Parallel CRISPR-Cas guided mutagenesis of chromosome and virulence plasmid in *Shigella flexneri* by cytosine base editing

Atin Sharma[a,b,c], Ruqiya Omer Aden[a,b,c], Andrea Puhar[a,b,c#], David A. Cisneros[a,b#]

[a]Department of Molecular Biology, Umeå University, Umeå, Sweden

[b]Umeå Centre for Microbial Research, Umeå University, Umeå, Sweden

[c]The Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, Umeå, Sweden

Running Head: Parallel programmable CRISPR-Cas mutations in *Shigella*

#Address correspondence to Andrea Puhar, andrea.puhar@umu.se, or David A. Cisneros, david.cisneros@umu.se

Andrea Puhar and David A. Cisneros contributed equally to this work as senior authors

Word count for Abstract: 239

Word count for Importance: 144

Word count for the main text (Excluding Acknowledgement, References, Figure legends, Tables, and Supplementary Material): 4925

Abstract

Shigella is a Gram-negative bacterium that invades the human gut epithelium. The resulting infection, shigellosis, is the deadliest bacterial diarrhoeal disease. Much of the information about the genes dictating the pathophysiology of *Shigella*, both on the chromosome and the virulence plasmid, was obtained by classical reverse genetics. However, technical limitations of the prevalent mutagenesis techniques restrict the generation of mutants in a single reaction to a small number, preventing large scale targeted mutagenesis of *Shigella* and the subsequent assessment of phenotype. We adopted a CRISPR-Cas dependent approach, where a nickase Cas9 and cytidine deaminase fusion is guided by sgRNA to introduce targeted C→T transitions, resulting in internal STOP codons and premature termination of translation. In proof-of-principle experiments using an *mCherry* fluorescent reporter assay, we were able to generate loss-of-function mutants in both *E. coli* and *Shigella* with up to 100% efficacy. Using a modified fluctuation assay, we determined that under optimized conditions, the frequency of untargeted mutations introduced by the Cas9-deaminase fusion is in the same range as spontaneous mutations, making our method a safe choice for bacterial mutagenesis. Further, we programmed the method to mutate well-characterized chromosomal and plasmid-borne *Shigella* genes and found the mutant phenotype to be similar to that of the reported gene deletion mutants, with no apparent polar effects. This method can be used in a 96-well plate format to increase the throughput and generate an array of targeted loss-of-function mutants in a few days.

Importance:

Loss-of-function mutagenesis is critical in understanding the physiological role of genes. Therefore, high-throughput techniques to generate such mutants are important for facilitating the assessment of gene function at a pace that matches system biology approaches. However,

to our knowledge, no such method was available for generating an array of single gene mutants in an important enteropathogen – *Shigella*. This pathogen causes high morbidity and mortality in children and antibiotic resistant strains are quickly emerging. Therefore, determination of the function of unknown *Shigella* genes is of utmost importance to develop effective strategies to control infections. Our present work will bridge this gap by providing a rapid method for generating loss-of-function mutants. The highly effective and specific method can be programmed to generated multiple mutants in a single, massively parallel, reaction. By the virtue of plasmid compatibility, this method can be extended to other members of Enterobacteriaceae.

Introduction

***S**higella* is a Gram-negative bacterium that belongs to the Enterobacteriaceae family. *Shigella flexneri* is the most prevalent species resulting in the disease – shigellosis or bacillary dysentery (1). Shigellosis is mostly a self-limiting disease in healthy adults but poses a major threat to children, elderly and the immunocompromised (2). With an annual death toll of 212,000, of which 63,713 are children below the age of 5, shigellosis is the deadliest bacterial diarrheal disease (3). Very low infection doses (10-100 bacteria), easy person-to-person spread and the acute inflammatory response are responsible for such high rates of incidence and mortality, especially in low- and medium-income countries (4). Being an intracellular enteric pathogen, *Shigella* gains access to the intestinal epithelial cells by attaching and inducing its own uptake. Once inside the cells, it lyses the vacuole, replicates, and by using actin-based motility, disseminates to neighboring cells (reviewed in (5)). Virulence of *Shigella* requires a virulence plasmid-encoded type III secretion system (T3SS) which injects proteins into the host cell aiding the bacterium in establishment of infection thorough invasion, subversion of host defenses, and dissemination (reviewed in (6)).

We recently sequenced the genome of *S. flexneri* serotype 5a M90T, a widely used lab reference strain, and mapped global transcription start sites (7). Unsurprisingly, many transcription start sites led to identification of genes that await functional characterization to reveal new and interesting information about *Shigella* physiology. Loss-of-function mutagenesis, and subsequent assessment of phenotype, is a common and perhaps the most convenient way to determine the function of unknown genes. The classic allele replacement technique, first described decades ago by Datsenko et al (8), is still the method of choice for targeted mutations in *Shigella* (9, 10) but it has some limitations. Since the success of this technique is directly dependent on transformation of a linear DNA construct, the requirement of a large amount of DNA and electroporation limit its throughput. The two-step process, allele

replacement and subsequent removal, leaves a scar sequence which, if not optimized for being in-frame, results in polar effects that affect phenotype assessment (11). Although PCR-mediated construction of a linear DNA construct is easy, it requires expensive long primers and electroporation which adds to the final cost and acts as a technical bottleneck. Additionally, the loss of virulence plasmid in *Shigella*, which could be enhanced by growth at higher temperatures, is a deterrent for large scale parallel mutagenesis in *Shigella*.

CRISPR-Cas, a bacterial defense system against phages, has gained much attention for its use in genetic modifications (12). The *Streptococcus pyogenes* Cas9, the most commonly used version, can be directed, using guide RNA (gRNA), to introduce double-stranded (ds) DNA breaks (13). In eukaryotes, this is subsequently followed by DNA repair, resulting in generation of mutants, but in prokaryotes, a dsDNA break is lethal (14). Therefore, CRISPR-Cas based mutagenesis became the method of choice for targeted mutations in the eukaryotes but its usage in prokaryotes was initially restricted to increasing the overall efficiency of allele replacement by counterselection of false positives (15, 16). However, later modifications resulted in advent of base editors or catalytically impaired Cas9 proteins fused to deaminases, resulting in base substitution and, eventually, loss-of-function mutations (17, 18). These methods avoid dsDNA breaks and result in marker-less mutations. Moreover, the requirement of plasmid borne gRNA and effector proteins (Cas9 variants) eliminates the need of electroporation, thereby increasing the efficacy of the process, making it possible to multiplex or generate massively parallel mutations (19–21).

Here, we describe a CRISPR-Cas mediated base editing method, using an established cytidine base editor (19, 21) that can be programmed to generate loss-of-function mutations in *S. flexneri*. By mutating well-characterized *S. flexneri* genes of known phenotype/function, encoded on both the virulence plasmid and the chromosome, we show that the one-step method is easy, fast, and highly effective in inactivating genes. Utilizing this method, multiple genes

in *S. flexneri*, can be mutated in parallel arrays making this method a great choice of high-throughput mutagenesis.

Results

A two plasmid-based system for expression of sgRNA and nCas9-AID was constructed

As an effector protein, we selected a fusion of catalytically inactive Cas9 – nickase Cas9 ($nCas9^{D10A}$) and an ortholog of activation-induced cytidine deaminase (AID) that has been shown to be an effective base editor in a Gram-positive industrial bacterium – *Corynebacterium glutamicum* (21). The fusion protein is guided to the target gene by sgRNA (single guide RNA – 20 nt target-specific 'spacer' fused to a gRNA scaffold) where, within a window of -20 to -16 from the protospacer adjacent motif (PAM), the AID deaminates cytosine (C) to Uracil (U) and the nCas9 nicks the unmodified strand resulting in its replacement and C→T (G→A on opposite strand) transition (Fig. 1B). This leads to conversion of internal CAG, CGA, and CAA codons on the coding strand or CCA on the non-coding strand into TAG, TGA, and TAA (STOP codons), premature termination of translation, and loss-of-function mutation. The nCas9-AID fusion, with *tac* promoter and the LacI repressor, was combined with the plasmid backbone from pSU19 (22) to generate a low-copy plasmid – pnCas9-AID (Fig. 1A). SacB encoding gene was cloned to facilitate curing of this plasmid (23).

To express the sgRNA, the gRNA expression cassette from pgRNA_ccdB (21), was combined with the plasmid backbone of pBluescript SK. In the resulting high copy plasmid, pgRNA_AT, spacers can be cloned under the J23119 promoter resulting in a transcriptional fusion with the gRNA scaffold and hence, a fully functional sgRNA.

The *tac* promoter, used to drive the expression of nCas9-AID, is already known to function in *Shigella* (24). However, to determine J23119 activity in *Shigella*, superfolder GFP (sfGFP) with a synthetic ribosome binding site was cloned in pgRNA_AT and transformed in *S. flexneri* (Supplementary Fig. S1). Expression of green fluorescence verified the strong constitutive

activity of J23119 promoter in *S. flexneri*.

To program a desired mutation, we used an online program – CRISPR-CBEI (25). Using sequence information, it determines mutable sites and lists spacers required for the desired mutation (Fig. 1C). We used a *ccdB* counterselection-based high efficiency cloning strategy (21) to clone spacers forming new sgRNA expression plasmids (26). The spacers from CRISPR-CBEI output were synthesized, with overhangs complementary to the ends of BsaI digested pgRNA_AT to ensure directional cloning (Fig. 1D). The oligos were hybridized and phosphorylated prior to ligation and transformed in *E. coli* DH5α. All the transformants obtained carried the required recombinant plasmid as the cells transformed with non-recombinant plasmid are killed (21, 26).

Base editing effectively inactivates chromosomal genes in *E. coli*

With optimal tools in hand, we wanted to benchmark our method in *E. coli*, where base editing has been shown to be effective (20). We generated a chromosomally tagged fluorescent strain of *E. coli* (*E. coli*::*mCherry*) for an easy-to-read phenotype. Spacer oligos that introduce premature STOP codons in *mCherry* at three different positions – 205, 340, and 294 nt, were cloned in pgRNA_AT resulting in pgRNA_m2, pgRNA_m3, and pgRNA_m4. A plasmid, pgRNA-X, expressing a non-binding random spacer oligo was also constructed. As a testament to the high efficacy of the cloning technique, all the recombinant sgRNA plasmids sequenced had oligos cloned in the correct orientation.

The pnCas9-AID and sgRNA plasmids were co-transformed in *E. coli*::*mCherry*. pSU19 and pgRNA-X were used as control plasmids to account for Cas9-independent untargeted random mutagenesis (Fig. 2A). Actively growing cultures of the cotransformants were induced with IPTG and 10-fold dilutions of these cultures were spotted on LA plates at different time points post-induction. The resulting colonies were counted and scanned for red fluorescence. As expected, strains expressing sgRNA (targeting or non-targeting) alone were still fluorescent

(Fig. 2B). The colonies lost fluorescence only when both nCas9-AID and gene-specific sgRNA were expressed, indicating high specificity and low background mutation rates. The frequency of mutation (*f*) was high, even after a short interval of 2h, and some gRNA were more effective than the others (Fig. 2C). With median *f* = 100% at 2h for both gRNA_m2 and gRNA_m4, they were more effective than gRNA_m3, whose median *f* was 87.8%. Interestingly, the frequency was high even in the absence of the inducer, IPTG, and only a meager increase in case of suboptimal gRNA_m3 was seen at 2h (median 87.8% vs 91.6%). At later time points, however, IPTG induction improved the frequency of mutation for gRNA_m3 (median 100% at 6h and 20h post-induction). The impact of nCas9-AID expression on bacterial viability was minimal (Supplementary Fig. S2). A slight decrease in the viability was only observed at early time points (2h), which was exacerbated by the addition of IPTG, but diminished at later time points. These results indicated that nCas9-AID could be programmed to direct mutagenesis in *E. coli*, without greatly affecting the bacterial viability. The mutagenesis was rather quick and therefore, long incubation times were not necessary.

Base editing effectively inactivates chromosomal genes in *S. flexneri*

To determine the efficacy of base editing in *S. flexneri*, we generated a strain (*S. flexneri*::*mCherry*) constitutively expressing *mCherry* from its chromosome and carried out mutagenesis using the same guide RNA expressing plasmids and conditions that were used in case of *E. coli* (Fig. 3A).

Like the high-specificity pattern observed in *E. coli*, fluorescence was lost only when nCas9-AID and gene-specific sgRNA were expressed together (Fig. 3B). Also, different sgRNA resulted in varying frequency of mutation and the addition of IPTG didn't seem to significantly alter it. However, the frequency of mutation, with the same sgRNA, was significantly lower in *S. flexneri* as compared to that in *E. coli*. At 2h post induction, gRNA_m4 was only about half as effective in *S. flexneri* (Fig. 3C). The sharpest decline was, however, seen in case of

gRNA_m3. The median frequency for it was highest at only 16.7%, even after 6h with IPTG induction. Although the efficacy of all the three sgRNA at later points was mostly similar in *E. coli*, the subtle differences were more apparent in *S. flexneri*, underlining the fact that despite belonging to the same family, the outcomes of molecular processes are much varied. The best performing sgRNA in *S. flexneri*, gRNA_m2, had median f = 85.7% at 2h which saw only a marginal increase to 92.3% upon addition of IPTG. These results suggested that although the mutagenesis works with similar specificity in *S. flexneri* as it does in *E. coli*, efficacies tend to vary and using certain sgRNA, high efficacies are achievable. Longer incubation times, after induction with IPTG, aren't necessary and IPTG induction itself only has a minor effect on bacterial viability (Supplementary Fig. S3).

nCas9-AID expressing strains are not hypermutators

Although it was evident from robust results that the base editing mutagenesis was highly specific, this apparent high specificity could have been an outcome of accelerated rate of mutagenesis in nCas9-AID expressing strains due to uncontrolled activity of the fused AID domain (27, 28). We performed the classic fluctuation assay using nalidixic acid to determine the basal rate of mutation in such strains (29). Known point mutations in *gyrA*, the gyrase encoding gene, render *Shigella* resistant to nalidixic acid (Supplementary Table S5) (30). Such mutants can be selected on antibiotic containing plates and the frequency of their appearance calculated for each strain. We designed and cloned a sgRNA (gRNA-gyrA) that would cause G→A transition at position 259, resulting in Asp_{87} → Asn_{87} substitution and resistance to nalidixic acid. pgRNA-gyrA was transformed in *S. flexneri* already carrying the pnCas9-AID plasmid. To distinguish between the rate of mutation caused by the presence or absence of pnCas9-AID, *S. flexneri* strains expressing and not expressing nCas9-AID (transformed with backbone plasmid pSU19 instead) with a non-targeting sgRNA, gRNA-X, were constructed (Fig 4A). Mutagenesis was carried out as in case of *mCherry*, but in addition to spotting on

TSA plates, 10-fold serial dilutions of the cultures were also spotted on TSA plates containing nalidixic acid. The nalidixic acid plates were incubated at 37°C for 24h and the resulting colonies were counted.

The background rate of *gyrA* mutation in absence of nCas9-AID, or a proxy measure of overall random mutation frequency, was in the order of 10^{-8} after 2h of IPTG induction (Fig. 4B). It was not significantly different when IPTG addition was omitted. The expression of gRNA-gyrA alone or a gRNA-X alone (in absence of nCas9-AID) made no difference to the mutation frequency even when IPTG was added. Importantly, the expression of nCas9-AID, in presence of a non-targeting sgRNA (gRNA-X), also did not affect the mutation frequency. Even at later time points, the frequency of mutation for strains expressing nCas9-AID with gRNA-X was similar to those not expressing nCas9-AID at all. It was only in the presence of a *gyrA*-specific sgRNA (gRNA-gyrA), that the nCas9-AID was able to introduce nalidixic acid resistance mutations at a high frequency. About 5 log-fold increase in the frequency of mutation verified the high specificity of our method once again. There was no significant change in the frequency of specific mutations at the three different time points, which confirmed that longer incubation times were not necessary. However, we observed an increase, of almost an order of magnitude, in the background frequency of mutation at later time points (4h and 6h) which made the case against longer incubations even stronger. These results, therefore, suggested that high frequencies of mutation achieved using specific sgRNA in presence of nCas9-AID were the result of specific interactions and over expression of nCas9-AID was not mutagenic by itself. This also indicated that, under the right conditions, the probability of introduction of unwanted off-target mutations by nCas9-AID was in the same order of magnitude as spontaneous mutations.

Mutations on *S. flexneri* chromosome and virulence plasmid can be made with equal ease using nCas9-AID

The large *Shigella* plasmid is essential for the bacterium's virulence. Therefore, its study has been determinant to understand *Shigella* pathogenesis. Although *Shigella* is historically mutated using the low-efficiency allele exchange method, the additional propensity of bacteria to lose the large virulence plasmid frequently presents a significant challenge to mutate and study the effect of plasmid-encoded genes. We decided to use base editing to mutate well-characterized virulence genes on both the chromosome and the virulence plasmid parallelly (Table 1). To generate these mutations, we included different possible combinations of gRNA: multiple gRNA targeting the same site, only one gRNA per gene, and multiple guides targeting different sites (Supplementary Table S6). Mutagenesis was performed as described above, but only for 2h after IPTG induction and the cultures were spotted on TSA plates supplemented with 0.01% Congo red, to facilitate selection of virulent colonies, and 5% sucrose, to cure the pnCas9-AID plasmid. Four red colonies (white in case of *mxiD*$^-$) were selected at random and sequenced to verify the mutation. All colonies sequenced turned out to be the desired mutants, with only one specific mutation introduced via base editing, resulting in 100% efficacy (Fig. 5A). However, we could not obtain any *icsA* mutants with gRNA_3, even after sequencing ten colonies, which, reminiscent of gRNA_m3, reflects that some gRNAs perform poorly. Nonetheless, we were able to generate multiple mutants in *S. flexneri*, on both the chromosome and the virulence plasmid, in a single experiment.

Loss-of-function mutants generated by base editing have the same phenotype as gene deletion mutants

Although the mutations in *S. flexneri* genes were confirmed by sequencing, we wished to verify their phenotype using established experimental assays.

Since *mxiD* mutants fail to invade epithelial cells, we carried out a gentamicin protection assay (GPA) (31, 32). Wild-type and *S. flexneri mxiD*$^-$ were allowed to infect TC7 intestinal epithelial cells *in vitro* and the extracellular bacteria were killed by gentamicin treatment. Bacterial

invasion was quantified by lysing the cells, to release intracellular bacteria that were protected from gentamicin, and spotting log-fold dilutions of the lysates on TSA plates (Fig. 5B). WT bacteria successfully invaded TC7 cells, but we did not observe any invasion by *mxiD*$^{-}$, within our limit of observation (Fig. 5C). Since *mxiD* mutants fail to assemble a functional T3SS, they are unable to secrete the T3SS substrates (31). Indeed, we did not recover any secreted proteins in the supernatant of *mxiD*$^{-}$, induced for secretion by Congo red (Fig 5D) (33).

***i**csA* and *vacJ* have been reported to be necessary for intracellular dissemination of *Shigella* but have no involvement in invasion (34, 35). By performing another GPA, we verified that *S. flexneri icsA*$^{-}$ and *vacJ*$^{-}$ had no invasion defect (Fig. 5E). To assess intracellular dissemination, we used the plaque formation assay where cells are infected with bacteria but the bacteria can only disseminate intracellularly due to the presence of gentamicin in the medium (36, 37). Successful intracellular dissemination results in spread of bacteria and subsequent cell death, which appears as a zone of clearance in the cell monolayer or a plaque (Fig 5F). TC7 cells were infected with the bacteria in 6-well plates and were enclosed in an agarose overlay containing gentamicin and serum. The agarose layer was peeled off after 48-72h of incubation and the cell layers were stained with Giemsa stain. Successful invasion and subsequent dissemination were evident in case of WT *S. flexneri* as numerous plaques of similar diameter were observed (Fig. 5G). However, no such plaques were visible in case of *icsA*$^{-}$ and *vacJ*$^{-}$, even at higher MOIs, indicating an obvious dissemination defect.

Taken together, these results show that the phenotype of the mutants generated by our base editing method was consistent with that reported in the literature.

Base editing results in non-polar mutations

Polar effects, or decreased expression of genes downstream the site of mutation, is a common mutagenesis problem in bacteria where the mRNA is polycistronic (11). They lead to misinterpretation of mutant phenotype as seen in case of *icsB*, an important plasmid-encoded

virulence factor in *Shigella*. Polar effects of *icsB* mutation have been shown to impair intracellular dissemination in *Shigella* (9, 38, 39). We decided to mutate *icsB* by base editing and determine the polarity of the mutant by assessing intracellular dissemination. We generated two *icsB* mutants (*icsB_m5*$^{-}$ and *icsB_m7*$^{-}$*)* with the premature STOP codon introduced at two different sites (Fig. 5A). From GPA, it was clear that *icsB*$^{-}$ had no apparent invasion defect (Fig. 5H) and there was no significant difference in the number and size of the plaques formed by both *icsB* mutants with respect to WT bacteria (Fig. 5G and 5I). This indicates that the mutants generated by base editing do not have any polar effect on the downstream genes when *in vitro* infection assays are used as readout.

Discussion

Functional characterization of genes is a primary step in understanding the physiology and pathogenesis of bacteria/microbes, which is made possible by powerful tools such as high-throughput reverse genetics (40). However, the success of high-throughput reverse genetics is entirely dependent on the ability to generate loss-of-function libraries with an equally high throughput (41). A great body of information about *Shigella* biology has so far been deciphered with the help of classic allele replacement mutagenesis, which is rather slow, cumbersome, and inefficient (5, 8). Generating transposon libraries, on the other hand, is fast and easy but it still requires verification of phenotypes by generating individual loss-of-function mutants (42). Within this context, a number of chromosomal genes in *Shigella* are transcribed with no function assigned to them and no high-throughput effort, to the best of our knowledge, has been undertaken to characterize these genes (7). We sought to bridge this gap by adopting a recently improved CRISPR-Cas base editing method, shown to work in *C. glutamicum*, for generating targeted loss-of-function mutants with higher throughput, in *Shigella* (21).

We designed a two-plasmid expression system where the base editing effector protein, nCas9-AID, is expressed from a low-copy curable plasmid which is compatible with a high-copy

plasmid pgRNA, expressing target-specific sgRNA. High efficiency cloning of short 25 nt spacer oligos (Fig. 1) already offers an improvement over long DNA oligos used in prevalent mutagenesis (8). The use of type II restriction enzyme BsaI not only ensures directional cloning of spacer oligos but also means that the parent plasmid needs to be digested only once for all cloning reactions. Substituting electroporation with chemical transformation in 96-well plates with low-density competent cells further decreases the cost and increases the overall throughput (43), ensuring pure colonies of several desired mutants prepared in parallel, in just 5-6 days from the receipt of spacer oligos (Fig. 6).

High frequency of mutation can also achieved by combining the expression of catalytically inactive Cas9, dCas9-AID, and a uracil DNA glycosylase inhibitor (20, 44), but we avoided using this combination as the uracil DNA glycosylase inhibitor can potentially induce hypermutation, owing to its ability to impair DNA repair (45). We were able to achieve high frequency of mutagenesis (up to 100%, Fig. 2C and 3C) using nCas9D10A-AID, consistent with previous reports (21). We observed only slight decreases in the viability of nCas9-AID expressing strains (Supplementary Fig. S2 and S3). Coupling nCas9-AID expression with a sgRNA, targeting or non-targeting, might have decreased the availability of free nCas9-AID, preventing it from binding to non-specific sites on genomic DNA and introducing unwanted and potentially toxic mutations (46). Indeed, we observed that the frequency of non-specific mutations in presence of nCas9-AID was as low as in its absence (Fig. 4). This also indicates at low off-target specificity, while maintaining on-target specificity (5-log folds higher than background). Nevertheless, as long as the transformation efficiency is high, small decreases in viability might not be of great concern.

Since the virulence plasmid is not necessary for bacterial survival, improved yield of mutants by counter-selecting the non-modified cells using CRISPR-Cas9 and recombineering methods together, is not applicable to mutation of plasmid-borne genes (15, 16). However, we were able

to mutate plasmid and chromosomal genes in *S. flexneri* in a single parallel experiment (Fig. 6) and were able to verify that the phenotype of the mutants conformed to that reported in the literature, using classic assays (31, 34, 35, 39). Using *icsB*$^{-}$, we were also able to verify the non-polar nature of the base edited mutants. This is accordance with a previous report, suggesting that mutations that don't alter the reading frame result in non-polar mutations in *Shigella* (47), and our recent results that mutations generated by CRISPR-Cas mutagenesis do not show polar effects at the phenotype level (48).

Although 2h of IPTG induction had a meagre impact on the frequency of specific mutation (from 85% to 92%, Fig. 3C) and non-specific mutations (in the order of 10^{-8}, Fig. 4B), we still decided against IPTG exclusion. Given the variations in efficacies of sgRNA, small increases in frequency can help decrease the number of colonies needed to be sequenced to verify the mutation (Fig. 6). However, depending on the availability of the resources at hand, IPTG addition can be avoided.

In addition to generating parallel single gene mutations, a spacer oligo library can be cloned into pgRNA in a single reaction resulting in a pgRNA plasmid library which can then be transformed into bacteria expressing nCas9-AID, generating an equally large, pooled library of loss-of-function mutants. Since the pgRNA (ColE1-derived origin) propagates stably (49), cloned spacer oligos can function as barcodes, aiding in high-throughput screening of gene function (50) and base editing can be subsequently used for targeted mutagenesis of selected candidates.

Despite many apparent advantages, the most obvious limitation of the base editing method is the availability of mutable sites in a particular gene. Certain genes, for example *mxiD* in this study, have only one mutable site. As seen from our results, certain sgRNAs are more effective than others and not having any alternative, therefore, can be limiting. However, with the discovery of Cas9 variants with varying PAM specificity, in the future there is a possibility of

combining them with deaminases to expand the arsenal of effector proteins to overcome this limitation (51).

In conclusion, we have successfully adapted a programmable CRISPR-Cas9 base editing method to generate loss-of-function mutants in *S. flexneri* and *E. coli*. With the aid of CRISPR-CBEI program, high-efficiency cloning, and parallel rapid transformation, the time, cost, and effort required for generating mutants in *S. flexneri* is drastically reduced. The loss-of-function phenotype, upon base editing, is indifferentiable from gene deletion with the no apparent polar effects and high specificity, without affecting the basal rate of spontaneous mutagenesis, is an additional benefit. This method can allow functional characterization of unknown *S. flexneri* genes by high-throughput reverse genetics and could be extended to other members of Enterobacteriaceae owing to plasmid and promoter compatibility.

Methods

Bacterial strains, cells, and culture conditions

Bacterial strains and cells used in this study have been summarized in the Supplementary Table S1. All *E. coli* strains were grown in LB, *S. flexneri* in TSB, and TC7 cells were gown in DMEM. Plasmids used in this study and antibiotic concentrations required to maintain them are mentioned in Supplementary Table S2 and Supplementary Table S3, respectively. *S. flexneri* colonies were selected on agar plates supplemented with 0.01% Congo red.

Construction of fluorescent strains

Fluorescent strains of *E. coli* and *S. flexneri*, constitutively expressing *mCherry* from the chromosome, were constructed using Tn7-mediated transposition (52). *mCherry* was cloned in pGRG25 and transformed into the host bacteria for transposition resulting in a stable insertion downstream the *glmS* gene. The detailed methodology is mentioned in Supplementary Methods S1.

Plasmid construction

All the oligos used for recombinant procedures are mentioned in Supplementary Table S4. pgRNA_AT was generated by combining parts from pgRNA-ccdB and a pBluescript SK variant that lacks the BsaI site in the β-lactamase gene (21). pnCas9-AID was generated by combining parts of pSU19 and pdCas9-AID by *in-vivo* assembly (IVA) cloning (21, 53). Plasmid construction is explained in detail in Supplementary Methods S1.

Guide RNA design and cloning

The guide RNA for mutagenesis were designed using the program CRISPR_CBEI (25) (Supplementary Table S6). Details of cloning and gRNA design are mentioned in Supplementary Methods S1.

Mutagenesis

pnCas9-AID and respective pgRNA were co-transformed in ultra-competent *E. coli* using the heat shock method (54). The co-transformants were selected on LA plates supplemented with chloramphenicol, carbenicillin, and glucose. Since co-transformation was not very efficient in *S. flexneri*, pnCas9-AID and various sgRNA plasmids were sequentially transformed.

Bacteria carrying pnCas9-AID and respective sgRNA plasmids were grown overnight at 37°C in broth supplemented with chloramphenicol, carbenicillin, and glucose. The overnight culture was diluted 100 times in fresh medium and the bacteria were grown till OD_{600} = 0.2. 1 mM IPTG was added to induce the expression of nCas9-AID and the cultures were incubated at 37°C for further 2h (longer incubation times of were used in case of *mCherry* and *gyrA* mutagenesis). 10-fold serial dilutions of the cultures were made in sterile PBS and 10 µl of each dilution was spotted on agar plates containing no antibiotics. For *gyrA* mutagenesis, the agar plates were supplemented with nalidixic acid (30 µg/ml) and in addition to 10-fold dilutions, a 10-fold concentrated culture was also spotted. The agar plates were incubated overnight at 37°C and the colonies were enumerated. In case of *mCherry* mutagenesis, the plates were imaged in a Typhoon scanner (GE Healthcare). To verify mutagenesis, four random

colonies were selected in each case and the site of mutagenesis was sequenced by colony PCR.

Bacterial invasion assay and plaque formation assay

The ability of bacteria to invade epithelial cells was determined by the GPA as described earlier (32). All mutant *Shigella* strains were transformed with pAfaE, a derivative of pIL22 (55) expressing AFA I adhesin. Plaque formation assay was performed as described in (37) with a small modification to the MOI and incubation times. The TC7 cells were infected at a MOI of 1:500 (bacteria:cells), and additionally at 1:150 for *vacJ*$^{-}$ and *icsA*$^{-}$. The infected cells were incubated for 48-72h before being fixed and stained.

Secretion assay

The secretion assay was performed as described earlier with modifications (56). Overnight cultures of bacteria were diluted 200 times and were grown till OD_{600} = 0.5. Cells were collected from 4 ml culture, by centrifugation, and washed twice with PBS. The pellet was finally resuspended in 1 ml prewarmed PBS containing 0.01% Congo red and incubated for an hour at 37°C. Bacteria were separated by centrifugation and the proteins in the supernatants were precipitated at 4°C for 10 minutes using 20% trichloroacetic acid. 1 mg/ml solution of BSA was used as a control for precipitation. The precipitated proteins were washed twice with ice-cold acetone and resuspended in Laemmli buffer before being resolved on 12% polyacrylamide gel. The gel was soaked overnight in 50% ethanol to remove Congo red and subsequently rehydrated gradually. The gel was stained with SimplyBlue SafeStain (Thermo Fisher Scientific) to visualize the proteins.

96-well mutagenesis

50 µl chemically competent *S. flexneri* carrying pnCas9-AID were pipetted in wells of a pre-chilled 96-well plate. sgRNA plasmids were added to individual wells and transformed by the heat-shock method. 150 µl of recovery medium (SOC) was added to each well and the plate was incubated at 37°C for 1h. 100 µl of cell suspension and 1/10th dilution from each well were

spread on selective plates. The plates were incubated overnight at 37°C and single colonies were picked up for mutagenesis. In a 96-well deep-well plate (Eppendorf), single colonies were inoculated in 500 µl of TSB with antibiotics and grown overnight at 37°C. 10 µl of overnight culture was added to 500 µl of fresh medium in a new 96-well deep well plate, which was incubated till OD_{600} = 0.2. 1 mM IPTG was added to the wells and the plate was incubated for another 2h before serial dilutions (10 µl of 10^{-5} and 10^{-6} dilution, specifically) of the cultures were spotted on Congo red plates. The plates were incubated at 37°C overnight and single colonies were checked for desired mutation by colony PCR and DNA sequencing.

Acknowledgement

This study was supported by financial grant from the Swedish Research Council (Vetenskapsrådet), grant number #2016-06598, to the Puhar lab. AP was funded by institutional support from Umeå University, Knut and Alice Wallenberg Foundation grant KAW 2015.0225, and Carl Kempe Foundation grant JCK-2031.3. AS was funded by MIMS Excellence by Choice Postdoctoral Programme stipend under the patronage of Emmanuelle Charpentier grant SMK-1532.2 and Svenska Sällskapet för Medicinsk Forskning (SSMF) Postdoctoral grant PD20-0022. DAC acknowledges support from the Carl Tryggers Stiftelse för Vetenskaplig Forskning (project CTS 18-65) and the Kempestiftelserna (project SMK 1860). DAC was supported in part also by funds from the Novo Nordisk Foundation (Grant no. NNF17OC0026486) awarded to Dr. Emmanuelle Charpentier at MIMS, The Laboratory for Molecular Infection Medicine Sweden).

The authors are thankful to Yu Wang and Jibin Sun, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, China for kindly gifting the pnCas9-AID-YU and pgRNA-ccdB plasmids; Javier Pizarro-Cerdá, Institut Pasteur, Paris, France for providing pSU2.1rp-mCherry plasmid; and Justin L. Sonnenburg, Stanford University School of Medicine, Stanford, USA for providing sfGFP carrying *Bacteroides thetaiotaomicron* strain.

Author contributions as per CASRAI are as follows. Conceptualization: AS, AP, DAC. Formal analysis: AS, ROA. Funding acquisition: AS, AP, DAC. Investigation: AS, ROA. Methodology: AS, ROA, AP, DAC. Project administration: AP, DAC. Resources: AP, DAC. Supervision: AP, DAC. Visualization: AS, ROA. Writing – original draft: AS. Writing – reviewing and editing: AP, DAC, AS.

References

1. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, Wu Y, Sow SO, Sur D, Breiman RF, Faruque AS, Zaidi AK, Saha D, Alonso PL, Tamboura B, Sanogo D, Onwuchekwa U, Manna B, Ramamurthy T, Kanungo S, Ochieng JB, Omore R, Oundo JO, Hossain A, Das SK, Ahmed S, Qureshi S, Quadri F, Adegbola RA, Antonio M, Hossain MJ, Akinsola A, Mandomando I, Nhampossa T, Acácio S, Biswas K, O'Reilly CE, Mintz ED, Berkeley LY, Muhsen K, Sommerfelt H, Robins-Browne RM, Levine MM. 2013. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Lancet 382:209–222.

2. Anderson M, Sansonetti PJ, Marteyn BS. 2016. Shigella Diversity and Changing Landscape: Insights for the Twenty-First Century. Front Cell Infect Microbiol 6:45.

3. Khalil IA, Troeger C, Blacker BF, Rao PC, Brown A, Atherly DE, Brewer TG, Engmann CM, Houpt ER, Kang G, Kotloff KL, Levine MM, Luby SP, MacLennan CA, Pan WK, Pavlinac PB, Platts-Mills JA, Qadri F, Riddle MS, Ryan ET, Shoultz DA, Steele AD, Walson JL, Sanders JW, Mokdad AH, Murray CJL, Hay SI, Reiner RC. 2018. Morbidity and mortality due to shigella and enterotoxigenic Escherichia coli diarrhoea: the Global Burden of Disease Study 1990–2016. The Lancet Infectious Diseases 18:1229–1240.

4. Kotloff KL, Riddle MS, Platts-Mills JA, Pavlinac P, Zaidi AKM. 2018. Shigellosis. The

Lancet 391:801–812.

5. Mattock E, Blocker AJ. 2017. How Do the Virulence Factors of Shigella Work Together to Cause Disease? Front Cell Infect Microbiol 7:64.

6. Schnupf P, Sansonetti PJ. 2019. Shigella Pathogenesis: New Insights through Advanced Methodologies. Microbiology Spectrum 7:7.2.28.

7. Cervantes-Rivera R, Tronnet S, Puhar A. 2020. Complete genome sequence and annotation of the laboratory reference strain Shigella flexneri serotype 5a M90T and genome-wide transcriptional start site determination. BMC Genomics 21:285.

8. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. PNAS 97:6640–6645.

9. Campbell-Valois F-X, Sachse M, Sansonetti PJ, Parsot C. 2015. Escape of Actively Secreting Shigella flexneri from ATG8/LC3-Positive Vacuoles Formed during Cell-To-Cell Spread Is Facilitated by IcsB and VirA. mBio 6:e02567-02514.

10. Li Z, Liu W, Fu J, Cheng S, Xu Y, Wang Z, Liu X, Shi X, Liu Y, Qi X, Liu X, Ding J, Shao F. 2021. Shigella evades pyroptosis by arginine ADP-riboxanation of caspase-11. Nature 599:290–295.

11. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2:2006.0008.

12. Charpentier E, Doudna JA. 2013. Rewriting a genome. Nature 495:50–51.

13. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013.

RNA-guided human genome engineering via Cas9. Science 339:823–826.

14. Bikard D, Euler CW, Jiang W, Nussenzweig PM, Goldberg GW, Duportet X, Fischetti VA, Marraffini LA. 2014. Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. Nat Biotechnol 32:1146–1150.

15. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. 3. Nature Biotechnology 31:233–239.

16. Jiang Y, Chen B, Duan C, Sun B, Yang J, Yang S. 2015. Multigene Editing in the Escherichia coli Genome via the CRISPR-Cas9 System. Appl Environ Microbiol 81:2506–2514.

17. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. 7603. Nature 533:420–424.

18. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR. 2017. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature 551:464–471.

19. Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, Tabata M, Mochizuki M, Miyabe A, Araki M, Hara KY, Shimatani Z, Kondo A. 2016. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science 353:aaf8729.

20. Banno S, Nishida K, Arazoe T, Mitsunobu H, Kondo A. 2018. Deaminase-mediated multiplex genome editing in Escherichia coli. Nat Microbiol 3:423–429.

21. Wang Y, Liu Y, Liu J, Guo Y, Fan L, Ni X, Zheng X, Wang M, Zheng P, Sun J, Ma Y. 2018. MACBETH: Multiplex automated Corynebacterium glutamicum base editing method.

Metabolic Engineering 47:200–210.

22. Bartolomé B, Jubete Y, Martínez E, de la Cruz F. 1991. Construction and properties of a family of pACYC184-derived cloning vectors compatible with pBR322 and its derivatives. Gene 102:75–78.

23. Gay P, Le Coq D, Steinmetz M, Berkelman T, Kado CI. 1985. Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. J Bacteriol 164:918–921.

24. Tobe T, Sasakawa C, Okada N, Honma Y, Yoshikawa M. 1992. vacB, a novel chromosomal gene required for expression of virulence genes on the large plasmid of Shigella flexneri. Journal of Bacteriology 174:6359–6367.

25. Yu H, Wu Z, Chen X, Ji Q, Tao S. 2020. CRISPR-CBEI: a Designing and Analyzing Tool Kit for Cytosine Base Editor-Mediated Gene Inactivation. mSystems 5:e00350-20.

26. Bernard P, Gabant P, Bahassi EM, Couturier M. 1994. Positive-selection vectors using the F plasmid ccdB killer gene. Gene 148:71–74.

27. Dickerson SK, Market E, Besmer E, Papavasiliou FN. 2003. AID Mediates Hypermutation by Deaminating Single Stranded DNA. Journal of Experimental Medicine 197:1291–1296.

28. Kuscu C, Adli M. 2016. CRISPR-Cas9-AID base editor is a powerful gain-of-function screening tool. Nat Methods 13:983–984.

29. Luria SE, Delbrück M. 1943. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. Genetics 28:491–511.

30. Ghosh A, N. S, Saha S. 2020. Survey of drug resistance associated gene mutations in Mycobacterium tuberculosis, ESKAPE and other bacterial species. Sci Rep 10:8957.

31. Allaoui A, Sansonetti PJ, Parsot C. 1993. MxiD, an outer membrane protein necessary for the secretion of the Shigella flexneri lpa invasins. Mol Microbiol 7:59–68.

32. Sharma A, Puhar A. 2019. Gentamicin Protection Assay to Determine the Number of Intracellular Bacteria during Infection of Human TC7 Intestinal Epithelial Cells by Shigella flexneri. Bio-protocol 9:e3292–e3292.

33. Bahrani FK, Sansonetti PJ, Parsot C. 1997. Secretion of Ipa proteins by Shigella flexneri: inducer molecules and kinetics of activation. Infect Immun 65:4005–4010.

34. Bernardini ML, Mounier J, d'Hauteville H, Coquis-Rondon M, Sansonetti PJ. 1989. Identification of icsA, a plasmid locus of Shigella flexneri that governs bacterial intra- and intercellular spread through interaction with F-actin. Proc Natl Acad Sci U S A 86:3867–3871.

35. Carpenter CD, Cooley BJ, Needham BD, Fisher CR, Trent MS, Gordon V, Payne SM. 2014. The Vps/VacJ ABC Transporter Is Required for Intercellular Spread of Shigella flexneri. Infection and Immunity 82:660–669.

36. Oaks EV, Wingfield ME, Formal SB. 1985. Plaque formation by virulent Shigella flexneri. Infect Immun 48:124–129.

37. Sharma A, Puhar A. 2019. Plaque Assay to Determine Invasion and Intercellular Dissemination of Shigella flexneri in TC7 Human Intestinal Epithelial Cells. Bio-protocol 9:e3293–e3293.

38. Allaoui A, Mounier J, Prévost M-C, Sansonetti PJ, Parsot C. 1992. icsB: a Shigella flexneri

virulence gene necessary for the lysis of protrusions during intercellular spread. Molecular Microbiology 6:1605–1616.

39. Rathman M, Jouirhi N, Allaoui A, Sansonetti P, Parsot C, Tran Van Nhieu G. 2000. The development of a FACS-based strategy for the isolation of Shigella flexneri mutants that are deficient in intercellular spread. Molecular Microbiology 35:974–990.

40. Fels U, Gevaert K, Van Damme P. 2020. Bacterial Genetic Engineering by Means of Recombineering for Reverse Genetics. Frontiers in Microbiology 11:2161.

41. Brochado AR, Typas A. 2013. High-throughput approaches to understanding gene function and mapping network architecture in bacteria. Current Opinion in Microbiology 16:199–206.

42. Freed NE, Bumann D, Silander OK. 2016. Combining Shigella Tn-seq data with gold-standard E. coli gene deletion data suggests rare transitions between essential and non-essential gene functionality. BMC Microbiology 16:203.

43. Wu N, Matand K, Kebede B, Acquaah G, Williams S. 2010. Enhancing DNA electrotransformation efficiency in Escherichia coli DH10B electrocompetent cells. Electronic Journal of Biotechnology 13:21–22.

44. Wang L, Xue W, Yan L, Li X, Wei J, Chen M, Wu J, Yang B, Yang L, Chen J. 2017. Enhanced base editing by co-expression of free uracil DNA glycosylase inhibitor. Cell Res 27:1289–1292.

45. Bridges BA. 2001. Hypermutation in bacteria and other cellular systems. Phil Trans R Soc Lond B 356:29–39.

46. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. 2014. DNA interrogation by

the CRISPR RNA-guided endonuclease Cas9. Nature 507:62–67.

47. Ménard R, Sansonetti PJ, Parsot C. 1993. Nonpolar mutagenesis of the ipa genes defines IpaB, IpaC, and IpaD as effectors of Shigella flexneri entry into epithelial cells. J Bacteriol 175:5899–5906.

48. Graffeuil A, Uhlin BE, Cisneros DA. 2020. "Polar mutagenesis of bacterial transcriptional units using Cas12a." bioRxiv https://doi.org/10.1101/2020.02.21.959866.

49. Hove-Jensen B. 2008. Two-step method for curing Escherichia coli of ColE1-derived plasmids. Journal of Microbiological Methods 72:208–213.

50. Wang T, Wei JJ, Sabatini DM, Lander ES. 2014. Genetic screens in human cells using the CRISPR-Cas9 system. Science 343:80–84.

51. Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, Gonzales APW, Li Z, Peterson RT, Yeh J-RJ, Aryee MJ, Joung JK. 2015. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature 523:481–485.

52. McKenzie GJ, Craig NL. 2006. Fast, easy and efficient: site-specific insertion of transgenes into Enterobacterial chromosomes using Tn7 without need for selection of the insertion event. BMC Microbiol 6:39.

53. García-Nafría J, Watson JF, Greger IH. 2016. IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly. Sci Rep 6:27459.

54. Inoue H, Nojima H, Okayama H. 1990. High efficiency transformation of Escherichia coli with plasmids. Gene 96:23–28.

55. Clerc P, Sansonetti PJ. 1987. Entry of Shigella flexneri into HeLa cells: evidence for

directed phagocytosis involving actin polymerization and myosin accumulation. Infect Immun 55:2681–2688.

56. 2008. Cytoplasmic targeting of IpaC to the bacterial pole directs polar type III secretion in Shigella. The EMBO Journal 27:447–457.

Figures and figure legends

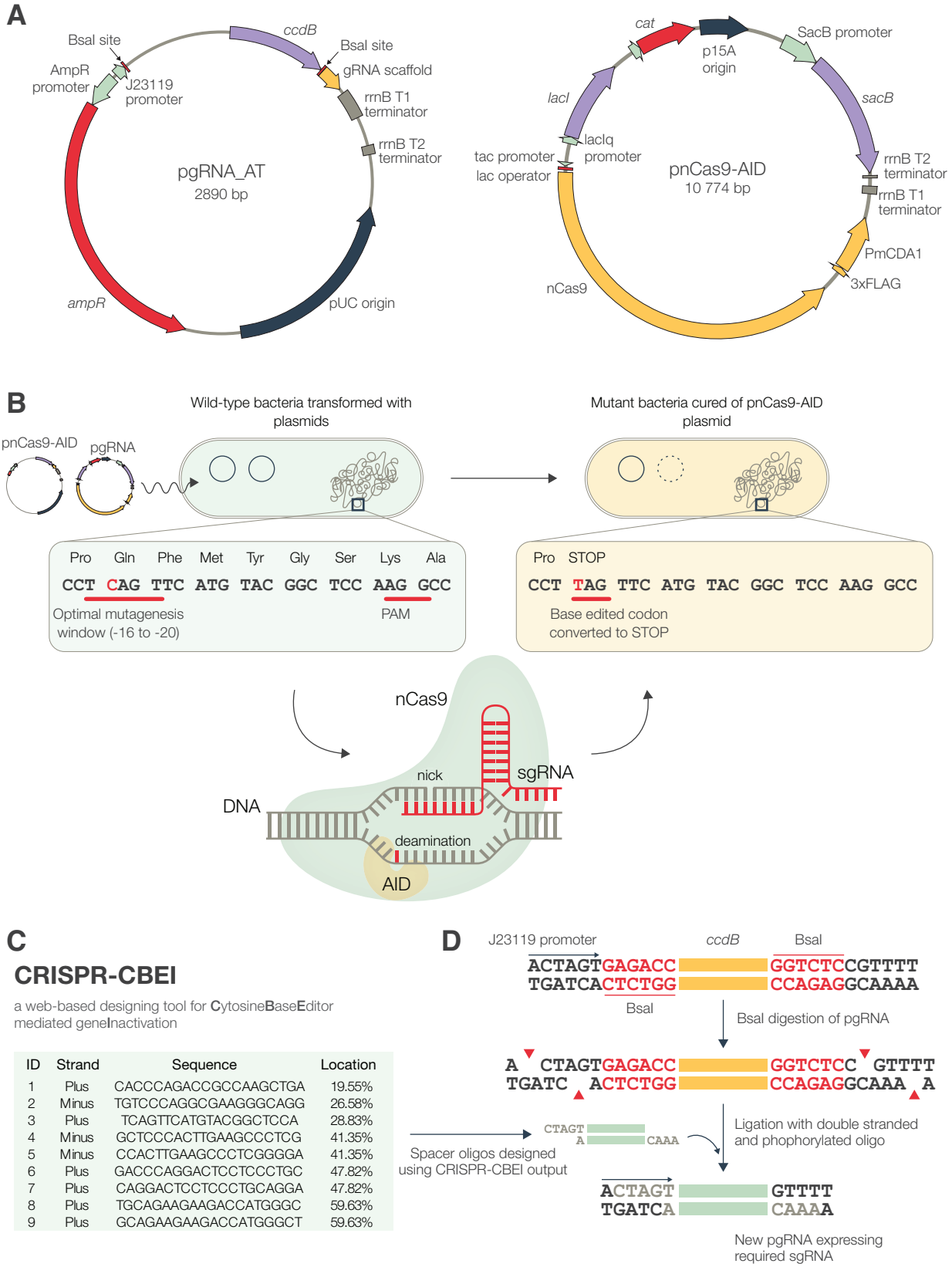

Figure 1: Overview of the CRISPR-Cas guided base editing method.

(A) Maps of plasmids pgRNA and pnCas9-AID used for mutagenesis in *E. coli* and *S. flexneri*. (B) Generation of mutants using *mCherry* as example. After co-transformation with pnCas9-AID and pgRNA, expression of Cas9-AID and a specific gRNA introduces a premature stop codon in the target sequence. The PAM and the editable window are underlined with the target C nucleotide shown in red. (C) Output from CRISPR-CBEI for *mCherry* and (D) the general scheme for high-efficiency cloning of the spacers determined by CRISPR-CBEI.

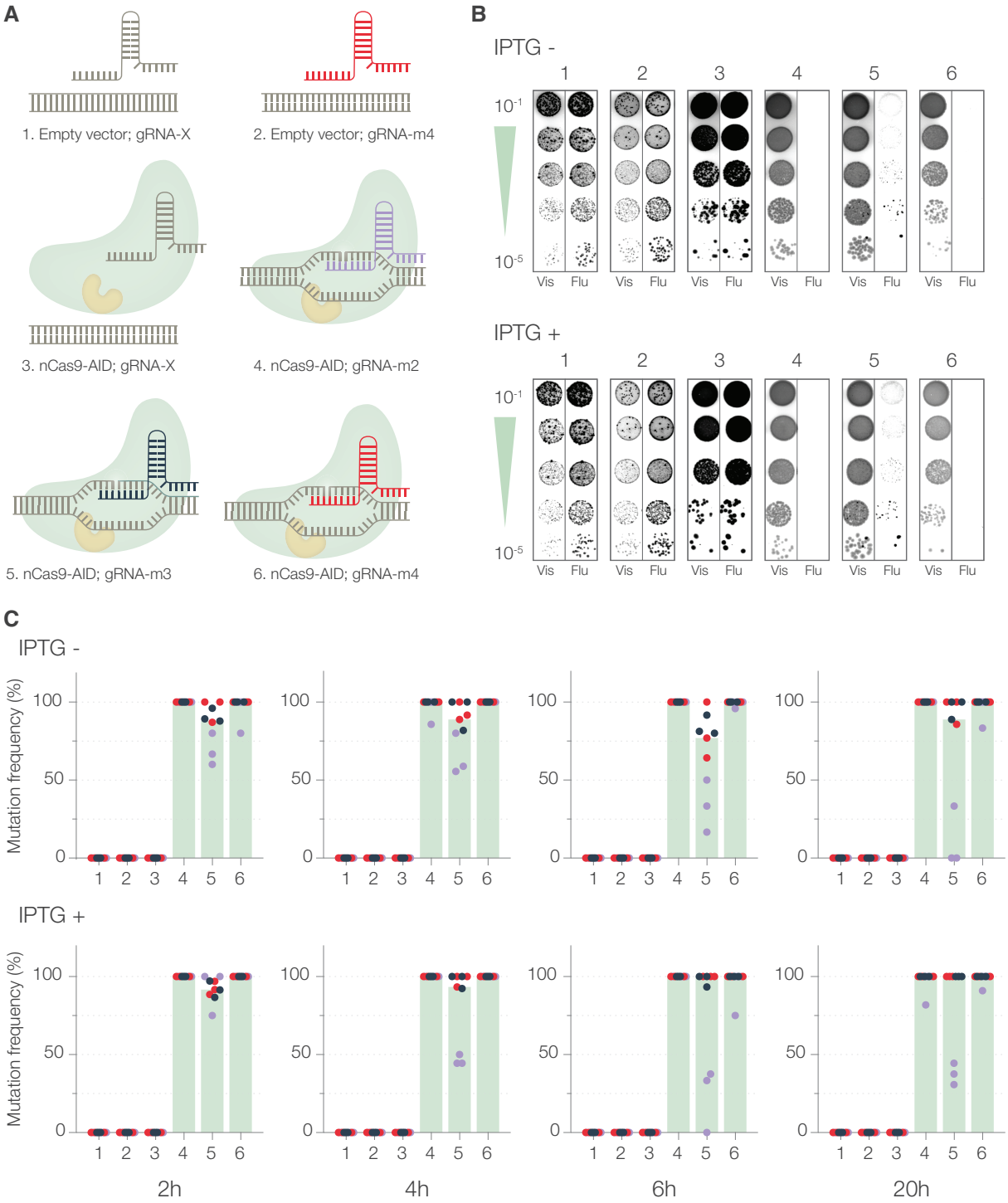

Figure 2: Mutagenesis of chromosomally encoded *mCherry* in *E. coli*.

(A) Schematic representation of the six different conditions used in the experiment. (B) Phenotypic determination of loss of fluorescence in *E. coli*. Representative visible (Vis) and fluorescence (Flu) images of colonies from dilution series of bacteria plated after 2h of IPTG induction. (C) Mutation frequency, determined as the ratio of the of number of non-fluorescent

(mutant) colonies to the total number of colonies, after 2, 4, 6, and 20h of mutagenesis with or without IPTG induction. Each dot represents a technical replicate and is color coded to represent a biological replicate. The bars represent the median of three independent experiments.

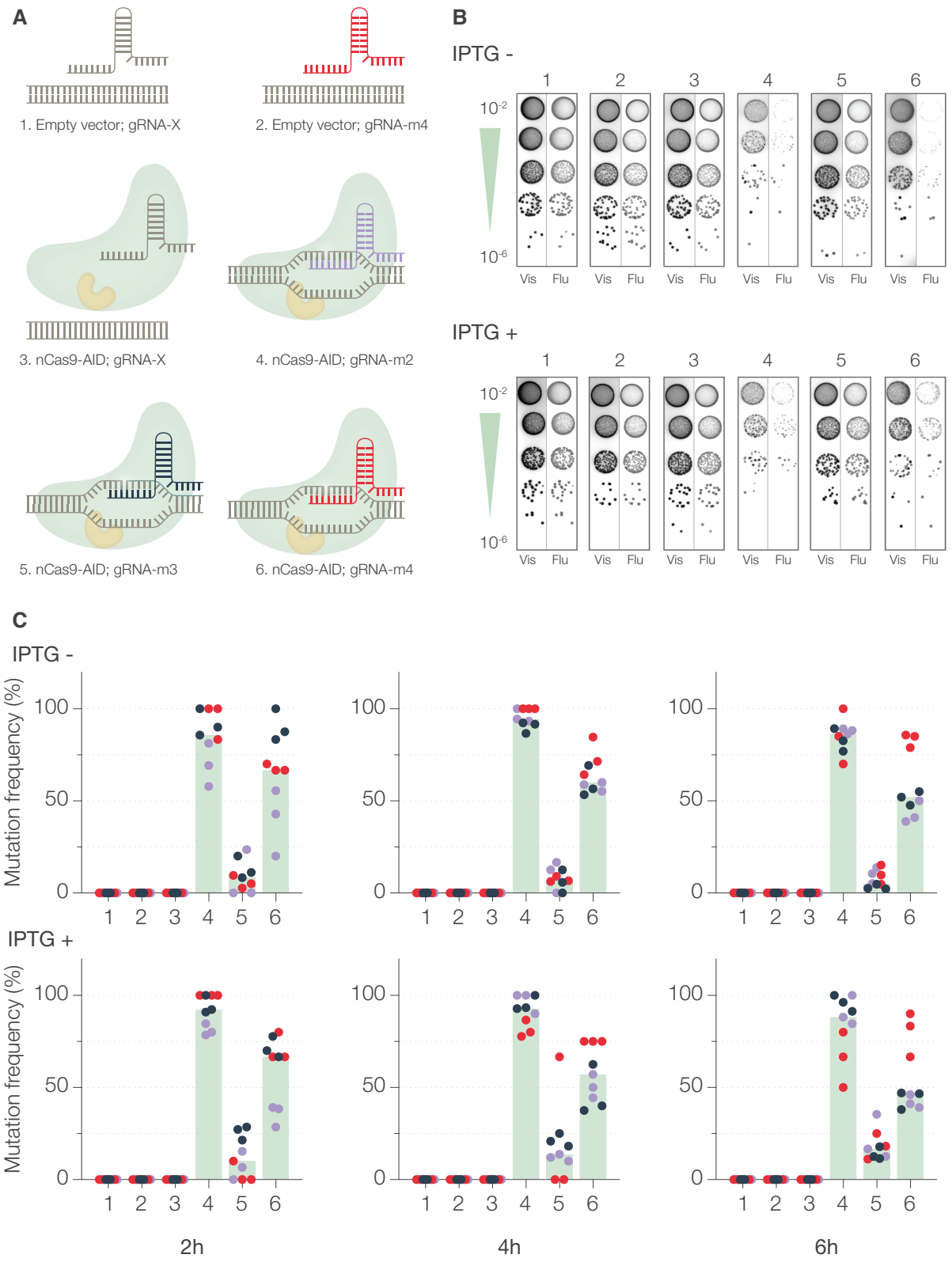

Figure 3: Mutagenesis of chromosomally encoded *mCherry* in *S. flexneri*

(A) Schematic representation of the six different conditions used in the experiment. (B) Phenotypic determination of loss of fluorescence in *S. flexneri*. Representative visible (Vis)

and fluorescence (Flu) images of colonies from dilution series of bacteria plated after 2h of IPTG induction. (C) Mutation frequency, determined as the ratio of the of number of non-fluorescent (mutant) colonies to the total number of colonies, after 2, 4, and 6h of mutagenesis with or without IPTG induction. Each dot represents a technical replicate and is color coded to represent a biological replicate. The bars represent the median of three independent experiments.

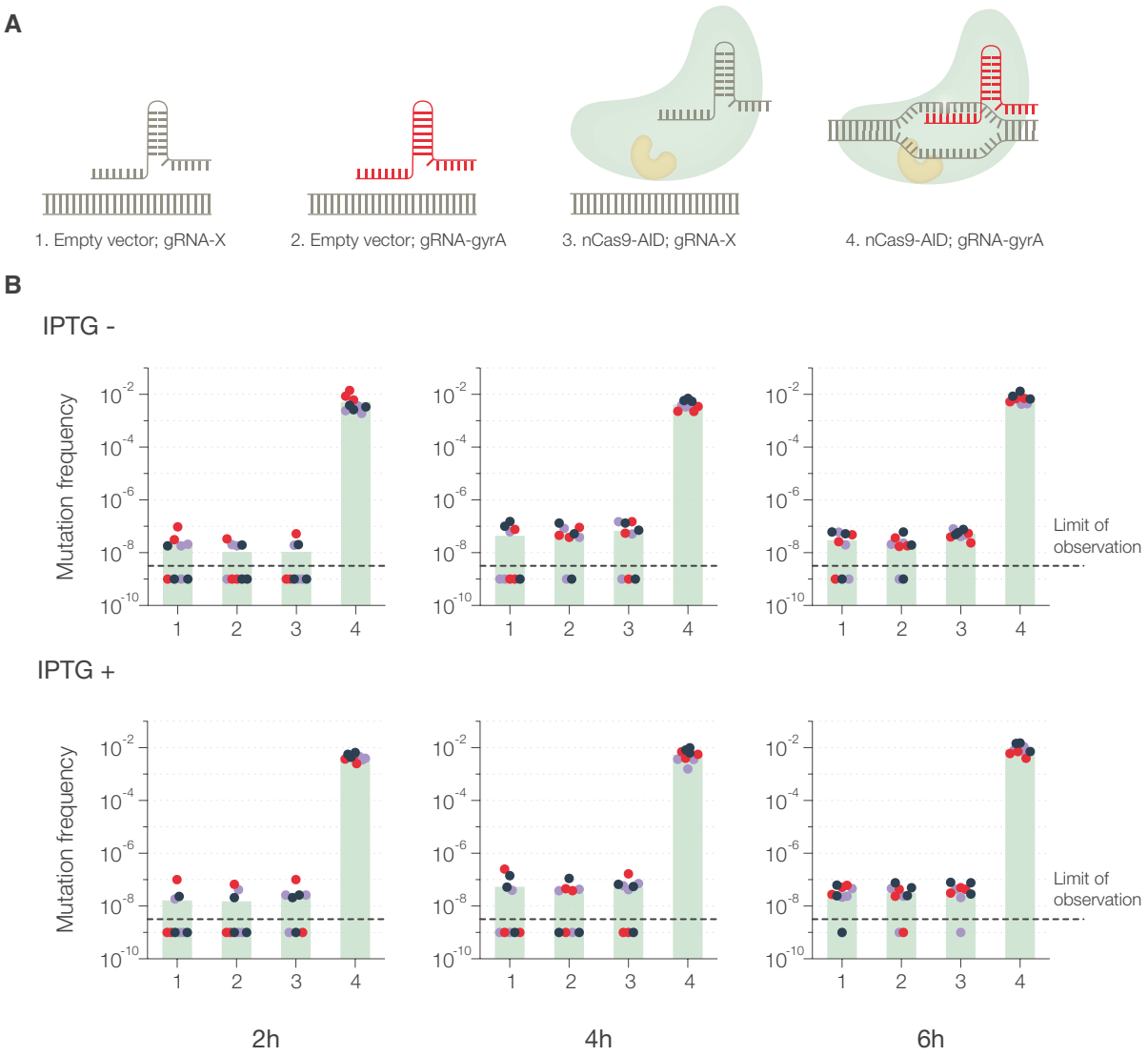

Figure 4: Nalidixic acid resistance assay to compare the frequency of guided and random mutations.

(A) Schematic representation of the four different conditions used in the experiment. (B)

Mutation frequency, determined as the ratio of the number of nalidixic acid resistant (mutant) colonies to the total number of colonies, after 2, 4, and 6h of mutagenesis with and without IPTG induction. Each dot represents a technical replicate and is color coded to represent a biological replicate. In some technical replicates, the mutation frequency was below the technically possible limit of observation ($1x10^{-9}$) and hence not reported. The bars represent the median of three independent experiments.

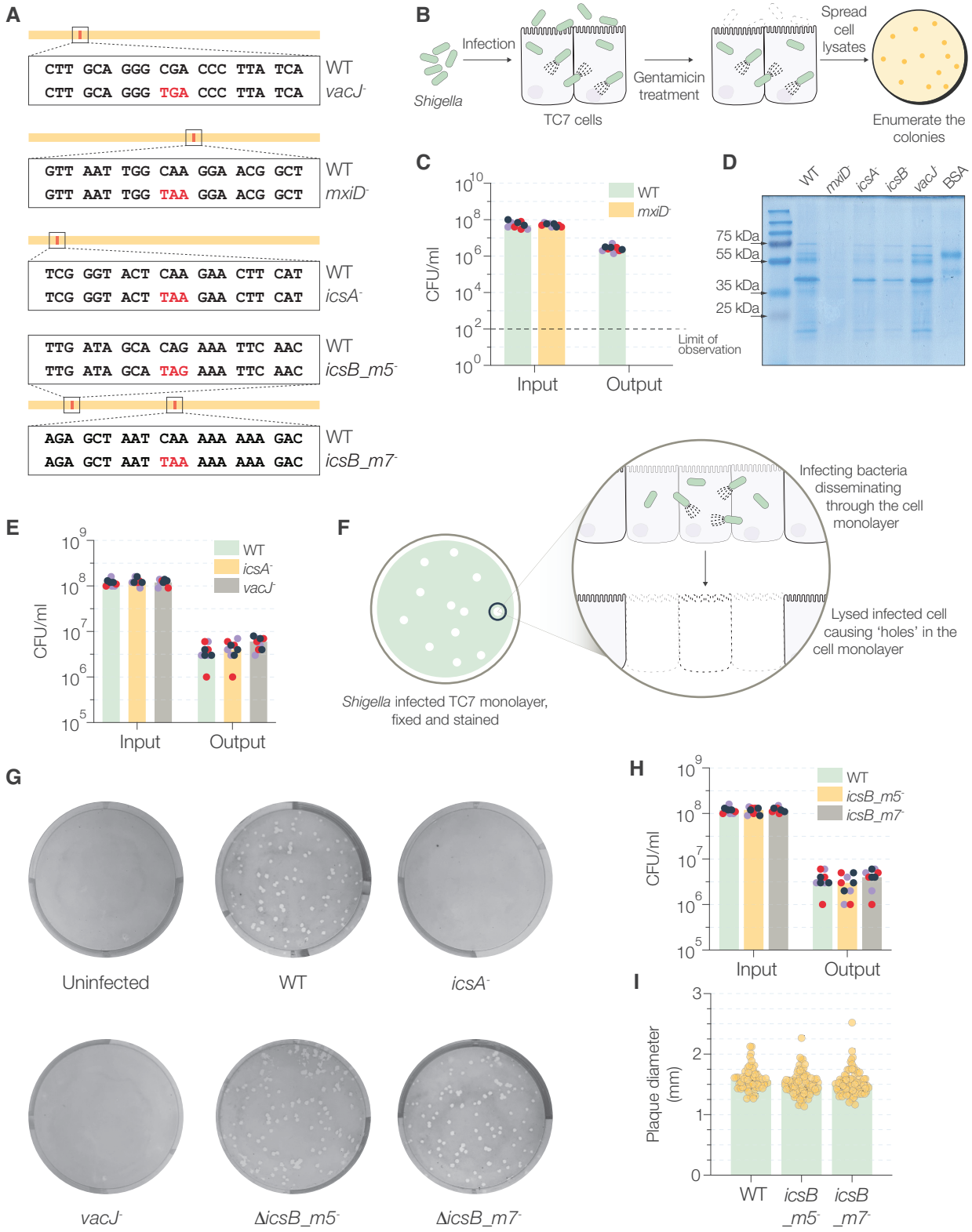

Figure 5: CRISPR/Cas-guided mutagenesis of characterized *S. flexneri* genes and assessment of their phenotype.

(A) Representation of the STOP codons (highlighted in red) introduced by base-editing. The total length of the gene is indicated by a line and the red mark indicates the location of the

mutable target site. (B) Schematic representation of the gentamicin protection assay used to assess the invasiveness of *Shigella*. (C) Invasiveness of WT and *mxiD*$^{-}$ *S. flexneri*, determined by the number of bacterial CFU used to infect TC7 cells (input) and the number of bacterial CFU that were successful at invading them (output). In case of *S. flexneri mxiD*$^{-}$, the CFU count was not above the technical limit of observation (10^2 CFU/ml). (D) SD-PAGE gel showing the total secreted proteins precipitated from the supernatants of indicated bacterial cultures after induction of secretion by Congo red. BSA was used as a control for protein precipitation. (E) Invasiveness of WT, *icsA*$^{-}$, and *vacJ*$^{-}$ *S. flexneri* determined by the gentamicin protection assay. (F) Schematic representation of the plaque formation assay. (G) Intracellular dissemination of WT and mutant *S. flexneri* determined by plaque formation assay. The images depict the stained monolayer of TC7 cells in a 6-well plate with the plaques seen as zones of clearance. (H) Invasiveness of WT and the two strains of *icsB*$^{-}$ *S. flexneri* determined by gentamicin protection assay. (I) Measurement of the diameters of the plaques from one representative plaque formation assay. In all the plots, each dot represents a technical replicate and is color coded to represent a biological replicate. The bars represent the mean of three independent experiments (except in (I)).

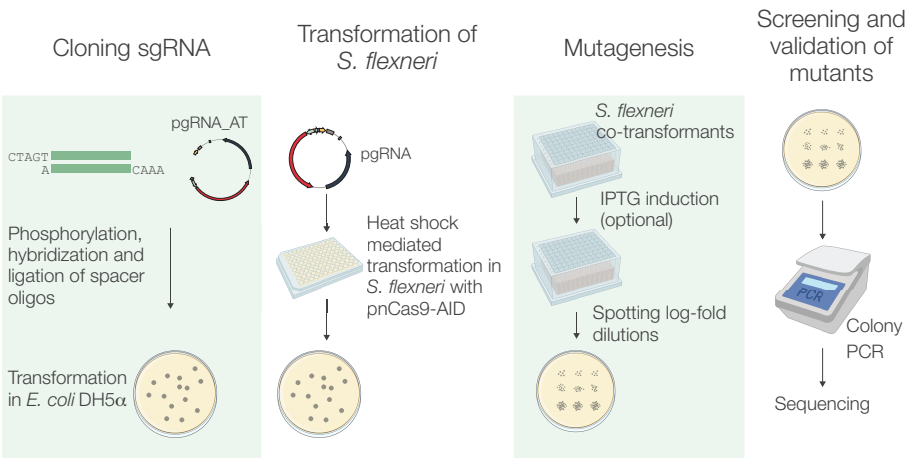

Figure 6: Overview of the workflow to generate multiple parallel mutations in *Shigella* in high throughput.

Cloning the spacer oligos in pgRNA and transformation in E. coli DH5α is

done on Day 1. The transformants obtained can be inoculated to extract plasmids later on Day 2 or overnight cultures can be made for plasmid extraction on Day 3. pgRNA are transformed on Day 3 and *S. flexneri* colonies carrying both plasmids are selected on Day 4 after overnight incubation. Inoculations can be made to carry out mutagenesis later on the same day (Day 4) or overnight cultures can be made for mutagenesis on Day 5. Depending upon the incubation, colony PCRs can be performed on Day 5 or Day 6 (the day following mutagenesis).

Tables

Table 1. Details of characterized *S. flexneri* genes that were selected for mutagenesis. The selected genes are found both on the virulence plasmid and the chromosome. The variation in the number of spacers and their corresponding mutable sites is also mentioned.

Gene	Location	Spacers	Function	Mutant phenotype	Reference
vacJ	Chromosome	Three spacers; same mutable site	Lysis of protrusion during intracellular dissemination	Dissemination defect	(35)
icsA	Virulence plasmid	Two spacers; different mutable sites	Actin-based motility	Dissemination defect	(34)
icsB	Virulence plasmid	Two spacers; different mutable sites	Post-invasion virulence factor	Dissemination defect (?)	(38, 39)
mxiD	Virulence plasmid	One spacer	Structural component of the injectisome	Invasion defect	(31)

654

Supplementary Material

Supplementary Methods S1. Detailed methodology of construction of plasmids, strains and cloning guide RNA.

Supplementary Figure S1. Determination of J23119 promoter activity in *S. flexneri*. (A) The map of pgRNA showing the restriction sites used to clone sfGFP and obtain pgRNA-(R)-GFP. The sfGFP has a synthetic ribosome binding site and is cloned under the J23119 promoter. (B) *S. flexneri* strains carrying pgRNA-X as non-fluorescent control and pgRNA-(R)-GFP streaked on a TSA plate supplemented with carbenicillin (50 µg/ml).

Supplementary Figure S2. Viability of *E. coli*::*mCherry* during mutagenesis determined by counting the colonies spotted after 2, 4, 6, and 20h of IPTG induction.

Supplementary Figure S3. Viability of *S. flexneri*::*mCherry* during mutagenesis determined by counting the colonies spotted after 2, 4, and 6h of IPTG induction.

Supplementary Table S1. Bacterial strains and cells used in this study

Supplementary Table S2. Plasmids used in this study.

Supplementary Table S3. Antibiotic concentrations used for maintenance of plasmids.

Supplementary Table S4. Oligonucleotides used in this study

Supplementary Table S5. Mutations in *Shigella gyrA* known to result in nalidixic acid resistance as determined from DRAGdb database. The mutation selected for fluctuation assay has been highlighted.

Supplementary Table S6. List of all the spacer sequences used in this study.