Nucleic acid cleavage with a hyperthermophilic Cas9 from an uncultured Ignavibacterium

Stephanie Tzouanas Schmidt[a], Feiqiao Brian Yu[a,b,c], Paul C. Blainey[d,e], Andrew P. May[c], and Stephen R. Quake[a,c,f,1]

[a]Department of Bioengineering, Stanford University, Stanford, CA 94305; [b]Department of Electrical Engineering, Stanford University, Stanford, CA 94305; [c]Chan Zuckerberg Biohub, San Francisco, CA 94158; [d]Department of Biological Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02142; [e]Broad Institute of Harvard and MIT, Cambridge, MA 02142; and [f]Department of Applied Physics, Stanford University, Stanford, CA 94305

Contributed by Stephen R. Quake, September 5, 2019 (sent for review March 15, 2019; reviewed by Gang Bao and Jacob Corn)

Clustered regularly interspaced short palindromic repeats (CRISPR)-associated 9 (Cas9) systems have been effectively harnessed to engineer the genomes of organisms from across the tree of life. Nearly all currently characterized Cas9 proteins are derived from mesophilic bacteria, and canonical Cas9 systems are challenged by applications requiring enhanced stability or elevated temperatures. We discovered IgnaviCas9, a Cas9 protein from a hyperthermophilic Ignavibacterium identified through mini-metagenomic sequencing of samples from a hot spring. IgnaviCas9 is active at temperatures up to 100 °C in vitro, which enables DNA cleavage beyond the 44 °C limit of *Streptococcus pyogenes* Cas9 (SpyCas9) and the 70 °C limit of both *Geobacillus stearothermophilus* Cas9 (GeoCas9) and *Geobacillus thermodenitrificans* T12 Cas9 (ThermoCas9). As a potential application of this enzyme, we demonstrate that IgnaviCas9 can be used in bacterial RNA-seq library preparation to remove unwanted cDNA from 16s ribosomal rRNA without increasing the number of steps, thus underscoring the benefits provided by its exceptional thermostability in improving molecular biology and genomic workflows. IgnaviCas9 is an exciting addition to the CRISPR-Cas9 toolbox and expands its temperature range.

CRISPR/Cas9 | genomics | extremophiles

The application of clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins has revolutionized molecular biology by making genome editing facile in both prokaryotes and eukaryotes (1, 2). Constituting the heritable and adaptive immune system of prokaryotes, CRISPR-Cas9 systems are present in archaea and bacteria from diverse environments (3). A wide variety of CRISPR-Cas9 systems exist, and class 2 systems, particularly type II systems, have been well characterized and broadly implemented, in part because these systems rely on a single effector protein, Cas9, and an RNA duplex, which can be replaced by a single guide RNA (sgRNA). CRISPR-Cas9 systems, particularly that from *Streptococcus pyogenes*, have been leveraged to edit genomes across organisms, create new tools for sequencing applications, and more (4).

Nearly all Cas9 proteins have been derived from mesophilic hosts and thus cannot be used in applications in which elevated temperatures and robust stability are required. Two Cas9 proteins from thermophiles that provide enhanced stability in in vivo environments and enable genome editing in thermophilic organisms have been reported recently (5, 6). These 2 proteins, GeoCas9 and ThermoCas9, were identified by sequencing environmental samples, and their hosts live at temperatures of 65 and 70 °C, respectively.

More broadly, thermostable enzymes have led to important advances in biotechnology, such as the polymerase chain reaction (PCR) assay. The ability to use CRISPR-Cas9 at even higher temperatures than those reported could enable exciting new molecular biology techniques. Our group recently used mini-metagenomic sequencing to discover a distinct CRISPR-Cas9 system in a hyperthermophilic bacterium from Yellowstone National Park's Lower Geyser Basin, where temperatures exceed 90 °C (7).

Motivated by the potential of CRISPR-Cas9 at such elevated temperatures, we expressed, purified, and characterized this type II-C Cas9 protein, which we call IgnaviCas9. IgnaviCas9 is active at temperatures up to 100 °C; its active temperature range is the widest yet reported for a CRISPR-Cas9 system, opening up potential molecular biology applications. As one such potential application, we demonstrate the reduction of undesirable 16s rRNA library molecules in bacterial samples being prepared for RNA-seq.

Results

Identification, Phylogenetic Characterization, and Expression of IgnaviCas9. Microfluidic mini-metagenomic sequencing of a sediment sample from the Lower Geyser Basin of Yellowstone National Park yielded a complete metagenome assembled genome (7). Comprising a single 3.4-Mb contig representing a novel lineage in the Ignavibacteriae phylum, this genome was found to contain a full CRISPR array. The temperature of the sample was recorded as 55 °C and that of the hot spring as >90 °C. That genome editing could occur in a bacterium residing in a near-boiling point ecosystem intrigued us, and we examined the sequence of the CRISPR array for insight into its properties.

The CRISPR array contains a Cas9 protein, Cas1 protein, and Cas2 protein along with 38 unique spacers. The absence of a

Significance

Cas9 proteins have revolutionized biotechnology by enabling flexible and facile targeted cleavage of nucleic acids. Nearly all of these proteins are active only at moderate, near-physiological temperatures. Through mini-metagenomic sequencing of hot spring samples from Yellowstone National Park, we discovered and characterized a novel hyperthermophilic Cas9 protein from an unculturable Ignavibacterium. This Cas9 protein, IgnaviCas9, expands the temperature range at which targeted nucleic acid cleavage is possible, thus speeding the development of new biotechnological techniques. We demonstrated one such application by using IgnaviCas9 to deplete undesired amplicons during the amplification step of library preparation in sequencing workflows. This Cas9 protein underscores the exciting applications that can be made possible by exploring nature's diversity.

Author contributions: S.T.S., F.B.Y., A.P.M., and S.R.Q. designed research; S.T.S. and F.B.Y. performed research; P.C.B. and S.R.Q. contributed new reagents/analytic tools; S.T.S., F.B.Y., A.P.M., and S.R.Q. analyzed data; and S.T.S. and S.R.Q. wrote the paper.

Reviewers: G.B., Rice University; and J.C., ETH Zurich.

Competing interest statement: A patent concerning the technology disclosed in this publication has been filed.

Data deposition: The IgnaviCas9 expression plasmid is available at Addgene (sequence 127595). Sequence data are available via Google Drive at bit.ly/2MdQoB8.

[1]To whom correspondence may be addressed. Email: quake@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1904273116/-/DCSupplemental.

First published October 28, 2019.

Csn2 and Cas4 protein suggested that the Ignavibacterium has a type II-C system (8), which was confirmed by phylogenetic comparison of IgnaviCas9 to other type II Cas9 proteins (Fig. 1*A*). In brief, multiple sequence alignment of amino acid sequences of representative type II Cas9 proteins was performed using MAFFT, and a maximum-likelihood phylogenetic tree was constructed using RAxML with the PROTGAMMALG substitution model and 100 bootstrap samplings (9). Like GeoCas9 and ThermoCas9,

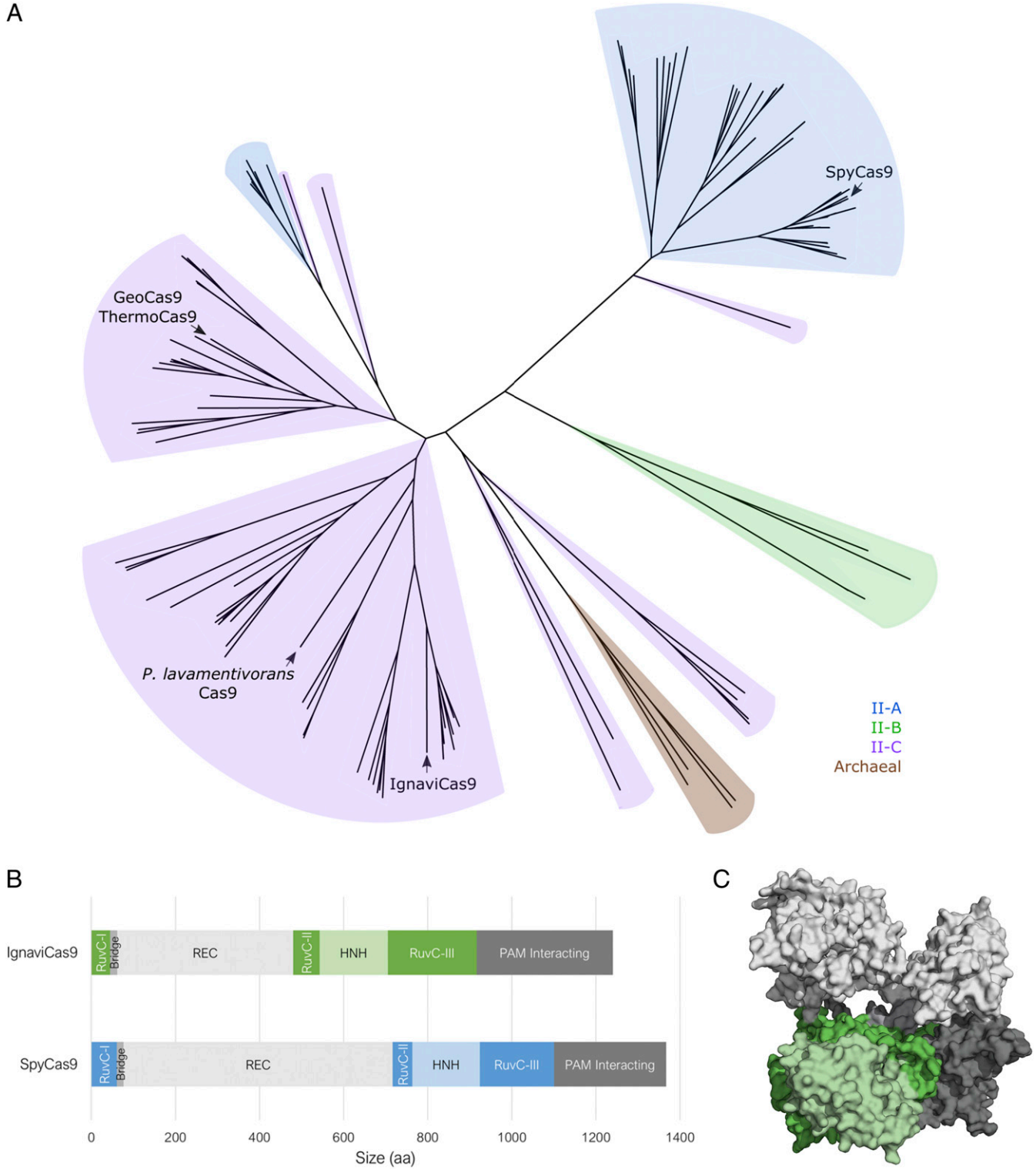

Fig. 1. Phylogenetic classification and structural prediction of IgnaviCas9. (*A*) Phylogenetic tree of representative Cas9s from type II systems. (*B*) Architectural domains of IgnaviCas9 and SpyCas9, where REC is the recognition lobe. (*C*) A homology model of IgnaviCas9 with colors indicating the domains annotated in *B*. The model was generated using Phyre2 (12).

IgnaviCas9 is a type II-C Cas9. However, IgnaviCas9 is located in an entirely different clade, suggesting that it is highly divergent from the 2 thermostable Cas9s reported thus far. The in vitro validated type II-C Cas9 to which it is most similar is that of *Parvibaculum lavamentivorans* (10), a mesophilic bacterium with an optimal growth temperature of 30 °C (11).

At 1,241-aa long, IgnaviCas9 is certainly shorter than SpyCas9 (1,368 aa) but longer than ThermoCas9 (1,082 aa) or GeoCas9 (1,087 aa). Through homology modeling and sequence alignment to other type II Cas9 proteins (*SI Appendix*, Fig. S1), the smaller size of IgnaviCas9 compared with SpyCas9 was found to arise from its reduced recognition (REC) lobe (Fig. 1*B*), which is consistent with other smaller Cas9s (10). While larger than other in vitro validated type II-C Cas9 proteins, IgnaviCas9 is shorter than SpyCas9, which can be advantageous for applications involving its delivery via adeno-associated viruses (13).

Having examined IgnaviCas9 phylogenetically, we next sought to produce and test the protein. Because the bacterium from which IgnaviCas9 comes has not been isolated and likely cannot be grown in standard laboratory culture conditions, we codon-optimized the sequence of IgnaviCas9 and cloned it into a Cas9 expression vector. BL21 *Escherichia coli* cells were transformed with this plasmid and cultured to express IgnaviCas9. Subsequent purification provided 12 mg of IgnaviCas9 from 4 L of culture for downstream experiments.

IgnaviCas9 sgRNA Engineering. That IgnaviCas9 falls within the type II-C classification proved helpful in designing its sgRNA based on computational prediction of its CRISPR RNA (crRNA) and *trans*-activating crRNA (tracrRNA) from the CRISPR array sequence. This approach was necessary since the bacterium from which IgnaviCas9 was isolated was not available and not likely culturable. We looked for combinations of potential crRNA and tracrRNA sequences that together allow for formation of the lower stem, bulge, upper stem, nexus, and hairpin features (Fig. 2*A*).

RNA secondary structure prediction of the designed sgRNA showed that all desired features remained present at temperatures of 60 °C for default NUPACK program settings, underscoring the potential of IgnaviCas9 to cleave DNA at temperatures outside the mesophilic range. We transcribed the sgRNA sequence preceded by 25 nt of spacer sequence for use in preliminary experiments.

IgnaviCas9 PAM Determination and sgRNA-Spacer Match Length Refinement. The protospacer adjacent motif (PAM), the sequence directly downstream of a nucleic acid target cleavable by CRISPR systems, varies among species and prevents the host genome from being attacked (14). As an initial approach, we designed double-stranded linear DNA containing a spacer sequence followed by a PAM from an in vitro validated type II-C CRISPR system (8). We performed cleavage assays by incubating the assorted DNA substrates with a ribonucleoprotein complex (RNP) of IgnaviCas9 and sgRNA targeting the spacer sequence at 52 °C for 30 min. We found that IgnaviCas9 cleaved the DNA substrate with the PAM CCACATCGAA, containing the NNNCAT motif from *P. lavamentivorans* (Fig. 3*A*). A control reaction used as a point of reference differed, in that the sgRNA included contained a scrambled version of the spacer sequence.

We then used the 38 spacers found in the IgnaviCas9 CRISPR array to isolate possible protospacers from the environmental sample in which IgnaviCas9 was found. Using BLAST to search the environmental sequences, we collected 10-bp sequences flanking the spacer that differed from the repeat sequence by an edit distance of at least 5. The sequence logo created using unique sequences meeting these criteria suggested that the PAM was likely to be adenine-rich (*SI Appendix*, Fig. S2*A*); positions 7 through 10 are not shown, because they were ultimately determined to not impact the PAM.

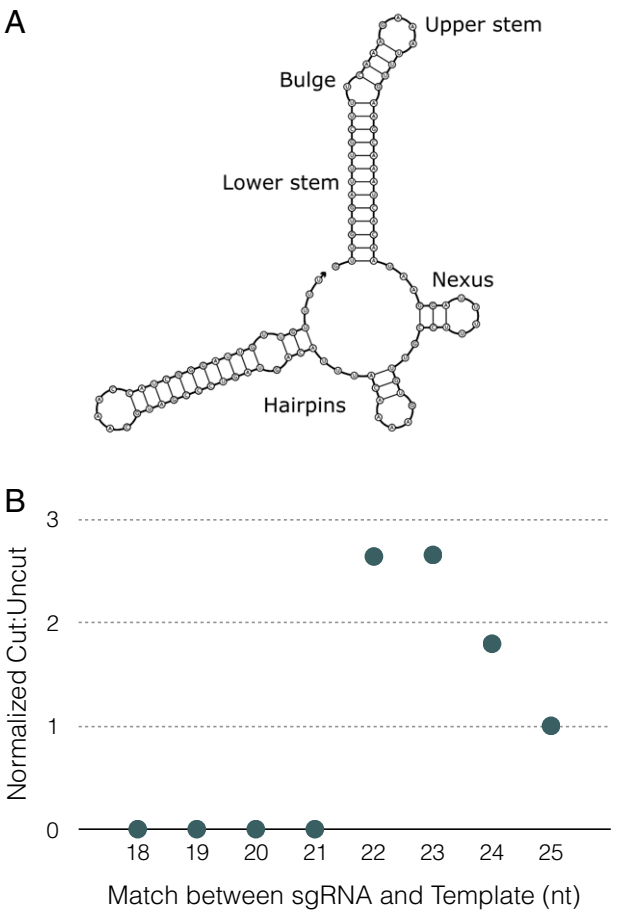

Fig. 2. sgRNA structure and spacer length. (*A*) Representation of the determined sgRNA with important structural features labeled. (*B*) Testing of the preferred spacer length was conducted by comparing cleavage at 52 °C of templates targeted by truncated versions of the initial spacer. The cut-to-uncut ratio was normalized to that corresponding to 25 nt (the length used for preliminary experiments).

We designed a new DNA substrate by modifying the aforementioned DNA substrate that was cut by IgnaviCas9 to include AGACATGAAA, an adenine-rich version of the *P. lavamentivorans* PAM. This choice was also informed by the results of a randomer depletion experiment. In brief, a template containing a 10-bp-long randomer was used as the DNA substrate in a cleavage reaction. The resulting mixture of fragments underwent sequencing, and a sequence logo was generated using randomers depleted relative to their presence in the starting library (*SI Appendix*, Fig. S2*B*). In a cleavage reaction performed as before, IgnaviCas9 was able to better cleave the DNA substrate containing the refined PAM (Fig. 3*B*).

We finalized the PAM recognized by IgnaviCas9 by testing DNA substrates containing the adenine-rich *P. lavamentivorans* PAM with single nucleotide substitutions at each of the 10 positions directly downstream of the spacer. To concisely convey the performance of IgnaviCas9 in cleaving the DNA template, we calculated a metric that we term the "cut-to-uncut" ratio by dividing the area under the larger cut peak by the area under the uncut peak (*SI Appendix*, Fig. S3). Disruption of IgnaviCas9 cleavage by a particular substitution demonstrated that the position of the substitution was important to the PAM, and that the nucleotide was not part of the PAM. We found that NRRNAT is the PAM recognized by IgnaviCas9; all substitutions at positions

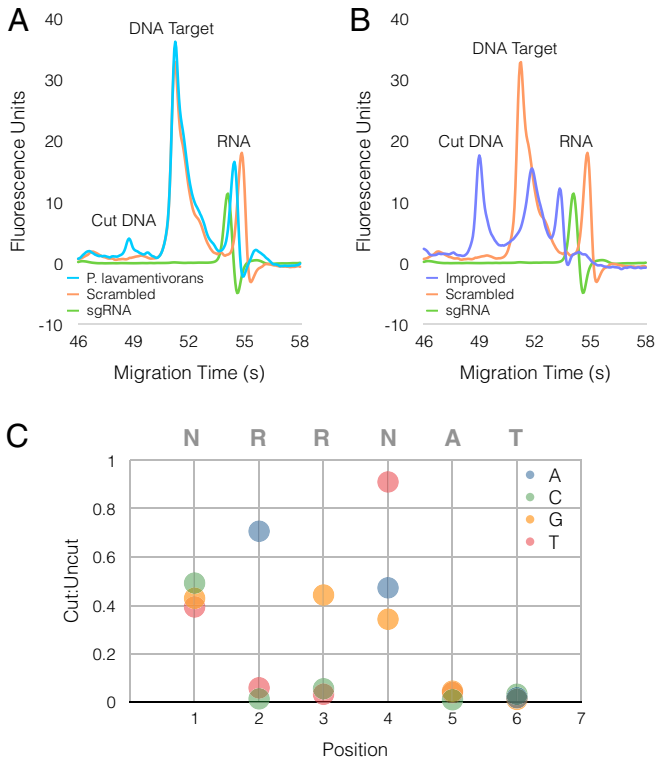

Fig. 3. PAM determination. (*A*) Electropherogram showing cleavage of the template containing the PAM from *P. lavamentivorans* compared with the control reaction with scrambled sgRNA and to sgRNA from the experimental condition. (*B*) Electropherogram showing cleavage of the template containing the PAM from *P. lavamentivorans* with adjustments informed by leads from bulk sequencing data. Curves from control reaction with scrambled sgRNA and from experimental condition sgRNA are included for comparison. (*C*) Performance of IgnaviCas9 in cleaving DNA templates with the indicated substitutions at the specified positions for the starting sequence of AGACAT. Substitutions abolishing cleavage activity enabled PAM refinement.

past 6 bp downstream of the spacer sequence were tolerated (Fig. 3*C*).

Having established the PAM of IgnaviCas9, we varied the length of the spacer included in the sgRNA to determine which lengths were optimal. We demonstrated that IgnaviCas9 cleaves DNA when the sgRNA includes spacer lengths of 22 to 25 nt, with improved performance for 22- and 23-nt spacer lengths (Fig. 2*B*). Cleavage does not occur for sgRNA with shorter spacer lengths. The spacer lengths that IgnaviCas9 prefers overlap with those favored by ThermoCas9 (19 to 25 nt) and GeoCas9 (21 or 22 nt) but are slightly larger than the 20 nt typically used with SpyCas9.

Active Temperature Range Assessment. By conducting the PAM determination experiments at 52 °C, we confirmed that IgnaviCas9 is active at temperatures above those of the active range of SpyCas9, which has been reported as between 20 and 44 °C (15, 16). To quantify the efficiency of IgnaviCas9 in cleaving DNA, we defined its efficiency as the area under the larger cut peak from the test condition divided by the area under the uncut peak from the scrambled control (*SI Appendix*, Fig. S4). We characterized the temperature range over which IgnaviCas9 is active by performing cleavage assays between 5 °C and 100 °C (Fig. 4*A* and *SI Appendix*, Fig. S5). We found that its performance in cutting various DNA targets, including longer templates like plasmid DNA, extended across the range tested, which reaches beyond the upper active temperature limit of other thermostable Cas9 proteins (Fig. 4*B*).

That IgnaviCas9 remains active at high temperatures and across a wide thermal range (Fig. 4*C*) indicates that it is particularly stable. Like ThermoCas9 (6), its spacer-protospacer mismatch tolerance does increase with temperature (*SI Appendix*, Fig. S6). As its mismatch tolerance is comparable to that of other thermostable Cas9 proteins (5, 6), we expect that IgnaviCas9 is likely more specific in its targeting than SpyCas9. More generally, IgnaviCas9 is more sensitive to mismatches proximal to the PAM than mismatches distal to PAM, which is consistent with the behavior of other Cas9 proteins.

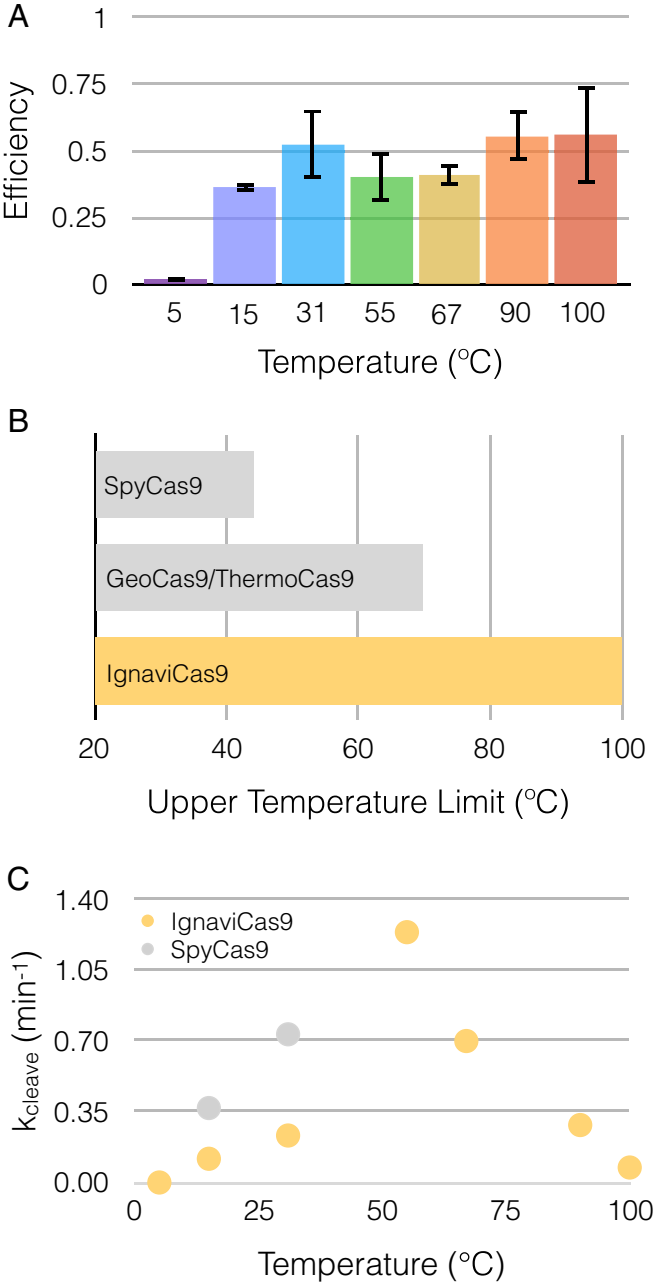

Fig. 4. Temperature evaluation of IgnaviCas9. (*A*) The efficiency of IgnaviCas9 in cleaving DNA templates is compared over a range of temperatures. The average and SD at each temperature tested is shown ($n = 3$). (*B*) The upper temperature limit of Cas9 homologs. (*C*) IgnaviCas9's rate of DNA cleavage compared with that of SpyCas9 over a range of temperatures.

Implementation of IgnaviCas9 to Remove Undesired Amplicons. The wide active temperature range of IgnaviCas9 is a unique property that can be harnessed for a host of molecular biology applications. In particular, its activity at both moderate and high temperatures led us to consider how it could be integrated into molecular biology and genomic workflows to eliminate undesired amplicons. Given our previous work in mini-metagenomics, we were interested in using IgnaviCas9 to reduce the amplification of library molecules derived from 16s rRNA in bacterial RNA-Seq.

When performing RNA-seq of actively growing bacterial strains or generating metatranscriptomic data from environmental samples, reads from 16s rRNA genes are typically highly abundant and reduce the sequencing bandwidth of expression profiles of interest. With this in mind, we deployed IgnaviCas9 during the PCR step of the sequencing library preparation workflow to cleave library fragments derived from 16s rRNA, thus reducing their presence in the final library without adding steps to the workflow. Previous work using mesophilic Cas9 in an additional workflow step before amplification has shown that this general idea has powerful applications (17), and we demonstrate that targeted depletion with IgnaviCas9 can be achieved during amplification, thus offering a more streamlined workflow and avoiding the additional clean-up step required by existing methods.

To this end, we designed sgRNA to target cDNA resulting from 16s rRNA and included IgnaviCas9 complexed to these sgRNAs in a PCR. Through sequencing, we demonstrated that compared with the starting library and a control reaction with SpyCas9, IgnaviCas9 targeting during amplification reduced the contribution of libraries derived from 16s rRNA, thus enriching the portion containing transcripts of interest (Fig. 5). More broadly, our approach could be used to eliminate other unwanted amplicons (e.g., primer dimers) as they are generated. Such implementations of IgnaviCas9 underscore its utility in improving widely used existing techniques in genomics and molecular biology.

Discussion

Taken together, this work expands the possible range of CRISPR-Cas9 to temperatures as high as 100 °C. While nearly all characterized Cas proteins are active at mesophilic temperatures, IgnaviCas9 is clearly hyperthermophilic while also showing activity at lower temperatures as well. Its natural propensity to cleave DNA across such a wide temperature range circumvents the need for protein engineering to create such a Cas9 and underscores the value of environmental metagenomics. By exploring the rich diversity of microbes present in extreme environments, we were able to identify a promising CRISPR-Cas9 system for further study and use. IgnaviCas9 is highly thermostable, enabling a wide variety of important applications. The highest previously reported active temperature for a Cas9 is 70 °C.

That IgnaviCas9 is able to bind and cleave DNA at such high temperatures underscores its stability, a feature that could make it well suited for in vivo use. In particular, increased stability suggests that IgnaviCas9 will have a longer lifetime in plasma compared with canonical variants and thus will be more effective in such applications as gene therapies (18) and lineage tracing in complex organisms (19). While organisms dwelling at higher temperatures are typically unicellular microorganisms, these microbes can catalyze important high-temperature industrial processes like fermentation. The improved ability to engineer thermophilic bacteria by means of IgnaviCas9 will facilitate the development and broader implementation of these processes. Finally, the ability to edit DNA at elevated temperatures beyond the active range of previously reported Cas9s is key to both improving existing and creating new molecular biology applications in which temperature serves as a means of control.

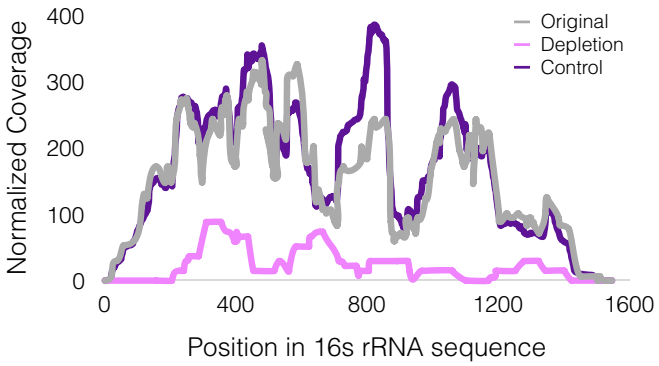

Fig. 5. Reduction of targeted sequence by IgnaviCas9. Coverage plot for 16s rRNA sequence targeted by IgnaviCas9 during PCR amplification. Normalized coverage given as per-base coverage divided by average whole genome coverage.

Methods

IgnaviCas9 Identification, Expression, and Purification. IgnaviCas9 was found through mini-metagenomic sequencing of a sediment sample taken from Mound Spring in the Lower Geyser Basin area of Yellowstone National Park under permit YELL-2009-SCI-5788. The sample was placed in 50% ethanol in a 2-mL tube without any filtering and kept frozen until its return from Yellowstone to Stanford University, at which time tubes containing the samples were transferred to −80 °C for long-term storage.

To compare IgnaviCas9 with other Cas9s (9), multiple sequence alignment of type II Cas9s was performed using MAFFT (20) and a maximum-likelihood phylogenetic tree was constructed using RAxML with the PROTGAMMALG substitution model and 100 bootstrap samplings (21). Its DNA sequence was codon-optimized for expression in *E. coli* and then synthesized (Integrated DNA Technologies). The resulting DNA was cloned into a pET-based vector with an N-terminal hexahistidine, maltose-binding protein, and tobacco etch virus sequence and C-terminal nuclear localization sequences.

IgnaviCas9 was expressed in BL21 strain *E. coli* (Agilent). After cultures reached an OD_{600} of 0.5, expression was induced by adding isopropyl β-D-1-thiogalactopyranoside to give a final concentration of 0.5 mM. The cultures were allowed to incubate for 7 h at 16 °C. Cells were harvested via centrifugation, and IgnaviCas9 was purified using ion exchange and size exclusion chromatography as described previously (17). IgnaviCas9-containing fractions were pooled, supplemented with glycerol to a final concentration of 50%, and stored at −80 °C until use.

sgRNA Design and Transcription. The crRNA and tracrRNA were identified from the IgnaviCas9 CRISPR locus by searching for complementarity between candidate sequences allowing for the formation of the requisite features when linked by a 5′-GAAA-3′ tetraloop (22). Possible sgRNA sequences were tested through secondary structure prediction using NUPACK (23).

DNA corresponding to the sgRNA including the target of interest was placed under control of a T7 promoter and synthesized (Integrated DNA Technologies). sgRNAs were transcribed using the MEGAshortScript T7 Transcription Kit (Thermo Fisher Scientific) with overnight incubation and purified using the MEGAclear Transcription Clean-Up Kit (Thermo Fisher Scientific).

In Vitro Cleavage Assays. The purified IgnaviCas9 and transcribed sgRNA were used to cleave DNA targets at desired temperatures. Templates ∼100 bp long used in the PAM determination experiments and temperature range testing were synthesized (Integrated DNA Technologies). Plasmid templates for additional temperature range testing were generated by linearizing the pwtCas9 plasmid (24) using XhoI (New England Biolabs).

IgnaviCas9 and the appropriate sgRNA were incubated together in reaction buffer at 37 °C for 10 min before the DNA target was added to the reaction. The reaction was immediately transferred to a thermocycler preset at the specified temperature and incubated for 30 min. The final composition of each reaction was 5 nM substrate DNA, 100 nM IgnaviCas9, 150 nM sgRNA, 20 mM Tris·HCl pH 7.6, 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, and 5% glycerol (vol/vol).

Each reaction was quenched using 6× Quench Buffer (15% glycerol, 100 mM EDTA) and then subjected to Proteinase K digestion at room temperature for 20 min before being loaded into a chip for fragment analysis

using a Bioanalyzer system (Agilent). The library resulting from the PAM depletion experiment in which a template containing 10-bp randomer was targeted underwent sequencing using an Illumina NextSeq 500 benchtop sequencer. Kinetic constants were calculated from time course activity data using Prism (GraphPad Software) with a one-phase exponential decay model as described previously (5, 25).

16s rRNA Depletion in Bacterial RNA-Seq Libraries. Four different sgRNAs were designed to target cDNA arising from 16s rRNA sequences. The sgRNA complexed with IgnaviCas9 as described above was added to cDNA derived from *E. coli* RNA that underwent reverse transcription and amplification using the ScriptSeq Complete Gold Kit for Epidemiology (Epicentre).

HiFi HotStart ReadyMixPCR Mix (Kapa Biosystems) was used for the combined amplification and targeted depletion reaction, which was composed of 25 μL of HiFi HotStart ReadyMixPCR Mix, 1 μL of ScriptSeq Index PCR Primer (Epicentre), 1 μL of Reverse PCR Primer (Epicentre), 1 ng of cDNA library, 2.5 μL of 5.5 μM IgnaviCas9, 15 μL of 1400 nM sgRNA, 5 μL of IgnaviCas9 reaction buffer, and water to a total volume of 50 μL. The control reaction included 25 μL of HiFi HotStart ReadyMixPCR Mix, 1 μL of ScriptSeq Index PCR Primer, 1 μL of Reverse PCR Primer, 1 ng of cDNA library, 2.2 μL of 6.2 μM SpyCas9 (New England Biolabs), 4.9 μL of 4,200 nM SpyCas9 sgRNA, 2.5 μL of Buffer 3.1 (New England Biolabs), and water to a total volume of 50 μL. The cycling protocol used was as follows: 95 °C for 3 min, 30 cycles of 98 °C for 20 s and 75 °C for 30 s, and 72 °C for 1 min.

A MiSeq (Illumina) Micro run was performed to sequence the original library and the test reaction that underwent concurrent amplification and targeted depletion. The resulting sequence reads were quality-filtered and trimmed using bbduk, aligned to the 16s rRNA sequence using bowtie2, and then sorted and indexed using samtools. Positional sequence coverage was determined using bedtools and then compared between samples by normalizing to the average whole genome coverage in each sample.

ACKNOWLEDGMENTS. We thank Emily Crawford (Chan Zuckerberg Biohub) for the plasmid used in protein expression and Richard Pfuetzner (Stanford University) for assistance with protein purification. We also thank Anastasia Nedderton (Stanford University) and National Park Service staff at Yellowstone National Park, including Research Coordinator Christie Hendrix, for assistance with different aspects of the sample collection, preservation, and characterization processes. S.T.S. is supported by the Fannie and John Hertz Foundation Fellowship, National Science Foundation Graduate Research Fellowship, Siebel Scholarship, and Gabilan Stanford Graduate Fellowship. P.C.B. was supported by the Burroughs Welcome Fund via a Career Award at the Scientific Interface. This work was supported by a Boundaries of Life program grant from the Templeton Foundation and by the Chan Zuckerberg Biohub (S.T.S., F.B.Y., and S.R.Q.).

1. M. Jinek *et al*., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
2. L. Cong *et al*., Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
3. E. V. Koonin, K. S. Makarova, F. Zhang, Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* **37**, 67–78 (2017).
4. H. Wang, M. La Russa, L. S. Qi, CRISPR/Cas9 in genome editing and beyond. *Annu. Rev. Biochem.* **85**, 227–264 (2016).
5. L. B. Harrington *et al*., A thermostable Cas9 with increased lifetime in human plasma. *Nat. Commun.* **8**, 1424 (2017).
6. I. Mougiakos *et al*., Characterizing a thermostable Cas9 for bacterial genome editing and silencing. *Nat. Commun.* **8**, 1647 (2017).
7. F. B. Yu *et al*., Microfluidic-based mini-metagenomics enables discovery of novel microbial lineages from complex environmental samples. *eLife* **6**, e26580 (2017).
8. A. Mir, A. Edraki, J. Lee, E. J. Sontheimer, Type II-C CRISPR-Cas9 biology, mechanism, and application. *ACS Chem. Biol.* **13**, 357–365 (2018).
9. D. Burstein *et al*., New CRISPR-Cas systems from uncultivated microbes. *Nature* **542**, 237–241 (2017).
10. F. A. Ran *et al*., In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186–191 (2015).
11. D. Schleheck, B. J. Tindall, R. Rosselló-Mora, A. M. Cook, Parvibaculum lavamentivorans gen. nov., sp. nov., a novel heterotroph that initiates catabolism of linear alkylbenzenesulfonate. *Int. J. Syst. Evol. Microbiol.* **54**, 1489–1497 (2004).
12. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
13. Z. Wu, H. Yang, P. Colosi, Effect of genome size on AAV vector packaging. *Mol. Ther.* **18**, 80–86 (2010).
14. F. J. Mojica, C. Díez-Villaseñor, J. García-Martínez, C. Almendros, Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
15. J. Wiktor, C. Lesterlin, D. J. Sherratt, C. Dekker, CRISPR-mediated control of the bacterial initiation of replication. *Nucleic Acids Res.* **44**, 3801–3810 (2016).
16. I. Mougiakos *et al*., Efficient genome editing of a facultative thermophile using mesophilic spCas9. *ACS Synth. Biol.* **6**, 849–861 (2017).
17. W. Gu *et al*., Depletion of Abundant Sequences by Hybridization (DASH): Using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. *Genome Biol.* **17**, 41 (2016).
18. C. Long *et al*., Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. *Science* **345**, 1184–1188 (2014).
19. S. T. Schmidt, S. M. Zimmerman, J. Wang, S. K. Kim, S. R. Quake, Quantitative analysis of synthetic cell lineage tracing using nuclease barcoding. *ACS Synth. Biol.* **6**, 936–942 (2017).
20. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
21. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
22. A. E. Briner, E. D. Henriksen, R. Barrangou, Prediction and validation of native and engineered Cas9 guide sequences. *Cold Spring Harb. Protoc.*, 10.1101/pdb.prot086785 (2016).
23. J. N. Zadeh *et al*., NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.* **32**, 170–173 (2011).
24. L. S. Qi *et al*., Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
25. S. C. Strutt, R. M. Torrez, E. Kaya, O. A. Negrete, J. A. Doudna, RNA-dependent RNA targeting by CRISPR-Cas9. *eLife* **7**, e32724 (2018).