Systematic evasion of the restriction-modification barrier in bacteria

Christopher D. Johnston[a,b,c,1], Sean L. Cotton[b], Susan R. Rittling[b,c], Jacqueline R. Starr[b,c], Gary G. Borisy[b,c,1], Floyd E. Dewhirst[b,c], and Katherine P. Lemon[b,d]

[a]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109; [b]The Forsyth Institute, Cambridge, MA 02142; [c]Harvard School of Dental Medicine, Boston, MA 02115; and [d]Division of Infectious Diseases, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115

Contributed by Gary G. Borisy, April 9, 2019 (sent for review November 28, 2018; reviewed by James J. Collins and George A. O'Toole)

Bacteria that are recalcitrant to genetic manipulation using modern in vitro techniques are termed genetically intractable. Genetic intractability is a fundamental barrier to progress that hinders basic, synthetic, and translational microbiology research and development beyond a few model organisms. The most common underlying causes of genetic intractability are restriction-modification (RM) systems, ubiquitous defense mechanisms against xenogeneic DNA that hinder the use of genetic approaches in the vast majority of bacteria and exhibit strain-level variation. Here, we describe a systematic approach to overcome RM systems. Our approach was inspired by a simple hypothesis: if a synthetic piece of DNA lacks the highly specific target recognition motifs for a host's RM systems, then it is invisible to these systems and will not be degraded during artificial transformation. Accordingly, in this process, we determine the genome and methylome of an individual bacterial strain and use this information to define the bacterium's RM target motifs. We then synonymously eliminate RM targets from the nucleotide sequence of a genetic tool in silico, synthesize an RM-silent "SyngenicDNA" tool, and propagate the tool as minicircle plasmids, termed SyMPL (SyngenicDNA Minicircle Plasmid) tools, before transformation. In a proof-of-principle of our approach, we demonstrate a profound improvement (five orders of magnitude) in the transformation of a clinically relevant USA300 strain of *Staphylococcus aureus*. This stealth-by-engineering SyngenicDNA approach is effective, flexible, and we expect in future applications could enable microbial genetics free of the restraints of restriction-modification barriers.

restriction modification | genetic intractability | SyngenicDNA | minicircle | transformation

Genetic engineering is a powerful approach for harnessing bacterial abilities and for discovering fundamental aspects of bacterial function. In recent years, the genetic toolkit at our disposal has massively expanded (1, 2). The application of these tools is largely limited to bacterial strains with high transformation efficiency (3). However, relative to the wealth and diversity of known bacterial species, there are currently only a small number of such highly genetically tractable strains. A strain that is not amenable to alterations of its genome or to the introduction of new genetic information during genetic engineering is termed genetically intractable. At present, genetic intractability is a pervasive and widespread problem across all fields of microbiology; the vast majority of bacteria that can be grown in a laboratory remain beyond the power of genetics for elucidating function or engineering for human use. Even within species that are genetically tractable, this tractability is often restricted to a small number of domesticated strains, while new primary isolates of the species with disparate phenotypic traits of interest are either poorly tractable or currently intractable. As a result, researchers have had to engage in expensive generation of ad hoc genetic systems for each distinct species, often with further laborious modifications for each distinct wild strain isolate.

In their natural environment, bacteria acquire new genetic information through horizontal gene transfer (HGT) by three distinct means: conjugation, transduction, and transformation. During conjugation, DNA is transferred from one organism to another by direct cell-to-cell contact. During transduction, DNA is carried by bacteriophages, viruses that invade by injecting DNA into host bacterial cells. These two processes involve multifaceted interactions requiring complex machinery and therefore are of limited value in modern bacterial genetics where DNA should ideally be easily and rapidly transferable into any given bacterial strain (4). During transformation, however, naked DNA is directly acquired and incorporated into the host genome by recombination with homologous sequences or, in the case of plasmids, by establishing a new episome (extrachromosomal DNA that replicates autonomously), resulting in genetic alteration of the cell. Genetic competence is the cellular state that enables bacteria to undergo natural transformation, a transient "window of opportunity" for DNA internalization (5). However, while there are over 6,600 validated cultured type strains of bacterial species (6) and ~30,000 formally named species in pure culture (7), natural transformation and competence have been observed in only a handful of ~80

Significance

Genetic engineering is a powerful approach for discovering fundamental aspects of bacterial physiology, metabolism, and pathogenesis as well as for harnessing the capabilities of bacteria for human use. However, the full power of genetic engineering can only be applied to a few model organisms. Biological diversity and strain-level variation in restriction-modification systems are critical barriers keeping most bacteria beyond the full potential of genetics. We have designed a systematic approach to effectively evade restriction-modification systems and successfully applied this approach to a clinically relevant USA300 strain of the human pathogen *Staphylococcus aureus*. Our results demonstrate the simplicity and effectiveness of this stealth-by-engineering approach, which could enable microbial genetic system design not restrained by innate restriction-modification defense mechanisms.

Author contributions: C.D.J., G.G.B., F.E.D., and K.P.L. designed research; C.D.J. and S.L.C. performed research; C.D.J., S.R.R., and J.R.S. analyzed data; and C.D.J. and K.P.L. wrote the paper.

Reviewers: J.J.C., Massachusetts Institute of Technology; and G.A.O., Geisel School of Medicine at Dartmouth.

Conflict of interest statement: C.D.J. discloses that he has filed and is inventor on pending patent applications (USSN: 62/408,693 and 62/802,016) entitled "Compositions and methods for evading bacterial defense mechanisms" and "Production of differentially methylated DNA in *E. coli*," respectively, relating to the SyngenicDNA and SyMPL methodologies developed and applied in this paper.

Data deposition: Genome/methylome sequence data for *Escherichia coli* MC_Forsyth and *Staphylococcus aureus* USA300 JE2_Forsyth have been deposited to the REBASE database, http://rebase.neb.com/rebase/rebase.html (under strain nos. 21741 and 21742, respectively).

[1]To whom correspondence may be addressed. Email: johnston@fredhutch.org or gborisy@forsyth.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1820256116/-/DCSupplemental.

Published online May 16, 2019.

bacterial species (5). This may even be an overestimation, as in several cases only a single report documents transformation, and molecular evidence of natural transformation is lacking (5). For all remaining cultivated bacterial species that are of interest, microbiologists must instead develop "artificial" transformation and individualized genetic systems, often at the strain level: a process continually stymied by genetically intractable phenotypes.

Restriction-modification (RM) systems are the most common underlying cause of genetic intractability in bacterial species. Found in ∼90% of sequenced bacterial genomes, RM systems enable bacteria to distinguish self- from nonself-DNA via two enzymatic activities: a restriction endonuclease (REase) and a modification methyltransferase (MTase). The REase recognizes the methylation status of DNA at a highly specific DNA target sequence and degrades unmethylated or inappropriately methylated (i.e., nonself) targets. Its cognate MTase protects the same target sequence across the host's genome via addition of a methyl group, marking each site as self. RM systems originally evolved as defense mechanisms against invading xenogeneic DNA (8), which is primarily encountered during bacteriophage infection. Consequently, most, if not all, of the currently available approaches to overcome them during genetic engineering are inspired by bacteriophage antirestriction mechanisms (9, 10). Bacteriophage mechanisms that involve methyl-modification of the phage genome to subvert the host's RM activities have already been translated into in vitro engineering approaches (9, 10). These can all be referred to as mimicry-by-methylation, as they essentially seek to modify the methylation pattern of a genetic tool to match the desired host and achieve molecular mimicry. There are two common mimicry-by-methylation approaches. (*i*) Methylate target sites on tools by using in vitro methylation with recombinant MTase enzymes (10), which are currently commercially available for only 37 of >450 known targets (11). (*ii*) Alternatively, achieve in vivo methylation by passaging a plasmid through a related strain that is either restriction enzyme-deficient (10) or a surrogate strain that has been extensively engineered to match the methylation profile of the strain of interest, i.e., the plasmid artificial modification (PAM) technique (12). Although these are very effective in some cases (13), owing to the labor-intensive and rigid nature of their underlying design they are often not readily adaptable to other strains due to RM system diversity (*SI Appendix*, Text S1) and, accordingly, are unsuitable for rapid application to a wide diversity of bacteria (10).

We therefore sought to design a versatile strategy to overcome RM barriers, one suitable for use in a broad range of bacterial species. The problem to be overcome is that in any given bacterial genus of interest the number of RM systems present and the target sequences recognized are hypervariable and highly species-specific, often even strain-specific (14). RM systems are also extremely diverse and can be differentiated into four types (type I, II, III, and IV), based on their recognized target and, also, subunit composition, cleavage position, cofactor requirements, and substrate specificity (15). Additionally, RM target motifs themselves vary greatly in sequence and length, ranging from 4 to 18 base pairs (bp), with >450 different motifs identified to date (11). It is clear, therefore, that a broadly applicable strategy to overcome RM barriers to genetic engineering will need to be adept at adjusting for RM system variation across different bacterial strains.

Importantly, all REase enzymes demonstrate exquisite specificity in target sequence recognition. This specificity is crucial, as REases are toxic to their hosts' genome in the absence of their cognate MTases and, consequently, seldom deviate from their recognition sequence (15). In the context of bacterial genetic engineering, this is a critical weakness underpinning the effectiveness of all RM systems. To rapidly exploit the inherent weakness of high target specificity, we designed a stealth-based strategy to evade RM system activities entirely. Our approach was inspired by a simple hypothesis: if a piece of DNA lacks the highly specific target recognition motifs for a host's RM systems, then it is invisible to these systems and will not be degraded upon artificial transformation. As RM defenses recognize genetic tools as xenogeneic DNA by virtue of the methylation status of highly specific target motifs (8), the systematic identification and elimination of such target motifs from the nucleotide sequence of a genetic tool should therefore facilitate the engineering of an artificial syngeneic DNA molecule that is RM-silent upon transformation. To succinctly encapsulate our approach, we coined the term "SyngenicDNA" (*SI Appendix*, Text S2).

One example of the tremendous effort, resources, and time it takes to expand genetic tractability is *Staphylococcus aureus*, a pathogen with significant relevance to public health, which accounts for over 10,000 deaths per year in the United States (16). Numerous papers describe mimicry-by-methylation approaches that seek to expand tractability to more clinically relevant strains, (e.g., refs. 17 and 18). Here, based on its public health importance, we selected *S. aureus* JE2, a derivative of the epidemic methicillin-resistant *S. aureus* (MRSA) USA300 LAC strain (19) to demonstrate proof-of-principle for our stealth-based approaches. We expect these approaches will be adopted by the broader microbiological community, enabling genetic system design no longer restrained by microbial restriction-modification defense mechanisms.

Results

Systematic Generation of SyngenicDNA-Based Genetic Tools. There are four basic steps to produce SyngenicDNA-based genetic tools (Fig. 1): (*i*) target identification, (*ii*) in silico tool assembly, (*iii*) in silico sequence adaptation, and (*iv*) DNA synthesis and assembly. Target identification requires the delineation of each methylated site, with single-base resolution, across an entire bacterial genome (i.e., the methylome) and starts with single-molecule real-time (SMRT) genome and methylome sequencing (SMRTseq) (14). Using methylome data, we delineate each of the recognition motifs protected by the MTases of the host's RM systems and infer the targets recognized and degraded by their cognate REases (*SI Appendix*, Text S3). This yields a concise list of a host microbes' RM targets to be eliminated from the DNA sequence of a selected genetic tool.

In silico tool assembly requires complete annotation of a genetic tool's sequence with respect to plasmid chassis, replication origins, antibiotic-resistance cassettes, promoters, repressors, terminators, and functional domains to avoid adverse changes to these structures during subsequent adaptation steps. Ideally, a complete and minimalistic genetic tool with previous demonstrable functionality in a genetically tractable strain is used for initial experiments, allowing for subsequent addition of DNA parts to increase functionality after successful transformation is achieved.

In silico sequence adaptation of the genetic tool is the most crucial step of the SyngenicDNA approach, and it is here where we exploit the intrinsic evolutionary weakness of high target-sequence specificity present in all RM systems. In this step, we first screen the complete nucleotide sequence of the genetic tool for the presence of RM targets identified by SMRTseq. We then recode the nucleotides of each RM target in silico to eliminate the target while preserving the functionality of the sequence. In noncoding regions, targets are removed, changing a single nucleotide (creating a SNP). In coding regions, the sequence of the target is removed using synonymous codon substitution. A single-nucleotide switch is generally sufficient to remove RM targets, but multiple switches can also be used. The preferential codon bias of the desired host is used to avoid introducing rare or unfavorable codons during the synonymous switch (*SI Appendix*, Text S4). Upon complete removal of all RM targets in silico, the recoded DNA sequence has been rendered RM-silent with respect to the host, termed SyngenicDNA, and is ready for de novo DNA synthesis.

Synthesis and assembly of RM-silent genetic tools is carried out using commercially available de novo DNA synthesis and standard assembly approaches, ensuring that any laboratory can construct SyngenicDNA tools. In commercial DNA synthesis, sequences are typically inserted into an *Escherichia coli* plasmid replicon and propagated to yield large amounts of the synthetic DNA. This *E. coli* replicon is convenient, but might include RM targets that could lead to degradation of the overall circular tool

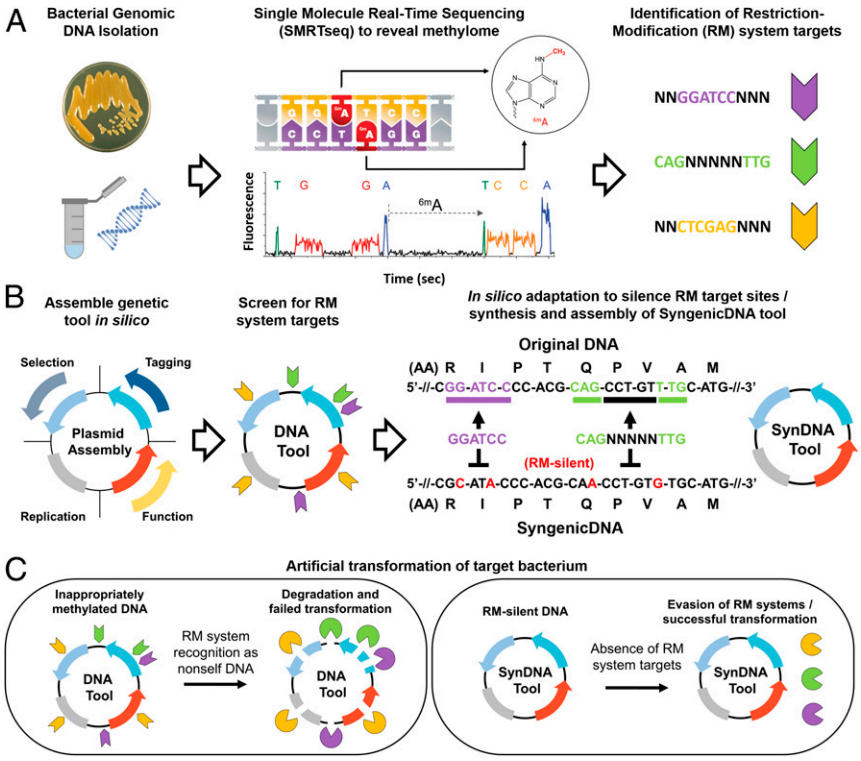

Fig. 1. Schematic representation of the SyngenicDNA approach. (*A*) Identification of RM system target motifs by SMRTseq. Methylome analysis of polymerase kinetics during sequencing permits detection of methylated sites at single-nucleotide resolution across the genome, revealing the exact motifs targeted by innate RM systems (indicated by colored nucleotides; N is any nucleotide) (kinetic trace image adapted from www.pacb.com). (*B*) Assembly in silico of a genetic tool with a desired functionality, followed by screening for the presence of RM target sequences and sequence adaptation, using SNPs or synonymous codon substitutions in coding regions, to create an RM-silent template which is synthetized de novo to assemble a SyngenicDNA tool. (*C*) Artificial transformation of the bacterium of interest target bacterium. Inappropriately methylated target motifs of the original genetic tool are recognized as nonself-DNA and degraded by RM systems. In contrast, the SyngenicDNA variant retains the form and functionality of the genetic tool, but is uniquely designed at the nucleotide level to evade the RM systems and can operate as desired within the target bacterial host.

after transformation into the host strain. We have developed two solutions to this potential issue. One solution is to generate a SyngenicDNA *E. coli* plasmid backbone for each specific microbial host strain (Fig. 1*B*). However, in routine applications this will increase costs of SyngenicDNA synthesis, and, moreover, the *E. coli* replicon itself becomes redundant after propagation in *E. coli*, as it is typically nonfunctional in other bacterial species after transformation. Our alternative solution, therefore, is to remove the *E. coli* replicon entirely, using minicircle DNA technology, rather than recode it. This approach also increases flexibility because the same *E. coli* replicon can be used to generate tools for multiple different microbial strains.

SyngenicDNA Minicircle Plasmid Tools. Minicircles (MCs) are minimalistic circular expression cassettes devoid of a plasmid backbone (20), which are primarily used in gene therapy applications to drive stable expression of transgenes in eukaryotic hosts. MCs are produced by attaching a parental plasmid (PP) to a transgene cassette, cultivating this construct in an *E. coli* host grown to high-cell density, inducing construct recombination to form an isolated transgene MC and a separate, automatically degraded, PP containing the *E. coli* replicon. MCs are then isolated by standard plasmid methods (20). Because any DNA sequence can take the place of the transgene, we hypothesized that MC technology could be repurposed to carry entire microbial plasmids and facilitate the removal of superfluous *E. coli* replicons from shuttle vectors. We demonstrated that the incorporation of SyngenicDNA sequences into a PP allowed us to create SyngenicDNA Minicircle Plasmid (SyMPL, pronounced "simple") tools (*SI Appendix*, Fig. S1). SyMPL tools include replication, selection, and functional domains for operation in a specific non-*E. coli* host, but lack an *E. coli* replicon despite being isolated at high concentrations from the MC-producing *E. coli* strain. In our SyMPL strategy, we attach a synthesized (and assembled) SyngenicDNA tool to the nonSyngenicDNA *E. coli* PP and propagate this construct in an MC-producing *E. coli* strain. The induction of MCs via recombination, with concurrent induction of a specific endonuclease that eliminates the PP, allows for easy isolation of a minimalistic SyngenicDNA-based genetic tool ready to transform into the desired host strain (*SI Appendix*, Fig. S1*C*).

The majority of laboratory *E. coli* strains, including the MC-producing *E. coli* host used in this study, contain three active MTases (Dam, Dcm, and HsdM) that introduce methylation modifications to specific target sites on the host genome (*SI Appendix*, Fig. S2). The Dam MTase modifies the adenine residue (^{m6}A) within the sequence GATC, the Dcm MTase modifies the internal cytosine residue (^{m5}C) of the sequence CCWGG (where W is A or T), and the HsdM MTase modifies the internal adenine residue (^{m6}A) of the sequence AACN$_6$GTGC. Therefore, plasmid tools propagated within such *E. coli* strains, including the minicircle (MC) producing strain (ZYCY10P3S2T), are modified at these target sequences.

The presence of methylated sites on SyngenicDNA-based tools could activate type IV RM systems upon artificial transformation. Generally, unintentional activation of methyl-targeting type IV systems is avoided by the propagation of plasmids within methyl-deficient *E. coli* strains such as JM110 (*dam*-, *dcm*-, *hsdRMS*+) or ER2796 (*dam*-, *dcm*-, *hsdRMS*-), thus preventing recognition and degradation via these systems. However, such methyl-free *E. coli* strains are unable to produce MCs since construction of the *E. coli* MC-producing strain (20) required complex engineering to stably express a set of inducible minicircle-assembly enzymes (the øC31-integrase and the I-SceI homing-endonuclease for induction of MC formation and degradation of the parental plasmid replicon, respectively).

Accordingly, when we repurposed MC technology for bacterial applications, it was also necessary to engineer *E. coli* MC-producer strains that generate various forms of methylation-free MCs (*SI Appendix*, Figs. S3–S5). Although a completely methylation-free MC producer could be required when working against type IV systems targeting both adenine- and cytosine-methylated DNA, bacterial RM systems exist with targets that specifically match the *E. coli* Dam MTase motif (GATC), such as Pin25611FII in *Prevotella intermedia* (14). These systems digest unmethylated Dam sites on genetic tools propagated within methyl-free strains; hence, Dam methylation is protective in these cases. Therefore, we created a suite of *E. coli* strains capable of producing distinct types of methyl-free MC DNA to account for the inherent variation of RM systems in bacteria and

maximize the applicability of our SyMPL approach. We applied iterative CRISPR-Cas9 genome editing to sequentially delete MTase genes from the original *E. coli* MC-producer strain (*dam*+, *dcm*+, *hsdM*+) (*SI Appendix*, Fig. S4). These strains produce methylcytosine-free MC DNA (*E. coli* JMC1; *dam*+, *dcm*-, *hsdM*+), methylcytosine- and methyladenine-free MC DNA except for Dam methylation (*E. coli* JMC2; *dam*+, *dcm*-, *hsdM*-), and completely methyl-free MC DNA (*E. coli* JMC3; *dam*-, *dcm*-, *hsdM*-). Depending upon the type IV RM systems identified within a desired bacterial host, one of these strains can be selected and utilized for production of SyMPL tools.

Application of SyngenicDNA and SyMPL Approaches to a Bacterial Pathogen. RM systems are a known critical barrier to genetic engineering in most strains of *Staphylococcus aureus* (21). Based on its public health importance, we selected *S. aureus* JE2, a derivative of the epidemic methicillin-resistant *S. aureus* (MRSA) USA300 LAC strain (19), to demonstrate the efficacy of our stealth-by-engineering approaches. First, we determined the methylome of JE2 using SMRT sequencing and identified this strain's RM targets. SMRTseq and REBASE analysis of JE2 confirmed the presence of two type-I RM systems recognizing the bipartite target sequences AGGN$_5$GAT and CCAYN$_6$TGT (the modified base within each motif is underlined, and n = any base) (*SI Appendix*, Table S1) and a type-IV system, previously shown to target cytosine methylation within the sequence SCNGS (where S = C or G) (21).

We then applied our SyngenicDNA approach to the *E. coli*–*S. aureus* shuttle vector pEPSA5 (Fig. 2 *A* and *B*). The pEPSA5 plasmid (*SI Appendix*, Text S4 and Fig. S1) contains a 2.5-kb *E. coli* replicon (ampicillin-resistance gene with a p15a origin for autonomous replication) and a 4.3-kb *S. aureus* replicon (chloramphenicol-resistance gene, pC194-derived origin, and a xylose repressor protein gene, *xylR*) (*SI Appendix*, Fig. S6*A*). The *S. aureus* replicon is nonfunctional when pEPSA5 is maintained and propagated within *E. coli*, and vice versa. Therefore, we modified *S. aureus* JE2 RM targets occurring within the coding region of the pEPSA5 *E. coli* replicon with synonymous substitutions adhering to *E. coli* codon bias. We synthesized, assembled, and propagated pEPSA5SynJE2 (Fig. 2*C*), a variant of pEPSA5 that differed by only six nucleotides (99.91% identical at nucleotide level), eliminating three RM target motifs present in the original sequence. We demonstrated an ~70,000-fold ($P = 7.76 \times 10^{-306}$) increase in transformation efficiency (cfu/μg DNA), using the entirely RM-silent pEPSA5SynJE2Dcm- (propagated in *dcm- E. coli*), compared with the original pEPSA5 plasmid (propagated in *dcm*+ *E. coli*) (Fig. 2*D* and *SI Appendix*, Text S5).

Subsequently, we sought to determine whether a further increase in transformation efficiency could be achieved using the SyMPL (minicircle) approach. We used the *dcm*- strains *E. coli* ER2796 and *E. coli* JMC1 to carry out the minicircle (MC) experiments independently of the type IV system in *S. aureus* JE2. We generated a SyngenicDNA pEPSA5 minicircle for JE2 (pEPSA5SynJE2MC); 38% smaller than pEPSA5 and free of the original *E. coli* replicon (Fig. 3A and *SI Appendix*, Fig. S7).

Most of the *S. aureus* JE2 RM system targets present on pEPSA5 are in the *E. coli* replicon (type I: $n = 2$, and type IV: $n = 8$) with only a single type I target in the *S. aureus* replicon (*SI Appendix*, Fig. S6*A*); thus the MC approach eliminates two of three type I targets. We investigated (*i*) whether the SyMPL approach achieves equal or perhaps even greater efficiency than the SyngenicDNA approach and (*ii*) whether removal of all type I targets is required to achieve appreciable gains in transformation efficiency (compared with a partially SyngenicDNA plasmid that has a single type I target remaining). The original plasmid pEPSA5 (Dcm+) was included in experiments only as a control for accurate final comparison of efficiencies and was not considered a primary comparison. The pEPSA5SynJE2MC variant achieved ~2×10^7 transformants per microgram DNA, a further 3.5-fold increase ($P = 1.78 \times 10^{-9}$) in transformation efficiency over pEPSA5SynJE2 and a >100,000-fold increase ($P = 1.97 \times 10^{-284}$) compared with the original unmodified pEPSA5 plasmid (propagated in *dcm*+ *E. coli*) (Fig. 3*B* and *SI Appendix*, Tables S2 and S3).

In SyMPL experiments, by reducing the overall size of MC plasmids, we also increased the number of *S. aureus* replicons present within the micrograms of DNA used for transformations compared with the micrograms used for full-length plasmids. Increasing the yield of functional replicons per microgram of DNA might be an additional advantage of the MC approach. Thus, to more accurately compare transformation efficiencies between MCs and full-length plasmids, we performed a secondary analysis to adjust the transformation efficiencies from cfu/μg DNA to cfu/pmol DNA (Fig. 3*C* and *SI Appendix*, Table S4). On a cfu/pmol DNA basis, the MC variant pEPSA5MCDcm- achieved a 436-fold increase in transformation efficiency over the original pEPSA5Dcm- ($P \leq 1.0 \times 10^{-306}$). The increase could be due to the elimination of the two type I motifs along with the *E. coli* replicon in the MC variant (*SI Appendix*, Fig. S7), or the smaller MCs passing more readily through reversible pores formed in the *S. aureus* cell envelope during electroporation, or a combination of both. The relatively small 2.3-fold ($P = 1.29 \times 10^{-4}$) increase in transformation efficiency achieved by MC variant pEPSA5SynJE2MC over the plasmid pEPSA5SynJE2, both of which are completely RM-silent in JE2, favors the first possibility. In contrast, pEPSA5MC and pEPSA5SynJE2MC differed only by the presence or absence of a single type I target, respectively (Fig. 3*A*). Eliminating this single-target sequence resulted in a modest 1.5-fold ($P = 1.01^{-14}$) increase in transformation efficiency. Importantly, this suggests that in future applications of the SyngenicDNA approach, if a single target exists in an unadaptable region of DNA, such as an origin of replication or a promoter, its inclusion on an otherwise RM-silent plasmid might have minimal impact on the overall transformation efficiency.

Discussion

We report the development of an approach to circumvent the most common cause of genetic intractability, RM barriers, during microbial genetic engineering. In contrast to current mimicry-by-methylation approaches, ours involves stealth-by-engineering (*SI Appendix*, Fig. S8). We identify the precise targets of the RM systems within a poorly tractable (or intractable) bacterial strain, eliminate these targets from the DNA sequence template of a genetic tool in silico, via single-nucleotide polymorphisms (SNPs) or synonymous nucleotide modifications, and synthesize a tailor-made version of the tool that is RM-silent with respect to that specific host. This stealth-based SyngenicDNA approach allows for simple reworking of currently available genetic tools and DNA parts to permit them to efficiently operate in bacteria with active RM defenses. Additionally, we have repurposed minicircle technology to generate SyngenicDNA Minicircle Plasmid (SyMPL) tools, which are free from components required for propagation in *E. coli* but superfluous in the target host. Using a clinically relevant USA300 strain of *S. aureus*, we have demonstrated the profound improvement in transformation efficiency that can be achieved by systematic evasion of RM systems using these SyngenicDNA and SyMPL approaches.

In future applications, we expect that SyngenicDNA will be most readily applied to genetic tools that are functional in tractable strains, to modify them for use in related strains that are currently intractable or poorly tractable due to RM barriers, e.g., a newly emerging epidemic strain (22) or a newly recognized strain with biotechnological potential. In addition, SyngenicDNA could also facilitate synthetic biology approaches aimed at modular design/assembly of new genetic tools for intractable species where no genetically accessible strain is available (14). Synthetic biology focuses on the construction of biological parts that can be understood, designed, and tuned to meet specific criteria, with the underlying principle that genetic tools should be minimalistic, constructed of modularized parts, and sequence-optimized to allow for compatibility. Standardized formats for genetic tool assembly already exist to facilitate the simple implementation of synthetic tools and distribution of physical

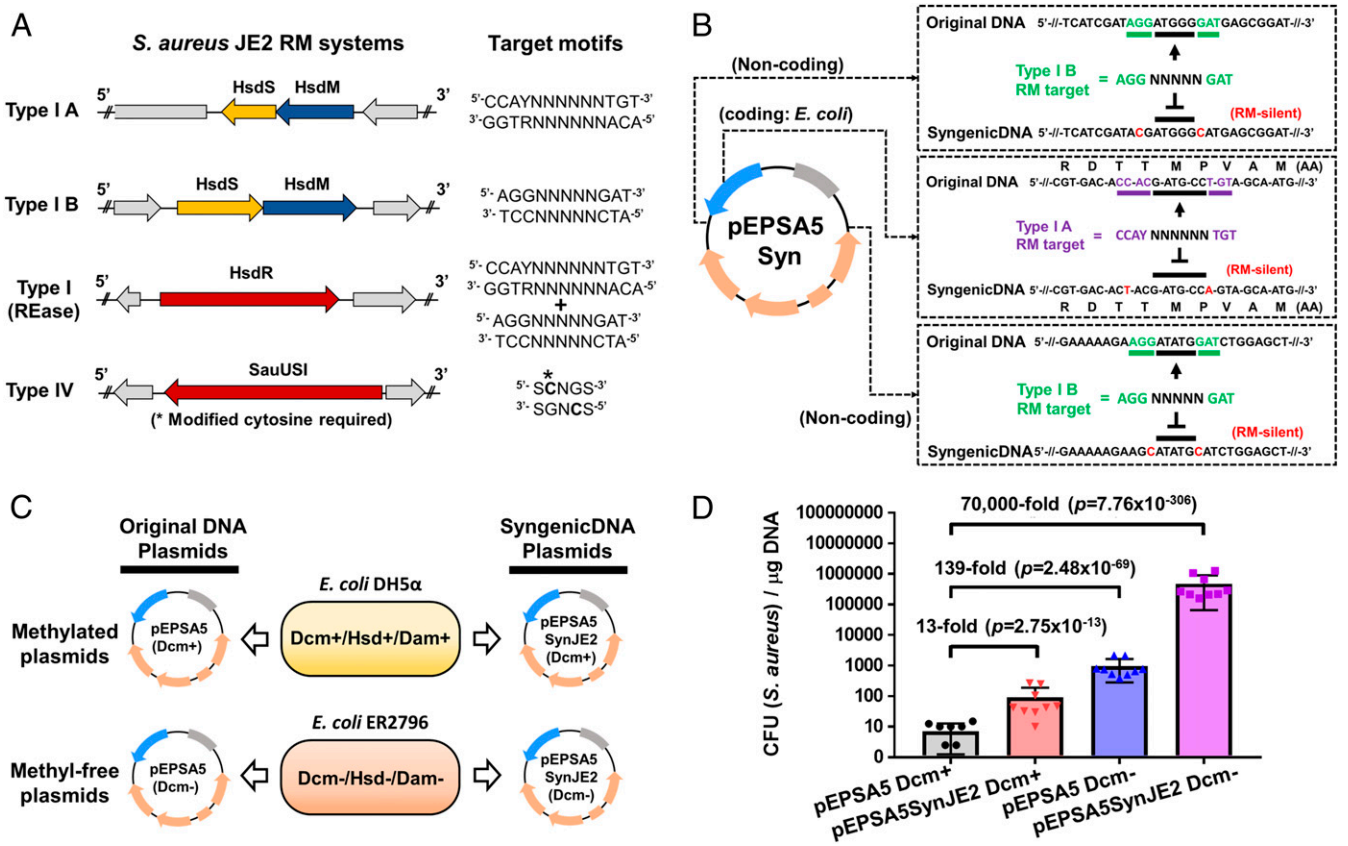

Fig. 2. The SyngenicDNA approach applied to *Staphylococcus aureus* JE2. (*A*) JE2 maintains two type I RM systems and a type IV restriction system. REase (HsdR and SauUSI) and MTase (HsdM) genes are shown in red and blue, respectively. Specificity subunit (HsdS) genes are shown in yellow. RM systems and their corresponding target motifs were identified by SMRTseq and REBASE analysis. (*B*) Construction of pEPSA5SynJE2, an RM-silent variant of the pEPSA5 plasmid tailored to JE2. Six nucleotide substitutions (two synonymous codon substitutions and four SNPs) eliminated all type I RM system targets from pEPSA5 sequence. (*C*) Plasmid propagation scheme. *E. coli* host strains produce DNA susceptible (DH5α; Dcm+) or resistant (*E. coli* ER2796; Dcm-) to the JE2 type IV restriction system. (*D*) Comparison of plasmid transformation efficiency (cfu/μg DNA) with pEPSA5 and the SyngenicDNA variant pEPSA5SynJE2.

parts between different laboratories (23). However, owing to RM systems variation between different strains of the same bacterial species (14), the design of reusable DNA parts that require physical reassembly for different bacteria is generally not applicable for intractable or poorly tractable strains with active RM systems. SyngenicDNA and SyMPL approaches should change that.

We adopted the core principles of synthetic biology, modularity and compatibility, but also accounted for variation in bacterial RM systems between strains by removing the need for physical assembly of reused parts propagated in other bacterial species. Because SyngenicDNA-based genetic tools require DNA synthesis de novo in the later step, the in silico tool assembly step could be utilized to augment plasmid backbones with additional useful parts (e.g., antibiotic-resistance cassettes, promoters, repressors, terminators, and functional domains, such as transposons or fluorescent markers) or create new tools. Additionally, because there is no requirement for a laboratory to

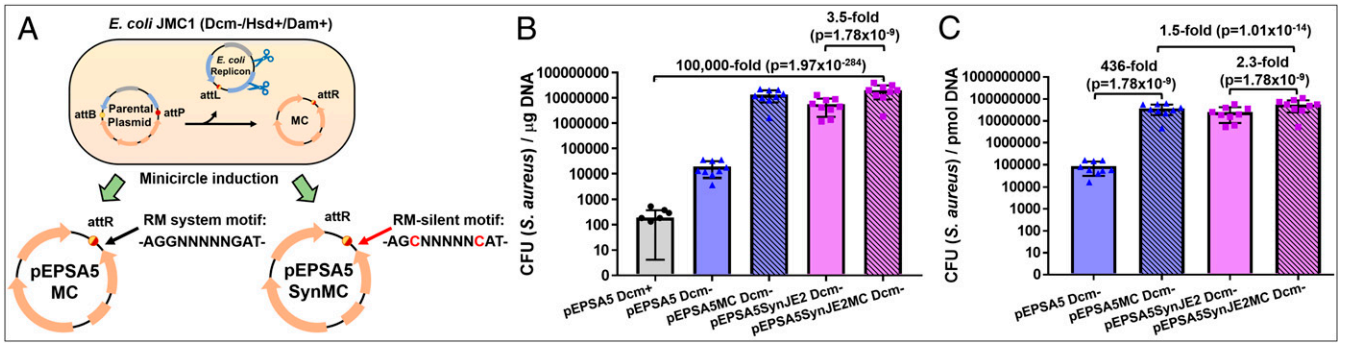

Fig. 3. The SyMPL approach applied to *Staphylococcus aureus* JE2. (*A*) Propagation of minicircles (pEPSA5MC and pEPSA5SynJE2MC) lacking Dcm-methylated sites within SyMPL-producer strain *E. coli* JMC1. (*B*) Comparison of SyngenicDNA and pEPSA5-based SyMPL plasmid transformation efficiency (cfu/μg DNA) with JE2. (*C*) Secondary analysis of SyngenicDNA and pEPSA5-based SyMPL plasmid transformation efficiencies in cfu/pmol DNA. Data are means ± SEM from nine independent experiments (three biological replicates with three technical replicates each).

physically obtain template DNA for PCR amplification of these additional parts, researchers would only need access to the publicly available DNA sequences of new parts to integrate them into a SyngenicDNA-based genetic tool, which could then be synthetized de novo in context. Notably, compatible replication origins and accessory elements for many cultivable bacterial phyla can be obtained from (*i*) the NCBI Plasmid Genome database, containing >50,000 complete DNA sequences of bacterial plasmids and associated genes (24), or (*ii*) the ACLAME database (25) (A Classification of Mobile genetic Elements), which maintains an extensive collection of mobile genetic elements including microbial plasmids from various sources.

In addition to impeding the biotechnological and commercial development of "probiotic" bacterial species (26) and the use of bacteria within industrial biofuel production or industrial processes (27), the limited genetic tractability of many major disease-causing bacteria of relevance to clinical and public health obstructs research in multiple fields. Our SyngenicDNA and SyMPL methods are effective, flexible, and we expect can now be applied to a wide range of bacteria to circumvent innate RM barriers, the most common underlying cause of genetic intractability (*SI Appendix*, Text S6). Finally, the fundamental methodology developed here will also likely be useful for evasion of other microbial defense mechanisms if they rely on distinct target recognition sequences to discriminate self- from nonself-DNA.

Materials and Methods

Microbial Strains and Reagents. *E. coli* NEBalpha competent cells were purchased from New England Biolabs (NEB). *E. coli* ER2796 was provided by the laboratory of Rich Roberts (NEB). *E. coli* MC (ZYCY10P3S2T; original minicircle-producing strain) was purchased from System Biosciences (SBI). A full list of reagents is provided in *SI Appendix*.

SMRTseq and RM System Identification. SMRTseq was carried out on a PacBioRSII (Pacific Biosciences) with P6/C4 chemistry at Johns Hopkins Deep Sequencing and Microarray Core Facility. Additional details on SMRTseq and RM system identification are in *SI Appendix*.

Bioinformatics and SyngenicDNA Adaptation in Silico. DNA sequence analysis and manipulations were performed using the Seqbuilder program of the DNASTAR software package (DNASTAR). Details of the bioinformatic tools used for adaptation are provided in *SI Appendix*.

DNA Synthesis and Assembly of SyngenicDNA Plasmids. A SyngenicDNA variant of the pEPSA5 plasmid (pEPSA5Syn) was assembled by replacing a 3.05-kb fragment of the original plasmid, encompassing three JE2 RM target sites, with a de novo synthesized DNA fragment that was RM-silent with respect to *S. aureus* JE2 (Fig. 2 and *SI Appendix*, Fig. S6). Details of assembly protocols are provided in *SI Appendix*.

Genome Editing of *E. coli* MC-Producer Strain. A CRISPR-Cas9/λ-Red multigene editing strategy was used for scarless MTase gene deletions in *E. coli* MC (ZYCY10P3S2T). Details on construction of a modified anhydrotetracycline inducible CRISPR-Cas9/λ-Red system and subsequent genome editing of the *E. coli* MC strain are provided in *SI Appendix*.

Production of SyMPL Tools. The 4.3-kb *S. aureus* replicons of both pEPSA5 plasmids (pEPSA5 and the pEPSA5SynJE2) were PCR-amplified and spliced to the MC parental plasmid (pMC; Systems Biosciences) to form pEPSA5P and pEPSA5SynJE2P. Primers and full details are provided in *SI Appendix*.

***S. aureus* Transformations.** Full details of competent cell preparations and electroporation protocols are provided in *SI Appendix*.

Statistical Analysis and Data Availability. Statistical analyses were carried out using Graphpad Prism (version 7.04; GraphPad Software) and Stata version 12.1 (StataCorp. 2011. Stata Statistical Software: Release 12. College Station, TX: StataCorp LP). Means with SE (SEM) are presented in each graph. Full details on statistical analyses and data availability are provided in *SI Appendix*.

ACKNOWLEDGMENTS. The authors wish to acknowledge the assistance of Richard J. Roberts (New England Biolabs) during REBASE analysis and Isabel Fernandez Escapa (Forsyth Institute) for technical input and assistance during *Staphylococcal* transformations. This research is supported by the National Institute of Dental and Craniofacial Research (NIDCR) of the National Institutes of Health (NIH) and the NIH Office of the Director (OD) under NIH Director's Transformative Research Award R01DE027850, and under Award NIDCR R01DE022586, and by the National Institute of Allergy and Infectious Diseases under Award R01AI101018. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

1. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 31:233–239.
2. van Opijnen T, Bodi KL, Camilli A (2009) Tn-seq: High-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods* 6:767–772.
3. Selle K, Barrangou R (2015) Harnessing CRISPR-Cas systems for bacterial genome editing. *Trends Microbiol* 23:225–232.
4. Aune TEV, Aachmann FL (2010) Methodologies to increase the transformation efficiencies and the range of bacteria that can be transformed. *Appl Microbiol Biotechnol* 85:1301–1313.
5. Johnston C, Martin B, Fichant G, Polard P, Claverys JP (2014) Bacterial transformation: Distribution, shared mechanisms and divergent control. *Nat Rev Microbiol* 12:181–196.
6. Floyd MM, Tang J, Kane M, Emerson D (2005) Captured diversity in a culture collection: Case study of the geographic and habitat distributions of environmental isolates held at the American type culture collection. *Appl Environ Microbiol* 71:2813–2823.
7. Dykhuizen D (2005) Species numbers in bacteria. *Proc Calif Acad Sci* 56(Suppl 1):62–71.
8. Vasu K, Nagamalleswari E, Nagaraja V (2012) Promiscuous restriction is a cellular defense strategy that confers fitness advantage to bacteria. *Proc Natl Acad Sci USA* 109:E1287–E1293.
9. Tock MR, Dryden DT (2005) The biology of restriction and anti-restriction. *Curr Opin Microbiol* 8:466–472.
10. Suzuki H (November 28, 2012) Host-mimicking strategies in DNA methylation for improved bacterial transformation. *Methylation-From DNA, RNA and Histones to Diseases and Treatment*, 10.5772/51691.
11. Roberts RJ, Vincze T, Posfai J, Macelis D (2015) REBASE–A database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Res* 43:D298–D299.
12. Yasui K, et al. (2009) Improvement of bacterial transformation efficiency using plasmid artificial modification. *Nucleic Acids Res* 37:e3.
13. Costa SK, Donegan NP, Corvaglia AR, François P, Cheung AL (2017) Bypassing the restriction system to improve transformation of *Staphylococcus epidermidis*. *J Bacteriol* 199:e00271-17.
14. Johnston CD, Skeete CA, Fomenkov A, Roberts RJ, Rittling SR (2017) Restriction-modification mediated barriers to exogenous DNA uptake and incorporation employed by *Prevotella intermedia*. *PLoS One* 12:e0185234.
15. Vasu K, Nagaraja V (2013) Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol Mol Biol Rev* 77:53–72.
16. Lee BY, et al. (2013) The economic burden of community-associated methicillin-resistant *Staphylococcus aureus* (CA-MRSA). *Clin Microbiol Infect* 19:528–536.
17. Monk IR, Foster TJ (2012) Genetic manipulation of Staphylococci-breaking through the barrier. *Front Cell Infect Microbiol* 2:49.
18. Jones MJ, Donegan NP, Mikheyeva IV, Cheung AL (2015) Improving transformation of *Staphylococcus aureus* belonging to the CC1, CC5 and CC8 clonal complexes. *PLoS One* 10:e0119487.
19. Fey PD, et al. (2013) A genetic resource for rapid and comprehensive phenotype screening of nonessential *Staphylococcus aureus* genes. *MBio* 4:e00537-12.
20. Kay MA, He CY, Chen ZY (2010) A robust system for production of minicircle DNA vectors. *Nat Biotechnol* 28:1287–1289.
21. Sadykov MR (2016) Restriction-modification systems as a barrier for genetic manipulation of *Staphylococcus aureus*. *Methods Mol Biol* 1373:9–23.
22. Manara S, et al. (2018) Whole-genome epidemiology, characterisation, and phylogenetic reconstruction of *Staphylococcus aureus* strains in a paediatric hospital. *Genome Med* 10:82.
23. Silva-Rocha R, et al. (2013) The standard European vector architecture (SEVA): A coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res* 41:D666–D675.
24. Shintani M, Sanchez ZK, Kimbara K (2015) Genomics of microbial plasmids: Classification and identification based on replication and transfer systems and host taxonomy. *Front Microbiol* 6:242.
25. Leplae R, Lima-Mendez G, Toussaint A (2010) ACLAME: A CLAssification of mobile genetic elements, update 2010. *Nucleic Acids Res* 38:D57–D61.
26. Sun Z, Baur A, Zhurina D, Yuan J, Riedel CU (2012) Accessing the inaccessible: Molecular tools for bifidobacteria. *Appl Environ Microbiol* 78:5035–5042.
27. Leang C, Ueki T, Nevin KP, Lovley DR (2013) A genetic system for *Clostridium ljungdahlii*: A chassis for autotrophic production of biocommodities and a model homoacetogen. *Appl Environ Microbiol* 79:1102–1109.