Computational design of CRISPR guide RNAs to enable strain-specific control of microbial consortia

Austin G. Rottinghaus[1], Steven Vo[2], and Tae Seok Moon[1,2,*]

[1]Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis, United States
[2]Division of Biology and Biomedical Sciences, Washington University in St. Louis, St. Louis, United States

* Correspondence
Tae Seok Moon
One Brookings Dr., Box 1180
St. Louis, MO 63130, USA
Tel: +1 (314) 935-5026
Email: tsmoon7@gmail.com

Abstract

Microbes naturally coexist in complex, multi-strain communities. However, extracting individual microbes from and specifically manipulating the composition of these consortia remains challenging. The sequence-specific nature of CRISPR guide RNAs can be leveraged to accurately differentiate microorganisms and facilitate the creation of tools that can achieve these tasks. We developed a computational program, ssCRISPR, which designs strain-specific CRISPR guide RNA sequences with user-specified target strains, protected strains, and guide RNA properties. We experimentally verify the accuracy of the strain-specificity predictions in both *Escherichia coli* and *Pseudomonas spp.* and show that up to three nucleotide mismatches are often required to ensure perfect specificity. To demonstrate the functionality of ssCRISPR, we apply computationally designed CRISPR-Cas9 guide RNAs to two applications: the purification of specific microbes through one- and two-plasmid transformation workflows and the targeted removal of specific microbes using DNA-loaded liposomes. For strain purification, we utilize gRNAs designed to target and kill all microbes in a consortium, except the specific microbe to be isolated. For strain elimination, we utilize gRNAs designed to target only the unwanted microbe, while protecting all other strains in the community. ssCRISPR will be of use in diverse microbiota engineering applications.

Keywords

computational CRISPR RNA design; microbial consortium engineering; targeted microbial killing; targeted strain isolation from microbiota; machine learning; liposome-mediated delivery

Significance Statement

Modifying microbial consortia with strain-specificity is critical for maintaining stable and healthy microbiota. However, consortium engineering tools with strain-specificity have yet to be developed. Here, we describe the development and validation of a novel computational program, which designs strain-specific CRISPR guide RNAs that can be utilized to modify complex consortia. As a proof of concept, we applied the program to two novel applications: the isolation of specific microbes from consortia and the removal of specific microbes from consortia. This new technique has many practical applications, including addressing the problem of antibiotic-resistant microbes and isolating useful microbes from the environment.

Main

Microbes naturally co-exist in complex and dynamic communities. These microbial consortia cooperate to influence the health of the environment, domestic animals, humans, and plants(1, 2). Efforts to create synthetic microbial communities have led to advances in fields, including metabolic engineering and bioremediation(3). Numerous microbes have been extracted from natural consortia with highly specialized and essential functions(4, 5). However, identifying and purifying these microbes remains challenging(6). Pathogens also inhabit these communities and opportunistically disrupt host health. Modern methods of removing them, including antibiotics, are highly disruptive to the survival of homeostatic, beneficial microbes and have led to the global emergence of deadly antibiotic- and bactericide-resistant pathogens(7, 8). Recent advances in phage engineering and plasmid conjugation have allowed microbes to be targeted and killed in a strain-specific manner, causing minimal impact on the stability of the microenvironment(9-11). Microbes have also been engineered with novel functions and introduced into natural microbiomes to improve the health of the host(12, 13) and engineered to selectively colonize specific microenvironments(14, 15). However, exogenously provided microbes often have a difficult time penetrating consortia, finding a niche, and persisting for a long term(16). As an alternative to supplementing microbiota with engineered microbes, microbes can instead be engineered *in situ* using external DNA delivery methods, increasing the endurance of the added functionality(6, 17, 18). However, methods for engineering microbes *in situ* often lack strain specificity, and instead introduce the exogenous DNA randomly into the microbiota(6).

CRISPR-Cas systems can be tuned to recognize specific genetic loci by modulating the sequence of the guide RNA (gRNA), providing opportunities for strain recognition in microbial consortia. This functionality has been harnessed for applications in strain-specific microbial engineering(17, 18) and elimination(19, 20). Numerous programs have been developed to help design gRNAs with high cutting efficiency and low off-target cleavage rates using machine learning and deep learning models that consider the sequence and thermodynamic characteristics of the gRNA sequence(21-24). However, programs for designing gRNAs specific to individual microbial strains are lacking. One recent work achieved this goal with an effective and accessible website(10). However, the program lacks strain selection options, cannot be utilized for diverse CRISPR systems beyond Cas9, and defines a strain-specific gRNA as one with at least one nucleotide (nt) mismatch in the non-target strains, which has been shown to be insufficient to prevent cleavage(25).

In this work, we created a program, ssCRISPR, which computationally designs strain-specific CRISPR gRNAs from user-defined target and non-target strains without the common deficiencies in current programs. Genome sequences for strain options were extracted from the expansive National Center for Biotechnology Information (NCBI) genome repository, giving users over 27,000 strain selection options, or can be provided by the user. Users of ssCRISPR can also input their desired protospacer adjacent motif (PAM) sequence, target sequence length, and PAM-target orientation, giving the program the customizability required for use with any CRISPR-Cas system. Furthermore, users can select their desired criteria for specificity, from 1-4 nt, as the application will dictate the required stringency. However, we show that to ensure complete strain-specificity, at least 3 nt mismatches in the target sequence relative to the genomes of all non-target strains may be required. We also demonstrated two potential applications of ssCRISPR-designed gRNAs: first, the purification of a single strain from a microbial consortium using a single plasmid transformation; second, the depletion of a single strain from a microbial consortium using liposomal delivery of strain-specific CRISPR-Cas9 cassettes. ssCRISPR can be downloaded and run locally either as a Python script or as an all-encompassing executable application. In either

case, users can take advantage of the user-friendly graphical interface to operate the program without programming expertise.

Results

ssCRISPR Identifies efficient gRNAs for target strains

We sought to create a program to computationally design strain-specific gRNAs through four sequential stages (Fig. 1). In the first stage, the user can select criteria for the program, allowing enough customizability for any consortium and Cas protein. In the second stage, the program identifies gRNA target sequences in each target strain and keeps sequences present in all strains. In the third stage, the program identifies gRNA target sequences in the non-target strains and removes these sequences from the list of gRNAs identified in the second stage. When specified, the list of non-target strain gRNAs is expanded to include each possible iteration with 1, 2, or 3 nucleotide changes. In the final stage, for use with specific Cas proteins, the program ranks the gRNAs from most to least efficient using machine learning models based on the sequence and thermodynamic properties.

To develop ssCRISPR, we first needed a reference database of genome sequences. We selected the NCBI genome repository, which at the time of last download included 27,569 complete bacterial genome sequences. The database is rapidly expanding to include newly sequenced genomes. The sequences can be quickly extracted from NCBI using the sequence reference number which eliminates the burdensome need to maintain the full sequences locally and allowing for easy future updates. To use the repository, we downloaded the table of strain names and corresponding sequence reference numbers and packaged the table file with the developed gRNA design program. The user then has the option to select target strains and protected, non-target strains for gRNA identification (Fig. 1). However, if a desired strain is not provided as an option, users can also provide their own sequences.

Having obtained an expansive database of strain selections, we next sought to make ssCRISPR generalizable across any CRISPR-Cas system. To achieve this goal, we created user inputs for the following characteristics: target sequence length, PAM sequence, and PAM orientation relative to the target sequence. These inputs allow the user to apply the program to CRISPR-Cas systems ranging from *Streptococcus pyogenes* Cas9, which has a 20 nt target sequence, an NGG PAM sequence, and a 5'-target-PAM-3' orientation(26), to *E. coli* Cas3, which has a 32 nt target sequence, AWG/NAG/ATG PAM sequence, and 5'-PAM-target-3' orientation(27). ssCRISPR applies these criteria to sequentially search the genomes of all selected target strains for the specified PAM sequences and extract the corresponding target sequences. Native plasmids are not considered viable gRNA target sites as they are mostly inessential for cell survival. However, if multiple unique chromosomes exist, all are considered for possible gRNA target sites. After searching each selected strain, ssCRISPR compares the lists of identified target sequences, and only gRNA sequences with exact matches between all target strains are maintained (Fig. 1).

To evaluate the program, we determined the number of CRISPR-Cas9 gRNA target sites shared between all 2,068 sequenced *E. coli* genomes using reverse alphabetical order. ssCRISPR identified 1,441 broad-targeting *E. coli* gRNA sequences (Fig. 2A). We repeated the process for all 1,020 sequenced strains of *Pseudomonas spp*. and identified 142 total gRNA target sites. The program run for *Pseudomonas spp*. gRNAs eliminated viable gRNAs more rapidly than the run for *E. coli*, with over 99% of potential gRNAs removed after just two strains (*P. zhaodongensis*

A252 and *P. zeae* OE 48.2) versus 880 *E. coli* strains. This observation can be explained by the larger genetic diversity between *Pseudomonas* species than same-species *E. coli* strains. We repeated the analysis for several additional Cas proteins, including variants for Cpf1 (*Lachnospiraceae bacterium* Cas12a; 23 nt target length, TTTV PAM, and 5'-PAM-target-3' orientation(28)), *E. coli* Cas3 (32 nt target length, AWG, ANG, or ATG PAM, and 5'-PAM-target-3' orientation(27)), and *Alicyclobacillus acideoterrestris* Cas12b (20 nt target length, TTN PAM, and 5'-PAM-target-3' orientation(29)) (Fig. S1). Cas proteins with more stringent PAM sequences and longer target sequences generally had fewer potential gRNA target sites in both *E. coli* and *Pseudomonas spp*. Furthermore, *Pseudomonas spp*., which have a higher GC content (~60%) than *E. coli* (~50%), had a 2-fold larger reduction in the number of predicted gRNAs for Cpf1 relative to Cas9 due to the AT rich Cpf1 PAM sequences.

We next sought a method to select the best possible gRNAs from the list of identified sequences. To achieve this goal, we adapted and incorporated a relative cleavage efficiency prediction model previously demonstrated by Guo *et al*.(21) We used the dataset of ~56,000 CRISPR-Cas9 gRNA sequences assessed by Guo *et al*. to train and optimize a gradient boosting regression machine learning model from the following 396 sequence composition and energetic properties: total A, T, C, G and GC content, T content of the four PAM-adjacent nucleotides, presence of an A, T, C, or G in each of the 20 PAM adjacent nucleotides (80 properties), presence of each nucleotide dimer (NN) in each of the 20 PAM adjacent nucleotides (304 properties), minimum free energy for the 12 PAM adjacent nucleotides and the full gRNA sequence, the melting temperature for the five PAM adjacent nucleotides, next eight nucleotides, remaining nucleotides, and the full gRNA sequence. The GC content, sequence of the PAM-adjacent seed region, and thermodynamic properties of the RNA and DNA-RNA complex were found to be the most important features of the model (Fig. S2). To evaluate the accuracy of the model, we compared the predicted and actual efficiency rankings (Fig. 2B). The ranking comparison displayed a moderate relationship, with a Spearman's rank correlation coefficient of 0.56, which is in line with other gRNA efficiency machine learning models(21, 22). To see if the model was generalizable across Cas proteins, we obtained a gRNA efficiency dataset for Cpf1 from Kim *et al*. and applied the model(24). The model showed no correlation between predicted and actual gRNA efficiency (Fig. S3A). As such, we used the same sequence composition and energetic properties to train a new machine learning model for Cpf1 gRNAs. The new model showed a moderate correlation, with a spearman's rank correlation coefficient of 0.57 (Fig. S3B). As such, ssCRISPR can present users with high efficiency gRNAs for Cas9 and Cpf1. However, new models will need to be created for alternative proteins as experimental datasets become available.

To experimentally validate the program, we selected four gRNAs that target all tested *E. coli* strains and four gRNAs that target all tested *Pseudomonas* strains with the highest predicted efficiency (Table S1). Plasmids for each gRNA target sequence were constructed with constitutive promoters driving gRNA expression. We next transformed *E. coli* DH10B, Nissle 1917, MG1655, and BL21(DE3), each harboring a Cas9 expression plasmid or an empty vector, with a non-targeting control plasmid or the gRNA plasmids; the following four transformation combinations were performed for each strain: strain with Cas9 transformed with gRNA, strain with Cas9 transformed with non-targeting control, strain without Cas9 transformed with gRNA, and strain without Cas9 transformed with non-targeting control. To obtain the efficiency of each gRNA, we calculated the ratio of the number of colonies obtained from each gRNA plasmid to the number of colonies obtained from the control plasmid with Cas9 present divided by the ratio of the number of colonies obtained from each gRNA plasmid to the number of colonies obtained from the control plasmid without Cas9 present. Each gRNA plasmid demonstrated a killing efficiency (see Methods) of 3- to 4-$\log_{10}$ in all four tested strains (Fig. 2C). We observed similar results with the *Pseudomonas spp*. gRNAs, with killing efficiencies of 2- to 4-$\log_{10}$ achieved for each gRNA in all

four strains (Fig. 2D). These results demonstrate that ssCRISPR identifies gRNAs with efficient target sites in multiple organisms.

Three nucleotide mismatches are required for optimal strain specificity

We next wanted to incorporate strain protection into the program by allowing the user to select non-target strains that lack the gRNA target site. However, criteria for what makes a gRNA sequence strain-specific were required (Fig. 1). It has been previously demonstrated that nucleotide mismatches in the PAM and 10-12 nt PAM-adjacent seed region cause the largest reduction in cleavage efficiency(25, 30, 31). As such, we first defined a strain-specific gRNA to be one that possesses at least one nucleotide mismatch in the PAM site or the 10 nt seed region compared to all specified non-target strains. The program applied the same method described above to identify all gRNA target sequences in the specified non-target strains. Any sequence in the identified list of broad-targeting gRNAs that contained a seed region perfectly matching a gRNA from the non-target strains was removed.

To assess this function, we tested the efficiency of four gRNAs, one specific to each of *E. coli* DH10B, Nissle 1917, MG1655, and BL21(DE3), in each of the four *E. coli* strains. Each gRNA efficiently killed its cognate strain (Fig. 3A, left). However, a gRNA efficiency of greater than 1-$\log_{10}$ was also observed in 4/12 non-cognate combinations. To improve specificity and reduce the likelihood of off-target cleavage, we increased the stringency of the program to require two mismatches in the same region. Specificity was improved but remained imperfect, with all four gRNAs demonstrating efficient activity in their cognate strain and significant activity observed in 1/12 non-cognate combinations (Fig. 3A, middle). We further increased the stringency to 3 nt but found that gRNA options were rapidly eliminated after considering each non-target strain. We determined that requiring three mismatches in the 10 nt seed region, but ignoring the rest of the gRNA sequence, led to a high probability of each gRNA sequence occurring in any given random nucleotide sequence (Fig. S4). To alleviate this issue, we expanded the considered region to be the full 12 nt seed region. This criterion successfully identified strain-specific gRNAs, with all four tested gRNAs demonstrating efficient activity in their cognate strain and no activity in their non-cognate strains (Fig. 3A, right).

Upon further analysis, we determined that the probability of a 12 nt gRNA seed sequence randomly occurring in any given sequence remained too high for considering many non-target strains. Specifically, 99% of gRNA sequences are eliminated by a random 80,000,000 nt sequence, corresponding to approximately 16 average size microbial genomes(32). As such, we expanded the considered region to a 20 nt target sequence. Using this criterion, over 1,000,000 strains worth of random DNA are required to eliminate 99% of gRNAs, with less than 1% of gRNAs eliminated after over 1,000 strains worth of random DNA. However, we found screening tens of thousands of gRNAs for 3 nt of specificity to be very computationally intensive. As such, if more than 5,000 gRNAs are identified with 2 nt of specificity, 5,000 are randomly selected for further analysis (Fig. S5A). However, we found this number to be more than sufficient. ssCRISPR identified thousands of gRNAs with specificity to each of the four considered *E. coli* and *Pseudomonas* strains when the set of four was considered exclusively (Figs. S6A, 6B). The number of viable gRNA sequences was reduced when all other *E. coli* or all other *Pseudomonas* strains were specified as non-target strains (Figs. S6C, S6D and Table S2). However, at least one gRNA was identified with 3 nt of specificity for all strains except *E. coli* MG1655. This result may be caused by its frequent use and analysis, as many of the sequenced strains in the reference genome database may be derived from *E. coli* MG1655.

We selected and tested the four best predicted gRNAs with specificity to each of the four *E. coli* strains (16 total gRNAs). All 16 gRNAs maintained perfect specificity, with no significant activity observed in any non-cognate combination (Fig. 3B). To further validate the program, we tested an additional predicted 16 strain-specific gRNAs in the four *Pseudomonas* strains. Again, all 16 gRNAs demonstrated perfect strain-specific activity (Fig. 3C). While we showed that 3 nt mismatches in a 20 nt gRNA target sequence allows for perfect strain specificity, ssCRISPR allows the user to specify the desired number of nucleotide mismatches (from 1-4), as fewer may be sufficient for some applications. Notably, when 4 nt of specificity are desired, the number of gRNAs with 3 nt tested is limited to 100 (Fig. S5B).

Purifying single strains from microbial consortia using ssCRISPR gRNAs

We next wanted to apply ssCRISPR to isolate and engineer a single strain from a microbial consortium. Modern methods of microbial engineering employ lambda Red-mediated recombination to engineer a strain of interest and CRISPR-Cas gRNAs that target the unmodified recombination site to select for successfully modified strains(33, 34). To utilize this system to isolate and engineer microbes, we created a workflow where strain-specific gRNAs, designed using ssCRISPR, target the genomes of non-desired strains, rather than the site of recombination in the desired strain. A consortium containing the desired strain can be transformed with the Cas9/lambda Red plasmid, cultured, and transformed again with the integration cassette and strain-specific gRNA plasmid (Fig. 4A). To negate the need for a gRNA that targets the integration site, an antibiotic resistance gene can be included in the integration cassette for selection during this initial round of engineering. The antibiotic resistance gene can be later replaced with any DNA of interest using a gRNA that targets the antibiotic resistance gene. Alternatively, a two-gRNA system can be employed, where one gRNA targets the genome of non-desired strains and a second targets the engineered site. If the user wants the integration to occur in multiple strains, they can also design the second gRNA with ssCRISPR by providing a sequence file for the desired integration region in one or more of the target strains.

To validate the one-gRNA system, we used ssCRISPR to design a gRNA that protects *E. coli* Nissle 1917 while targeting *E. coli* DH10B, MG1655, and BL21(DE3). We next created an integration cassette harboring a kanamycin resistance gene that targets the *lacZ* locus in *E. coli* Nissle 1917. The *E. coli* Nissle 1917 *lacZ* sequence is 99% identical to the *lacZ* sequences in the other *E. coli* strains, suggesting that any strain-specificity by the system would be a result of strain-specific genomic cleavage from the gRNA, and not differences in homologous recombination efficiency caused by nucleotide mismatches in the homologous arms. We tested the system using cultures of each strain individually and in an equal-part consortium. *E. coli* BL21(DE3) yielded no colonies when transformed with the Cas9/lambda Red plasmid and was therefore excluded from this experiment. When we transformed the integration cassette with a control plasmid, colonies of all three strains were observed (Fig. 4B). However, in the microbial mixture, *E. coli* MG1655 and Nissle 1917 outcompeted *E. coli* DH10B due to their higher growth rates. When the strains were transformed with the strain-specific gRNA plasmid, only engineered colonies of *E. coli* Nissle 1917 were observed. This demonstrates that ssCRISPR can facilitate the isolation and engineering of specific microbes from a consortium. We next attempted to use the system to isolate and engineer *E. coli* Nissle 1917 from murine fecal samples. We previously obtained murine fecal samples from mice gavaged with 10^8 CFUs of *E. coli* Nissle 1917(20). When we transformed the Cas9/lambda Red plasmid into the fecal consortium, 100% of the resulting colonies were from *E. coli* Nissle 1917. This result suggests that the plasmid is not compatible with other microbial genera and can therefore be leveraged alone to purify *E. coli* from complex consortia.

When only strain isolation is desired, Cas9 and strain-specific gRNAs can be paired on a single plasmid, and a single transformation can be used to isolate the strain (Fig. 4C). Furthermore, multiple gRNAs can be expressed in an array from a single promoter and post-transcriptionally processed using intergenic RNA cleavage sites(35) or in multiple independent and non-repetitive cassettes(36). To demonstrate this idea, we used a p15A origin plasmid, which only replicates in *Enterococcus spp.*(18), to constitutively express Cas9 and a gRNA ELSA array(36). The gRNA array consisted of six non-repetitive gRNA cassettes that target different subsets of *Enterococci* but protect *E. coli* Nissle 1917 with at least 1 nt of specificity (Table S3). We individually tested two gRNAs from each strain group to identify ones with the desired specificity (Fig. S7). When we transformed a mixture of *E. coli* strains with a control plasmid and the test plasmid, we observed a substantially higher (p<0.0001) relative abundance of *E. coli* Nissle 1917 in the population that received the test plasmid (95%) compared to the population that received the control plasmid (13%; Fig. 4D). We then used the same plasmid to purify *E. coli* Nissle 1917 from a more complex strain mixture composed of *P. putida* F1, *Salmonella typhimurium*, and *Rhodococcus opacus* PD630. Transforming the strain mixture with the test plasmid significantly depleted *P. putida* F1 (p=0.0006) and *S. typhimurium* (p=0.0062), while increasing the abundance of *E. coli* Nissle 1917 (p<0.0001). *R. opacus* PD630, which is an incompatible host for p15A origin plasmids, was not detected after either transformation. We created a similar construct for the purification of *P. putida* F1 from a consortium of *Pseudomonas* strains and demonstrated a strong increase (p<0.0001) in the abundance of *P. putida* F1 in the population that received the test plasmid (85%) compared to the population that received the control plasmid (<1%; Fig. S8). Collectively, these data show that gRNAs designed using ssCRISPR can be utilized to isolate microbes from consortia in a single transformation.

Liposome delivery of strain-specific CRISPR-Cas9 antimicrobials

ssCRISPR also has the potential to be used to selectively remove microbes from a consortium. To accomplish this goal, we selected a gRNA that specifically targets *E. coli* Nissle 1917 and inserted it on the p15A plasmid with the constitutive Cas9 cassette. When we transformed an equal-part, multi-strain *E. coli* consortia with the control plasmid and test plasmid, we observed a 3.8-$\log_{10}$ reduction in *E. coli* Nissle 1917 CFUs for the test plasmid compared to the control plasmid treated populations (Fig. 5A). *E. coli* DH10B, MG1655, and BL21(DE3) also showed lower CFUs in response to transformation with the test plasmid compared to those transformed with the control plasmid, but to a significantly smaller degree than *E. coli* Nissle 1917 (p<0.0001). These changes may have been a result of differences in the transformation efficiency of the competent cells or plasmids. Alternatively, the lower CFUs may have been a result of the inherent toxicity of constitutive Cas9 and gRNA expression. Optimization of the Cas9 and gRNA expression levels may reduce the toxicity and eliminate the gRNA sequence-independent CFU differences. We applied the same protocol to remove *E. coli* Nissle 1917 from murine fecal samples. Prior to transformation, we quantified the amount of *E. coli* Nissle 1917 in the samples and determined that the strain made up approximately 2% of the aerobically-culturable microbes (Fig. S9). Transformation of the fecal consortia with the control plasmid increased the relative CFUs of *E. coli* Nissle to approximately 10% of the total aerobically-culturable microbes (Fig. 5B). However, transformation of the fecal consortia with the test plasmid eliminated *E. coli* Nissle 1917. These results show that ssCRISPR gRNAs can be used to selectively target and eliminate microbes in consortia.

ssCRISPR gRNAs can also be used to create strain-specific CRISPR antimicrobials by pairing them with a non-specific DNA delivery method. Several methods of non-specific delivery of

biologics have been demonstrated in bacteria, including plasmid conjugation(18), bacteriophage infection(37), and liposome delivery(38). To date, bacteriophages(9) and plasmid conjugation(10, 11) have been used to deliver strain-specific antimicrobials *in situ*. We sought to instead package plasmid DNA carrying Cas9 and ssCRISPR gRNAs in liposomes that non-specifically fuse with microbes and deliver the DNA payload which is lethal only to strains harboring the gRNA target sequence (Fig. 5C). We constructed liposomes and packaged them post-synthesis with the control and *E. coli* Nissle-1917-killing test plasmid described above and optimized the liposome synthesis and plasmid-packaging protocols (Fig. S10). We next incubated an equal-part, multi-strain *E. coli* consortium with the liposomes for 30 minutes and quantified the number of cells that survived plasmid delivery (Fig. 5D). The *E. coli* DH10B, MG1655, and BL21(DE3) populations treated with the test and control plasmids showed similar CFUs after plasmid delivery. However, *E. coli* Nissle 1917 showed a 2-$\log_{10}$ reduction in viable CFUs when comparing the control and test treated populations. On average, less than 1 CFUs/mL was observed in populations treated with either the control or test plasmids without liposome packaging, showing the importance of the liposomes for plasmid delivery (Fig. S11). We next wanted to determine the feasibility of using ssCRISPR to design strain-specific gRNAs for *in vivo* applications, such as pathogen elimination, where significantly more complex consortia would be encountered. We designed gRNAs that target all *E. coli* strains, all *Staphylococcus aureus* and *Staphylococcus epidermidis*, or all *Clostridioides difficile* strains, while protecting strains from all other genera with at least 1 nt of specificity. We identified 189 *E. coli*-specific, 124 *S. aereus*- and *S. epidermidis*-specific, and more than 5,000 *C. difficile*-specific gRNAs. Furthermore, for cases where the exact strain causing a bacterial infection is unknown, a gRNA array could also be used to eliminate several of the pathogens that may be causing the infection. Together, these results show that ssCRISPR can be used to design gRNAs that target microbes in consortia with high selectivity and efficiency.

Discussion

Manipulating microbial consortia with strain specificity can facilitate significant advances in medicine, agriculture, and climate control(6, 13, 39). However, a method for reliably distinguishing strains is essential to minimize unwanted side effects(6). Current programs for designing strain-specific gRNAs lack selectable strain options, cannot be customized for different CRISPR systems, and insufficiently define the characteristics that make a gRNA strain-specific(10). As described here, we created the ssCRISPR program to design CRISPR gRNAs with reliable strain-specific cleavage profiles. To ensure accuracy, we comprehensively tested selectivity criteria in multiple microbial strains. In addition, to allow for wide-spread use of ssCRISPR, we incorporated a wide array of user-defined parameters and more than 27,000 selectable strain options (Fig. 1). We showed that ssCRISPR accurately predicts gRNAs with efficient and specific activity in all selected target strains (Fig. 2) and minimal activity in selected non-target strains (Fig. 3). Furthermore, we demonstrated two applications of ssCRISPR: first, to purify specific microbes from defined consortia (Fig. 4); second, to remove individual microbes from defined and complex consortia using broad-spectrum delivery methods such as liposomes (Fig. 5).

Purifying a specific microbe from a consortium can be a difficult task using standard modern methods such as targeted enrichment in tailored complex media and serial plating(40). However, this process can be simplified using strain-specific gRNAs designed with ssCRISPR. To use ssCRISPR to purify a microbe from a consortium, a degree of knowledge about the strains in the mixture is required. If the consortium is defined, designing gRNAs using ssCRISPR to target strains is a simple process. However, it is still essential that the genetic parts, such as the origin of replication and promoters, are compatible with the organisms to facilitate the purification; the

origin needs to be functional in the strain of interest, and the promoters driving expression of the Cas protein and gRNAs need to be functional in any organism with origin compatibility. Furthermore, for more complex consortia, experiments such as 16S rRNA sequencing may be required to first characterize the composition of the mixture and identify relevant strains. However, the isolation process can be improved by carefully selecting origins with narrow compatibility groups (Fig. 4D) and by selecting growth conditions favorable for the desired microbe(41).

Creating technologies to remove specific microbes from a consortium is essential to combat the growing issues of antibiotic- and bactericide-resistant pathogens in domesticated animals(42), humans(7), and plants(8). Identifying gRNAs for strain-specific removal is simpler than for purification, as microbial diversity becomes an advantage. For this application, genetic parts only need to be functional in the selected target strains. However, for the delivery of strain-specific CRISPR antimicrobials, factors including delivery efficiency and genetic remnants need to be considered. Recent advances in plasmid conjugation allow for a significantly higher transfer and delivery rates of the CRISPR cassettes(43). However, genetic materials transferred via bacteriophages, viral vectors, and plasmid conjugation are permanent once introduced into the environment, and wide-spread delivery of this replicating genetic material into native microbes can have adverse biological consequences(44, 45). Here, as a proof of concept, we used plasmid-packaged liposomes to deliver the CRISPR payload but experienced a low uptake efficiency. However, liposomes have the potential to deliver antimicrobial CRISPR systems in non-permanent forms, including as RNA and proteins, which are degraded intracellularly. Furthermore, RNA- and protein-based payloads may have a higher delivery efficiency than plasmids when packaged in liposomes, as both can be engineered to penetrate a cell membrane more easily than plasmids in the event that the liposome only fuses with the outer membrane(46, 47). Finally, the CRISPR delivery efficiency by the liposomes may be improved through alternative liposome production methods. Producing the CRISPR-loaded liposomes through microfluidic or controlled ethanol injection approaches could result in smaller, unilamellar liposomes with better microbial fusion efficiencies(48).

The ssCRISPR program is not without limitations. The selectable strain options in ssCRISPR are derived from the NCBI genome repository and can be easily updated to include the rapidly accumulating new microbial genomes. However, the number of strains with sequenced genomes pales in comparison to the 10^{12} microbial species predicted to exist on Earth(49). As such, the true specificity of the gRNAs designed by the program will never be completely defined until all microbial genomes have been sequenced. In addition, although the ssCRISPR efficiency predictions for Cas9 and Cpf1 gRNAs are comparable to numerous other machine learning models, they fall behind recent deep learning models in accuracy(23, 24). Fortunately, in most applications of ssCRISPR, only a highly active gRNA, rather than the best gRNA, is needed. To this end, when considering the top 5% most efficient gRNAs in a defined group, ssCRISPR predicts 96% (Cas9) or 98% (Cpf1) of the subset to be above the true median efficiency (Fig. 2B and Fig. S3B). Therefore, ssCRISPR efficiency predictions are sufficient to select for highly effective gRNAs. Lastly, we showed that ssCRISPR can be used to design efficient gRNAs for strain-specific targeting, isolation, and removal in different strains of *E. coli* and *Pseudomonas*. However, when applying the program to design gRNAs for more complex consortia, the results should be validated in more diverse organisms to ensure that the outputs are accurate. Furthermore, Cas proteins are not functional in all microbial strains. As such, when selecting a CRISPR system for an application, data mining or experimental validation may be required to ensure functionality in the strains of interest.

In summary, we developed ssCRISPR, a user-friendly program for computationally designing strain-specific gRNAs for diverse microbes and CRISPR systems. We validated our

computational tool by testing gRNAs with a wide array of target and non-target strain profiles in *E. coli* and *Pseudomonas spp*. Furthermore, we demonstrated two applications of the program, including the strain-specific isolation and removal of individual microbes from consortia. However, the program can facilitate numerous additional applications in microbiome engineering in humans and the environment(6, 13, 17, 18, 43). ssCRISPR is easily accessible and can be downloaded and run locally as a Python script or as a single package executable application without programming knowledge through the user interface. ssCRISPR will be a valuable tool for managing the health of livestock, plants, and humans, identifying microbes with novel characteristics, exploring dynamics of microbial communities, and tailoring microbiota for improved functions.

Methods

Generating strain selection options and obtaining genome sequences

All programming was performed using Python 3.7, Spyder IDE, and Anaconda software package. A list of bacterial strain names and sequence reference numbers was downloaded from NCBI (https://www.ncbi.nlm.nih.gov/genome/browse#!/prokaryotes/). Strains were filtered for complete genomes to remove partial or incomplete sequences and for bacteria to remove archaea. The list was then imported into the Python program. To create selectable strain choices, the list was sorted alphabetically, and duplicates removed, only maintaining the first sequence in the downloaded list. Genome sequences for the selected target and non-target strains are then individually extracted from the NCBI server using Entrez.efetch and the genome reference numbers. To account for short temporary lapses in the NCBI servers, genome calls are attempted 10 times before drawing an error.

Identifying strain-specific guide RNAs

To generate gRNAs with target sites in all selected target strains, genome sequences are individually extracted from the NCBI database. Locations of all PAM sites are then identified in the genome of the first selected target strain. Next, the specified number of PAM adjacent nucleotides are extracted with the specified orientation relative to the PAM site to generate a string with the gRNA sequence. All identified gRNA sequences are compiled in a list. This gRNA target site identification process is then repeated for the second selected target strain. The two lists of gRNA sequences are then compared and only sequences present in both lists are maintained. This process is repeated for all remaining target strains to generate a list of gRNA sequences, termed here as perfect gRNAs, present in all selected target strains with perfect homology.

To protect strains from gRNA cleavage, the program extracts genome sequences from the NCBI database in batches of 25 strains. Locations for the PAM sequences are then identified from the combined genomes and the respective gRNA sequences extracted and compiled in a list of non-target strain gRNAs. To generate a list of strain-specific gRNAs, gRNA sequences shared between the perfect gRNAs list and the non-target gRNAs list are first removed from the list of perfect gRNAs, resulting in a list of gRNAs with at least 1 nt of specificity. If additional nucleotides of specificity are required, the remaining list of perfect gRNAs are sequentially input into functions that generate lists of all sequence permutations with 1, 2, and 3 nt mismatches and the shared

sequences removed from the list of perfect gRNAs until the desired degree of specificity is reached.

Predicting relative guide RNA cleavage efficiencies

We altered a method of gRNA efficiency predictions previously described by Guo *et al*.(21) The set of 56,335 Cas9 gRNA sequences assessed by Guo *et al*.(21) and 15,000 Cpf1 gRNA sequences assessed by Kim *et al*.(24) were independently analyzed for the following 396 sequence composition and energetic properties: total A, T, C, G and GC content, T content of the four PAM-adjacent nucleotides, presence of an A, T, C, or G in each of the 20 PAM adjacent nucleotides (80 properties), presence of each nucleotide dimer (NN) in each of the 20 PAM adjacent nucleotides (304 properties), minimum free energy for the 12 PAM adjacent nucleotides and the full gRNA sequence, and the melting temperature for the five PAM adjacent nucleotides, next eight nucleotides, remaining nucleotides, and the full gRNA sequence. The resulting property array and the corresponding experimental gRNA cleavage rates were used to train gradient boosting regression machine learning models with a 90:10 split between training group and test group. The models were optimized by tuning the following parameters until the minimum sum squared error was reached for the test groups: the number of boosting stages, the minimum number of samples required to split an internal node, the maximum depth of the tree, and the learning rate.

Plasmids, strains, and growth conditions

The *Pseudomonas* pCas9-RK2K and pSEVA-gRNAT plasmids were purchased from GenScript (catalog numbers MC_0000261 and MC_0000262)(50). Plasmids were designed using SnapGene and assembled in *E. coli* DH10B using the Gibson Assembly (100 mM Tris-HCl, 10 mM $MgCl_2$, 0.2 mM dNTPs, 10 mM DTT, 5% PEG-8000, 1 mM NAD^+, 4 U/μL Taq DNA ligase, 4 U/mL T5 exonuclease, and 25 U/mL Phusion DNA polymerase) or Golden Gate Assembly (1X T4 ligase buffer, 1X Cutsmart buffer, 40 U/μL T4 ligase, 1 U/μL SapI, and 1 U/μL DpnI) methods. Plasmids lethal to *E. coli* DH10B were instead assembled in *E. coli* Nissle 1917. Plasmids harboring both Cas9 and gRNA expression cassettes were assembled in strains expressing AttJ, a TetR-like transcription factor, to repress the P_{attKLM}-*cas9* cassette and minimize toxicity(51). Plasmid DNA was isolated using the PureLink Quick Plasmid Miniprep Kit (K210011, Invitrogen) or PureLink HiPure Plasmid Midiprep Kit (K210005, Invitrogen), and polymerase chain reaction (PCR) products were extracted from electrophoresis gels using the Zymoclean Gel DNA Recovery Kit (D4008, ZYMO research). Chemicals were purchased from Millipore Sigma (Burlington, MA, USA). Enzymes were purchased from New England Biolabs (Ipswich, MA, USA). All Sanger and next-generation sequencing was performed by Genewiz (South Plainfield, NJ, USA). Primers were purchased from Integrated DNA Technologies (Coralville, IA, USA). All plasmids and parts constructed and used in this work are summarized in Tables S4 and S5, respectively.

All strains of *E. coli* used in the study, including DH10B, MG1655, Nissle 1917, and BL21(DE3) were cultured in LB medium at 37°C with 250 rpm shaking unless otherwise stated. Cultures derived from mouse fecal samples were also cultured in LB medium at 37°C with 250 rpm shaking. Medium was supplemented with the following concentrations of antibiotics as necessary: 100 μg/ml ampicillin, 20 μg/ml kanamycin, and 100 μg/ml spectinomycin (Gold Biotechnology, Olivette, MO, USA). *Pseudomonas* strains *P. putida* F1, *P. putida* KT2440, *P. stutzeri* JM300, and *P. syringae* pv. tomato DC3000 were cultured in LB medium with 250 rpm shaking. Cultures

561 containing exclusively *P. putida* F1, *P. putida* KT2440, or *P. stutzeri* JM300 were grown at 30°C.
562 Cultures containing exclusively *P. syringae* pv. tomato DC3000 or mixtures containing multiple
563 *Pseudomonas* strains were grown at 28°C. Medium was supplemented with the following
564 concentrations of antibiotics as necessary: 10 µg/ml gentamycin and 50 µg/ml (*P. putida* F1, *P.*
565 *syringae* pv. tomato DC3000, or *P. stutzeri* JM300) or 200 µg/ml (*P. putida* KT2440 or strain
566 mixtures) tetracyclin (Gold Biotechnology, Olivette, MO, USA).
567
568
569 **gRNA efficiency assays**
570
571 *E. coli*-specific gRNAs were assessed for cleavage efficiency using a chemical transformation cell
572 death assay. Strains were first transformed with a plasmid harboring a constitutive P_{tet}-*cas9*
573 expression cassette but lacking *tetR*. The strains were then incubated overnight in 5 mL of LB in
574 14 mL round bottom tubes (14-959-11B, Fisher Scientific) at 37°C and 250 rpm. Cultures were
575 then diluted 50X into fresh LB supplemented with the relevant antibiotic for the Cas9 plasmid in
576 250 mL baffled Erlenmeyer flasks. Cultures were incubated for ~1.5 h to an OD600 of 0.4 and
577 distributed in 1 mL aliquots in 1.7 mL centrifuge tubes (20383, GeneMate). The tubes were
578 centrifuged at 3000xg for 2 min, the supernatant removed, and the pellets resuspended in 100 µL
579 ice cold 0.1 M $CaCl_2$. Each tube was supplemented with 10 ng of the control plasmid or a gRNA
580 plasmid, gently mixed, and chilled on ice for 20 min. Each tube was then heat shocked in a 42°C
581 water bath for 60 sec and supplemented with 900 µL SOC (5 g/L yeast extract, 20 g/L tryptone,
582 0.5 g/L NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, and 20 mM Glucose). The transformed cells were
583 incubated for 60 min at 37°C and 250 rpm. Culture dilutions were then plated on LB-agar plates
584 with the relevant antibiotics and incubated overnight for CFU quantification. To obtain the
585 efficiency of each gRNA, we calculated the ratio of the number of colonies obtained from each
586 gRNA plasmid (+gRNA) to the number of colonies obtained from the control plasmid (-gRNA) with
587 Cas9 present (+Cas9) divided by the ratio of the number of colonies obtained from each gRNA
588 plasmid to the number of colonies obtained from the control plasmid without Cas9 present (-Cas9)
589 (Equation 1).
590

$$Efficiency = \frac{CFUs_{+Cas9}^{+gRNA}/CFUs_{+Cas9}^{-gRNA}}{CFUs_{-Cas9}^{+gRNA}/CFUs_{-Cas9}^{-gRNA}} \quad (1)$$

591
592 *Pseudomonas*-specific gRNAs were assessed for cleavage efficiency using an electroporation
593 cell death assay. Strains were first transformed with the pCas9-RK2K plasmid which harbors a
594 constitutive Cas9 expression cassette. The strains were then incubated overnight in 5 mL of LB
595 in 14 mL round bottom tubes at 28°C (*P. syringae* pv. tomato DC3000) or 30°C (*P. putida* F1, *P.*
596 *putida* KT2440, or *P. stutzeri* JM300) and 250 rpm. Cultures were then diluted 25X into 50 mL
597 fresh LB supplemented with the relevant antibiotic for the Cas9 plasmid in 250 mL baffled
598 Erlenmeyer flasks. Cultures were incubated for ~2 h to an OD600 of 0.4, centrifuged at 4000xg
599 for 12 min, and washed three times with 50 mL of 3 mM HEPES. The pellet was resuspended in
600 500 µL of 3 mM HEPES and 50 µL aliquots transferred to 1.7 mL centrifuge tubes. Each tube was
601 supplemented with 250 ng of the control plasmid or a gRNA plasmid, gently mixed, electroporated
602 at 2.5 kV (12358-346, Bulldog Bio; Eporator 4309, Eppendorf), and resuspended in 950 µL SOC.
603 The transformed cells were incubated for 2.5 h at 28 or 30°C and 250 rpm. Culture dilutions were
604 then plated on LB-agar plates with the relevant antibiotics and incubated overnight for CFU
605 quantification.
606
607
608 ***E. coli* strain-specific recombineering**

To construct engineered *E. coli* variants, we utilized lambda red-mediated recombineering as previously described(33). The dsDNA insert was obtained by constructing a plasmid with a kanamycin-resistance cassette flanked by 500 bp arms homologous to the *lacZ* insertion region. The full product (both arms and insert DNA) were PCR amplified and purified by gel extraction. *E. coli* MG1655, DH10B, and Nissle 1917 were individually transformed with the pMP11 plasmid containing constitutive Cas9 and arabinose-inducible lambda Red expression cassettes. Individual colonies of each strain were incubated overnight in 5 mL of LB in 14 mL round bottom tubes at 30°C and 250 rpm. Cultures were then mixed and diluted 50X in 50mL of LB supplemented with 2% arabinose in 250 mL baffled Erlenmeyer flasks. Cultures were incubated at 30°C and 250 rpm for an ~2 h to an OD600 of 0.4. Cultures were chilled and washed three times in 50 mL ice cold water, resuspended in 300 µL ice cold water, and 50 µL aliquots transferred to chilled 1.7 mL centrifuge tubes. Tubes were supplemented with 100 ng of the dsDNA insert and 100 ng of either a control plasmid or the strain-selection gRNA plasmid. The cells were electroporated at 2.5 kV, suspended in 950 µL SOC, and incubated at 30°C and 250 rpm for 3 h. Cultures were plated on LB-agar supplemented with spectinomycin and kanamycin to select for cells that received both the control or gRNA plasmid and the integration cassette, respectively. The resulting strains were identified by colony PCR and sequencing.

Isolating or killing specific strains from microbial consortia

For same-genus strain mixtures, all strains were individually incubated overnight in 5 mL of LB in 14 mL round bottom tubes at 37°C (*E. coli*) or 28°C (*Pseudomonas spp.*) and 250 rpm. For *E. coli* and fecal mixtures, cultures were combined and diluted 50X into 50 mL of fresh LB in 250 mL baffled Erlenmeyer flasks and incubated for ~1.5 h to an OD600 of 0.4. For *Pseudomonas spp.*, cultures were combined and diluted 25X into 50 mL of fresh LB in 250 mL baffled Erlenmeyer flasks and incubated for ~2 h to an OD600 of 0.4. The multi-strain cultures were chilled and washed three times in 50 mL ice cold water (*E. coli and fecal*) or 3 mM HEPES (*Pseudomonas spp.*) and resuspended in 500 µL ice cold water (*E. coli and fecal*) or 3 mM HEPES (*Pseudomonas spp.*), and 50 µL aliquots were transferred to chilled 1.7 mL centrifuge tubes. The multi-strain cells were then transformed with 10 ng (*E. coli*) or 250 ng (*Pseudomonas spp.*) of the control plasmid or relevant test plasmid harboring *cas9* and strain-specific gRNA cassettes and resuspended in 950 µL SOC. After a 60 min (*E. coli*) or 2.5 h (*Pseudomonas spp.*), the transformations were plated for the specified cell quantification method.

For NGS strain quantification, transformations were plated onto LB-agar plates supplemented with spectinomycin (*E. coli*) or gentamycin (*Pseudomonas*) and incubated overnight at the respective temperature. All colonies were mixed together and resuspended in 5 mL of LB. The resuspension was then used as a template for a mixed colony PCR with primers harboring NGS adapter sequences (Table S6). PCR products were gel purified and submitted to Genewiz for Amplicon-EZ sequencing. For antibiotic-based quantification, transformations were serially diluted and each plated onto four LB-agar plates with antibiotics matching the resistances of the four strains.

For multi-genus strain mixtures, each strain was individually incubated overnight in 5 mL of LB in 50 mL glass culture tubes (47729-586, VWR International) at 30°C and 250 rpm. Cultures were combined at an OD600 ratio of 1:1:1:1, diluted 50X into fresh LB, and incubated for 2 h. Cultures were then chilled, washed three times with 50 mL ice cold water, resuspended in 500 µL water, aliquoted at 50 µL, and transformed with 100 ng of the control plasmid or test plasmid. Transformations were resuspended in 950 µL SOC, incubated for 2.5 h at 30°C and 250 rpm, and

plated for qPCR-based strain quantification. After 24 h of incubation at 30°C, all resulting colonies were combined, and the genomic DNA extracted using the ZR Fungal/Bacterial DNA MidiPrep kit (D6105, Zymo Research). The genomic DNA was used as the template for quantitative PCR (qPCR) reactions using qPCR primers for each strain (Table S7). qPCR primer pairs for each strain were designed following previously described guidelines(52). SsoAdvanced Universal SYBR Green Supermix (1725270, BioRad), Simi-Skirted 96-well PCR Plates (T-3070-1, GeneMate), and the standard suggested CFX Connect Real-Time System (Bio-Rad) protocol were used for the qPCR reactions. The $2^{-\Delta\Delta CT}$ analysis method was then used to quantify relative population values across samples.

Collecting murine fecal samples

All mouse experiments were approved by the Washington University in St. Louis School of Medicine Institutional Animal Care and Use Committee (Protocol number 21-0160). Mouse experiments were conducted in compliance with the Washington University in St. Louis Biological and Chemical Safety Committee. Finally, mouse experiments were also performed in AAALAC-accredited facilities in accordance with the National Institute of Health guide for the case and use of laboratory animals.

Mouse experiments were performed with 8-week old female C57BL/6 mice (Jackson Labs C57BL/6J, RRID:IMSR_JAX:000664) in a specific pathogen free barrier facility maintained by WUSM DCM. Mice were provided feed (Purina Conventional Mouse Diet (JL Rat/Mouse 6F Auto) #5K67) and water *ad libitum*. Oral gavage of mice was performed using 18ga x 38mm plastic feeding tubes (FTP-18-38, Instech). Mice were administered 20 mg streptopmycin sulfate salt prior to EcN gavage to ablate the native microbiome. 24 hours after streptomycin administration, mice were orally gavaged with 10^8 CFU EcN in 100 µL phosphate buffered saline. 24 hours after EcN administration, fecal samples were collected in sterile 2 mL microtubes and frozen at -80 °C until ready for use.

Liposome synthesis, packaging, and killing assays

Liposomes were generated as previously described(38, 53). The neutral lipid 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (DOPE; 76548, Millipore Sigma) and cationic lipid *N*-[1-(2,3-dioleoyloxy)propyl]-*N,N,N*-trimethylammonium chloride (DOTAP; D6182, Millipore Sigma) were individually dissolved in chloroform at a concentration of 5 mM. The two lipids were then mixed at a 1:1 molar ratio in a 250 mL Büchner flask. The chloroform was removed under a vacuum overnight. The lipid film was rehydrated in 20 mM HEPES at a final concentration of 5 mM of each lipid. The mixture was vortexed for 1 min and sonicated in a 40°C bath sonicator (Branson M3800H) for 30 min. Half of the mixture was removed after 5 min of sonication for protocol optimization experiments. Liposomes were stored at 4°C until used. To package the liposomes with plasmid DNA, liposomes were diluted to the specified concentration in 1 mL of 20 mM HEPES and mixed with 1 µg of plasmid. The mixture was then subjected to five 1-2 min freeze-thaw cycles between liquid nitrogen and a 40°C water bath(54, 55).

To assess the antimicrobial activity of the DNA-loaded liposomes, *E. coli* MG1655, DH10B, BL21(DE3), and Nissle 1917, each harboring a plasmid with a different antibiotic resistance gene, were individually incubated overnight in 5 mL of LB in 14 mL round bottom tubes at 37°C and 250 rpm. Cultures were combined and diluted 40X into 40 mL of fresh LB in 250 mL baffled Erlenmeyer flasks and incubated for an additional ~2 h to an OD600 of ~0.6. 0.5 mL of the exponential phase cultures were aliquoted into 1.7 mL centrifuge tubes and centrifuged at 3000xg for 2 min. The

supernatant was then removed, and the pellet washed with 1 mL 20 mM HEPES. The tube was again centrifuged at 3000xg for 2 min, and the supernatant was removed. The pellet was then resuspended in 0.5 mL of the DNA-loaded liposome mixture. The liposome-*E. coli* mixture was incubated at 37°C and 250 rpm for 30 min. The centrifuge tubes were supplemented with 0.5 mL of SOC medium and returned to the incubator for an additional 60 min. For CFU quantification and cell type identification, cultures were plated onto four LB-agar plates, each supplemented with a different antibiotic.

Next-generation sequencing (NGS)

Amplicon-EZ next generation sequencing was performed by Genewiz to sequence individual DNA strands from purified colony PCR samples obtained from pooled cell samples. The resulting Fastq.gz files were analyzed using custom Python scripts. Two Fastq.gz files were obtained for each sequencing sample (one forward and one reverse); however, only forward reads were analyzed to avoid double counting. Individual sequencing reads were extracted from the files and assessed for read length and sequence. Only sequences of at least 240 nucleotides long were considered. Sequences were compared to the wildtype sequences and counted for the relevant strains: *E. coli* Nissle 1917, MG1655, DH10B, and BL21(DE3) or *P. putida* F1, *putida* KT2440, *stutzeri* JM300, and *syringae* DC3000 (Table S6). Only sequencing reads with a perfect match to one of the strains of interest were counted.

Quantification of the frequency of gRNA target sequences in random DNA

The probability that a gRNA target sequence, including the PAM sequence, will appear in a randomly generated nucleotide sequence was calculated using Equation 2. The equation inaccurately assumes that every nucleotide in the sequence is independently generated without bias. This results in an overestimation in the probability of random occurrence relative to in practice when multiple sequence-similar strains are considered.

$$P = 1 - (1 - (0.25)^{PAM+gT})^{N-PAM-gT} \quad (2)$$

Where

$P =$ Probability that the gRNA target sequence is present in a random nucleotide sequence

$PAM =$ Number of non-random nucleotides in the PAM sequence

$gT =$ Number of nucleotides in the gRNA target site being considered for specificity

$N =$ Length of the random nucleotide sequence

Quantification and statistical analysis

All statistical tests were performed using GraphPad Prism or Excel. All statistical details of experiments, including definition of center, significance criteria, and sample size can be found in the figure legends, in the Results section, or in the Source Data file. Sample sizes were chosen based on our previous work(20, 56) and the literature, and represent sample sizes routinely used for these methods. No sample size calculations were performed during the design of experiments. Samples were randomized during group assignment in all experiments. No samples were excluded from analyses. The Investigators were not blinded to allocation during experiments and outcome assessment.

Data availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by Lead Contact, Tae Seok Moon (tsmoon@wustl.edu).

Materials availability

Plasmids generated in this paper are available upon request from the Lead Contact. This study did not generate additional new unique reagents.

Data availability

All plasmid maps and NGS data were deposited to Mendeley Data (doi: 10.17632/gpgyytwgb5.2; https://data.mendeley.com/datasets/gpgyytwgb5/2). Source data has been provided as a source data file. Any additional information is available from the Lead Contact upon request.

Code availability

All code has been deposited to GitHub (https://github.com/Austin-Rottinghaus/ssCRISPR/). A stand-alone, executable version of the software can be downloaded from Mendelay Data (doi: 10.17632/gpgyytwgb5.2; https://data.mendeley.com/datasets/gpgyytwgb5/2).

Acknowledgments

We thank Prof. Brian Pfleger for the gift of the pMP11 plasmid. We thank Prof. Laura Jarboe and Prof. Barbara Kunkel for their gifts of *P. putida* KT2440 and *P. syringae* pv. tomato DC3000, respectively. We thank Aura Ferreiro of Gautam Dantas' lab for the fecal samples from mice. We also thank members of the Moon lab for helpful suggestions and comments on this work and manuscript. This work was supported by the National Institutes of Health (R01 AT009741), the Office of Naval Research (N00014-17-1-2611, N00014-19-1-2357, and N00014-21-1-2206), the United States Department of Agriculture (2020-33522-32319), National Science Foundation (CBET-1350498 and MCB-2001743), and U.S. Environmental Protection Agency (84020501). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

Author contributions

A.G.R. and T.S.M. conceived the project. A.G.R., S.V., and T.S.M. designed experiments and analyzed the data. A.G.R. and S.V. performed the experiments. A.G.R. and T.S.M. wrote the manuscript.

Declaration of interests

The authors filed a provisional application with the US Patent and Trademark Office on this work.

References

1. B. S. O'Banion, L. O'Neal, G. Alexandre, S. L. Lebeis, Bridging the Gap Between Single-Strain and Community-Level Plant-Microbe Chemical Interactions. *Mol Plant Microbe Interact* **33**, 124-134 (2020).
2. L. H. Morais, H. L. t. Schreiber, S. K. Mazmanian, The gut microbiota-brain axis in behaviour and brain disorders. *Nat Rev Microbiol* **19**, 241-255 (2021).
3. S. Che, Y. Men, Synthetic microbial consortia for biosynthesis and biodegradation: promises and challenges. *J Ind Microbiol Biotechnol* **46**, 1343-1358 (2019).
4. F. Y. Chang *et al.*, Gut-inhabiting Clostridia build human GPCR ligands by conjugating neurotransmitters with diet- and human-derived fatty acids. *Nat Microbiol* **6**, 792-805 (2021).
5. P. Strandwitz *et al.*, GABA-modulating bacteria of the human gut microbiota. *Nat Microbiol* **4**, 396-403 (2019).
6. T. S. Moon, SynMADE: synthetic microbiota across diverse ecosystems. *Trends Biotechnol* **40**, 1405-1414 (2022).
7. F. Prestinaci, P. Pezzotti, A. Pantosti, Antimicrobial resistance: a global multifaceted phenomenon. *Pathog Glob Health* **109**, 309-318 (2015).
8. G. W. Sundin, L. F. Castiblanco, X. Yuan, Q. Zeng, C. H. Yang, Bacterial disease management: challenges, experience, innovation and future prospects: Challenges in Bacterial Molecular Plant Pathology. *Mol Plant Pathol* **17**, 1506-1518 (2016).
9. R. J. Citorik, M. Mimee, T. K. Lu, Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. *Nature biotechnology* **32**, 1141-1145 (2014).
10. A. Reuter *et al.*, Targeted-antibacterial-plasmids (TAPs) combining conjugation and CRISPR/Cas systems achieve strain-specific antibacterial activity. *Nucleic Acids Res* **49**, 3584-3598 (2021).
11. R. Lopez-Igual, J. Bernal-Bayard, A. Rodriguez-Paton, J. M. Ghigo, D. Mazel, Engineered toxin-intein antimicrobials can selectively target and kill antibiotic-resistant bacteria in mixed populations. *Nat Biotechnol* **37**, 755-760 (2019).
12. M. B. Amrofell, A. G. Rottinghaus, T. S. Moon, Engineering microbial diagnostics and therapeutics with smart control. *Current opinion in biotechnology* **66**, 11-17 (2020).
13. J. Ke, B. Wang, Y. Yoshikuni, Microbiome Engineering: Synthetic Biology of Plant-Associated Microbiomes in Sustainable Agriculture. *Trends Biotechnol* **39**, 244-261 (2021).
14. T. Chien *et al.*, Enhancing the tropism of bacteria via genetically programmed biosensors. *Nat Biomed Eng* **6**, 94-104 (2022).
15. C. L. Ho *et al.*, Engineered commensal microbes for diet-mediated colorectal-cancer chemoprevention. *Nat Biomed Eng* **2**, 27-37 (2018).
16. L. M. Kaminsky, R. V. Trexler, R. J. Malik, K. L. Hockett, T. H. Bell, The Inherent Conflicts in Developing Soil Microbial Inoculants. *Trends Biotechnol* **37**, 140-151 (2019).
17. B. E. Rubin *et al.*, Species- and site-specific genome editing in complex bacterial communities. *Nature microbiology* **7**, 34-47 (2022).
18. C. Ronda, S. P. Chen, V. Cabral, S. J. Yaung, H. H. Wang, Metagenomic engineering of the mammalian gut microbiome in situ. *Nat Methods* **16**, 167-170 (2019).

19. A. A. Gomaa *et al.*, Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. *mBio* **5**, e00928-00913 (2014).
20. A. G. Rottinghaus, A. Ferreiro, S. R. S. Fishbein, G. Dantas, T. S. Moon, Genetically stable CRISPR-based kill switches for engineered microbes. *Nat Commun* **13**, 672 (2022).
21. J. Guo *et al.*, Improved sgRNA design in bacteria via genome-wide activity profiling. *Nucleic Acids Res* **46**, 7052-7069 (2018).
22. L. Wang, J. Zhang, Prediction of sgRNA on-target activity in bacteria by deep learning. *BMC Bioinformatics* **20**, 517 (2019).
23. H. K. Kim *et al.*, SpCas9 activity prediction by DeepSpCas9, a deep learning-based model with high generalization performance. *Sci Adv* **5**, eaax9249 (2019).
24. H. K. Kim *et al.*, Deep learning improves prediction of CRISPR-Cpf1 guide RNA activity. *Nat Biotechnol* **36**, 239-241 (2018).
25. P. D. Hsu *et al.*, DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827-832 (2013).
26. M. Jinek *et al.*, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y* **337**, 816-821 (2012).
27. R. P. Hayes *et al.*, Structural basis for promiscuous PAM recognition in type I-E Cascade from E. coli. *Nature* **530**, 499-503 (2016).
28. H. K. Kim *et al.*, In vivo high-throughput profiling of CRISPR-Cpf1 activity. *Nat Methods* **14**, 153-159 (2017).
29. S. Shmakov *et al.*, Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol Cell* **60**, 385-397 (2015).
30. H. Feng, J. Guo, T. Wang, C. Zhang, X. H. Xing, Guide-target mismatch effects on dCas9-sgRNA binding activity in living bacterial cells. *Nucleic Acids Res* **49**, 1263-1277 (2021).
31. E. Semenova *et al.*, Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A* **108**, 10098-10103 (2011).
32. M. Land *et al.*, Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics* **15**, 141-161 (2015).
33. C. R. Mehrer, M. R. Incha, M. C. Politz, B. F. Pfleger, Anaerobic production of medium-chain fatty alcohols via a beta-reduction pathway. *Metab Eng* **48**, 63-71 (2018).
34. W. Jiang, D. Bikard, D. Cox, F. Zhang, L. A. Marraffini, RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology* **31**, 233-239 (2013).
35. C. C. Campa, N. R. Weisbach, A. J. Santinha, D. Incarnato, R. J. Platt, Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. *Nat Methods* **16**, 887-893 (2019).
36. A. C. Reis *et al.*, Simultaneous repression of multiple bacterial genes using nonrepetitive extra-long sgRNA arrays. *Nat Biotechnol* **37**, 1294-1301 (2019).
37. P. Tao *et al.*, In vitro and in vivo delivery of genes and proteins using the bacteriophage T4 DNA packaging machine. *Proc Natl Acad Sci U S A* **110**, 5846-5851 (2013).
38. S. Pereira *et al.*, Lipoplexes to Deliver Oligonucleotides in Gram-Positive and Gram-Negative Bacteria: Towards Treatment of Blood Infections. *Pharmaceutics* **13** (2021).
39. M. B. N. Albright *et al.*, Solutions in microbiome engineering: prioritizing barriers to organism establishment. *ISME J* **16**, 331-338 (2022).

40. W. D. Xian *et al.*, Network-directed efficient isolation of previously uncultivated Chloroflexi and related bacteria in hot spring microbial mats. *NPJ Biofilms Microbiomes* **6**, 20 (2020).
41. E. J. Stewart, Growing unculturable bacteria. *J Bacteriol* **194**, 4151-4160 (2012).
42. E. Palma, B. Tilocca, P. Roncada, Antimicrobial Resistance in Veterinary Medicine: An Overview. *Int J Mol Sci* **21** (2020).
43. K. Neil *et al.*, High-efficiency delivery of CRISPR-Cas9 by engineered probiotics enables precise microbiome editing. *Mol Syst Biol* **17**, e10335 (2021).
44. Z. Hosseinidoust, T. G. van de Ven, N. Tufenkji, Evolution of Pseudomonas aeruginosa virulence as a result of phage predation. *Appl Environ Microbiol* **79**, 6110-6116 (2013).
45. V. L. Waters, Conjugative transfer in the dissemination of beta-lactam and aminoglycoside resistance. *Front Biosci* **4**, D433-456 (1999).
46. D. H. Burke, Cell-penetrating RNAs: new keys to the castle. *Mol Ther* **20**, 251-253 (2012).
47. K. Kanekura *et al.*, Characterization of membrane penetration and cytotoxicity of C9orf72-encoding arginine-rich dipeptides. *Sci Rep* **8**, 12740 (2018).
48. D. Carugo, E. Bottaro, J. Owen, E. Stride, C. Nastruzzi, Liposome production by microfluidics: potential and limiting factors. *Sci Rep* **6**, 25876 (2016).
49. K. J. Locey, J. T. Lennon, Scaling laws predict global microbial diversity. *Proc Natl Acad Sci U S A* **113**, 5970-5975 (2016).
50. J. Sun *et al.*, Genome editing and transcriptional repression in Pseudomonas putida KT2440 via the type II CRISPR system. *Microb Cell Fact* **17**, 41 (2018).
51. H. B. Zhang, L. H. Wang, L. H. Zhang, Genetic control of quorum-sensing signal turnover in Agrobacterium tumefaciens. *Proc Natl Acad Sci U S A* **99**, 4638-4643 (2002).
52. S. A. Bustin *et al.*, The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical chemistry* **55**, 611-622 (2009).
53. B. K. Kim *et al.*, DOTAP/DOPE ratio and cell type determine transfection efficiency with DOTAP-liposomes. *Biochim Biophys Acta* **1848**, 1996-2001 (2015).
54. T. Gjetting *et al.*, A simple protocol for preparation of a liposomal vesicle with encapsulated plasmid DNA that mediate high accumulation and reporter gene activity in tumor tissue. *Results Pharma Sci* **1**, 49-56 (2011).
55. W. L. Fotoran, R. Santangelo, B. N. M. de Miranda, D. J. Irvine, G. Wunderlich, DNA-Loaded Cationic Liposomes Efficiently Function as a Vaccine against Malarial Proteins. *Mol Ther Methods Clin Dev* **7**, 1-10 (2017).
56. A. G. Rottinghaus, C. Xi, M. B. Amrofell, H. Yi, T. S. Moon, Engineering ligand-specific biosensors for aromatic amino acids and neurochemicals. *Cell Syst* 10.1016/j.cels.2021.10.006 (2021).

Figure captions

Fig. 1: ssCRISPR program logic flowchart for strain-specific gRNA design.
The user first inputs the desired non-target strains, target strains, nucleotides of specificity (1-4 nt), PAM sequences and orientation (5' or 3'), and target length (grey). The program searches the first selected target strain for all potential gRNA target sites using the user-specified PAM sequence, PAM orientation, and target length. Next, the program iterates through all additional selected target strains and identifies the gRNA target sequences that are perfectly shared between the strains (green). The program then identifies gRNA target sites in batches of non-target strains and eliminates gRNAs that have less than the specified nucleotides of specificity to the any non-target strain (blue). Finally, for Cas9 and Cpf1 gRNAs, the program predicts the relative efficiencies of the determined gRNAs using 396 sequence composition and energetic properties. The gRNAs are ranked by their relative efficiency and a full report of the results is provided to the user (yellow). The number of gRNAs tested for specificity is capped at *10,000 for 2 nt mismatches and **100 for 3 nt mismatches due to limits in computation power.

Fig. 2: Computational design of gRNAs with broad strain specificity.
(A) The number of gRNAs that broadly target different amounts of each of the 2,068 *E. coli* and 1,020 *Pseudomonas* strains. **(B)** Actual versus predicted efficiency rankings for 56,335 Cas9 gRNAs. Actual efficiency values were obtained from Guo et al(21). Predicted efficiency rankings were determined using a modified approach from Guo et al (See methods). **(C and D)** Efficiency values for the top four predicted gRNAs that target (C) all *E. coli* strains (gRNAs All-E1, All-E2, All-E3, and All-E4) in *E. coli* DH10B, Nissle 1917, MG1655, and BL21(DE3) or (D) all *Pseudomonas* strains (gRNAs All-P1, All-P2, All-P3, and All-P4) in *P. putida* F1, *P. putida* KT2440, *P. stutzeri* JM300, and *P. syringae* pv. tomato DC3000. Efficiency values were obtained using cell death transformation assays. Efficiency values are the ratio of the number of colonies obtained from each gRNA plasmid to the number of colonies obtained from the control plasmid with Cas9 present divided by the ratio of colonies obtained from each gRNA plasmid to the number of colonies obtained from the control plasmid without Cas9 present (see Methods). Values and error bars are the average and standard deviation of biological triplicate, respectively. Source data are provided as a source data file.

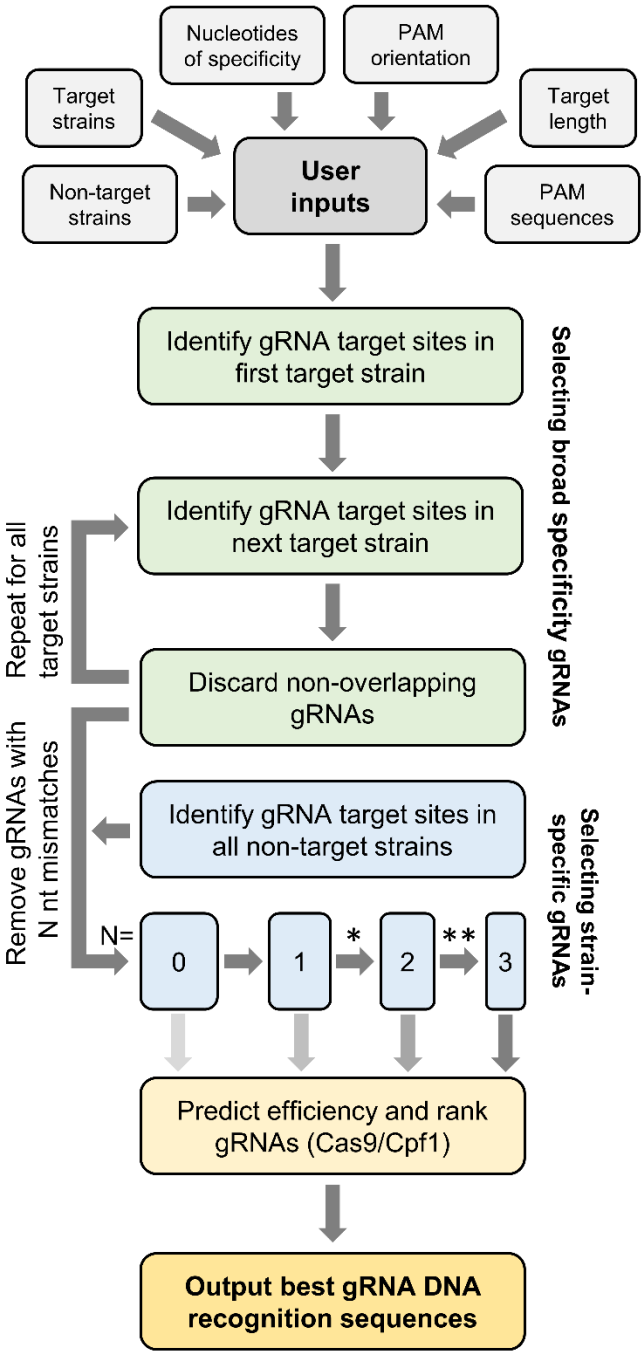
Fig. 3: Computational design of strain-specific gRNAs.
(A) Efficiency of the top scoring strain-specific gRNAs with at least one mismatched nucleotide (nt) in the PAM or at least (left) 1 mismatched nucleotide in the 10 nt PAM-adjacent target region, (middle) 2 mismatched nucleotides in the 10 nt PAM-adjacent target region, or (right) 3 mismatched nucleotides in the 12 nt PAM-adjacent target region. gRNAs were designed to selectively target *E. coli* DH10B, Nissle 1917, MG1655, or BL21(DE3). (**B and C)** Efficiency of strain-specific gRNAs with at least one mismatched nucleotide in the PAM or at least 3 mismatched nucleotides in the 20 nt PAM-adjacent target region. gRNAs were designed to selectively target (B) *E. coli* DH10B, Nissle 1917, MG1655, or BL21(DE3), or (C) *P. putida* F1, *P. putida* KT2440, *P. stutzeri* JM3000, or *P. syringae* pv. tomato DC3000. The top four predicted gRNAs for each strain were selected from the program and tested for killing efficiency using a transformation assay. Efficiency values are the ratio of the number of colonies obtained from each gRNA plasmid to the number of colonies obtained from the control plasmid with Cas9 present divided by the ratio of the number of colonies obtained from each gRNA plasmid to the number of colonies obtained from the control plasmid without Cas9 present (see methods). Each value is the average of biological duplicate (A) or triplicate (B and C). Source data are provided as a source data file.

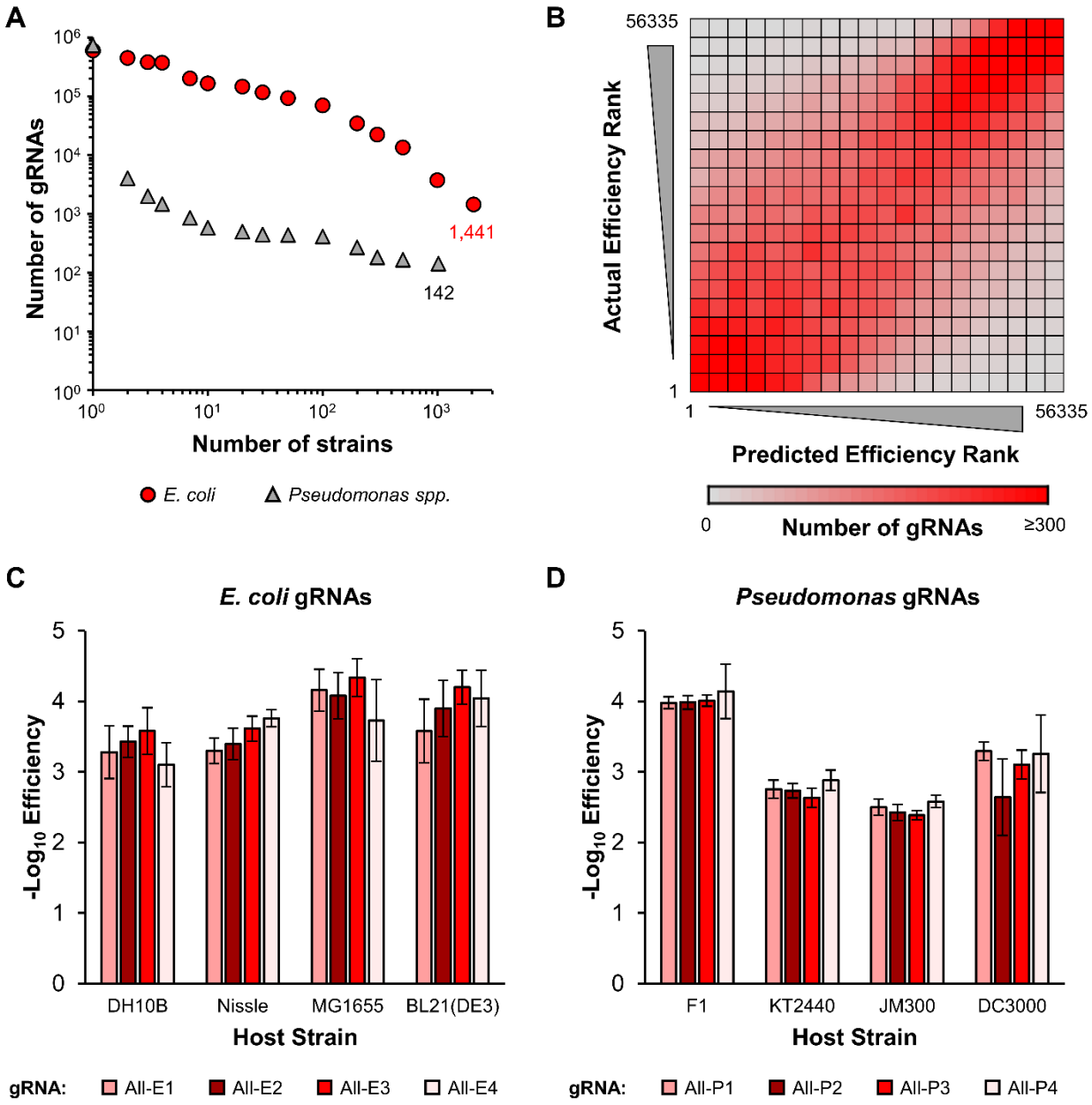
Fig. 4: Isolation of specific bacteria from microbial consortia using strain-specific CRISPR-Cas9 selection

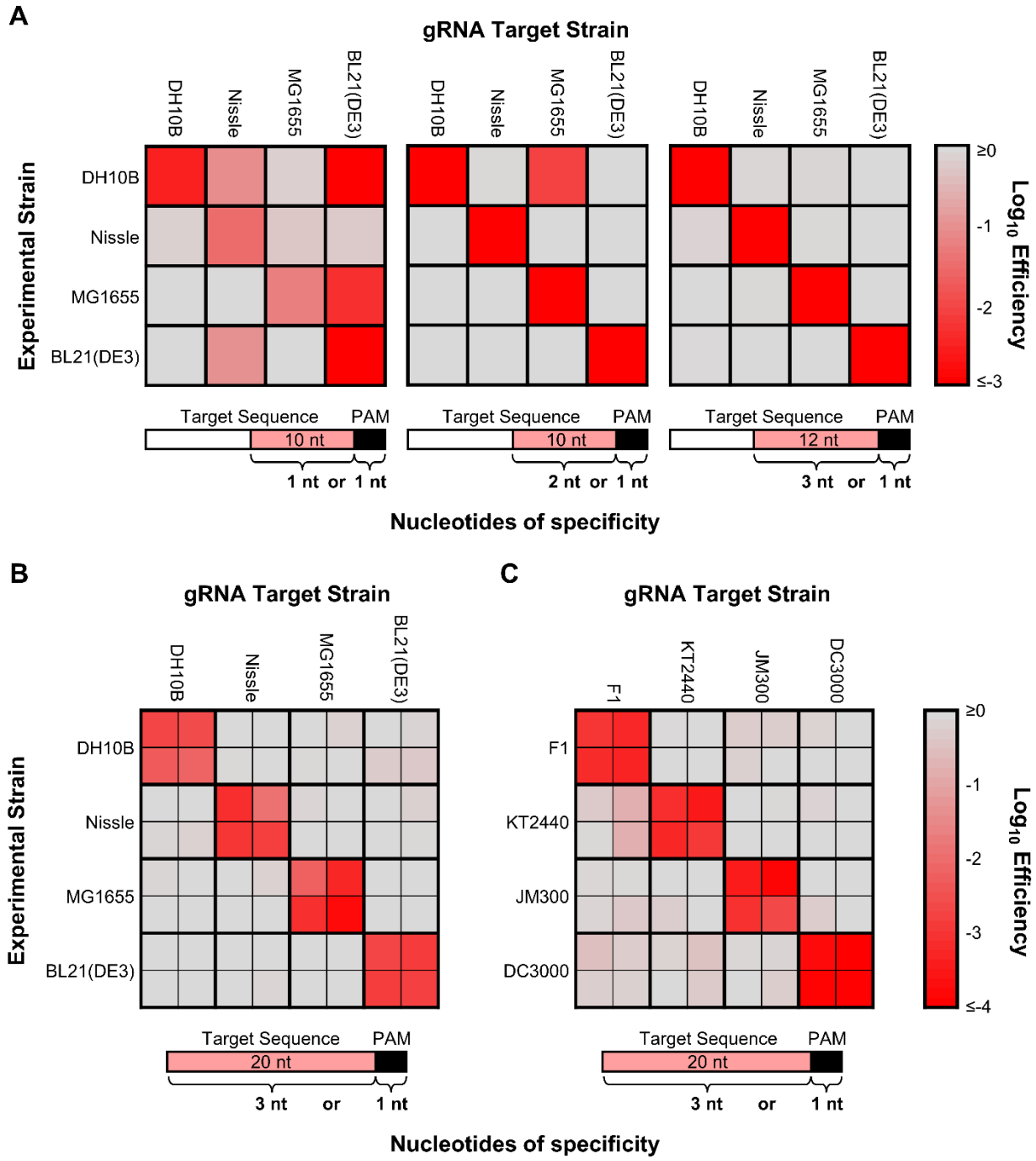
(A) Procedural schematic for isolating specific strains from a consortium. Selected consortia are transformed with a Cas9- and lambda red-containing CRISPR plasmid. The strain mixture with the CRISPR plasmid is then transformed with a strain-specific gRNA plasmid, designed to target selected non-desired strains, and double stranded DNA carrying an antibiotic resistance gene (ARG). The ARG is integrated into the genome by lambda red recombinase to yield antibiotic-resistant microbes. Recombinants are then isolated by plating on agar plates containing the relevant antibiotic. The transformed gRNA plasmid selectively kills non-desired strains, leaving viable colonies only of the desired microbe. If the purified strain is desired to be further engineered, a second round of recombination can be performed using an ARG-specific gRNA to replace the ARG with any DNA of interest. **(B)** Isolation of *E. coli* Nissle 1917 from a three-microbe consortium with *E. coli* DH10B and MG1655. Cultures of each of the three microbes alone and together at a 1:1:1 ratio were transformed with a kanamycin-resistance cassette and either a control plasmid (gRNA –) or a plasmid harboring a gRNA designed to target *E. coli* DH10B and MG1655, but not *E. coli* Nissle 1917. **(C)** Plasmid schematic for strain-specific isolation of microbes from consortia. Six gRNAs were designed to target different subsets of the *Enterobacteriaceae* family, while protecting *E. coli* Nissle 1917. Each gRNA is expressed in its own unique cassette with nonrepetitive constitutive promoters, Cas9 hairpins, terminators, and spacer regions. **(D)** Isolation of *E. coli* Nissle from defined single-genus and multi-genus microbial consortia. Cultures of *E. coli* DH10B, MG1655, BL21(DE3), and Nissle 1917 or *E. coli* Nissle 1917, *P. putida* F1, *S. typhimurium*, and *R. opacus* PD630 were mixed at a 1:1:1:1 ratio and transformed with an empty control plasmid or a plasmid harboring a constitutive Cas9 cassette and an *Enterobacteriaceae*-targeting but *E. coli* Nissle-protecting gRNA array. Strains were quantified by next-generation amplicon sequencing (left) or qPCR (right). Values and error bars are the average and standard deviation of biological triplicate, respectively. Statistical comparisons between the control plasmid and gRNA array plasmid were performed using two-sided two-way ANOVA with Sidak's multiple comparisons (***, $p < 0.001$; ****, $p < 0.0001$). Source data and *p*-values are provided as a source data file.

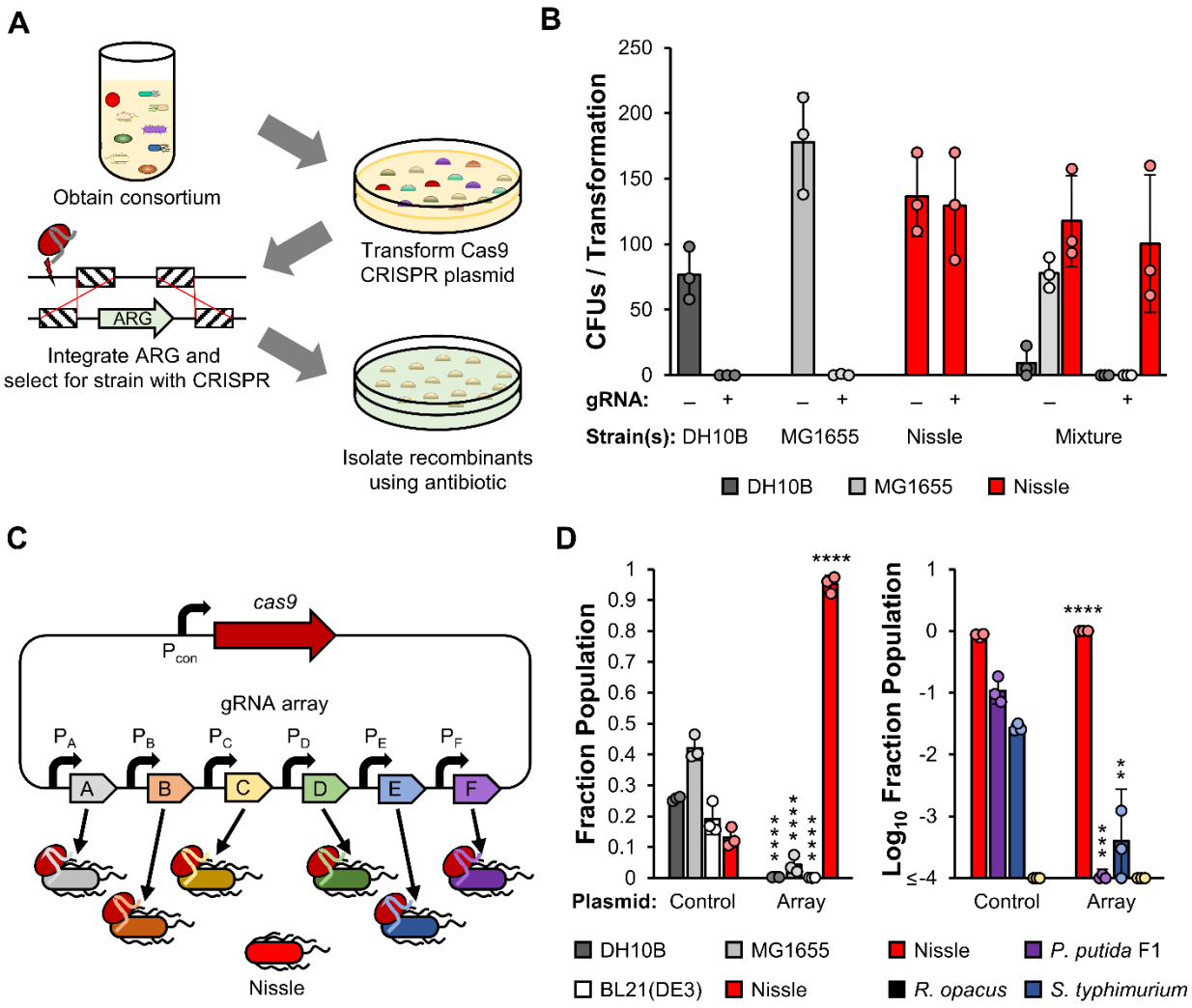
Fig. 5: Removal of individual microbes from consortia using strain-specific CRISPR-Cas9 DNA antimicrobials

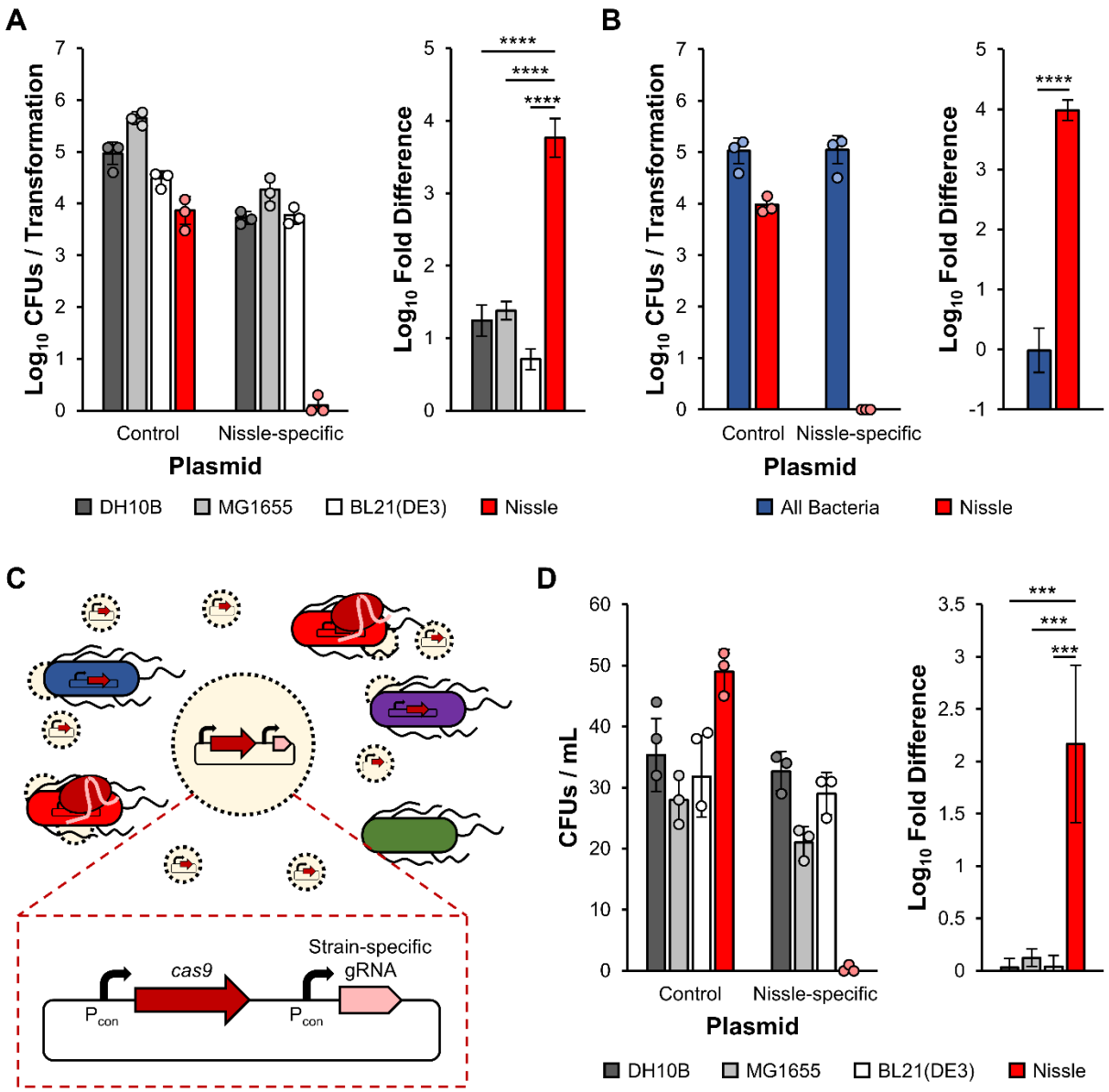
(A and B) Plasmids harboring constitutive Cas9 and *E. coli* Nissle-specific gRNAs selectively remove *E. coli* Nissle from microbial consortia. Defined consortia of (A) a 1:1:1:1 mixture of *E. coli* DH10B, MG1655, BL21(DE3), and Nissle 1917 or (B) mouse fecal samples containing ~2% *E. coli* Nissle were transformed with a control plasmid or an *E. coli* Nissle-specific targeting plasmid, and the strains were identified by antibiotic plating (see Methods). The fold difference in CFUs between transformation with the control plasmid and *E. coli* Nissle-specific plasmid was then quantified. **(C)** Schematic of strain-specific antimicrobial liposomes. Cationic liposomes packaged with plasmids harboring Cas9 and strain-specific gRNA cassettes are delivered to complex microbial consortia. Liposomes nonspecifically fuse with microbes, delivering the payload. Microbes harboring the gRNA target sequence have their genome inactivated by Cas9 cleavage, causing cell death. **(D)** CFUs and fold difference of *E. coli* DH10B, MG1655, BL21(DE3), and Nissle 1917 that received control plasmid and *E. coli* Nissle-specific plasmid payloads after incubation with DNA-loaded liposomes. Values and error bars are the average and standard deviation of biological triplicate, respectively. Statistical comparisons between the control plasmid and the Nissle-specific plasmid were performed using two-sided one-way ANOVA with Tukey's Honest Significant Difference post-hox test (***, $p < 0.001$; ****, $p < 0.0001$). Source data and *p*-values are provided as a source data file.


Nucleotides of specificity
PAM orientation
Target strains
Target length
Non-target strains
User inputs
PAM sequences
Identify gRNA target sites in first target strain
Identify gRNA target sites in next target strain
Repeat for all target strains
Discard non-overlapping gRNAs
Selecting broad specificity gRNAs
Remove gRNAs with N nt mismatches
Identify gRNA target sites in all non-target strains
N=
0
1
*
2
**
3
Selecting strain-specific gRNAs
Predict efficiency and rank gRNAs (Cas9/Cpf1)
Output best gRNA DNA recognition sequences


A
Number of gRNAs
Number of strains
1,441
142
E. coli
Pseudomonas spp.
B
56335
Actual Efficiency Rank
1
1
56335
Predicted Efficiency Rank
0
Number of gRNAs
≥300
C
E. coli gRNAs
-Log10 Efficiency
DH10B
Nissle
MG1655
BL21(DE3)
Host Strain
gRNA:
All-E1
All-E2
All-E3
All-E4
D
Pseudomonas gRNAs
-Log10 Efficiency
F1
KT2440
JM300
DC3000
Host Strain
gRNA:
All-P1
All-P2
All-P3
All-P4


A
gRNA Target Strain
DH10B
Nissle
MG1655
BL21(DE3)
Experimental Strain
Target Sequence
PAM
10 nt
1 nt or 1 nt
2 nt or 1 nt
12 nt
3 nt or 1 nt
Nucleotides of specificity
Log10 Efficiency
≥0
-1
-2
≤-3
B
C
F1
KT2440
JM300
DC3000
20 nt
3 nt or 1 nt
-3
≤-4


A
Obtain consortium
ARG
Integrate ARG and
select for strain with CRISPR
Transform Cas9
CRISPR plasmid
Isolate recombinants
using antibiotic
B
CFUs / Transformation
250
200
150
100
50
0
gRNA: – + – + – + – +
Strain(s): DH10B MG1655 Nissle Mixture
DH10B MG1655 Nissle
C
cas9
P con
gRNA array
P A P B P C P D P E P F
A B C D E F
Nissle
D
Fraction Population
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
****
Plasmid: Control Array
DH10B MG1655 Nissle P. putida F1
BL21(DE3) Nissle R. opacus S. typhimurium
Log10 Fraction Population
1
0
-1
-2
-3
≤-4
****
Control Array


A
Log10 CFUs / Transformation
Log10 Fold Difference
****
****
****
Control
Nissle-specific
Plasmid
DH10B
MG1655
BL21(DE3)
Nissle
B
Log10 CFUs / Transformation
Log10 Fold Difference
****
Control
Nissle-specific
Plasmid
All Bacteria
Nissle
C
cas9
Strain-specific gRNA
Pcon
Pcon
D
CFUs / mL
Log10 Fold Difference
***
***
***
Control
Nissle-specific
Plasmid
DH10B
MG1655
BL21(DE3)
Nissle