

Article

Not peer-reviewed version

Cloning, Expression, and Functional Analysis of the Compact *Anoxybacillus flavithermus* Cas9 Nuclease

[Anastasiya Matveeva](#) , Alexander Ryabchenko , Viktoria Petrova , Daria Prokhorova , [Evgenii Zhuravlev](#) , Alexander Zakabunin , Artem Y. Tikunov , [Grigory Stepanov](#) *

Posted Date: 31 October 2023

doi: 10.20944/preprints202310.2063.v1

Keywords: CRISPR/Cas9; genome engineering; whole-genome sequencing; PAM identification



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Cloning, Expression, and Functional Analysis of the Compact *Anoxybacillus flavithermus* Cas9 Nuclease

Anastasiya Matveeva, Alexander Ryabchenko, Viktoria Petrova, Daria Prokhorova, Evgenii Zhuravlev, Alexander Zakabunin, Artem Tikunov, and Grigory Stepanov *

Institute of Chemical Biology and Fundamental Medicine, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, 630090 Russia

* Correspondence: stepanovga@niboch.nsc.ru; Tel.: +7(383)363-51-89

Abstract: Research on Cas9 nucleases from different organisms holds great promise for advancing genome engineering and gene therapy tools as it could provide novel structural insights into CRISPR editing mechanisms, expanding its application area in biology and medicine. The current study focuses on generating a construct to express a compact Cas9 nuclease (AnoCas9) from the thermophilic microorganism *Anoxybacillus flavithermus*. Next, distinctive AnoCas9 properties are investigated. AnoCas9 gene is expressed in *E.coli* producing a polypeptide fused with the maltose-binding protein (MBP-AnoCas9). His-Tag in its structure enables purification using metal-chelate chromatography followed by the cleavage of the polypeptide by TEV protease. Bioinformatical analysis of the CRISPR array found in the genome of an *Anoxybacillus flavithermus* strain helps predict a functional PAM sequence for AnoCas9, which is supported by *in vitro* experiments. The purified protein demonstrates nuclease activity in the presence of crRNA:tracrRNA duplex in the 37-60 °C range, with maximum activity observed at 45-55 °C. The analysis of FAM-labeled dsDNA substrate cleavage has allowed us to determine the functional AnoCas9 PAM motif as 5'-NNNNCDAA-3'. Thus, AnoCas9 adds to the repertoire of thermophilic Cas9 effectors and its properties suggest application in areas requiring the presence of thermostable CRISPR/Cas systems.

Keywords: CRISPR/Cas9; genome engineering; whole-genome sequencing; PAM identification

1. Introduction

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems provide microbes with RNA-based mechanism for innate adaptive immunity against foreign genetic elements. Typically, their genomic context is comprised of the CRISPR locus containing spacer sequences interspersed with repeats, and arrays of CRISPR-associated (*cas*) genes encoding functional proteins. In the presence of foreign DNA, CRISPR/Cas integrates its short fragment into the CRISPR locus as a new spacer. Expression of the CRISPR locus with further processing leads to the formation of mature CRISPR RNAs (crRNAs). Cas proteins expressed from the *cas* array then bind to crRNA to form ribonucleoprotein complexes (RNPs). crRNA specifically targets invading nucleic acid and guides RNP in order to disrupt it [1–3]. Class 2 systems are characterized by the presence of a single RNA-bound Cas effector that recognizes and cleaves target DNA [4,5]. Their programmable nature was employed to develop novel strategies for specific genome editing using CRISPR/Cas9 (class 2 type II), CRISPR/Cas12 (class 2 type V), and, later, CRISPR/Cas13 (class 2 type VI) systems [5–12]. The expansion of versatile tools and approaches based on CRISPR/Cas gave rise to the robust evolution of genome engineering with applications in medicine, biotechnology, diagnostics, and fundamental research [13–24].

A key component of the class 2 type II CRISPR/Cas systems is a single multidomain enzyme Cas9. It functions using either the duplex of guide crRNA and trans-activating crRNA (tracrRNA) or the product of their artificial fusion in the form of a single-guide RNA (sgRNA) [5,25]. Cas9 RNP complex utilizes two nuclease sites, RuvC and HNH, to generate a double-strand break (DSB) in a target DNA sequence. DSB sites are determined through the complementarity of the target DNA to the 20-nt protospacer sequence in guide RNA (crRNA or sgRNA). The re-programming of Cas9 to

aim the cleavage of the DNA of choice, however, is limited by the requirement of a specific protospacer adjacent motif (PAM) sequence. The presence of PAM in the target DNA region is necessary both for the recognition and the initiation of cleavage. Each Cas9 effector associates with a specific PAM sequence making it a unique signature [4,5]. The most well-known and extensively used for the programmed genome editing CRISPR/Cas9 system is derived from *Streptococcus pyogenes* (SpyCas9). It has been adapted to suit multiple applications requiring the introduction of DSBs or single-strand breaks (nicks) into target DNA, or its binding [6,26–30]. SpyCas9 prefers a 5'-NGG-3' sequence as its PAM motif, thus setting restrictions for the range of possible DNA targets. Therefore, there is a need to expand the Cas9 toolkit with novel proteins exhibiting diverse PAM specificities. Moreover, the discovery and characterization of other naturally occurring CRISPR/Cas9 systems are quite promising in terms of obtaining Cas effectors with unique recognition and cleavage properties that would allow for conditional regulation of the editing process and expand the application of CRISPR in multiple areas including fundamental structural studies [31,32].

Routine analyses of large sets of sequencing data revealed that multiple prokaryotic genomes incorporate CRISPR arrays, including CRISPR/Cas9 [31,33]. Interestingly, almost 50% of the described Cas9 proteins fall into the type II-C category. This type has been presumed to display a lesser nuclease activity in comparison to the type II-A proteins, with SpyCas9 being the most prominent exemplar. Nevertheless, recent studies demonstrate a larger diversity in both the structure and properties of the type II-C proteins. This subset of Cas9 proteins is characterized by a smaller size, which is advantageous for their delivery into cells using vector-based approaches, and enhanced specificity due to the elongated PAM sequence or, in some cases, due to the greater length of the crRNA:target DNA heteroduplex [34]. The majority of organisms that have been assigned as type II-C CRISPR hosts are typically classified as extremophiles inhabiting hot springs, wastewaters, etc. The most well-known representative of this type is Cas9 derived from *Geobacillus stearothermophilus*, which has demonstrated editing activity in mammalian cells, a wide range of working temperatures, and increased thermostability alongside enhanced lifetime in human plasma [34,35].

Anoxybacillus species are frequently discovered in thermophilic conditions. As an example, *Anoxybacillus flavithermus* inhabits hot springs and is characterized as a gram-positive, endospore-forming anaerobic microorganism [36]. To this date, there are only seven whole genomes for *Anoxybacillus flavithermus* strains deposited in genome databases. Furthermore, sequence analysis revealed the presence of a common gene group playing a role in the adaptation to environmental conditions, such as growth in the alkaline environment under high temperatures. The search for CRISPR/Cas systems resulted in the discovery of genes associated with class 1 (types I and III) and class 2 (type V) for some of the strains; however, corresponding arrays have not been yet studied extensively to assess the functioning of CRISPR immunity in these microorganisms [37]. Thus, further investigations into thermophilic microorganisms hold great promise for expanding the CRISPR toolbox.

To identify putative functioning CRISPR systems, the present study analyzes the shotgun whole genome sequencing data for the set of microorganisms deposited in the Collection of extremophilic microorganisms and type cultures (Institute of Chemical Biology and Fundamental Medicine, Novosibirsk). The *Anoxybacillus flavithermus* strain obtained in the Kamchatka region has been selected since its genome contained the Cas9 ORF sequence as well as the adjacent CRISPR array. Further optimization of the protein-coding sequence enabled the successful expression of the *Anoxybacillus flavithermus* Cas9 (AnoCas9) in *E. coli*. AnoCas9 demonstrated stability and nuclease activity at high temperatures. The CRISPR array was analyzed to predict the PAM sequence, and the prediction was further supported with *in vitro* experiments based on the cleavage of fluorescently labeled dsDNA substrates. The 5'-NNNNCDAA-3' sequence was identified as the functional PAM motif. Characterization of yet another thermophilic and thermostable Cas9 holds great promise for a variety of applications, and further investigations could provide structural insights, contributing to our fundamental understanding of Cas9 binding and cleavage mechanisms.

2. Materials and Methods

2.1. Materials

Gene amplification for the subsequent cloning was conducted using Q5 polymerase (NEB). Routine amplification and clone testing were performed using the BioMaster HS-Taq PCR-Color (2×)

mix (Biolabmix LTD, Russia). Sanger sequencing of the AnoCas9 ORF was carried out using the BigDye™ Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific), with products being further analyzed using the ABI 3130XL Genetic Analyser at the SB RAS Genomics Core Facility. Oligonucleotide primers were synthesized by Biosset LTD. Table 1 represents oligonucleotide sequences.

Table 1. Amplification and sequencing primers used for cloning.

Name	Structure (5'-3')
Cas9-4025_ext_F	cgcagcgggtcacccgaagcttatgagatataaaataggattagacatag
Cas9-4025_ext_R	tggtgggtgctcgagcagaagcttatgactaattgattgtaacgaatg
1F	cgaaaacctgtattttcagggcggtaccatgcgttataaaattggttagacatc
2R	cgtaaaacaccttcacggactaaaagacggcgaatacgttctaacgatgtttg
3F	cgtcttttagtccgtgaagggtg
4R	tcagtcgaaagactgggccttcgccggggctaattcaatgactaattgattgtaacgaa tg
268F	atgcgtccggcgtaga
269R	gctagtattgctcagcgggtg
5F	agactgtcgatgaagccctg
5R	tgccgaactcagaagtgaaac
Seq_Cas9(Anoav)_1	actccatcgcagaagctgc
Seq_Cas9(Anoav)_2	gcaaatgtcttgaacaagtttatgg
Seq_Cas9(Anoav)_3	cgtcttttagtccgtgaagggtg
Seq_Cas9(Anoav)_4	ccagctgagtatttaggattcg
Seq_Cas9(Anoav)_5	ggtccaatcattcgactgtg
Seq_Cas9(Anoav)_6	cagtaccgttacaccataattcac
Seq_Cas9(Anoav)_7	acgcatcaacacgcaactgc
Seq_Cas9(Anoav)_8	ctctttgaattcttgttctctgc
Seq_Cas9(Anoav)_9	caagttcttttccatcagcacg
Seq_Cas9(Anoav)_10	tcaaaaacatcggtgcgaacg

Cloning was performed using the NEBuilder HiFi DNA Assembly Master Mix kit (NEB). Plasmid restriction was done using HindIII, XmaI, and KpnI endonucleases (SibEnzyme Ltd, Russia). Electrocompetent *E. coli* cells TOP10 or BL21 (DE3) were transformed using MicroPulser (Bio-Rad). Plasmid DNA was isolated using the Plasmid DNA Isolation Maxi kit (Biolabmix LTD, Russia). Bacterial strains and plasmids used in the study are presented in Table 2. Vector design, primer design, and sequencing analysis were conducted using the Unipro UGENE tool [38].

Table 2. Strains and plasmids used to generate the construct expressing AnoCas9.

Strains	Relevant Genotype	Source
<i>Anoxybacillus flavithermus</i>	Wild type	CEMTC (ICBFM SB RAS)
<i>E. coli</i> TOP10	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL endA1 nupG	Thermo Scientific™
<i>E. coli</i> BL21 (DE3)	F- ompT hsdSB (rB-, mB-) gal dcm (DE3)	Thermo Scientific™
<i>E. coli</i> Rosetta 2 (DE3)	F-ompT hsdSB(rB- mB-) gal dcm (DE3) pRARE2 (Cam ^R)	Novagen (Sigma-Aldrich)
Plasmids	Description	Source
pET36b(+)	PT7lac, Kan ^R	Novagen

pETm-Cas9	PT7lac, Kan ^R	This study
pUC19-210Cas9	pBR233ori-F, Amp ^R	ICBFM SB RAS
pD441-HMBP	PT5lac, Kan ^R	DNA TwoPointO Inc. («ATUM»)
pD441-HMBP-Cas9	PT5lac, Kan ^R	This Study

Chromatography was performed using the IMAC SepFast™ и SP SepFast™ 6HF columns (Biotoolomics, Great Britain). HMBP-Cas9 was cleaved using TEV-protease (Biolabmix LTD, Russia). Proteins were analyzed using electrophoresis in the Mini-Protean Tetra System (Bio-Rad) and Precision Plus Protein Dual Color Standards (Bio-Rad) as markers.

The genome region encoding tracrRNA was amplified using the BioMaster HS-Taq PCR-Color (2×) mix (Biolabmix LTD, Russia). Oligonucleotide primers were synthesized at the Laboratory of Synthetic Biology (ICBFM SB RAS). Table 3 represents oligonucleotide sequences. tracrRNA was transcribed *in vitro* using the High-yield T7 RNA synthesis kit (Biolabmix LTD, Russia). DNA template and tracrRNA were isolated using the DR Reaction mixtures DNA isolation kit (Biolabmix LTD, Russia).

Table 3. Amplification primers used to generate the dsDNA template and crRNA sequences.

Name	Structure (5'-3')
Ano_tracrRNA_T7	atgcagctaatacgactcactataggtcagggttactatgataagg
Ano_tracrRNA_R	atggggagtgcccttaaagg
Ano_ANXA_cr	cagggaugcauuuguggccauguaauaguuccccuga

The pCMV6 plasmid carrying the ANXA6 cDNA insert (pANXA6, OriGene #RC202086) was used in the Cas9 cleavage assays. The pMJ806 plasmid encoding *S. pyogenes* Cas9 (SpyCas9) carrying an N-terminal His6 tag, maltose-binding protein sequence, and TEV protease cleavage site was obtained from Addgene (#39312). SpyCas9 protein was overexpressed and purified according to a previously described protocol [39] for further usage as a reference Cas9 nuclease.

FAM-labeled double-stranded DNA (dsDNA) PAM templates were generated through routine amplification using the BioMaster HS-Taq PCR-Color (2×) mix (Biolabmix LTD, Russia). Oligonucleotide primers were synthesized at the Laboratory of Synthetic Biology (ICBFM SB RAS). Table S1 represents oligonucleotide sequences. DNA templates were isolated using the DR Reaction mixtures DNA isolation kit (Biolabmix LTD, Russia). Cas9 cleavage products were purified using the MagBlood DNA isolation kit (Biolabmix LTD, Russia).

2.2. Whole Genome Sequencing and Data Analysis

A number of bacterial strains were selected from the Collection of Extremophilic Microorganisms and Type Cultures (Institute of Chemical Biology and Fundamental Medicine, Novosibirsk) and cultivated. Bacterial DNA was isolated and purified using DNA and RNA isolation and precipitation kit (Biolabmix, Novosibirsk, Russia) and used to construct libraries with TruSeq DNA PCR-Free kit (Illumina) and TruSeq Illumina adaptors (Illumina). Synthesized libraries were sequenced on Illumina NovaSeq 6000 with NovaSeq 6000 S2 Reagent Kit v1.5 (2x150 cycles). Libraries were sequenced in biological duplicates and raw data was demultiplexed with bcl2fastq v2.20.0.422 using default parameters. Four data archives were generated for each biological sample. Filtration by quality and adapter removal were performed with Trimmomatic v.0.39. Trimmed reads were then used to assemble genome contigs using SPADes v.3.15.4. All contigs longer than 500 bp were annotated for the presence of ORFs and CRISPR arrays with the RAST server.

2.3. Gene Sources and Cloning Strategy

Anoxybacillus flavithermus strain 4025 from the Collection of extremophilic microorganisms and type cultures (Institute of Chemical Biology and Fundamental Medicine, Novosibirsk) was selected as a model microorganism expressing thermophilic Cas9. Genomic DNA was used as a gene source.

The *anoCas9* gene was amplified using Cas9-4025_ext_F/R primers (Table 1). Oligos contained an overlap with a modified pET36b(+) vector [40]. *anoCas9* was cloned into the vector pre-digested

with HindIII. Bacterial colonies containing pETm-Cas9 plasmids were selected using PCR with flanking primers 268F и 269R (Table 1), with the presence of elongated products (3566 bp) as the selection criterion.

Subsequently, the *anoCas9* gene was codon-optimized for the expression in *E.coli* using the [http://atgme.org/web tool](http://atgme.org/web_tool) [41] and the *E. coli* strain W3110 [gbtct]: 4332 CDSs as a reference genome (1372057 codons). As a result, rare codons found in the first 210 bp from the 5'-end were replaced. The optimized DNA fragment was synthesized *de novo* in the Laboratory of Synthetic Biology (ICBFM SB RAS) and cloned into the pUC19 vector. The novel 5'-fragment of *anocas9* was used to assemble a partially optimized *anoCas9* gene and clone it into the expressing vector pD441-HMBP.

anoCas9 was cloned into pD441-HMBP using KpnI and XmaI digestion sites. To amplify the optimized gene fragment (210 bp), 1F/2R primers were used (Table 1), containing an overlap with the pD441-HMBP vector and wild-type *anoCas9*. The remaining *anoCas9* fragment was amplified using 3F/4R primers (Table 1). The oligos were overlapping with the optimized 5'-fragment and pD441-HMBP. The NEBuilder kit was used to assemble three fragments – a hydrolyzed vector and two PCR products. Colonies were screened for the presence of the full-size product (3580 bp) using PCR with 5F/5R primers.

Finally, construct sequences were verified through the Sanger sequencing in the SB RAS Genomics Core Facility using Seq_Cas9(Anoav)_1-10 primers (Table 1).

2.4. *Escherichia coli* Cultivation Conditions

Genome engineering procedures involving *Escherichia coli* were conducted using the standard LB medium with an addition of kanamycin (30 µg/ml).

Biomass for protein isolation was cultivated using the corresponding vector and Rosetta 2 (DE3) as a production strain. 100 ml of cell culture was incubated with kanamycin (30 µg/ml) and chloramphenicol (25 µg/ml) overnight at 37°C, 200 rpm. The next day, 7.5 ml of inoculate was transferred into 2.5 L Ultra Yield Thomson flasks (Thomson LTD) containing 750 ml of fresh ZYM-505 media (no microelements or lactose addition) [42], added with the abovementioned antibiotics. Cells were cultivated at 37°C, 200 rpm in the Excella E25 incubator-shaker (Eppendorf) to reach the OD600 of 0.6-0.8 o.u./ml (total media volume equaled to 6 L, divided into 8 flasks). Next, 0.4 ml of 1M isopropyl β-D-1-thiogalactopyranoside (IPTG) was added into flasks to reach the final concentration of ~0.5 mM. Afterward, cells were cultivated for 18 hours at 20°C, 200 rpm. Cells were harvested via centrifugation for 20 min at ~10,000 g, 4°C (Hitachi CR22GIII, R9A rotor). The biomass was stored at -80°C until protein isolation.

2.5. *AnoCas9* Purification

2.5.1. Collection of the Cell Extract

Biomass (25±5 g) was resuspended in 250 ml of Buffer A (50 mM Tris-HCl (pH 8.0 at 25°C); 300 mM NaCl), adding PMSF to the final concentration of 1mM. Cell suspension was placed into the ice bath and sonicated for 1 min three times (with pauses for 2 min). Cell debris was pelleted for 30 min at ~32,000 g, 4°C (Hitachi CR22GIII, R19A rotor), and the supernatant was used for chromatography.

2.5.2. IMAC-Sepharose Chromatography

Immobilized metal affinity chromatography (IMAC) was used as the first step of protein purification. The supernatant was applied to the column with resin that had been equilibrated in advance with Buffer A, flowing at a speed of 1-2 ml/min. The column was then washed with 100 ml of Buffer A. The target precursor protein HMBP-AnoCas9 was eluted with a linear imidazole gradient (0-300 mM). The gradient was prepared using Buffers A and B (50 mM Tris-HCl (pH 8.0 at 25°C); 300 mM NaCl, 300 mM imidazole) so that the volume ratio of Buffer B increased from 0% to 100% within three volumes of resin. Collected fractions were analyzed for the presence of the HMBP-Cas9 precursor protein. Next, the purest positive fractions were combined and the A280 adsorption was measured to assess protein concentration. The HMBP-AnoCas9 protein was treated with TEV-protease at ~1,000 u per 50 mg of protein ratio. The reaction mixture was transferred into a dialysis bag to perform dialysis at 4°C overnight (buffer contained 20 mM Tris-HCl (pH 8.0 at 25°C), 100 mM

NaCl, 1 mM EDTA, 1 mM dithiothreitol, 10% glycerol). The next day, the reaction mixture was centrifuged for 30 min at ~32,000 g, 4°C, and the supernatant was used for SP-Sepharose purification.

2.5.3. SP-Sepharose Chromatography

The supernatant was applied to the resin column (volume 15 ml) pre-equilibrated with Buffer C (20 mM Tris-HCl (pH 8.0 at 25°C), 100 mM NaCl) at a 1-2 ml/min speed. The column was subsequently washed with 80 ml of Buffer A. The target AnoCas9 was eluted with a sodium chloride linear gradient (100-1000 mM). The gradient was prepared using Buffers C and D (20 mM Tris-HCl (pH 8.0 at 25°C); 1 M NaCl) so that the volume ratio of Buffer D increased from 0% to 100% within three volumes of the resin. Collected fractions were analyzed for the presence of the AnoCas9 protein. The purest positive fractions were combined and concentrated at 4,000 rpm, 4°C, to the final volume of 0.5-1.0 ml using the Eppendorf Centrifuge 5810R. 20 ml of the conservation buffer (20 mM Tris-HCl (pH 7.4 at 25°C), 600 mM NaCl, 0.2 mM EDTA, 2 mM dithiothreitol) was added and the concentration procedure was repeated. Finally, the concentrate was transferred into a clean Eppendorf tube, and an equal volume of glycerol was added dropwise. For the final solution, electrophoretic purity was determined, OD280 measured, and AnoCas9 concentration assessed (1 o.u. was considered equal to 1.17 mg/ml or 9.2 pmol/μl). The protein was stored at -20°C in the final solution containing 10 mM Tris-HCl (pH 7.4 at 25°C), 300 mM NaCl, 0.1 mM EDTA, 1 mM dithiothreitol, and 50% glycerol. Proteins and cell lysates were visualized by 10-12% denaturing SDS-PAGE with Coomassie Brilliant Blue G-250 staining.

2.6. CRISPR Array Analysis

AnoCas9 nucleotide and amino acid sequences were analyzed using BLAST alignment tools. Whole genome sequencing data was analyzed using the RAST server to determine the corresponding CRISPR array. Spacer sequences were annotated and further used to predict the PAM sequence using the CRISPRTarget web tool [43]. Spacer sequences were aligned to the data from Genbank-Phage, RefSeq-plasmid, and IMGVR databases. The highest-scoring alignments were further analyzed to determine a set of putative PAM motifs. Data were combined using WebLogo [44].

2.7. CRISPR RNA Synthesis

Genomic DNA was used as a template to amplify the tracrRNA-encoding region. A short dsDNA template was isolated and used subsequently to generate tracrRNA through *in vitro* transcription. CRISPR RNAs (crRNAs) with varying spacer sequences were synthesized *de novo* in the Laboratory of Synthetic Biology (ICBFM SB RAS).

2.8. In Vitro DNA Plasmid Cleavage with Cas9

In vitro cleavage assays were performed as described [45]. Prior to the experiment, crRNA was hybridized with tracrRNA (2 μM each) in deionized water at 95°C for 30 s followed by 3 min on ice. SpyCas9 or AnoCas9 ribonucleoprotein (RNP) complex was assembled (final concentrations: 0.4 μM Cas9 (Sp or Ano), 2 μM tracrRNA and 2 μM crRNA) in a reaction buffer (20 mM HEPES pH 7.5, 100 mM KCl, 5% glycerol, 1 mM DTT, 0.5 mM EDTA, 2 mM MgCl₂) at 37°C for 20 min. The target pANXA6 plasmid was cleaved in the presence of a 50-fold excess of Cas9/crRNA:tracrRNA and a 1:5 molar ratio of Cas9 to the crRNA:tracrRNA complex. Samples were incubated at the indicated temperatures (37–75 °C) for 1 hour and then quenched by adding 10 μL of proteinase K (60 mM EDTA, 4 M urea, 0.4 mg/ml proteinase K) and incubating at 37°C for 15 min. Products were resolved by the electrophoresis in 1% agarose gel, stained with 0.5 μg/ml ethidium bromide, and visualized using a UV imager.

2.9. In Vitro PAM Library Digestion with Cas9

Rapid identification of functional PAM motifs for AnoCas9 was performed through the cleavage of a fluorescently labeled dsDNA fragment library. The method adapts the standard approach using radiolabeled dsDNA substrates [46]. A set of oligonucleotides, containing sequences complementary to the spacer and variable nucleotides at different positions in the PAM motif, was synthesized in the Laboratory of Synthetic Biology ICBFM SB RAS (referred to as Spacer-PAM-RN, Table S1). The single-

stranded template was turned into dsDNA using the second universal oligonucleotide carrying the spacer sequence at its 3'-end and thus complementary to the Spacer-PAM-RN 3'-end (Uni-Spacer-F, Table S1). Partial oligonucleotide duplexes were elongated through routine three-step amplification. PCR products were then utilized as templates to generate elongated FAM-labeled dsDNA substrates using a Uni-Spacer-F oligo and a universal 5'-FAM-labeled oligonucleotide extending the generated dsDNA template downstream PAM motif (FAM-Uni-R, Table S1). Three-step amplification allowed us to produce the final library of dsDNA substrates containing the upstream extension, spacer sequence, PAM motif of choice, downstream extension, and FAM label. DNA templates were isolated, OD₂₆₀ was measured, and dsDNA concentration was assessed.

In vitro cleavage assays were performed as described in 2.8. Briefly, the cleavage of the target dsDNA template was carried out in the presence of a 25-fold excess of Cas9/crRNA:tracrRNA and a 1:5 molar ratio of Cas9 to crRNA:tracrRNA complex. The samples were incubated at 50°C for 1 hour and then quenched by adding proteinase K. The fragments were purified and further resolved using the Nanophore 05 gene analyzer (Syntol, Russia). The ratio of cleaved DNA to the total dsDNA amount was assessed and used to quantify editing efficiency.

3. Results

3.1. Identification of *Anoxybacillus Flavithermus* Cas9

Twenty-one bacteria strains were selected from the Collection of Extremophilic Microorganisms and Type Cultures (Institute of Chemical Biology and Fundamental Medicine, Novosibirsk). The strains were cultivated and genomic DNA was isolated for further sequencing. After the analysis of RAST annotations, eight samples were found to contain sequences encoding various Cas proteins with only seven out of them containing full-sized CRISPR-Cas arrays consisting of *cas* genes and spacer-repeat units. Further analysis informed the selection of the *Anoxybacillus flavithermus* strain 4025 as the Cas9 ORF (further referred to as *anoCas9*) was found and annotated in its genome sequencing data.

The *Anoxybacillus flavithermus* Cas9 ORF is 3264-nt long and encodes a protein of 1087 aa in length. BLASTn analysis using the whole-genome shotgun contigs collection revealed a single fully matching nucleotide sequence that had already been annotated as a Cas9 nuclease. It was found through a similar bioinformatical analysis of the genome of another *Anoxybacillus flavithermus* strain (GenBank WP_099668921.1). BLASTp analysis of *AnoCas9* demonstrated its high homology with a few other *Anoxybacillus* Cas9, as well as with other species (81.86-97.33%). Interestingly, all of these proteins were characterized by a smaller size than commonly used SpyCas9 (Figure 1A). *AnoCas9* was also the same length as *Geobacillus stearothermophilus* Cas9 (*GeoCas9*), one of the most prominent and well-described thermostable Cas effectors [35]. Domain structure of *GeoCas9* was established previously, with the REC domain being considerably smaller compared to SpyCas9 (Figure 1B). Although high sequence homology between *AnoCas9* and *GeoCas9* suggested similar properties, *in vitro* experiments were further required to properly characterize the *AnoCas9* effector.

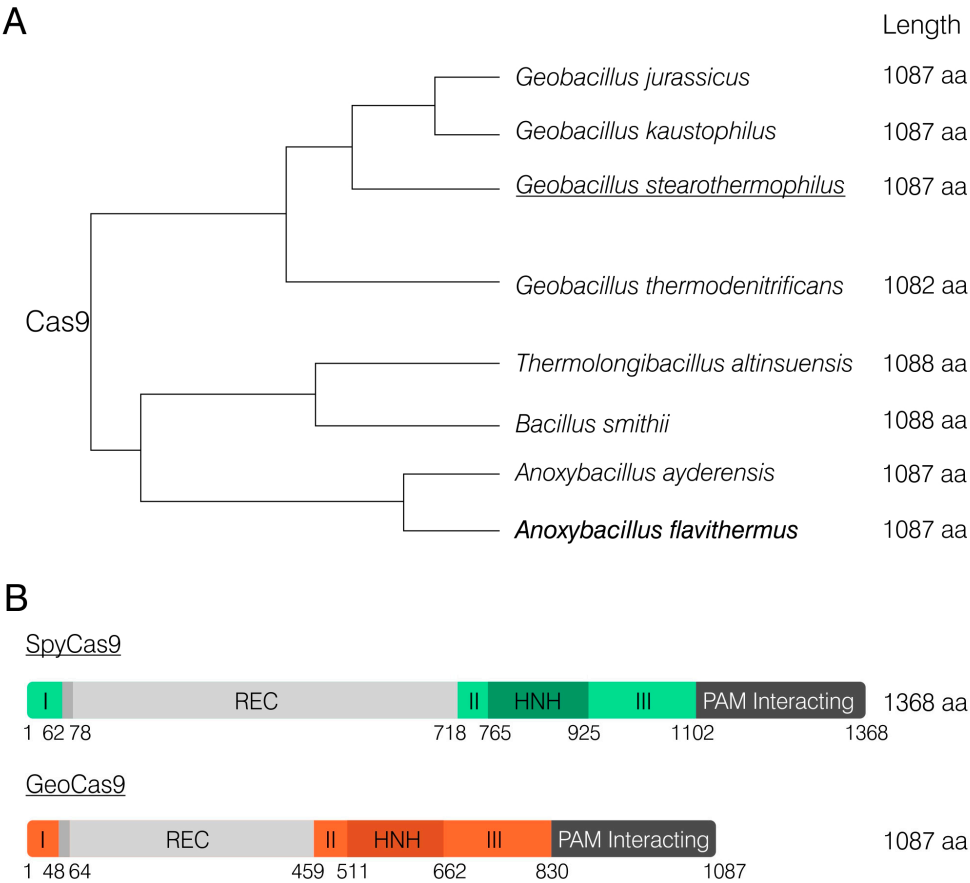


Figure 1. Schematical representation of the AnoCas9 BLASTp alignment analysis (A). Comparison of SpyCas9 and GeoCas9 domain structures as previously described [35] (B).

3.2. Obtaining AnoCas9 Recombinant Protein

3.2.1. Cloning anoCas9 Gene into pETm Vector

The *anoCas9* ORF was first cloned into a pETm vector – a variant of pET36b(+) in which the original ORF between start- and stop-codons is replaced with the modified ORF that contains restriction sites and DNA fragment encoding 8 histidine amino acid residues required for further IMAC purification procedures [40]. The *anoCas9* gene was amplified from the *Anoxybacillus flavithermus* genomic DNA using a set of primers overlapping with pETm, pre-cleaved by HindIII. Routine Gibson assembly was then carried out followed by the transformation of electrocompetent *E. coli* cells. The colonies containing pETm with *anoCas9* insertion (pETm-AnoCas9) were selected and used to inoculate LB media for overnight incubation and subsequent plasmid DNA isolation. Plasmid DNA was then Sanger sequenced to verify the insertion sequence (Figure S1).

anoCas9 codons were analyzed using the ATGme tool for the expression in *E. coli* cells. Its ORF was found to contain a lot of rare *E. coli* codons, and thus Rosetta 2 (DE3) was selected for the expression as a more prolific strain. Sanger-verified pETm-AnoCas9 vector was used to transform Rosetta 2 (DE3) cells. Colonies were screened and tested for protein expression, with the expected molecular weight of AnoCas9-8xHis being ~130.5 kDa. However, for 20 clones, no significant difference was observed between samples before and after the induction with IPTG, and there was no band corresponding to the expected recombinant AnoCas9. We suggest that the lack of expression might be caused by the presence of rare *E. coli* codons in the wild-type *anoCas9* ORF.

3.2.2. 5'-end AnoCas9 Codon Optimization

Sequence analysis demonstrated the presence of multiple rare *E. coli* codons in the 5'-end region of *anoCas9* ORF. Moreover, rare codons were found in repeats, potentially causing premature termination of protein elongation. To increase the recombinant AnoCas9 expression efficiency in *E. coli*, the first 70 codons were optimized using the ATGme tool (Figure 2).



Figure 2. The structure of the first 70 *anoCas9* codons before (A) and after the optimization (B). Rare codons (frequency from 5 to 10) are highlighted in orange and very rare codons (frequency less than 5) are in red.

The optimized DNA fragment was synthesized *de novo* and used to assemble a full-length optimized *anoCas9*. *anoCas9* gene was also fused with the maltose-binding protein (MBP) gene to increase the efficiency of chimeric recombinant polypeptide MBP-AnoCas9 synthesis.

3.2.3. Cloning Partially Optimized AnoCas9 Gene into pD441-HMBP and Producing Recombinant AnoCas9

To generate the construct encoding the fused *MBP-anoCas9* gene, the pD441-HMBP vector was selected. The vector contained a DNA fragment encoding 6xHis at the 5'-end of the MBP gene (HMBP-tag) and a DNA fragment downstream the MBP gene encoding TEVp protease that enabled the cleavage of the purified protein in order to obtain HMBP and AnoCas9.

Routine Gibson assembly was carried out using three fragments: two amplicons and a digested pD441-HMBP vector (section 2.3 Materials and Methods). A ligation mixture was then used to transform electrocompetent TOP10 *E. coli* cells. PCR screening with the flanking primers allowed us to select colonies carrying the pD441-HMBP-AnoCas9 vector (Figure S2). Clones of choice were further cultivated to isolate plasmid DNA followed up with Sanger sequencing to verify the structure of the DNA fragment encoding the optimized *anoCas9* gene (Figure S3).

The generated pD441-HMBP-AnoCas9 vector was used to transform Rosetta 2 (DE3) cells. Clones were selected with PCR screening and their ability to produce the protein was assessed. The molecular weight (Mw) of the recombinant chimeric HMBP-AnoCas9 polypeptide was expected to be ~173 kDa. IPTG induction resulted in the formation of a band with mobility corresponding to the expected Mw, thus indicating the presence of the target HMBP-AnoCas9 polypeptide.

Selected clones were then used to grow the biomass for the recombinant protein production, as described in section 2.4, and to purify the obtained recombinant protein, as described in section 2.5 (Materials and Methods). After the first purification stage using the IMAC Sepharose, a fused HMBP-AnoCas9 polypeptide was obtained (Figure 3).

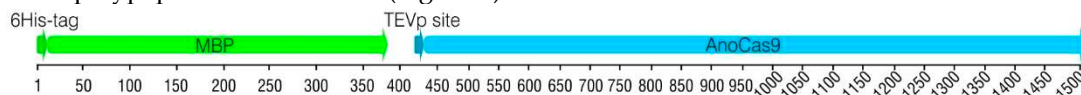


Figure 3. Schematic representation of the HMBP-AnoCas9 structure.

The polypeptide was then cleaved with TEV protease to separate the HMBP-tag and AnoCas9 protein, with the latter being further purified using SP Sepharose chromatography and concentrated as described in 2.5.3 to produce the final solution (Figure 4).

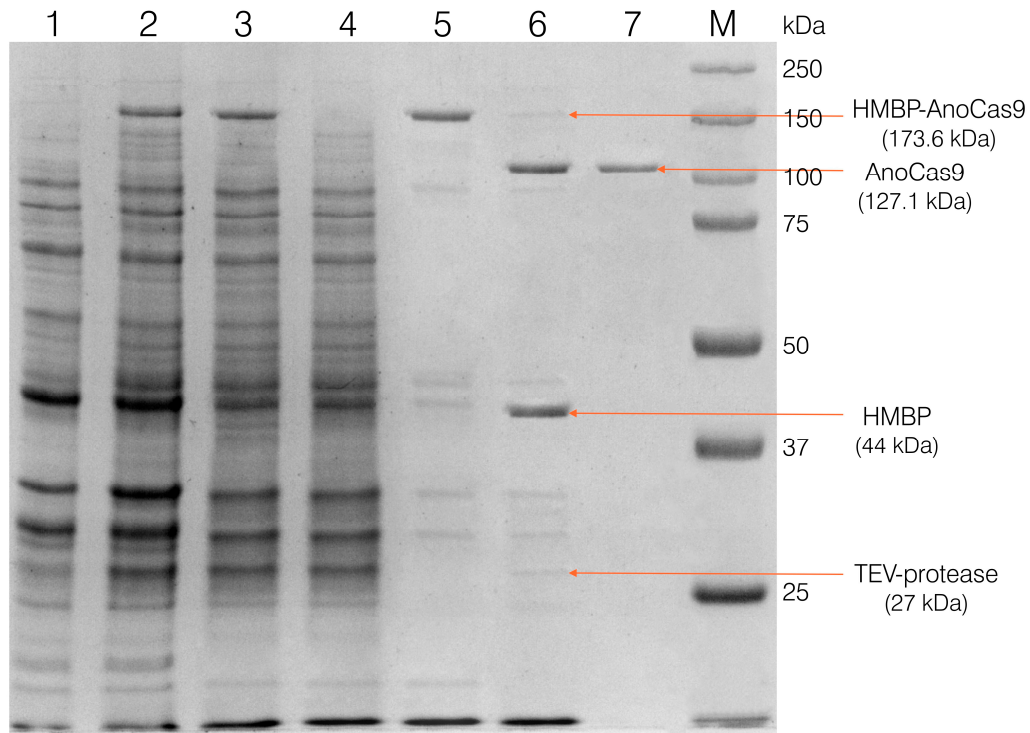


Figure 4. Analysis of the fractions at different purification stages in 10% SDS-PAGE. Lane 1 refers to cells with no IPTG induction; lane 2 – to cells induced with IPTG (biomass for the protein isolation); lane 3 – cell extract (supernatant after sonication and centrifugation); lane 4 – the unretained fraction after IMAC chromatography; lane 5 – purified HMBP-AnoCas9 after the IMAC; lane 6 – reaction mixture after incubation with TEV protease; lane 7 – purified AnoCas9 after SP Sepharose chromatography (final solution).

3.3. CRISPR Array Analysis

To assess the biological activity of the produced AnoCas9 protein, its guide RNA structure had to be determined as well as a functional PAM sequence crucial for Cas9 nuclease activity. Whole genome sequencing data for the *Anoxybacillus flavithermus* strain of choice was analyzed using the RAST and CRISPR-Finder tools to identify the CRISPR array associated with AnoCas9. The array consisted of 16 spacers and repeats (Table S2).

The domain structure of GeoCas9 was previously established, with the PAM-interacting (PI) domain being located in the C-terminal region (831-1087 aa) [35]. Alignment of AnoCas9 and GeoCas9 peptide sequences revealed 52 mismatches in the PI domain, 25 of which were non-conservative (Figure 5A). This discovery led to the assumption that AnoCas9 might require a PAM motif different from the GeoCas9 motif (5'-NNNNCRAA-3', where N is any nucleotide). While multiple approaches are used nowadays to identify the PAM, most of them rely on *in vitro* experiments, often including the usage of libraries and a subsequent NGS analysis. The CRISPRTarget tool, on the other hand, helps predict the PAM through the bioinformatical analysis of spacer sequences. The analysis was carried out using the standard parameters to align spacer sequences with the data from Genbank-Phage, RefSeq-plasmid, and IMGVR databases. The 20 most reliable alignments were generated, and all of them used data from IMGVR. Eight targets with >90% match were selected; however, they referred to two spacers only – spacer 3 and spacer 13 (Figure S4). The analysis of the 3'-adjacent region resulted in the selection of putative PAM sequences (Figure 5B). Subsequent analysis using WebLogo allowed us to predict the overall PAM motif – 5'-NNRNCNCRAN-3'. Interestingly, this sequence aligned with the GeoCas9 motif 5'-NNNNCNA-3' (Figure 5C).

Given the high similarity observed between AnoCas9 and GeoCas9, we hypothesized that guide RNA structure might also bear a resemblance. The sequence of tracrRNA for the GeoCas9 nuclease found previously [35] was aligned to the whole-genome sequencing data for the *Anoxybacillus flavithermus* strain using BLASTn. As a result, a homologous sequence with 78% similarity was observed in the CRISPR array vicinity and was further used as the AnoCas9 tracrRNA sequence (Figure 6A). This sequence (93 nt) was found to be shorter than GeoCas9 tracrRNA (98 nt). By analyzing tracrRNA, spacer sequences, and GeoCas9 sgRNA structure described in previously published data, we predicted the structure of crRNA for AnoCas9 with the 21-nt guide region (Figure 6B). While experiments involving high-throughput sequencing of small RNA fraction could be carried out in the future to support this prediction, for the purposes of the current study, the predicted crRNA and tracrRNA structures (Figure 6C) were synthesized and used to assess AnoCas9 properties *in vitro*.

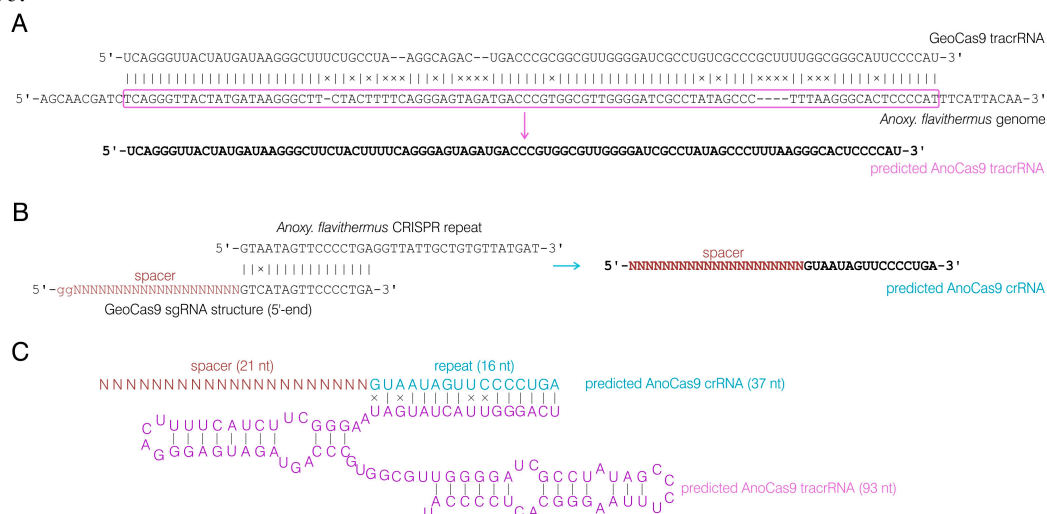


Figure 6. Analysis of the AnoCas9 CRISPR array. GeoCas9 tracrRNA and *A. flavithermus* whole-genome shotgun sequencing alignment and predicted AnoCas9 tracrRNA structure (A). *A. flavithermus* CRISPR repeat alignment with the GeoCas9 sgRNA and predicted AnoCas9 crRNA structure (B). Overall structure of the functional AnoCas9 crRNA:tracrRNA duplex (C).

3.4. Recombinant AnoCas9 Nuclease Activity Assessment

To assess the cleavage activity of AnoCas9, we used SpyCas9 as the control nuclease and the plasmid substrate pANXA6 from our previous study [45]. 5'-TGTTCAAA-3' PAM motif fitting the described GeoCas9 motif (5'-NNNNCRAA-3') was selected to construct crRNA. This design implied that sequence and structural homology between AnoCas9 and GeoCas9 would result in similar activity. We tested the efficiency of SpyCas9 and AnoCas9 RNPs within a wide temperature range (37-75°C, Figure 7). The reaction was carried out on the target plasmid in the presence of a 50-fold excess of Cas9/crRNA:tracrRNA and stopped after 60 min. SpyCas9 was most active at 37 and 40°C, and the percentage of cleaved plasmid reached approximately 100% (Figure 7A). At the same time, an increase in temperature led to a marked slowdown of the plasmid cleavage, which is consistent with previous data [35]. In contrast, AnoCas9 retained its nuclease activity at temperatures up to 60°C (Figure 7B). Interestingly, the percentage of cleaved plasmid notably increased at temperatures in the 45-55 °C range, with AnoCas9 activity reduced to a low level at temperatures above 65°C. Thus, AnoCas9 represents another thermostable Cas9 ortholog.

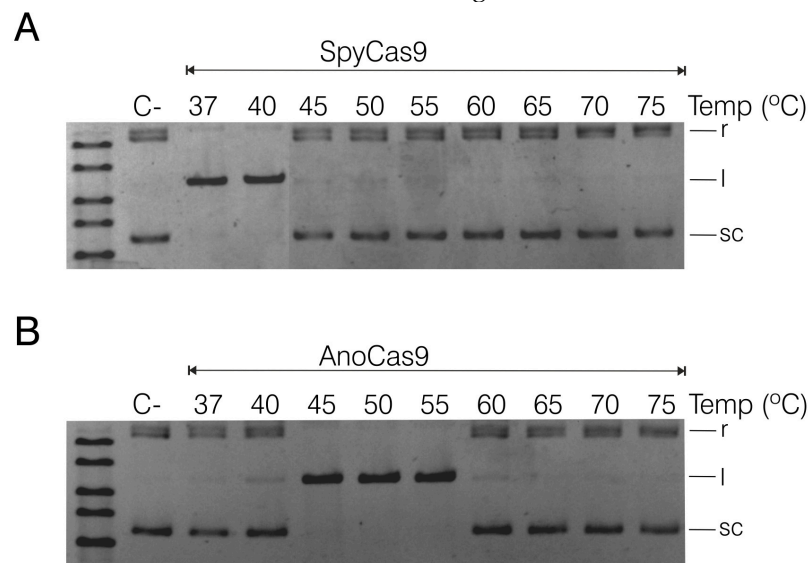


Figure 7. Assessment of SpyCas9 (A) and AnoCas9 (B) activity. Reaction mixtures were resolved in the 1% agarose gel after incubation at the indicated temperature and quenching. Plasmid DNA without Cas9 treatment (C-) was used as a negative control. The top two bands represent the cleavage products (relaxed [r] and linear [l] form), while the bottom band represents the substrate (supercoiled [sc] form). The molar ratio of Cas9 RNP to target plasmid was 50:1, and the cleavage reactions were stopped after 60 min.

3.5. In Vitro PAM Analysis

While we successfully demonstrated the nuclease activity of AnoCas9 under mesophilic conditions, the question of PAM preferences was still left for discussion. In order to support CRISPRTarget findings, we constructed an array of FAM-labeled dsDNA templates with the spacer sequence used in the previous step as a target in the pANXA6 plasmid. The PAM motif 5'-TGTTCAAA-3', present in pANXA6 and used for AnoCas9-mediated cleavage, was chosen as a template sequence. Given that CRISPRTarget analysis generated 5'-NNRNCRAAN-3' as the general structure for the AnoCas9 CRISPR array, cytidine in position 5 and adenosine in position 7 were further locked and other positions were randomized in pairs (Table S3).

Templates were synthesized through two consequent heteroduplex elongation and amplification reactions as described in section 2.9 (Materials and Methods). AnoCas9 RNP complexes were pre-formed and incubated with dsDNA substrates at 50°C for 1 h, as it was considered an optimal temperature. The reaction was then stopped and resolved using a gene analyzer. Uncleaved substrate length was expected to be 126 bp, while AnoCas9 cleavage led to the formation of a FAM-labeled product with the expected length of 84 bp (Figure 8A). The area ratio, cut to total (cut+uncut), was calculated to assess AnoCas9 editing efficiency (Figure 8B). The analysis revealed PAM preference for 5'-NNNNCDAA-3', thus supporting the bioinformatical prediction described above and indicating that AnoCas9 belongs to the GeoCas9 subclass of thermophilic Cas9 nucleases.

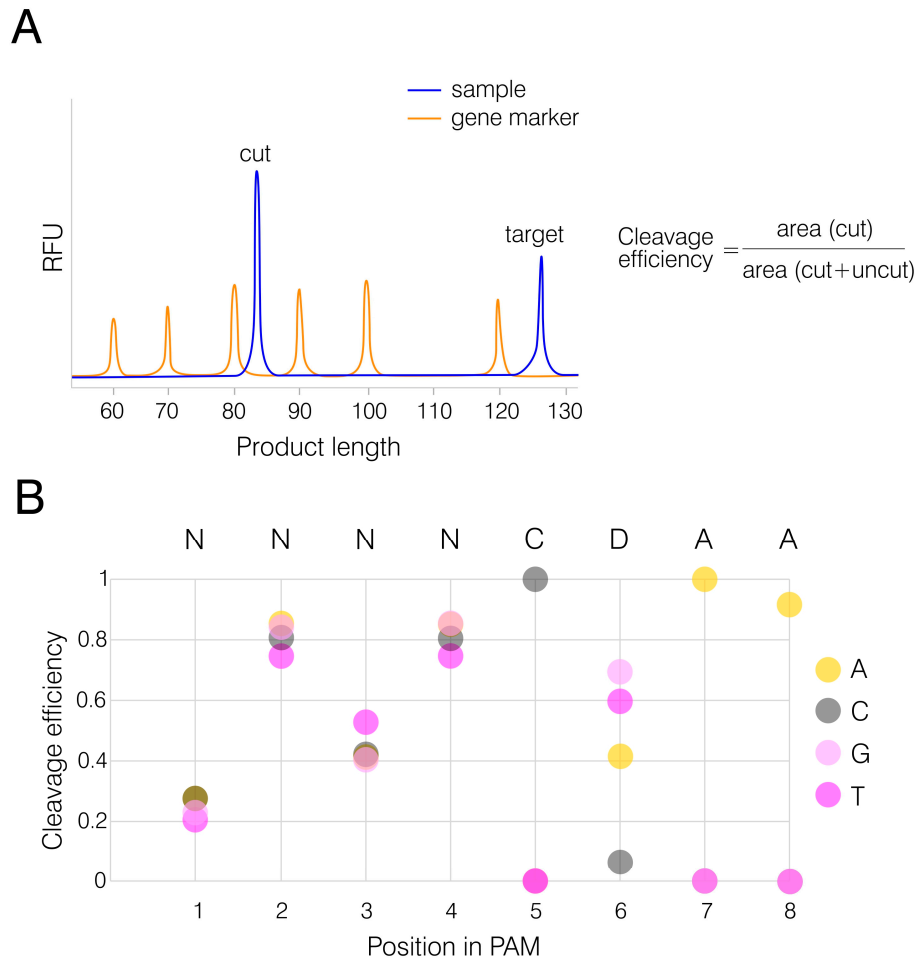


Figure 8. PAM determination. Electropherogram demonstrating the cleavage of target dsDNA substrate containing a specific PAM sequence (A). FAM-labeled dsDNA (“target”) was digested by Cas9. The presence of a functional PAM in the substrate structure resulted in the formation of short fragments (“cut”), with areas of peaks referring to the “cut” and “uncut” products measured to assess the cleavage efficiency. Analysis of PAM preferences for AnoCas9 based on the previously established TGTTCAAA sequence and the CRISPRTarget prediction (B). Substrates with variable nucleotides in different positions were tested. The average cleavage efficiency was assessed for each nucleotide position in pairs, except for locked positions 5 and 7.

4. Discussion

The results obtained demonstrate the reconstruction and characterization of yet another compact thermophilic Cas9 ortholog, AnoCas9. Although it shows high similarity to GeoCas9, there are also differences such as temperature range, which is slightly wider for GeoCas9, and the PAM sequence, which is slightly broader for AnoCas9. This distinguishes AnoCas9 as an individual nuclease, thereby adding to the repertoire of thermostable and thermophilic Cas effectors.

Currently, researchers are still exploring a pool of thermostable Cas9 nucleases, with novel proteins being discovered and characterized. However, this subclass of Cas effectors has already found its application in editing genomes of thermophilic microorganisms and fungi for fundamental research and biotechnology [47–50]. Aside from GeoCas9, a few examples include Cas9 from *Acidothermus cellulolyticus* (AceCas9), *Geobacillus thermodenitrificans* T12 (ThermoCas9), *Ignavibacteriae phylum* (IgnaviCas9), and others. AceCas9 is used for gene editing in model thermophilic bacteria for biofuel processing. This nuclease is also found to be sensitive to the methylation state of the first but not the second cytosine base, making it a potential epigenomic detector [51–53]. ThermoCas9 has demonstrated its editing activity in model thermophilic organisms and human cells [50,54–56]. IgnaviCas9 retains its activity at temperatures up to 100°C which is proposed to be leveraged for the depletion of molecules derived from 16S rRNA in bacterial RNA-Seq libraries [57]. Further research and discoveries of novel thermophilic Cas9 orthologs will broaden the existing collection, thus

expanding the application of the CRISPR-based tools (nickases, base editors, etc.) to multiple areas. Given its working temperature range, AnoCas9, for example, could potentially be employed for allele-specific isothermal amplification as a helper nuclease cleaving the non-target allele template to prevent nonspecific amplification [58–60].

PAM recognition is a key stage for Cas9 activity, which determines targets available for CRISPR-mediated editing. Currently, studies on obtaining Cas9 variants with varying PAM preferences typically employ site-directed mutagenesis [61–64]. The characterization of abundant naturally occurring Cas9 nucleases, on the other hand, is beneficial not only for the expansion of the CRISPR toolkit but also for the structural studies on mechanisms of PAM recognition and cleavage initiation [65–68]. A few discovered thermophilic Cas9 proteins demonstrate varying stringency in PAM preferences (Table 4). AnoCas9 nuclease fits into this group as one of the smallest Cas9 effectors compared to SpyCas9 (1368 aa). It also demonstrates the highest similarity in PAM preference to ThermoCas9 and GeoCas9, which is consistent with the close relation of the *Anoxybacillus* genus to *Geobacillus*. Further investigations on thermophilic microorganisms in order to obtain novel thermostable Cas9 variants will be helpful for the development of a site-directed mutagenesis strategy to generate an effective thermostable Cas9 protein with relaxed PAM preferences or even PAMless editing ability.

Table 4. Most prominent thermophilic Cas9 nucleases, their sources, and PAM preferences.

Name	Source	Size, aa	PAM sequence (5'-3')	Reference
ThermoCas9	<i>Geobacillus thermodenitrificans</i> T12	1082	NNNNCVAA NNNNCCCA	[54]
GeoCas9	<i>Geobacillus stearothermophilus</i>	1087	NNNNCRAA	[35]
CaldoCas9	<i>Geobacillus</i> sp. LC300	1087	NNNNGNMA	[69]
AceCas9	<i>Acidothermus cellulolyticus</i>	1138	NNNCC	[51,52]
IgnaviCas9	<i>Ignavibacteriae</i> phylum	1241	NRRNAT	[57]
AtCas9	<i>Alicyclobacillus tengchongensis</i>	1147	NNNNCNNN NNNNRNNA	[70]
AnoCas9	<i>Anoxybacillus flavithermus</i>	1087	NNNNCDAA	

5. Conclusions

The current study presents the procedure for reconstruction and expression of a thermophilic Cas9 ortholog from *Anoxybacillus flavithermus* in *E.coli* cells. The AnoCas9 protein was isolated and then functionally characterized through *in vitro* experiments. AnoCas9 exhibited nuclease activity at temperatures up to 60°C and PAM preference of 5'-NNNNCDAA-3', fitting into the family of thermophilic *Geobacillus* Cas9 variants. This nuclease might find its place in the CRISPR toolkit for editing the genomes of thermophilic organisms, as well as in diagnostic approaches requiring thermostable Cas9 effectors.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Author Contributions: Conceptualization, A.M., A.T. and G.S.; Data curation, E.Z. and A.T.; Formal analysis, A.M., E.Z. and A.T.; Funding acquisition, G.S.; Investigation, A.M., A.R., V.P., D.P., E.Z. and A.Z.; Methodology, A.M., A.R., V.P., D.P. and A.Z.; Project administration, G.S.; Resources, A.R., A.Z. and A.T.; Supervision, G.S.; Validation, A.M., V.P. and D.P.; Visualization, A.M., A.R., V.P., D.P. and A.Z.; Writing – original draft, A.M., A.R. and D.P.; Writing – review & editing, A.M., A.R. and G.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-15-2021-1085).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data generated or analyzed during this study are included in this article.

Acknowledgments: The genome sequencing experiments were performed using the equipment of the Resource Center of Genomic Research of Sirius University of Science and Technology.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; collection, analysis, or interpretation of data; writing of the manuscript; or decision to publish the results.

References

1. Barrangou, R.; Fremaux, C.; Deveau, H.; Richards, M.; Boyaval, P.; Moineau, S.; Romero, D.A.; Horvath, P. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* (1979) **2007**, 315, 1709–1712, doi:10.1126/science.1138140.
2. Brouns, S.J.J.; Jore, M.M.; Lundgren, M.; Westra, E.R.; Slikhuis, R.J.H.; Snijders, A.P.L.; Dickman, M.J.; Makarova, K.S.; Koonin, E. V.; van der Oost, J. Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science* **2008**, 321, 960–964, doi:10.1126/science.1159689.
3. Westra, E.R.; Swarts, D.C.; Staals, R.H.J.; Jore, M.M.; Brouns, S.J.J.; van der Oost, J. The CRISPRs, They Are A-Changin': How Prokaryotes Generate Adaptive Immunity. *Annu Rev Genet* **2012**, 46, 311–339, doi:10.1146/annurev-genet-110711-155447.
4. Hsu, P.D.; Scott, D.A.; Weinstein, J.A.; Ran, F.A.; Konermann, S.; Agarwala, V.; Li, Y.; Fine, E.J.; Wu, X.; Shalem, O.; et al. DNA Targeting Specificity of RNA-Guided Cas9 Nucleases. *Nat Biotechnol* **2013**, 31, 827–832, doi:10.1038/nbt.2647.
5. Ann Ran, F.; Hsu, P.D.; Wright, J.; Agarwala, V.; Scott, D.A.; Zhang, F. Genome Engineering Using the CRIPR-Cas9 System. *Nature* **2013**, doi:10.1038/nprot.2013.143.
6. Ran, F.A.; Hsu, P.D.; Lin, C.-Y.; Gootenberg, J.S.; Konermann, S.; Trevino, A.E.; Scott, D.A.; Inoue, A.; Matoba, S.; Zhang, Y.; et al. Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **2013**, 154, 1380–1389, doi:10.1016/j.cell.2013.08.021.
7. Doudna, J.A.; Charpentier, E. The New Frontier of Genome Engineering with CRISPR-Cas9. *Science* (1979) **2014**, 346, doi:10.1126/science.1258096.
8. Fonfara, I.; Richter, H.; Bratovič, M.; Le Rhun, A.; Charpentier, E. The CRISPR-Associated DNA-Cleaving Enzyme Cpf1 Also Processes Precursor CRISPR RNA. *Nature* **2016**, 532, 517–521, doi:10.1038/nature17945.
9. Kleinstiver, B.P.; Tsai, S.Q.; Prew, M.S.; Nguyen, N.T.; Welch, M.M.; Lopez, J.M.; McCaw, Z.R.; Aryee, M.J.; Joung, J.K. Genome-Wide Specificities of CRISPR-Cas Cpf1 Nucleases in Human Cells. *Nat Biotechnol* **2016**, doi:10.1038/nbt.3620.
10. Abudayyeh, O.O.; Gootenberg, J.S.; Konermann, S.; Joung, J.; Slaymaker, I.M.; Cox, D.B.T.; Shmakov, S.; Makarova, K.S.; Semenova, E.; Minakhin, L.; et al. C2c2 Is a Single-Component Programmable RNA-Guided RNA-Targeting CRISPR Effector. *Science* (1979) **2016**, 353, doi:10.1126/science.aaf5573.
11. Jones, M.K.; Grau, K.R.; Costantini, V.; Kolawole, A.O.; Graaf, D.; Freiden, P.; Graves, C.L.; Koopmans, M.; Wallet, S.M.; Tibbetts, S.A.; et al. Two Distinct RNase Activities of CRISPR-C2c2 Enable Guide-RNA Processing and RNA Detection. **2016**, 10, 1939–1947, doi:10.1038/nprot.2015.121.Human.
12. Reardon, S. Welcome to the CRISPR Zoo. *Nature* **2016**, 531, 160–163, doi:10.1038/531160a.
13. Uddin, F.; Rudin, C.M.; Sen, T. CRISPR Gene Therapy: Applications, Limitations, and Implications for the Future. *Front Oncol* **2020**, 10.
14. Kordyś, M.; Sen, R.; Warkocki, Z. Applications of the Versatile CRISPR-Cas13 RNA Targeting System. *WIREs RNA* **2022**, 13, e1694, doi:https://doi.org/10.1002/wrna.1694.
15. Butt, H.; Eid, A.; Momin, A.A.; Bazin, J.; Crespi, M.; Arold, S.T.; Mahfouz, M.M. CRISPR Directed Evolution of the Spliceosome for Resistance to Splicing Inhibitors. *Genome Biol* **2019**, 20, 73, doi:10.1186/s13059-019-1680-9.
16. Yao, R.; Liu, D.; Jia, X.; Zheng, Y.; Liu, W.; Xiao, Y. CRISPR-Cas9/Cas12a Biotechnology and Application in Bacteria. *Synth Syst Biotechnol* **2018**.
17. Myhrvold, C.; Freije, C.A.; Gootenberg, J.S.; Abudayyeh, O.O.; Metsky, H.C.; Durbin, A.F.; Kellner, M.J.; Tan, A.L.; Paul, L.M.; Parham, L.A.; et al. Field-Deployable Viral Diagnostics Using CRISPR-Cas13. *Science* (1979) **2018**, 360, 444–448, doi:10.1126/science.aas8836.
18. Loutre, R.; Heckel, A.-M.; Smirnova, A.; Entelis, N.; Tarassov, I. Can Mitochondrial DNA Be CRISPRized: Pro and Contra. *IUBMB Life* **2018**, 70, 1233–1239, doi:10.1002/iub.1919.
19. Labuhn, M.; Adams, F.F.; Ng, M.; Knoess, S.; Schambach, A.; Charpentier, E.M.; Schwarzer, A.; Mateo, J.L.; Klusmann, J.-H.; Heckl, D. Refined SgRNA Efficacy Prediction Improves Large- and Small-Scale CRISPR-Cas9 Applications. *Nucleic Acids Res* **2018**, 46, 1375–1385, doi:10.1093/nar/gkx1268.

20. Haque, E.; Taniguchi, H.; Hassan, Md.M.; Bhowmik, P.; Karim, M.R.; Śmiech, M.; Zhao, K.; Rahman, M.; Islam, T. Application of CRISPR/Cas9 Genome Editing Technology for the Improvement of Crops Cultivated in Tropical Climates: Recent Progress, Prospects, and Challenges. *Front Plant Sci* **2018**, doi:10.3389/fpls.2018.00617.
21. Adli, M. The CRISPR Tool Kit for Genome Editing and Beyond. *Nat Commun* **2018**.
22. Lee, M. Deep Learning in CRISPR-Cas Systems: A Review of Recent Studies. *Front Bioeng Biotechnol* **2023**, *11*.
23. Teng, Y.; Jiang, T.; Yan, Y. The Expanded CRISPR Toolbox for Constructing Microbial Cell Factories. *Trends Biotechnol* **2023**, doi:https://doi.org/10.1016/j.tibtech.2023.06.012.
24. Lau, C.-H.; Suh, Y. In Vivo Epigenome Editing and Transcriptional Modulation Using CRISPR Technology. *Transgenic Res* **2018**, *27*, 489–509, doi:10.1007/s11248-018-0096-8.
25. Hsu, P.D.; Scott, D. a; Weinstein, J. a; Ran, F.A.; Konermann, S.; Agarwala, V.; Li, Y.; Fine, E.J.; Wu, X.; Shalem, O.; et al. Rationally Engineered Cas9 Nucleases with Improved Specificity. *Nat Biotechnol* **2013**, *31*, 827–832, doi:10.1038/nbt.2647.
26. Knight, S.C.; Tjian, R.; Doudna, J.A. Genomes in Focus: Development and Applications of CRISPR-Cas9 Imaging Technologies. *Angewandte Chemie - International Edition* **2018**, *57*, 4329–4337.
27. Reeder, S.B.; Hu, H.H.; Sirlin, C.B.; Group, L.I.; Diego, S. Targeting Genomic Rearrangements in Tumor Cells Using Cas9- Mediated Insertion of a Suicide Gene. *Nat Biotechnol* **2017**, *36*, 1011–1014, doi:10.1002/jmri.23741.Proton.
28. Liszczak, G.P.; Brown, Z.Z.; Kim, S.H.; Oslund, R.C.; David, Y.; Muir, T.W. Genomic Targeting of Epigenetic Probes Using a Chemically Tailored Cas9 System. *Proceedings of the National Academy of Sciences* **2017**, *114*, 681–686, doi:10.1073/pnas.1615723114.
29. Hirose, M.; Fujita, Y.; Parr, C.J.C.; Hayashi, K.; Kashida, S.; Hotta, A.; Woltjen, K.; Saito, H. Cell-Type-Specific Genome Editing with a MicroRNA-Responsive CRISPR-Cas9 Switch. *Nucleic Acids Res* **2017**, *45*, doi:10.1093/nar/gkx309.
30. Spencer, N.Y.; Yan, Z.; Cong, L.; Zhang, Y.; Engelhardt, J.F. Definitive Localization of Intracellular Proteins- Novel Approach Using CRISPR-Cas9 Genome Editing with Glucose 6-Phosphate Dehydrogenase as a Model. **2017**, 55–67, doi:10.1016/j.ab.2015.11.002.Definitive.
31. Burstein, D.; Harrington, L.B.; Strutt, S.C.; Probst, A.J.; Anantharaman, K.; Thomas, B.C.; Doudna, J.A.; Banfield, J.F. New CRISPR–Cas Systems from Uncultivated Microbes. *Nature* **2017**, *542*, 237–241, doi:10.1038/nature21059.
32. Karvelis, T.; Gasiunas, G.; Young, J.; Bigelyte, G.; Silanskas, A.; Cigan, M.; Siksnys, V. Rapid Characterization of CRISPR-Cas9 Protospacer Adjacent Motif Sequence Elements. *Genome Biol* **2015**, *16*, 253, doi:10.1186/s13059-015-0818-7.
33. Koonin, E. V; Gootenberg, J.S.; Abudayyeh, O.O. Discovery of Diverse CRISPR-Cas Systems and Expansion of the Genome Engineering Toolbox. *Biochemistry* **2023**, doi:10.1021/acs.biochem.3c00159.
34. Mir, A.; Edraki, A.; Lee, J.; Sontheimer, E.J. Type II-C CRISPR-Cas9 Biology, Mechanism, and Application. *ACS Chem Biol* **2018**, *13*, 357–365, doi:10.1021/acscchembio.7b00855.
35. Harrington, L.B.; Paez-Espino, D.; Staahl, B.T.; Chen, J.S.; Ma, E.; Kyrpides, N.C.; Doudna, J.A. A Thermostable Cas9 with Increased Lifetime in Human Plasma. *Nat Commun* **2017**, *8*, 1424, doi:10.1038/s41467-017-01408-4.
36. da Rosa, D.F.; Macedo, A.J. The Genus *Anoxybacillus*: An Emerging and Versatile Source of Valuable Biotechnological Products. *Extremophiles* **2023**, *27*, 22, doi:10.1007/s00792-023-01305-3.
37. Schultz, J.; Parise, M.T.D.; Parise, D.; Medeiros, L.G.; Sousa, T.J.; Kato, R.B.; Uetanabaro, A.P.T.; Araújo, F.; Ramos, R.T.J.; de Castro Soares, S.; et al. Unraveling the Genomic Potential of the Thermophilic Bacterium *Anoxybacillus Flavithermus* from an Antarctic Geothermal Environment. *Microorganisms* **2022**, *10*, 1673, doi:10.3390/microorganisms10081673.
38. Okonechnikov, K.; Golosova, O.; Fursov, M. Unipro UGENE: A Unified Bioinformatics Toolkit. *Bioinformatics* **2012**, *28*, 1166–1167, doi:10.1093/bioinformatics/bts091.
39. Anders, C.; Jinek, M. In Vitro Enzymology of Cas9. In; 2014; pp. 1–20.
40. Ryabchenko, A. V.; Karavaev, V.S.; Beklemishev, A.B. Comparative Structural and Immunochemical Analysis of Recombinant OspC Antigens of Novosibirsk *Borrelia garinii* and *Borrelia afzelii* Isolates. *Bulletin of SB RAMS* **2010**, 6–12.

41. Daniel, E.; Onwukwe, G.U.; Wierenga, R.K.; Quaggin, S.E.; Vainio, S.J.; Krause, M. ATGme: Open-Source Web Application for Rare Codon Identification and Custom DNA Sequence Optimization. *BMC Bioinformatics* **2015**, *16*, 303, doi:10.1186/s12859-015-0743-5.
42. Studier, F.W. Protein Production by Auto-Induction in High-Density Shaking Cultures. *Protein Expr Purif* **2005**, *41*, 207–234, doi:https://doi.org/10.1016/j.pep.2005.01.016.
43. Biswas, A.; Gagnon, J.N.; Brouns, S.J.J.; Fineran, P.C.; Brown, C.M. CRISPRTarget. *RNA Biol* **2013**, *10*, 817–827, doi:10.4161/rna.24046.
44. Crooks, G.E.; Hon, G.; Chandonia, J.-M.; Brenner, S.E. WebLogo: A Sequence Logo Generator: Figure 1. *Genome Res* **2004**, *14*, 1188–1190, doi:10.1101/gr.849004.
45. Prokhorova, D. V.; Vokhtantsev, I.P.; Tolstova, P.O.; Zhuravlev, E.S.; Kulishova, L.M.; Zharkov, D.O.; Stepanov, G.A. Natural Nucleoside Modifications in Guide RNAs Can Modulate the Activity of the CRISPR-Cas9 System *In Vitro*. *CRISPR J* **2022**, *5*, 799–812, doi:10.1089/crispr.2022.0069.
46. Wright, A. V.; Sternberg, S.H.; Taylor, D.W.; Staahl, B.T.; Bardales, J.A.; Kornfeld, J.E.; Doudna, J.A. Rational Design of a Split-Cas9 Enzyme Complex. *Proceedings of the National Academy of Sciences* **2015**, *112*, 2984–2989, doi:10.1073/pnas.1501698112.
47. Le, Y.; Sun, J. CRISPR/Cas Genome Editing Systems in Thermophiles: Current Status, Associated Challenges, and Future Perspectives. In; 2022; pp. 1–30.
48. Lau, M.S.H.; Sheng, L.; Zhang, Y.; Minton, N.P. Development of a Suite of Tools for Genome Editing in *Parageobacillus Thermoglucosidasius* and Their Use to Identify the Potential of a Native Plasmid in the Generation of Stable Engineered Strains. *ACS Synth Biol* **2021**, *10*, 1739–1749, doi:10.1021/acssynbio.1c00138.
49. Singh, V.; Raheja, Y.; Basotra, N.; Sharma, G.; Tsang, A.; Chadha, B.S. CRISPR/Cas9 Mediated Gene Editing of Transcription Factor ACE1 for Enhanced Cellulase Production in Thermophilic Fungus *Rasamsonia Emersonii*. *Fungal Biol Biotechnol* **2023**, *10*, 18, doi:10.1186/s40694-023-00165-y.
50. Gallo, G.; Mougiakos, I.; Bianco, M.; Carbonaro, M.; Carpentieri, A.; Illiano, A.; Pucci, P.; Bartolucci, S.; van der Oost, J.; Fiorentino, G. A Hyperthermoactive-Cas9 Editing Tool Reveals the Role of a Unique Arsenite Methyltransferase in the Arsenic Resistance System of *Thermus Thermophilus* HB27. *mBio* **2021**, *12*, doi:10.1128/mBio.02813-21.
51. Hand, T.H.; Das, A.; Li, H. Directed Evolution Studies of a Thermophilic Type II-C Cas9. In; 2019; pp. 265–288.
52. Tsui, T.K.M.; Hand, T.H.; Duboy, E.C.; Li, H. The Impact of DNA Topology and Guide Length on Target Selection by a Cytosine-Specific Cas9. *ACS Synth Biol* **2017**, *6*, 1103–1113, doi:10.1021/acssynbio.7b00050.
53. Das, A.; Hand, T.H.; Smith, C.L.; Wickline, E.; Zawrotny, M.; Li, H. The Molecular Basis for Recognition of 5'-NNNCC-3' PAM and Its Methylation State by *Acidothermus Cellulolyticus* Cas9. *Nat Commun* **2020**, *11*, 6346, doi:10.1038/s41467-020-20204-1.
54. Mougiakos, I.; Mohanraju, P.; Bosma, E.F.; Vrouwe, V.; Finger Bou, M.; Naduthodi, M.I.S.; Gussak, A.; Brinkman, R.B.L.; van Kranenburg, R.; van der Oost, J. Characterizing a Thermostable Cas9 for Bacterial Genome Editing and Silencing. *Nat Commun* **2017**, *8*, 1647, doi:10.1038/s41467-017-01591-4.
55. Ganguly, J.; Martin-Pascual, M.; van Kranenburg, R. CRISPR Interference (CRISPRi) as Transcriptional Repression Tool for *Hungateiclostridium Thermocellum* DSM 1313. *Microb Biotechnol* **2020**, *13*, 339–349, doi:10.1111/1751-7915.13516.
56. Trasanidou, D.; Barendse, P.; Bouzetos, E.; de Haan, L.; Bouwmeester, H.; Staals, R.H.J.; Mougiakos, I.; van der Oost, J. Efficient Genome and Base Editing in Human Cells Using ThermoCas9. *CRISPR J* **2023**, *6*, 278–288, doi:10.1089/crispr.2023.0005.
57. Schmidt, S.T.; Yu, F.B.; Blainey, P.C.; May, A.P.; Quake, S.R. Nucleic Acid Cleavage with a Hyperthermophilic Cas9 from an Uncultured Ignavibacterium. *Proceedings of the National Academy of Sciences* **2019**, *116*, 23100–23105, doi:10.1073/pnas.1904273116.
58. Chen, J.; Qiu, T.; Mauk, M.G.; Su, Z.; Fan, Y.; Yuan, D.J.; Zhou, Q.; Qiao, Y.; Bau, H.H.; Ying, J.; et al. Programmable Endonuclease Combined with Isothermal Polymerase Amplification to Selectively Enrich for Rare Mutant Allele Fractions. *Chinese Chemical Letters* **2022**, *33*, 4126–4132, doi:10.1016/j.cclet.2021.11.065.
59. Zou, R.S.; Gavrillov, M.; Liu, Y.; Rasoloson, D.; Conte, M.; Hardick, J.; Shen, L.; Chen, S.; Pekosz, A.; Seydoux, G.; et al. Improving the Specificity of Nucleic Acid Detection with Endonuclease-Actuated Degradation. *Commun Biol* **2022**, *5*, 290, doi:10.1038/s42003-022-03242-x.

60. Chen, J.; Qiu, T.; Mauk, M.G.; Fan, Y.; Jiang, Y.; Ying, J.; Zhou, Q.; Qiao, Y.; Bau, H.H.; Song, J. CRISPR Cas9-Mediated Selective Isothermal Amplification for Sensitive Detection of Rare Mutant Alleles. *Clin Chem* **2021**, *67*, 1569–1571, doi:10.1093/clinchem/hvab163.
61. Xie, L.; Hu, Y.; Li, L.; Jiang, L.; Jiao, Y.; Wang, Y.; Zhou, L.; Tao, R.; Qu, J.; Chen, Q.; et al. Expanding PAM Recognition and Enhancing Base Editing Activity of Cas9 Variants with Non-PI Domain Mutations Derived from XCas9. *FEBS J* **2022**, *289*, 5899–5913, doi:10.1111/febs.16457.
62. Walton, R.T.; Christie, K.A.; Whittaker, M.N.; Kleinstiver, B.P. Unconstrained Genome Targeting with Near-PAMless Engineered CRISPR-Cas9 Variants. *Science (1979)* **2020**, *368*, 290–296, doi:10.1126/science.aba8853.
63. Legut, M.; Daniloski, Z.; Xue, X.; McKenzie, D.; Guo, X.; Wessels, H.-H.; Sanjana, N.E. High-Throughput Screens of PAM-Flexible Cas9 Variants for Gene Knockout and Transcriptional Modulation. *Cell Rep* **2020**, *30*, 2859–2868.e5, doi:10.1016/j.celrep.2020.02.010.
64. Zhang, D.; Zhang, B. SpRY: Engineered CRISPR/Cas9 Harnesses New Genome-Editing Power. *Trends in Genetics* **2020**, *36*, 546–548, doi:10.1016/j.tig.2020.05.001.
65. Leenay, R.T.; Maksimchuk, K.R.; Slotkowski, R.A.; Agrawal, R.N.; Gomaa, A.A.; Briner, A.E.; Barrangou, R.; Beisel, C.L. Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Mol Cell* **2016**, *62*, 137–147, doi:10.1016/j.molcel.2016.02.031.
66. Karvelis, T.; Gasiunas, G.; Siksnys, V. Methods for Decoding Cas9 Protospacer Adjacent Motif (PAM) Sequences: A Brief Overview. *Methods* **2017**, *121–122*, 3–8, doi:10.1016/j.ymeth.2017.03.006.
67. Karvelis, T.; Gasiunas, G.; Siksnys, V. Harnessing the Natural Diversity and in Vitro Evolution of Cas9 to Expand the Genome Editing Toolbox. *Curr Opin Microbiol* **2017**, *37*, 88–94, doi:10.1016/j.mib.2017.05.009.
68. Ma, D.; Xu, Z.; Zhang, Z.; Chen, X.; Zeng, X.; Zhang, Y.; Deng, T.; Ren, M.; Sun, Z.; Jiang, R.; et al. Engineer Chimeric Cas9 to Expand PAM Recognition Based on Evolutionary Information. *Nat Commun* **2019**, *10*, 560, doi:10.1038/s41467-019-08395-8.
69. Adalsteinsson, B.T.; Kristjansdottir, T.; Merre, W.; Helleux, A.; Dusaucy, J.; Tourigny, M.; Fridjonsson, O.; Hreggvidsson, G.O. Efficient Genome Editing of an Extreme Thermophile, *Thermus Thermophilus*, Using a Thermostable Cas9 Variant. *Sci Rep* **2021**, *11*, 9586, doi:10.1038/s41598-021-89029-2.
70. Shi, Y.-J.; Duan, M.; Ding, J.-M.; Wang, F.-Q.; Bi, L.-L.; Zhang, C.-X.; Zhang, Y.-Z.; Duan, J.-Y.; Huang, A.-H.; Lei, X.-L.; et al. DNA Topology Regulates PAM-Cas9 Interaction and DNA Unwinding to Enable near-PAMless Cleavage by Thermophilic Cas9. *Mol Cell* **2022**, *82*, 4160–4175.e6, doi:10.1016/j.molcel.2022.09.032.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.