Manuscript

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], Alexander 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

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

*These authors contributed equally to this work.

[†]To whom correspondence should be addressed: J.v.d.O. (john.vanderoost@wur.nl)

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: an RNA-guided DNA-endonuclease 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 the 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.

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 as SpCas9 is not active at elevated temperatures[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, 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). 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 (Supplementary Fig. 1)[35]. 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[11,35]. 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.

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[36]. As only two hits were obtained with phage genomes (Supplementary Fig. 2A), it was decided to proceed with an *in vitro* PAM determination approach. The predicted sgRNA sequence was generated by *in vitro* transcription, including a spacer that should allow for ThermoCas9-based targeting of linear dsDNA substrates with a matching protospacer. The protospacer was flanked at its 3'-end by randomized 7-base pair (bp) sequences. After performing ThermoCas9-based cleavage assays at 55°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[17,37]. For this purpose, and taking into account the results from the *in silico* ThermoCas9 PAM prediction (Supplementary Fig. 2), 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. 2B) also suggested the significance of a cytosine at the 5th and an adenine at the 8th PAM positions.

To further clarify the ambiguity of the PAM at the 6th and 7th PAM positions, we generated a set of 16 different target DNA fragments in which the matching protospacer was flanked by 5'-CCCCCNNA-3'PAMs. Cleavage assays of these fragments (each with a unique combination of the 6th and 7th nucleotide) were performed in which the different components (ThermoCas9, sgRNA guide, dsDNA target) were pre-heated separately at different temperatures (20, 30, 37, 45, 55 and 60°C) for 10 min before combining and incubating them for 1 hour at the corresponding assay temperature. When the assays were performed at temperatures between 37°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. 3), making these sequences the most preferred PAMs. Our findings demonstrate that at its lower temperature limit, ThermoCas9 only cleaves fragments with a preferred PAM. This characteristic could be exploited during *in vivo* editing processes, for example to avoid off-target effects in eukaryotic Cas9-based genome editing.

Metal ion dependency, thermostability and truncations

Previously characterized, mesophilic Cas9 endonucleases employ divalent cations to catalyze the generation of DSBs in target DNA[14,38]. 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[14,20,39]. 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.

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 cleavage efficiency drops significantly when a spacer of 18 nt is used.

In vivo, the ThermoCas9:sgRNA RNP complex is probably formed within seconds. 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 3C). This finding confirmed our assumption that the ThermoCas9 stability strongly correlates with the association of an appropriate sgRNA guide[40].

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 previous analysis[28,29], the mesophilic SpCas9 was inactive above 45°C (Figure 3D); above this temperature SpCas9 activity rapidly decreased to undetectable levels. In contrast, ThermoCas9 cleavage activity could be detected between 25 and 65°C (Figure 3D). 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[40,39], we tested the activity of ThermoCas9 on ssDNA substrates. However, no cleavage was observed, indicating that ThermoCas9 is a dsDNA nuclease (Supplementary Fig. 5B).

ThermoCas9-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*[24,25,41,42]. 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[24,28,31,43]. In order to use a minimum of genetic parts, we followed a single plasmid approach. We constructed a set of pNW33n-based pThermoCas9 plasmids containing the *thermocas9* gene under the control of the native *xylL* promoter (P_{xylL}), a homologous recombination template for repairing Cas9-induced double stranded DNA breaks within a gene of interest, and a sgRNA expressing module under control of the constitutive *pta* promoter (P_{pta}) from *Bacillus coagulans* (Figure 4A).

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[28,43]. 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 4B), 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 or post Cas9 cleavage[11,28,44]. 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[11,28,44]. As the promoter we use here for *thermocas9*-expression is not sufficiently controllable and HR is inefficient in *B. smithii*[28,43], 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 Cas9 was very tightly controlled by the growth temperature rather than by gene expression. This allowed for extended time for the cells to perform HR prior to Cas9 counter-selection, resulting in a higher *pyrF* deletion efficiency[28]. We anticipate that the use of a tightly controlled promoter will increase efficiencies of the ThermoCas9-system.

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[45,46]. 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 Pm-promoter, a homologous recombination template for deletion of the *pyrF* gene and a sgRNA expressing module under the control of the constitutive P3

promoter. After transformation of *P. putida* KT2440 cells and PCR confirmation of plasmid integration, a colony was inoculated in selective liquid medium for overnight culturing at 37°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 heterologously produced, purified and used for an *in vitro* cleavage assay with the same DNA target as used in the aforementioned ThermoCas9 assays; no cleavage was observed confirming the catalytic inactivation of the nuclease.

Towards the development of a ThermodCas9-based CRISPRi tool, we aimed for the transcriptional silencing of the highly expressed *ldhL* gene from the genome of *B. smithii* ET 138. We constructed the pNW33n-based vectors pThermoCas9i_*ldh*L 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 4C). The pThermoCas9i_*ldh*L 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[18,47], 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, as described previously[31]. The growth of the pThermoCas9i_*ldh*L cultures was 50% less than the growth of the pThermoCas9i_ctrl cultures (Figure 4E). We have previously shown that deletion of the *ldhL* gene leads to severe growth retardation in *B. smithii* ET 138 due to a lack of Ldh-based NAD^+-regenerating capacity under micro-aerobic conditions[43]. 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_*ldh*L cultures compared to the pThermoCas9i_ctrl cultures (Figure 4E).

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[5,48], 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, a strain that we previously 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 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[49].

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 Δ*sigF* Δ*hsdR*[28] was used for the gene editing and silencing experiments using ThermoCas9. It was grown in LB2 medium[43] at 55°C. 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. 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, after which incubation was continued at 16°C. For cloning PAM constructs for 6th and 7th, and 8th positions, DH5a 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 [52] 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.*[53]. For all *E. coli* strains, if required chloramphenicol was used in concentrations of 25 mg/L and kanamycin in 50 mg/L. *Pseudomonas 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 and 3-methylbenzoate in a concentration of 3 mM.

ThermoCas9 expression and purification

ThermoCas9 was PCR-amplified from the genome of *G. thermodenitrificans* T12, then cloned and heterologously expressed in *E. coli* Rosetta (DE3) and purified using FPLC by a combination of Ni^{2+}-affinity, ion exchange and gel filtration chromatographic steps. The gene sequence was inserted into plasmid pML-1B (obtained from the UC Berkeley MacroLab, Addgene #29653) by ligation-independent cloning using oligonucleotides (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 OD_{600nm} of 0.5-0.6. Expression was induced by the addition of IPTG to a final concentration of 0.5 mM and incubation was continued at 16°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 16,000×*g* 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 incubating them at the various assay temperatures in a cleavage buffer (100 mM sodium phosphate buffer (pH=7), 500 mM NaCl, 25 mM $MgCl_2$, 25 (V/V%) glycerol, 5 mM dithiothreitol (DTT)) for 1 hour. Each cleavage reaction contained 160

nM of ThermoCas9 protein, 4 nM of substrate DNA, and 150 nM of synthetized sgRNA. Reactions were stopped by adding 6x loading dye (NEB) and run on 1.5% agarose gels. Gels were stained with SYBR safe DNA stain (Life Technologies) and imaged with a Gel DocTM EZ gel imaging system (Bio-rad).

Library construction for *in vitro* PAM screen

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

PAM screening assay

The PAM screening of thermoCas9 was performed using *in vitro* cleavage assays, which consisted of (per reaction): 160 nM of ThermoCas9, 150 nM in vitro transcribed sgRNA, 4 nM of DNA target, 4 µl of cleavage buffer (100 mM sodium phosphate buffer pH 7.5, 500 mM NaCl, 5 mM DTT, 25% glycerol) and MQ water up to 20 µl final reaction volume. The PAM containing cleavage fragments from the 55°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 [54].

***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 transformed of *B. smithii* ET 138 electro-competent cells, which were prepared according to a previously described protocol [43]. 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 (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 P3 promoter was amplified from pSW_I-SceI (BG2373, BG2374). This promoter fragment was fused to the sgRNA fragment in an overlap extension PCR using primers BG2372 and BG2373 making use of the overlaps of primers BG2371 and BG2374. ThermoCas9 was amplified from the pThermoCas9_ctrl plasmid (BG2375, BG2376). The inducible Pm-XylS system, to be used for 3-methylbenzoate induction of ThermoCas9 was amplified from pSW_I-SceI (BG2377, BG2378).

Editing protocol for *P. putida*

Transformation of the plasmid to *P. putida* was performed according to Choi *et al.*[55]. 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.

RNA isolation

RNA isolation was performed by the phenol extraction based on a previously described protocol [56]. Overnight 10 mL cultures were centrifuged at 4°C and 4816×*g* for 15 min and immediately used for RNA isolation. After removal of the medium, cells were suspended in 0.5 mL of ice-cold TE buffer (pH 8.0) and kept on ice. All samples were divided into two 2 mL screw-capped tubes containing 0.5 g of zirconium beads, 30 µL of 10% SDS, 30 µL of 3 M

sodium acetate (pH 5.2), and 500 µL of Roti-Phenol (pH 4.5−5.0, Carl Roth GmbH). Cells were disrupted using a FastPrep-24 apparatus (MP Biomedicals) at 5500 rpm for 45 s and centrifuged at 4°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 of chloroform−isoamyl alcohol (Carl Roth GmbH) was added, after which samples were centrifuged at 4 °C and 18 400×*g* for 3 min. 300 µL of the aqueous phase was transferred to a new tube and mixed with 300 µL of the lysis buffer from the high pure RNA isolation kit (Roche). Subsequently, the rest of the procedure from this kit was performed according to the manufacturer's protocol, except for the DNase incubation step, which was performed for 45 min. The concentration and integrity of cDNA was determined using Nanodrop-1000 Integrity and concentration of the isolated RNA was checked on a NanoDrop 1000.

Quantification of mRNA by RT-qPCR

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

HPLC

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

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 determine the targets of 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 metabolic 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. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520,** 186–191 (2015).
36. 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).
37. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J.-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24,** 1012–9 (2014).
38. 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).
39. 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).
40. 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).
41. 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).
42. 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).

43. Bosma, E. F. *et al.* Establishment of markerless gene deletion tools in thermophilic Bacillus smithii and construction of multiple mutant strains. *Microb. Cell Fact.* **14,** 99 (2015).
44. Oh, J.-H. & van Pijkeren, J.-P. CRISPR-Cas9-assisted recombineering in Lactobacillus reuteri. *Nucleic Acids Res.* **42,** e131 (2014).
45. 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).
46. 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).
47. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8,** 2180–2196 (2013).
48. Weinberger, A. D., Wolf, Y. I., Lobkovsky, A. E., Gilmore, M. S. & Koonin, E. V. Viral diversity threshold for adaptive immunity in prokaryotes. *MBio* **3,** e00456-12 (2012).
49. Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science (80-.).* **343,** 1247997–1247997 (2014).
50. 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).
51. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33,** 1870–1874 (2016).
52. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular cloning : a laboratory manual*. (Cold Spring Harbor Laboratory, 1989).
53. *Current Protocols in Molecular Biology*. (John Wiley & Sons, Inc., 2001). doi:10.1002/0471142727
54. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Res.* **14,** 1188–1190 (2004).
55. 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).
56. 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).

Acknowledgements

We would like to thank Koen Weenink, Steven Aalvink and Bastienne Vriesendorp for their technical assistance. R.v.K. is 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. R.B. is financially supported by Corbion. A patent application has been filed related to this work.

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

The authors declare no competing financial interests.

Data availability

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

Corresponding author

Correspondence and requests for materials should be addressed to J.v.d.O. (john.vanderoost@wur.nl).

Figures

Figure 1. The *Geobacillus thermodenitrificans* T12 type-IIC CRISPR-Cas locus encodes a thermostable Cas9 homolog, 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 [50] is shown, with different colours for the domains.

(B) Phylogenetic tree of Cas9 orthologues those are highly identical to ThermoCas9. Evolutionary analysis was conducted in MEGA7[51].

(C) SDS-PAGE of ThermoCas9 after purification by metal-affinity chromatography and gel filtration. The migration of the obtained single band is consistent with the theoretical molecular weight of 126 kD of the apo-ThermoCas9.

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. 3.

Figure 3. ThermoCas9 is active at a wide temperature range and its thermostability increases when bound to sgRNA.

(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 at least two biological replicates are shown, with error bars representing S.D.

(C) To identify the maximum temperature, endonuclease activity of ThermoCas9:sgRNA RNP complex was assayed after incubation at 60˚C, 65˚C and 70˚C for 5 or 10 min. The pre-heated DNA substrate was added and the reaction was incubated for 1 hour at the corresponding temperature.

(D) Comparison of active temperature range of ThermoCas9 and SpCas9 by activity assays conducted after 5 min of incubation at the indicated temperature. The pre-heated DNA substrate was added and the reaction was incubated for 1 hour at the same temperature.

Figure 4. ThermoCas9-based genome engineering in thermophiles.

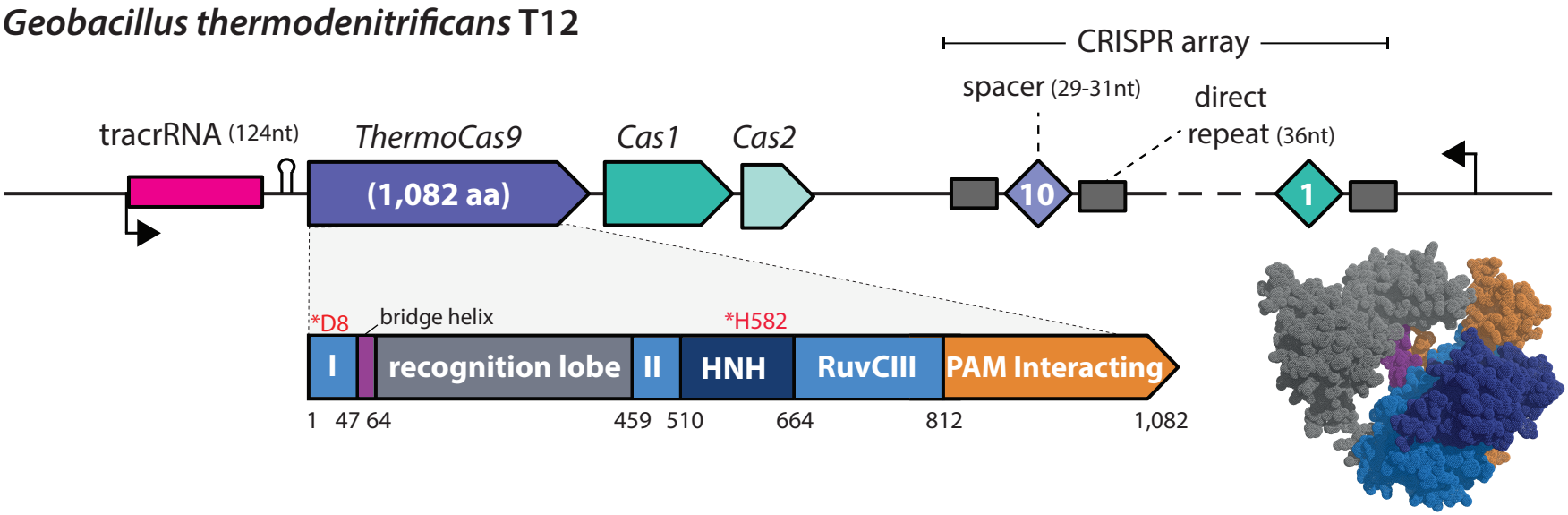
(A) Schematic overview of the basic pThermoCas9_Δgene-of-interest (goi) construct. The *thermocas9* gene was introduced either to the pNW33n (*B. smithii*) or to the pEMG (*P. putida*) vector. Homologous recombination flanks were introduced upstream *thermocas9* and encompassed the 1kb (*B.smithii*) or 0.5kb (*P. putida*) upstream and 1kb or 0.5 kb downstream region of the gene of interest (goi) in the targeted genome. A sgRNA-expressing module was introduced downstream the *thermocas9* gene. As the origin of replication (ori), replication protein (rep), antibiotic resistance marker (AB) and possible 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.

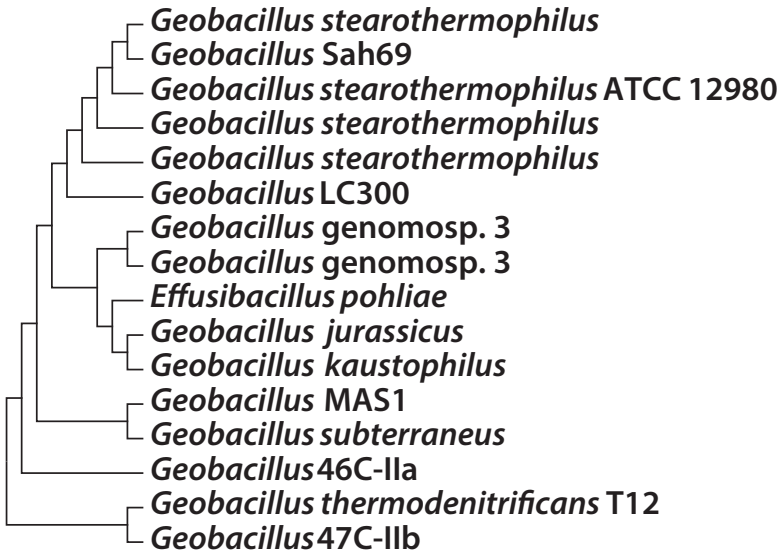
(C) Schematic overview of the basic pThermoCas9i_goi construct. Aiming for the expression of a catalytically inactive ThermoCas9 (ThermodCas9: D8A, H582A mutant), the corresponding mutations were introduced to create the *thermodcas9* gene. The *thermodcas9* gene was introduced to the pNW33n vector. A sgRNA-expressing module was introduced downstream the *thermodcas9.*

(D) Graphical representation of the production, growth and RT-qPCR results from the *ldhL* silencing experiment using ThermodCas9. The graphs represent the lactate production, optical density at 600nm and percentage of *ldhL* transcription in the repressed cultures compared to the control cultures. Average values from at least two biological replicates are shown, with error bars representing S.D.

a.


Geobacillus thermodenitrificans T12
CRISPR array
spacer (29-31nt)
direct repeat (36nt)
tracrRNA (124nt)
ThermoCas9
(1,082 aa)
Cas1
Cas2
10
1
*D8
bridge helix
*H582
I
recognition lobe
II
HNH
RuvCIII
PAM Interacting
1 47 64 459 510 664 812 1,082

b.


Geobacillus stearothermophilus
Geobacillus Sah69
Geobacillus stearothermophilus ATCC 12980
Geobacillus stearothermophilus
Geobacillus stearothermophilus
Geobacillus LC300
Geobacillus genomosp. 3
Geobacillus genomosp. 3
Effusibacillus pohliae
Geobacillus jurassicus
Geobacillus kaustophilus
Geobacillus MAS1
Geobacillus subterraneus
Geobacillus 46C-IIa
Geobacillus thermodenitrificans T12
Geobacillus 47C-IIb

c.

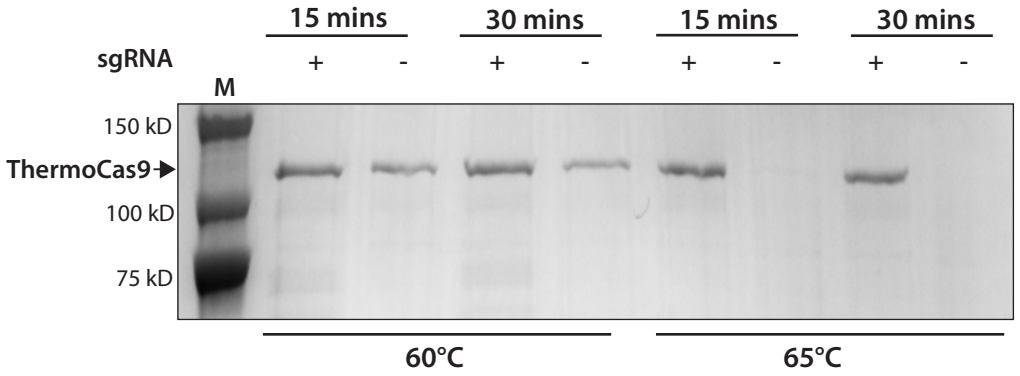
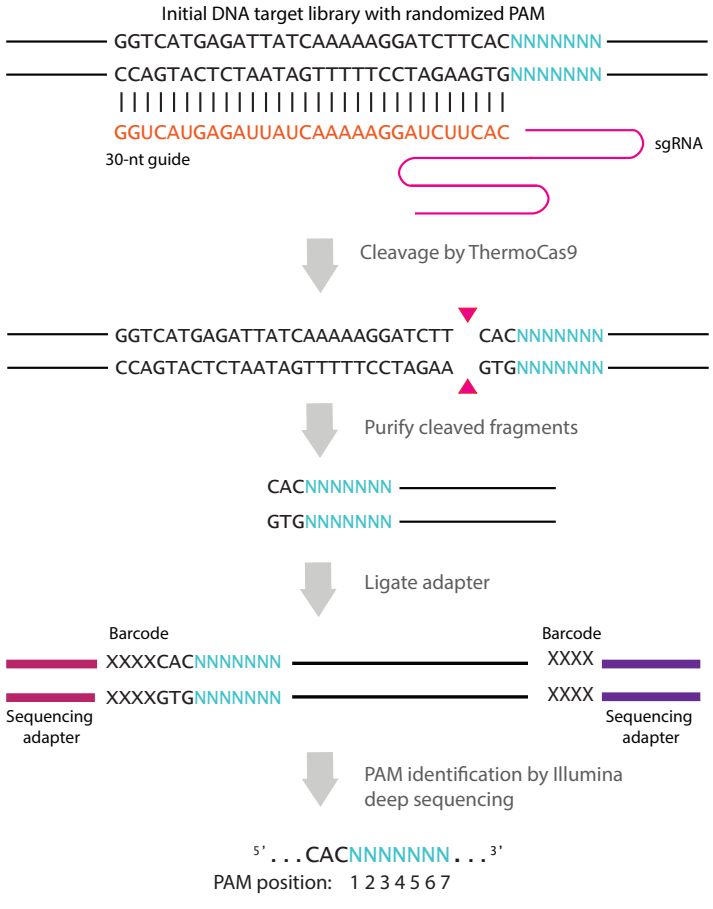
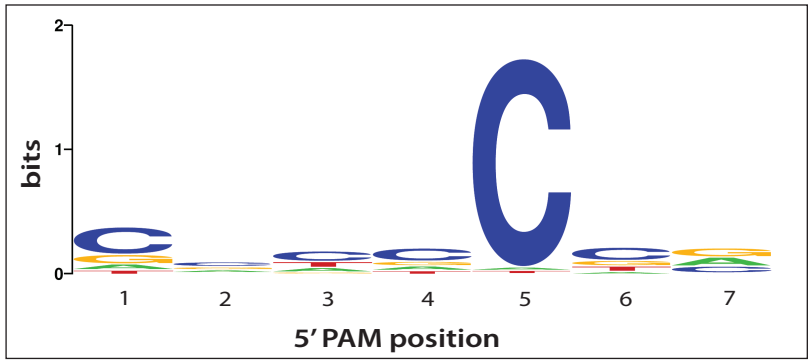

15 mins
30 mins
15 mins
30 mins
sgRNA
+ - + - + - + -
M
150 kD
ThermoCas9
100 kD
75 kD
60°C
65°C

Figure 1

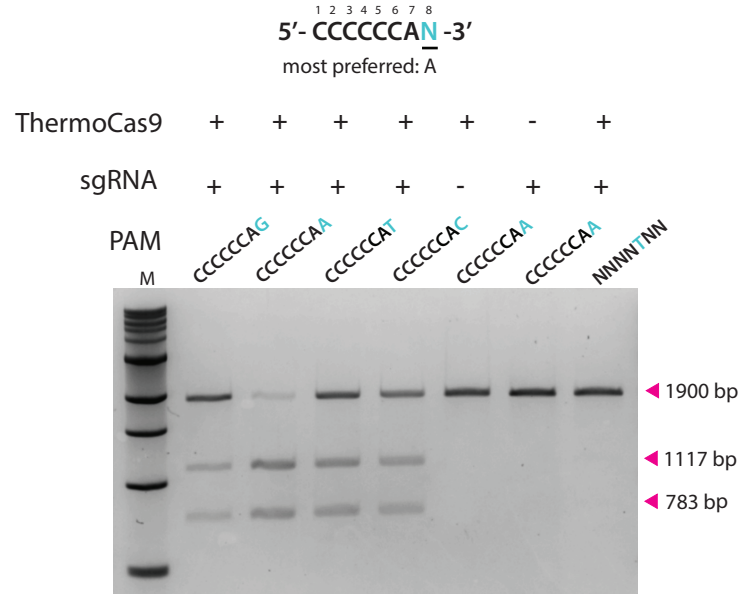
a.



b.



c.



d.

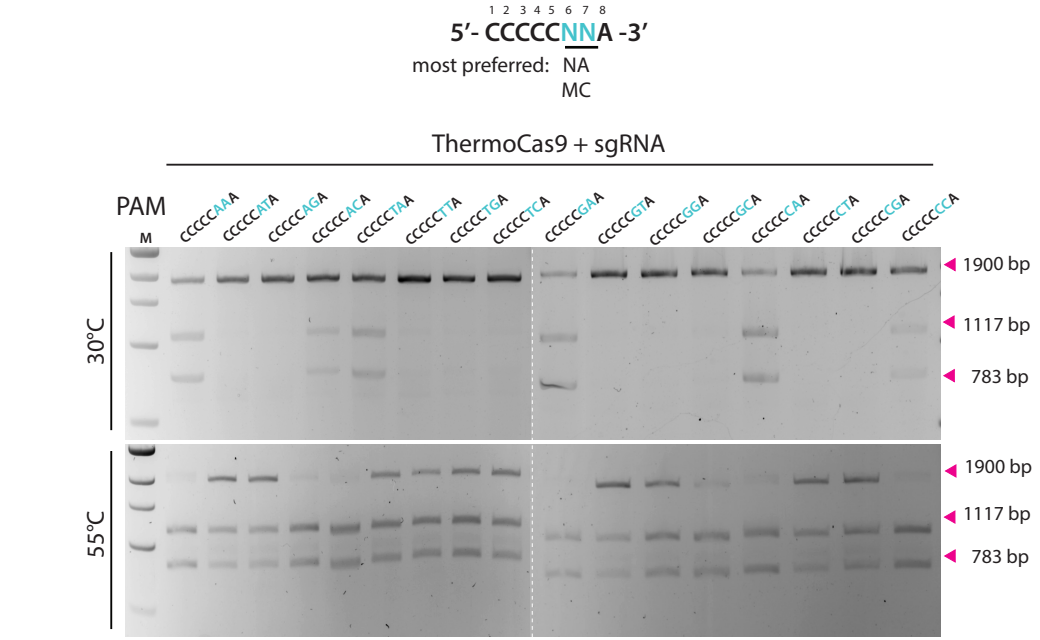


Figure 2

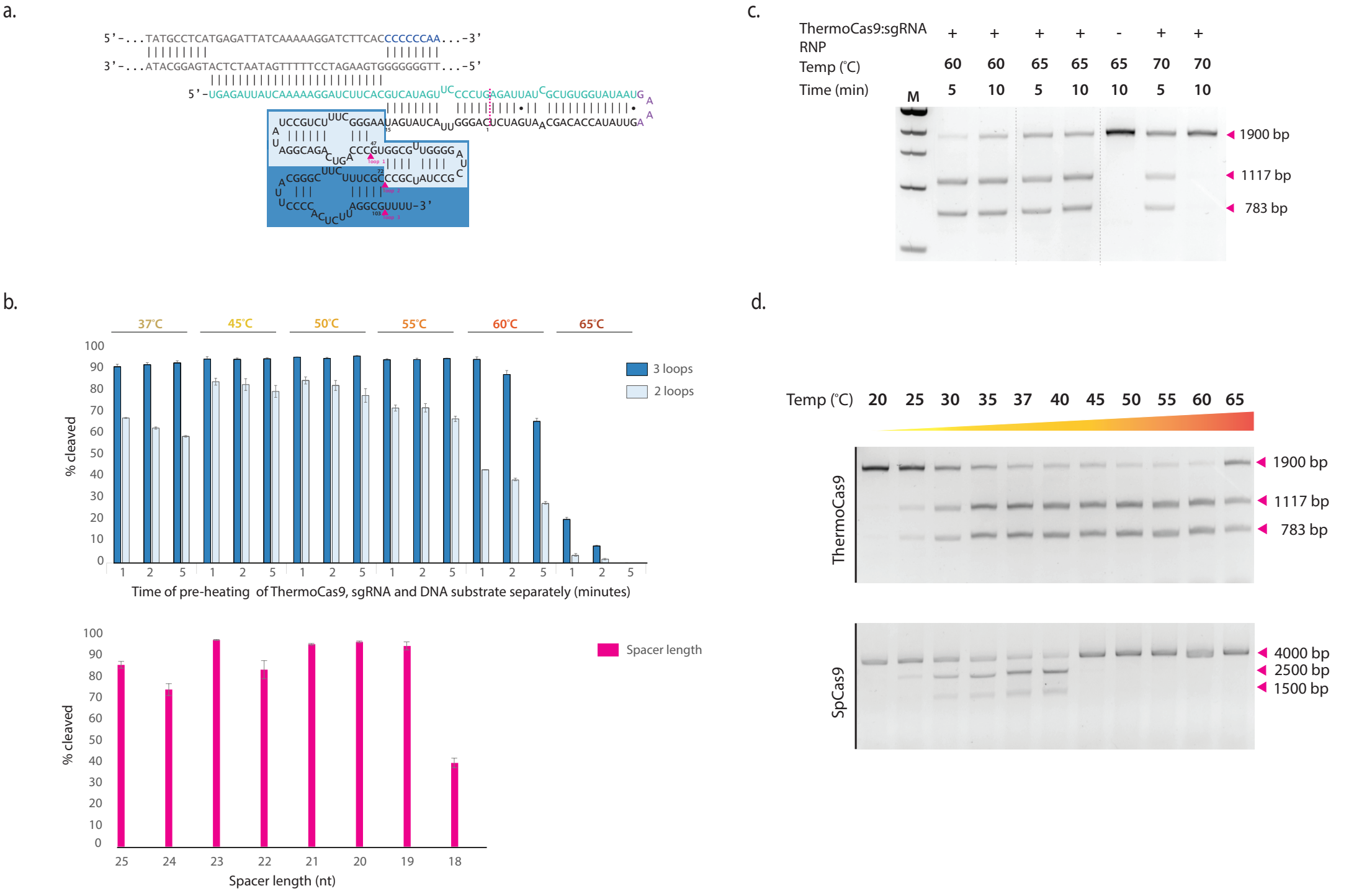

a.
5'-...TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCCCCCAA...-3'
3'-...ATACGGAGTACTCTAATAGTTTTTCCTAGAAGTGGGGGGGTT...-5'
5'-UGAGAUUAUCAAAAAGGAUCUUCACGUCAUAGUUCCCCUCAGAUUAUCGCUGUGGUAUAAUGA
UCCGUCUUUCGGGAAUAGUAUCAUUGGGACUCUAGUACGACACCAUAUUGA
AGGCAGACUGACCCGUGGCGUUGGGGAU
ACGGGCUUCUUUCGCCCGCUAUCCGC
UCCCCACUCUUAGGCGUUUU-3'
b.
37°C
45°C
50°C
55°C
60°C
65°C
% cleaved
3 loops
2 loops
Time of pre-heating of ThermoCas9, sgRNA and DNA substrate separately (minutes)
Spacer length
% cleaved
Spacer length (nt)
c.
ThermoCas9:sgRNA RNP
Temp (°C)
Time (min)
M
1900 bp
1117 bp
783 bp
d.
Temp (°C) 20 25 30 35 37 40 45 50 55 60 65
ThermoCas9
1900 bp
1117 bp
783 bp
SpCas9
4000 bp
2500 bp
1500 bp

Figure 3

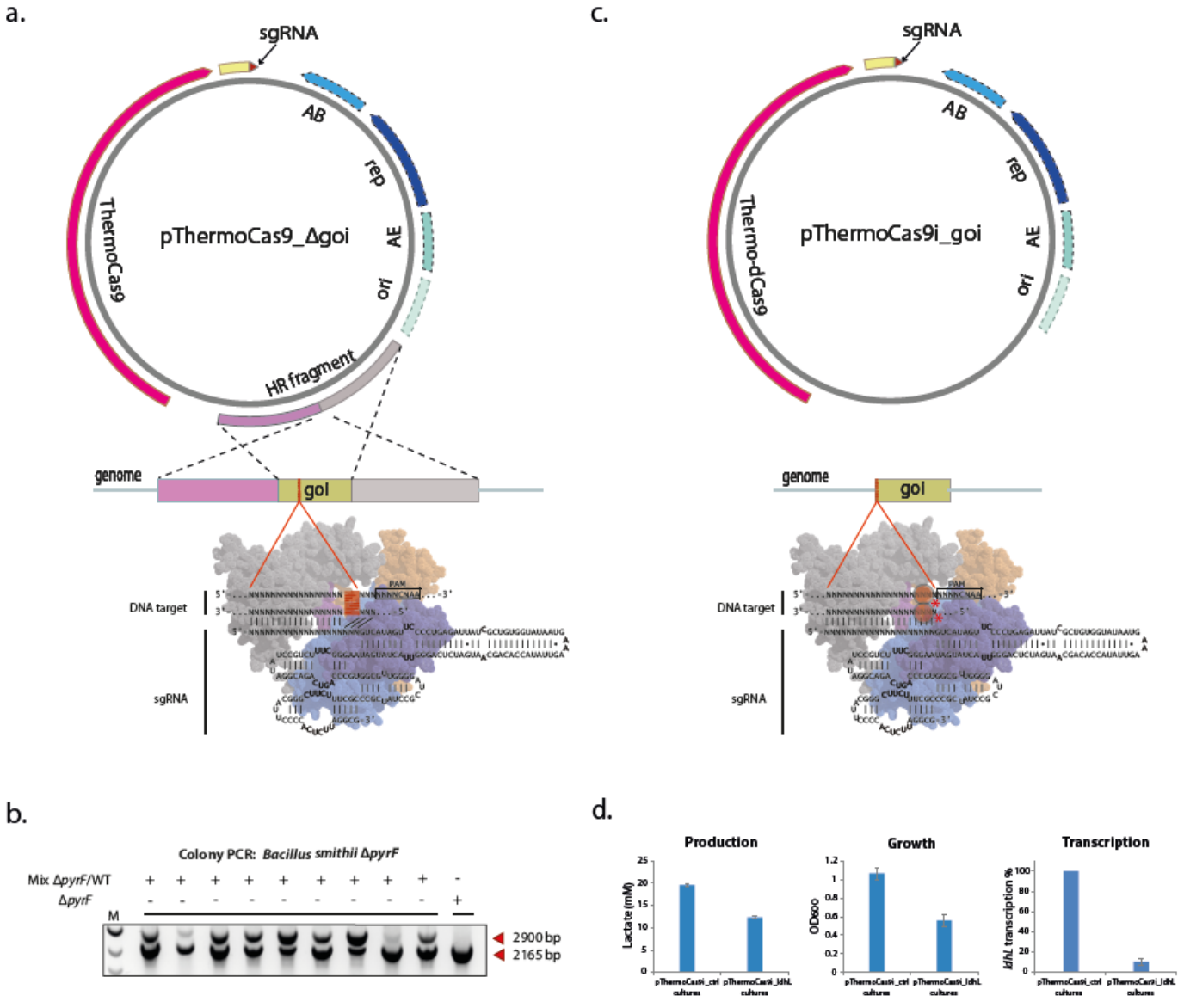

a.
sgRNA
AB
rep
AE
ori
HR fragment
ThermoCas9
pThermoCas9_Δgoi
genome
goI
PAM
DNA target
sgRNA
c.
sgRNA
AB
rep
AE
Ori
Thermo-dCas9
pThermoCas9i_goi
genome
goI
PAM
DNA target
sgRNA
b.
Colony PCR: Bacillus smithii ΔpyrF
Mix ΔpyrF/WT
ΔpyrF
M
2900 bp
2165 bp
d.
Production
Lactate (mM)
Growth
OD600
Transcription
ldhL transcription %
pThermoCas9i_ctrl cultures
pThermoCas9i_ldhL cultures

Figure 4