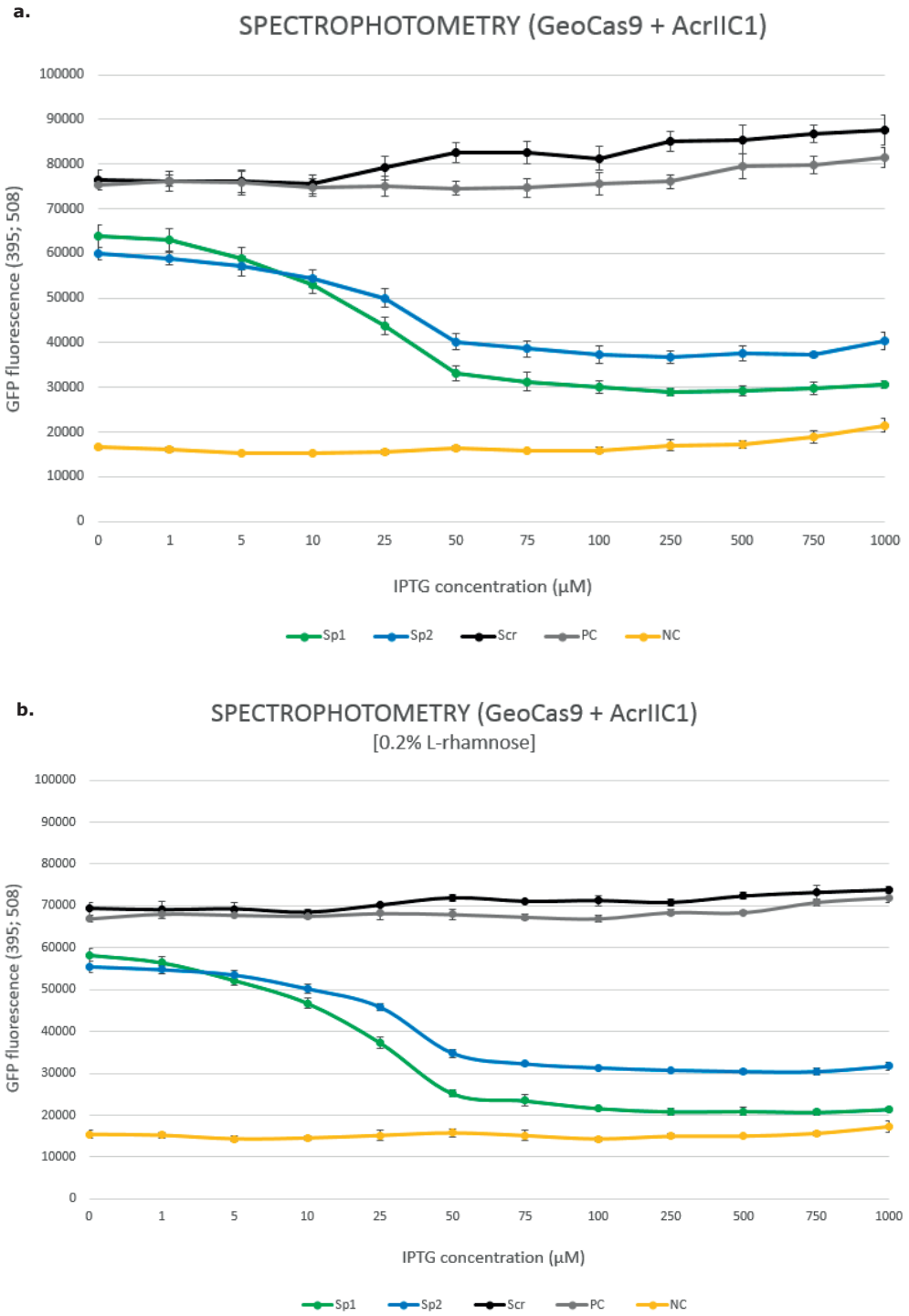
Sup. Fig. S8 | Thermo-dCas9 and Geo-dCas9-based silencing in *E. coli* DH10B_*gfp* (flow cytometry-based fluorescence loss assays). a. Complete data for Thermo-dCas9-mediated silencing in *E. coli* DH10B_*gfp*. **b.** Complete data for Geo-dCas9-mediated silencing in *E. coli* DH10B_*gfp*.



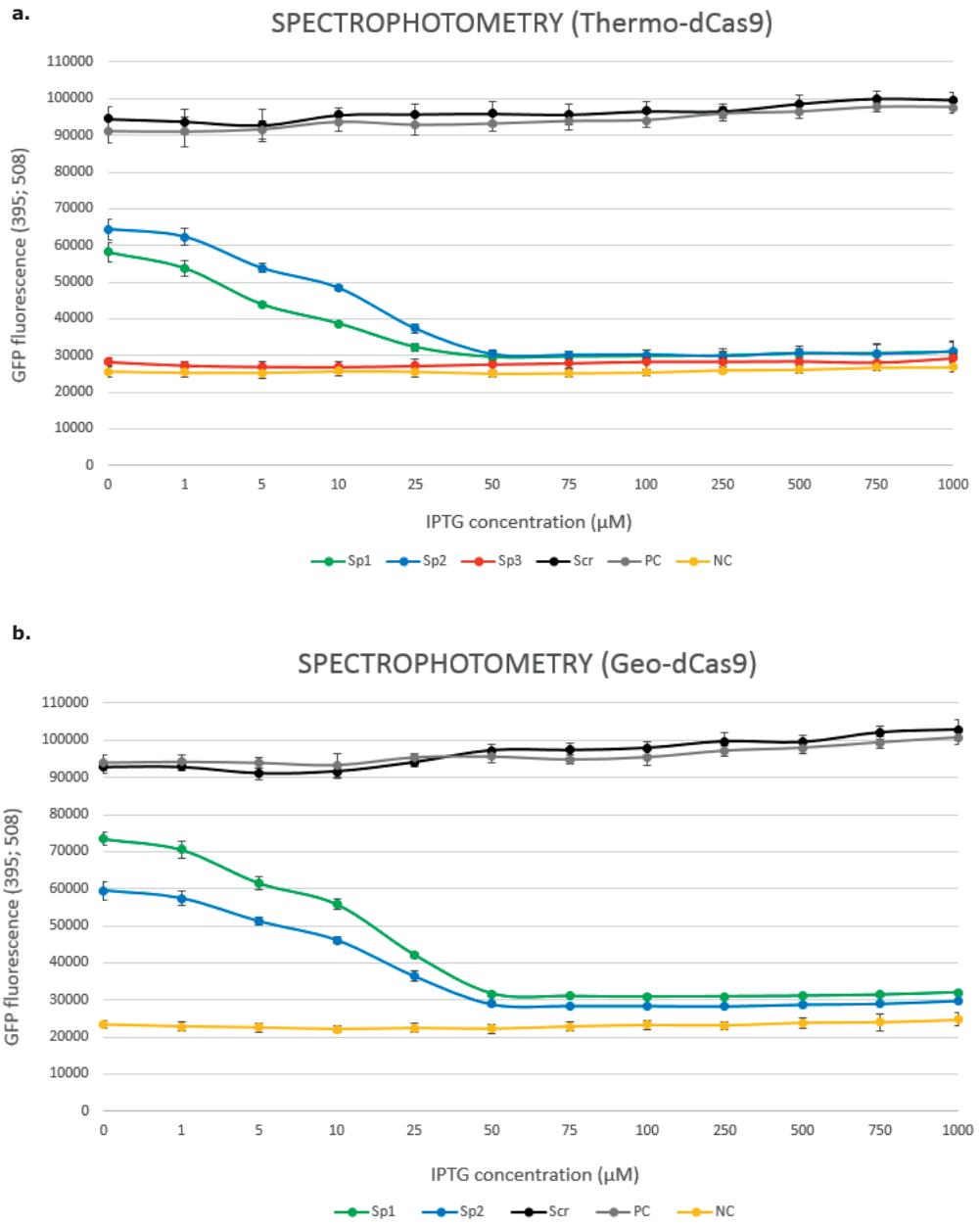
Sup. Fig. S9 | GFP intensity (405-512/25) curves indicate single populations in all tested samples (flow cytometry-based fluorescence loss assays). Different colours of the peaks indicate different IPTG concentrations for the induction of the expression of GeoCas9, ThermoCas9, and ThermoCas9$_{L551H}$. All graphs are representative of biological triplicates.



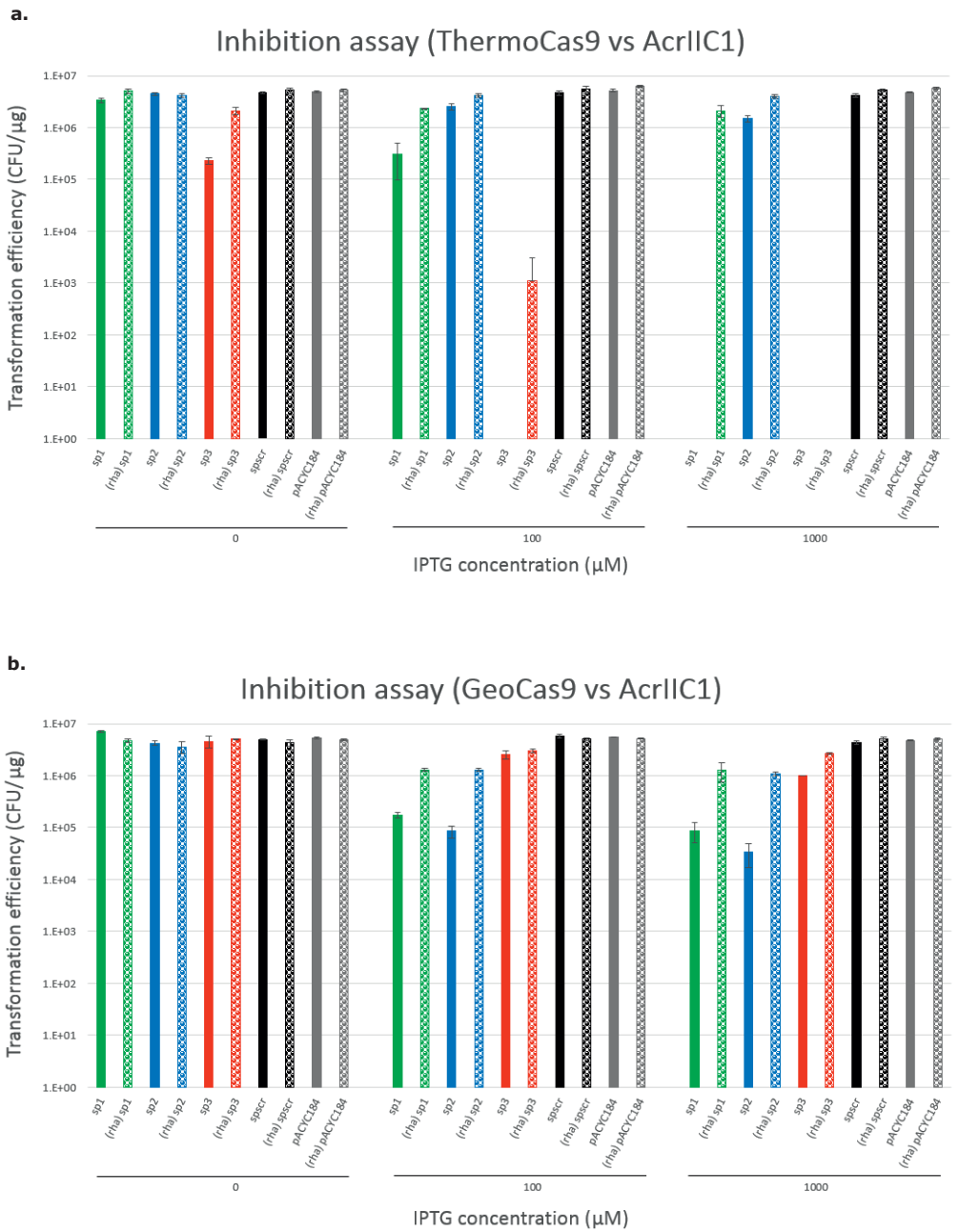
Sup. Fig. S10 | ThermoCas9/ThermoCas9$_{L551H}$:AcrIIC1$_{Nme}$-based silencing in *E. coli* DH 10B_*gfp* (spectrophotometry-based fluorescence loss assays). a. Complete data for ThermoCas9/ThermoCas9$_{L551H}$–mediated silencing, when the expression of AcrIIC1$_{Nme}$ is not induced (no supplementation 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).



Sup. Fig. S11 | 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).



Sup. Fig. S12 | Thermo-dCas9 and G eo-dCas9-based silencing in *E. coli* DH10B_*gfp* (spectrophotometry-based fluorescence loss assays). a. Complete data for Thermo-dCas9-mediated silencing in *E. coli* DH10B_*gfp*. **b.** Complete data for Geo-dCas9-mediated silencing in *E. coli* DH10B_*gfp*.



Sup. Fig. S13 | AcrIIC1$_{Nme}$-mediated inhibition of ThermoCas9, ThermoCas9$_{L551H}$, and GeoCas9 in *E. coli* DH10B_*gfp* (single-vector system). a. Complete data for inhibition of ThermoCas9 and ThermoCas9$_{L551H}$ cleavage activity, when the expression of AcrIIC1$_{Nme}$ is either not induced (no supplementation of L-rhamnose) or induced (supplementation of 0.2% L-rhamnose). **b.** Complete data for inhibition of GeoCas9 cleavage activity, when the expression of AcrIIC1$_{Nme}$ is either not induced (no supplementation of L-rhamnose) or induced (supplementation of 0.2% L-rhamnose).

REFERENCES

1. Haimovich, A. D., Muir, P., & Isaacs, F. J. (2015). Genom es by design.Nature Reviews Genetics, 16(9), 501.
2. Reyrat, J. M., Pelicic, V., Gicquel, B., & Rappuoli, R. (1998). Counterselectable m arkers: untapped tools for bacterial genetics and pathogenesis. Infection and immunity, 66(9), 4011-4017.
3. Schweizer, H. P. (2008). Bacterial genetics: past achievem ents, present state of the field, and future challenges. Biotechniques, 44(5), 633-641.
4. F Bosm a, E., van der Oost, J., M de Vos, W., & van Kranenburg, R. (2013). Sustainable production of bio-based chem icals by extrem ophiles.Current Biotechnology, 2(4), 360-379.
5. Esvelt, K. M., & Wang, H. H. (2013). Genome-scale engineering for system s and synthetic biology. Molecular systems biology, 9(1), 641.
6. Pines, G., Freed, E. F., Winkler, J. D., & Gill, R. T. (2015). Bacterial recom bineering: genom e engineering via phage-based hom ologous recom bination.ACS synthetic biology, 4(11), 1176-1185.
7. Frazier, C. L., San Filippo, J., Lam bowitz, A. M., & Mills, D. A. (2003). Genetic manipulation of Lactococcus lactis by using targeted group II introns: generation of stable insertions without selection. Applied and environmental microbiology, 69(2), 1121-1128.
8. Yao, J., & Lam bowitz, A. M. (2007). Gene targeting in gram-negative bacteria by use of a m obile group II intron ("targetron") expressed from a broad-host-range vector. Applied and environmental microbiology, 73(8), 2735-2743.
9. Wang, Y., Li, X., Milne, C. B., Janssen, H., Lin, W., Phan, G., Hu, H., Jin, Y.S., Price, N. D., & Blaschek, H. P. (2013). Developm ent of a gene knockout system using m obile group II introns (Targetron) and genetic disruption of acid production pathways in *Clostridium beijerinckii*. Applied and environmental microbiology, AEM-00971.
10. Mougiakos, I., Bosm a, E. F., de Vos, W. M., van Kranenburg, R., & van der Oost, J. (2016). Next generation prokaryotic engineering: the CRISPR-Cas toolkit. Trends in biotechnology, 34(7), 575-587.
11. Donohoue, P. D., Barrangou, R., & May, A. P. (2017). Advances in industrial biotechnology using CRISPR-Cas system s.Trends in biotechnology.
12. Mougiakos, I., Bosma, E. F., Ganguly, J., van der Oost, J., & van Kranenburg, R. (2018). Hijacking CRISPR-Cas for high-throughput bacterial m etabolic engineering: Advances and prospects. Current opinion in biotechnology, 50, 146-157.
13. Stout, E., Klaenham m re, T., & Barrangou, R. (2017). CRISPR-Cas technologies and applications in food bacteria. Annual review of food science and technology, 8, 413-437.
14. Barrangou, R., & Horvath, P. (2017). A decade of discovery: CRISPR functions and applications. Nature microbiology, 2(7), 17092.
15. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A program m able dual-RNA–guided DNA endonuclease in adaptive bacterial im m unity. Science, 1225829.
16. Deltcheva, E., Chylinski, K., Sharm a, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., & Charpentier, E. (2011). CRISPR RNA m aturation by tran-sencoded sm all RNA and host factor RNase III. Nature, 471(7340), 602.
17. Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Frem aux, C., Boyaval, P., Rom ero, D. A., Horvath, P., & Moineau, S. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. Journal of bacteriology, 190(4), 1390-1400.
18. Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J., & Alm endros, C. (2009). Short m otif sequences determ ine the targets of the prokaryotic CRISPR defence system Microbiology, 155(3), 733-740.
19. Aravind, L., & Koonin, E. V. (200 1). Prokaryotic hom ologs of the eukaryotic DNA-end-binding protein Ku, novel dom ains in the Ku protein and prediction of a prokaryotic double-strand break repair system .Genome research, 11(8), 1365-1374.
20. Iliakis, G., Wang, H., Perrault, A. R., Boecker, W., Rosidi, B., Windhofer, F., Wu, W., Guan, J., Terzoudi, G., & Pantelias, G. (2004). Mechanism s of DNA double strand break repair and chrom osom e aberration form ation.Cytogenetic and genome research, 104(1-4), 14-20.
21. Bowater, R., & Doherty, A. J. (2006). Making ends m eet: repairing breaks in bacterial DNA by non-hom ologous end-joining. PLoS genetics, 2(2), e8.
22. Cui, L., & Bikard, D. (2016). Consequences of Cas9 cleavage in the chrom osom e of *Escherichia coli*. Nucleic acids research, 44(9), 4243-4251.
23. Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genom es using CRISPR-Cas system s.Nature biotechnology, 31(3), 233.
24. Barrangou, R., & vanPijkeren, J. P. (2016). Exploiting CRISPR–Cas im m une system s for genome editing in bacteria. Current opinion in biotechnology, 37, 61-68.

25. Wang, Y., Zhang, Z. T., Seo, S. O., Choi, K., Lu, T., Jin, Y. S., & Blaschek, H. P. (2015). Markerless chrom osom al gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system . Journal of biotechnology, 200, 1-5.
26. Huang, H., Chai, C., Li, N., Rowe, P., Minton, N. P., Yang, S., Jiang, W., & Gu, Y. (2016). CRISPR/ Cas9-based efficient genom e editing in *Clostridium ljungdahlii*, an autotrophic ga-sferm enting bacterium .ACS synthetic biology, 5(12), 1355-1361.
27. Li, Q., Chen, J., Minton, N. P., Zhang, Y., Wen, Z., Liu, J., Yang, H., Zeng, Z., Ren, X., Yang, J., Gu, Y., Jiang, W., Jiang, Y., & Yang, S (2016). CRISPR-based genom e editing and expression control system s in *Clostridium* acetobutylicum and *Clostridium beijerinckii*. Biotechnology journal, 11(7), 961-972.
28. Li, Y., Lin, Z., Huang, C., Zhang, Y., Wang, Z., Tang, Y. J., Chen, T & Zhao, X. (2015). Metabolic engineering of *Escherichia coli* using CRISPR–Cas9 m editated genom e editing .Metabolic engineering, 31, 13-21.
29. Tong, Y., Charusanti, P., Zhang, L., Weber, T., & Lee, S. Y. (2015). CRISPR-Cas9 based engineering of actinom ycetal genom es.ACS synthetic biology, 4(9), 1020-1029.
30. Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J., & Yang, S. (2015). Multigene editing in the *Escherichia coli* genom e using the CRISPR-Cas9 system .Applied and environmental microbiology, AEM-04023.
31. Wang, Y., Zhang, Z. T., Seo, S. O., Lynn, P., Lu, T., Jin, Y. S., & Blaschek, H. P. (2016). Bacterial genom e editing with CRISPR-Cas9: deletion, integration, single nucleotide m odification, and desirable "clean" m utant selection in*Clostridium beijerinckii* as an exam ple.ACS synthetic biology, 5(7), 721-732.
32. Yan, Q., & Fong, S. S. (2017). Challenges and advances for genetic engineering of non-m odel bacteria and uses in consolidated bioprocessing. Frontiers in microbiology, 8, 2060.
33. Mougiakos, I., Bosma, E. F., Weenink, K., Vossen, E., Goijvaerts, K., van der Oost, J., & van Kranenburg, R. (2017). Efficient genom e editing of a facultative therm ophile using m esophilic spCas 9.ACS synthetic biology, 6(5), 849-861.
34. Richter, F., Fonfara, I., Bouazza, B., Schum acher, C. H., Bratovič, M., Charpentier, E., & Möglich, A. (2013). Engineering of tem perature-and light-switchable Cas9 variants. Nucleic acids research, 44(20), gkw930.
35. Dow, L. E., Fisher, J., O'rourke, K. P., Muley, A., Kastenhuber, E. R., Livshits, G., Tschaharganeh, D. F., & Lowe, S. W. (2015). Inducible *in vivo* genom e editing with CRISPR-Cas9. Nature biotechnology, 33(4), 390.
36. Polstein, L. R., & Gersbach, C. A. (2015). A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. Nature chemical biology, 11(3), 198.
37. Nihongaki, Y., Kawano, F., Nakajim a, T., & Sato, M. (2015). Photoactivatable CRISPR-Cas9 for optogenetic genom e editing.Nature biotechnology, 33(7), 755.
38. Wright, A. V., Sternberg, S. H., Taylor, D. W., Staahl, B. T., Bardales, J. A., Kornfeld, J. E., & Doudna, J. A. (2015). Rational design of a split-Cas9 enzym e com plex. Proceedings of the National Academy of Sciences, 112(10), 2984-2989.
39. Zetsche, B., Volz, S. E., & Zhang, F. (2015). A split- Cas9 architecture for inducible genom e editing and transcription m odulation.Nature biotechnology, 33(2), 139.
40. Nuñez, J. K., Harrington, L. B., & Doudna, J. A. (2016). Chem ical and biophysical m odulation of Cas9 for tunable genom e engineering.ACS chemical biology, 11(3), 681-688.
41. Cao, J., Wu, L., Zhang, S. M., Lu, M., Cheung, W. K., Cai, W., Gale, M., Xu, Q., & Yan, Q. (2016). An easy and efficient inducible CRISPR/Cas9 platform with im proved specificity for m ultiple gene targetin g.Nucleic acids research, 44(19), e149-e149.
42. Liu, K. I., Ram li, M. N.B., Woo, C. W. A., Wang, Y., Zhao, T., Zhang, X., Yim , G. R., Chong, B. Y., Gowher, A., Chua, M. Z., Jung, J., Lee, J. H., & Tan, M. H.(2016). A chem ical-inducible CRISPR– Cas9 system for rapid control of genom e editing.Nature chemical biology, 12(11), 980.
43. Bondy-Denom y, J., Pawluk, A., Maxwell, K. L., & Davidson, A.R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial im m une system.Nature, 493(7432), 429.
44. Pawluk, A., Bondy-Denom y, J., Cheung, V. H., Maxwell, K. L., & Davidson, A. R. (2014). A new group of phage anti-CRISPR genes inhibits the type IE CRISPR-Cas system of Pseudomonas aeruginosa. MBio, 5(2), e00896-14.
45. Pawluk, A., Staals, R. H., Taylor, C., Watson, B. N., Saha, S., Fineran, P. C., Maxwell, K. L., & Davidson, A.R. (2016). Inactivation of CRISPR-Cas system s by anti-CRISPR proteins in diverse bacterial species. Nature microbiology, 1(8), 16085.
46. Pawluk, A., Am rani, N., Zhang, Y., Garcia, B., Hidalgo-Reyes, Y., Lee, J., Edraki, A., Shah, M., Sontheim er, E. J., Maxwell, K. L., & Davidson, A. R. (2016). Naturally occurring off-switches for CRISPR-Cas9. Cell, 167(7), 1829-1838.
47. Hynes, A. P., Rousseau, G. M., Lem ay, M. L., Horvath, P., Rom ero, D. A., Frem aux, C., & Moineau, S.(2017). An anti-CRISPR from a virulent streptococcal phage inhibits *Streptococcus pyogenes* Cas9. Nature microbiology, 2(10), 1374.

48. He, F., Bhoobalan-Chitty, Y., Van, L. B., Kjeldsen, A. L., Dedola, M., Makarova, K. S., Koonin, E. V., Brodersen, D. E., & Peng, X. (2018). Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype ID immu nity. Nature microbiology, 3(4), 461.
49. Marino, N. D., Zhang, J. Y., Borges, A. L., Sousa, A. A., Leon, L. M., Rauch, B. J., Walton, R. T., Berry J. D., Joung, J. K., Kleinsteiver, B. P., & Bondy-Denom y, J. (2018). Discovery of widespread Type I and Type V CRISPR-Cas inhibitors. Science, eaau517.
50. Watters, K. E., Fellm ann, C., Bai, H. B., Ren, S. M., & Doudna, J. A. (2018). System atic discovery of natural CRISPR-Cas12a inhibitors. Science, eaau5138.
51. Bondy-Denom y, J., Garcia, B., Strum , S., Du, M., Rollins, M. F., Hidalg-oReyes, Y., Wiedenheft, B., Maxwell, K. L., & Davidson, A. R. (2015). Multiple m echanism s for CRISP–RCas inhibition by anti-CRISPR proteins. Nature, 526(7571), 136.
52. Maxwell, K. L., Garcia, B., Bondy-Denom y, J., Bona, D., Hidalgo-Reyes, Y., & Davidson, A. R. (2016). The solution structure of an anti-CRISPR protein. Nature communications, 7, 13134.
53. Chowdhury, S., Carter, J., Rollins, M. F., Golden, S. M., Jackson, R. N., Hoffm ann, C., Nosaka, L., Bondy-Denom y, J., Maxwell, K. L., Davidson, A. R., Fischer, E. R. Lander, G. C., & Wiedenheft, B. (2017). Structure reveals m echanism s of viral suppressors that intercept a CRISPR RNA-guided surveillance com plex. Cell, 169(1), 47-57.
54. Peng, R., Xu, Y., Zhu, T., Li, N., Qi, J., Chai, Y., Wu, M., Zhang, X., Shi, Y., Wang, P., Wang, J., Gao, N., & Gao, G. F. (2017). Alternate binding m odes of anti-CRISPR viral suppressors AcrF1/2 to Csy surveillance com plex revealed by cryo-EM structures. Cell research, 27(7), 853.
55. Guo, T. W., Bartesaghi, A., Yang, H., Falconieri, V., Rao, P., Merk, A., Eng, E. T., Raczkowski, A. M., Fox, T., Earl, L. A., Patel, D. J., & Subram aniam , S(.2017). Cryo-EM structures reveal m echanism and inhibition of DNA targeting by a CRISPR-Cas surveillance com plex. Cell, 171(2), 414-426.
56. Dong, D., Guo, M., Wang, S., Zhu, Y., Wang, S., Xiong, Z., Yang, J., Xu, Z., & Huang, Z. (2017). Structural basis of CRISPR–SpyCas9 inhibition by an anti-CRISPR protein. Nature, 546(7658), 436.
57. Shin, J., Jiang, F., Liu, J. J., Bray, N. L., Rauch, B. J., Baik, S. H., Nogales, E., Bondy-Denom y, J., Corn, J. E., & Doudna, J. A. (2017). Disabling Cas9 by an anti-CRISPR DNA m im ic. Science advances, 3(7), e1701620.
58. Yang, H., & Patel, D. J. (2017). Inhibition m echanism of an ant-iCRISPR suppressor AcrIIA4 targeting SpyCas9. Molecular cell, 67(1), 117-127.
59. Kim , I., Jeong, M., Ka, D., Han, M., Kim , N. K., Bae, E., & Suh, J. Y. (2018). Solution structure and dynam ics of anti-CRISPR AcrIIA4, the Cas9 inhibitor. Scientific reports, 8(1), 3883.
60. Harrington, L. B., Doxzen, K. W., Ma, E., Liu, J. J., Knott, G. J., Edraki, A., Garcia, B., Am rani, N., Chen, J. S., Cofsky, J. C., Kranzusch, P. J., Sontheim er, E. J., Davidson, A. R., Maxwell, K. L., & Doudna, J. A. (2017). A broad-spectrum inhibitor of CRISPR-Cas9. Cell, 170(6), 1224-1233.
61. Rauch, B. J., Silvis, M. R., Hultquist, J. F., Waters, C. S., McGregor, M. J., Krogan, N. J., & Bondy-Denom y, J. (2017). Inhibition of CRISPR-Cas9 with bacteriophage proteins. Cell, 168(1-2), 150-158.
62. Basgall, E. M., Goetting, S. C., Goeckel, M. E., Giersch, R. M., Roggenkam p, E., Schrock, M. N., Halloran, M., & Finnigan, G. C. (2018). Gene drive inhibition by the anti-CRISPR proteins AcrIIA2 and AcrIIA4 in *Saccharomyces cerevisiae*. Microbiology, 164(4), 464.
63. Wang, X., Yao, D., Xu, J. G., Li, A. R., Xu, J., Fu, P., Zhou, Y., & Zhu, Y. (2016). Structural basis of Cas3 inhibition by the bacteriophage protein AcrF3. Nature structural & molecular biology, 23(9), 868.
64. Mougiakos, I., Mohanraju, P., Bosm a, E. F., Vrouwe, V., Bou, M. F., Naduthodi, M. I., Gussak, A., Brinkm an, R. B. L., & Oost, J. (2017). Characterizing a therm ostable Cas9 for bacterial genom e editing and silencing. Nature communications, 8(1), 1647.
65. Harrington, L. B., Paez-Espino, D., Staahl, B. T., Chen, J. S., Ma, E., Kyrpides, N. C., & Doudna, J. A. (2017). A therm ostable Cas9 with increased lifetim e in hum an plasm a. Nature Communications, 8(1), 1424.
66. Banno, S., Nishida, K., Arazoe, T., Mitsunobu, H., & Kondo, A. (2018). Deam inase-m ediated m ultiplex genom e editing in *Escherichia coli*. Nature microbiology, 3(4), 423.
67. Zheng, K., Wang, Y., Li, N., Jiang, F. F., Wu, C. X., Liu, F., Chen, H. C., & Liu, Z. F. (2018). Highly efficient base editing in bacteria using a Cas9-cytidine deam inase fusion. Communications Biology, 1(1), 32.
68. Gu, T., Zhao, S., Pi, Y., Chen, W., Chen, C., Liu, Q., Li, M., Han, D., & Ji, Q. (2018). Highly efficient base editing in Staphylococcus aureus using an engineered CRISPR RNA-guided cytidine deam inase. Chemical science, 9(12), 3248-3253.
69. Wang, Y., Liu, Y., Liu, J., Guo, Y., Fan, L., Ni, X., Zheng, X., Wang, M., Zheng, P., Sun, J., & Ma, Y. (2018). MACBETH: Multiplex autom ated *Corynebacterium glutamicum* base editing m ethod. Metabolic engineering, 47, 200-210.
70. Chen, W., Zhang, Y., Zhang, Y., Pi, Y., Gu, T., Song, L., Wang, Y., & Ji, Q. (2018). CRISPR/Cas9-based genom e editing in Pseudomonas aeruginosa and cytidine deam inase-m ediated base editing in Pseudomonas species. iScience.
3729. Green, R., & Rogers, E. J. (2013). Chem ical transform ation of E. coli. Methods in enzymology, 529,

SUPPLEMENTARY TABLES

Supplementary Table 1: Spacers used in this study for ThermoCas9, ThermoCas9L551H , and GeoCas9				
Spacer ID	**Target strand**	**Targeting location in PlacUV5 (5' -> 3')**	**PAM (5' -> 3')**	**Sequence (23 nt) (5' -> 3')**
1	+	16 - 38 (P_{lac}UV5)	NNNNCACA	TCCGGCTCGTATAATGTGTGGGG
2	-	60 - 38 (P_{lac}UV5)	NNNNCACA	TAGAGGGAAACCGTTGTGGTCTC
3	-	82 - 60 (*gfp* gene)	NNNNCCAA	AGAATTTATGCCCATTCACATCC
scr	X	X	X	CTAGATCCGCAGTAACCCCATGG

Supplementary Table 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 pAcrIIC1Nm e	Lab stock
pET_MBP_Geo_st	-	-	PCR template for pGCas9_sp1, pGCas9_sp2, pGCas9_sp3, pGCas9_spscr	[63]
pPtet_Acr_Prha_ TCas9	-	-	PCR template for pGCas9_sp1, pGCas9_sp2, pGCas9_sp3, pGCas9_spscr, pTCas9_sp1, pTCas9_sp2, pTCas9_sp3, pTCas9_spscr	Lab stock
pPtet_Acr_Prha_ TCas9L551H	-	-	PCR template for pTCas9L551H_sp3, pTCas9L551H_spscr	Lab stock
pPrha_TCas9 _sp1	-	-	PCR template for pGCas9_sp1	Lab stock
pPrha_TCas9 _sp2	-	-	PCR template for pGCas9_sp2	Lab stock
pPrha_TCas9 _sp3	-	-	PCR template for pGCas9_sp3	Lab stock
pPrha_TCas9 _spscr	-	-	PCR template for pGCas9_spscr	Lab stock
pPrha_GCas9 _sp1	-	-	PCR template for pAcr	Lab stock
pMK-RQ- _Acr_coK12_coT12	-	-	PCR template for pAcr	Lab stock
pThermoCas9i	-	-	PCR template for pT-dCas9_sp1, pT-dCas9_sp2, pT-dCas9_sp3, and pT-dCas9_spscr	[62]
pGCas9_sp1	Gibson assem bly	Fragm ent 1:*geoCas9* gene from pET_MBP_Geo_st amplified with primers BG13307and BG13310	GeoCas9 and sgRNA (spacer 1) targeting the genome-integrated PlacUV5	This study

		Fragment 2: 5' UTR, lac operator, Ptet, PlacI, and *lacI* gene from pPtet_Acr_Prha_ TCas9 amplified with primers BG13311 and BG13312 Fragment 3: terminator, sgRNA 1, terminator, backbone, and terminator from pPrha_TCas9 _sp1 amplified with primers BG13313 and BG13309	targeting assays, inhibition assays (dual-vector system), spectophotometry- and flow cytometry-based fluorescence loss assays. Moreover, PCR template for pTCas9_sp1, pG-dCas9_sp1, and pAcr_GCas9_sp1.	
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, Ptet, PlacI, and *lacI* gene from pPtet_Acr_Prha_ TCas9 amplified with primers BG13311 and BG13312 Fragment 3: terminator, sgRNA 2, terminator, backbone, and terminator from pPrha_TCas9 _sp2 amplified with primers BG13313 and BG13309	GeoCas9 and sgRNA (spacer 2) targeting the genome-integrated PlacUV5 in targeting assays, inhibition assays (dual-vector system), spectophotometry- and flow cytometry-based fluorescence loss assays. Moreover, PCR template for pTCas9_sp2, pG-dCas9_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, Ptet, PlacI, and *lacI* gene from pPtet_Acr_Prha_ TCas9 amplified with primers BG13311 and BG13312 Fragment 3: terminator, sgRNA 3, terminator, backbone, and terminator from pPrha_TCas9 _sp3 amplified with primers BG13313 and BG13309	GeoCas9 and sgRNA (spacer 3) targeting the genome-integrated PlacUV5 in targeting assays, inhibition assays (dual-vector system), spectophotometry- and flow cytometry-based fluorescence loss assays. Moreover, PCR template for pTCas9_sp3, pTCas9L551H_sp3, pG-dCas9_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, Ptet, PlacI, and *lacI* gene from pPtet_Acr_Prha_ TCas9 amplified with primers BG13311 and BG13312 Fragment 3: terminator, sgRNA scr, terminator, backbone, and terminator from pPrha_TCas9 _spscr amplified with primers BG13313 and BG13309	GeoCas9 and sgRNA (spacer scr) non-targeting the genome-integrated PlacUV5 in targeting assays, inhibition assays (dual-vector system), spectophotometry- and flow cytometry-based fluorescence loss assays. Moreover, PCR template for pTCas9_spscr, pTCas9L551H_spscr, pG-dCas9_spscr, and pAcr_GCas9_spscr.	This study
pTCas9_sp1	Gibson assembly	Fragment 1: 5'UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, sgRNA 1, terminator, backbone, and terminator from pGCas9_sp1 amplified with primers BG13459 and BG13460 Fragment 2: *thermoCas9* gene from pPtet_Acr_Prha_ TCas9 amplified with primers BG13461 and BG13462	ThermoCas9 and sgRNA (spacer 1) targeting the genome-integrated PlacUV5 in targeting assays, inhibition assays (dual-vector system), spectophotometry- and flow cytometry-based fluorescence loss assays. PCR template for pT-dCas9_sp1, and pAcr_TCas9_sp1.	This study
pTCas9_sp2	Gibson assembly	Fragment 1: 5'UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, sgRNA 2, terminator, backbone, and terminator from pGCas9_sp2 amplified with primers BG13459 and BG13460	ThermoCas9 and sgRNA (spacer 2) targeting the genome-integrated PlacUV5 in targeting assays, inhibition assays (dual-vector system), spectophotometry- and flow	This study

		Fragment 2: *thermoCas9* gene from pPtet_Acr_Prha_ TCas9 amplified with primers BG13461 and BG13462	cytometry-based fluorescence loss assays. PCR template for pT-dCas9_sp2, and pAcr_TCas9_sp2.	
pTCas9_sp3	Gibson assembly	Fragment 1: 5'UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, sgRNA 3, terminator, backbone, and terminator from pGCas9_sp3 amplified with primers BG13459 and BG13460 Fragment 2: *thermoCas9* gene from pPtet_Acr_Prha_ TCas9 amplified with primers BG13461 and BG13462	ThermoCas9 and sgRNA (spacer 3) targeting the genome-integrated PlacUV5 in targeting assays, inhibition assays (dual-vector system), spectophotometry- and flow cytometry-based fluorescence loss assays. PCR template for pT-dCas9_sp3, and pAcr_TCas9_sp3.	This study
pTCas9_spscr	Gibson assembly	Fragment 1: 5'UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, sgRNA scr, terminator, backbone, and terminator from pGCas9_spscr amplified with primers BG13459 and BG13460 Fragment 2: *thermoCas9* gene from pPtet_Acr_Prha_ TCas9 amplified with primers BG13461 and BG13462	ThermoCas9 and sgRNA (spacer scr) non-targeting the genome-integrated PlacUV5 in targeting assays, inhibition assays (dual-vector system), spectophotometry- and flow cytometry-based fluorescence loss assays. PCR template for pT-dCas9_spscr, and pAcr_TCas9_spscr.	This study
pAcr	Gibson assembly	Fragment 1: Prha, and 5'UTR from pPrha_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, ThermoCas9L551H 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, Ptet, PlacI, *lacI* gene, terminator, sgRNA 3, terminator, backbone, and terminator from pGCas9_sp3 amplified with primers BG13459 and BG13460 Fragment 2: *thermoCas9L551H* gene from pPtet_Acr_Prha_ TCas9L551H amplified with primers BG13461 and BG13462	ThermoCas9L551H and sgRNA (spacer 3) targeting the genome-integrated PlacUV5 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, Ptet, PlacI, *lacI* gene, terminator, sgRNA scr, terminator, backbone, and terminator from pGCas9_spscr amplified with primers BG13459 and BG13460 Fragment 2: *thermoCas9L551H* gene from pPtet_Acr_Prha_ TCas9L551H amplified with primers BG13461 and BG13462	ThermoCas9L551H and sgRNA (spacer scr) non-targeting the genome-integrated PlacUV5 in inhibition assays (dual-vector system), spectophotometry- and flow cytometry-based fluorescence loss assays.	This study
pG-dCas9_sp1	Gibson assembly	Fragment 1: *Geo-dCas9* gene from pGCas9_sp1 amplified with primers BG13680 and BG13681	Geo-dCas9 and sgRNA (spacer 1) targeting the genome-integrated PlacUV5	This study

		Fragment 2: *Geo-dCas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, sgRNA 1, terminator, backbone, and terminator from pGCas9_sp1 amplified with primers BG13682 and BG13683	in spectophotometry- and flow cytometry-based fluorescence loss assays.	
pG-dCas9_sp2	Gibson assembly	Fragment 1: *Geo-dCas9* gene from pGCas9_sp2 amplified with primers BG13680 and BG13681 Fragment 2: *Geo-dCas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, sgRNA 2, terminator, backbone, and terminator from pGCas9_sp1 amplified with primers BG13682 and BG13683	Geo-dCas9 and sgRNA (spacer 2) targeting the genome-integrated PlacUV5 in spectophotometry- and flow cytometry-based fluorescence loss assays.	This study
pG-dCas9_sp3	Gibson assembly	Fragment 1: *Geo-dCas9* gene from pGCas9_sp3 amplified with primers BG13680 and BG13681 Fragment 2: *Geo-dCas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, sgRNA 3, terminator, backbone, and terminator from pGCas9_sp1 amplified with primers BG13682 and BG13683	Geo-dCas9 and sgRNA (spacer 3) targeting the genome-integrated PlacUV5 in spectophotometry- and flow cytometry-based fluorescence loss assays.	This study
pG-dCas9_spscr	Gibson assembly	Fragment 1: *Geo-dCas9* gene from pGCas9_spscr amplified with primers BG13680 and BG13681 Fragment 2: *Geo-dCas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, sgRNA scr, terminator, backbone, and terminator from pGCas9_sp1 amplified with primers BG13682 and BG13683	Geo-dCas9 and sgRNA (spacer scr) targeting the genome-integrated PlacUV5 in spectophotometry- and flow cytometry-based fluorescence loss assays.	This study
pT-dCas9_sp1	Gibson assembly	Fragment 1: *Thermo-dCas9* gene from pThermoCas9i amplified with primers BG12625 and BG13684 Fragment 2: 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, sgRNA 1, terminator, backbone, and terminator from pTCas9_sp1 amplified with primers BG13311 and BG13685	Thermo-dCas9 and sgRNA (spacer 1) targeting the genome-integrated PlacUV5 in spectophotometry- and flow cytometry-based fluorescence loss assays.	This study
pT-dCas9_sp2	Gibson assembly	Fragment 1: *Thermo-dCas9* gene from pThermoCas9i amplified with primers BG12625 and BG13684 Fragment 2: 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, sgRNA 2, terminator, backbone, and terminator from pTCas9_sp2 amplified with primers BG13311 and BG13685	Thermo-dCas9 and sgRNA (spacer 2) targeting the genome-integrated PlacUV5 in spectophotometry- and flow cytometry-based fluorescence loss assays.	This study
pT-dCas9_sp3	Gibson assembly	Fragment 1: *Thermo-dCas9* gene from pThermoCas9i amplified with primers BG12625 and BG13684 Fragment 2: 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, sgRNA 3, terminator, backbone, and terminator from pTCas9_sp3 amplified with primers BG13311 and BG13685	Thermo-dCas9 and sgRNA (spacer 3) targeting the genome-integrated PlacUV5 in spectophotometry- and flow cytometry-based fluorescence loss assays.	This study
pT-dCas9_spscr	Gibson assembly	Fragment 1: *Thermo-dCas9* gene from pThermoCas9i amplified with primers BG12625 and BG13684	Thermo-dCas9 and sgRNA (spacer scr) targeting the genome-integrated PlacUV5	This study

		Fragment 2: 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, sgRNA scr, terminator, backbone, and terminator from pTCas9_spscr amplified with primers BG13311 and BG13685	in spectophotometry- and flow cytometry-based fluorescence loss assays.	
pAcr_GCas9_sp1	Gibson assembly	Fragment 1: Prha, and AcrIIC1$_{Nme}$ from pAcr amplified with primers BG14000 and BG14001 Fragment 2: Terminator, GeoCas9, 5' UTR, lac operator, Ptet, backbone, PlacI, and partly *lacI* gene from pGCas9_sp1 amplified with primers BG12705 and BG14002 Fragment 3: Partly *lacI* gene, terminator, PJ23119, sgRNA 1, terminator, and backbone from pGCas9_sp1 amplified with primers BG12722 and BG14003	GeoCas9 and sgRNA (spacer 1) targeting the genome-integrated PlacUV5 and AcrIIC1$_{Nme}$ inhibiting GeoCas9, in inhibition assays (single-vector system).	This study
pAcr_GCas9_sp2	Gibson assembly	Fragment 1: Prha, and AcrIIC1$_{Nme}$ from pAcr amplified with primers BG14000 and BG14001 Fragment 2: Terminator, GeoCas9, 5' UTR, lac operator, Ptet, backbone, PlacI, and partly *lacI* gene from pGCas9_sp2 amplified with primers BG12705 and BG14002 Fragment 3: Partly *lacI* gene, terminator, PJ23119, sgRNA 2, terminator, and backbone from pGCas9_sp2 amplified with primers BG12722 and BG14003	GeoCas9 and sgRNA (spacer 2) targeting the genome-integrated PlacUV5 and AcrIIC1$_{Nme}$ inhibiting GeoCas9, in inhibition assays (single-vector system).	This study
pAcr_GCas9_sp3	Gibson assembly	Fragment 1: Prha, and AcrIIC1$_{Nme}$ from pAcr amplified with primers BG14000 and BG14001 Fragment 2: Terminator, GeoCas9, 5' UTR, lac operator, Ptet, backbone, PlacI, and partly *lacI* gene from pGCas9_sp3 amplified with primers BG12705 and BG14002 Fragment 3: Partly *lacI* gene, terminator, PJ23119, sgRNA 3, terminator, and backbone from pGCas9_sp3 amplified with primers BG12722 and BG14003	GeoCas9 and sgRNA (spacer 3) targeting the genome-integrated PlacUV5 and AcrIIC1$_{Nme}$ inhibiting GeoCas9, in inhibition assays (single-vector system).	This study
pAcr_GCas9_spscr	Gibson assembly	Fragment 1: Prha, and AcrIIC1$_{Nme}$ from pAcr amplified with primers BG14000 and BG14001 Fragment 2: Terminator, GeoCas9, 5' UTR, lac operator, Ptet, backbone, PlacI, and partly *lacI* gene from pGCas9_spscr amplified with primers BG12705 and BG14002 Fragment 3: Partly *lacI* gene, terminator, PJ23119, sgRNA scr, terminator, and backbone from pGCas9_spscr amplified with primers BG12722 and BG14003	GeoCas9 and sgRNA (spacer scr) targeting the genome-integrated PlacUV5 and AcrIIC1$_{Nme}$ inhibiting GeoCas9, in inhibition assays (single-vector system).	This study
pAcr_TCas9_sp1	Gibson assembly	Fragment 1: Prha, and AcrIIC1$_{Nme}$ from pAcr amplified with primers BG14000 and BG14001 Fragment 2: Terminator, ThermoCas9, 5' UTR, lac operator, Ptet, backbone, PlacI, and partly *lacI* gene from pTCas9_sp1 amplified	ThermoCas9 and sgRNA (spacer 1) targeting the genome-integrated PlacUV5 and AcrIIC1$_{Nme}$ inhibiting GeoCas9, in inhibition assays (single-vector system).	This study

		with primers BG12705 and BG14002 Fragment 3: Partly *lacI* gene, terminator, PJ23119, sgRNA 1, terminator, and backbone from pTCas9_sp1 amplified with primers BG12722 and BG14003		
pAcr_TCas9_sp2	Gibson assembly	Fragment 1: Prha, and AcrIIC1$_{Nme}$ from pAcr amplified with primers BG14000 and BG14001 Fragment 2: Terminator, ThermoCas9, 5' UTR, lac operator, Ptet, backbone, PlacI, and partly *lacI* gene from pTCas9_sp2 amplified with primers BG12705 and BG14002 Fragment 3: Partly *lacI* gene, terminator, PJ23119, sgRNA 2, terminator, and backbone from pTCas9_sp2 amplified with primers BG12722 and BG14003	ThermoCas9 and sgRNA (spacer 2) targeting the genome-integrated PlacUV5 and AcrIIC1$_{Nme}$ inhibiting GeoCas9, in inhibition assays (single-vector system).	This study
pAcr_TCas9_sp3	Gibson assembly	Fragment 1: Prha, and AcrIIC1$_{Nme}$ from pAcr amplified with primers BG14000 and BG14001 Fragment 2: Terminator, ThermoCas9, 5' UTR, lac operator, Ptet, backbone, PlacI, and partly *lacI* gene from pTCas9_sp3 amplified with primers BG12705 and BG14002 Fragment 3: Partly *lacI* gene, terminator, PJ23119, sgRNA 3, terminator, and backbone from pTCas9_sp3 amplified with primers BG12722 and BG14003	ThermoCas9 and sgRNA (spacer 3) targeting the genome-integrated PlacUV5 and AcrIIC1$_{Nme}$ inhibiting GeoCas9, in inhibition assays (single-vector system).	This study
pAcr_TCas9_spscr	Gibson assembly	Fragment 1: Prha, and AcrIIC1$_{Nme}$ from pAcr amplified with primers BG14000 and BG14001 Fragment 2: Terminator, ThermoCas9, 5' UTR, lac operator, Ptet, backbone, PlacI, and partly *lacI* gene from pTCas9_spscr amplified with primers BG12705 and BG14002 Fragment 3: Partly *lacI* gene, terminator, PJ23119, sgRNA scr, terminator, and backbone from pTCas9_spscr amplified with primers BG12722 and BG14003	ThermoCas9 and sgRNA (spacer scr) targeting the genome-integrated PlacUV5 and AcrIIC1$_{Nme}$ inhibiting GeoCas9, in inhibition assays (single-vector system).	This study

Supplementary Table 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	pAcrIIC1Nme	Ampicillin	This study

Supplementary Table 4: Primers/Oligonucleotides used in this study			
	Primer/ Oligo ID	**Sequence (5' -> 3')**	**Function**
Targeting, inhibition	BG13307	tcataagcaggccattttgtctgatTTAGTCACGAGTAGATTGCAGTG	Construction of pGCas9_sp1, pGCas9_sp2, pGCas9_sp3, pGCas9_spscr
	BG13310	ttgctaacgcagtcaggcaccgtgtATGCGTTATAAGATTGGCCTGG	
	BG13311	ACACGGTGCCTGACTG	
	BG13312	tttttggtccatcagtcgattctgaTCACTGCCCGCTTTCC	
	BG13313	TCAGAATCGACTGATGGACC	
	BG13309	ATCAGACAAAATGGCCTGC	
	BG13459	tatcaagaccgattttatacttcatACACGGTGCCTGACTG	Construction of pTCas9_sp1, pTCas9_sp2, pTCas9_sp3, pTCas9_spscr
	BG13460	cggggaaactatccgtccgttataaATCAGACAAAATGGCCTGCTTATG	
	BG13461	TTATAACGGACGGATAGTTTCCC	
	BG13462	ATGAAGTATAAAATCGGTCTTGATATCGG	
Inhibition (dual-vector approach),	BG13427	aagcgggcagtgagcgcaacgcaatCACCACAATTCAGCAAATTGTG	Construction of pAcr
	BG13428	ATGTATATCTCCTTCTTATAGTTAAACAAAATTATTTCTAG	
	BG13429	ttaactataagaaggagatatacatATGAATAAAACATATAAAATCGGAAAAAATG CGG	
	BG13430	gctcagcggtggcggccgctctaggTCATAGTTCAACAAACTCCCAACATG	
	BG13431	CCTAGAGCGGCCGCCACCGCTGAGCAATAACTAGCATAACCCCTTG GGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG	
	BG13432	CAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCT AGTTATTGCTCAGCGGTGGCGGCCGCTCTAGG	
	BG13433	taaacgggtcttgaggggttttttgTTAAGCCAGCCCCGAC	
	BG13434	ATTGCGTTGCGCTCAC	
	BG13459	tatcaagaccgattttatacttcatACACGGTGCCTGACTG	Construction of pTCas9L551H_sp3 pTCas9L551H_spscr
	BG13460	cggggaaactatccgtccgttataaATCAGACAAAATGGCCTGCTTATG	
	BG13461	TTATAACGGACGGATAGTTTCCC	
	BG13462	ATGAAGTATAAAATCGGTCTTGATATCGG	
Spectrophotometry	BG13680	gatcacaGCatctacttcaacgtaac	Construction of pG-dCas9_sp1, pG-dCas9_sp2, pG-dCas9_sp3, pG-dCas9_spscr
	BG13681	gcctggccatcggtattac	
	BG13682	gtaataccgatggccaggc	
	BG13683	gttacgttgaagtagatGCtgtgatc	
	BG12625	atcttattaatcagataaaatatttGTCGACAGGATTAAACAAAAATGGC	Construction of pT-dCas9_sp1, pT-dCas9_sp2, pT-dCas9_sp3, pT-dCas9_spscr
	BG13684	ttgctaacgcagtcaggcaccgtgtATGAAGTATAAAATCGGTCTTGCTATCG	
	BG13311	ACACGGTGCCTGACTG	
	BG13685	aaatattttatctgattaataagatgatcttcttgagatcg	
Inhibition assays	BG14000	atcttattaatcagataaaatatttCACCACAATTCAGCAAATTGTG	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
	BG14001	gccatttttgtttaatcctgtcgacTCATAGTTCAACAAACTCCCAACATG	
	BG12705	GTCGACAGGATTAAACAAAAATGGCC	
	BG14002	CAATCAGCAACGACTGTTTGC	
	BG12722	GCAAACAGTCGTTGCTGATTG	
	BG14003	AAATATTTTATCTGATTAATAAGATGATCTTCTTGAGATCG	
Sequencing	BG12706	GGTAGCTCAGAGAACCTTCG	Sequencing of pGCas9_sp1, pGCas9_sp2, pGCas9_sp3, pGCas9_spscr, pG-dCas9_sp1, pG-dCas9_sp2, pG-dCas9_sp3, pG-dCas9_spscr, *pAcr_GCas9_sp1, *pAcr_GCas9_sp2, *pAcr_GCas9_sp3, *pAcr_GCas9_spscr
	BG12712	ACGCGATGGATATGTTCTG	
	BG13157	GCGAATTCGGCAACATGTC	
	BG13158	CAGGGTGAAGTCTACTCCTC	
	BG13159	GCTACATCAGCCGCTTCTTC	
	BG13160	GTCTGCTGGAGCATAACAAC	
	BG13161	GCACCCTGAAACGTTTCG	
	BG13163	CGAAACGTGTAATCTTCGGTC	
	BG13164	GGCCAGTTCCTTGTTCTG	
	BG13165	CGGTTTTCATCCTGTTCC	
	BG13166	CTGGAATGCTTGTTCATACAGC	
	BG13167	CGCTCCAGACGGTGTTTAC	
	BG12721	CTGCGTTAGCAATTTAACTGTG	
	BG12722	GCAAACAGTCGTTGCTGATTG	
	BG12723	GTTAACCACCATCAAACAGG	
	BG12718	CAGTCGCGTACCGTCTTC	

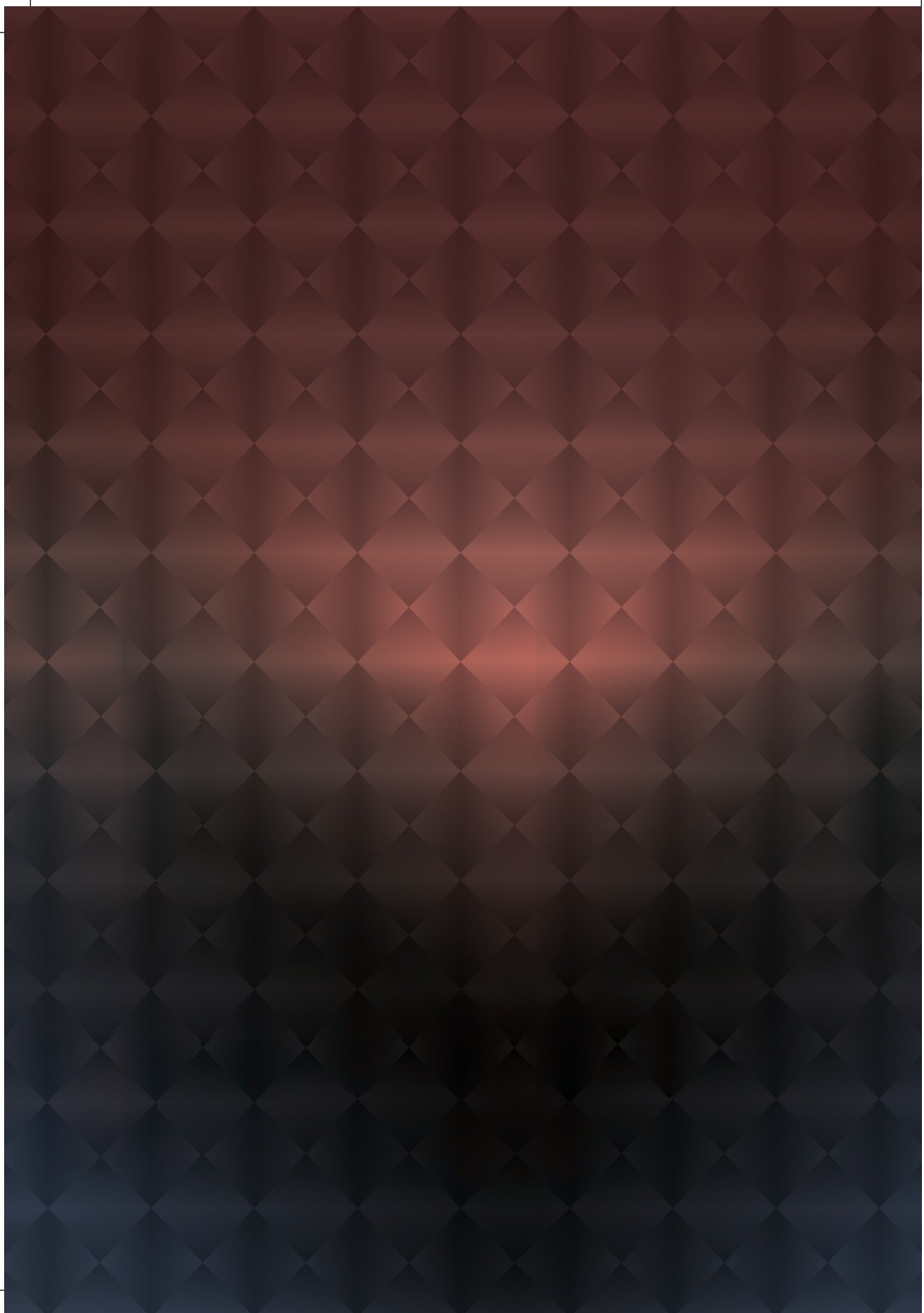
	BG12719	GTGCATGCAAGGAGATGG	
	*BG12705	GTCGACAGGATTAAACAAAAATGGCC	
	BG12706	GGTAGCTCAGAGAACCTTCG	Sequencing of pTCas9_sp1, pTCas9_sp2, pTCas9_sp3, pTCas9_spscr, pTCas9L551H_sp3, pTCas9L551H_spscr, pT-dCas9_sp1, pT-dCas9_sp2, and pT-dCas9_sp3, pT-dCas9_spscr, *pAcr_TCas9_sp1, *pAcr_TCas9_sp2, *pAcr_TCas9_sp3, and *pAcr_TCas9_spscr
	BG12707	GTCGTATCGATGATTTTGATTGTTCG	
	BG12708	GCGTGAGATATAACGGGTATC	
	BG12709	GTAGACTTCGCCTTGTTCC	
	BG12710	GTTCCCATATTCGCGCTG	
	BG12712	ACGCGATGGATATGTTCTG	
	BG12713	CGAAGATTTAGGTGTCCGC	
	BG12714	CATACACATTCCAGTCCTTCACC	
	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	

Supplementary Table 5: DNA sequences of Cas9 and anti-CRISPR genes used in this study

Gene	DNA sequence (5'->3')
thermocas9	ATGAAGTATAAAATCGGTCTTGATATCGGCATTACGTCTATCGGTTGGGCTGTCATTAATTTGG ACATTCCTCGCATCGAAGATTTAGGTGTCCGCATTTTTGACAGAGCGGAAAACCCGAAAACCG GGGAGTCACTAGCTCTTCCACGTCGCCTCGCCCGCTCCGCCCGACGTCGTCTGCGGCGTCGCAA ACATCGACTGGAGCGCATTCGCCGCCTGTTCGTCCGCGAAGGAATTTTAACGAAGGAAGAGCT GAACAAGCTGTTTGAAAAAAAGCACGAAATCGACGTCTGGCAGCTTCGTGTTGAAGCACTGGA TCGAAAACTAAATAACGATGAATTAGCCCGCATCCTTCTTCATCTGGCTAAACGGCGTGGATTT AGATCCAACCGCAAGAGTGAGCGCACCAACAAAGAAAACAGTACGATGCTCAAACATATTGA AGAAAACCAATCCATTCTTTCAAGTTACCGAACGGTTGCAGAAATGGTTGTCAAGGATCCGAA ATTTTCCCTGCACAAGCGTAATAAAGAGGATAATTACACCAACACTGTTGCCCGCGACGATCTT GAACGGGAAATCAAACTGATTTTCGCCAAACAGCGCGAATATGGGAACATCGTTTGCACAGAA GCATTTGAACACGAGTATATTTCCATTTGGGCATCGCAACGCCCTTTTGCTTCTAAGGATGATA TCGAGAAAAAAGTCGGTTTCTGTACGTTTGAGCCTAAAGAAAAACGCGCGCCAAAAGCAACAT ACACATTCCAGTCCTTCACCGTCTGGGAACATATTAACAAACTTCGTCTTGTCTCCCCGGGAGG CATCCGGGCACTAACCGATGATGAACGTCGTCTTATATACAAGCAAGCATTTCATAAAAATAA AATCACCTTCCATGATGTTCGAACATTGCTTAACTTGCCTGACGACACCCGTTTTAAAGGTCTT TTATATGACCGAAACACCACGCTGAAGGAAATGAGAAAGTTCGCTTCCTTGAACTCGGCGCC TATCATAAAATACGGAAAGCGATCGACAGCGTCTATGGCAAAGGAGCAGCAAAATCATTTCGT CCGATTGATTTTGATACATTTGGCTACGCATTAACGATGTTTAAAGACGACACCGACATTCGCA GTTACTTGCGAAACGAATACGAACAAAATGGAAAACGAATGGAAAATCTAGCGGATAAAGTC TATGATGAAGAATTGATTGAAGAACTTTTAAACTTATCGTTTTCTAAGTTTGGTCATCTATCCCT TAAAGCGCTTCGCAACATCCTTCCATATATGGAACAAGGCGAAGTCTACTCAACCGCTTGTGA ACGAGCAGGATATACATTTACAGGGCCAAAGAAAAAACAGAAAACGGTATTGCTGCCGAACA TTCCGCCGATCGCCAATCCGGTCGTCATGCGCGCACTGACACAGGCACGCAAAGTGGTCAATG CCATTATCAAAAAGTACGGCTCACCGGTCTCCATCCATATCGAACTGGCCCGGGAACTATCAC AATCCTTTGATGAACGACGTAAAATGCAGAAAGAACAGGAAGGAAACCGAAAGAAAAACGAA ACTGCCATTCGCCAACTTGTTGAATATGGGCTGACGCTCAATCCAACTGGG**CTT**GACATTGTGA AATTCAAACTATGGAGCGAACAAAACGGAAAATGTGCCTATTCACTCCAACCGATCGAAATCG AGCGGTTGCTCGAACCAGGCTATACAGAAGTCGACCATGTGATTCCATACAGCCGAAGCTTGG ACGATAGCTATACCAATAAAGTTCTTGTGTTGACAAAGGAGAACCGTGAAAAAGGAAACCGC ACCCCAGCTGAATATTTAGGATTAGGCTCAGAACGTTGGCAACAGTTCGAGACGTTTGTCTTGA CAAATAAGCAGTTTTCGAAAAAGAAGCGGGATCGACTCCTTCGGCTTCATTACGATGAAAACG AAGAAAATGAGTTTAAAAATCGTAATCTAAATGATACCCGTTATATCTCACGCTTCTTGGCTAA CTTTATTCGCGAACATCTCAAATTCGCCGACAGCGATGACAAACAAAAAGTATACACGGTCAA CGGCCGTATTACCGCCCATTTACGCAGCCGTTGGAATTTTAACAAAAACCGGGAAGAATCGAA TTTGCATCATGCCGTCGATGCTGCCATCGTCGCCTGCACAACGCCGAGCGATATCGCCCGAGTC ACCGCCTTCTATCAACGGCGCGAACAAAACAAAGAACTGTCCAAAAAGACGGATCCGCAGTTT CCGCAGCCTTGGCCGCACTTTGCTGATGAACTGCAGGCGCGTTTATCAAAAAATCCAAAGGAG AGTATAAAAGCTCTCAATCTTGGAAATTATGATAACGAGAAACTCGAATCGTTGCAGCCGGTT

	TTTGTCTCCCGAATGCCGAAGCGGAGCATAACAGGAGCGGCTCATCAAGAAACATTGCGGCGT TATATCGGCATCGACGAACGGAGCGGAAAAATACAGACGGTCGTCAAAAAGAAACTATCCGA GATCCAACTGGATAAAACAGGTCATTTCCCAATGTACGGGAAAGAAAGCGATCCAAGGACAT ATGAAGCCATTCGCCAACGGTTGCTTGAACATAACAATGACCCAAAAAAGGCGTTTCAAGAGC CTCTGTATAAACCGAAGAAGAACGGAGAACTAGGTCCTATCATCCGAACAATCAAAATCATCG ATACGACAAATCAAGTTATTCCGCTCAACGATGGCAAAACAGTCGCCTACAACAGCAACATCG TGCGGGTCGACGTCTTTGAGAAAGATGGCAAATATTATTGTGTCCCTATCTATACAATAGATAT GATGAAAGGGATCTTGCCAAACAAGGCGATCGAGCCGAACAAACCGTACTCTGAGTGGAAGG AAATGACGGAGGACTATACATTCCGATTCAGTCTATACCCAAATGATCTTATCCGTATCGAATT TCCCCGAGAAAAAACAATAAAGACTGCTGTGGGGGAAGAAATCAAAATTAAGGATCTGTTCG CCTATTATCAAACCATCGACTCCTCCAATGGAGGGTTAAGTTTGGTTAGCCATGATAACAACTT TTCGCTCCGCAGCATCGGTTCAAGAACCCTCAAACGATTCGAGAAATACCAAGTAGATGTGCT AGGCAACATCTACAAAGTGAGAGGGGAAAAGAGAGTTGGGGTGGCGTCATCTTCTCATTCGAA AGCCGGGGAAACTATCCGTCCGTTATAA
geocas9	ATGCGTTATAAGATTGGCCTGGACATCGGTATTACCTCTGTTGGTTGGGCAGTCATGAACCTGG ATATCCCTCGTATCGAAGATCTGGGCGTGCGCATTTTTGACCGTGCGGAGAACCCGCAGACCG GTGAATCTCTGGCTCTGCCGCGTCGTCTGGCACGTAGCGCACGCCGCCGCCTGCGTCGTCGTAA ACACCGTCTGGAGCGTATTCGTCGCCTGGTTATTCGTGAAGGCATCCTGACGAAAGAAGAACT GGATAAACTGTTCGAAGAAAAACACGAGATCGACGTATGGCAGCTGCGTGTAGAAGCCCTGG ACCGTAAGCTGAACAACGACGAACTGGCGCGTGTCCTGCTGCATCTGGCAAAGCGTCGTGGCT TCAAATCTAACCGTAAATCTGAACGCTCCAATAAAGAGAACTCCACTATGCTGAAACATATTG AGGAGAACCGTGCAATTCTGTCTAGCTACCGTACCGTGGGCGAAATGATTGTTAAAGACCCGA AATTCGCACTGCATAAGCGTAACAAAGGCGAAAACTACACCAACACCATTGCACGCGATGACC TGGAACGTGAAATCCGTCTGATTTTCTCCAAACAGCGCGAATTCGGCAACATGTCTTGCACCGA AGAATTCGAAAACGAATATATTACCATTTGGGCATCTCAGCGTCCGGTGGCGTCTAAAGATGA TATCGAAAAAAAAAGTAGGCTTTTGTACTTTCGAACCGAAGGAAAAACGTGCGCCGAAAGCCAC CTATACCTTCCAGTCTTTTATCGCGTGGGAACATATCAACAAACTGCGTCTGATTTCTCCGTCTG GCGCCCGCGGCCTGACCGACGAAGAACGTCGTCTGCTGTATGAACAAGCATTCCAGAAAAACA AAATTACCTACCACGATATTCGTACCCTGCTGCATCTGCCGGACGACACCTACTTCAAGGGCAT CGTTTACGATCGCGGTGAATCTCGTAAGCAGAACGAAAACATTCGTTTCCTGGAACTGGATGC ATACCACCAGATCCGTAAAGCTGTAGATAAAGTTTACGGCAAGGGTAAATCCAGCAGCTTCCT GCCGATCGACTTTGATACCTTCGGTTACGCGCTGACCCTGTTTAAAGACGATGCGGATATCCAC TCTTACCTGCGCAACGAGTACGAACAGAACGGCAAACGTATGCCTAACCTGGCTAACAAAGTT TACGATAACGAGCTGATTGAAGAACTGCTGAACCTGTCCTTCACTAAATTCGGTCACCTGTCTC TGAAAGCTCTGCGTTCCATCCTGCCGTATATGGAACAGGGTGAAGTCTACTCCTCCGCTTGTGA ACGTGCAGGCTACACCTTCACCGGTCCGAAAAAGAAGCAAAAAACTATGCTGCTGCCGAACAT CCCGCCGATTGCGAACCCTGTAGTAATGCGTGCACTGACCCAGGCGCGCAAAGTAGTCAACGC GATCATCAAAAAGTACGGCAGCCCGGTTTCCATCCATATCGAACTGGCGCGCGACCTGAGCCA GACTTTTGACGAGCGTCGTAAAACTAAAAAGGAACAGGATGAAAACCGTAAAAAAAACGAAA CCGCGATCCGCCAGCTGATGGAATACGGTCTGACTCTGAACCCTACTGGTCACGATATTGTGA AGTTCAAGCTGTGGTCTGAACAGAACGGTCGCTGTGCTTACTCTCTGCAGCCGATCGAGATCG AACGTCTGCTGGAGCCAGGTTACGTTGAAGTAGATCATGTGATCCCGTACTCCCGCTCTCTGGA TGATTCTTATACCAACAAAGTTCTGGTTCTGACTCGCGAAAACCGTGAGAAAGGCAACCGCAT CCCAGCTGAATATCTGGGTGTTGGCACTGAGCGTTGGCAACAGTTCGAAACCTTCGTCCTGACC AATAAACAGTTCTCTAAAAAGAAACGTGACCGTCTGCTGCGTCTGCACTACGATGAAAACGAA GAGACTGAATTCAAAAACCGTAACCTGAACGATACTCGCTACATCAGCCGCTTCTTCGCAAAC TTCATTCGTGAACACCTGAAATTTGCGGAATCCGACGATAAACAGAAAGTTTATACCGTAAAC GGCCGTGTTACCGCCCACCTGCGTTCTCGCTGGGAGTTCAACAAGAACCGTGAGGAAAGCGAT CTGCACCACGCTGTTGACGCCGTTATTGTGGCGTGCACCACCCCAAGCGATATCGCTAAGGTG ACCGCATTCTACCAGCGTCGTGAGCAGAACAAGGAACTGGCCAAAAAAACCGAACCGCATTTT CCGCAGCCGTGGCCGCACTTCGCGGACGAACTGCGTGCTCGTCTGTCCAAACATCCTAAAGAA AGCATCAAAGCTCTGAACCTGGGTAACTACGATGACCAAAAACTGGAATCTCTGCAGCCGGTG TTTGTCAGCCGTATGCCGAAACGTTCTGTTACTGGCGCTGCGCACCAGGAAACGCTGCGCCGTT ACGTGGGCATCGACGAACGCTCCGGTAAAATCCAGACCGTAGTAAAAACCAAACTGTCCGAG ATTAAACTGGATGCATCCGGCCACTTCCCAATGTACGGTAAAGAATCCGATCCACGCACTTAT GAAGCCATCCGCCAGCGTCTGCTGGAGCATAACAACGACCCGAAAAAGGCATTCCAGGAGCC TCTGTACAAACCGAAAAAAAACGGCGAACCGGGCCCGGTAATCCGTACTGTAAAAATTATCGA CACGAAAAACCAGGTGATCCCTCTGAACGATGGTAAAACCGTGGCCTACAATTCCAACATCGT TCGCGTGGACGTGTTCGAAAAAGATGGTAAATACTACTGTGTACCGGTGTATACCATGGACAT CATGAAAGGCATTCTGCCGAACAAAGCGATTGAACCGAACAAGCCGTACTCTGAATGGAAAG AAATGACCGAAGATTACACGTTTCGTTTCAGCCTGTATCCGAACGACCTGATCCGCATCGAACT GCCGCGTGAAAAAACCGTTAAAACCGCTGCAGGCGAAGAAATTAACGTGAAAGACGTGTTCG TTTACTATAAAACGATCGACTCCGCAAACGGCGGCCTGGAACTGATTTCTCACGACCACCGTTT CTCTCTGCGTGGCGTTGGCTCTCGCACCCTGAAACGTTTCGAGAAATATCAAGTTGATGTTCTG GGTAACATCTATAAAGTGCGTGGCGAGAAACGTGTCGGTCTGGCGTCCTCCGCACACAGCAAA CCTGGCAAAACCATTCGTCCACTGCAATCTACTCGTGACTAA

acrIIC1N$_{me}$	ATGAATAAAACATATAAAATCGGAAAAAATGCGGGATATGACGGCTGCGGGTTATGTTTAGCC GCGATCTCAGAAAATGAAGCTATTAAAGTTAAGTATCTGCGCGATATCTGTCCAGACTATGAC GGCGACGACAAAGCTGAGGACTGGCTGAGATGGGGAACGGATAGCCGCGTGAAAGCTGCGGC TTTAGAAATGGAGCAATATGCTTATACGTCGGTTGGGATGGCGTCATGTTGGGAGTTTGTTGAA CTATGA



Chapter 7

HIJACKING CRISPR-CAS FOR HIGH-THROUGHPUT BACTERIAL METABOLIC ENGINEERING: ADVANCES AND PROSPECTS

Ioannis Mougiakos[1#], Elleke Bosma[2#], Joyshree Ganguly[3#], John van der Oost[1], Richard van Kranenburg[1,3]*

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

[2] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet B220, 2800 Kgs. Lyngby, Denmark

[3] Corbion, Arkelsedijk 46, 4206 AC Gorinchem, The Netherlands.

#Contributed equally

*Corresponding author

Chapter adapted from publication:

Curr Opin Biotechnol. 2018 Apr;50:146-157. doi: 10.1016/j.copbio.2018.01.002

ABSTRACT

High engineering efficiencies are required for industrial strain development. Due to its user-friendliness and its stringency, CRISPR-Cas-based technologies have strongly increased genome engineering efficiencies in bacteria. This has enabled more rapid metabolic engineering of both the model host *Escherichia coli* and non-model organisms like *Clostridia*, *Bacilli*, *Streptomycetes* and cyanobacteria, opening new possibilities to use these organisms as improved cell factories. The discovery of novel Cas9-like systems from diverse microbial environments will extend the repertoire of applications and broaden the range of organisms in which it can be used to create novel production hosts. This review analyzes the current status of prokaryotic metabolic engineering towards the production of biotechnologically relevant products, based on the exploitation of different CRISPR-related DNA/RNA endonuclease variants.

KEYWORDS

CRISPR/Cas, bacteria, metabolic engineering, genome editing, genome silencing

INTRO DUCTIO N

The transition towards a bio-based economy demands the development of fermentation-based processes economically competitive with the currently employed unsustainable production processes[1]. Unfortunately, only very few natural organisms are suitable for their direct application in an industrial process. Therefore, efficient metabolic engineering via targeted genome engineering is required and the development and use of simple and high-throughput genome engineering tools generally applicable to many model and non-model organisms is of great importance[2]. Most prokaryotes possess homology directed repair (HDR) systems, which have since long been exploited in a great variety of microorganisms for targeted chromosomal integrations of desired modifications[3-5]. In the HDR-based systems, plasmid-borne homologous recombination templates, which often harbor selection markers for screening purposes, are introduced into the genome through double or sequential single crossover events. The HDR-based approach is based Cre-lox or FLP-FRT systems, for excision of the markers from the genomes for recycling purposes. However, these systems leave genomic scars that could be the cause of unwanted chromosomal rearrangements[6,7]. Alternatively, for a small number of bacteria, markerless genomic modifications are possible via recombineering systems[8]. These systems are based on bacteriophage recombinases and ssDNA, dsDNA or plasmid-borne DNA fragments with sequence homology to the genomic target. However, due to the absence of marker-based selection, these systems mostly result in low mutation efficiencies[8]. The construction and screening of mutants in all these HDR-based approaches is time consuming, rendering these tools suboptimal for extensive metabolic engineering, particularly in non-model organisms with low transformability and recombination rates.

A breakthrough moment in the molecular microbiology field was the discovery of bacterial adaptive immune systems that are based on genomic Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) - the memory of the systems - and CRISPR–associated (Cas) proteins[9-11].The repurposing of the RNA-guided DNA endonuclease from the type IIa CRISPR-Cas system of *Streptococcus pyogenes* (spCas9) and of other Cas9 orthologues as genome editing tools brought an unprecedented revolution to the life sciences field[2,12,13]. The basis of he Cas9 engineering tools is the simple way in which Cas9 nucleases can be guided to the desired DNA target, denoted as protospacer, by a CRISPR–RNA:trans-activating CRISPR RNA (crRNA:tracrRNA) hybrid complex. For this purpose, the 5'-end of the crRNA module, denoted as spacer, has to be complementary to the

selected protospacer[14] and a specific short DNA motif, denoted as protospacer adjacent motif (PAM), has to be present at the 3'-end of the denoted as spacer, has to be complementary to the selected protospacer[14] and a specific short DNA motif, denoted as protospacer adjacent motif (PAM), has to be present at the 3'-end of the selected protospacer[15,16]. For further simplification of the engineering processes, the crRNA:tracrRNA complex can be combined into a chimeric single guide RNA (sgRNA)[17].

CRISPR-Cas-based genome editing is now broadly used in a variety of organisms,including human cells, zebrafish, plants, yeast and bacteria[18,19]. The successful application of Cas9-based genome editing in eukaryotic cells is based on the error-prone correction of Cas9-induced double-strand DNA breaks (DSBs) by the efficient eukaryotic Non-Homologous End Joining (NHEJ) mechanism. Contrary to eukaryotes, most prokaryotes do not have an active NHEJ system3 and Cas9-induced DSBs cannot be repaired, resulting in cell death. Recent studies have shown that engineering efficiencies in prokaryotes were strongly increased (often up to 100%) upon combining the existing homologous recombination and recombineering systems with Cas9-targeting; the Cas9-induced DNA breaks served, simultaneously, as recombination inducers and counter-selection tools[12,20,21]. Moreover, catalytically inactive variants of Cas9 orthologues and variants fused with transcriptional activating factors have been developed and used for transcriptional regulation. Altogether, CRISPR-Cas9 orthologues and catalytically inactive mutants have accelerated the construction and screening of *in silico* designed strains, facilitating metabolic engineering of a wide range of bacterial species for industrial cell factory development (Figure 1).

In this review, we summarize recent bacterial metabolic engineering studies that focused on the construction and improvement of microbial cell factories, making use of CRISPR-Cas-based technologies. Additionally, we explore newly developed CRISPR-Cas tools and argue on how their application could improve the currently available technologies. Finally, we discuss how screening diverse environments can lead to discovery of new Cas-related variants to extend the repertoire of applications and create novel production hosts.

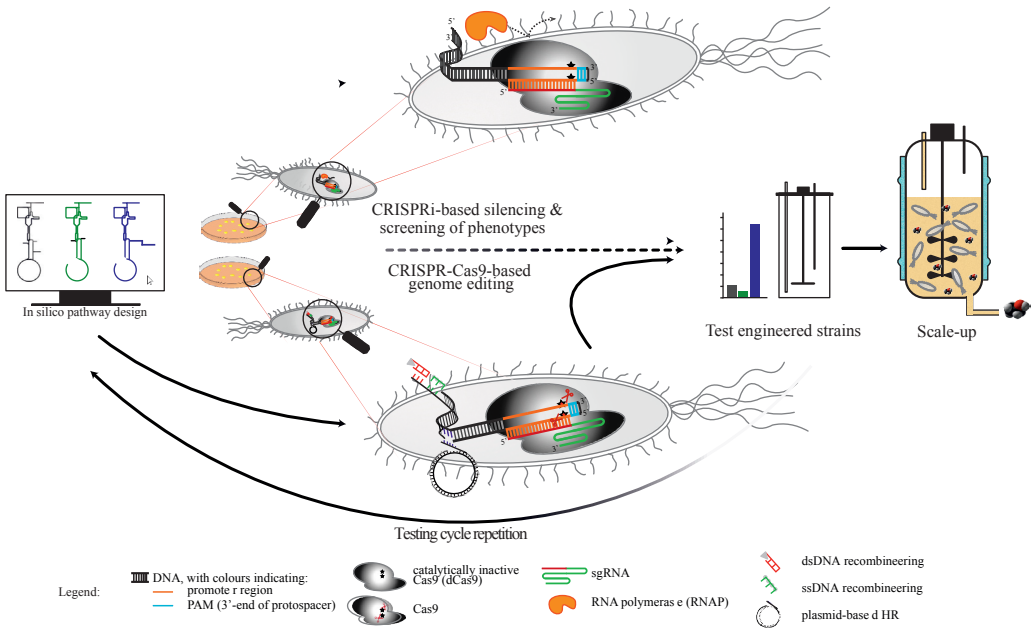


Figure 1. CRISPR-Cas9-based metabolic engineering of bacterial cell factories. Using Cas9 as counter-selection tool for traditional genome editing processes such as recombineering or plasmid-based homologous recombination has led to efficient metabolic engineering towards a wide range of products. Additionally, dCas9 can be used for CRISPRi to rapidly screen *in silico* predicted phenotypes prior to Cas9-based engineering.

BACTERIAL METABOLIC ENGINEERING AND CRISPR-CAS TECHNOLOGIES

Many chemicals such as terpenoids, alcohols, amino acids, organic acids and antibiotics have high commercial value in the pharmaceutical and nutritional industry, and as fuels and building block chemicals. Most of these compounds result from multi-step metabolic pathways and are often tightly regulated in their natural genomic context[22,23]. Model organisms, like *E. coli*, have well-studied metabolisms, extensive and high-throughput molecular toolboxes and detailed *in silico* metabolic models. Hence, the use of such organisms minimizes the number of required engineering steps and maximizes engineering efficiencies. Nevertheless, the demands for engineering work remain high before efficient production strains are constructed due to the complexity of the metabolic pathways for many commercially interesting products[2]. Moreover, model organisms are often suboptimal as production hosts and the use of alternative organisms could benefit the production of many valuable chemicals. Cas-based genome engineering and silencing tools have enabled and accelerated complex metabolic engineering and systems-level understanding of metabolic pathways in a wide range of organisms[2,12,13,18] (Figure 1).

CRISPR-CAS EDITING

Metabolic engineering strategies include plasmid-based expression or, preferably, chromosomal integration of heterologous metabolic pathways, and/or targeted genome editing and adaptive evolution for flux redistribution through native metabolic pathways (Table 1). In their pioneering work, Li *et al.*[24] combined Cas9-induced targeted DSBs with ss or ds λ-RED recombineering for the introduction of a heterologous β-carotene biosynthetic pathway into the *E. coli* genome. They further substituted the promoters and ribosome binding sites (RBSs) from the native MEP pathway genes to achieve different levels of overexpression of the corresponding enzymes. Further engineering steps, including numerous deletions and promoter/RBS substitutions of central carbon metabolism genes, improved pyruvate and glyceraldehyde-3P supply and lead to the construction of a highly improved β-carotene producing strain[24]. This extensive study was possible only due to the development of the Cas9-based tools, revealing their great potential for efficient and diverse manipulation of genomic DNA.

The Cas9-recombineering method was further exploited with the development of the CRISPR-enabled trackable genome engineering (CREATE) tool[25]. Application of this tool in *E. coli* cells allowed their simultaneous transformation with multiple libraries of plasmid-borne recombination templates, each designed to introduce easily trackable mutations at different genomic loci[25]. The CREATE tool was employed to introduce multiple RBS variations for each of the genes in a genomically integrated isopropanol production pathway in *E. coli*, leading to the time-efficient construction and testing of ~1000 strains. The isopropanol titer of the best strain was 1.5-fold higher compared to the initial integration strain, but still lower compared to the plasmid-based overexpression approach[26]. Cas9-based downregulation or deletion of competing pathways in strains already overexpressing heterologous pathways towards the desired product could further improve titers. This has been proven successful for many compounds, including n-butanol in *E. coli*[27] and *Clostridium saccharoperbutylacetonicum* [28], isopropanol-butanol-ethanol in *Clostridium acetobutylicum*[29], succinic acid in *Synechococcus elongatus*[30], γ-amino-butyric acid (GABA) in C*orynebacterium glutamicum*[31], and 5-aminolevulinic acid in *E. coli*[32] (Table 1). However, the use of natural producers can substantially reduce the complexity of engineering steps towards production and tolerance build-up. The tolerance of *Streptomyces* species to antibiotics has been exploited for the production of antibiotic and antitumor compounds simply by Cas9-facilitated genomic integration of multiple biosynthetic gene cluster (BGC)

copies[33,34]. In *C. glutamicum*, the natural proline production was enhanced 6.5-fold through aa codon saturation mutagenesis approach to relieve product inhibition[35]. Noteworthy, this work was performed using the Cas12a (formerly Cpf1) RNA-guided endonuclease, making it the first application of a non-Cas9-based CRISPR-Cas/recombineering genome editing tool in bacteria for metabolic engineering purposes (Table 1).

Product		Species	Strategy				Chromosomal modifications made using CRISPR-Cas-editing	Ref.
			Product pathway overexpression	Chromosomal deletions	Chromosomal insertions	Chromosomal substitutions		
Terpenoids	β-carotene	*Escherichia coli*	Heterologous, chromosomal	Competing pathways	Product pathway	Promoters, RBSs	Knock-in *crtE-crtB-crtI-crtY* + knock-out *ldhA*, knock-in *gps*, combinatorial promoter/RBS replacement of 9 MEP pathway genes, combinatorial overexpressions and deletions of 8 central carbon metabolism genes, knock-in 2nd copies of selected MEP and β-carotene pathway genes	[24]
			Native, plasmid; Heterologous, chromosomal		Product pathway		Knock-in *almgs* under control of various regulatory parts (in a β-carotene production strain)	[36]
Alcohols	Isopropanol	*Escherichia coli*	Heterologous, chromosomal		Product pathway	RBSs	Knock-in and RBS replacement of *thl, atoDA, adc, adh*	[26]
	n- Butanol		Heterologous, plasmid			5'-UTR of competing pathway gene	Modification of *gltA* 5'-UTR for expression reduction	[27]
		Clostridium saccharoperbutylacetonium		Competing pathways			Knock-out *pta, buk*	[28]
	Isopropanol-butanol-ethanol	*Clostridium acetobutylicum*	Heterologous and native, chromosomal		Product pathway		Knock-in *ctfAB, adc, adh*	[29]
Amino acids	L-proline	*Corynebacterium glutamicum*	-			Codons (to relieve product inhibition)	Codon saturation mutagenesis *γ-glutamyl kinase* by CRISPR-Cpf1	[35]
	γ-amino-butyric acid (GABA)		Heterologous, plasmid	Transporters, degradation pathway			Knock-out *Ncgl1221, gabT, gabP* and various combinations thereof	[31]
	5-Amino-levulinic acid	*Escherichia coli*	Heterologous, plasmid	Competing pathways		Promoter, codons (to relieve product inhibition)	*coaA* point mutation (R106A), *serA* promoter replacement and C-terminal residues deletion, knockout *sucCD*, *hemB* translational downregulation by start codon substitution	[32]

Table 1. CRISPR-Cas-mediated metabolic engineering of bacteria for industrial products.

Product		Species	Strategy				Chromosomal modifications made using CRISPR-Cas-editing	Ref.
			Product pathway overexpression	Chromosomal deletions	Chromosomal insertions	Chromosomal substitutions		
Org. acids	Succinic acid	*Synechococcus elongatus*	Native, chromosomal	Competing pathways	Product pathway		Knock-out *glgc* and knock-in *gltA-ppc* (under nitrogen starvation conditions)	[30]
Antibiotics/anti-tumor	Pristinamycin I (PI)	*Streptomyces pristinaespiralis*	Native, chromosomal	Repressor, competing pathway	Product pathway		Knock-out *snaEI, snaE2* and *papR3*, knock-in PI biosynthetic gene cluster	[33]
	Pristinamycin II (PII)		Native, chromosomal		Product pathway		Knock-in of artificial bacteriophage attachment/integration (*attB*) sites in which the biosynthetic pathway is subsequently inserted	[34]
	Chloramphenicol	*Streptomyces coelicolor*	Heterologous, chromosomal		Product pathway			
	Anti-tumor compound YM-216391		Heterologous, chromosomal		Product pathway			
Lipid	Fatty acids	*Escherichia coli*	Heterologous, chromosomal	Competing pathways	Product pathway		Knock-in *fadR, delta9* and *acc* (deletions made previously)	[37]

Table 1(cont). CRISPR-Cas-mediated metabolic engineering of bacteria for industrial products.

Finally, Cas9-based editing tools were successfully employed for membrane engineering purposes in *E. coli*. The β-carotene storage capacity of *E. coli* cell membranes was increased by chromosomal integration of heterologous membrane-bending protein genes using plasmid-borne homologous recombination and Cas9-targeting[36]. Furthermore, the Cas9-recombineering tool proved efficient for the enhancement of the *E. coli* lipid content by simultaneous chromosomal integration of a heterologous fatty acid regulatory transcription factor gene together with a delta9 desaturase and an acetyl-CoA carboxylase gene[37] (Table 1).

CRISPRi

Next to the integration or deletion of genes and pathways, an important metabolic engineering strategy is the fine-tuning of gene expression. Whereas in eukaryotic systems siRNA-techniques have since long enabled transcriptional control, for prokaryotes such silencing tools have only recently become available with the CRISPR interference (CRISPRi) tool, which is based on dCas9: the catalytically inactive variant of the Cas9 endonuclease[38]. This tool can be used for complete or partial expression; repression strength can be tuned by altering the position of the selected protospacer within the targeted gene (Figure 2), or the inducer concentration when an inducible promoter is employed for the expression of the Cas9 or the sgRNA module. This is crucial when targeting essential genes, competing pathways (which can also be biomass-producing) or regulators for which a basal expression level is re-

quired (Figure 2a,b,c,d). It can also be used as a quick alternative to the often laborious RBS/promoter-engineering to tune production pathway activity to either modulate amounts and properties of the target product (Figure 2e)[39,40] or prevent accumulation of toxic intermediates[41] (Figure 2f). Although CRISPRi does not lead to the construction of stably genetically modified strains, it is a powerful method for quick evaluation of the possible effects of genetic modifications to the metabolism of a microorganism, allowing to design genome editing approaches and to gain insights into microbial metabolisms (Figure 1, Table 2).

Similar to CRISPR-Cas-based editing, model organisms such as *E. coli* and *C. glutamicum* were the first organisms for which CRISPRi-based metabolic engineering was applied (Table 2). A heterologous polyhydroxyalkanoate (PHA) biosynthesis pathway was introduced into *E. coli* for the production of poly(3-hydroxybutyrate-co-4-hydroxybutyrate) [P(3HB-co-4HB)][42]. The 4HB fraction of the polymer was enhanced via CRISPRi-based downregulation of multiple TCA cycle genes aiming to increase the supply of the 4HB-precursor succinate semialdehyde[42] (Figure 2e). Subsequent works in *E. coli* and in the natural PHA-producer *Halomonas* focused on silencing cell morphology genes to increase the storage capacity of the cells for PHAs[40,43] (Figure 2d), as well as biosynthetic pathway genes to control PHA content and chemical properties such as molecular weight and polydispersity[39,40] (Figure 2d). Cell morphology engineering, through CRISPRi-based repression of *E. coli* cell division and shape genes, combined with expression of a heterologous production pathway, was also used for the production of the phytochemical zeaxanthin[44] (Table 2). Several other studies used CRISPRi-based repression of competing pathway or repressor genes for the enrichment of precursor pools (Figure 2a, c), aiming at the enhancement of the natural amino acid production by *C. glutamicum*[45] and the *E. coli*-based heterologous production of various phytochemicals including naringenin[46,47], resveratrol[48], pinosylvin[49], anthocyanin[50], as well as medium chain fatty acids (MCFAs)[51] and n-butanol[52] (Table 2) Additionally, CRISPRi-based repression of essential genes was used to minimize carbon loss towards biomass formation. Notably, many of these studies employed multiplex silencing using an sgRNA approach, whereas Cress *et al.* used a dual RNA (crRNA/tracrRNA) approach, developing a rapid CRISPR-array assembly method denoted as CRISPathBrick[47]. This tool could facilitate multiplex CRISPRi-based silencing in non-model organisms with limited toolboxes.

CRISPRi-based repression has already been used for metabolic engineering purposes in non-model organisms such as in *Clostridium cellulovorans* and *Klebsiella pneu-*

Product		Species	Product pathway overexpression method	CRISPRi-based repression of:[1]	Genes targeted by CRISPRi	M[2]	Ref.
Polyhydroxyalkanoates (PHA)	P(3HB-co-4HB) with enhanced 4HB content	*Escherichia coli*	Heterologous, plasmid	Competing pathways	*sad/gadB, sucCD, sdhAB*	+	[42]
	Polyhydroxy-butyrate (PHB)		Heterologous, plasmid	Cell morphology	*ftsZ, mreB*	-	[43]
	PHB		Heterologous, plasmid	Product pathway	*phaC*	-	[39]
	3-hydroxybutyrate/ 3-hydroxyvalerate (PHBV) and PHB	*Halomonas* sp.	-	Cell morphology Product pathway Competing pathways	*ftsZ, prpC, gltA*	-	[40]
Phytochemicals	Zeaxanthin	*Escherichia coli*	Heterologous, plasmid (+sensor for dynamic control) and chromosomal	Cell morphology	*ftsZ, mreB, pbp, rodZ*	-	[44]
	Naringenin		Heterologous, plasmid	Competing pathways	*eno, adhE, mdh, fabB, fabF, sucC, fumC*	+	[46]
	Naringenin		Heterologous, plasmid		*fadR, fumC, sucABCD, scpC*		
	Resveratrol		Heterologous, plasmid	Competing pathways	*fabD, fabH, fabB, fabF, fabI*	-	[48]
	Pinosylvin		Heterologous, plasmid	Competing pathways	*eno, adhE, fabB, sucC, fumC, fabF*	+	[49]
	Anthocyanin		Heterologous, plasmid	Repressor of product pathway	*metJ*	-	[50]
Fatty acids	Medium chain fatty acids (MCFAs)	*Escherichia coli*	Heterologous and native, plasmid	Competing pathways	*rdA, adhE, ldhA, poxB, pta*	-	[51]
Terpenoids	(-)-α-bisabolol, isoprene, lycopene	*Escherichia coli*	Heterologous, plasmid	Product pathway Competing pathway: biomass	*mvaK1, mvaE, ispA*	-	[41]
	Mevalonate		Heterologous, plasmid	Competing pathway: biomass	*pyrF, oriC, dnaA*	-	[53]
Amino acids	L-lysine, L-glutamate	*Corynebacterium glutamicum*	-	Competing pathways	*pgi, pck, pyk*	-	[45]
Alcohols	Acetone-butanol-ethanol	*Clostridium cellulovorans* & *Clostridium beijerinckii*	Native, plasmid	Competing pathways (as well as knock-out)	putative hydrogenase in *C. cellulovorans*	-	[54]
	n-Butanol	*Klebsiella pneumoniae*	Heterologous, plasmid	Competing pathways	*ilvB, ilvI, metA, alaA*	-	[55]
		Escherichia coli	Heterologous, plasmid	Competing pathways	*pta, frdA, ldhA, adhE*	+	[52]
	Fatty alcohols	*Synechocystis* sp.	Heterologous, plasmid	Competing pathways	*plsX, aar, ado, plsC, lplat*	+	[56]
Organic acids	Succinate	*Synechococcus elongatus*	-	Competing pathways (under nitrogen starvation conditions)	*glgc, sdhA, sdhB*	-	[57]
	Lactate	*Synechococcus* sp.	-	Competing pathway to accumulate activator metabolite	*glnA*	-	[58]

Table 2. CRISPRi-based metabolic engineering of bacteria for industrial products. [1] See Figure 2 for a visualization of these strategies. Modifications other than CRISPRi are shown in brackets. [2] M: multiplexing.

moniae for alcohol production[54,55], *Synechococcus elongatus* for succinate production[57] (Table 2), as well as in *Clostridium acetobutylicum* to relieve carbon catabolite repression for sugar co-utilization[59]. The CRISPRi tool is particularly useful in cyanobacteria, in which genome editing is complicated and time consuming due to slow growth and multiple chromosome copies[57,60]. Furthermore, the ability to fine-tune expression levels using CRISPRi was exploited in *Synechococcus* sp., where repression of nitrogen assimilation gene glnA was shown to increase the pool of α-ketoglutarate[58]. A moderate increase of this metabolite enhanced glycolytic flux and lactate production, whereas a too large increase resulted in a decrease in protein production[58]. CRISPRi-based multiplex gene repression was established in *Synechocystis* sp.[60] and subsequently used to determine optimal gene repression combinations for fatty alcohol production[56] (Table 2). As in all studies using CRISPRi in cyanobacteria, dCas9 and sgRNAs were chromosomally integrated into the genome. This resulted in stable repression strains in the absence of selective pressure for single sgRNAs, but the use of repetitive promoter elements resulted in undesired recombination events when multiplexing was attempted[56], highlighting the potential advantage of using a dual RNA approach. Finally, it was observed that a targeted gene with a very distant transcription start site (TSS) from the start codon could be efficiently repressed by CRISPRi only when employing multiple sgRNAs targeting within the gene or the preceding operon genes[60]. Hence, this study revealed that the efficient application of the CRISPRi tool is strongly connected with the precise identification of TSS for the targeted genes.

FUTURE PERSPECTIVES

In prokaryotes, CRISPR-Cas based genome editing has strongly increased engineering efficiencies by adding a powerful counter-selection method to existing engineering systems or by enhancing recombination efficiencies through induction of cellular DNA repair mechanisms[12,20]. In model organism *E. coli* high-throughput tools such as crMAGE[7] and CREATE[25] made multiplex engineering possible upon combining the Cas9 and recombineering tools. It is expected that these high-throughput tools will be Further developed into automated pipelines for rapid industrial strain development, while the thorough exploitation of their potential requires the additional development of rapid and easy screening and read-out systems.CRISPR-Cas-based counter-selection tools have increased editing efficiencies in many non-model organisms. However, further improvement of these tools still strongly depends on the development of basics, such as well-characterized inducible promoters[61]. The tight control of the Cas9-expression would allow for efficient integration of

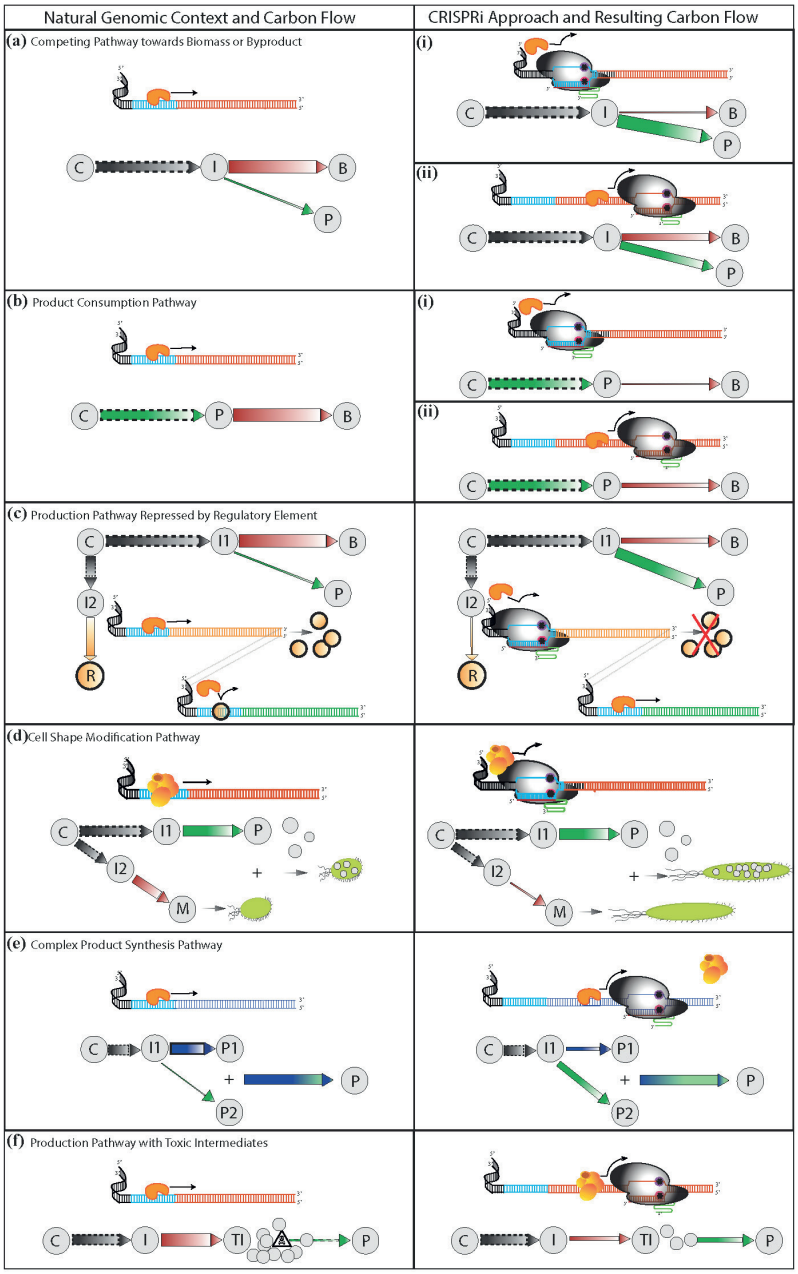


Figure 2. Overview of CRISPRi-based metabolic engineering strategies to increase production of the target product (P). Abbreviations: C: carbon source; I: intermediate metabolite; B: byproduct or biomass; P: target product; TI: toxic intermediate; R: repressor; M: morphology. See legend of Figure 1 for graphic legend. Arrows represent intracellular carbon flows and their thickness corresponds to the flow rate. Arrows with dashed outlines represent merged pathways. **a)** Repression of competing pathway that leads to byproduct or biomass formation with i) indicating dCas9 targeting the promoter region, resulting in stronger repression than in ii), where the coding region is targeted. **b)** Repression of competing pathway that leads to product consumption with i) indicating dCas9 targeting the promoter region, resulting in stronger repression than in ii), where the coding region is targeted. **c)** Repression of the product pathway to change product com position or properties. **d)** Repression of the product pathway to prevent accumulation of toxic intermediates. **e)** Repression of repressor of the target product pathway. **f)** Repression of cell shape/morphology genes to increase cell size and storage capacity for the target product.

an employed homologous recombination template prior to the counter-selection step. The use of intrinsic Cas9-properties, such as temperature-sensitivity, can substitute the requirement for inducible promoters in organisms that can grow under conditions outside the Cas9 activity range[62].

The use of alternative Class 2 CRISPR systems will further extend the Cas-based engineering toolbox. Recently, the Cpf1 (Cas12) RNA-guided DNA endonucleases from the type V CRISPR-Cas systems of *Francisella novicida*[35] and *Acidaminococcus* sp.[63] have been repurposed for bacterial genome editing and silencing. Cpf1 does not require a tracrRNA and can process its own precursor crRNA via its intrinsic RNAse activity. Hence, the use of Cpf1 for the development of a multiplex engineering tool can prevent the issues encountered when using Cas9 and multiple sgRNAs for multiplex engineering[63]. The use of a Cpf1-recombineering tool in *C. glutamicum*, for which only very low levels of Cas9 expression can be tolerated[31,35,64], resulted in screenable editing efficiencies[35], while a DNase-dead Cpf1 (ddCpf1) variant was recently employed for multiplex silencing in *E. coli*[63]. The newly discovered Class 2 Type VI system Cas13 (C2c2) RNA-guided RNA-nuclease can be used for silencing via the degradation of transcripts, or for tracking of transcripts using fluorescent-coupled catalytically inactive variants[65,66]. The RNA-guided RNA endonuclease from the type VI CRISPR-Cas systems of *Leptotrichia shahii* and *Leptotrichia wadei* have already been successfully repurposed for RNA interference in *E. coli*[66,67]. Furthermore, the repurposing of native CRISPR-Cas systems for genome editing[68], has been proved efficient and holds promise for organisms with low transformation efficiencies[69]. It is anticipated that the development of easy and rapid characterization techniques[70-73] will accelerate the exploitation of novel CRISPR-Cas systems for the development of prokaryotic engineering tools. These tools will further expand the number of target sites, the range of easy-to-engineer organisms and they will increase the engineering speed by simultaneous usage of different Cas systems for genome editing and plasmid curing[65,74] similar to the recently developed EXIT-circuit approach that combines Cas9-base d editing and I-SceI-based plasmid curing[75].

Screening natural resources for novel CRISPR-Cas systems will further expand the applications and range of organisms and environments in which CRISPR-Cas-based editing can be applied[76-78]. A recent example of this approach was the identification of a thermostable Cas9-orthologue in the genome of a *Geobacillus thermodenitrificans* strain isolated from a compost sample, which was further characterized and employed to establish

the first Cas9-based engineering tool for thermophilic bacteria[79]. The robustness of thermostable Cas9-based tools can be further exploited for applications in extreme environments, as was recently shown for another thermostable Cas9-orthologue with prolonged life time in blood plasma[80]. The discovery of novel Cas nucleases with different properties, such as tolerance to alkaline or acidic pH and high saline concentrations, would be possible by screening selected environmental samples and metagenomic libraries. The characterization of these nucleases could lead to the development of engineering tools with wide applicability to biotechnologically relevant but currently unexploited extremophilic organisms.

CONCLUSIONS

Altogether, the developments in CRISPR-Cas-based bacterial genome engineering and increase insight into metabolism on a systems level and enable more rapid strain engineering, which is crucial for the development of a bio-based economy using microbial cell factories. Rapid current developments and future applications, which will further expand the range of organisms and applications of CRISPR-Cas-based editing for metabolic engineering, consist of fine-tuning of the tools, their adaptation to different hosts, their extension into combinations with other active components such as proteases, markers or activators, as well as the discovery and development of novel Cas-like systems.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

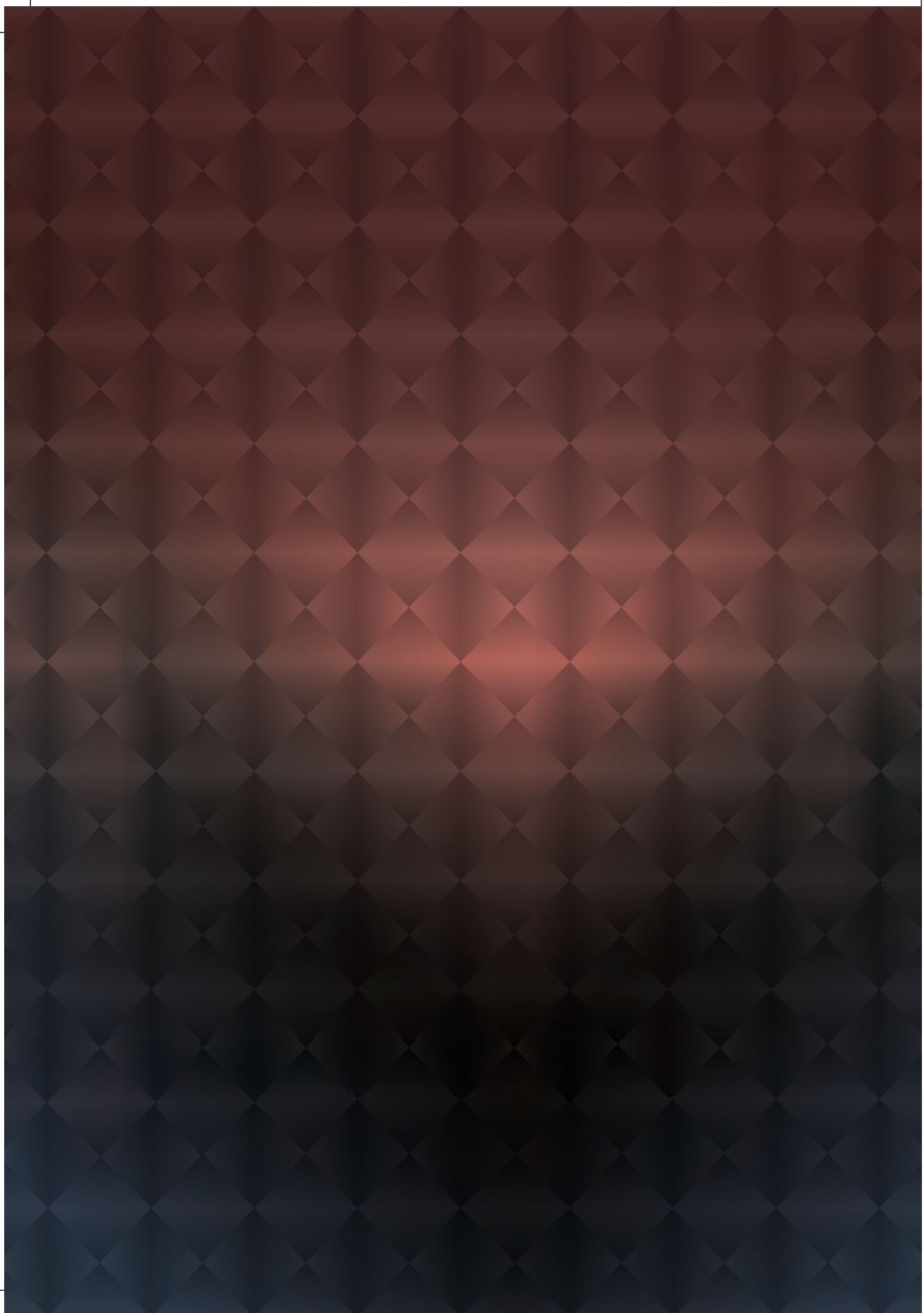
Funding: This work was supported by Corbion BV, The Netherlands; The European Union Marie Skłodowska-Curie Innovative Training Networks (ITN) [contract number 642068]; Dutch Technology Foundation STW, which is part of The Netherlands Organization for Scientific Research (NWO) and which is partly funded by the Ministry of Economic Affairs (grant number 13938); and Marine Biotechnology ERA-NET “ThermoFactories” (grant number 5178-00003B) and the Novo Nordisk Foundation, Denmark.

REFERENCES

1. Hillmyer, M.A., The promise of plastics from plants. Science, 2017. **358**(6365): p. 868-870.
2. Donohoue, P.D., R. Barrangou, and A.P. May, Advances in industrial biotechnology using CRISPR-Cas systems. Trends in biotechnology, 2017.
3. Bowater, R. and A.J. Doherty, Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. PLoS genetics, 2006. **2**(2): p. e8.
4. Cromie, G.A., J.C. Connelly, and D.R. Leach, Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. Molecular cell, 2001. **8**(6): p. 1163-1174.
5. Court, D.L., J.A. Sawitzke, and L.C. Thomason, Genetic engineering using homologous recombination. Annual review of genetics, 2002. **36**(1): p. 361-388.
5. Olorunniji, F.J., S.J. Rosser, and W.M. Stark, Site-specific recombinases: molecular machines for the Genetic Revolution. Biochemical Journal, 2016. **473**(6): p. 673-684.
6. Ronda, C., et al., CRMAGE: CRISPR Optimized MAGE Recombineering. Scientific Reports. 2016. 19452.
7. Pines, G., et al., Bacterial recombineering: genome engineering via phage-based homologous recombination. ACS synthetic biology, 2015. **4**(11): p. 1176-1185.
8. Barrangou, R., et al., CRISPR provides acquired resistance against viruses in prokaryotes. Science, 2007. **315**(5819): p. 1709-1712.
9. Brouns, S.J., et al., Small CRISPR RNAs guide antiviral defense in prokaryotes. Science, 2008. **321**(5891): p. 960-964.
10. Garneau, J.E., et al., The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature, 2010. **468**(7320): p. 67-71.
11. Mougiakos, I., et al., Next Generation Prokaryotic Engineering: The CRISPR-Cas Toolkit. Trends Biotechnol, 2016. **34**(7): p. 575-87.
12. Barrangou, R., Cas9 targeting and the CRISPR revolution. Science, 2014. **344**(6185): p. 707-708.
13. Deltcheva, E., et al., CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature, 2011. **471**(7340): p. 602-+.
15. Deveau, H., et al., Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J Bacteriol, 2008. **190**(4): p. 1390-400.
16. Mojica, F., et al., Short motif sequences determin e the targets of the prokaryotic CRISPR defence system. Microbiology, 2009. **155**(3): p. 733-740.
17. Jinek, M., et al., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity .Science, 2012. **337**(6096): p. 816-21.
18. Komor, A.C., A.H. Badran, and D.R. Liu, CRISPR-based technologies for the manipulation of eukaryotic genomes. Cell, 2016.
19. Barrangou, R. and P. Horvath, A decade of discovery: CRISPR functions and applications. Nature microbiology, 2017. **2**: p. 17092.
20. Jiang, W., et al., RNA-guided editing of bacterial genomes using CRISPR-Cas systems. 2013. **31**(3): p. 233-239.
21. Oh, J.-H. and J.-P. van Pijkeren, CRISPR–Cas9-assisted recombineering in Lactobacillus reuteri. Nucleic Acids Research, 2014. **42**(17): p. e131-e131.
22. Mewalal, R., et al., Plant-derived terpenes: A feedstock for specialty biofuels. Trends in biotechnology, 2017. **35**(3): p. 227-240.
23. Delmulle, T., S.L. De Maeseneire, and M. De Mey, Challenges in the microbial production of flavonoids. Phytochemistry Reviews, 2017: p. 1-19.
24. Li, Y.F., et al., Metabolic engineering of *Escherichia coli* using CRISPR-Cas9 meditated genome editing. Metabolic Engineering, 2015. **31**: p. 13-21.
25. Garst, A.D., et al., Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. 2017. **35**(1): p. 48-55.
26. Liang, L.Y., et al., CRISPR EnAbled Trackable genome Engineering for isopropanol production in *Escherichia coli*. Metabolic Engineering, 2017. **41**: p. 1-10.
27. Heo, M.J., et al., Controlling Citrate Synthase Expression by CRISPR/Cas9 Genome Editing for n-Butanol Production in *Escherichia coli*. Acs Synthetic Biology, 2017. **6**(2): p. 182-189.
28. Wang, S.H., et al., Genome Editing in *Clostridium saccharoperbutylacetonicum* N1-4 with the CRISPR-Cas9 System. Applied and Environmental Microbiology, 2017. **83**(10).
29. Wasels, F., et al., A two-plasmid inducible CRISPR/Cas9 genome editing tool for *Clostridium acetobutylicum*. Journal of Microbiological Methods, 2017. **140**: p. 5-11.
30. Li, H., et al., CRISPR-Cas9 for the genome engineering of cyanobacteria and succinate production. Metabolic Engineering, 2016. **38**: p. 293-302.
31. Cho, J.S., et al., CRISPR/Cas9-coupled recombineering for metabolic engineering of *Corynebacterium glutamicum*. Metabolic Engineering, 2017. 42: p. 157-167.

32. Ding, W.W., et al., 5-Aminolevulinic acid production from inexpensive glucose by engineering the C4 pathway in *Escherichia coli*. Journal of Industrial Microbiology & Biotechnology, 2017. 44(8): p. 1127-1135.
33. Meng, J., et al., Improvement of pristinamycin I (PI) production in *Streptomyces pristinaespiralis* by metabolic engineering approaches. Synthetic and Systems Biotechnology, 2017. 2(2): p. 130-136.
34. Li, L., et al., Multiplexed site-specific genome engineering for overproducing bioactive secondary metabolites in actinomycetes. Metabolic Engineering, 2017. 40: p. 80-92.
35. Jiang, Y., et al., CRISPR-Cpf1 assisted genome editing of *Corynebacterium glutamicum*. Nature Communications, 2017. 8.
36. Wu, T., et al., Membrane engineering-A novel strategy to enhance the production and accumulation of β-carotene in *Escherichia coli*. Metabolic engineering, 2017. 43: p. 85-91.
37 . Xia, J., et al., Expression of Shewanella frigidimarina fatty acid metabolic genes in E-coli by CRISPR/ cas9-coupled lambda Red recombineering. Biotechnology Letters, 2016. 38(1): p. 117-122.
38. Qi, L.S., et al., Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell, 2013. 152(5): p. 1173-1183.
39. Li, D., et al., Controlling microbial PHB synthesis via CRISPRi. Applied Microbiology and Biotechnology, 2017. 101(14): p. 5861-5867.
40. Tao, W., L. Lv, and G.Q. Chen, Engineering Halomonas species TD01 for enhanced polyhydroxyalkanoates synthesis via CRISPRi. Microbial Cell Factories, 2017. 16.
41. Kim, S.K., et al., CRISPR interference-guided balancing of a biosynthetic mevalonate pathway increases terpenoid production. Metabolic Engineering, 2016. 38: p. 228-240.
42. Lv, L., et al., Application of CRISPRi for prokaryotic metabolic engineering involving multiple genes, a case study: Controllable P(3HB-co-4HB) biosynthesis. Metabolic Engineering, 2015. 29: p. 160-168.
43. Elhadi, D., et al., CRISPRi engineering *E. coli* for morphology diversification. Metabolic Engineering, 2016. 38: p. 358-369.
44. Shen, H.J., et al., Dynamic control of the mevalonate pathway expression for improved zeaxanthin production in *Escherichia coli* and comparative proteome analysis. Metabolic Engineering, 2016.38:p.180-190.
45. Cleto, S., et al., *Corynebacterium glutamicum* Metabolic Engineering with CRISPR Interference (CRISPRi). ACS Synthetic Biology, 2016. 5(5): p. 375-385.
46. Wu, J.J., et al., Enhancing flavonoid production by systematically tuning the central metabolic pathways based on a CRISPR interference system in *Escherichia coli*. Scientific Reports, 2015. 5.
47. Cress, B.F., et al., CRISPathBrick: Modular Combinatorial Assembly of Type II-A CRISPR Arrays for dCas9-Mediated Multiplex Transcriptional Repression in *E. coli*. Acs Synthetic Biology, 2015.4(9):p.987-1000.
48. Wu, J.J., et al., Efficient de novo synthesis of resveratrol by metabolically engineered *Escherichia coli*. Journal of Industrial Microbiology & Biotechnology, 2017. 44(7): p. 1083-1095.
49. Wu, J.J., et al., Rational modular design of metabolic network for efficient production of plant polyphenol pinosylvin. Scientific Reports, 2017. 7.
50. Cress, B.F., et al., CRISPRi-mediated metabolic engineering of E-coli for O-methylated anthocyanin production. Microbial Cell Factories, 2017. 16.
51. Wu, J.J., et al., A systematic optimization of medium chain fatty acid biosynthesis via the reverse beta oxidation cycle in *Escherichia coli*. Metabolic Engineering, 2017. 41: p. 115-124.
52. Kim, S.K., et al., CRISPR interference-guided multiplex repression of endogenous competing pathway genes for redirecting metabolic flux in *Escherichia coli*. Microbial cell factories, 2017. 16(1): p. 188.
53. Li, S.Y., et al., Enhanced protein and biochemical production using CRISPRi-based growth switches. Metabolic Engineering, 2016. 38: p. 274-284.
54. Wen, Z.Q., et al., Enhanced solvent production by metabolic engineering of a twin-clostridial consortium. Metabolic Engineering, 2017. 39: p. 38-48.
55. Wang, M.M., et al., CRISPRi based system for enhancing 1-butanol production in engineered Kiebsiella pneumoniae. Process Biochemistry, 2017. 56: p. 139-146.
56. Kaczmarzyk, D., et al., Diversion of the long-chain acyl-ACP pool in Synechocystis to fatty alcohols through CRISPRi repression of the essential phosphate acyltransferase PlsX. Metabolic engineering, 2017.
57. Huang, C.H., et al., CRISPR interference (CRISPRi) for gene regulation and succinate production in cyanobacterium S. elongatus PCC 7942. Microbial Cell Factories, 2016. 15.
58. Gordon, G.C., et al., CRISPR interference as a titratable, trans-acting regulatory tool for metabolic engineering in the cyanobacterium Synechococcus sp. strain PCC 7002. Metabolic Engineering, 2016. 38:p.170-179.
59. Bruder, M.R., et al., Extending CRISPR-Cas9 Technology from Genome Editing to Transcriptional Engineering in the Genus *Clostridium*. Applied and Environmental Microbiology, 2016. 82(20):
p. 6109-6119.

60. Yao, L., et al., Multiple gene repression in cyanobacteria using CRISPRi. ACS synthetic biology, 2015. 5(3): p. 207-212.
61. Yan, Q. and S.Y. Fong, Challenges and advances for genetic engineering of non-model bacteria and uses in consolidated bioprocessing. Frontiers in Microbiology, 2017. 8: p. 2060.
62. Mougiakos, I., et al., Efficient Genome Editing of a Facultative Thermophile Using Mesophilic spCas9. ACS Synthetic Biology, 2017.
63. Zhang, H., et al., A Novel and Efficient Method for Bacteria Genome Editing Employing both CRISPR/Cas9 and an Antibiotic Resistance Cassette. Frontiers in Microbiology, 2017. **8**.
64. Peng, F., et al., Efficient gene editing in *Corynebacterium glutamicum* using the CRISPR/Cas9 system. Microbial cell factories, 2017. **16**(1: p. 201.
65. Nakade, S., T. Yamamoto, and T. Sakuma, Cas9, Cpf1 and C2c1/2/3-What's next? Bioengineered, 2017. **8**(3: p. 265-273.
66. Abudayyeh, O.O., et al., C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science, 2016. **353**(6299.
67. Abudayyeh, O.O., et al., RNA targeting with CRISPR–Cas13. Nature, 2017. 550(7675): p. 280-284.
68. Li, Y., et al., Harnessing Type I and Type III CRISPR-Cas systems for genome editnig. Nucleic Acids Research, 2016. 44(4): p. e34-e34.
69. Stout, E., T. Klaenhammer, and R. Barrangou, CRISPR-Cas Technologies and Applications in Food Bacteria. Annual Review of Food Science and Technology, 2017. 8: p. 413-437..
70. Jenkinson, E. and P. Krabben, Targeted mutations. 2015, GB2530831.
71. Cong, L., et al., Multiplex genome engineering using CRISPR/Cas systems. Science, 2013. 339(6121): p. 819-23.
72. Esvelt, K.M., et al., Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nat Methods, 2013. 10(11): p. 1116-21.
73. Mitsunobu, H., et al., Beyond Native Cas9: Manipulating Genomic Information and Function. Trends in Biotechnology. 35(10): p. 983-996.
74. Yan, M.Y., et al., CRISPR-Cas12a-Assisted Recombineering in Bacteria. Appl Environ Microbiol, 2017. 83(17).
75. Tang, Q., C. Lou, and S.-J. Liu, Construction of an easy-to-use CRISPR-Cas9 system by patching anewly designed EXIT circuit. Journal of biological engineering, 2017. 11(1): p. 32.
76. Murugan, K., et al., The Revolution Continues: Newly Discovered Systems Expand the CRISPR-Cas Toolkit. Molecular cell, 2017. 68(1): p. 15-25.
77. Karvelis, T., G. Gasiunas, and V. Siksnys, Harnessing the natural diversity a*nd in* vitro evolution of Cas9 to expand the genome editing toolbox. Current opinion in microbiology, 2017. 37: p. 88-94.
78. Koonin, E.V., K.S. Makarova, and F. Zhang, Diversity, classification and evolution of CRISPR-Cas systems. Current Opinion in Microbiology, 2017. 37(Supplement C): p. 67-78.
79. Mougiakos, I., et al., Characterizing a thermostable Cas9 for bacterial genome editing and silencing.Nature Communications, 2017. 8(1): p. 1647.
80. Harrington, L., et al., A thermostable Cas9 with increased lifetime in human plasma. bioRxiv, 2017.



CHAPTER 8

EXPLORING AND ENHANCING THE POTENTIAL OF *BACILLUS SMITHII* ET 138 AS A GREEN CHEMICALS PRODUCTION PLATFORM ORGANISM

Ioannis Mougiakos[1], Elleke Bosma[1], Antonius H. P. van de Weijer[1], Meliawati Meliawati[1], Agnès Oromí Bosch[1], Sjef Boeren[2], John van der Oost[1], Richard van Kranenburg[1,3,*]

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

[2] Laboratory of Biochemistry, Wageningen University, Stippeneng 4, 6708 WE Wageningen, The Netherlands.

[3] Corbion, Arkelsedijk 46, 4206 AC Gorinchem, The Netherlands.

*Corresponding author

ABSTRACT

Modern society highly resides on fossil fuel resources for energy and chemicals. The study and exploitation of non-model microbial production platforms, along with the currently model microorganisms like *Escherichia coli* and *Saccharomyces cerevisiae*, will accelerate the transition to an economy that promotes the environmental friendly production of green chemicals and fuels by microbial fermentation of renewable biomass. *Bacillus smithii* ET 138 is a recently isolated bacterial strain with great potential as a biorefinery-based production platform. Amongst its other features, it is able to grow in a wide temperature range, it is genetically accessible, able to grow and produce organic acids under oxygen rich and limited conditions and tolerates low pH environments. In this study, we employed our SpCas9- based genome editing tool to substitute the native NAD+-dependent glyceraldehyde-3P dehydrogenase (*gapA*) gene, in the genome of the previously constructed lactate dehydrogenase deletion mutant *B. smithii* ET 138 Δ*ldhL* (TKO) strain, by the $NADP^+$-dependent *gapC* gene from *Clostridium acetobutylicum* ATCC 824. This substitution aimed to relieve the NAD+ deficiency of the TKO strain under oxygen-limited conditions, due to the deletion of its main NAD+ regenerative pathway, and resulted in moderate enhancement of its growth. We further employed the recently developed ThermoCas9-based genome editing tool to study the non-canonical acetate production pathway of *B. smithii* ET 138, by deleting the acetyl-CoA synthetase (*acs*) and acetyl-CoA hydrolase (*ach*) genes. Both deletions did not affect the acetate production of the mutant strains. Moreover, we cultured *B. smithii* ET 138 under controlled conditions at a wide temperature range aiming to reveal differences in growth, production and protein content. Finally, we engineered a *B. smithii* ET 138 strain that shows enhanced production of dicarboxylic acids and reduced production of acetate. This study underlined the interesting features of the *B. smithii* ET 138 metabolism and its potential to become a green chemicals production host, paving the way for further engineering and culturing studies.

KEYWORDS

CRISPR/Cas9, *Bacillus smithii*, thermophiles, metabolic engineering, temperature shift

INTRODUCTION

A bio-based economy aims towards the substitution of non-renewable energy and chemical resources by renewable biomass-based substrates for microbial fermentations in biorefineries[1,2]. Examples of valuable microbial products that can replace oil-based chemicals are dicarboxylic acids such as succinate and malate, which can be used in pharmaceuticals and cosmetics and as building blocks for bio-plastics. The development of microbial workhorses, that produce compounds of interest at high yields and titers, is based on the thorough understanding of their metabolism and on high-throughput metabolic engineering approaches with efficient genome engineering tools. Hence, till recently, only a handful of microbes, such as *Escherichia coli* and *Saccharomyces cerevisiae*, could serve as production platform organisms. Nonetheless, the recent development of generally applicable and efficient CRISPR-Cas genome engineering tools for a wide variety of bacteria can accelerate studies on their metabolism and unlock their biotechnological potential[3,4].

The use of moderately thermophilic instead of mesophilic bacteria in biorefineries presents several advantages, such as reduced cooling costs, elimination of contamination risks from mesophilic contaminants and reduced requirements for thermoactive saccharolytic enzymes[5]. The development of the thermotolerant ThermoCas9-based genome editing and silencing tools, applicable at temperatures at least up to 55°C, can facilitate the metabolic exploration and biotechnological exploitation of a wide range of thermophilic bacteria. *Bacillus smithii* ET 138 was the first thermophilic bacterium edited with the ThermoCas9-based tool[6] and it was selected for its genetic accessibility, wide growth temperature range, tolerance to pH fluctuations, ability to grow anaerobically and its interesting metabolism[7]. Bosma et al. previously demonstrated that *B. smithii* ET 138 mainly produces L-lactate under oxygen-limited and acetate under oxygen-rich culturing conditions[8]. Lactate production from pyruvate is catalyzed by L-lactate dehydrogenase (LdhL), which in parallel oxidizes an NADH molecule into NAD+. Deleting the *ldhL* gene from the *B. smithii* ET 138 genome[9] resulted in the complete elimination of L-lactate production under every tested condition. It also led to severe growth reduction under oxygen-limited conditions, most likely due to the lack of an alternative efficient NAD+-regenerating pathway[8]. Growth under aerobic conditions is not hampered by the *ldhL* deletion, with acetate being the main product under these conditions8. However, it is unknown how *B. smithii* produces acetate. None of the three currently sequenced *B. smithii* genomes, incl. ET 138, encompasses any phosphotransacetylase (*pta*) and acetate kinase (*ack*), or pyruvate oxidase (pox) genes, which constitute the common acetate production pathways in bacteria[10].

Altogether, to plan further metabolic engineering steps and enable the development of *B. smithii* into a platform organism it is crucial to 1) reconstitute the growth of *B. smithii* ET 138 Δ*ldhL* under oxygen-limited conditions, and 2) identify the alternative acetate production pathway employed by *B. smithii* ET 138. In this study, we aimed to achieve this by 1) substituting the native NAD^+-dependent glyceraldehyde 3P-dehydrogenase (*gapA*) gene for the $NADP^+$-dependent glyceraldehyde 3P-dehydrogenase (i) from *Clostridium acetobutylicum* ATCC 824 and 2) deleting the genes of corresponding enzymes that could catalyse acetate-producing reactions, namely acetyl-CoA hydrolase (*ach*) and acetyl-CoA synthetase (*acs*). We performed culturing experiments with the constructed mutant strains to evaluate the effect of the deletions on *B. smithii* ET 138 metabolism. To further improve our understanding of *B. smithii* metabolic pathways and fluxes, and optimize culture conditions, we additionally explored differences in growth, product profiles and proteome at temperatures between 37 and 55°C, between which the acetate and dicarboxylic acid production patterns showed striking differences. Finally, we engineered a strain with reduced carbon flux towards acetate and enhanced flux towards dicarboxylic acids under the optimized culturing conditions.

RESULTS & DISCUSSION

RECONSTITUTING *B. SMITHII* ET 138 Δ*LDHL* GROWTH UNDER OXYGEN-LIMITED CONDITIONS

Background knowledge

Lactate is the main fermentation product of *B. smithii* as the conversion of pyruvate to L-lactate via L-lactate dehydrogenase (LdhL) is the main NAD^+-regenerative route under oxygen-limited culturing conditions[8]. Earlier studies strongly suggested that the growth reduction of *B. smithii* ET 138 ΔldhL compared to the wild type under oxygen-limited conditions is the result of NAD^+-deficiency[8]. This hypothesis was supported by partial reconstitution of the *B. smithii* ET 138 ΔldhL growth under oxygen-limited conditions via plasmid-based expression of the NAD^+- and D-lactate producing- D-lactate dehydrogenase (LdhA) from *Lactobacillus delbrueckii* subsp. *bulgaricus*[8]. Metabolomics data on *B. smithii* ET 138 Δ*ldhL* fermentations revealed the intracellular accumulation of glyceraldehyde-3P (G3P), an upper glycolytic intermediate which is normally metabolised into 1,3-bisP-glycerate (1,3bPG) by the NAD^+-dependent glyceraldehyde 3P-dehydrogenase gene (*gapA*) or the $NADP^+$-dependent glyceraldehyde 3P-dehydrogenase gene (*gapB*)[8] (Fig. 1).

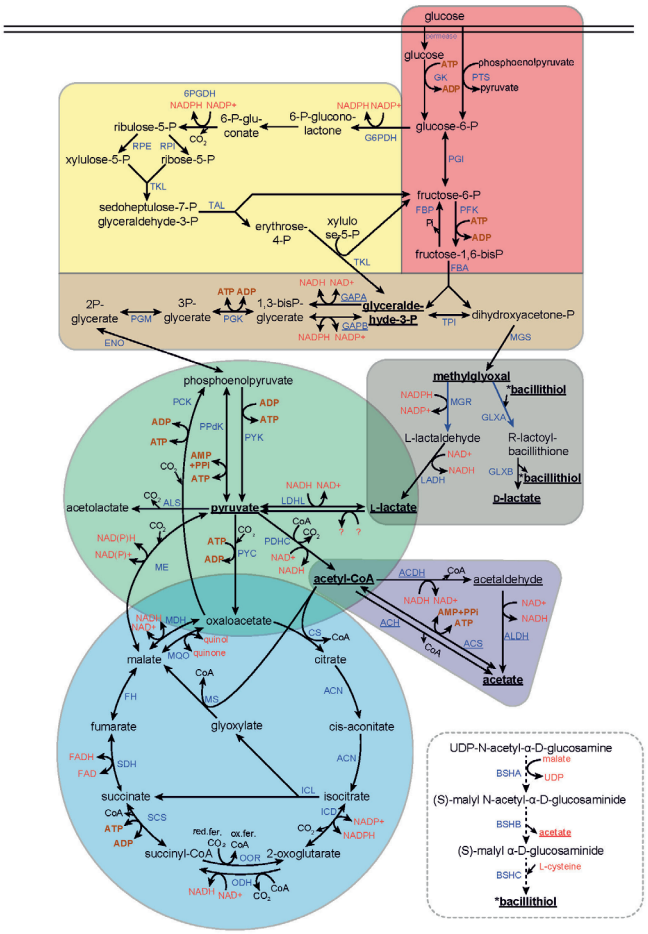


Figure 1. Schematic representation of the central glucose metabolism of *B. smithii* ET 138 along with candidate acetate production pathways and the bacillithiol production pathway. Coloured boxes indicate: red rectangular: upper glycolysis; orange rectangular: lower glycolysis; brown rectangular: pentose phosphate pathway; grey rectangular: methylglyoxal pathway; green eclipse: reactions connected to pyruvate metabolism; purple triangle: reactions examined in this study that could be responsible for the acetate production; blue circle: tricarboxylic acid (TCA) cycle; dashed square: bacillithiol production pathway. Blue arrows indicate reactions catalyzed by enzymes of which the EC-number was identified via domainome analysis while dashed arrows indicate reactions catalyzed by enzymes which have not been identified in the *B. smithii* ET 138 genome. Intermediates or producs denoted in bold and underlined have been discussed in this study. Enzymes underlined have been engineered and/or discussed in this study. Abbreviations: PTS: phosphotransferase system; GK: glucokinase; PGI: glucose-6-phosphate isomerase; G6PDH: glucose-6-phosphate dehydrogenase; 6PGDH: 6-phosphogluconate dehydrogenase; RPI: phosphopentose isomerase; RPE: phosphopentose epimerase; TKL: transketolase; TAL: transaldolase; FBP: fructose bisphosphatase; PFK: phosphofructokinase; FBA: fructose bis-phosphate aldolase; TPI: triosephosphate isomerase; GAP: glyceraldehyde 3-phosphate dehydrogenase; PGK: phosphoglycerate kinase; PGM: phosphoglycerate mutase; ENO: enolase; MGS: methylglyoxal synthase; MGR: methylglyoxal reductase; LADH: lactaldehyde dehydrogenase; GLXA: glyoxalase A; GLXB: glyoxalase B; PCK: phosphoenol pyruvate carboxykinase; PPC: phosphoenol pyruvate carboxylase; PYK: pyruvate kinase; PYC: pyruvate carboxylase; CS: citrate synthase; ACN: aconitase; ICL: isocitrate lyase; MS: malate synthase; ICD: isocitrate dehydrogenase; OOR: 2-oxoglutarate reductase; ODH: 2-oxoglutrate dehydrogenase; SCS: succinyl-CoA synthetase; SDH: succinate dehydrogenase; FH: fumarate hydratase; ME: malic enzyme; MDH: malate dehydrogenase; MQO: malate:qu inone oxidoreductase; ALS: acetolactate synthase; LDHL: L-lactate dehydrogenase; PDHC: pyruvate dehydrogenase complex; ACDH: acetyl-CoA dehydrogenase; ALDH: acetaldehyde dehydrogenase; ACS: acetyl-CoA synthetase; ACH: acety-CoA hydratase; BSHA: L-malate glucosyltransferase; BSHB: malyl-N-acetyl-D-glucosamine N-acetyl hydrolase; BSHC: bacillithiol synthase.

Transcriptomics data on the same *B. smithii* ET 138 Δ*ldhL* fermentations showed the expected transcriptional upregulation of the *gapA* and *gapB* genes, as well as the upregulation of the triosphosphate isomerase (*tpiA*) gene compared to the wild type control fermentations. The *tpiA* upregulation suggests that *B. smithii* ET 138 ΔldhL either purposely promotes the isomerisation of the formed dihydroxyacetone-phosphate (DHAP) into G3P, in its attempt to enhance the "stuck" lower glycolysis, or it tries to channel the accumulating G3P towards the methylglyoxal pathway, with D/L-lactate as final product, via its isomerization to DHAP (Fig. 1). Nonetheless, the transcriptomics data did not reveal any transcriptional upregulation of the methylglyoxal synthase (mgs) gene and the HPLC analysis of the *B. smithii* ET 138 Δ*ldhL* fermentation products did not show lactate accumulation, supporting the former scenario.

RECONSTITUTING *B. SMITHII* ET 138 Δ*LDHL* GROWTH UNDER OXYGEN-LIMITED CONDITIONS

Strain development via genetic engineering

The aim of this study was to overcome the metabolic bottleneck of the NAD+-deficiency, that results in intracellular G3P accumulation in *B. smithii* ET 138 Δ*ldhL* fermentations, and further expand the variety of *B. smithii* ET 138 fermentation products. We decided to substitute the NAD+-dependent *gapA* gene in the genome of *B. smithii* ET 138 with the NADP+-dependent *gapC* gene from *Clostridium acetobutylicum* ATCC 824. This substitution would result in the elimination of any NAD+-dependent reaction in glycolysis, relieving the NAD+-deficiency in *B. smithii* ET 138 Δ*ldhL* fermentations and allowing us to evaluate if the NADP+-availability is high enough to reconstitute growth and glucose consumption. Moreover, it was previously shown that the *gapA* to *gapC* substitution is an effective method for changing cofactor availability and redirecting metabolic pathways[11].

We used the *B. smithii* Δ*sigF*_Δ*ldhL*_Δ*hsdR* mutant strain[12], from now onwards denoted as TKO (Triple Knock-Out), as platform organism for the *gapA* substitution by *gapC*. Next to the inability to produce lactate, the TKO strain has also lost the ability to sporulate, and it does not encompass any active restriction/modification system that could hamper plasmid-based genome engineering experiments. The previously developed plasmid-based homologous recombination (HR)-SpCas9 counter selection system[12] was employed for the substitution of the *gapA* by the *gapC* gene in the *B. smithii* ET 138 genome (Fig. 2a). After *in silico* analysis[13] and confirmation that the that the codon usage landscape of *gapC* gene in its native genomic context would be comparable to the codon usage landscape in *B. smithii*, we proceeded with immediate PCR amplification of the *gapC* from the *C. acetobutylicum* ge-

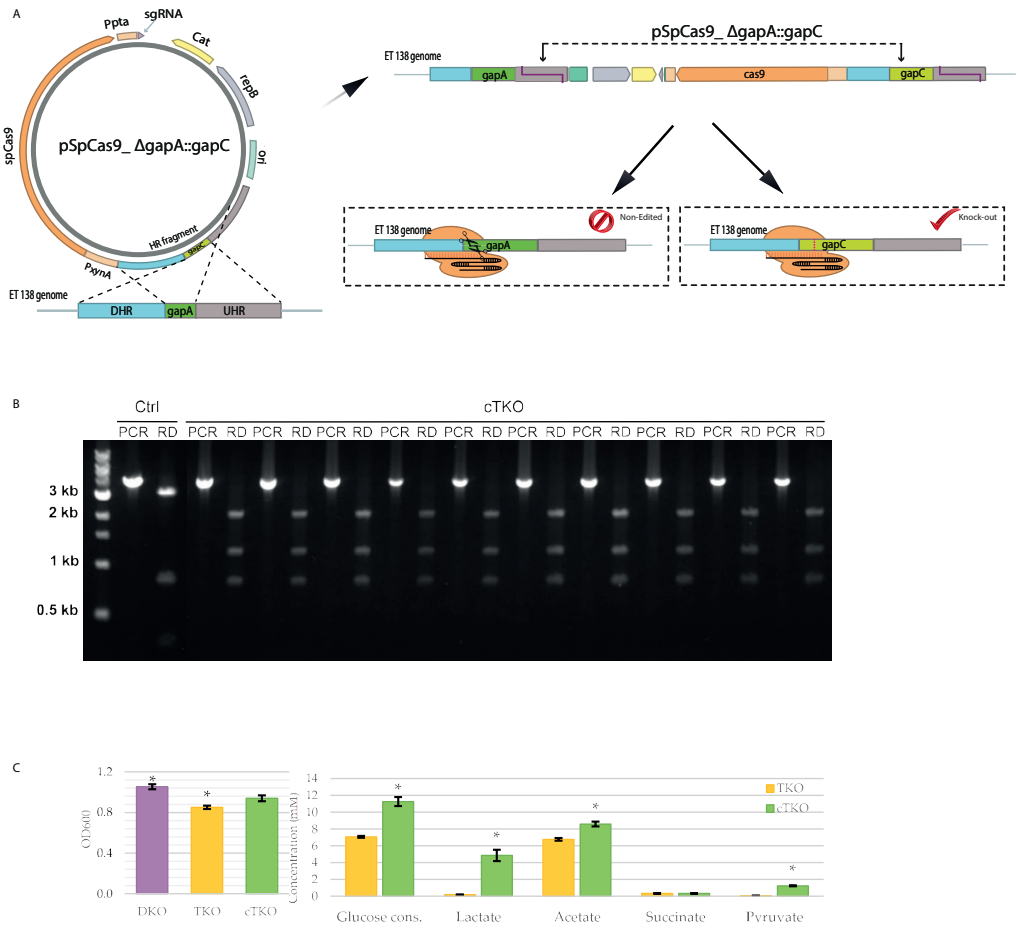


Figure 2. A) Schematic overview of the pSpCas9_ ΔgapA::gapC based editing process. The spcas9 gene was cloned into the pNW33n vector under the control of PxynA promoter, followed by the sgRNA-expressing module encompassing a spacer targeting the gapA gene within the genome of *B. smithii* ET 138. The sgRNA module is under the transcriptional control of the Ppta from B. coagulans. The 1kb upstream and 1kb downstream genomic region of the gapA gene are cloned in the pSpCas9_ ΔgapA::gapC vector upstream of the spcas9 and interspaced by the gapC gene from *Clostridium acetobutylicum* ATCC 824. After a first single crossover event, the vector is integrated into the genome. A second crossover event will either result in a wild type revertant genome, which is recognised and targeted by the spCas9:sgRNA ribonucleoprotein (RNP), or a genome with the desired insertion. The RNP-induced lethal double stranded DNA breaks eliminate the wild type cells and allow the rapid and easy screening of the mutant cells. B) Gel-electrophoresis of the PCR and restriction analysis products from cTKO *B. smithii* ET 138 strains. Genomic DNA was extracted from 10 cTKO cultures grown at 50˚C, and 4 kb fragments from the gapC genomic regions were PCR amplified using genome specific primers. The PCR products were digested with EcoRI, giving fragments of 315 bp, 809 bp and 2931 bp for the wild-type genotype and 809 bp, 1202 bp and 2041 bp for the mutated genotype.Genomic DNA from the wild-type strain was used as negative control. RD: Restriction digested product, PCR: PCR product. C) Growth and production profile of DKO, TKO and cTKO strains under micro-aerobic conditions. Samples were taken after 24 h of growth in 25 mL TVMY medium in 50 mL Greiner tubes at 50˚C. Data represent means of 3 replicates and error bars indicate standard deviations. Significant differences with the control strains were determined by Student's t test; *, P <0.05.

nome. Moreover, the 1 kb upstream (US) and downstream (DS) genomic regions of the *gapA* were PCR amplified from the *B. smithii* genome and all three fragments were cloned into the pSpCas9 vector, creating the pSpCas9_ ΔgapA::gapC vector (Fig. 2a). Moreover, the cloning was designed in such way that the *gapC* gene would be flanked by the US and DS regions as the *gapA* is flanked in the *B. smithii* genome. The spacer introduced into the sgRNA expressing module of the pSpCas9_ ΔgapA::gapC vector was designed to lead SpCas9 to introduce double stranded DNA breaks (DSB) in *gapA* upon incubation of the cells at 37°C, facilitating the selection of HR-based mutated cells over the wild-type cells (Fig. 2a). After transformation of TKO with the pSpCas9_ΔgapA::gapC vector and subsequent HR and Cas9-counter-selection, 28 randomly selected colonies were PCR-screened for clean *gapA* to *gapC* substitutions with genome-specific primers. To discriminate between *gapA* and *gapC* that generate PCR products of similar sizes, each PCR amplification product was subjected to restriction digestion with EcoRI. Out of the 28 PCR products, 27 had the restriction pattern corresponding to the clean mutant genotype (Fig. 2b). T he *gapA* to *gapC* genotype of 10 mutant colonies was confirmed via sequencing of the corresponding PCR-amplified fragments. Finally, plasmid curing was successfully performed by non-selective subculturing, resulting in the TKO _Δ*gapA*::*gapC* mutant strain, from now onwards denoted as cTKO.

RECONSTITUTING *B. SMITHII* ET 138 Δ*LDHL* GROWTH UNDER OXYGEN-LIMITED CONDITIONS

Phenotypic characterization of the *gapA* to *gapC* substitution strain

The effect of the chromosomal *gapA* to *gapC* replacement on growth and organic acids production was assessed by culturing the *B. smithii* cTKO, TKO and Δ*sigF*_Δ*hsdR* (from here onwards denoted as DKO) strains under oxygen-limited conditions (25 ml TVMY medium supplemented with 10 g/l glucose in 50 ml falcon tubes) for 24 h at 50°C. The culturing temperature was set at the optimal growth temperature of *C. acetobutylicum* ATCC 824, from which the *gapC* originates[14] and which is within the optimal growth temperature range of *B. smithii*[7]. Under the tested conditions, the growth of the cTKO cultures was enhanced compared to the TKO cultures, but they still did not reach the growth levels of the DKO cultures (Fig. 2c). The glucose consumption of the cTKO cultures was 41% higher compared to the TKO cultures (Fig. 2c). The higher cTKO glucose consumption was accompanied by a dramatic increase in lactate production, a moderate increase in acetate production and a minor increase in pyruvate production compared to the TKO cultures. These results indicate

that the NADP+-based G3P conversion to 1,3bPG in the cTKO strain affects the metabolism by changing the redox cofactors availability, increasing the glucose consumption and channeling carbon towards the methyglyoxal pathway.

As described in the background knowledge part, the major metabolic bottleneck of Δ*ldhL* cultures under oxygen-limited conditions is the NAD^+-dependent conversion of G3P to 1,3bPG by GapA. Here we show that the substitution of *gapA* by the $NADP^+$-dependent *gapC* in the cTKO genome relieves -to some extend- upper glycolysis from this bottleneck step and allows for higher glucose consumption and better growth of the cTKO cultures, compared to the TKO cultures. In line with this observation, it was previously reported for *E. coli* that the *gapA* to *gapC* substitution significantly enhanced the flux through the glycolytic pathway, whereas the flux through the Pentose Phosphate (PP) pathway was strongly reduced[11], as this important pathway for biomass production heavily relies on $NADP^+$. Hence, the incomplete growth restoration of the TKO cultures, compared to the DKO cultures, could be attributed to reduced flux through the PP pathway.

The cTKO cultures produce substantial amounts of lactate that most probably originates from the methylglyoxal pathway, considering the background deletion of the L-lactate dehydrogenase gene from the cTKO genome as well as the lack of a D-lactate dehydrogenase gene from the *B. smithii* ET 138 genome[10]. As previously mentioned, there was no indication of a transcriptionally active methylglyoxal pathway in the *B. smithii* ET 138 Δ*ldhL* fermentations, while surprisingly the cTKO fermentations in this study indicate activation of this pathway. While this remains to be experimentally proven, a possible explanation is the cofactor coupling of the $NADP^+$-consuming GapC to the $NADP^+$-producing methylglyoxal reductase (MgR) (Fig. 1): the conversion of methylglyoxal to L-lactaldehyde via MGR is accompanied by the production of $NADP^+$, which can be further used by GapC during glycolysis. Notably, the conversion of the toxic L-lactaldehyde to L-lactate requires an NAD^+ molecule[15], probably creating another NAD^+-deficiency issue. This might explain that the observed lactate production in the cTKO fermentations was higher than the TKO fermentations but much lower compared to the theoretical maximum of 2 lactate molecules produced per glucose molecule consumed. Moreover, the accumulation of toxic methylglyoxal could also negatively affect the growth of the cTKO cultures, unless it is further converted into D-lactate in a process that requires bacillithiol as cofactor (Fig. 1).

Lastly, Martinez *et al.*[11] demonstrated for the *E. coli* Δ*gapA*::*gapC* strain that the flux through the lower TCA cycle was decreased, as the conversion of isocitrate to oxoglutarate is an $NADP^+$-dependent process. This allowed for enhanced acetyl-CoA

conversion into acetate. Accordingly, the observed increase in acetate production from the cTKO fermentations (Fig. 2C) might be the result of lower acetyl-CoA incorporation into the TCA cycle (Fig. 1).

Altogether, the substitution of *gapA* by *gapC* in the TKO genome had a positive impact on the growth and glucose consumption of the TKO strain under oxygen-limited conditions, it redirected flux towards the methylglyoxal pathway, but likely introduced a NADP(H) redox imbalance or a methylglyoxal toxicity problem. Further metabolomics studies would be needed to confirm this, showing either lack of PP pathway intermediates or methylglyoxal accumulation. On the metabolic engineering aspect, the cTKO is a NADPH-producing strain that could serve as platform for metabolic engineering studies aiming at NADPH-dependent synthesis production pathways. Additionally, as it was recently demonstrated for *E. coli*, deletion of the *tpiA* gene from the cTKO genome could increase the flux towards the methylglyoxal pathway[16], while a limited number of additional and simple engineering steps could further result in simultaneous cofactor regeneration and production of biotechnologically interesting compounds, such as 1,2-propenediol or 1-propanol[17]. Alternatively, a study on *B. smithii* methylgyoxal synthase (*mgsA*) deletion mutants would be of great interest if aiming to direct the total consumed carbon towards lower glycolysis and induce the production of for example TCA cycle intermediates. Additionally, methylglyoxal has been shown to enhance catabolite repression in *E. coli*; acknowledging the differences between the catabolite repression system of Gram positive and Gram negative bacteria, it would be of great interest to study if the Δ*mgsA* deletion in combination with the deactivation of the phosphotransferase system could facilitate sugar co-metabolism in *B. smithii* ET 138[18]. Finally, the observed increase in acetate production for the cTKO fermentations stresses the major importance of identifying the acetate production pathway in *B. smithii* ET 138 for further engineering steps.

GENETIC BACKGROUND INVESTIGATION OF *B. SMITHII* ET 138 ACETATE PRODUCTION PATHWAY

Deletion of *ach* and *acs*

Acetate is the main product of *B. smithii* ET 138 under aerobic conditions, and the main by-product under microaerobic conditions[7]. In this part of our study, we attempted to reveal the *B. smithii* ET 138 genes responsible for the acetate production, since the canonical pathways based on the phosphotransacetylase (*pta*) - acetyl-CoA kinase (*ack*) or the pyruvate oxidase (*pox*) genes are not present in the B. *smithii* ET 138

genome. We reasoned to proceed by deleting two of the theoretically most probable candidate genes: the acetyl-CoA hydrolase (*ach*) and acetyl-CoA synthetase (*acs*) genes. Acetyl-CoA hydrolase (AcH) catalyses the hydrolysis of acetyl-CoA into acetate and CoA (Fig. 1), and it was previously found to be highly transcribed in *B. smithii* ET 138 cultures under aerobic but not under oxygen-limited conditions (unpublished data). Acetytl-CoA synthetase (AcS) catalyses the ATP-dependent reversible activation of acetate into acetyl-CoA (Fig. 1). Although the reaction favours the acetyl-CoA formation, the Gibbs free energy (ΔG°') of the acetate formation is -4.5 kJ/mol, hence thermodynamically feasible. Additionally, it was recently demonstrated that the AMP- and acetyl-CoA-forming ACS from *Syntrophus aciditrophicus* catalyses the production of acetate and ATP when the AMP-to-ATP ratio is high[19].

For the *ach* and *acs* deletions, we again used the TKO strain as platform strain, but this time we employed our previously developed homologous recombination (HR) - ThermoCas9-based counter selection system (Fig. 3a). The genomic regions 1 kb upstream of the start codon and 1 kb downstream of the stop codon of each target gene were PCR amplified and each pair was cloned as a fused 2kb long HR fragment into the editing vectors. For each deletion, we evaluated 4 different spacers, creating two 4-member libraries of editing vectors (pThermoCas9_Δach_ sp1/2/3/4 and pThermoCas9_Δacs_sp1/2/3/4). TKO was transformed with the two editing libraries, yielding similar numbers of colonies, on which we performed genome-specific colony PCR. Out of the 9 colonies from the Δ*ach* editing library, only 2 (both from the transformation with the pThermoCas9_Δach_sp1 vector) gave a colony PCR amplification product that corresponded to the expected Δ*ach* deletion size. However, both colonies appeared to be a mixture of wild-type and Δ*ach* cells (Fig. S1a). Therefore, we employed them both as the starting material for a sequential re-streaking process on LB2_7Cm plates until we obtained clean Δ*ach* mutant colonies, from now onwards denoted as QKO_Δ*ach* (Quadruple Knock-Out *ach*). On the contrary, 6 out of the 9 colonies transformed with the Δ*acs* editing library (all 6 originating from the transformations with the pThermoCas9_Δacs_sp1/2/3 vectors) gave a single PCR amplification product corresponding to the expected Δ*acs* deletion size (Fig. S1b), indicating that all 6 are clean Δ*acs* mutants, from now onwards denoted as QKO_Δ*acs* (Quadruple Knock-Out *acs*). Finally, plasmid curing was successfully performed by non-selective subculturing for both mutant strains, and a last round of genome specific PCR on genomic DNA isolated from four plasmid cured cultures per mutant confirmed the desired deletions (Fig. 3b).

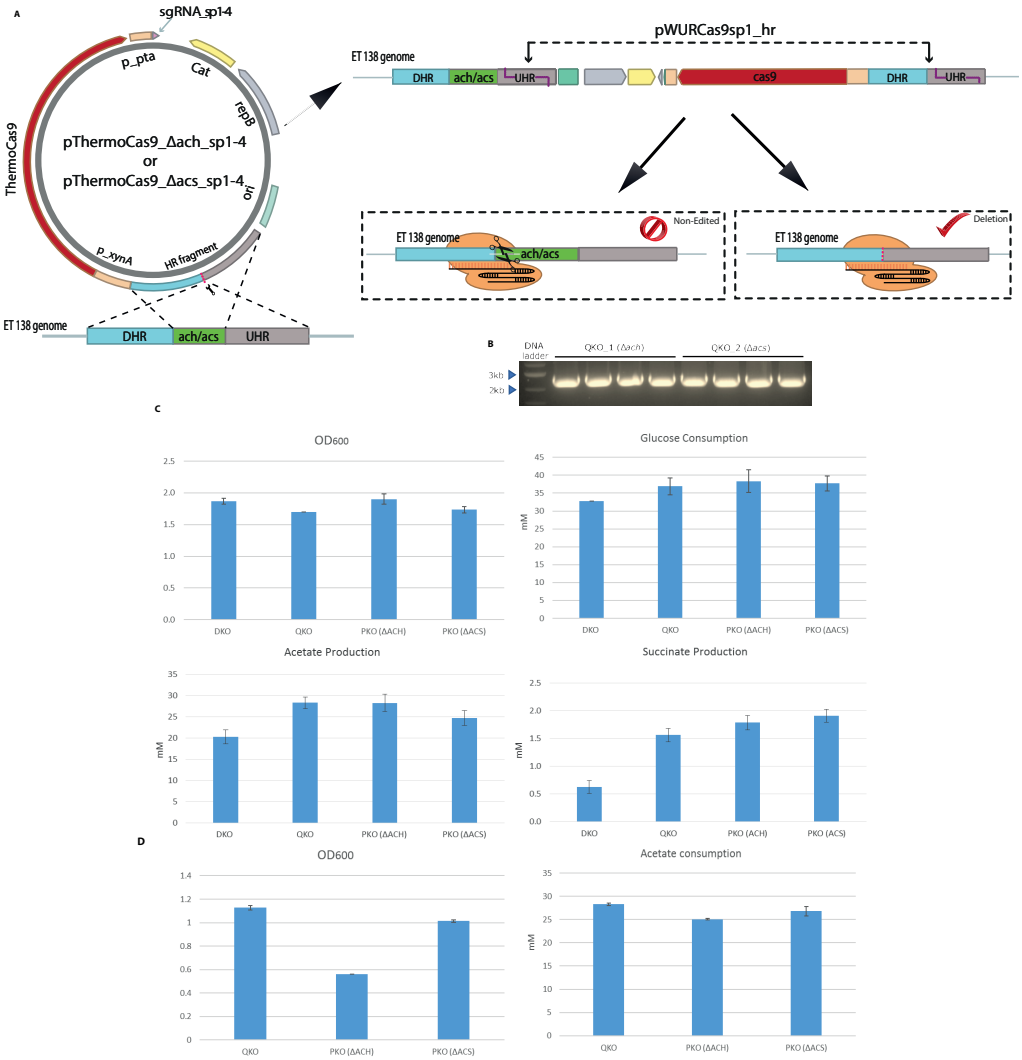


Figure 3. A) Schematic overview of the pThermoCas9_ Δach_sp or pThermoCas9_ Δacs_sp - based editing process. The *thermoCas9* gene was cloned into the pNW33n vector under the control of P_{xynA} promoter, followed by the sgRNA-expressing module encompassing a spacer targeting the *acs* or *ach* gene within the genome of *B. smithii* ET 138. The sgRNA module is under the transcriptional control of the P_{pta} from *B. coagulans*. The 1kb upstream and 1kb downstream genomic regions of the *ach* or the *acs* genes are cloned as fused fragments in either the pThermoCas9_ Δach_sp or the pThermoCas9_ Δacs_sp vectors respectively, upstream of the *thermocas9* gene. After a first single crossover event, the vectors are integrated into the genome. A second crossover event will either result in wild type revertant genomes, which is recognised and targeted by the spCas9:sgRNA ribonucleoprotein (RNP), or genomes with the desired deletions. The RNP-induced lethal double stranded DNA breaks eliminate the wild type cells and allow the rapid and easy screening of the mutant cells. B) Gel-electrophoresis of genome specific PCR products from plasmid cured QKO_Δ*ach* and QKO_Δ*acs* strains. Δ*ach* genotype is 2442 bp while the expected size for the wild-type genotype is 3963 bp. The expected size for the PCR fragments of the Δ*acs* genotype is 2494 bp while the expected size for the wild-type genotype is 4207 bp. C) Growth and production profile of the DKO, TKO, QKO_Δ*ach* and QKO_Δ*acs* strains, after 24 hours of culturing in 50 mL Greiner tubes containing 10 mL TVMY medium supplemented with 10 g/L glucose. Data shown are average of triplicates. D) Growth and acetate consumption profile of the TKO, QKO_Δ*ach* and QKO_Δ*acs* strains, after 24 hours of culturing in 50 mL Greiner tubes containing 10 mL TVMY medium supplemented with 5 g/L ammonium acetate. Data shown are average of triplicates.

GENETIC BACKGROUND INVESTIGATION OF *B. SMITHII* ET 138 ACETATE PRODUCTION PATHWAY

Phenotypic characterization of the Δ*ach* and Δ*acs* strains

The constructed QKO_Δ*ach* and QKO_Δ*acs* strains were cultured under aerobic/acetate-producing conditions: in 50ml falcon tubes containing 10 mL TVMY medium supplemented with 10 g/L glucose as carbon source. The DKO and the TKO strains were used as controls. The growth and glucose consumption of the DKO, TKO, QKO_Δ*ach* and QKO_Δ*acs* were comparable (Fig. 3c).Both QKO strains produced substantially higher amounts of acetate compared to the DKO control and comparable amounts to the TKO control, indicating that the *ach* and *acs* genes in the genome of *B. smithii* are not key elements of the acetate production pathway(s).

To see if the *ach* and *acs* genes are involved in acetate assimilation we looked at growth on acetate. Activation of acetate into acetyl-CoA is required for its further assimilation. It was previously reported that the growth of *Saccharomyces cerevisiae*[20] and *Neurospora crassa*[21] Δ*ach* mutant strains was not affected when glucose was used as the carbon source, while it was strongly inhibited when acetate was used as the carbon source. It has been shown that in *S. cerevisiae* acetyl-CoA hydrolase has a role in regulating the acetyl-CoA concentration, in order to prevent protein autoacetylation phenomena and generation of toxic metabolites[22]. In line with these studies, we decided to further investigate the ach and *acs* role in acetate metabolism.

The TKO, QKO_Δ*ach* and QKO_Δ*acs* strains were cultured as previously but in TVMY medium with 5 g/L ammonium acetate instead of glucose as the carbon source (Fig. 3d). The QKO_Δ*ach* growth, but not acetate consumption, was severely affected, only reaching a final OD_{600} of 0.5, while surprisingly the QKO_Δ*acs* growth and acetate consumption was comparable to the TKO strain. The deletion of the ach gene in the QKO_Δ*ach* strain in combination with the observed unaffected acetate consumption, when ammonium acetate was used as the carbon source, potentially led to toxic intracellular accumulation of acetyl-CoA that could have significantly affected the growth of the QKO_Δ*ach* cultures compared to the control cultures. Contrary to expectations, the QKO_Δ*acs* strain grew well in the medium with acetate as the carbon source. It was previously demonstrated for other microorganisms that disruption of the acs gene leads to very poor or even no growth on acetate[23,24]. As there is only one predicted acs homolog in the *B. smithii* ET 138 genome, the fact that QKO_Δ*acs* can grow on acetate indicates the existence of an alternative enzyme that overtakes the acetyl-CoA synthetase activity.

To date, only three enzymes have been shown to demonstrate acetate activation activity: the phosphotransacetylase-acetate kinase, propionyl-CoA synthetase and propionate kinase[25-29]. However, none of the corresponding genes are found in the genome of ET 138. Therefore, the pathway responsible for acetate activation in ET 138 remains to be elucidated.

Investigating the culturing temperature influence on the production profile of ET 138

The above results clearly indicate a lack of understanding of the metabolic pathways and fluxes in *B. smithii* ET 138. They also show strong differences in metabolism under different culturing conditions. Hence, we hypothesized that we could exploit these differences, shedding light on *B. smithii* ET 138 pathways and fluxes: since the acetate production pathway could not be identified via the gene deletions process, we aimed to discover and study high- and low-acetate producing culturing conditions. Not only does *B. smithii* ET 138 produce acetate under aerobic culturing conditions whereas lactate is its main product under oxygen-limited conditions, but also preliminary culturing experiments at different temperatures showed striking differences in product profiles. In this part of our study, we decided to explore how culturing temperature affects *B. smithii* ET 138 growth and organic acids production pattern under aerobic culturing conditions, with special focus on acetate and dicarboxylic acids.

The DKO strain was cultured at 37°C, 45°C, 50°C and 55°C in 1 L fermenters containing 500 mL TVMY liquid medium supplemented with 30 g/L glucose, under controlled pH and DO conditions (pH 6.5 and 10% DO). All the cultures were inoculated with precultures grown at the corresponding temperature, except for the 37°C culture for which the 45°C precultures were used due to the extremely long lag phase of the 37°C precultures. The maximum growth for the DKO 45°C, 50°C and 55°C cultures was observed 10 hours post-inoculation. Surprisingly, the 50°C cultures showed the highest growth, with maximum OD_{600} value at 5.35, and the 55°C cultures the lowest, with maximum OD_{600} value at 3.8 (Fig. 4a). The 37°C cultures showed an extended lag phase, as after 10 hours the growth of the DKO 37°C culture reached an OD_{600} value 1.20. The growth delay at 37°C could be attributed to the non-optimal growth temperature as well as to the use of precultures grown at 45°C. Despite the long lag phase, the 37°C cultures reached maximum OD600 value higher than the 45°C and 55°C cultures, but still lower compared to the 50°C cultures (Fig. 4a). The cell growth rate, the productivity and the glucose consumption rate are tightly interconnected in every bacterial culture. Here we observed that glucose consumption was increasing at hi-

gher temperatures (Fig. 4b). The DKO 55°C cultures consumed almost the entire amount of provided glucose between 24-48 hours of cultur ing, while the glucose consumption of the 50°C cultures was slightly lower. Since the final growth of the 55°C cultures was lower compared to the rest of the cultures, a relatively larger part of the consumed glucose was channeled towards products. After 10 hours of culturing, the 55°C cultures consumed 43 mM of glucose and produced 0.9 mol of acetate for every mol of consumed glucose (Fig. 4c). After 48 hours of culturing, however, the glucose consumption was more than triple (142 mM), there was no additional growth, and for every mol of consumed glucose there was still 0.9 mol of acetate produced together with 0.05 mol of succinate (Fig. 4c, 4d). This indicates a clear uncoupling of growth and production for the DKO cultures at 55°C, while the produced acetate is most likely the result of overflow metabolism. For the 50°C cultures the uncoupling of growth and production was also apparent but less pronounced. Interestingly, after 48 hours of culturing there was only 0.3 mol of acetate produced per mol of consumed glucose, while the succinate yield on glucose was double compared to the 55°C cultures (Fig. 4c, 4d). Additionally, after 48 hours the final glucose consumption of the 45°C and 37°C cultures was 60% and 70%, respectively, which is lower compared to the glucos consumption of the 55°C cultures and they produce negligible amounts of organic acids (Fig. 4a, 4b, 4c, 4d). Finally, except for acetate and succinate, there were no other organic acids produced at all tested temperatures.

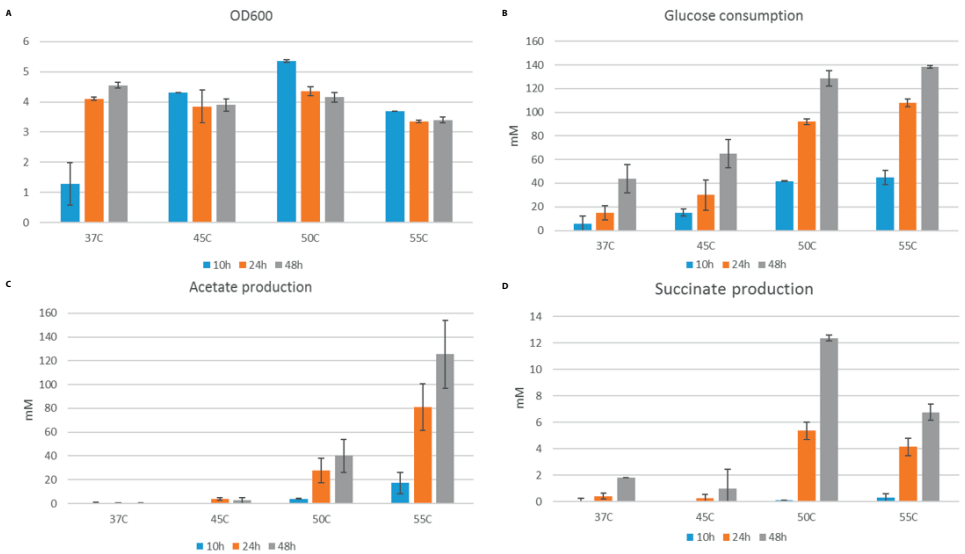


Figure 4. Culturing of DKO at 37°C, 45°C, 50°C and 55°C in 1 L fermenters containing 500 mL TVMY liquid medium supplemented with 30g/L glucose, under controlled pH and DO inoculation. conditions A) (pH Cell 6.5 growth and 10% (OD600), DO). B) Glucose Sampling points consumption, at 10, C) 24 Acetate and 48 production, hours post D) Succinate production. The data shown here are average of duplicates for the 37°C and 45°C cultures, triplicates for the 50°C and 55°C cultures, with error bars depicting the standard deviations

COMPARATIVE PROTEOMICS ANALYSIS BETWEEN ACETATE-PRODUCING AND NON-PRODUCING CONDITIONS AT 37°C AND 55°C

Our culturing experiments at different temperatures showed that the main *B. smithii* ET 138 product at 50°C and 55°C under aerobic conditions is acetate. On the contrary, there was no acetate production detected at the 37°C and 45°C *B. smithii* ET 138 cultures. Since the *B. smithii* ET 138 acetate production pathway still remains to be elucidated, we decided to follow a comparative proteomics analysis approach using two of the above described DKO cultures. Samples were taken from the 37°C and 55°C DKO cultures, 24 hours post inoculation, and the total soluble protein (SP) fraction was prepared from the corresponding cell-free extracts. The two samples were selected due to the significant difference detected in acetate production (0.07 mM of acetate at 37°C, 80 mM of acetate at 55°C) (Fig. 4c). The prepared SP fractions were processed by SDS-PAGE, in-gel digestion, μColumns-purification and LC-MS detection analysis. The comparative proteomics analysis revealed that there were 74 and 46 proteins more than 100-fold upregulated at 37°C and 55°C, respectively (Supp. Figure 3, Supp. table 1). From the proteins found to be significantly upregulated at 55°C, tw o are known to have direct or indirect connection with formation of acetate: the acetyl-CoA synthetase and the bacillithiol synthase. As in this study we already demonstrated that deletion of acetyl-CoA synthetase does not eliminate the acetate formation in *B. smithii* ET 138 cultures, we focused on the possible connection between the bacilithiol and acetate production processes.

Bacillithiol is the major low molecular weight thiol in bacilli and related firmicutes, and it plays an important role in cell physiology by maintaining the reducing cytoplasmic environment and by protecting cysteine residues from oxidation[30,31]. Moreover, it provides resistance to aldehydes and ketones, amongst which is the toxic methylglyoxal32. We earlier hypothesized that methylglyoxal could be an abundant intracellular intermediate in the cTKO fermentations at 55°C. Notably, these fermentations showed enhanced acetate production compared to the TKO control fermentations. Along the bacillithiol production process, acetate is formed as by-product (Fig. 1). Nonetheless, it has been demonstrated that the intracellular concentration of bacillithiol in other *Bacilli* does not usually exceed the level of 100 μM[30], hence it cannot justify the observed acetate production. Another possible reason for the triggering of the bacilithiol production at higher temperature cultures is the simultaneous production of acetolactate. Acetolactate synthases (AcS) are responsible for the conversion of pyruvate to acetolactate (Fig. 1). Some acetolactate synthases though, such as the ones from *E. coli*[33] and *S. typhimurium*, perform[34] oxygenase side reactions

producing peracetate and peroxide. Thiol peroxidases, that require thiols as cofactors[35], are responsible for the intra-cellular elimination of the highly reactive peroxide molecules, providing an explanation for the upregulation of the bacillithiol biosynthesis in *B. smithii* ET 138 cultures. Previously acquired transcriptomics data on *B. smithii* ET 138 cultures at 55°C[8] showed that transcription of the thiol peroxidase and the acetolactate synthase genes is enhanced under oxygen-rich compared to oxygen-limited conditions. Moreover, the comparative proteomics analysis data show a 4-fold higher expression of the thiol peroxidase in the 55°C culture compared to 37°C culture. However, the same analysis also showed that the production of the acetolactate synthase at the 37°C culture is higher compared to the 55°C culture, reducing the chances that the described acetate production scenario is the most probable.

In an alternative scenario, the acetate production in 55°C *B. smithii* ET 138 cultures can be connected with a metabolic "jamming" due to the rapid glucose consumption rate. As the TCA cycle reaches its maximum capacity, the acetyl-CoA can no longer be incorporated to the cycle thus accumulates inside the cells. Then, the excess acetyl-CoA can be further converted into acetaldehyde by the acetaldehyde dehydrogenase (AcdH) (Fig. 1). As acetaldehyde is a highly reactive and toxic compound, which either needs a compound as bacillithiol to be counteracted or has to be converted into something less toxic such as acetate, via for instance an aldehyde dehydrogenase (AldH) (Fig. 1). Nonetheless, the comparative proteomics analysis data though show that aldehyde dehydrogenases are significantly more abundant in the 37°C *B. smithii* ET 138 culture compared to the 55°C culture. Alternatively, it has been suggested that the autooxidation of acetaldehyde to peracetic acid is possible, followed by the subsequent reaction of the peracetic acid with acetaldehyde to form a peroxide. The peroxide can be further decompose to acetic acid and acetic anhydride[36]. Moreover, the peroxides formation might explain the upregulation of bacillithiol synthesis.

Finally, the comparative proteomics analysis data revealed the overexpression of the aminomethyltransferase protein (the "T-protein" of the glycine cleavage system) at the 55°C culture, which can be triggered only under high glycine concentration conditions. This is an interesting observation, as the proteomics data also suggest nitrogen-deprivation at the 55°C culture, indicated by the over production of the nitrogen regulatory protein PII37. Thus, we expect that glycine is abundant at the 55°C

culture despite the nitrogen-limitation. Hence there is a possible connection between glycine and acetate accumulation. L-threonine aldolase is an enzyme that catalyses the degradation of threonine acetaldehyde. As has been discussed, acetaldehyde can be converted to acetate. However, there is no predicted threonine aldolase gene in the genome of *B smithii* ET 138.

Even though we did not yet manage to elucidate the *B. smithii* ET 138 acetate production pathway, it is more than clear that follow-up experiments as well as further examination of the already acquired proteomics data is a priority for future work. Till date, a limited number of comparative proteomics studies have been performed in thermophiles, and especially in facultative thermophiles, on cultures incubated at significantly differing temperatures. Such studies could shed light on for example thermo-adaptation, of which little is currently known. In a recent study, *Bacillus licheniformis* was cultured at 42°C, 50°C and 60°C and the reported quantitative proteomics results indicated the overexpression at high temperatures of chaperones involved in facilitating correct protein folding, and of enzymes involved in the production of cell wall components, as well as growth inducing and stress response factors, as part of the amino acid and fatty acid metabolism[38]. Interestingly, the comparative proteomics data for *B. smithii* ET 138 present a similar trend in the type of overexpressed enzymes at 55°C, supporting the possible existence of a thermo-adaptation enzymatic machinery in moderate thermophilic bacteria. Furthermore, taking into account that the acetate production is apparently crucial for the *B. smithii* ET 138 metabolism, it would not be surprising if the acetate production enzymatic machinery is constantly expressed in the cells, aiming to allow for rapid adjustment to fluctuating environments based on allosteric regulations. In that case, the comparative proteomics approach is of little use and further construction and phenotypic characterization of additional *B. smithii* ET 138 mutants should be considered.

ENGINEERING AND CULTURING OF *B. SMITHII* ET 138 STRAINS TOWARDS PRODUCTION OF DICARBOXYLIC ACIDS

The previously described culturing experiments showed that DKO culturing under aerobic conditions at 50°C gives the best succinate to acetate production ratio. In this part of our study, we aimed to improve the dicarboxylic acids production and reduce the carbon loss in acetate production, without affecting the *B. smithii* ET 138 growth.

In large scale fermenters, even if they operate under oxygen-controlled conditions, the formation of oxygen-limited "pockets" is frequent. In these pockets, *B. smithii* ET 138 would metabolise the provided substrate into L-lactate. In order to entirely avoid carbon losses towards L-lactate, we decided to evaluate the production performance of the TKO strain, instead of the DKO, under oxygen-rich conditions. Moreover, since we are interested in the

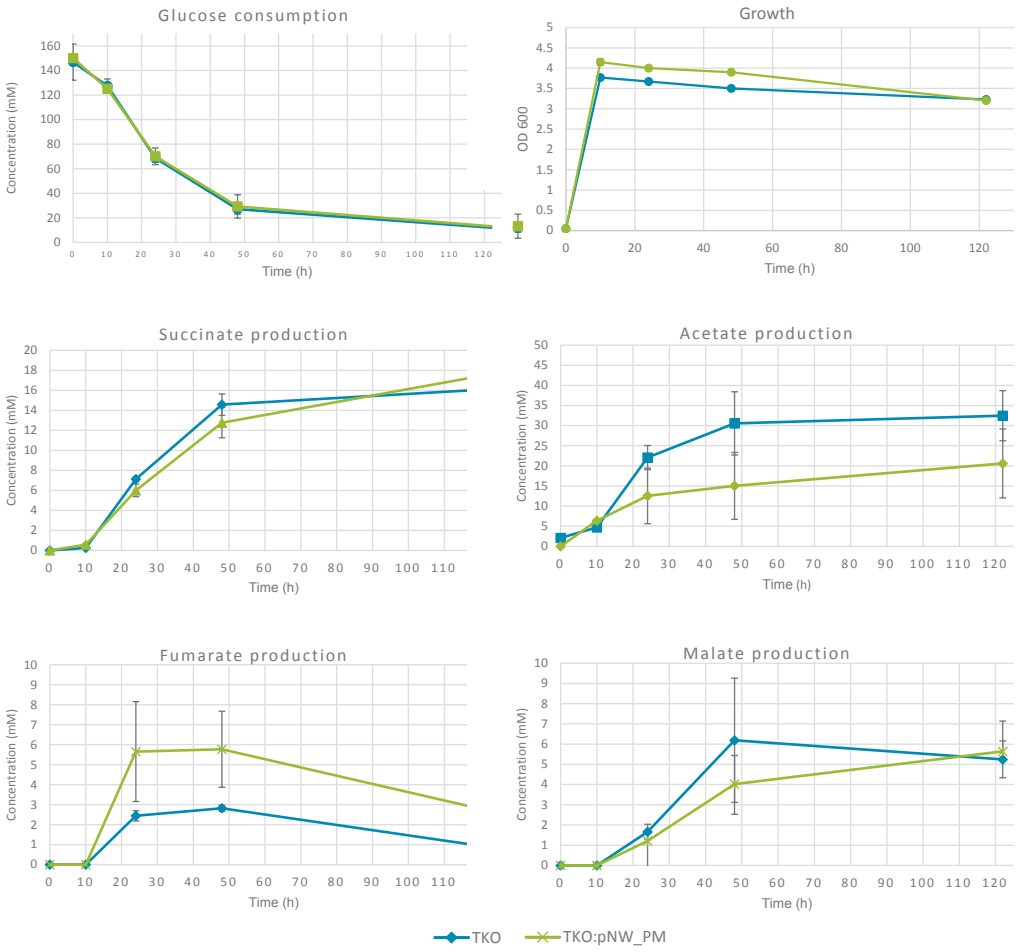


Figure 5. Aerobic cultures of the TKO and TKO:pNW_PM strains. Growth, glucose consumption and product profiles of the strains cultured in 500 mL TVMY supplemented with 30 g/L glucose at 50°C, under controlled pH and DO conditions (pH 6.5 and 10% DO). The data shown are average of triplicates with error bars depicting the standard deviations.

production of dicarboxylic acids, we reasoned to test the performance of a strain with an "enhanced" upper part of the reductive TCA cycle. Hence, we transformed the TKO strain with the pNW_PM vector that was constructed by employing the pNW33n *E. coli-Bacilli* shuttle vector as backbone and cloning the *B. smithii* ET 138 native py-

ruvate carboxylase (*pyc*) and the malate dehydrogenase (*mdh*) genes, as an operon, under the transcriptional control of the strong P_{pta} (phosphate acetyltransferase) promoter.

We cultured the TKO and TKO:pNW_PM strains in 1 L fermenters containing 500 mL TVMY liquid medium supplemented with 30g/L glucose, under controlled pH and DO conditions (pH 6.5 and 10% DO) at 50°C for 122h. Growth and glucose uptake rates of both strains were similar (Fig. 5a, 5b), however each strain produced different amounts of organic acids (Fig. 5c, 5d, 5e, 5f). TKO produced acetate and succinate tosimilar levels as previously reported for DKO under the same culturing conditions, with succinate production though being slightly enhanced (Fig. 5c, 5d). Surprisingly, TKO produced low amounts of malate and fumarate, becoming the first *B. smithii* ET 138 strain to produce these compounds (Fig. 5e, 5f). TKO:pNW_PM also produced acetate, succinate, malate and fumarate (Fig. 5c, 5d, 5e, 5f). However, the TKO:pNW_PM acetate production was significantly reduced compared to the DKO and TKO strains; only 21 mM of acetate produced after 122 hours of fermentation (Fig. 5c). At the same time point, TKO:pNW_PM produced 20 mM of succinate with a yield of 0.14 mol succinate/mol glucose (Fig. 5d). This is the highest reported succinate yield by a *B. smithii* ET 138 strain, nonetheless it is still far lower than the theoretical maximum yield of 1.7 mol succinate/mol glucose[39]. Moreover, TKO:pNW_PM produced similar amounts of malate compared to the TKO strain, while it produced 3 times higher amounts of fumarate after 48 hours (Fig. 5e, 5f). Apparently, the pyc and mdh overexpression enhanced the reductive part of the TCA cycle, with fumarate being its final product due to the lack of a fumarate reductase in the genome of *B. smithii* ET 138. A large part of the produced fumarate appeared to be consumed during the remaining course of the culturing process (Fig. 5f).

The TKO:pNW_PM strain developed in this experiment shows the same growth profile as the DKO and TKO strains under oxygen-rich conditions, while its acetate production yield is reduced and its dicarboxylic acids production yield is enhanced. Therefore, the TKO:pNW_PM strain should become the platform strain for further metabolic engineering experiments, especially those focusing on the identification of the acetate production pathway. It has been demonstrated that the deletion of the succinate dehydrogenase (*sdhA*), the isocitrate dehydrogenase (*icd*) and the isocitrate lyase repressor (*iclR*) genes in *E. coli* resulted in the construction of an efficient succinate-producing strain under oxygen-rich conditions[47]. We previously attempted to delete the genes for the different subunits of the succinate dehydrogenase complex without success. This was not surprising as the lack of a fumarate reductase gene from the *B. smithii* ET 138 genome makes the succinate dehydro

genase complex indispensable. The deletion of the *icd* and *iclR* genes though, along with overexpression of the native *pyc* and *mdh* genes, may increase the carbon flux into the glyoxylate shunt and hence increase the production of malate and fumarate.

CONCLUSIONS

Aiming to exploit the potential of nature towards the development of an sustainable economy, special attention should be paid on studying the metabolism of non-model microorganisms. This work underlined the very interesting differentiations in *B. smithii*, metabolism at different temperatures, as well as the great potential in improving its production profile towards dicarboxylic acids. The experimental work was based on our previously developed SpCas9- and ThermoCas9-based genome editing tools for gene deletions and substitutions, as well as culturing under controlled temperature and pH conditions and comparative proteomics analysis of selected samples. We achieved 1) the partial reconstitution of the *B. smithii* ET 138 Δ*ldhL* growth under oxygen-limited conditions and we simultaneously constructed a NADPH-producing strain for future expansion of the *B. smithii* product variety, 2) the elimination of a number of metabolic pathways candidates for the still unknown acetate production pathway of *B. smithii*, and 3) the revealing of strains and culturing conditions that favour the production of dicarboxylic acids on top of other by-products. Finally, the existence of efficient genome engineering tools and of a mutants collection that could be further tested under diverse conditions, strongly facilitates further fundamental and metabolic engineering studies of *B. smithii*.

MATERIALS AND METHODS

BACTERIAL STRAINS AND GROWTH CONDITIONS

Strains constructed and used in this study are listed in Table 1.

Strain	Relevant genotype	Reference
***E. coli* DH5α**	Cloning host for all the plasmid constructions	NEB
***B. smithii* ET 138 strains:**		
DKO	Δ*sigF* Δhsd*R*	[12]
TKO	Δ*sigF* Δ*hsdR* Δ*ldhL*	[12]
TKO:pNW_PM	Δ*sigF* Δ*hsdR* Δ*ldhL:* pNW_PM	This study
cTKO	Δ*sigF* Δ*hsdR* Δ*gapA*::*gapC*	This study
QKO_Δach	Δ*sigF* Δ*hsdR* Δ*ldhL* Δ*ach*	This study
QKO_Δacs	Δ*sigF* Δ*hsdR* Δ*ldhL* Δ*acs*	This study

Table 1. Strains used in this study

E. coli DH5α chemically competent cells (NEB) were employed as cloning hosts and routinely cultured in LB medium at 37°C and 200 rpm. LB agar plates were prepared by adding 15 g of agar (Difco) per liter of LB medium. When required, chloramphenicol was added up to 25 μg/mL final concentration. All *B. smithii* strains were routinely cultured at 55° C and 150rpm,unless stated otherwise, in TVMY or LB2 medium. The media were prepared as described previously[9] and the corresponding agar plates were prepared by adding 30 g of agar (Difco) per liter of medium. Electro-competent *B. smithii* cells were prepared and transformed with plasmid DNA as previously described[7]. For the *B. smithii* ET 138 culturing experiments in tubes, 10 ml TVMY medium cultures were inoculated with the desired cells from the corresponding glycerol stocks and cultured overnight in 50 mL Greiner tubes. After overnight growth, 100 μL of each pre-culture was transferred to 10 mL fresh TVMY medium -in 50 mL Greiner tube- for experiments under oxygen-rich conditions and 250 μL of each pre-culture was transferred to 25 mL fresh TVMY medium -in 50 mL Greiner tube- for experiments under oxygen-limited conditions. When required, chloramphenicol was added up to 7 μg/ mL final concentration.

PLASMID CONSTRUCTION AND *B. SMITHII* EDITING

All the plasmids constructed and/or used in this study are listed in Supp. table 2 together with the primer numbers and the templates used for the construction process. The primers were designed with appropriate overhangs for NEBuilder HiFi DNA assembly (NEB) and their sequences are listed in Supp. table 3. The fragments for assembling the plasmids were obtained through PCR with Q5 Polymerase (NEB), 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), according to the manufacturer's protocol. Upon overnight incubation of the transformation mixture on LB agar plates supplemented with antibiotic, single colonies were inoculated in LB medium supplemented antibiotic. Upon overnight incubation, plasmid material was isolated from each culture using the GeneJet plasmid miniprep kit (Thermo Fisher Scientific). Each plasmid was sequence verified (Macrogen, Inc) employing primers listed in Supp. table 4 and 1 μg from each verified construct was transformed into the desired *B. smithii* ET 138 strain. The MasterPure Gram Positive DNA Purification Kit (Epicentre) was used for genomic DNA isolation from *B. smithii* cultures, while the Phire Plant Direct PCR kit (Thermo Scientific) was used for colony PCR on *B. smithii* ET 138 colonies, according to the manufacturers' protocols.

The construction of the cTKO strain was based on the pSpCas9_ ΔgapA::gapC editing vector. Competent TKO cells were prepared and transformed with the pSpCas9_ ΔgapA::gapC vector. The transformation reaction was plated on LB2 selection plate supplemented with 7 μg/ml chloramphenicol (LB2_7Cam) at 55°C. Three of the formed colonies were used for inoculation of 10ml TVMY liquid medium supplemented with 7 μg/ml chloramphenicol (TVMY_7Cam) and the cultures were incubated at 55°C for 8 hours. 100 μl from each culture were used for inoculation of fresh 10 ml TVMY_7Cam liquid media and the cultures were incubated at 45°C, 150rpm for 8hours. These first two incubation steps constitute the homologous recombination phase of the process. The incubation step was repeated once more with incubation temperature set at 37°C. This last incubation step constitutes the counter selection phase of the process. 100 μl from each culture were plated on TVMY_7Cam agar plates and after 24 h randomly selected colonies were PCR screened for clean *gapA* to *gapC* substitutions with genome specific primers (Supp. table 5) and the PCR fragments were subjected to restriction digestion with the EcoRI restriction enzyme (NEB). The PCR and digestion products were subjected to 1% agarose gel electrophoresis and the PCR products with a digestion pattern that corresponds to the desired substitution were purified using Zymogen gel DNA recovery kit (Zymo Research) and sequence verified (Macrogen, Inc).

The construction of the QKO_Δ*ach* and QKO_Δ*acs* strains was based on the pThermoCas9_ Δach_sp1-4 and pThermoCas9_ Δacs_sp1-4 editing vectors. Competent TKO cells were prepared and transformed with the pThermoCas9_ Δach_sp1-4 and pThermoCas9_ Δacs_sp1-4 vectors. The transformation reactions were plated on LB2_7Cam plates and incubated at 55°C. Randomly selected colonies were PCR screened for Δ*ach* and Δ*acs* genotypes with genome specific primers (Supp. table 5) and the PCR products were subjected to 1% agarose gel electrophoresis. The PCR products with size that corresponds to the desired deletion were purified using Zymogen gel DNA recovery kit (Zymo Research) and sequence verified (Macrogen, Inc).

PLASMID CURING

For the plasmid curing processes, colonies that contained the desired genotypes were inoculated in 10 mL LB2 medium in 50 mL Greiner tubes. The cultures were incubated at 55˚C, and every 24 hours 100 μL of culture were re-inoculated in fresh media. After 5 days, 200 μL of 10^{-4} culture dilutions were plated on LB2 agar plates. After overnight incubation at 50-55˚C, single colonies were streaked on LB2 and LB2_7Cam plates. Subsequently, colo-

nies growing only in the LB2 plates were subjected to cPCR with *spcas9*- and *thermocas9*-specific primers (Supp. table 6). The PCR products were subjected to 1% agarose gel electrophoresis, and colonies that resulted in cPCR without product were considered cured from plasmid. Cured colonies were further cultured in LB2 and LB2_7Cam media to further confirm the absence of the plasmid.

ANALYTICAL METHODS

B. smithii cell growth was monitored by measuring the optical density at 600 nm (OD_{600}) using a UV-1800 UV-VIS spectrophotometer (Shimadzu). Acidification of the cultures in the tube culturing experiments was determined by measuring the pH using a ProLine benchtop pH meter (ProSense B.V.). Extracellular concentrations of glucose and fermentation products were determined using the high pressure liquid chromatography (HPLC) system ICS-5000 (Thermo Scientific). The system was equipped with an Aminex HPX 87H column (Bio-Rad Laboratories), a UV1000 detector operating at 210 nm and a RI-150 refraction index detector at 40°C. The mobile phase consisted of 0.016 N H_2SO_4, and the column was operated at a flow rate of 0.8 mL/min at 80°C. For the preparation of the HPLC samples, 1 mL culture samples were removed at the desired time point from each culturing experiment and they were centrifuged for 10 min at room temperature (RT) at 13,000 rpm. The resulting supernatant was stored at -20˚C or diluted 4:1 with 10 mM DMSO in 0.04 N H_2SO_4 (internal standard) and directly used for HPLC analysis.

CULTURING IN BIOREACTORS

The culturing in bioreactors experiments were conducted in glass reactors of 1L working volume (Eppendorf) operated via the DASGIP® software. The bioreactors contained 500 mL of modified TVMY medium (TVMY without MOPS and supplemented with 30 g/L glucose), as well as antibiotic when required. The temperature was controlled to remain at the desired levels and the pH was controlled to maintain at 6.5 via addition of 3 M KOH. The precultures for the bioreactor culturing experiments were prepared upon inoculation of 10 mL TVMY medium in 50 mL Greiner tubes, supplemented with antibiotic when required, with cells from the glycerol stocks. After overnight culturing at the selected temperatures 100μl from the precultures were transferred to 20 mL fresh TVMY medium in 50 mL Greiner tubes to an initial OD_{600} of 0.08. Upon reaching an OD_{600} of 0.5-0.7 (exponential growth phase), the cultures were used as starting material for the inoculation of the bioreactors to an initial OD_{600} of approximately 0.015. Samples of about 2 mL were taken at several time points, OD_{600} was measured to monitor cell growth and glucose and fermentation products were quantified by

HPLC. Furthermore, when required, chloramphenicol was added to the bioreactors every 24 h in order to maintain the chloramphenicol concentration above 6 µg/mL. Chemical degradation of chloramphenicolwas calculated by assuming a half-life time of 100 h at 55˚C[40]. Concentration of chloramphenicol at a given time point can be determined based on the equation: $C(t) = C0 \times 0.5100$. Antifoam was added to the bioreactors in case of foaming.

PROTEOMICS

CELL-FREE EXTRACT (CFE PREPARATION)

Selected culture samples were sent for proteomics analysis. The samples were collected from fermenters and pelleted by centrifugation at >14000 rpm, 4°C and for 15 minutes. The cell pellets were further washed via resuspension in 500 µL ice-cold sterile phosphate buffer solution pH7 (PBS,. The resuspended pellets were then centrifuged, the supernatants were discarded and the washing step was repeated. Cell pellets from the second washing step were resuspended in 500 µL of 100 mM Tris-HCl buffer pH 7.6 containing 50 mM DTT. 250 µL of each resuspension mixture were stored at -20°C for further use, while the rest 250 µL were transferred into a sterile bead-beating tube supplemented with sterile beads. The cells were then disrupted by bead beating at 6000 rpm for 40 seconds, using the Percellys24 Homogenizer (Bertin Instruments). The lysis mixtures were centrifuged at >14000 rpm and 4°C for 10 minutes. The supernatants, which were the cell free extracts (CFE), were transferred to 2.0 mL low binding protein tubes and used for further analysis. The total protein content of the CFEs was determined using Roti®-Nanoquant (Roth), according to manufacturer's protocol.

SDS-PAGE

In a clean 1.5 mL low binding protein tube, the CFE samples were mixed with 4x sample loading buffer (Bio-Rad) containing 10% β-mercaptoethanol. The mixtures were incubated at 95°C for 10 minutes, then briefly centrifuged. 50 µL of each mixture were loaded in the wells of a 1 mm Mini-PROTEAN® TGXTM Precast Gel (Bio-Rad). The gel was subjected to electrophoresis (25 mA, 30 min) for protein separation using the PowerPacTM HC HigH-Current Power Supply (Bio-Rad) unit. At the end of the run, the gel was rinsed with miliQ water and incubated overnight in excess Commassie Blue stain with gentle shaking. Subsequently the gel was rinsed several times with miliQ water , until no trace of Commassie Blue stain could be observed. During the whole procedure it was important to keep the gel covered to prevent keratin contamination. The gel could then be directly used for the in-gel digestion step or stored at 4°C until further use.

IN-GEL DIGESTION

The previously described sds-page gel was incubated overnight at RT in 25 mL of 50 mM NH4HCO3 and 10 mM DTT pH 8 solution. Following the incubation, the gel was rinsed with miliQ water and washed with 25 mL of a solution containing 22,5 mL dH2O, 2.5 mL 1M Tris pH 8 and 20 mM iodoacetamide pH 8. Using a scalpel, the sample from each well was divided into 5 small fractions. Each gel fraction was again cut into smaller pieces then transferred to a clean 0.5 mL low protein low binding tube. Fresh 50 µL trypsin solution was added to each tube, followed by overnight incubation at RT. The following day, the pH of each mixture was adjusted to between 2 and 4, with the addition of 10% trifluoroacetic acid (TFA).

µCOLUMNS PREPARATION

The trypsinated samples were cleaned, using µColumns, for the liquid chromatography-mass spectrometry (LC-MS) analysis step. For the µColumns preparation, a small piece of C18 Empore disk (frit) was transferred to a 200 µl tip, then 200 µl methanol were added to the tip. 50% slurry LichroprepC18 column material in methanol was also prepared, and 4 µl of this mixture was added to the tip containing the frit and the methanol. The content of the µ-Column was then eluted either by hand, with a 10 mL plastic syringe, or with vacuum manifold. The µColumns washing step was repeated with 100 µl of methanol. Subsequently the µColumns were eluted with 100 µl of 1ml/L HCOOH in dH2O. The samples after the in-gel digestion were added to the µ-Columns and the µ-Columns were washed with 100 µl of 1ml/L HCOOH and transferred to clean 0.5 mL low binding tubes. The peptides were manually eluted from the µ-Columns upon addition of 50 µl of a solution containing 50% Acetonitrile and 50% 1ml/L HCOOH in water. The samples were then concentrated via incubation at 45°C for at least 2 hours, until the final volume was < 20 µl. Finally, the samples were adjusted to final volume of 50 µL and then sonicat ed for 5 seconds.

STATISTICAL ANALYSIS

Statistical significances of growth, glucose consumption and acids production differences between genetically engineered and control strains were evaluated by performing the two-sided Student's t-test. The values were presented as the mean ± standard error of 2 or 3 replicates. A P value <0.05 was considered to be statistically significant.

AUTHOR CONTRIBUTIONS

IM, EFB, AHPvdW, MM, AOB, were involved in the design and execution of the experiments and in the data analysis; IM wrote the manuscript with contributions from EFB, MM, AOB; SB was involved in the experimental design and the data analysis of the proteomics study; JvdO and RvK were involved in the design and coordination of the study and in the revision of the manuscript. All authors have read and approved the final manscript.

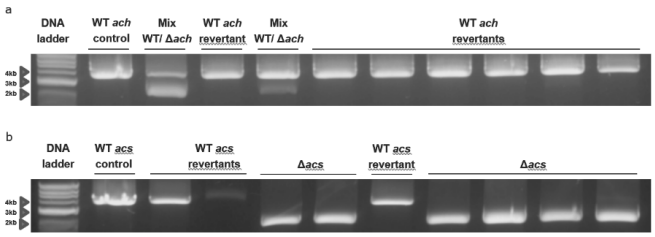
CONFLICT OF INTEREST

The authors declare to have no conflict of interest. RvK is employed by the company Corbion (Gorinchem, the Netherlands).

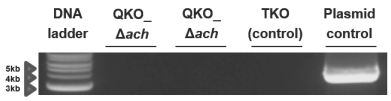
ACKNOWLEDGEMENTS

We thank Ioannis Kostopoulos (Wageningen University) for his contribution in the proteomics samples preparation and data analysis. We thank Mihris Naduthodi for the interesting and fruitful discussions.

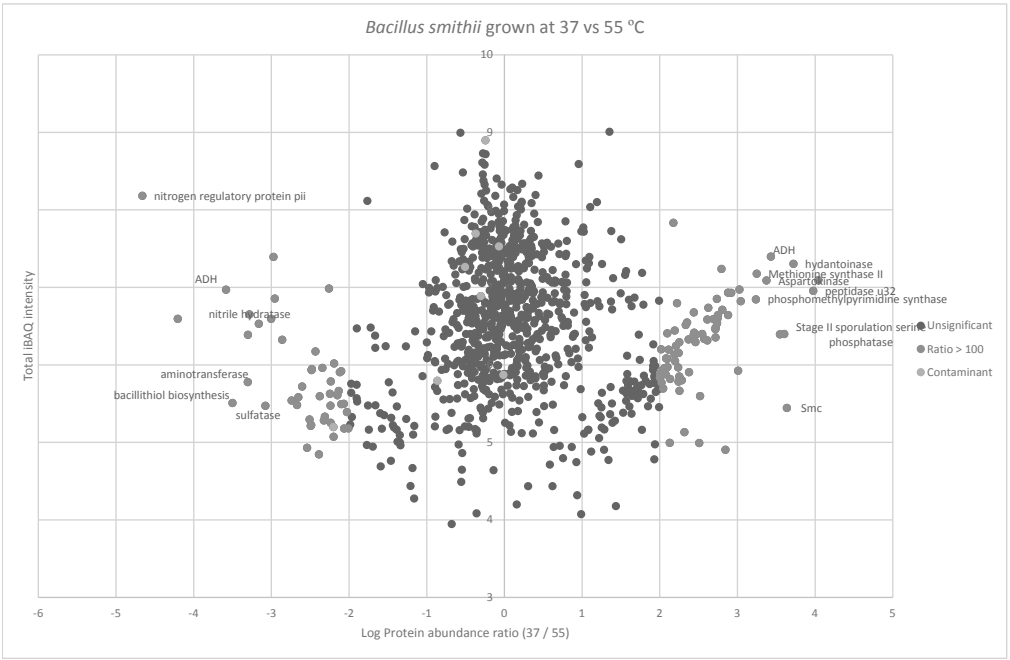
Supplementary Figures



Supp. Figure 1. Gel-electrophoresis of genome specific colony PCR products from *B. smithii* ET 138 colonies formed upon transformation with pThermo_Δach_sp1 and pThermo_Δacs_sp1/2/3 editing vectors. The Δ*ach* genotype corresponds to 2442 bp while the wild-type genotype corresponds to 3963 bp. The expected size for the PCR fragments of the Δ*acs* genotype is 2494 bp while the expected size for the wild-type genotype fragments is 4207 bp.



Supp. Figure 2. . Gel-electrophoresis of plasmid specific PCR products on DNA isolated from *B. smithii* ET 138 cultures. The QKO_Δ*ach* and QKO_Δ*acs* cultures were inoculated by colonies that resulted from the -described in this study- plasmid curing processes, while the TKO negative control culture was inoculated by a glycerol stock of already plasmid cured cells. The pThermo_Δach_sp1 vector was employed as positive control.



Supp. Figure 3. Graphical representation of the comparative proteomics data analysis. The graph depicts the logarithmic protein abundance ratio between proteins isolated from *B. smithii* ET 138 (strain DKO) cultures at 37°C and 55°C. The blue dots correspond to proteins with logarithmic abundance (37/55) ratio ≤2 and ≥-2 (insignificant difference between 37°C and 55°C), the orange dots correspond to proteins with logarithmic abundance (37/55) ratio ≥2 and ≤-2 (significant difference between 37°C and 55°C) and the grey dots correspond to contaminants. The proteins with the highest abundance at 37°C and 55°C are noted.

Supplementary tables

Supp. table 1. List of *B. smithii* ET 138 proteins significantly more abundant either at 37°C or 55°C, as extracted from the proteomics data analysis. The Log protein values correspond to the logarithmic protein abundance ratio between 37°C and 55°C. Only proteins with Log ratio ≥ 2 or ≤-2 are listed.

Name	Log Protein	Protein IDs
hydantoinase b/oxoprolinase	4.04	138update_0816
peptidase u32	3.98	138update_2403;A0A0H4P5Y9
hydantoinase/oxoprolinase	3.72	138update_0815
sm cs flexible hinge	3.64	138update_1453;A0A0H4NZ49
stage ii sporulation protein e	3.60	138update_0073;A0A0H4PC65
Hypothetical protein	3.55	138update_0268;A0A0H4NZU7
Alcohol dehydrogenase	3.43	A0A0H4NZ36;138update_0586
Aspartokinase	3.37	A0A0H4NW59;138update_0280
Methionine synthase II	3.25	A0A0H4PLN6;138update_3312
phosphom ethylpyrim idine synthase	3.24	138update_1819
Arginine decarboxylase	3.04	A0A0H4NXQ8;138update_0035
Foldase protein PrsA	3.02	A0A0H4NWW6;138update_0938; A0A0H4P7C4;138update_2819
cobalam in-independent m ethionine synthase	3.01	138update_3306;A0A0H4P7W7
transcription sporulation two-com ponent dna-binding phosphorylation sensory regulation	2.92	138update_2251;A0A0H4PIL5
Acetolactate synthase	2.88	A0A0H4NXP3;138update_0905
RNA polym erase sigm a factor	2.88	A0A0H4P1D5;138update_1403
Uncharacterized protein Bsub YpbR	2.84	A0A0H4P126;138update_2067
Quinolinate synthase	2.80	138update_2451;A0A0H4P651
Glycine cleavage system H protein	2.79	A0A0H4P618;138update_2797
O-succinylhom oserine sulfhydrylase	2.75	A0A0H4NZ30;138update_1687
Hypothetical protein	2.74	138update_0572
Anthranilate phosphoribosyltransferase	2.74	A0A0H4P0M6;138update_1929
Glyoxalase fam ily protein	2.73	A0A0H4P295;138update_1223
Dihydrolipoam ide acetyltransferase com ponent of pyruvate dehydrogenase com plex	2.72	138update_0537;A0A0H4P0K7
Acyl-coenzym e A synthetases/AMP-fattyacid ligases YtcI-like	2.72	A0A0H4PJK4;138update_2600
Carbon starvation protein	2.72	A0A0H4P4G6;138update_3136
Acetoin dehydrogenase E1 com ponent beta-subunit	2.69	A0A0H4PDK4;138update_0536
Uncharacterized protein	2.62	A0A0H4P8J6;138update_3191
acetyl-coa carboxylase carboxyl transferase beta subunit	2.61	138update_2575;A0A0H4P591

Name	Log Protein	Protein IDs
glycosyl transferases group 1	2.61	138update_2685
Diam inopim elate decarboxylas	2.57	A0A0H4P3Z2;138update_2183
Hydrolase HAD superfam ilyin cluster with DUF1447	2.55	A0A0H4P0N3;138update_1185
Two-com ponent sensor histidine kinase	2.52	A0A0H4NXN7;138update_1281
Transcriptional regulatorTetR fam ily	2.51	A0A0H4PF47;138update_1171
Aspartokinase	2.47	A0A0H4P324;138update_1901
Glycine oxidase ThiO	2.46	A0A0H4P0S2;138update_0628
saicar synthetase signature	2.44	138update_3029;A0A0H4PKT0
peptidase	2.44	138update_2826;A0A0H4P3F6
phosphorylase pyridoxal-phosphate attachm ent site	2.44	138update_2684;A0A0H4P5S9
s-adenosyl-l-m ethionine-dependent m ethyltransferase	2.44	138update_2966
tRNA uridine 5-carboxym ethylam inom ethyl m odificatio enzyme MnmG	2.38	A0A0H4PLT1;138update_3376
phospholipase/carboxylesterase/thioesterase	2.37	138update_1670
protein secy signature	2.35	138update_0145;A0A0H4NVM4
Anti-sigm a F factor antagonist	2.33	A0A0H4P0N1;138update_2189
Hypothetical protein	2.32	138update_2622;A0A0H4P5F1
six-hairpin glycosidase	2.28	138update_1941;A0A0H4P0N7
Stage V sporulation protein AF SpoVAF	2.25	A0A0H4P0M9;138update_2184
5,10-m ethylenetetrahydrofolate reductase	2.25	A0A0H4PDZ0;138update_0688
Succinyl-CoA:3-ketoacid-coenzym e A transferase	2.25	A0A0H4P1B5;138update_0819
Hydroxym ethylpyrim idine phosphate kinase	2.24	A0A0H4P1M7;138update_2214
asparagine synthase	2.24	138update_2659;A0A0H4P5Q5
Hypothetical protein	2.24	138update_2873;A0A0H4P2S3
ribosom al protein s6 signature	2.22	138update_3365;A0A0H4P559
Pili retraction protein pilT	2.21	A0A0H4NVH5;138update_0099
Glycosyltransferase	2.20	A0A0H4P127;138update_2319
fatty acid synthase	2.19	138update_1682;A0A0H4NZT6
Thiazole synthase	2.19	A0A0H4NW37;138update_0626
sporulation stage 0 protein	2.18	138update_2031;A0A0H4P3G9
Glycogen synthase	2.13	A0A0H4P2B5
Single-stranded-DNA-specific exonuclease RecJ	2.13	A0A0H4P625;138update_2432
Methylthioribulose-1-phosphate dehydratase	2.11	A0A0H4P6Z1;138update_3051
dehydrogenase e1 com ponent	2.10	138update_0535;A0A0H4NVY3
Hypothetical protein	2.10	138update_1655
Diami nopim elate epim erase	2.10	A0A0H4P0V6;138update_2014
heat shock protein	2.08	138update_0080;A0A0H4NXT5

Name	Log Protein	Protein IDs
Aldehyde dehydrogenase	2.08	A0A0H4P418;138update_1744
S1 RNA binding dom ain	2.06	A0A0H4P2L5;138update_2823
forma te-tetrahydrofolate ligase signature	2.04	138update_2058;A0A0H4P086
Carbamo yl-phosphate synthase sm all chain	2.02	A0A0H4NZ11;138update_1420
Aspartokinase	2.02	A0A0H4P533;138update_2533
Chrom osom e plasm id partitioning protein ParB Stage 0 sporulation protein	2.02	138update_3372;A0A0H4P950
Succinyl-CoA:3-ketoacid-coenzym e A transferase subunit A	2.02	A0A0H4PEB5;138update_0818
aminotransferase class iv	2.00	138update_0658
Cell shape-determ ining protein MreC	2.00	A0A0H4PJ66;138update_2469
denosylm ethionine-8-ami no-7-oxononanoate am inotransferase	-2.01	138update_2229;A0A0H4P1N9
Xanthine dehydrogenaseFAD binding subunit	-2.03	A0A0H4P4N2;138update_1967
Phosphopentom utase	-2.06	A0A0H4P1I3;138update_2192
Sialic acid utilization regulatorRpiR fam ily	-2.07	A0A0H4P453;138update_2220
1-deoxy-D-xylulose 5-phosphate reductoisom erase	-2.10	A0A0H4PG46;138update_1514
Sulfur carrier protein adenylyltransferase ThiF	-2.10	A0A0H4NZ78;138update_0625
Transition state regulatory protein AbrB	-2.12	A0A0H4NXR4;138update_0045
pyridine nucleotide-disulphide oxidoreductase class-ii	-2.13	138update_2152;A0A0H4P1C5
Am ino acid ABC transporter	-2.15	A0A0H4P671;138update_2838
Imidazoleglycerol-phosphate dehydratase	-2.19	A0A0H4NWL9;138update_0474
Acetyl-coenzym e A synthetase	-2.20	A0A0H4P214
Am inom ethyltransferase	-2.24	A0A0H4P4C1;138update_2285
Flagellar M-ring protein	-2.24	A0A0H4PG15;138update_1480
Coenzym e A biosynthesis bifunctional protein CoaBC	-2.24	138update_1432;A0A0H4NY73
ClpB protein	-2.25	A0A0H4P2H3;138update_1287
30S ribosom al protein	-2.26	A0A0H4NZ53;138update_1458
l-lactate perm ease	-2.31	138update_3342;A0A0H4P803
DNA polym erase X fam ily	-2.32	A0A0H4P6E3;138update_2542
transform ing protein p21 ras signature	-2.34	138update_3377;A0A0H4P955
Ribosom e-binding ATPase YchF	-2.38	A0A0H4P4C3;138update_3366
Am ino acid perm ease	-2.39	A0A0H4P3T5;138update_2921
Transcriptional repressor NrdR	-2.43	A0A0H4P6J0;138update_2559
RND multidrug efflux transporter Acriflavin resistance protein	-2.48	A0A0H4P8E8

Name	Log Protein	Protein IDs
Nucleotide-binding protein	-2.48	A0A0H4NWM5;138update_0482
90kda heat shock protein signature	-2.49	138update_1706
PTS system N-acetylglucosam ine-specific IIC com ponent	-2.49	A0A0H4P0C5;138update_0458
Cell division topological determ inant MinJ	-2.51	A0A0H4NYT1;138update_0447
Flagellar hook-associated protein 2	-2.54	A0A0H4NWH0;138update_0423
Two-com ponent response regulator SA1424	-2.60	A0A0H4P812;138update_3353
Cell wall lytic activity	-2.65	A0A0H4NYP2;138update_0396
DNA repair protein RecN	-2.67	A0A0H4P1R5;138update_2253
Niacin transporter NiaP	-2.74	A0A0H4P615
Uncharacterized protein	-2.86	A0A0H4PCL1;138update_0210
Am m onium transporter	-2.96	A0A0H4P8F1;138update_3143
Acyl carrier protein	-2.97	A0A0H4PFY0;138update_1451
Uncharacterized protein	-3.00	A0A0H4P2G9;138update_2772
sulfatase	-3.08	138update_3296;A0A0H4P7V3
nitrile hydratase alpha /thiocyanate hydrolase gam m a	-3.16	138update_1625
nitrile hydratase beta subunit	-3.28	138update_1624
Hypothetical protein	-3.30	138update_0518
bacillithiol biosynthesis bshc	-3.30	A0A0H4NZF9;138update_0188
bacillithiol biosynthesis; Putative cysteine ligase	-3.50	138update_1387;A0A0H4PFR9
Alcohol dehydrogenase	-3.58	A0A0H4P012;138update_0356
Hypothetical protein	-4.20	138update_3172
nitrogen regulatory protein pii	-4.66	138update_3142;A0A0H4PL50

Supp. table 2. List of plasm ids constructed and/or used during this study, together with the prim er num bers and the PCR tem plates em ployed for their construction via Gibson assem bly.

Plasmid	Fragment for Gibson assembly	PCR Template	Primer No.
pSpCas9_ ΔgapA::gapC	*gapA* spacer_ pnw33N backbone	pNW_spCas9 [12]	BG9870
			BG9871
	gapA Upstream region	*B. smithii* ET138 genome	BG9872
			BG9873
	gapC	*C. acetobutylicum* genome	BG9874
			BG9875
	gapA Downstream region	*B. smithii* ET138 genome	BG9876
			BG9877
	SpCas9_*gapA* spacer	pNW_spCas9[12]	BG9878
			BG9879

pThermoCas9_Δach_sp1	ThermoCas9_*ach*_sp1	pThermoCas9_ctrl [6]	BG11773
			BG11774
	*ach*_sp1_pNW33n backbone	pThermoCas9_ctrl [6]	BG11775
			BG11776
	ach Upstream region	*B. smithii* ET138 genome	BG11777
			BG11778
	ach Downstream region	*B. smithii* ET138 genome	BG11779
			BG11780
pThermoCas9_Δach_sp2	ThermoCas9_*ach*_sp2	pThermoCas9_ctrl [6]	BG11773
			BG11782
	*ach*_sp2_pNW33n backbone	pThermoCas9_ctrl[6]	BG11781
			BG11775
	ach Upstream region	*B. smithii* ET138 genome	BG11777
			BG11778
	ach Downstream region	*B. smithii* ET138 genome	BG11779
			BG11780
pThermoCas9_Δach_sp3	ThermoCas9_*ach*_sp3	pThermoCas9_ctrl [6]	BG11773
			BG11784
	*ach*_sp3_pNW33n backbone	pThermoCas9_ctrl [6]	BG11783
			BG11775
	ach Upstream region	*B. smithii* ET138 genome	BG11777
			BG11778
	ach Downstream region	*B. smithii* ET138 genome	BG11779
			BG11780
pThermoCas9_Δach_sp4	ThermoCas9_*ach*_sp4	pThermoCas9_ctrl [6]	BG11773
			BG11786
	*ach*_sp4_pNW33n backbone	pThermoCas9_ctrl [6]	BG11785
			BG11775
	ach Upstream region	*B. smithii* ET138 genome	BG11777
			BG11778
	ach Downstream region	*B. smithii* ET138 genome	BG11779
			BG11780
pThermoCas9_Δach_sp1	ThermoCas9_*acs*_sp1	pThermoCas9_ctrl [6]	BG11773
			BG11796
	*acs*_sp1_pNW33n backbone	pThermoCas9_ctrl [6]	BG11795
			BG11775
	acs Upstream region	*B. smithii* ET138 genome	BG11797
			BG11798
	acs Downstream region	*B. smithii* ET138 genome	BG11799
			BG11800

pThermoCas9_Δach_sp2	ThermoCas9_*acs*_sp2	pThermoCas9_ctrl [6]	BG11773 BG11802
	*acs*_sp2_pNW33n backbone	pThermoCas9_ctrl [6]	BG11801 BG11775
	acs Upstream region	*B. smithii* ET138 genome	BG11797 BG11798
	acs Downstream region	*B. smithii* ET138 genome	BG11799 BG11800
pThermoCas9_Δach_sp3	ThermoCas9_*acs*_sp3	pThermoCas9_ctrl [6]	BG11773 BG11804
	*acs*_sp3_pNW33n backbone	pThermoCas9_ctrl [6]	BG11803 BG11775
	acs Upstream region	*B. smithii* ET138 genome	BG11797 BG11798
	acs Downstream region	*B. smithii* ET138 genome	BG11799 BG11800
pThermoCas9_Δach_sp4	ThermoCas9_*acs*_sp4	pThermoCas9_ctrl [6]	BG11773 BG11806
	*acs*_sp4_pNW33n backbone	pThermoCas9_ctrl	BG11805 BG11775
	acs Upstream region	***B**. smithii* ET138 genome	BG11797 BG11798
	acs Downstream region	*B. smithii* ET138 genome	BG11799 BG11800
pNW_PM	[8]		

Supp. table 3. Sequences of the primers used for the plasmid construction processes.

Primer No.	**Sequence**
BG9870	GAACGGTTGTCATCATACCGGTTTTAGAGCTAGAAATAGCAAG
BG9871	TCATGACCAAAATCCCTTAAC
BG9872	tcacgttaagggattttggtcatgaGTACCAGACCTGCTGCAAATC
BG9873	AACCATTAATAGCTATCTTTGCCATTTTAAAAATTCCTCCTTCATAG

Primer No.	Sequence
BG9874	ATGGCAAAGATAGCTATTAATGGTTTTG
BG9875	AATTTCCAATATTCTTCAGTTTATACTATTTTGCTATTTTTGCAAAGTAAGC
BG9876	TATAAACTGAAGAATATTGGAAATTG
BG9877	attatcctcagctcactagcgccatATGTCTAAAGCTTCCCAATC
BG9878	ATGGCGCTAGTGAGCTG
BG9879	CGGTATGATGACAACCGTTCGTATAACGGTATCCATTTTAAGAATAATCC
BG11773	GAACATCCTCTTTCTTAGTTTG
BG11774	aacagcggtctcttcatccggcgGTATAACGGTATCCATTTTAAGAAT
BG11775	cgccggatgaagagaccgctgttGTCATAGTTCCCCTGAGATTATCG
BG11776	TCATGACCAAAATCCCTTAACGTG
BG11777	cacgttaagggattttggtcatgaGTAAACGTTCTTTTAAGCACGGAATC
BG11778	CAAAAATCCCCCCTTTGATTGTT
BG11779	aacaatcaaaggggggatttttgTAATTTTAAACTCCCGTTTCCAGACG
BG11780	caaactaagaaagaggatgttcATCGTGCAGATTATGATTACTCTGTC
BG11781	ctattcaatcagggatcggatccGTCATAGTTCCCCTGAGATTATCG
BG11782	ggatccgatccctgattgaatagGTATAACGGTATCCATTTTAAGAAT
BG11783	tccggatcaagaggaatgccgatGTCATAGTTCCCCTGAGATTATCG
BG11784	atcggcattcctcttgatccggaGTATAACGGTATCCATTTTAAGAAT
BG11785	cgtcttgaagaacttcagaatatGTCATAGTTCCCCTGAGATTATCG
BG11786	atattctgaagttcttcaagacgGTATAACGGTATCCATTTTAAGAAT
BG11795	aagatgagacacatatcgactttGTCATAGTTCCCCTGAGATTATCG
BG11796	aaagtcgatatgtgtctcatcttGTATAACGGTATCCATTTTAAGAAT

Primer No.	Sequence
BG11797	cacgttaagggattttggtcatgaCGATGTATTTTCCATAATCACGCAAG
BG11798	tagttcgaaatcgttgtttcattgtc
BG11799	gacaatgaaacaacgatttcgaactaCATTTGACCGATTCCCCCTTTC
BG11800	caaactaagaaagaggatgttcTGAAGTGCATCTCGAATTGAAGC
BG11801	tattcggaccttggttaatgggaGTCATAGTTCCCCTGAGATTATCG
BG11802	tcccattaaccaaggtccgaataGTATAACGGTATCCATTTTAAGAAT
BG11803	cacaatcacgttggatgctcccaGTCATAGTTCCCCTGAGATTATCG
BG11804	tgggagcatccaacgtgattgtgGTATAACGGTATCCATTTTAAGAAT
BG11805	tgtgtctcatcttcttgaatatcGTCATAGTTCCCCTGAGATTATCG
BG11806	gatattcaagaagatgagacacaGTATAACGGTATCCATTTTAAGAAT

Supp. table 4. List of primers used in this study for sequence verification of the employed editing plasmids.

Sequencing target	Primer No	Sequence
pSpCas9_ΔgapA::gapC	BG8478	TATTTACGTTGACGAATCTTG
	BG9216	AGGGCTCGCCTTTGGGAAG
	BG9874	ATGGCAAAGATAGCTATTAATGGTTTTG
	BG9875	AATTTCCAATATTCTTCAGTTTATACTATTTTGCTATTTTTGCAAAGTAAGC
	BG8487	GTATTTCCCAGAACCTTGAAC
	BG8472	GGAAATACAGACCGCCACAG
Commo n for pTherm oCas9_Δach_sp1-4 & pTherm oCas9_Δacs_sp1-4	BG3664	AGGGCTCGCCTTTGGGAAG
	BG6769	CAATCCAACTGGGCTTGAC
	BG6840	TTGCAGAAATGGTTGTCAAG
	BG6841	CAAGAACTTTATTGGTATAG
	BG11773	GAACATCCTCTTTCTTAGTTTG
Commo n for pTherm oCas9_Δach_sp1-4	BG11778	CAAAAATCCCCCCTTTGATTGTT
	BG11779	aacaatcaaagggggattttgTAATTTTAAACTCCCGTTTCCAGACG
	BG11780	caaactaagaaagaggatgttcATCGTGCAGATTATGATTACTCTGTC

pTherm oCas9_Δach_sp1	BG11774	aacagcggtctcttcatccggcgGTATAACGGTATCCATTTTAAGAAT
pTherm oCas9_Δach_sp2	BG11782	ggatccgatccctgattgaatagGTATAACGGTATCCATTTTAAGAAT
pTherm oCas9_Δach_sp3	BG11784	atcggcattcctcttgatccggaGTATAACGGTATCCATTTTAAGAAT
pTherm oCas9_Δach_sp4	BG11786	atattctgaagttcttcaagacgGTATAACGGTATCCATTTTAAGAAT
Common for pThermoCas9_Δacs_sp1-4	BG11798	tagttcgaaatcgttgtttcattgtc
	BG11799	gacaatgaaacaacgatttcgaactaCATTTGACCGATTCCCCCTTTC
	BG11800	caaactaagaaagaggatgttcTGAAGTGCATCTCGAATTGAAGC
pTherm oCas9_Δacs_sp1	BG11796	aaagtcgatatgtgtctcatcttGTATAACGGTATCCATTTTAAGAAT
pTherm oCas9_Δacs_sp2	BG11802	tcccattaaccaaggtccgaataGTATAACGGTATCCATTTTAAGAAT
pTherm oCas9_Δacs_sp3	BG11804	tgggagcatccaacgtgattgtgGTATAACGGTATCCATTTTAAGAAT
pTherm oCas9_Δacs_sp4	BG11806	gatattcaagaagatgagacacaGTATAACGGTATCCATTTTAAGAAT

Supp. table 5. List of prim ers used in this study for gene deletion or substitution PCR checks.

Sequencing target	Primer No	Sequence
ΔgapA::gapC m utant confirm ation	BG10051	TTCACTAACGCCGTACCTG
	BG10052	CTTCACGATATTTTTTGCTTGTTTC
Δach m utant confirm ation	BG11834	GTATCCAGCATCCTTATGGC
	BG11835	TGCTACAGCAATGCTACTG
	BG12128	GTCAGCCAAACGTACGCTAG
	BG12129	CGGAGTAACAAAAGCGTATTTGGC
Δacs m utant confirm ation	BG11836	TCGTTAAACCATACGCTTTTGC
	BG11837	GTTTGCCGTCTTGGACAATG
	BG12130	GGCTTTTCCACGATTCATTTGTC
	BG12131	CCCGAGCGATGATAATTCTTCC

Supp. table 6. List of plasm id specific prim ers used in this study for plasm id curing PCR checks.

Sequencing target	Primer No	Sequence
Spcas9 gene	BG8473	ATCTCATTGCTTTGTCATTG
	BG8485	CCTTTATCGACAACTTCTTC
Thermocas9 gene	BG8196	ATGAAGTATAAAATCGGTCTTG
	BG8197	TAACGGACGGATAGTTTC

References

1. Bauer, F., et al., Technological innovation systems for biorefineries: a review of the literature. Biofuels, Bioproducts and Biorefining, 2017. 11(3): p. 534-548.
2. Sheldon, R.A., Utilisation of biomass for sustainable fuels and chemicals: Molecules, methods and metrics. Catalysis Today, 2011. 167(1): p. 3-13.
3. Mougiakos, I., et al., Next Generation Prokaryotic Engineering: The CRISPR-Cas Toolkit. Trends Biotechnol, 2016. 34(7): p. 575-87.
4. Mougiakos, I., et al., Hijacking CRISPR-Cas for high-throughput bacterial metabolic engineering: advances and prospects. Curr Opin Biotechnol, 2018. 50: p. 146-157.
5. F. Bosma, E., et al., Sustainable Production of Bio-Based Chemicals by Extremophiles. Current Biotechnology, 2013. 2(4): p. 360-379.
6. Mougiakos, I., et al., Characterizing a thermostable Cas9 for bacterial genome editing and silencing. Nature Communications, 2017. 8(1): p. 1647.
7. Bosma, E.F., et al., Isolation and Screening of Thermophilic Bacilli from Compost for Electrotransformation and Fermentation: Characterization of *Bacillus smithii* ET 138 as a New Biocatalyst. Applied and Environmental Microbiology, 2015. 81(5): p. 1874-1883.
8. Bosma, E.F., Isolation, characterization and engineering of *Bacillus smithii* : a novel thermophilic platform organism for green chemical production. 2015, Wageni ngen UniversitWageningen.
9. Bosma, E.F., et al., Establishment of markerless gene deletion tools in thermophilic *Bacillus smithii* and construction of multiple mutant strains. Microbial Cell Factories, 2015. 14(1): p. 99.
10. Bosma, E.F., et al., Complete genome sequence of thermophilic *Bacillus smithii* type strain DSM 4216T. Standards in Genomic Sciences, 2016. 11(1): p. 52.
11. Martínez, I., et al., Replacing *Escherichia coli* NAD-dependent glyceraldehyde 3-phosphate dehydrogenase (GAPDH) with a NADP-dependent enzyme from *Clostridium acetobutylicum* facilitates NADPH dependent pathways. Metabolic Engineering, 2008. 10(6): p. 352-359.
12. Mougiakos, I., et al., Efficient Genome Editing of a Facultative Thermophile Using Mesophilic spCas9. Acs Synthetic Biology, 2017. 6(5): p. 849-861.
13. Claassens, N.J., et al., Improving heterologous membrane protein production in *Escherichia coli* by combining transcriptional tuning and codon usage algorithms. PLOS ONE, 2017. 12(9): p. e0184355.
14. Schreiber, W. and P. Durre, The glyceraldehyde-3-phosphate dehydrogenase of *Clostridium acetobutylicum*: isolation and purification of the enzyme, and sequencing and localization of the gap gene within a cluster of other glycolytic genes. Microbiology, 1999. 145 (Pt 8): p. 1839-47.
15. Subedi, K.P., et al., Role of GldA in dihydroxyacetone and methylglyoxal metabolism of *Escherichia coli* K12. FEMS Microbiology Letters, 2008. 279(2): p. 180-187.
16. McCloskey, D., et al., Adaptation to the coupling of glycolysis to toxic methylglyoxal production in tpiA deletion strains of *Escherichia coli* requires synchronized and counterintuitive genetic changes. Metab Eng, 2018. 48: p. 82-93.
17. Jain, R. and Y. Yan, Dehydratase mediated 1-propanol production in metabolically engineered *Escherichia coli*. Microb Cell Fact, 2011. 10: p. 97.
18. Yomano, L.P., et al., Deletion of methylglyoxal synthase gene (mgsA) increased sugar co-metabolism in ethanol-producing *Escherichia coli*. Biotechnol Lett, 2009. 31(9): p. 1389-98.
19. James, K.L., et al., Pyrophosphate-Dependent ATP Formation from Acetyl Coenzyme A in Syntrophus aciditrophicus, a New Twist on ATP Formation. MBio, 2016. 7(4).
20. Lee, F.J., L.W. Lin, and J.A. Smith, Acetyl-CoA hydrolase involved in acetate utilization in *Saccharomyces cerevisiae*. Biochim Biophys Acta, 1996. 1297(1): p. 105-9.
21. Connerton, I.F., W. McCullough, and J.R. Fincham, An acetate-sensitive mutant of Neurospora crassa deficient in acetyl-CoA hydrolase. J Gen Microbiol, 1992. 138(9): p. 1797-800.
22. van den Berg, M.A., et al., The two acetyl-coenzyme A synthetases of *Saccharomyces cerevisiae* differ with respect to kinetic properties and transcriptional regulation. J Biol Chem, 1996. 271(46): p. 28953-9.
23. De Virgilio, C., et al., Cloning and disruption of a gene required for growth on acetate but not on ethanol: The acetyl ‐ coenzyme a synthetase gene of Saccharmoyces cerevisiae. Yeast, 1992. 8(12): p. 1043-1051.
24. Gardner, J.G., et al., Control of acetyl-coenzyme A synthetase (AcsA) activity by acetylation/deacetylation without NAD+ involvement in Bacillus subtilis. Journal of bacteriology, 2006. 188(15): p. 5460-5468.
25. Brown, T., M. Jones-Mortimer, and H. Kornberg, The enzymic interconversion of acetate and acetyl-coenzyme A in *Escherichia coli*. Microbiology, 1977. 102(2): p. 327-336.
26. Knorr, R., M.A. Ehrmann, and R.F. Vogel, Cloning, expression, and characterization of acetate kinase from Lactobacillus sanfranciscensis. Microbiological research, 2001. 156(3): p. 267-277.

27. Horswill, A.R. and J.C. Escalante-Semerena, The prpE gene of Salmonella typhimurium LT2 encodes propionyl-CoA synthetase. Microbiology, 1999. 145(6): p. 1381-1388.

28. Bobik, T.A., et al., The Propanediol Utilization (pdu) Operon of Salmonella enterica Serovar Typhimurium LT2 Includes Genes Necessary for Formation of Polyhedral Organelles Involved in Coenzyme B12-Dependent 1, 2-Propanediol Degradation. Journal of bacteriology, 1999. 181(19): p. 5967-5975.

29. Heßlinger, C., S.A. Fairhurst, and G. Sawers, Novel keto acid formate‐lyase and propionate kinase enzymes are components of an anaerobic pathway in *Escherichia coli* that degrades L‐threonine to propionate. Molecular microbiology, 1998. 27(2):p. 477-492.

30. Newton, G.L., et al., Bacillithiol is an antioxidant thiol produced in Bacilli. Nature chemical biology, 2009. 5(9): p. 625.

31. Gaballa, A., et al., Biosynthesis and functions of bacillithiol, a major low-molecular-weight thiol in Bacilli. Proceedings of the National Academy of Sciences, 2010. 107(14): p. 6482-6486.

32. Chandrangsu, P., et al., Methylglyoxal resistance in Bacillus subtilis: contributions of bacillithiol‐dependent and independent pathways. Molecular microbiology, 2014. 91(4): p. 706-715.

33. Chipman, D.M., R.G. Duggleby, and K. Tittmann, Mechanisms of acetohydroxyacid synthases. Current opinion in chemical biology, 2005. 9(5): p. 475-481.

34. Tse, J.M. and J.V. Schloss, The oxygenase reaction of acetolactate synthase. Biochemistry, 1993. 32(39): p. 10398-10403.

35. Helmann, J.D., Bacillithiol, a new player in bacterial redox homeostasis. Antioxidants & redox signaling, 2011. 15(1): p. 123-133.

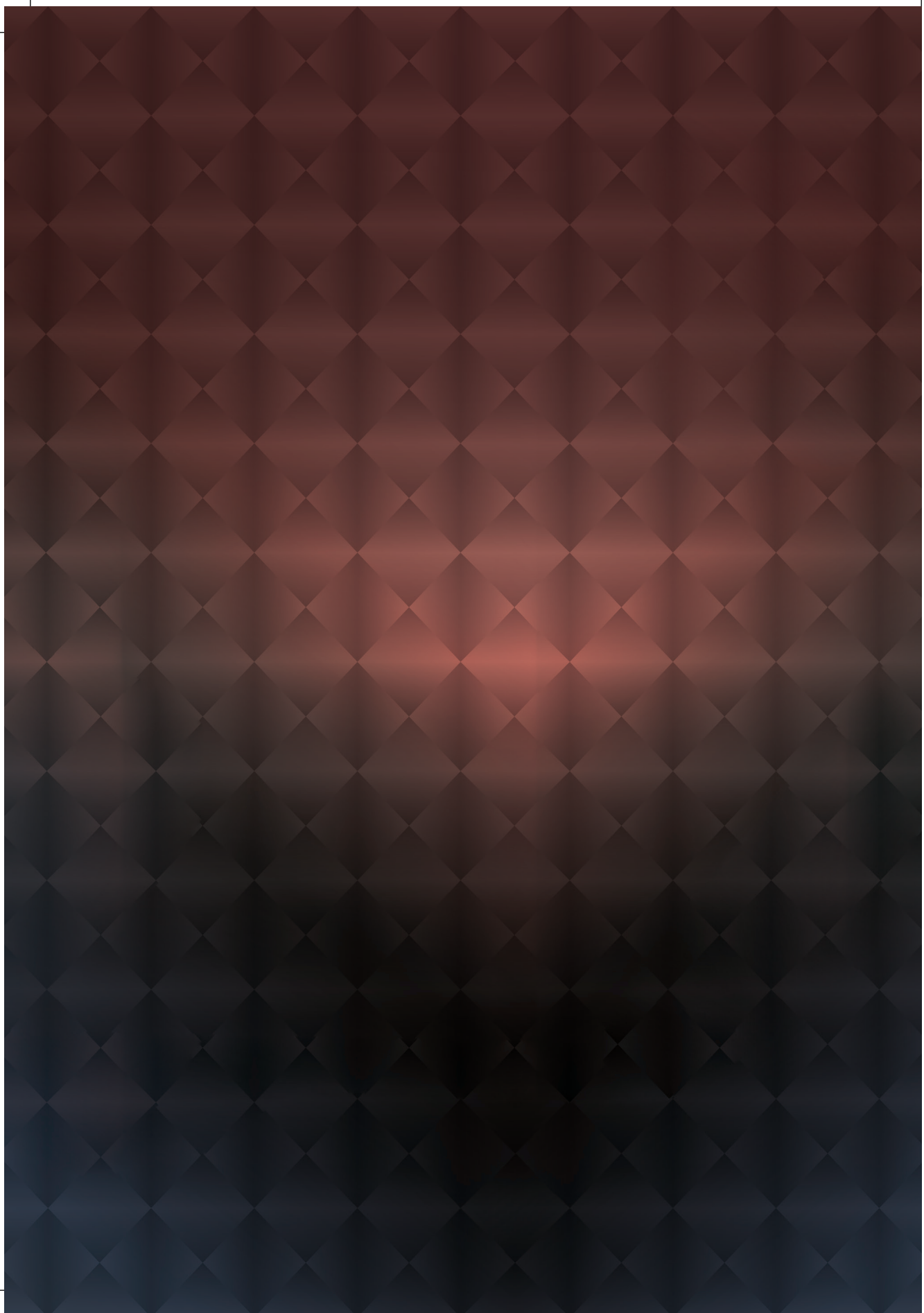
36. Bawn, C. and J. Williamson, The oxidation of acetaldehyde in solution. Part I.—The chemistry of the intermediate stages.Transactions of the Faraday Society, 1951. 47: p. 721-734.

37. Durand, A. and M. Merrick, *In vit*ro analysis of the Escherichia coli AmtB-GlnK complex reveals a stoichiometric interaction and sensitivity to ATP and 2-oxoglutarate. Journal of Biological Chemistry, 2006. 281(40): p. 29558-29567.

38. Dong, Z., et al., Tandem mass tag-based quantitative proteomics analyses reveal the response of Bacillus licheniformis to high growth temperatures. Annals of Microbiology, 2017. 67(7): p. 501-510.

39. Sánchez, A.M., G.N. Bennett, and K.-Y. San, Novel pathway engineering design of the anaerobic central metabolic pathway in Escherichia coli to increase succinate yield and productivity. Metabolic engineering, 2005. 7(3): p. 229-239.

40. Taylor, M., et al., Genetic tool development underpins recent advances in thermophilic whole‐cell biocatalysts. Microbial biotechnology, 2011. 4(4): p. 438-448.



Chapter 9

THESIS SUMMARY AND GENERAL DISCUSSION

THESIS SUMMARY

The global demand on chemicals and fuels is exponentially increasing. At the same time, the excessive exploitation of fossil-based resources for the coverage of this demand has a high environmental impact, motivating the production of green chemicals and biofuels from renewable resources. As described in **chapter 1**, the biotechnological potential of microbes for the production of food, beverages and –to a lesser extent- chemicals have been exploited throughout the human history. The last decades, the emerge of synthetic biology tools as well as of efficient, cost- and time-effective genome engineering tools, allowed the conduction of high-throughput microbial metabolic engineering studies that lead to the development of microbial platforms with high production capacities for green chemicals and biofuels. The genome engineering tools that brought a real revolution to the field of metabolic engineering are based predominantly on the RNA-guided DNA-endonuclease of the type II-A CRISPR-Cas adaptive immune system of *Streptococcus pyogenes*, namely SpCas9. This thesis mainly focused on the development of genetic tools and novel CRISPR-Cas9-based and antiCRISPR-based genome engineering tools for model and non-model mesophilic and thermophilic bacteria. Consequently, some of these tools were further employed for studying the metabolism and the biotechnological potential of the moderate thermophilic bacterium *Bacillus smithii* ET 138.

Chapter 2 reviews the first generation of CRISPR-Cas-based engineering tools, which is extensively used till date for genome editing as well as transcriptional silencing of both model and non-model prokaryotes. In this generation of editing tools, Cas9 endonucleases (more frequently the one originating from the type II-A CRISPR-Cas system of *S. pyogenes*, namely SpCas9) are combined either with plasmid-borne homologous recombination templates or recombineering systems (for the case of model organisms with already developed such systems), or more rarely with non-homologous end joining (NHEJ) systems, for gene deletions, gene insertions or nucleotide substitutions. Moreover, this chapter reviews the first generation of CRISPR-Cas transcriptional downregulation applications that employed either the catalytically inactive ("dead") SpdCas9 variant or a native type I CRISPR-Cas system lacking the Cas3 nuclease component. Additionally, it is underlined the importance of characterizing novel CRISPR-Cas systems, aiming to increase the engineering throughput and develop CRISPR-Cas applications for a wider variety of bacteria. Finally, a short overview of the *in silico*, *in vitro* and *in vivo* PAM identification approaches is provided.

Chapter 3 describes the development of the first SpCas9-based genome editing tool for a moderate thermophilic bacterium, namely *Bacillus smithii* ET 138. It is demonstrated that SpCas9 is *in vivo* inactive at temperatures above 42 °C and the wide growth temperature range of *B. smithii* ET 138 is employed for controllable induction of SpCas9 expression at temperatures below the 42°C threshold. Plasmid-borne editing templates are employed for homologous recombination (HR)-based introduction of the desired modifications to the genome of *B. smitiii* ET 138 at 45–55 °C. Transfer of the cultures to 37 °C induces the expression of active spCas9 from the same plasmid that carries the HR-templates. Subsequently, spCas9-based introduces lethal double-stranded DNA breaks to the non-edited cells, acting as a counter-selection mechanism. The developed method resulted in rapid and highly efficient genome editing applications on the *B. smithii* ET 138 genome, including the deletion of the pyrF (Orotidine 5'-phosphate decarboxylase) gene, knockout -via premature stop codons insertion- of the hsdR (Type I restriction enzyme R Protein) gene and insertion of the ldhL (L-lactate dehydrogenase) gene back in the genome of a *B. smithii* ET 138 Δ*ldhL* mutant strain.

The development of a spCas9-based genome editing tool for the biotechnologically relevant α-proteobacterium *Rhodobacter sphaeroides* is described in **chapter 4**. A highly efficient SpCas9-based targeting system is established for *R. sphaeroides* and it is subsequently combined with HR from plasmid-borne templates for gene deletions, insertions, as well as for single nucleotide substitution with high efficiencies. Additionally, this study further investigates the unique NHEJ mechanism of *R. sphaeroides*, that is comprised of one LigD and two Ku homologs. The SpCas9-based targeting system is combined with the native *R. sphaeroides* NHEJ system resulting in indel formations with efficiencies of up to 20%. Moreover, it is demonstrated the activity of the *R. sphaeroides* NHEJ mechanism is based on the LigD and not on the predicted Ku homologs, strongly suggesting the existence of an unprecedented bacterial NHEJ system.

The first aim of **chapter 5** is to describe the characterization of the Cas9 orthologue from the type IIc CRISPR-Cas system of the thermophilic bacterium *Geobacillus thermodenitrificans* T12 ThermoCas9. It is demonstrated that ThermoCas9 is *in vitro* active between 20-70 °C, has strict PAM-preference at the lower temperatures range, does not tolerate extensive spacer-protospacer mismatches, the structure of its sgRNA influences its activity at elevated temperatures and it cleaves more efficiently supercoiled-plasmid DNA compared to linear DNA. The *in vitro* characterization data are further employed for the development of Thermo

Cas9-based engineering tools for gene deletion (*pyrF*) and transcriptional silencing (*ldhL*) at 55 °C in *B. smithii* ET 138 and for gene deletion (*pyrF*) at 37 °C in *Pseudomonas putida*. This is one of the first reports that provides fundamental insights into a thermophilic CRISPR-Cas family member and the first report that establishes Cas9-based bacterial genome engineering at temperatures above 42°C.

The potential to harness the activity of an antiCRISPR proteins for genome engineering purposes is examined in **Chapter 6** of this thesis. The antiCRISPR protein from *Neisseria meningitidis* ($AcrIIC1_{Nme}$) is employed as "on/off-switch" of the thermo-active Cas9 orthologues from *Geobacillus thermodenitrificans* T12 (ThermoCas9) and *Geobacillus stearothermophilus* (GeoCas9). Initially it is demonstrated that both ThermoCas9 and GeoCas9 can introduce lethal dsDNA breaks in *E. coli* at 37°C in a tunable and spacer-dependent manner. Consequently it is 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" Thermo-dCas9 and Geo-dCas9 variants. Finally, this chapter describes the construction of the first reported single-vector, tightly controllable and highly efficient Cas9/$AcrIIC1_{Nme}$-based tool for coupled silencing and targeting in bacteria that could be further employed for genome editing and transcriptional regulation of academically and industrially relevant bacteria.

The number of metabolic engineering studies on model and non-model prokaryotes, towards the production of green chemicals and biofuels, that make use of CRISPR-Cas-based engineering technologies are exponentially increasing the last years. **Chapter 7** reviews these studies and justifies that the characterisation of novel CRISPR-Cas systems from diverse environments is crucial for the extension of CRISPR-Cas-based engineering applications to previously unexploited non-model organisms that inhabit harsh environments (extreme temperatures, pH, salt concentrations etc.) but have great biotechnological potential.

The work performed during this thesis on studying the metabolism of *B. smithii* ET 138, as well as on enhancing its production capacity in dicarboxylic acids, is presented in **Chapter 8**. The SpCas9- based genome editing tool, which was described in chapter 3, was employed for the substitution of the native NAD^+-dependent glyceraldehyde-3P dehydrogenase (*gapA*) gene by the NADP+-dependent *gapC* gene from *Clostridium acetobutylicum* ATCC 824. The substitution was performed in the genome of the *B. smithii* ET 138 Δ*ldhL* strain, it aimed to

relieve its NAD^+ deficiency under oxygen-limited culturing conditions, and it resulted in moderate growth enhancement. The ThermoCas9-based genome editing tool was employed for the deletion of the acetyl-CoA synthetase (*acs*) and acetyl-CoA hydrolase (*ach*) genes, aiming to reveal the non-canonical acetate production pathway of *B. smithii* ET 138. Nonetheless, none of these deletions affected the acetate production. Moreover, B. *smithii* ET 138 was cultured under oxygen rich conditions, at a wide temperatures (between 37°C and 55°C), aiming to investigate the influence of the growth temperature on its production profile. Strikingly, acetate production was completely abolished when culturing at 37°C, while it was maximum at 55°C. Comparative proteomics between cultures grown at these two temperatures did not provide clear information about the acetate production pathway. Finally, a *B. smithii* ET 138 strain, transformed with a plasmid overexpressing the native pyruvate carboxylase (*pyc*) and the malate dehydrogenase (*mdh*) genes, was cultured at 50°C and produced the highest reported amounts of dicarboxylic acids while keeping low the acetate production levels.

General Discussion

The exploitation of the biotechnological potential of non-model microorganisms could accelerate the transition to a green economy. This potential resides in the natural ability of microbes to produce valuable molecules[1,2], to tolerate harsh production conditions[3], to communicate with other microbes[4] and/or consume specific carbon sources[5]. The extent of metabolic engineering experiments on non-model microorganisms is restricted by their underdeveloped molecular/synthetic biology and genome engineering toolboxes, especially when compared to model microorganisms. The predominant bottlenecks that hamper metabolic engineering of non-model organisms are:

1. the low genetic accessibility (transformation efficiency) of the vast majority of the microorganisms
2. the lack of extensive genetic toolboxes*
3. the inefficient and time-consuming genome editing and transcriptional silencing tools*
4. the limited understanding of their metabolism

*Note that even though the genetic engineering tools are, by definition, included in the genetic toolbox of an organism, here it was decided that the independent analysis of the engineering tools would facilitate the flow of this discussion.

In the course of this thesis, this set of bottlenecks has been evaluated and addressed in experiments aiming to overcome them and to metabolically engineer *B. smithii* ET 138. This last chapter aims to further reflect on these bottlenecks and the results obtained regarding them, as well as to discuss several additional experiments and place the data in a broader context.

1) GENETIC ACCESSIBILITY

The first step in bacterial genome engineering processes is to evaluate its genetic accessibility by establishing a protocol for efficient introduction of exogenous DNA (generally in the form of a plasmid). This is commonly achieved either via transformation or via conjugation. During transformation, DNA is transferred from the extracellular space through the cell envelop either through induced disruption by electrical or chemical means, or via the native DNA-uptake systems (natural competence)[6]. Nonetheless, the process to establish a reliable transformation protocol for a microorganism is strain-specific, labor-intensive and commonly based on a trial-and-error approach[7,8]. Moreover, the existence of efficient restriction and modification (R/M) systems in the majority of the microorganisms may severely reduce transformation efficiencies, as also observed in Chapter 3 of this thesis. As for transformation approaches that are based on natural competence, these are restricted to the low number of bacterial strains de-

monstrated to possess an active DNA-uptake machinery[9,10]. During conjugation, DNA (usually in the form of a conjugative plasmid) is transferred from a "donor" to a "recipient" cells via pili-mediated direct contact[11]. This process is commonly employed by bacteria in nature for genetic information exchange, resulting in the spread of features, like stress tolerance[12], antibiotic resistance[13,14], proteolytic activity and carbohydrate utilization systems[15], between bacterial strains. Moreover, all conjugative plasmids code for the enzymatic machinery responsible for the conjugation and some encode anti-restriction proteins to circumvent the cellular R/M systems[16]. Conjugative plasmids have been extensively used for heterologous expression and genome engineering purposes, including *Rhodobacter sphaeroides* as described in Chapter 4 of this thesis. Nonetheless, conjugation is a process restricted to conditions compatible with the donor strain, and host factors influencing conjugation are still poorly understood[17].

The microorganisms used during the course of this thesis were genetically accessible via previously established protocols; *Pseudomonas putida* KT2440[18] and *Bacillus smithii* ET 138[7] are electro-transformable strains, while *Rhodobacter sphaeroides* ATC 35053[19] is genetically accessible via conjugation. In this thesis, the work to improve transformation efficiencies was limited to *B. smithii* ET 138. In chapter 3 the *hsdR* gene, which codes for the only restriction enzyme expressed in *B. smithii* ET 138, was deleted from the genome of the sporulation-deficient *B. smithii* ET 138 Δ*sigF* strain. The resulting *B. smithii* ET 138 Δ*sigF* Δ*hsdR* strain was extensively used in this thesis and even though the *hsdR* deletion did not improve the previously reported maximum transformation efficiency (in CFU/µg of transforming DNA) for a *B. smithii* ET 138 strain, the *B. smithii* ET 138 Δ*sigF* Δ*hsdR* strain was efficiently transformable with any pNW33n-based and pG2K-based construct of size up to -at least- 12.5 kb, without sequence limitations that were an obstacle prior to removal of the *hsdR*. This result facilitated the rapid and efficient realization of genome engineering experiment, which was not possible for the wild type *B. smithii* ET 138.

2) GENETIC TOOLBOXES

Since 1973, when the first transformation of *E. coli* with a plasmid vector took place[20], a wide variety of genetic tools has been developed for bacterial engineering, aiming to facilitate both fundamental research and metabolic engineering for strain development purposes. Amongst these tools belong: i) replicating or integrative vectors, employed for single genes and/or pathway expression and genome engineering purposes, ii) constitutive or inducible promoters, employed for the continuous or controlled transcription of associated genes,

iii) riboswitches, which upon binding to a suitable ligand molecule act as on/ off transcriptional/ translational switches or result in mRNA maturation, iv) ribosome binding sites (RBS), responsible for the translational efficiency of the associated genes, v) transcriptional terminators, responsible for the dislodging of the RNA polymerase from the mRNA transcript, and vi) reporter genes, such as genes that code for either antibiotic resistance (selection marker) and fluorescence proteins (phenotypic marker); the latter allow for measuring the activity of the previously mentioned regulatory elements. The modularity of the described genetic tools/parts set the basis for the development of the synthetic biology field that focuses on the building of artificial biological systems for research and engineering purposes. Part of the work performed during this thesis aiming to transform *B. smithii* ET 138 into a synthetic biology chassis, by extending its genetic toolbox with vectors and promoters, is described in chapters 2 and 3. Further details will be discussed below.

VECTORS

Prior to this thesis, pNW33n was the only available *B. smithii* ET 138 vector for genome engineering and protein expression purposes[21]. pNW33n is an *E. coli–Bacillus/Geobacillus* sp. shuttle vector, comprised of the *repB* gene, which codes for a rolling circle replication initiator protein, the ColE1 *E. coli* origin of replication and the *cat* gene, which codes for a thermostable chloramphenicol acetyltransferase confering resistance to chloramphenicol. The further expansion of the *B. smithii* ET 138 toolbox with additional vectors was assigned as a high priority task.

During this thesis, the transformation efficiency of *B. smithii* ET 138 with the recently developed *E. coli – Bacillus/Geobacillus* sp. shuttle vectors pG1C, pG1K and pG2K was evaluated[22]. The three tested vectors were selected due to their proven transformability into the Gram-positive thermophile *Parageobacillus thermoglucosidasius*[22]. All three vectors contained the ColE1 origin of replication. Both pG1C and pG1K code for the RepBST1 replication protein, while pG2K codes for the RepB replication protein (the same as pNW33n). Moreover, the pG1K and pG2K code for a thermostable kanamycin nucleotidyltransferase, and pG1C codes for a thermostable chloramphenicol acetyltransferase (the same as pNW33n). The restriction- and sporulation-deficient *B. smithii* ET 138 Δ*sigF* Δ*hsdR* strain was employed for the transformation experiments and only pG2K was transformable into this strain. The transformation efficiency achieved with the pG2K vector was comparable to the transformation efficiency achieved with pNW33n (10^5 CFU/ μg of plasmid DNA), as was expected since the two vectors share the same replication protein and have comparable sizes. Moreover, the co-transformation of *B. smithii* ET 138 Δ*hsdR* with the

pNW33n and pG2K was possible, with reduced efficiency (10^1 CFU/ μg of plasmid DNA). The inefficient expression of the RepBST1 protein by the pG1C and pG1K vectors was the most probable reason for the unsuccessful transformation of *B. smithii* ET 138 Δ*hsdR* with these vectors. Hence, testing alternative promoters and RBSs for the RepBST1 expression may further increase the number of vectors in the genetic toolbox of *B. smithii* ET 138.

A different approach for the introduction of additional plasmids into the genetic toolbox of an organism is the use of the origins of replication (ori) and of the replication proteins genes (*rep*) from the native plasmids of closely related species. In the case of *B. smithii* ET 138, the ori and rep from native plasmids of *Bacillus coagulans*, *Bacillus licheniformis* or other close relatives could be evaluated as backbones for the construction, and subsequent test transformation to *B. smithii* ET 138, of a "new" plasmid with the addition of the previously mentioned thermostable chloramphenicol acetyltransferase or kanamycin nucleotidyltransferase genes.

PROMOTERS

The development of inducible or repressible promoters, as well as of promoter libraries that cover a wide range of transcriptional strengths is of great importance for an efficient genetic toolbox of an organism. Aiming to identify xylose- or glucose-inducible promoters present in the *B. smithii* ET 138 genome, transcriptomics analysis was performed on total RNA samples isolated from *B. smithii* ET 138 cultures grown in LB2 rich medium up to the early exponential phase (3 hours at 55°C) in the absence or presence of a carbon source (glucose or xylose). The comparative data analysis led to the selection of the bidirectional P_{xylA} promoter for experimental evaluation of its inducibility.

The P_{xylA} promoter controls the transcription of the xylose isomerase (xylA) and transcriptional repressor (xylR) genes and the acquired transcriptomics data revealed that it is transcriptionally active only in the presence of xylose. Previous studies on other Gram-positive bacteria[23,24] showed that there are two regulation systems that control the transcription of P_{xylA} (Fig. 1a). The first system is based on the XylR transcriptional regulator; in the absence of xylose, XylR binds to the P_{xylA} promoter and hinders the binding of the RNA polymerase, while in the presence of xylose the XylR is subjected to a conformational change and abolishes the ability to bind the P_{xylA}, hence allowing for the xylA transcription and the further catabolism of xylose. The second regulation system is based on the presence of a catabolite-responsive element (CRE-box) in the P_{xylA}; the catabolic control protein A (CcpA) binds to the CRE-box when glucose is present, blocking transcription from PxylA in both directions. Based on previous reports for general CRE-box

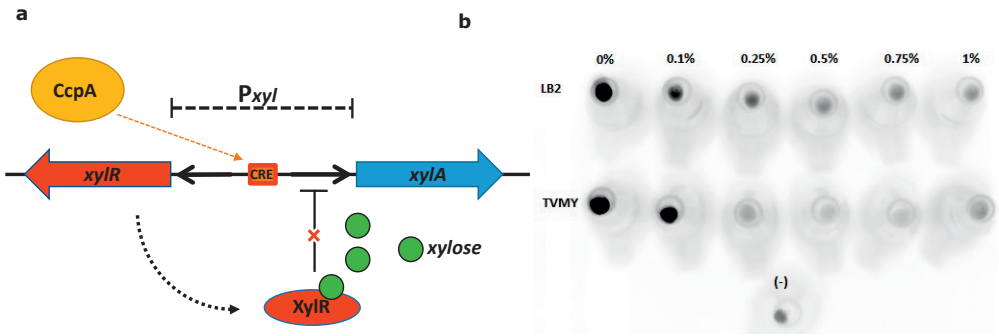


Figure 1a. Schematic representation of the P_{xylA} promoter, as located within the *B. smithii* ET 138 genome, and of the 2 systems responsible for the control of its transcriptional repression. **b.** B. UV exposed samples of *B. smithii* ET 138:: pNW_sfGFP cultures grown overnight in LB2 rich or TVMY poor media supplemented with different concentrations of glucose for repression of the P_{xylA} promoter that controls the transcription of the sfgfp gene.

sequences[25,26] an almost palindromic, CRE-box like sequence was identified 34bp upstream of the *xylA* gene within the P_{xylA} region. The construction of a plasmid-borne xylose inducible system for transcriptional regulation in *B. smithii* ET 138 would require the cloning of the P_{xylA} promoter in the pNW33n backbone together with the *xylR* gene, both right upstream of the gene-of-interest. Since the CRE-box based mode of regulation does not require the presence of XylR, it was decided to evaluate the efficiency of a plasmid-borne glucose repressible system for transcriptional regulation of a gene-of-interest. For this purpose, the thermostable *sfgfp* gene[27] was employed as the reporter gene and cloned in the pNW33n backbone under the transcriptional control of P_{xylA}, resulting in the pNW_sfGFP plasmid. *B. smithii* ET 138:: pNW_sfGFP precultures grown overnight in LB2 rich or TVMY poor media (10 ml in 50 ml falcon tubes, 55°C, 150 rpm) were employed for inoculation of LB2 and TVMY cultures respectively, supplemented with 0 to 1%(w/v) glucose (10 ml in 50 ml falcon tubes, 55°C, 150 rpm). Regardless the employed glucose concentration, the optical density (OD600) of the cultures from each medium was comparable after overnight growth. Moreover, analysis of culture samples (with cell densities normalized based on the OD600 measurements) under UV light showed that glucose concentrations of 0.25% and above result in almost complete elimination of fluorescence (Fig. 1b).

The P_{xylA} was employed in Chapters 4 and 8 of this thesis for tuning down the transcription of the *thermocas9* and *thermodcas9* genes. Nonetheless, the use of glucose during the ThermoCas9-based editing and silencing experiments did not affect the efficiency

of the tools, indicating that the P_{xylA} promoter is not tightly repressible. An attempt to obtain a more tightly controlled, xylose inducible and glucose repressible transcription system, could be to clone the *xylR* gene upstream of the P_{xylA} / gene-of-interest. Along the same line with the previously presented strategy, transcriptomics analysis on cultures grown under different environmental stimuli, such as higher or lower nitrogen concentrations, in the presence of high metal concentrations etc., could reveal other inducible or repressible promoters, as well as promoters of different strengths, present in the *B. smithii* ET 138 genome. These promoters could be further tested with reporter genes and added to the *B. smithii* ET 138 genetic toolbox. Additionally, a well-characterized *B. smithii* ET 138 promoter could be used for the construction of a promoter library, by randomizing the sequence between the -10, -35 promoter regions. This approach would eliminate the need for extensive characterization of many different promoters and at the same time could result in a number of closely related promoters with a wide range of transcriptional strengths.

3) GENOME ENGINEERING TOOLS

For many model and most non-model prokaryotes, the execution or throughput of metabolic engineering and systems biology experiments is hindered by the lack of efficient genome engineering tools. From 2013 onwards it has been demonstrated that for a wide variety of prokaryotes the activity of the Cas9 endonuclease from the type II−A CRISPR−Cas system of *Streptococcus pyogenes* (SpCas9) can be combined with homologous recombination and recombineering (recombination-mediated genome engineering) approaches for the development of easy to design, efficient and precise genome editing tools, as was reviewed in chapters 2 and 7 of this thesis. Moreover, it was recently demonstrated that the editing efficiencies of these tools can significantly and unpredictably vary amongst the same genomic targets[28], justifying the usefulness of developing different SpCas9-based editing tools. Additionally, many single effectors from other type II CRISPR-Cas systems, usually of comparable or smaller size than SpCas9, have been characterized[29-33] and tested in genome engineering applications[29,31,32] aiming to achieve higher efficiencies, to avoid encountered SpCas9 toxicity issues, to increase targeting specificities and to extend the number of available targets in a genome (via recognition of alternative protospacer-adjacent motifs (PAMs)).

The characterization of ThermoCas9, a thermotolerant Cas9 homologue that is smaller and has higher *in vitro* targeting specificity than SpCas9, has resulted in the develop-

ment of ThermoCas9-based engineering tools for thermophilic and mesophilic bacteria, as described in chapters 4 and 6 of this thesis. Additionally, the potential of two more CRISPR-Cas systems, or the effector components thereof, to efficiently target and engineer the genome of *B. smithii* ET 138 was examined. Further details on these works are discussed below.

CHARACTERIZING THE *B. SMITHII* ET 138 TYPE I-B CRISPR-CAS SYSTEM

There are a few examples of bacterial genome engineering tools designed to take advantage of native CRISPR-Cas systems[34-36], reducing the size of the employed plasmids and the requirements in genetic parts. *B. smithii* ET 138 contains one type I-B CRISPR-Cas adaptive immune system in its genome that lacks the *cas6* gene and which is followed by a CRISPR array of 32 spacers and a truncated type I-A CRISPR-Cas system that encompasses a *cas6* gene (Fig. 2). Transcriptomics data extracted from *B. smithii* ET 138 cultures grown under different aeration conditions, in different media and with different carbon sources suggested that the genes of the CRISPR-Cas I-B system are actively transcribed, especially at the late exponential phase of growth (data not shown). This finding motivated the further evaluation of the *in vivo* DNA targeting efficiency of the system, which would also indicate the use of the Cas6 component from the neighboring type I-A system. *In silico* mining for the preferred PAM sequences of the system, using a previously established approach reviewed in chapter 2[37,38], did not result in a convincing prediction. Hence, it was decided to proceed with an *in vivo* PAM determination approach.

Figure 2. Schematic representation of the type I-A and type I-B CRISPR/Cas systems within the *B. smithii* ET 138 genome. The two systems are interspaced by a single CRISPR array and each repeat of the array is depicted as a purple rectangular.

For the *in vivo* PAM determination, four spacers from the native *B. smithii* ET 138 CRISPR array (spacers 2, 12, 22 and 30 starting from the predicted leader sequence) were selected and cloned into the pNW33n vector in order to serve as protospacers in the transformation assays. Each protospacer was flanked at its 5′-end by a randomized 3-bp long sequence, which served as the PAM library. Wild-type *B. smithii* ET 138 cells were transformed with each plasmid library. Notably, the transformation efficiency of the cells with the library containing spacer 12 was three orders of magnitude lower com-

pared to the pNW33n control. The methylation analysis of the *B. smithii* ET 138 genome revealed the m6A modification of the 5'-CAGNNNNNNTGT-3' motif throughout the genome, and this motif was also present in the tested spacer 12. It was hypothesized that the described modification is most probably related to the R/M system of *B. smithii* ET 138 discussed above. Transformation of the spacer 12 plasmid library to sporulation- and restriction- deficient *B. smithii* ET 138 Δ*sigF* Δ*hsdR* strain reconstituted the transformation efficiency to control levels, confirming the hypothesis. Hence the observed drop in the transformation efficiency could not be attributed to targeting by the native CRISPR-Cas systems.

The transformation efficiency of wild type *B. smithii* ET 138 cells with the libraries containing spacers 2, 22 and 30 was comparable to the efficiency upon transformation with the pNW33n control. Plasmid material was isolated from all the surviving colonies and was analyzed by deep sequencing together with samples from the same plasmid libraries prior transformation to *B. smithii* ET 138. The comparison of the sequencing results showed no statistically relevant differences between the members of the plasmid libraries pre- and post-transformation. These results indicated that either the type I-B CRISPR-Cas system of *B. smithii* ET 138 is not efficiently expressed, under the tested conditions, or that the activity of the missing Cas6 of the system is not complemented by the Cas6 of the neighboring type I-A system. The existence of an antiCRISPR system that counteracts the activity of the CRISPR-Cas system is also a possibility that requires further *in silico* and *in vivo* studies.

TESTING AACAS12B AS A COUNTER SELECTION TOOL IN *B. SMITHII* ET 138

Chapters 2 and 3 describe the SpCas9- and ThermoCas9-based tools developed during this PhD project. One of the main objectives of this project was to further expand the CRISPR-Cas-based genome engineering toolbox for thermophilic bacteria. Previous studies have demonstrated that the Cas12b endonuclease from the type V−B CRISPR−Cas system of the thermophilic bacterium *Alicyclobacillus acidoterrestris* (AaCas12b) is active *in vitro* at 37°C and 50°C, as well as *in vivo* at 37°C in *E. coli*[33]. AaCas12b requires a crRNA-tracrRNA duplex for targeting, while the spacer sequence resides at the 3'-end of the crRNA and the PAM sequence (5'-TTN-3') that it recognizes resides at the 5'-end of the protospacer and is quite distinct to the PAM sequences of the currently characterized thermoactive Cas9 proteins. Moreover, it was shown that single nucleotide mismatches between the spacer of the employed sgRNA and the corresponding protospacer target hindered the AaCas12b *in vitro*

activity, suggesting the potential of AaCas12b for reduced *in vivo* off−target cleavage events. These facts motivated the development of a AaCas12b-based genome engineering toolbox for thermophilic bacteria.

The first step in this toolbox development process was the evaluation of the *in vivo* targeting activity of AaCas12b at 50°C (the optimal growth temperature of *A. acidoterrestris*) in *B. smithii* ET 138. As the codon usage landscapes of *B. smithii* ET 138 and *A. acidoterrestris* are comparable, the native *aacas12b* gene was cloned into the pNW33n vector under the transcriptional control of the P_{xylA} promoter, which was previously employed for the thermocas9 transcription. The native *A. acidoterrestris* tracrRNA transcriptional module as well as the corresponding CRISPR array were also cloned into the same plasmid. Four spacers (three targeting and one non-targeting control) were each individually introduced into the array creating a library of four plasmids with a single-spacer CRISPR array (leader-repeat-spacer-repeat) each. The plasmids were transformed to *B. smithii* ET 138 cells and the number of obtained CFUs upon transformation with the targeting constructs was compared to the number of obtained CFUs upon transformation with the control construct. The experiments were performed in biological duplicates and the transcription of the tracrRNA module and the *aacas12b* gene was confirmed via specific PCR amplifications using as template the constructed cDNA library from *B. smithii* ET 138 cultures transformed with the control plasmid. The transformation efficiency of *B. smithii* ET 138 was reduced by 1 order of magnitude only for one the targeting constructs, when compared to the control transformation efficiency. In attempt to improve the AaCas12b targeting efficiency, a similar plasmid-based *in vivo* interference assay was performed employing the same spacers incorporated at the 3'-end of sgRNA expressing modules, instead of crRNA:tracrRNA duplexes, as previously described in literature for AaCas12b *in vitro* targeting assays. However, this approach reduced the AaCas12b targeting efficiency even further.

The optimal AaCas12b targeting efficiency achieved during this experiment was comparable to the previously described results for *in vivo* AaCas12b−mediated targeting in *E. coli* at 37°C. The AaCas12b targeting efficiency might be improved by the use of different promoters with characterized transcriptional start sites, which could be used for and enhance the crRNA or sgRNA transcription. Nonetheless, the AaCas12b targeting efficiency in the currently designed system was much lower than the ThermoCas9 targeting efficiency, and way below the required efficiency levels for the development of an efficient tool.

THERMOCAS9 TARGETING EFFICIENCY IN *G. THERMODENITRIFICANS* T12

There is only a limited number of studies that make use of endogenous type II CRIPSR-Cas systems for bacterial genome editing purposes[39]. The *in vitro* characterization of the ThermoCas9 endonuclease from the type II−C CRISPR−Cas system of *Geobacillus thermodenitrificans* T12 strain was presented in chapter 3 of this thesis. The native *in vivo* efficiency of this system was further investigated, aiming to assess if it can be employed for the development of a "self-editing" system in *G. thermodenitrificans* T12. A minimal CRISPR array composed of the native leader and a repeat−spacer−repeat (RSR) sequence was cloned into the pNW33n *E. coli-Bacilli/Geobacilli* shuttle vector. The non-targeting control spacer was further substituted by spacers targeting different regions of the *ldhL* gene within the genome of *G. thermodenitrificans* T12, creating a library of 4 constructs (3 targeting and 1 non-targeting control). The constructs were separately transformed into *G. thermodenitrificans* T12 competent cells at 55°C, but no significant difference was observed in the transformation efficiencies between the cells transformed with the targeting and the non−targeting constructs. This result could be attributed to the weak transcription levels of the thermocas9 gene that were detected by RT-PCR, whereas there was no detectable transcription for the predicted tracrRNA module.

A new plasmid library was constructed aiming to circumvent the null/weak/conditional transcription of the native tracrRNA module. The previously employed spacers were separately cloned at the 5'-end of the ThermoCas9 sgRNA expression module in the pNW33n vector. Each sgRNA module was placed under the transcriptional control of four different promoters, previously shown to be functional in *G. thermodenitrificans* T12[40]; the P_{pta} from *B. coagulans*, the P_{rplS} from *P. thermoglucosidasius*, the native P_{uppT12} and the P_{pta} from P. *thermoglucosidasius*. Again, there was no significant difference observed in the transformation efficiencies between the cells transformed with the targeting and the non−targeting sgRNA expressing constructs, regardless the employed promoter. The described results indicate that either the native ThermoCas9 expression is not high enough to facilitate efficient targeting, or that there is an antiCRIPSR system within the genome of *G. thermodenitrificans* T12, which counteracts the activity of the native type II-C CRISPR-Cas system. Further *in silico* and *in vivo* studies are required in order to elucidate the existence of such an antiCRISPR system.

DEVELOPING THE 2ND GENERATION OF CRISPR-CAS BASED ENGINEERING TOOLS: PERSPECTIVES

Undoubtedly, the currently developed and employed CRISPR-Cas-based tools for bacterial genome engineering, predominantly of model bacteria such as *E. coli*, have unprecedentedly boosted the field of metabolic engineering both for fundamental and systems biology studies as well as for the production of green chemicals and biofuels. As reviewed in chapters 2 and 7, the plasmid-borne homologous recombination-Cas9 based counter-selection tool and the transcriptional silencing tool based on the catalytically inactivate dCas9 are the basic CRISPR-Cas applications for bacterial genome engineering and have been successfully adapted for few non-model bacteria. In model bacteria, such as *E. coli*[41], *P. putida*[42] and *Corynebacterium glutamicum*[43], the activity of CRISPR-Cas effectors has been combined with recombineering systems, which are phage derived enzymatic systems employed for the introduction of linear (ss- or ds-) DNA fragments into a genome of interest[44]. These fragments carry the desired mutations and the use of the CRISPR-Cas effectors as inducers of the recombination efficiencies and/or as counter-selection systems against the unedited genomes has substantially facilitated the selection of the mutated strains. Additionally, the CRISPR-Cas effectors have been combined with native or heterologous non-homologous end joining systems[45-48], as also presented in chapter 5, facilitating the construction of strain libraries with random indel mutations at the Cas-targeted sequences. Due to the simple designing basis of these combined systems, their adaptation for use in non-model bacteria can substantially enhance the throughput of their metabolic studies and possibly allow for multiplex genome editing.

The on-going development of a new generation of bacterial CRISPR-Cas based tools, such as base editors and EvolvR, as well as the exploitation of antiCRISPR systems, as proposed in chapter 6, promises to enhance engineering efficiencies and throughput, accelerating metabolic engineering studies for both model and non-model bacteria. As base editors are termed the fusions of nicking or catalytically inactive CRISPR-Cas effectors to DNA-deaminase enzymes, mainly cytidine deaminases (convert cytidine to thymidine) and adenosine deaminases (convert adenine to guanine), that introduce specific mutations within a certain window of the targeted protospacer[49-51]. Additionally, the fusion of the nicking Cas9 variant (nCas9) with a nick-translating, error-prone DNA polymerase resulted in the development of the EvolvR system that can diversify any nucleotide within an editing window of 350 nucleotides at millions-fold greater rates than the wild type mutation rates[52].

The use of these tools in bacteria allows either the introduction of predictable mutations, usually converting the codon of an amino-acid into a stop codon, or the construction of extensive mutant libraries, eliminating the need for homologous recombination templates carrying the desired mutations.

The CRISPR-Cas based revolution on bacterial genome engineering is only in its dawn. The possibilities that open up with the use of CRISPR-Cas RNA-guided DNA-endonucleases are unlimited and it is generally accepted that the development of new advanced genome engineering applications and synthetic biology tools will only accelerate. Metabolic engineering studies on model and non-model bacteria will benefit from this revolution and the transition to a green economy will hopefully soon turn from a dream into reality.

4) UNDERSTANDING *B. SMITHII* ET 138 METABOLISM FOR ENGINEERING PURPOSES

Chapter 7 of this thesis reviewed many important metabolic engineering studies on model and non-model prokaryotes that were engineered by CRISPR-Cas based genome engineering tools towards the production of green chemicals and biofuels. Chapter 8 of this thesis described the metabolic engineering steps employed to transform *B. smithii* ET 138 into a production platform of dicarboxylic acids, using the novel tools developed in this thesis. Next to genetic tool development as described above, a major bottleneck in using non-model organisms is the lack of understanding on their metabolisms, which usually present significant differences compared to the metabolisms of well-studied model organisms such as *E. coli* and *B. subtilis*. The limited knowledge and understanding on the metabolism of *B. smithii* ET 138 became apparent in the work presented in Chapter 8, as well as in several additional experiments that will be described and placed into further perspective in the final part of this chapter.

THE MYSTERY OF THE *B. SMITHII* ET 138 ACETATE PRODUCTION PATHWAY

As discussed in chapter 8, *B. smithii* ET 138 does not contain the genes of a canonical acetate production pathway, even though acetate is the main product under oxygen-rich culturing conditions. The work already performed and discussed in this thesis attempting to reveal the *B. smithii* ET 138 acetate production pathway includes the construction of acetyl-CoA synthase and aldehyde dehydrogenase deletion mutants. The mutants were cultured under acetate- producing conditions and the results showed that

none of these genes is crucial for the acetate production. Moreover, comparative proteomics analysis data presented in chapter 8 revealed the overproduction of the aminomethyltransferase protein (the "T-protein" of the glycine cleavage system) in the 55°C culture, which can be triggered only under high glycine concentration conditions[53]. This is an interesting observation, as the proteomics data also suggest nitrogen-deprivation for the 55°C culture, indicated by the overproduction of the nitrogen regulatory protein PII54. Thus, we expect that glycine is abundant in the 55°C acetaldehyde dehydrogenase. However, the experimental verification of the hypothesis is hindered by the lack of predicted threonine aldolase genes in the *B. smithii* ET 138 genome.

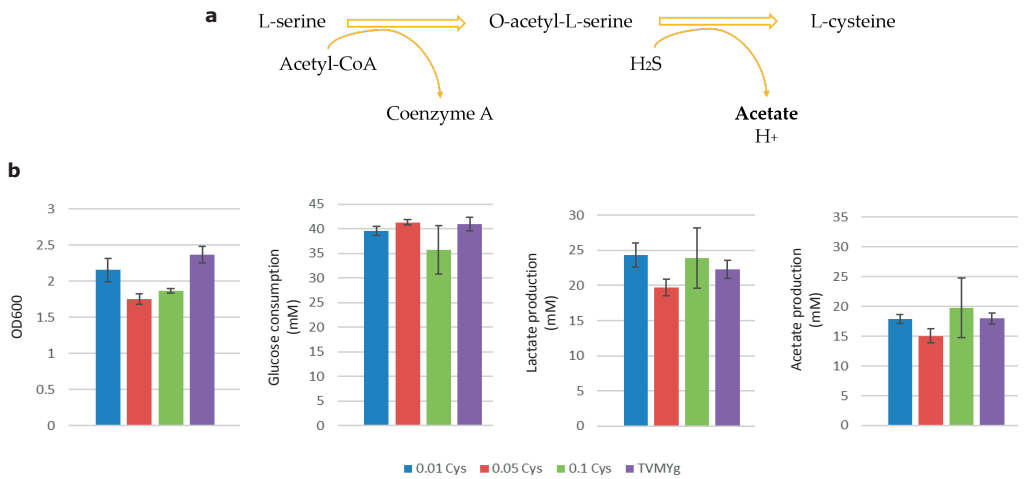


Figure 3a. Representation of the L-cysteine biosynthetic pathway. **b.** Effect of L-cysteine supplementation on *B. smithii* ET 138 growth and production profile. The cultures were grown in 10 ml TVMY medium supplemented with 10 g/L glucose and different concentrations of L-cysteine (0.01 g/l, 0.05 g/l, 0.1 g/l). Experiments were done in 50 ml Greiner tubes and harvested after 24 hours of culturing. Data shown were average of triplicates.

Another pathway that results in the production of acetate as by-product is the L-cysteine biosynthesis pathway (Fig. 3a), which is responsible for the conversion of L-serine to L-cysteine. The first reaction of this pathway is catalyzed by serine acetyltransferase, which transfers the acetyl-group of an acetyl-CoA molecule to L-serine, converting it to O-acetyl- L-serine. The second reaction is catalysed by cysteine synthase that converts O-acetyl- L-serine into L-cysteine and acetate in the presence of H_2S. Since there is no extracellular cysteine accumulation observed in the acetate-producing *B. smithii ET 138 c*ultures, the described pathway could not be the main acetate producing pathway. Nonetheless, its contribution to the acetate production by *B. smithii* ET 138 was investigated by culturing the *B. smithii* ET 138 Δ*sigF*-Δ*hsdR* strain in TVMY liquid medium supplemented with 10 g/l glucose, at 55°C, 150 rpm and in volumes previously shown to promote acetate production (10 ml medium in 50 ml

Greiner tubes). Moreover, the cultures were supplemented with either different L-cysteine concentrations (0.08-0.8mM), though low enough to avoid high levels of medium reduction, or different Na2SO4 concentrations (1.8-18mM). The hypothesis was that increasing Na2SO4 concentrations would increase the H2S availability, enhancing the L-cysteine and acetate production, whereas increasing L-cysteine supplementation would reduce the requirement for L-cysteine production leading to simultaneous decrement in acetate. 24 hours post inoculation,the optical density of the cultures at 600nm was measured, as indication of the growth, and samples from the culture supernatants were collected for HPLC analysis and determination of the type and amount of produced acids (Fig. 3b). It was observed that increasing L-cysteine concentrations negatively affected the growth of the *B. smithii* ET 138 Δ*sigF*-Δ*hsdR* cultures. This could be attributed to reduced oxygen availability caused by L-cysteine supplementation (even though the employed concentrations were low enough) or by the L-cysteine reaction with dihydroxyacetone phosphate (DHAP) and pyruvate, which forms thiohemiketal compounds[55]. This reaction reduces the amount of available pyruvate either for lactate formation, which is the main NAD^+-regeneration pathway in *B. smithii* ET 138 under oxygen limited conditions, or for introduction to the TCA cycle. Moreover, it has been shown that cysteine inhibits leucine, isoleucine, threonine, valine and RNA biosynthesis[56]. Notably, no statistically relevant effect on glucose consumption, as well as on the lactate and acetate production was observed for increasing L-cysteine supplementation. This result could be attributed to the low employed cysteine concentrations, compared to the glucose concentration, which most probably could not thermodynamically block the conversion of serine to cysteine via the examined pathway. Additionally, it was observed that Na_2SO_4 supplementation does not influence the acetate production, regardless the employed concentration (data not shown). Finally, from the acquired data it was not possible to determine the contribution of the L-cysteine biosynthesis pathway to the total *B. smithii* ET 138 acetate production. The transcriptional silencing of the L-cysteine biosynthesis pathway genes with the use of the ThermodCas9 tool could provide more insights in the future.

Acetolactate synthases (ALS) are involved in the conversion of pyruvate to branched-chain amino acids. They are evolutionary connected with pyruvate decarboxylases[57], which catalyse the conversion of pyruvate to acetaldehyde, but also present extensive sequence homology with pyruvate oxidases (POX), a class of enzymes responsible for the conversion of pyruvate to acetate or acetyl-phosphate[58]. Notably, the ALS large subunit from *B. smithii* ET 138 shows 28% identity to the POX from *Lactobacillus plantarum*, which converts pyruvate to acetyl-P[59,60]. Additionally, previous stu-

Lactobacillus plantarum, which converts pyruvate to acetyl-P[59,60]. Additionally, previous studies have demonstrated that some acetolactate synthases, such as the ones from E. coli61 and *S. typhimurium*[62] have an oxygenase catalytic centre that results in the production of peracetate and peroxide[62]. Thiol peroxidase can eliminate the highly reactive peroxide and the analysis of the comparative proteomics data presented in chapter 8 showed 4-fold higher thiol peroxidase expression for the 55°C *B. smithii* ET 138 culture, producing acetate, compared to the 37°C culture, producing no acetate. Moreover, previously acquired transcriptomic data from *B. smithii* ET 138 cultures grown at 55°C show that thiol peroxidase and acetolactate synthase are generally more expressed under oxygen-rich than oxygen-limited conditions. However, the comparative proteomics data show that acetolactate synthase is more abundant in the 37°C culture. Evaluation of a *B. smithii* ET 138 Δ*als* mutant strain or *als* knock-down strain, employing the developed ThermoCas9-based editing and silencing tools, could shed light on the influence of acetolactate synthase on the cell growth and production pattern.

SUCCINATE DEHYDROGENASE DELETION

The highest succinate production levels by *B. smithii* ET 138 were achieved under oxygen-rich conditions. Hence, it was hypothesized that the oxidative TCA cycle or/and the glyoxylate shunt were responsible for the produced succinate. Under oxygen-rich conditions, succinate is further oxidized into fumarate via a reaction catalysed by the succinate dehydrogenase complex. Deletions of the succinate dehydrogenase genes in *Corynebacterium glutamicum* and *Saccharomyces cerevisiae* has resulted in succinate-producing strains when cultured under oxygen-rich conditions[63-65]. A similar engineering approach has been executed in the *B. smithii* ET 138 metabolism, by deleting the genes encoding the different succinate dehydrogenase subunits using the developed Cas9-based editing tools. None of the attempted deletion trials was successful, strongly suggesting an essential role of the succinate dehydrogenase complex in *B. smithii* ET 138, at least under the tested conditions; this is in agreement with the absence of fumarate reductase genes in the *B. smithii* ET 138 genome. As previously performed for *E. coli*[66] and *Synechococcus elongatus*[67], the use of the ThermodCas9-based CRISPRi tool would allow for different levels of transcriptional silencing of the *B. smithii* ET 138 succinate dehydrogenase genes, in order to identify the minimal expression level of the succinate dehydrogenase complex required for growth, and to study the corresponding effects on the *B. smithii* ET 138 production pattern.

MALIC ENZYME DELETION

Previous studies on the metabolism of the *B. smithii* ET 138 Δ*ldhL* strain (lactate dehydrogenase-deletion mutant), under oxygen limited conditions, have indicated the existence of a futile cycle between pyruvate, oxaloacetate and malate via the catalytic activities of pyruvate carboxylase (*pyc*), malate dehydrogenase (*mdh*) and malic enzyme (*mae*)[68]. It was hypothesized that deletion of the me gene from the *B. smithii* ET 138 genome would disrupt this cycle, resulting either in the accumulation of either malate, or fumarate or other TCA cycle intermediates. Multiple attempts to delete a 400 bp long region from the me gene, including the start codon, employing the developed Cas9-based genome editing tools, did not work out as anticipated. The processes resulted in the construction of *B. smithii* ET 138 strains that contained the knock-out remnant of the *mae* gene in the native chromosomal position but also in a variation of the Cas9-based editing vector that carried the wild type *mae* gene in-between the employed homologous recombination flanks. Moreover, it was hypothesized that the vector was steadily integrated into the *B. smithii* ET 138 genome, or present as free plasmid in a very low number of copies, since the developed strains were showing resistance to the employed antibiotic after a few rounds of culturing in medium without antibiotic. Whole genome sequencing of the constructed strains could shed light on the genomic re-arrangements. Nonetheless, the failure to delete the *mae* gene from the genome of *B. smithii* ET 138 suggests that it is essential for viability. Hence, the use of the ThermodCas9-based CRISPRi technology for different levels of transcriptional silencing of the *mae* gene could facilitate the identification of the minimal required expression level of the enzyme and the study of the corresponding effects on the production pattern of *B. smithii* ET 138. Furthermore, these results indicate the importance of knockout and silencing studies in non-model organisms to improve understanding of their metabolism, as they reveal unexpected differences with related species. For example, in *B. subtilis*, deletion of any of its 4 predicted malic enzymes did not have any effect on growth on glucose but only on growth on malate[69]. This is in contrast with the findings of our study in *B. smithii*, where the deletion of the predicted malic enzymes is already problematic, even though its pyruvate node metabolism is rather similar to that of *B. subtilis*, which in turn shows striking differences from *E. coli*[70].

PHOSPHOENOLPYRUVATE CARBOXYKINASE OVEREXPRESSION

Aiming to improve the *B. smithii* ET 138 succinate production, it is essential to properly control the phosphoenolpyruvate (PEP)-pyruvate-oxaloacetate (OAA)

node since it fuels the TCA cycle, enables gluconeogenesis, and replenishes TCA cycle intermediates[70,71]. The *B. smithii* ET 138 PEP carboxykinase (PCK) catalyses the interconversion between OAA and PEP and is predicted to function in the gluconeogenic direction (unpublished data). We hypothesized that overexpression of native *B. smithii* ET 138 PCK could reverse its direction, as was previously shown for *E. coli*[72], contributing to ATP-generation, increased carbon flux towards OAA, and increased CO2 fixation. Plasmid-based overexpression of the native pck gene in *B. smithii* ET 138 cultures led to the formation of cell aggregates when put under the control of the strong P_{ldhL} promoter, while it did not have apparent effect on growth, production and substrate consumption when put under the control of the moderate P_{pta} promoter. In previous studies it was demonstrated that PCK overexpression in *E. coli* had positive effects on growth and succinate production when the medium was supplemented with $NaHCO_3$ at concentrations greater than 10 g/L[73]. The supplementation of *B. smithii* ET 138 cultures with 4 g/l $NaHCO_3$ negatively affected the growth, suggesting the requirement for further fine–tuning of CO_2 supplementation. Another possible engineering approach for increasing the intracellular PEP pool, hence forcing PCK function in the glycolytic direction, would be the knockout of the PEP-dependent phosphotransferase system (PTS), which is responsible for the glucose import to the cells, or the heterologous expression of a PCK known to function in the glycolytic direction, like the one from *Clostridium thermocellum*. Finally, it was reported in *Bacillus subtilis* that upon deletion of the pyruvate kinase (*pyk*) gene, the PCK operated in the reverse direction, converting PEP to OAA[71]; hence it would be of great interest to evaluate the effect of the *pyk* deletion from the *B. smithii* ET 138 genome on its metabolism.

CONCLUDING REMARKS

The fundamental studies on the metabolism of genetically accessible model and non-model microorganisms, as well as the development of accurate metabolic models that aim to facilitate the exploitation of their biotechnological potential, heavily relies on the development of extensive genetic toolboxes and time-efficient genome engineering tools. As reviewed in chapter 2, the first generation of CRISPR-Cas based bacterial genome engineering tools have facilitated and accelerated the studies on bacterial metabolisms, while a second upcoming generation of CRISPR-Cas tools will further enhance the throughput of these studies and will provide the potential for the development of novel applications.

Additionally, chapter 7 of this thesis reviewed the bacterial metabolic engineering studies that aimed for the development of green chemical- and biofuel-production workhorses.

The first aim of this thesis was to develop and further extend the existing genetic and genome engineering toolboxes of thermophilic and mesophilic bacteria, while the work towards that aim was mainly focused on the moderate thermophile *B. smithii* ET 138 and the mesophiles *R. sphaeroides* and *E. coli*. The expansion of the available *B. smithii* ET 138 and *R. sphaeroides* genetic toolboxes with vectors, promoters or other genetic elements was limited, as presented in chapters 3, 4, 5 and 9, while solutions were proposed for surpassing the identified bottlenecks in the tools development process in the near future. Nonetheless, the developed parts satisfactorily acted as the basis for the SpCas9- and ThermoCas9-based genome engineering tools presented in chapters 3, 4 and 5. This fact underlined the minimal requirements of these tools and suggested their high potential to be easily and efficiently adapted in other model and non-model organisms. Furthermore, chapter 6 demonstrated the potential of the antiCRISPR protein AcrIIC1 to interact with the ThermoCas9 endonuclease and form a ThermoCas9/antiCRISPR complex that *in vivo* in *E. coli* has similar activity as the catalytically inactive ThermodCas9 variant. This finding opens up the potential for the construction of a new generation of ThermoCas9-based genome engineering and synthetic biology tool. Finally, this chapter presented the targeting efficiencies of alternative CRISPR-Cas systems in *B. smithii* ET 138 and in *G. thermodenitrificans* T12, but none of them reached the desired efficiency levels for the development of genome engineering tools. Further studies on these systems, based on the proposed approaches, could either improve their targeting efficiencies or, in the cases of the native systems, could result in the identification of novel antiCRISPR systems.

The second aim of this thesis was to shed light on the unknown acetate production pathway of the *B. smithii* ET 138, as well as to enhance its dicarboxylic acids production profile. Even though the genetic background of the acetate production pathway was not elucidated, as presented in chapters 8 and 9, some possibilities were eliminated and further candidate pathways were proposed for future examination. Additionally, metabolic engineering studies presented in the same chapters, resulted in a strain with enhanced production in dicarboxylic acids and elucidated the essentiality of enzymes, such as the succinate dehydrogenase complex and the malic enzyme. Finally, studies on *B. smithii* ET 138 cultures at 37°C and 55°C revealed great differences in production profiles and enzymatic content, and hence suggested a temperature-based shift in the *B. smithii* ET 138 metabolism.

All in all, this thesis developed the first reported CRISPR-Cas based engineering tools for thermophilic bacteria, applicable also to mesophilic bacteria, and paved the way for the development of novel CRISPR-Cas and antiCRISPR based engineering tools. Additionally, this thesis deepen the knowledge on the *B. smithii* ET 138 metabolism, constructed a strain with improved dicarboxylic acids production profile and proposed future work for enhancement of its biotechnological potential.

References

1. Sheldon, R.A., Utilisation of biomass for sustainable fuels and chemicals: Molecules, methods and metrics. Catalysis Today, 2011. 167(1): p. 3-13.
2. Clardy, J., M.A. Fischbach, and C.T. Walsh, New antibiotics from bacterial natural products. Nature biotechnology, 2006. 24(12): p. 1541.
3. F Bosma, E., et al., Sustainable production of bio-based chemicals by extremophiles. Current Biotechnology, 2013. 2(4): p. 360-379.
4. Keller, L. and M.G. Surette, Communicationin bacteria: an ecological and evolutionary perspective. Nature Reviews Microbiology, 2006. 4(4): p. 249.
5. Head, I.M., D.M. Jones, and W.F. Röling, Marine microorganisms make a meal of oil. Nature Reviews Microbiology, 2006. 4(3): p. 173.
6. Dubnau, D., DNA uptake in bacteria. Annual Reviews in Microbiology, 1999. 53(1): p. 217-244.
7. Bosma, E.F., et al., Isolation and Screening of Thermophilic Bacilli from Compost for Electrotransformation and Fermentation: Characterization of a ET 138 as a New Biocatalyst. Applied and Environmental Microbiology, 2015. 81(5): p. 1874-1883.
8. Aune, T.E.V. and F.L. Aachmann, Methodologies to increase the transformation efficiencies and the range of bacteria that can be transformed. Applied microbiology and biotechnology, 2010. 85(5): p. 1301-1313.
9. Blokesch, M., Natural competence for transformation. Current Biology, 2016. 26(21): p. R1126-R1130.
10. Johnston, C., et al., Bacterial transformation: distribution, shared mechanisms and divergent control. Nature Reviews Microbiology, 2014. 12(3): p. 181.
11. Norman, A., L.H. Hansen, and S.J. Sørensen, Conjugative plasmids: vessels of the communal gene pool. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 2009. 364(1527): p. 2275-2289.
12. Shoeb, E., et al., Horizontal gene transfer of stress resistance genes through plasmid transport. World Journal of Microbiology and Biotechnology, 2012. 28(3): p. 1021-1025.
13. Huddleston, J.R., Horizontal gene transfer in the human gastrointestinal tract: potential spread of antibiotic resistance genes. Infection and drug resistance, 2014. 7: p. 167.
14. Gevers, D., G. Huys, and J. Swings, *In vitro* conjugal transfer of tetracycline resistance from Lactobacillus isolates to other Gram-positive bacteria. FEMS Microbiology Letters, 2003. 225(1): p. 125-130.
15. Kok, J., et al., The Evolution of gene regulation research in Lactococcus lactis. FEMS Microbiology Reviews, 2017. 41(Supp_1): p. S220-S243.
16. Frost, L.S., et al., Mobile genetic elements: the agents of open source evolution. Nature Reviews Microbiology, 2005. 3(9): p. 722.
17. Dahmane, N., et al., Impact of Cell Surface Molecules on Conjugative Transfer of the Integrative and Conjugative Element ICE of *Streptococcus thermophilus*. Applied and Environmental Microbiology, 2018. 84(5).
18. Choi, K.-H., A. Kumar, and H.P. Schweizer, A 10-min method for preparation of highly electrocompetent Pseudomonas aeruginosa cells: Application for DNA fragment transfer between chromosomes and plasmid transformation. Journal of Microbiological Methods, 2006. 64(3): p. 391-397.
19. Simon, R., U. Priefer, and A. Pühler, A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in gram negative bacteria. Nature biotechnology, 1983. 1(9): p. 784.
20. Cohen, S.N., et al., Construction of Biologically Functional Bacterial Plasmids *In Vitro*. Proceedings of the National Academy of Sciences, 1973. 70(11): p. 3240-3244.
21. Bosma, E.F., et al., Establishment of markerless gene deletion tools in thermophilic *Bacillus smithii* and construction of multiple mutant strains. Microbial Cell Factories, 2015. 14(1): p. 99.
22. Reeve, B., et al., The Geobacillus plasmid set: a modular toolkit for thermophile engineering. ACS synthetic biology, 2016. 5(12): p. 1342-1347.

23. Wieland, K.-P., B. Wieland, and F. Götz, A promoter-screening plasmid and xylose-inducible, glucose-repressible expression vectors for Staphylococcus carnosus. Gene, 1995. 158(1): p. 91-96.
24. Miyoshi, A., et al., A xylose-inducible expression system for Lactococcus lactis. FEMS Microbiology Letters, 2004. 239(2): p. 205-212.
25. Schumacher, M.A., et al., Structural basis for allosteric control of the transcription regulator CcpA by the phosphoprotein HPr-Ser46-P. Cell, 2004. 118(6): p. 731-741.
26. Marciniak, B.C., et al., High- and low-affinity cre boxes for CcpA binding in Bacillus subtilis revealed by genome-wide analysis. BMC Genomics, 2012. 13(1): p. 401.
27. Frenzel, E., et al., *In vivo* selection of sfGFP variants with improved and reliable functionality in industrially important thermophilic bacteria. Biotechnology for Biofuels, 2018. 11(1): p. 8.
28. Leenay, R.T., et al., Genome editing with CRISPR‐Cas9 in Lactobacillus plantarum revealed that editing outcomes can vary across strains and between methods. Biotechnology journal, 2018: p. 1700583.
29. Ran, F.A., et al., *In vivo* genome editing using *Staphylococcus aureus* Cas9. Nature, 2015. 520: p. 186.
30. Harrington, L.B., et al., A thermostable Cas9 with increased lifetime in human plasma. Nature Communications, 2017. 8(1): p. 1424.
31. Mougiakos, I., et al., Characterizing a thermostable Cas9 for bacterial genome editing and silencing. Nature Communications, 2017. 8(1): p. 1647.
32. Zetsche, B., et al., Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. Nature biotechnology, 2017. 35(1): p. 31.
33. Shmakov, S., et al., Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. Molecular cell, 2015. 60(3): p. 385-397.
34. Pyne, M.E., et al., Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *Clostridium*. Scientific reports, 2016. 6: p. 25666.
35. Zhang, J., et al., Exploiting endogenous CRISPR-Cas system for multiplex genome editing in *Clostridium tyrobutyricum* and engineer the strain for high-level butanol production. Metabolic Engineering, 2018. 47: p. 49-59.
36. Luo, M.L., et al., Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. Nucleic acids research, 2014. 43(1): p. 674-681.
37. Grissa, I., G. Vergnaud, and C. Pourcel, CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic acids research, 2007. 35(suppl_2): p. W52-W57.
38. Biswas, A., et al., CRISPRTarget: bioinformatic prediction and analysis of crRNA targets. RNA biology, 2013. 10(5): p. 817-827.
39. Jiang, W., et al., RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature Biotechnology, 2013. 31: p. 233.
40. Daas, M.J.A., et al., Engineering *Geobacillus thermodenitrificans* to introduce cellulolytic activity; expression of native and heterologous cellulase genes. BMC Biotechnology, 2018. 18(1)
41. Pyne, M.E., et al., Coupling the CRISPR/Cas9 system to lambda Red recombineering enables simplified chromosomal gene replacement in *Escherichia coli*. Applied and environmental microbiology, 2015: p. AEM. 01248-15.
42. Aparicio, T., V. de Lorenzo, and E. Martínez-García, CRISPR/Cas9-Based Counterselection Boosts Recombineering Efficiency in *Pseudomonas putida*. Biotechnology journal, 2018.
43. Cho, J.S., et al., CRISPR/Cas9-coupled recombineering for metabolic engineering of *Corynebacterium glutamicum.* Metabolic engineering, 2017. 42: p. 157-167.
44. Liu, P., N.A. Jenkins, and N.G. Copeland, A highly efficient recombineering-based method for generating conditional knockout mutations. Genome research, 2003. 13(3): p. 476-484.
45. Tong, Y., et al., CRISPR-Cas9 based engineering of actinomycetal genomes. ACS synthetic biology, 2015. 4(9): p. 1020-1029.
46. Sun, B., et al., A CRISPR-Cpf1-Assisted Non-Homologous End Joining Genome Editing System of *Mycobacterium smegmatis*. Biotechnology Journal, 2018. 13(9): p. 1700588.
47. Su, T., et al., A CRISPR-Cas9 Assisted Non-Homologous End-Joining Strategy for One-step Engineering of Bacterial Genome. Scientific Reports, 2016. 6: p. 37895.
48. Cui, L. and D. Bikard, Consequences of Cas9 cleavage in the chromosome of *Escherichia coli.* Nucleic acids research, 2016. 44(9): p. 4243-4251.
49. Kim, Y.B., et al., Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nature Biotechnology, 2017. 35: p. 371.
50. Gaudelli, N.M., et al., Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature, 2017. 551: p. 464.

51. Eid, A., S. Alshareef, and M.M. Mahfouz, CRISPR base editors: genome editing without double-stranded breaks. Biochemical Journal, 2018. 475(11): p. 1955-1964.
52. Halperin, S.O., et al., CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. Nature, 2018. 560(7717): p.248-252.
53. Kikuchi, G., The glycine cleavage system: composition, reaction mechanism, and physiological significance. Molecular and cellular biochemistry, 1973. 1(2): p. 169-187.
54. Aldehni, M.F., et al., Signal transduction protein PII is required for NtcA-regulated gene expression during nitrogen deprivation in the cyanobacterium Synechococcus elongatus strain PCC 7942. Journal of bacteriology, 2003. 185(8): p. 2582-2591.
55. Dugaiczyk, A., M.T. Malecki, and J.J. Eiler, The Effect of Cysteine on l-α-Glycerophosphate and Lactate Dehydrogenase Reactions. Journal of Biological Chemistr y, 1968. 243(9): p. 2236-2240.
56. Kari, C., et al., Mechanism of the growth inhibitory effect of cysteine on *Escherichia coli*. Microbiology, 1971. 68(3): p. 349-356.
57. Green, J.B., Pyruvate decarboxylase is like acetolactate synthase (ILV2) and not like the pyruvate dehydrogenase E1 subunit. FEBS letters, 1989. 246(1-2):p. 1-5.
58. Fagan, R.L., et al., 7.03 - Flavin-Dependent Enzymes, in Comprehensive Natural Products II2010, Elsevier: Oxford. p. 37-113.
59. Goffin, P., et al., Major role of NAD-dependent lactate dehydrogenases in aerobic lactate utilization in Lactobacillus plantarum during early stationary phase. Journal of bacteriology, 2004. 186(19): p. 6661-6666.
60. Lorquet, F., et al., Characterization and functional analysis of the poxB gene, which encodes pyruvate oxidase in Lactobacillus plantarum. Journal of bacteriology, 2004. 186(12): p. 3749-3759.
61. Chipman, D.M., R.G. Duggleby, and K. Tittmann, Mechanisms of acetohydroxyacid synthases. Current opinion in chemical biology, 2005. 9(5): p. 475-481.
62. Tse, J.M. and J.V. Schloss, The oxygenase reaction of acetolactate synthase. Biochemistry, 1993. 32(39): p. 10398-10403.
63. Litsanov, B., et al., Efficient aerobic succinate production from glucose in minimal medium with *Corynebacterium glutamicum*. Microbial biotechnology, 2012. 5(1): p. 116-128.
64. Arikawa, Y., et al., Effect of gene disruptions of the TCA cycle on production of succinic acid in *Saccharomyces cerevisiae*. Journal of bioscience and bioengineering, 1999. 87(1): p. 28-36.
65. Kubo, Y., H. Takagi, and S. Nakamori, Effect of gene disruption of succinate dehydrogenase on succinate production in a sake yeast strain. Journal of bioscience and bioengineering, 2000. 90(6): p. 619-624.
66. Lv, L., et al., Application of CRISPRi for prokaryotic metabolic engineering involving multiple genes, a case study: controllable P (3HB-co-4HB) biosynthesis. Metabolic engineering, 2015. 29: p. 160-168.
67. Huang, C.-H., et al., CRISPR interference (CRISPRi) for gene regulation and succinate production in cyanobacterium S. elongatus PCC 7942. Microbial cell factories, 2016. 15(1): p. 196.
68. Bosma, E.F., Isolation, characterization and engineering of *Bacillus smithii* : a novel thermophilic platform organism for green chemical production. 2015, Wageningen University: Wageningen.
69. Lerondel, G., et al., YtsJ has the major physiological role of the four paralogous malic enzyme isoforms in Bacillus subtilis. Journal of bacteriology, 2006. 188(13): p. 4727-4736.
70. Sauer, U. and B.J. Eikmanns, The PEP—pyruvate—oxaloacetate node as the switch point for carbon flux distribution in bacteria. FEMS microbiology reviews, 2005. 29(4): p. 765-794.
71. Zamboni, N., et al., The phosphoenolpyruvate carboxykinase also catalyzes C3 carboxylation at the interface of glycolysis and the TCA cycle of Bacillus subtilis. Metabolic engineering, 2004. 6(4): p.277-284.
72. Zhang, X., et al., Metabolic evolution of energy-conserving pathways for succinate production in *Escherichia coli*. Proceedings of the National Academy of Sciences, 2009: p. pnas. 0905396106.
73. Lee, S., P. Kim, and Y. Kwon, Influence of gluconeogenic phosphoenolpyruvate carboxykinase (PCK) expression on succinic acid fermentation in *Escherichia coli* under high bicarbonate condition. Journal of Microbiology and Biotechnology, 2006. 16: p. 1448-1452.

About the author

Ioannis Mougiakos was born on March 1st 1983 in Athens, Greece. He studied Biotechnological Applications and Technologies at the University of Ioannina (Greece) from which he obtained his BSc in 2008, and Acting at the Dilos School of Dramatic Arts (Athens, Greece), from which he obtained his BA in 2011. Between 2012 and 2014 he studied Cellular and Molecular Biotechnology at the Wageningen University and Research, from which he obtained his MSc. During his MSc studies he conducted his Major and Minor Thesis projects at the Molecular Ecology group of the Laboratory of Microbiology of the Wageningen University and Research, under the supervision of Dr. ir. Detmer Sipkema, on the heterologous expression and *in vitro* characterization of marine halogenases.

In October 2014 Ioannis initiated his TTW-funded PhD research project at the Bacterial Genetics and Bacterial Cell factories groups of the Laboratory of Microbiology of the Wageningen University and Research. During his project that was co-supervised by Prof. John van der Oost and Prof. Richard van Kranenburg Ioannis mainly focused on the development of CRISPR-Cas based genome engineering tools for thermophilic and mesophilic bacteria. Additionally, he employed the developed tools and engineered the metabolism of the biotechnologically relevant moderate thermophilic bacterium *Bacillus smithii* strain ET 138, aiming to enhance the production of dicarboxylic acids. Most of this work has been described in this thesis.

Currently, Ioannis is employed as a PostDoc researcher at the Bacterial Genetics group of the Laboratory of Microbiology of the Wageningen University and Research and he is focused on the development of high-throughput genetic and metabolic engineering tools for model and non-model prokaryotes and eukaryotes.

List of Publications

Mougiakos, I.*, Bosma, E. F.*, de Vos, W. M., van Kranenburg, R. & van dera Oost, J. **(**2016**)**. Next generation prokaryotic engineering: The CRISPR-Cas toolkit. Trends in Biotechnology 34, 575–587.

Mougiakos, I.*, Bosma, E. F.*, Weenink, K., Vossen, E., Goijvaerts, K., van der Oost, J., & van Kranenburg, R. (2017). Efficient genome editing of a facultative thermophile using mesophilic spCas9. ACS Synthetic Biology, 6(5), 849-861.

Mougiakos, I.*, Mohanraju, P.*, Bosma, E. F.*, Vrouwe, V., Bou, M. F., Naduthodi, M. I., Gussak, A., Brinkman, R. B. L., & van Kranenburg, R., & van der Oost, J. (2017). Characterizing a thermostable Cas9 for bacterial genome editing and silencing. Nature Communications, 8(1), 1647.

Mougiakos, I.*, Bosma, E. F.*, Ganguly J.*, van der Oost, J. & van Kranenburg, R. (2018) Hijacking CRISPR-Cas for high-throughput bacterial metabolic engineering: advances and prospects. Current Opinion in Biotechnology, 2018. 50: p. 146-157.

Mougiakos, I.*, Orsi, E.*, Ghiffary, M. R., de Maria, A., Perez, B. A., Fluitman, F., Kengen, S.W.M., Weusthuis, R.A.**, van der Oost, J.**. (2019) Efficient Cas9-based genome editing of Rhodobacter sphaeroides by native homologous recombination and non-homologous repair systems. Submitted

Trasanidou, D.*, Mohanraju P.*, Bouzetos, E., van Kranenburg, R., Mougiakos, I.**, van der Oost, J.** (2019). Characterizing an anti-CRISPR-based on/off switch for bacterial genome engineering. in preparation

List of Patent applications

J. van der Oost, R. van Kranenburg, E.F. Bosma, I. Mougiakos, P. Mohanraju. Thermostable Cas9 Nucleases. PCT/EP2017/082870. Priority date 14-12-16.

J. van der Oost, R. van Kranenburg, E.F. Bosma, I. Mougiakos, P. Mohanraju. Thermostable Cas9 Nucleases. PCT/EP2017/070796. Priority date 14-12-16.

J. van der Oost, R. van Kranenburg, E.F. Bosma, I. Mougiakos, P. Mohanraju. Thermostable Cas9 Nucleases. PCT/EP2017/070797. Priority date 14-12-16.

J. van der Oost, R. van Kranenburg, E.F. Bosma, I. Mougiakos. Microbial genome editing. PCT/EP2017/077975. Priority date 02-11-16.

Overview of completed training activities

Discipline specific activities

Meetings & conferences

- International Conference on Metabolic Engineering in Bacteria & BACELL, Amsterdam (NL). 2015
- Thermophilic Bacilli meeting, Bath (UK), 2015
- Molecular Genetics meeting, Lunteren (NL), 2015
- Microbial Biotechnology 2.0, Delft (NL), 2015
- Genome Engineering and Synthetic Biology: tools and technologies, Ghent (BEL), 2016
- Netherlands Biotechnology Conference (NBC-16), Wageningen (NL), 2016
- Thermophilic Bacilli meeting, Wageningen (NL), 2016
- BioSynSys conference, Bordeaux (FRA), 2016
- Microbial Biotechnology 3.0, Delft (NL), 2016
- 19th International Conference on Bacilli & Gram-Positive Bacteria, Berlin (GER), 2017
- Thermophilic Bacilli meeting, Nottingham (UK),2017
- Microbial Biotechnology 4.0, Delft (NL), 2017
- WUR Microbiology Centennial Symposium, Wageningen (NL), 2017
- CRISPR: From Biology to Technology and Novel Therapeutics, Sitges (ESP), 2017
- International Conference on CRISPR Technologies, Raleigh (NC, USA), 2017
- Metabolic engineering 12, Berlin (GER), 2018
- European Conference on Bacilli and Gram-Positive Bacteria, Bath (UK), 2018

Courses

- Metabolomics for Microbial Systems Biology, Campinas (BRA), 2015
- Metabolic engineering, Wageningen (NL), 2015
- Bioprocess Design, Delft (NL), 2017
- Microbial Physiology and Fermentation Technology, Delft (NL), 2018

General courses

- VLAG PhD week, Baarlo (NL), 2015
- Project and time management, Wageningen (NL), 2015
- Scientific Writing, Wageningen (NL), 2015
- Competence assessment, Wageningen (NL), 2017
- Career perspectives, Wageningen (NL), 2018

Optionals

- Preparation of research proposal
- PhD study trip to California (USA), 2015
- PhD study trip to Germany, Sweden, Denmark(organizing member), 2017
- Weekly Bacterial Genetics group meetings, Wa geningen (NL)
- Microbiology PhD meetings, Wageningen (NL)

ACKNOWLEDGEMENTS

What a journey... A 4 years long "PhD-Odyssey" of numerous PCRs, uncountable culturing experiments, a disturbing number of enzymatic assays, dozens of presentations and the 9 chapters presented here is officially completed. A challenging experience that forced me to meet my stress limits but also to empower my creativity. Most importantly, it was a journey with a lot of travelling companions who either helped remarkably in the outcomes of the project, or kept me mentally sound during the multiple ups-and-downs (some did both....). I couldn'thank enough all of you guys; I want you to know that the following few lines only represent a small fraction of how grateful I am to each one of you for helping me achieve my dream.

John and **Richard**, **Richard** and **John**. You were my two supervisors and promotors and I will address my thanks to you in an alphabetical order.

John, I cannot thank you enough for giving me the opportunity to become part of the Bacterial Genetics group and to start working on a promising, exciting but also very demanding topic. Even though you were initially not convinced about my determination, you were always approachable and available for discussions (especially if they involved crazy ideas!) which is extremely rare for a scientist at your level… As time passed by our relationship evolved, your doubts turned into trust and care, and this fact motivated me to work even harder towards becoming a better scientist and person. It was an honor and pleasure to be under your supervision and I consider myself extremely lucky. And thank you for allowing me to continue working in BacGen as a post-doc.

Richard you believed in me from the interview for this position and that did not change throughout the years. You have no idea how much your trust kept me standing on my feet through rough times. You were a great mentor, always open to discuss problems and results as well as to reduce stress levels when things were getting out of hands. My Mediterranean temperament still cannot understand how you can keep being so cool (but not indifferent) practically under any situation! It was always fun to spend time with you in dinners and trips, and our discussions greatly helped me to evolve both mentally and scientifically. Thanks you very much for these 4 years of fruitful co-operation and I hope that there are many more to come.

Elleke, I was in doubt whether I should thank you here or in the "family" section of my acknowledgements. From our first discussion, just before I accepted the PhD position offer, I sensed that we will co-operate very well. But I was wrong... Our co-operation was "ducking" awesome. We shared the same passion for work and the same curio-

sity-driven love for science. Even though you were the experienced post-doc you never made me feel inferior, being only helpful and kind. Our professional relationship quickly and effortlessly evolved into a solid friendship that will stand for the years to come. I feel also lucky that you introduced me to your great family (**Yolande, Klaas, Marjoleine** and **Mattijs**) who I deeply thank for always making me feel welcome, and of course your beloved husband. **Zeno**, thank you for all the great talks and experiences we shared throughout the years and of course for the amazing cover of this thesis! Elleke, I feel blessed that our lives crossed paths and I owe you a lot. Including a life supply of gin and tonics.

Prarthana, even though being late is your favorite hobby, you came into my PhD life just the right moment. Not only because of our enjoyable and fruitful collaboration but also because you allowed me to add into my life a charismatic friend. Almost from the first discussions we had we could communicate in a way that I can only communicate with my best friends. And I couldn't have enjoyed more our common trips, some of them with your amazing partner. **Ashwin**, keep taking good care of her and keep telling her your amazing jokes (she loves them…). Prarthana, one more common characteristic that sealed our bond is that we detest waking up early in the morning and we love working during the night. Hence we shared lots of evenings (and weekends) in the lab that I enjoyed a lot, but thank God that Helix closes at 10pm… Thanks for all Prat and I hope that there are still many more to come.

Willem, it was a great honor and pleasure to have you involved in my project and to write a paper together. You set a great role model, both as a scientist and as a person, for all of us in MIB. **Tom**, my crazy blue-eyed Dutch friend! I feel really lucky that I met you. Not only because most of the work on the bioreactors wouldn't had been done without your contribution but also because you are a great guy. You love pranking people, especially after having a couple of beers, but you also really care about your colleagues. Thank you for the great time we had as office-mates and for the way you supported me through rough times. I wish you and your family (**Juliette**, **Sophie** and **Coen**) all the best! **Mariska**, **Rudy** and **Bastienne**, my great Corbion collaborators! Mariska thank you for always being helpful and available to discuss technical problems, especially at the beginning of the project, but also for being such a pleasant company during the conferences we attended. Rudy and Bastienne, your contribution to the finalization of this thesis was extremely valuable and I will always be thankful and happy that I met and collaborated with you. And of course thanks the rest of the Corbion team and **Elrike** for the fun meetings

Enrico and **Ruud**, our collaboration gradually evolved into a really nice and respectfulrelation that I am thankful of. I'm really excited for our on-going projects and I can't wait for our scientific stories to come out soon. **Giovanni**, we started our co-operation exactly at the moment I submitted this thesis but our interaction during this short period of time made me profoundly appreciate your character and your working ethos.

Big part of the work described in this thesis was practically performed by a number of very talented and hard-working students as part of their thesis projects. And I am very grateful to all of you guys. **Kirsten** and **Alex** you were my first students and what a way to start my "supervision" career! With your hard work and remarkable abilities you substantially contributed to the success of two projects! **Eric**, I think you are the most chill-out German I have ever met, yet you successfully survived a not so chill-out thesis… I am happy that you are back to BacGen, as a colleague this time. **Koen**, you are a great guy and a smart but also creative scientist, and working with you was nothing but joy. **Mihris**, you are definitely the most positive-thinking person I have ever met and I couldn't be more proud for all your achievements. I am super happy to share the shame office with you as colleagues. **Agnes**, your crazy Catalan spirit and your remarkable abilities brought a fresh air to the metabolic engineering part of the project and made it even more interesting. **Valentijn**, I can't think of a more persistent person with great will to learn and improve; working with you evolved into a successful and joyful experience that taught me a lot. **Max**, your intelligence pushed the projects you worked on to a higher level. You are a great asset for the BacGen group. **Guus**, your enthusiasm and hard work made our discussions enjoyable and helped me understand a lot on the *in silico* "magic" you were working on. **Meli**, it was so much fun working with you and watching you being so calm but at the same time motivated and active; I'm still impressed on how you handled all these bioreactors at the same time. **Ghiffa**, your great determination to succeed and your organizational skills made working with you a great experience and I am happy we collaborated. **Alberto**, handling a hippy Italian student was the last thing I expected to do during my PhD, nonetheless it turned into an experience that I will remember with great joy. **Despoina**, you are one of the most motivated and hard-working persons I have ever met, with almost always a big smile on your face. I am extremely happy to continue working with you on your PhD project. **Maria**, even though you were not officially my student our long conversations on both scientific and personal stuff made me admire your spirit, capabilities and determination. **Mauricio**, **Eugene** and **James**, I had the pleasure to start co-operating with you at the very

end of my PhD (almost at the beginning of my post-doc) yet I really want to thank you for your kindness and hard work on the follow-up projects from this thesis. Guys, thank you all for your efforts and for your tolerance. Almost all of you have either started (or will soon start) a PhD or found good jobs, so I am a really happy and proud supervisor. Keep up the good work!

The last 4,5 years I was lucky enough to be part of a great scientific group that also feels a bit like a "scientific family". I want to cordially thank all the members of this family that have already maade the next move in their careers; **Teunke, Nico, Franklin, Tim, Tijn, Alex,** thanks for the fun times and the discussions. I also want to thank the still present or new members of BacGen; **Yifan, Sjoerd, Jorrit, Jeroen, Joyshree, Costas, Wen, Melvin, Tijs, Lione, Mamou, Thomas, Belen, Joep, Ismael, Janneke, Catarina, James, Carina, Rob, Serve** and **Raymond**. Even though I didn't have many common projects with you I really enjoyed our BacGen meetings, our dinners (especially the "Christmas" dinners that are usually more like "almost Easter" dinners), coffee breaks or spontaneous corridor conversations and I want to thank you for sharing our problems (scientific or personal) and experiences.

Except for the BacGen group people there are also many people that helped in making my everyday PhD life a bit more sufferable and they belong either to the MIB or the SSB groups. **Johanna** Goodlife! I can't even imagine this PhD without your presence, Miss Spongy… Thanks for the convos, the fun, the parties, your great contribution in organizing the PhD trip, for everything! **Detmer**, you were a great MSc thesis supervisor, you played an important role in the decision to get this PhD position offer and -most importantly- you are the first Dutch person I ever met who was actually fun to talk to… Thanks for all and for being an amazing PhD trip supervisor! Equally great PhD trip supervisor though was **Irene** (aka mama :-P), with her incredible Spanish-loud laugh and warm spirit. Thank you so much for our conversations and for the fun times we shared. **Peer** (aka daddy) I was humbly honored when you publicly chose me as the best "food option" in case of emergency and be sure that I will not forget it... **Cristina**, my crazy Italian friend! Grazie milla for the brunches, the dinners (the good food in general), the discussions and the parties. Miss you tanto! **Bastian**, you are by far my favorite German and I miss partying with you (especially when strippers are involved…). **Monika**, your crazy spirit and distinctive laugh is missed a lot at MIB. Can't wait to party with you again. **Sven**, it´s always fun to party with you and thanks for agreeing to take my defense photos. **Hanne**, it has always been so much fun interacting with you and I am glad you finally understood that BacGen is the right place for you... **Diana**, thanks for the nice

conversations and the fun time in parties. Fons, thank you for the coffee corner chats that we frequently had; many times proved to be quite didactic for me. **Caroline**, I really enjoyed our talks throughout the years. **Anja**, my bestest friend and **Heidi**, my 2nd bestest friend! You are the joy of life and I thank you very much for our fun conversations and your help on many practical and bureaucratic stuff. **Rob**, **Sjon**, **Philippe, Stephen**, **Tom**, **Ton**, **Ineke** and **Merlijn**, the technicians dream team! Thank you for taking care of the lab organization stuff and allow us to work efficiently and safely in a nice environment. **Juelli**, **Loowee**, **Anna**, **Klaudyna**, **Lara**, **Yuan**, **Yue** thank you for all the fun times and especially the amazing US trip experience. **Hugo**, **Aleks**, **Jie**, **Martijn**, **Ran**, **Sudarshan**, **Hikmah**, **Tika**, **Daan**, **Emmy**, **Indra**, **Erika, Ruben**, **Linde** and **Nong,** you were great companions during the European PhD trip and I am happy that I got to know you better. **Ruth**, **Leire, Carrie, Dorrett** and **Lyon,** thanks for the fun and funny conversations during lunches and drinks. **Enrike**, **Yahya** and **Daniel** thanks for the nice dinners and for introducing me to the Spanish Pop culture.

And now it's time to thank the Greek Wageningen gang. **Βιβή**, **Αγγελική**, **Ήβη**, **Σταμάτη** και **Γιάννη** (Τάμπακοπουλος) σας ευχαριστώ από τα βάθη της καρδιάς μου που με βοηθήσατε να αποφύγω την κατάθλιψη τα πρώτα δύσκολα χρόνια στο Βαχενίτσι. Ειδικά Βιβή και Αγγελική, η επικοινωνία μας με βοήθησε και τα επόμενα χρόνια να το (σχεδόν) καταφέρω... **Μάνο** και **Haimil**, κρατάει χρόνια αυτή η κολώνια αδέρφια και δε λέει να ξεθυμάνει... σας ευχαριστώ πολυ για την υποστήριξη, για τις ωραιές chill-out κουβέντες μας (κυρίως με τη Haimil γιατί που να μιλήσεις ήρεμα με τον τρελλοκρητικό) και για τα όνειρα που μοιραζόμαστε και θα συνεχίσουμε να μοιραζόμαστε. **Γιάννο** (Κους Κους), **Ελευθερία, Μένια**, **Κώστα**, **Νικόλα**, **Χρήστο**, **Προκόπη**, **Σταμάτη**, σας ευχαριστώ που όλα αυτά τα χρόνια διατηρήσατε ζωντανή τη φλόγα του Ελληνάρα μέσα μου (οριακά...). Σας αγαπώ όλους γλυκούλια. Άντε σύντομα στα δικά σας οι λεύτερες.

The international friends I made in Wageningen played an important role in my personal and scientific development. **Shreyans**, my dear friend with the kind heart. After a long period away from studies I would have never gotten back on track during the MSc program if I didn't have your help. I will be forever thankful and it is my great honour to be your best man in your upcoming wedding with **Jose**. I consider both of you part of my family and your happiness makes me happy too (and without the need for anything "recently acquired"). **Kal**, our dinners with the rest of the Indian gang revealed to me not only a nice and caring person but also an excellent cook! Thank you for our interesting (yet sometimes weird) conversations and for the amazing food. **Rob**, **Elke** and **Daniel** (aka Danonito) my gym brothers and sister, thanks for

helping me get back on track with my physical condition and I really appreciateour nice conversations about life and future. I'm really glad I met you and I really value our friendship. **Jerry** (aka Lolito) best housemate evaaaaa!!! I would love to write more about you, crazy Americano, but "I'm sooooo tiiireeeeed"... I will certainly miss you buddy.

Bas schatje, you were one of the most important supporters throughout my PhD. Despite our difficult times, your love and care kept me going at the moments I was losing hope. You were a great partner and friend, and this project owes a big part of its success to you. And love may change forms but still remains love... **Anja, Ann** and **Tanja**, my beautiful smiley ladies, and my friend **Mantas,** thank you for all the love, the support and for the fun, family dinners and drinks we had the last few years!

Γιάννη και **Δημήτρη** μου, **Δημήτρη** και **Γιάννη** μου. Είστε οι Φίλοι μου. Ήσασταν από τα βασικότερα ψυχικά μου στηριγματα σε όλην αυτήν την προσπάθεια (ίσως περισσότερο από όσο και οι ίδιοι έχετε καταλάβει) και 2 απο τους κυριότερους λόγους που χαιρόμαι να έρχομαι στην Ελλάδα. Σας ευχαριστώ για όλες τις εύκολες και δύσκολες στιγμές που περάσαμε μαζί και για όσες πρόκειται να περάσουμε στο μέλλον. Γιατί έρχονται πολλές...(κυρίως καλές ελπίζω). **Ιορδάνη**, **Ιωάννη** και **Γιώργο** χαίρομαι αφάνταστα που σας έχω στη ζωή μου και σας ευχαριστώ που προσέχετε και αγαπάτε τους πιο πάνω κυρίους και την ακόλουθη κυρία. Ότι και αν λένε, σας έχουν απόλυτη ανάγκη.

Ελίζα μου, Φούλα μου αγαπημένη. Έχουμε περάσει καλές αλλά και δύσκολες περιόδους, όμως γνώριζα και γνωρίζω πως πάντα θα χαιρόμαστε ο ένας με τις επιτυχίες του άλλου και θα λυπόμαστε με τις λύπες του. Χωρίς την υποστήριξή σου (πρακτική και ψυχολογική) και την αγάπη σου δε θα μπορούσα ποτέ να έχω ολοκληρώσει αυτή τη διαδρομή. Ήσουν και είσαι το άλλο μου μισό και δεν υπάρχει κάτι που να με έχει γεμίσει περισσότερη περηφάνεια από το ότι ειμαι ο αδερφός σου. , σε ευχαριστώ που την προσέχεις.

Μαμά και **Μπαμπά**, τα λόγια δε φτάνουν για να εκφράσω την αγάπη και την ευγνωμοσύνη μου. Όλη μου η διαδρομή μέχρι εδώ, με τα καλά αλλά και τα στραβοπατήματα, φέρει τη στήριξη και την υπογραφή σας. Περάσαμε πολλές δυσκολες στιγμές αλλά και πολλές όμορφες, ειδικά όταν η αγάπη κατάφερνε να κερδίσει τους εγωισμούς. Και να μάστε εδω, όλοι μαζί, χέρι χέρι. Γιατί χωρίς το δικό σας χέρι το δικό μου θα'ταν λειψό.

Γιαγιά, όλα για σένα, για πάντα...

The research presented in this thesis was financially supported by the by the Dutch Technology Foundation TTW, which is part of The Netherlands Organization for Scientific Research (NWO) and which is partly funded by the Dutch Ministry of Economic Affairs.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by: Zeno van den Broek

Printed by: Proefschriftmaken.nl || www.proefschriftmaken.nl