Cas9-mediated genome editing in the methanogenic archaeon Methanosarcina acetivorans

Dipti D. Nayak* and William W. Metcalfb,1

*Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801; and bDepartment of Microbiology, University of Illinois at Urbana–Champaign, Urbana, IL 61801

Edited by Mary E. Lidstrom, University of Washington, Seattle, WA, and approved February 1, 2017 (received for review November 9, 2016)

Although Cas9-mediated genome editing has proven to be a powerful genetic tool in eukaryotes, its application in Bacteria has been limited because of inefficient targeting or repair; and its application to Archaea has yet to be reported. Here we describe the development of a Cas9-mediated genome-editing tool that allows facile genetic manipulation of the slow-growing methanogenic archaeon Methanosarcina acetivorans. Introduction of both insertions and deletions by homology-directed repair was remarkably efficient and precise, occurring at a frequency of approximately 20% relative to the transformation efficiency, with the desired mutation being found in essentially all transformants examined. Off-target activity was not observed. We also observed that multiple single-guide RNAs could be expressed in the same transcript, reducing the size of mutagenic plasmids and simultaneously simplifying their design. Cas9-mediated genome editing reduces the time needed to construct mutants by more than half (3 vs. 8 wk) and allows simultaneous construction of double mutants with high efficiency, exponentially decreasing the time needed for complex strain constructions. Furthermore, coexpression of the nonhomologous end-joining (NHEJ) machinery from the closely related archaeon, Methanocella paludicola, allowed efficient Cas9-mediated genome editing without the need for a repair template. The NHEJ-dependent mutations included deletions ranging from 75 to 2.7 kb in length, most of which appear to have occurred at regions of naturally occurring microhomology. The combination of homology-directed repair-dependent and NHEJ-dependent genome-editing tools comprises a powerful genetic system that enables facile insertion and deletion of genes, rational modification of gene expression, and testing of gene essentiality.

Significance

Methanogenic archaea play a central role in the global carbon cycle, with profound implications for climate change, yet our knowledge regarding the biology of these important organisms leaves much to be desired. A key bottleneck that hinders the study of methanogenic archaea, especially those within the genus Methanosarcina, results from the time-consuming and often cumbersome tools that are currently available for genetic analysis of these microbes. The Cas9-mediated genome-editing approach for Methanosarcina acetivorans described in this study addresses this major constraint by streamlining the mutagenic process and enabling simultaneous introduction of multiple mutations. This work also sheds light on the distinct properties of homology-dependent repair and nonhomologous end-joining machinery in Archaea.

Author contributions: D.D.N. and W.W.M. designed research; D.D.N. performed research; D.D.N. and W.W.M. analyzed data; and D.D.N. and W.W.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession no. KY486376) and the Sequenced Reads Archive (SRA) in the National Center for Biotechnology Information (accession no. PRJNA352863).

1To whom correspondence should be addressed. Email: metcalf@illinois.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1618596114/-/DCSupplemental.

Cas9 | Archaea | methanogens | Methanosarcina | genetics

The CRISPR (clustered regularly interspaced palindromic repeats) array and associated cas genes are widespread in microbial genomes (1), where they confer acquired immunity to phage and foreign DNA elements (2). The type IIA system from Streptococcus pyogenes is especially well characterized and has been widely applied as a remarkably effective genome-editing tool (3). During genome editing, heterologous expression of the RNA-guided DNA endonuclease Cas9 and a chimeric single-guide (sg) RNA, comprised of a 20-bp spacer that targets the chromosome and an 80-bp scaffold that binds Cas9, leads to a lethal double-strand break (DSB) at all target sites within the genome that are flanked by a 3′ NGG protospacer adjacent motif (PAM) (4) (Fig. S1). In eukaryotes, the nonhomologous end-joining (NHEJ) repair pathway can mend the DSB by generating simple insertions or deletions at the sgRNA target site, thus preventing additional rounds of Cas9-mediated cleavage (3, 5). Alternatively, the native homology-dependent repair (HDR) pathway can repair the fatal DSB, so long as a repair template that modifies or removes the sgRNA target site is provided, again preventing additional rounds of Cas9-mediated cleavage (3, 5) (Fig. S1). Appropriately designed repair templates allow recovery of strains with precise insertions and deletions, allowing unprecedented ability to manipulate the genomes of these diploid (or polyploid) organisms (6). Although Cas9-mediated genome editing has been successfully and broadly implemented in eukaryotes (3), similar progress has not been achieved in prokaryotes, with Cas9-mediated genome editing having been demonstrated in only 10 bacterial genera (7–10); to our knowledge, it has not been applied in archaea.

Archaea have been recognized as a phylogenetically distinct group since the 1990s (11) and it is now well-established that they are prevalent in many environments, often providing keystone ecosystem functions (12, 13). As a result, archaea play a major role in the biogeochemical cycling of nitrogen, sulfur, and carbon (13). Methanogenic archaea are particularly noteworthy from this standpoint. These microorganisms are widely distributed in strictly anaerobic environments, such as waterlogged rice paddies, sewage treatment plants, and the digestive systems of numerous animals (14), where they generate the overwhelming majority of methane released in the atmosphere. As such, it is not surprising that they have a significant impact on climate change and the global carbon cycle. Members of the genus Methanosarcina are among the most abundant and metabolically versatile methanogens known (15). They are also genetically tractable (16) and have emerged as important model organisms for genetic analysis of methanogen biology. Although the range of genetic techniques available for use in Methanosarcina is fairly comprehensive (16, 17), slow-growth and fastidious cultivation requirements have dramatically affected the pace of genetic studies within this genus.

With this in mind, we explored whether the Cas9-mediated genome-editing technique could increase the efficiency, efficacy, and speed of genetic analysis in Methanosarcina. Our results show that Cas9-mediated editing driven by the native HDR machinery in
this archaeon is extremely rapid and efficient, even when multiple mutations are simultaneously introduced. Furthermore, although Methanosarcina species do not encode a native NHEJ pathway, coexpression of the NHEJ machinery from the closely related archaeon, Methanocella paludicola, along with the Cas9–sgRNA complex, allowed robust template-independent repair.

**Results**

**Development of a Cas9-Dependent Genome-Editing System for Methanosarcina acetivorans.** To determine whether the appropriate components for genome editing from *S. pyogenes* are functional in *M. acetivorans*, we constructed a Methanosarcina/Escherichia coli shuttle vector that expresses the *S. pyogenes* Cas9 ORF from a tetracycline inducible Methanosarcina promoter (Fig. 1A). We also constructed a derivative of this plasmid that employs a methanol-inducible promoter to express an sgRNA that targets Cas9 to ssuC, a gene required for uptake of the methanogenesis inhibitor bromoethane sulfonic acid (BES) (18) (Fig. 1A and B). *M. acetivorans* was readily transformed with the Cas9-only plasmid pDN206 (78,900 ± 9,940 PurR transformants); however, pDN208, which contains the ssuC-targeting sgRNA in addition to Cas9, produced only 4 ± 3 PurR transformants. This difference in plating efficiency of more than four orders-of-magnitude strongly suggests that Cas9 is not toxic by itself, but that the Cas9–sgRNA complex from *S. pyogenes* is capable of generating a lethal DSB in *M. acetivorans*. Similar results were obtained with and without the inducers methanol and tetracycline.

Next, we determined the ability of the native HDR machinery in *M. acetivorans* to repair the lethal DSB generated by the sgRNA–Cas9 complex. To this end, repair templates of varying size were added to the ssuC-targeting vector. These repair templates generate a 34-bp deletion/frameshift mutation within ssuC that removes the targeting site while simultaneously introducing a diagnostic NotI restriction endonuclease site (Fig. 1B). Addition of repair templates with 1-kb homology arms to the plasmids relieved the lethal effect of targeting Cas9 to ssuC, generating nearly 20,000 PurR transformants per 2 μg DNA. A similar plasmid with 0.5-kbp homology arms generated roughly half as many transformants. Significantly, the 103-fold higher transformation efficiency for pDN211 relative to pDN208 indicated that 99.9% of the PurR transformants are likely to be mutants (i.e.,

---

**Fig. 1.** Cas9-mediated genome editing in *M. acetivorans*. (A) Key elements of the pDN_CRISPR plasmid series include the Cas9 ORF from *S. pyogenes* fused to the tetracycline-inducible PmcrB(teto1) promoter (in green), sgRNA(s) fused to the methanol-inducible PmtaCB1 promoter (in pink), a homology repair template (in orange), and the entire pC2A plasmid replicon containing an autonomous Methanosarcina origin of replication (in gray). The puromycin transacylase (pac) marker enables selection of puromycin resistant (PurR) transformants and the hypoxanthine phosphoribosyltransferase (hpt) marker facilitates plasmid curing by counter selection on medium containing BAPD. Note: The *E. coli* replicon and resistance marker genes have not been shown. (B) Expression of sgRNA with a 20-bp target sequence identical to a region of the WT ssuC locus (in blue) flanked by a 3′ NGG PAM (in red) with Cas9 generates a DSB at the ssuC locus. A region of the plasmid pDN211 contains a homology repair (HR) template to abolish the target site by generating a 34-bp deletion and simultaneously introducing a diagnostic NotI restriction endonuclease site in the ssuC ORF (in orange). (C) The chromosomal ssuC locus amplified from 20 PurR transformants containing pDN211 as well as the parent strain (WWM60) and subjected to restriction digest with NotI. Upon digestion, 1.1-kbp and 1.3-kbp fragments are observed for all PurR transformants (lanes 2–21), whereas a single 2.4-kbp fragment corresponding to the WT locus is observed for WWM60 (lane 22).
only 1 of every 1,000 PurR transformants would still contain the WT locus. To validate this hypothesis, 20 of these transformants were genotyped by a performing a NotI digest of a PCR amplicon containing the edited ssuC locus: all tested positive for the introduced ssuC locus. To validate the plasmid curing system, a concentration lethal to the parent strain. Genome editing was also observed when plasmids were integrated into the chromosome using a ΦC31 integrase system (17).

The initial gene-edited strains produced in these experiments retain the targeting machinery; thus, we constructed plasmid derivatives that include a counter selectable marker (hpt) to facilitate curing of gene-editing vector. This marker confers sensitivity to the purine analog 8-aza-2,6-diaminopurine (8ADP) in strains that lack the native hpt gene (19). To validate the plasmid curing system, which has not previously been attempted in Methanosarcina, we selected 8ADP<sup>R</sup> clones from three independent Pur<sup>R</sup> transformants constructed using the counter selectable vectors. All 8ADP<sup>R</sup> isolates analyzed were Pur<sup>R</sup> and also contain the frameshift mutation at the ssuC locus (Fig. S2A). PCR-based screening with plasmid-specific primers showed that the vector was indeed cured from these strains (Fig. S2B). These proof-of-principle experiments show that a Cas9-mediated genome editing technique can be used to effectively introduce unmarked mutations in <em>M. acetivorans</em>.

**Optimization of the Cas9-Dependent Genome-Editing Technique in <em>M. acetivorans</em>.** To determine the optimal expression levels for the genome-editing machinery, we varied the transcription of Cas9 by selecting transformants on media with increasing concentrations of tetracycline, and of the sgRNA by plating on media with either methanol (induced) or trimethylamine (TMA; repressed) as growth substrates. Surprisingly, no significant difference in genome-editing efficiency was observed (Fig. 2A). In fact, the basal level of transcription provided by the two promoters in the absence of the inducers was sufficient for effective editing (Fig. 2A). A control vector identical to pDN211 but lacking the sgRNA (pDN207) was used to estimate the efficiency of genome editing. The efficiency of genome editing was measured as the ratio of mutant recovery (i.e., plating efficiency of pDN211 relative to the plating efficiency of the control vector and was estimated on media with either methanol or TMA as growth substrates. Significantly, genome editing in these experiments was particularly efficient, with edited strains being obtained at frequencies of approximately 20–25% relative to the control (i.e., one in four cells that receive the plasmid undergo gene conversion) (Fig. 2B).

To examine the maximum size of deletions that can be reliably generated by a single sgRNA, we tested repair templates with 1-kb homology arms placed at varying distance from the sgRNA-directed DSB (Fig. 2C). The transformation efficiency remained steady for templates that are ≤250 bp away from each end of the DSB, but declined precipitously when the distance increased beyond this point (Fig. 2C). Thus, a single sgRNA can be reliably used to delete up to 0.5 kbp of the chromosome, although larger deletions (up to 1 kbp) can be produced at the expense of efficiency.

**Multiplex Expression of sgRNAs in <em>M. acetivorans</em> Enables Simultaneous Introduction of Multiple Mutations.** To explore the possibility of using multiple Cas9-mediated DSBs to create larger deletions, or to simultaneously introduce more than one mutation, we tested two alternate arrangements for the expression of multiple sgRNAs. In the first arrangement, sgRNAs were expressed individually, whereas in the second they were expressed as a single transcript separated by a 30-bp linker sequence (Fig. 3A). Plasmids with sgRNAs in either arrangement were equally efficient in generating strains with complete deletions (approximately 2 kbp) of the <em>mtmCB1</em> and <em>mtmCB2</em> loci, which are highly homologous genes encoding monomethylamine methyltransferase isozymes (Fig. 3B and Fig. S3). Thus, to reduce the size of mutagenic plasmids and
Independent Repair in M. acetivorans

Heterologous Expression of NHEJ Genes Leads to Template-Independent Repair in M. acetivorans. A lethal phenotype for plasmids expressing a Cas9-sgRNA complex in the absence of a repair template was uniformly observed across a wide range of sgRNAs tested in this study, suggesting that NHEJ does not occur in M. acetivorans (Table 1). This result is consistent with the absence of genes related to Ku and LigD in the completely sequenced genome (22). Nevertheless, in some circumstances HDR-independent gene editing would be very useful. Therefore, we examined whether NHEJ could be established in Methanosarcina for use in conjunction with the Cas9-sgRNA complex. For this purpose, we chose the NHEJ machinery from the closely related methanogen M. paludicola, which has previously been reconstituted in vitro (23, 24). An artificial operon encoding four M. paludicola NHEJ proteins [DNA ligase (Lig), poly-merase (Pol), phosphoesterase (PE), and Ku] was synthesized and transcriptionally fused to the moderately expressed sorC promoter (25) to allow transcription in M. acetivorans (Fig. S7). This cassette was then added to the Cas9 ssuC-targeting vector without a repair template. Transformation with this plasmid was approximately 100-fold less efficient than the corresponding HDR vector, but approximately 10-fold higher than with plasmids lacking the NHEJ system (Fig. 4A). Therefore, expression of the M. paludicola NHEJ machinery overcame the lethal effect of the Cas9 ssuC-targeting vector without a repair template. Molecular analysis of the ssuC locus in these transformants revealed deletions ranging from 75 bp to 2.7 kb in length, often occurring at naturally occurring regions of microhomology 6–11 bp in length (Fig. 4B). Thus, the combined Cas9/NHEJ system provides the opportunity to generate a variety of mutations surrounding a single target site. Importantly, these plasmids are much simpler to construct, requiring only addition of target-specific sgRNA. We also tested whether addition of two sgRNAs targeting DNA sequences approximately 450 bp apart in conjunction with NHEJ could be used to generate precise deletions without a repair template. Interestingly, attempts to construct ssuC deletions via this method were not successful: only a handful of colonies were obtained (6 ± 3 per 2 µg plasmid) and none had the precise deletion desired. We examined 20 transformants obtained by this method. Two contained a 1.3-kb deletion of the ssuC locus, which occurred at a region of microhomology (Fig. 4B). The remainder had WT copies of the ssuC gene and, thus, are likely to be so-called escape mutants in which the Cas9 gene or sgRNA has mutated on the targeting plasmid (26).

Discussion

The Cas9-based tools developed in this study will have a transformative impact on the speed, scope, and scale of research that can be accomplished in the methanogenic archaeon, M. acetivorans.
Most notably, multiplexed gene-editing plasmids will enable generation of strains with multiple mutations, ranging from SNPs to large indels, in a matter of weeks versus years. These tools will enable researchers to swiftly tag genes at their native loci on the host chromosome, allowing the study of context-specific gene expression, “pull-down” experiments to establish protein–protein and protein–DNA interaction networks, and purification of proteins that contain unique amino acids (27) or novel posttranslational modifications (28). Furthermore, deleting a gene of interest using the NHEJ-based technique is very cost-effective, as it simply requires the insertion of a commercially synthesized DNA fragment containing the appropriate sgRNAs into pDN243, the vector containing Cas9 and the NHEJ machinery. Thus, studies that were previously inconceivable, such as constructing a library of strains with single-gene deletions in every nonessential gene on the M. acetivorans chromosome [as described in Wang et al. (29)], will now become feasible, in terms of both time and cost. Finally, we expect that minor modifications will enable the application of this approach to a broad range of methanogens and other archaea.

Certain features of Cas9-mediated genome editing in M. acetivorans are particularly unique and noteworthy. For example, unlike eukaryotes (30), targeting of the Cas9–sgRNA complex to a particular chromosomal region in M. acetivorans is remarkably precise, and no off-target activity was observed upon resequencing multiple, independently genome-edited mutants (Table S1). Because the M. acetivorans genome (approximately 5.75 Mbp) is 10- to 100-fold smaller in comparison with eukaryotic genomes, it is possible that fewer off-target sites are present. However, no off-target activity could be detected, despite our intentional choice of highly similar sgRNA targets in the mtmcB isozymes (Fig. S3). Thus, it is likely that properties of the Cas9–sgRNA complex, including target specificity, vary significantly across domains of life, perhaps because of differences in chromosomal organization and DNA repair machinery. Notably, unlike the Cas9-mediated genome editing in bacteria (7–9), we observe a high rate of HDR for the Cas9-mediated DSB and a very low frequency of “escape” mutants. These key distinctions are likely to stem from evolutionarily distinct HDR machinery. Archaeal DNA repair involves homologs of the eukaryotic proteins Mre11 and Rad50, and two other unique proteins HerA and NurA, which perform end-resection after a DSB occurs (31). Subsequently, the RecA orthologs RadA and RadE homologously recombine with the target DNA, and Ku homologs interact with eukaryotes, mediate strand invasion (31). Finally, Hjc, unrelated to the RuvABC complex in bacteria (32), is involved in the resolution of the Holliday junction (31). Thus, it is tempting to speculate that coexpression of archaeal HDR machinery along with Cas9 might overcome some of the obstacles that have been inferred from the regions of microhomology at which repair occurred (Fig. 4B). Moreover, DNA repair mediated by MMEJ is almost completely abolished when two sgRNAs were simultaneously expressed, suggesting that the repair mechanism has the ability to distinguish breaks that occur at discrete loci.

Finally, one might ask why we chose to use the well-established S. pyogenes Cas9–sgRNA complex for genome-editing purposes over the native type I or type III CRISPR/Cas systems that are commonly found in Methanosarcina spp. (35), as was done in Sulfolobus islandicus (36). First, the CRISPR/Cas subtypes vary significantly across the genus Methanosarcina, even within strains belonging to the same species (35). Hence a genome-editing technique reliant on the native CRISPR/Cas machinery for one strain might not work in other closely related strains. Recent studies across a wide-range of bacteria have revealed that anti-CRISPR proteins to silence the native CRISPR/Cas system are also often encoded on the chromosome (37). Although no anti-CRISPR proteins have been detected in Methanosarcina, it is possible that they exist and might potentially complicate use of the native CRISPR/Cas machinery for genome editing. Finally, tweaking the native CRISPR/Cas machinery for genome editing purposes is likely to impact organismal physiology in an unpredictable fashion and skew genetic analyses downstream. Thus, we chose to deploy the simple, modular Cas9-mediated genome editing machinery on a vector that will be transiently maintained in M. acetivorans.

Materials and Methods

Strains, Media, and Growth Conditions. All chemicals were purchased from Sigma-Aldrich unless otherwise specified. M. acetivorans strains were grown in single-cell morphology (38) at 37 °C in bicarbonate-buffered high-salt (HS) liquid medium containing 125 mM NaCl and 30 mM TMA hydrochloride in Balch tubes with Na2CO3 (80/20). Plating solid medium was conducted in an anaerobic glove chamber (Coy Laboratory Products) as described previously (9). Solid media were incubated anaerobically at 37 °C for 10 days. A gblocks library (R. I. Chemicals) and BES were added to a final concentration of 2 μg/ml, 20 μg/ml, 0.4 μm, respectively, from sterile, anaerobic stock solutions. Anaerobic, sterile stocks of tetracycline hydrochloride in deionized water were prepared fresh shortly before use and added to a final concentration as indicated. E. coli strains were grown in LB broth at 37 °C with standard antibiotic concentrations. WM4489, a DH10B derivative engineered to control copy-number of oriV-based plasmids (40), was used as the host strain for all plasmids generated in this study (Table S2). Plasmid copy number was increased by adding sterile rhamnose to a final concentration of 10 mM.

Plasmids. All plasmids used in this study are listed in Table S2. The plasmid pMJ0806 was obtained from Jennifer Doudna, University of California, Berkeley, CA (Addgene plasmid # 39312). The S. pyogenes (Spy) Cas9 ORF was fused to the P tetO1 promoter in plasmid pJK027A (17), and linearized with Ndel and HindIII by the Gibson assembly method, as described previously (41). The DNA segments containing sgRNA flanked by putative mtc81 promoter and terminator sequences from M. acetivorans were synthesized as double-stranded DNA fragments (“gblocks”) from integrated DNA Technologies and used for cloning purposes per the manufacturer’s instructions. A 3.25-kbp artificial operon with the NHEJ polymerase (Mcp_2125), DNA ligase (Mcp_2126), phosphoesterase (Mcp_2127), and Ku (Mcp_0581) genes from Methanococcus paludicola was ordered from the GeneArt gene synthesis service (Life Technologies). All synthetic DNA fragments and repair templates were introduced in the appropriate pJK027A-derived vector (carrying the appropriate vector backbone linearized with either AscI or PmeI by the Gibson assembly method, as described previously (41). The entire pC2A plasmid was introduced in the appropriate pJK027A-derived vector (carrying the αattB site) by retrofusing with pAMG40 (carrying the αattP site) using the BP Clonase II master mix (Invitrogen) per the manufacturer’s instructions. WM4489 was transformed by electroporation at 1.8 kV using an E. coli Gene Pulser (Bio-Rad). Standard techniques were used for the isolation and manipulation of plasmid DNA. All pJK027A-derived plasmids were verified by Sanger sequencing at the Roy J. Carver Biotechnology Center, University of Illinois at Urbana–Champaign, and all pAMG40 cointegrates were verified by restriction endonuclease analysis. Primers used in this study are listed in
Table S3. The plasmid sequence and annotations for pDN211 have been submitted to GenBank (accession no. KU436376).

In Silico Design of Target Sequences. All target sequences used in this study are listed in Table S4. Target sequences were designed using the CRISPR site finder tool in Geneious version R9 (42). The Transformation of M. acetivorans. All M. acetivorans strains used in this study are listed in Table S5. Liposome-mediated transformation was used for M. acetivorans, as described previously (43), and 10 ml of late-exponential phase culture of M. acetivorans and 2 μg of plasmid DNA were used for each transformation.

Genome Sequencing and Analysis. Genomic DNA from M. acetivorans was extracted using a protocol described previously (44). DNA libraries were prepared with the Hyper Library construction kit (Kapa Biosystems) and quantified using qPCR. All libraries were sequenced on one lane of an Illumina MiSeq v2 (Illumina) at the Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign using a 500 cycles v2 sequencing kit (Illumina). Trimmed, paired end 250 nt reads were mapped to the M. acetivorans reference genome (NC_003552) using default parameters for bresseq v0.25 (45). Trimmed genome sequencing reads have been deposited in the Sequenced Reads Archive at the National Center for Biotechnology Information under accession no. PRJNA352863.

ACKNOWLEDGMENTS. We thank Dr. Mary Elizabeth Metcalf for technical assistance. This work was supported in part by the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the US Department of Energy Grant DE-FG02-02ER15296 (to W.W.M.), and the Carl R. Woese Institute for Genomic Biology postdoctoral fellowship (to D.D.N.). D.D.N. is currently a Simons Foundation fellow of the Life Sciences Research Foundation.