Systematic genome engineering approaches to investigate mutational effects and evolutionary processes

Ph.D. Thesis

Ákos József Nyerges

Supervisor: Csaba Pál, Ph.D.

Doctoral School of Biology

Biological Research Centre of the Hungarian Academy of Sciences

Institute of Biochemistry

Synthetic and Systems Biology Unit

University of Szeged

Faculty of Natural Sciences and Informatics

Szeged

2019

Table of Contents

List of abbreviations

gDNA   genomic DNA

MIC   minimum inhibitory concentration

IC75   75% inhibitory concentration

OD   optical density

TRM   trimethoprim

TALEN   transcription activator-like effector nuclease

CRISPR   clustered regularly interspaced short palindromic repeats

MAGE   multiplex automated genome engineering

TM-MAGE   transient mutator multiplex automated genome engineering

ALE   adaptive laboratory evolution

MMR   methyl-directed mismatch repair

GRO   genomically recoded organism

GSK   GlaxoSmithKline plc

R&D   research and development

FoR   frequency-of-resistance

PMF   proton motive force

MO-MAGE   microarray-oligonucleotide multiplex automated genome engineering

ssDNA   single stranded DNA

dsDNA   double stranded DNA

TE buffer   Tris and ethylenediaminetetraacetic acid buffer

HPLC   high-performance liquid chromatography

LB^L lysogeny broth Lennox

TB   terrific broth

MS   Minimal Salt medium

MHBII   Mueller Hinton broth II

rpm   rotation per minute

UPEC	Uropathogenic *Escherichia coli*
DHFR	dihydrofolate reductase
NMR	nuclear magnetic resonance
ATCC	American Type Culture Collection
HNCMB	Hungarian National Collection of Medical Bacteria
IUPAC	International Union of Pure and Applied Chemistry
PCR	polymerase chain reaction
NCBI	The National Center for Biotechnology Information
EUCAST	European Committee on Antimicrobial Susceptibility Testing
FALCOR	Fluctuation AnaLysis CalculatOR
SMRT	Single Molecule Real-Time
DIvERGE	directed evolution with random genomic mutations

Introduction

1. Synthetic biology

The ability to 'read' and rationally 'write' genetic information is transforming basic and applied research. This ability defined a new field roughly 44 years ago when Wacław Szybalski, a Polish geneticist coined the term "synthetic biology" [1]. As Szybalski noted, "*Up to now we are working on the descriptive phase of molecular biology.* […] *But the real challenge will start when we enter the synthetic biology phase of research in our field. We will then devise new control elements and add these new modules to the existing genomes or build up wholly new genomes. This would be a field with unlimited expansion potential.*" Szybalski was right. His pioneering vision has become reality through the singular development of molecular biology, chemistry, and informatics [2]. This has evolved into the modern field of synthetic biology, which is broadly defined as a discipline for "designing and constructing biological modules, biological systems, and biological machines, or the re-design of existing biological systems for useful purposes" [3].

Although a detailed review of synthetic biology is largely beyond the scope of this thesis, since Szybalski's first notes on synthetic biology, the ability to rationally engineer biological systems has led to a plethora of novelties which have transformed both biotechnology and basic research. Briefly, biological engineering has paved the way for unprecedented advancements, including the production of antimalarial artemisinin in baker's yeast [4], the *de novo* design of enzymes with novel catalytic activities [5,6], as well as the manufacturing of artificial protein and DNA shells to specifically deliver payloads [7,8], and the construction of programmable genetic circuits and memories within living bacteria, yeasts, and mammalian cells [9,10].

Synthetic biology is inherently multidisciplinary and has relied on advances in distinct fields, including DNA sequencing, chemistry, informatics, and analytics. However, most directly, the development of this field is a result of breakthroughs in the synthesis and editing of genetic material [11]. The advent of technologies suitable to edit and write DNA sequences has created the subfield of genome engineering within synthetic biology [11–14]. Genome engineering aims the construction and rational modification of organismal genomes to construct genotypes that give rise to a desired function, i.e. a specific phenotype [11,15].

Given the sheer size of bacterial chromosomes and the rapidly decreasing cost of *de novo* DNA manufacturing, the construction of entire organismal genomes is now possible [14]. This "bottom up" approach of genome engineering gave the first evidence of the feasibility of generating infectious viruses (first the polio [16], later the "Spanish flu" H1N1 influenza [17],

and horsepox viruses [18]), entirely from computationally designed and stored DNA sequences. Then, by building on advancements in massively parallel DNA synthesis and assembly, in 2010 scientists from the J. Craig Venter Institute (JCVI, Rockville, USA) reported the preparation of the first self-replicating bacterial cells with an entirely synthetic genome [19]. By transplanting a computationally designed and *in vitro* synthesized *Mycoplasma mycoides* genome into *Mycoplasma capricolum* cells, Daniel Gibson and his coworkers at JCVI booted up the first bacterium whose genome was designed on a computer hardware [19,20]. However, while this *de novo*, "bottom up" construction of functional genomes holds far-reaching promise and will likely enable powerful applications in the future, it also comes with several drawbacks. First and most importantly, synthesizing entire genomes remains a substantial technical challenge. The 1.08×10^6 base pairs of the *Mycoplasma mycoides* genome represented hundreds of man-years of intense laboratory work, costing millions of US dollars. Also, scaling up *de novo* construction of entire genomes to the genome size of most biotechnologically applied bacterial species or yeasts turned out to be cumbersome and extremely time consuming [11,21]. Two ongoing endeavors, namely (I) the Synthetic Yeast 2.0 project which aims at the *de novo* construction of the entire chromosomal assembly of baker's yeast (*Saccharomyces cerevisiae*) with a genome size of approximately 12.1 megabase pairs [21], and (II) rEcoli-57, an *Escherichia coli* strain with a genome size of 3.97 megabase pairs in which the codon use is being reprogrammed to use only 57 distinct codons [22], have demonstrated that totally *de novo* synthesis and assembly of multi megabase pairs of DNA is frequently hampered by synthesis errors and the unpredictability of biological designs. As a notable example, a single base pair deletion in the essential gene *dnaA* resulted in the temporal inability to generate a viable *Mycoplasma* genome [14,19]. Thus, designing, constructing, and booting up genomes with perfect fidelity has remained a tremendous technical challenge so far.

Instead of synthesizing genetic material *de novo,* "top down" synthetic biology approaches are focusing on re-designing existing biological systems [11]. This approach has several advantages over "bottom up" construction. First, it does not necessitate the labor-intensive synthesis of long and costly genetic elements. Second, as it allows the construction of desired edits in living cells, it inherently couples modifications to survival and organismal fitness. Therefore, it provides a straightforward way to identify and counter-select deleterious changes and designs. This way, variants with a non-viable genotype will fail to survive during the construction process, and thus can be avoided. Additionally, deleterious changes will result in an immediately detectable growth defect which can be mitigated later. Finally, and most importantly, engineering living biological systems *in vivo* enables the simultaneous prototyping of billions of variants. This ability, in turn, greatly increases our chances to identify variants

with the desired trait. In living cells genotype is inherently linked to a phenotype, thus the desired traits of engineered cell populations can be isolated and continuously enriched by screening and selection. This approach enables the continuous directed evolution of desired traits – a method acknowledged by the Nobel prize in Chemistry in 2018, awarded to Frances Arnold for random mutagenesis and the directed evolution of enzymes and to George Smith and Sir Gregory Winter for the phage display of peptides and antibodies [23].

Thus, *in vivo*, "top down" construction of novel biological functions is a powerful and less failure-prone alternative to *de novo* synthesis.

2. Methods of microbial genome engineering

The availability of the first microbial genomic sequences at the end of the 20th century created the opportunity for functional genomics, to understand the genotypic background of phenotypes. By systematically interrogating gene functions at the genome scale, the first genome engineering endeavors focused on the systematic analysis of genotypic information. Based on the systematic assessment of these phenotype-to-genotype relationships, synthetic biology then exploited this knowledge to engineer novel biological functions.

Ideally, a method for genome engineering would allow the precise construction of multiple genotypic changes with high efficiency. Thereby, it minimizes the time and effort required to identify a specific genotype, and allows of generating multiple edits to efficiently explore the sequence space. These features directed method-developments for genome engineering towards achieving higher precision, higher number of simultaneous edits, and multiplexability within the same cell [24]. Techniques incorporating zinc-finger nucleases, transcription activator-like effector nucleases (TALENs), single- or double-stranded DNA-recombineering, RNA-directed nucleases (e.g. CRISPR-Cas9) and base-editors have been essential tools for the efficient modification of DNA in numerous species, including viral, bacterial, yeast and mammalian cells [24].

Although RNA-directed nucleases have greatly extended the range of organisms which can be subjected to genomic changes, there seems to be an upper practical limit when it comes to utilizing these techniques for simultaneous modifications of multiple loci [25,26]. In contrast, multiplex genome editing is required for explicit genotype-phenotype mapping, as well as for the modification of protein complexes and biosynthetic pathways. Given the plethora of currently available genome engineering techniques, rather than presenting a detailed introduction of one-by-one which has already been done by others [26–28] and our research group as well [24,29], I would focus on recent developments in recombination-

mediated genetic engineering (recombineering) reckoned as the most high-throughput and multiplexable genetic engineering technique to date [30,31].

Recombineering utilizes homologous recombination to integrate linear DNA strands into chromosomal or episomal DNA [32–34]. Linear DNA strands can be either double-stranded (ds) or single-stranded (ss). Among those, dsDNA-recombineering has been widely utilized to create gene deletions or the insertion of novel genetic material, i.e. new biosynthetic pathways or sensory circuits [34–36]. However, the advent of recombineering in genome engineering originates from the development of ssDNA-recombineering [37]. Compared to the dsDNA-based recombineering, ssDNA-recombineering uses short, computationally designable and mass-manufacturable ssDNA oligonucleotides (oligos) to introduce user-defined DNA alterations to the target locus of interest [30]. The most notable advantage of ssDNA-recombineering is that it can use short ssDNA oligonucleotides resulting in editing efficiencies up-to 35% without the necessary selection of the desired genotype, simultaneously at multiple positions on microbial chromosomes [38]. Moreover, recent advances have allowed to improve the synthesis of such oligos, reaching a throughput of hundreds of thousands of user-definable sequences in a single synthesis reaction [31]. These features, in conjunction with the straightforward and automatable execution of recombineering experiments (Figure 1), have allowed continuous recombineering and thus engineering billions of microbial variants within laboratory timescale [38].

Thus, single-stranded DNA oligonucleotide-mediated recombineering is an especially powerful tool to perform large-scale genome editing in bacteria. In *Enterobacteriaceae*, a diverse and both biotechnologically and medically relevant family of Gram-negative *Proteobacteria*, oligo-mediated recombineering utilizes elements derived from the λ bacteriophage in order to carry out genetic modifications [39–41]. Specifically, it is based on Beta, the ssDNA annealing protein of the λ Red recombination machinery. This so-called λ Red recombination system is RecA-independent and consists of 3 phage-derived proteins: Gam, Beta, and Exo. The first protein, Gam, prevents the degradation of linear dsDNA by blocking the cellular action of endogenous RecBCD nuclease. Exo then degrades the ends of recombinogenic linear dsDNA with its 5′ - 3′ exonuclease activity, and thereby generates ssDNA regions. Finally, Beta catalyzes the annealing of ssDNA fragments to the lagging strand at the open replication forks of bacterial chromosomes [34,42]. While dsDNA recombineering requires both Exo, Beta, and Gam, ssDNA recombineering requires only Beta to be present in the cell [37,40]. Beta in itself mediates the annealing of ssDNA oligonucleotides to the lagging strand of the replication fork. Following annealing, these oligonucleotides become incorporated into the replicating chromosome as Okazaki fragments. By exploiting the remarkable efficiency of Beta-mediated ssDNA annealing, small

modifications including single-nucleotide changes, insertions and deletions up to a length of 30 nucleotides can be encoded within a single oligo, which can be repeatedly and simultaneously introduced to target DNA to achieve targeted reprogramming at multiple loci of the host genome [43].

3. Multiplex Automated Genome Engineering

Recombineering can be multiplexed, i.e. multiple oligos are capable of modifying their genomic targets simultaneously and independently. These modifications may range from single individual mutations to large-scale genome refactoring, involving thousands of simultaneously edited targets in a cell population [31,38]. The straightforward computational design and cost-effective synthesis of the mutagenizing ssDNA strands (i.e. 90-nucleotide-long oligonucleotides) have made ssDNA recombineering a fundamental tool for synthetic biology [11,30]. This technique – called multiplex automated genome engineering (MAGE) [38] – is currently the most versatile and high-throughput approach available for real genome-scale editing, which allows the continuous generation of billions of genetic variants within a population (Figure 1). MAGE utilizes λ Red ssDNA-recombineering to simultaneously incorporate multiple ssDNA oligonucleotides in a rapid, automated, and high-throughput manner, and thereby rapidly creates the desired alleles and combinatorial mutational libraries.
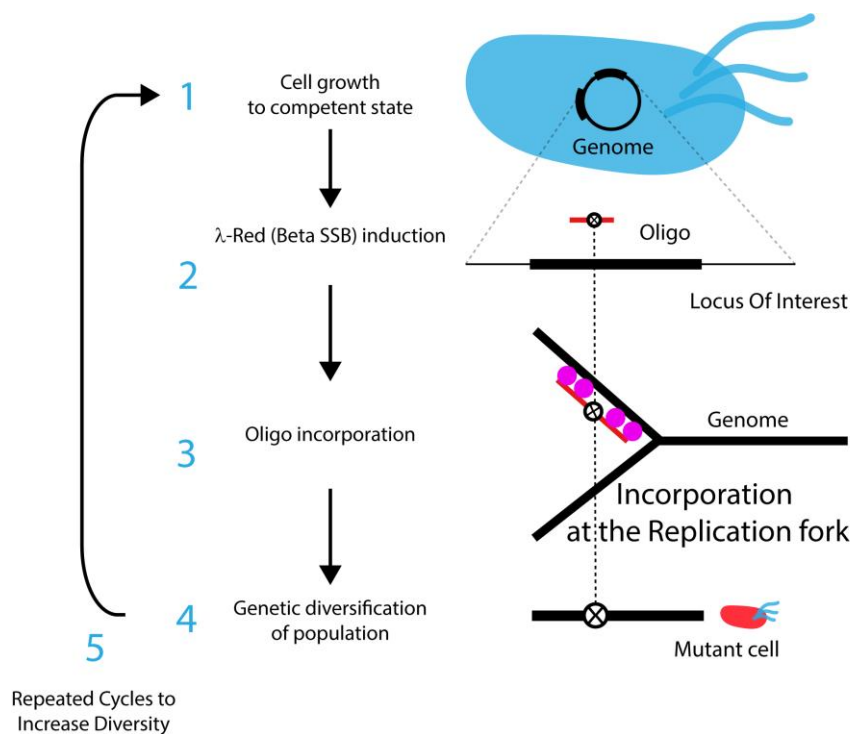

Figure 1. Overview of Multiplex Automated Genome Engineering. Chemical DNA oligo synthesis precisely determines the sequence of the editing template (ssDNA oligos). These

oligos are complementary to the genomic lagging-strand and introduce mutations into their target after annealing and subsequent integration as Okazaki-fragments. Iterative repetition of this process, consisting of cell growth (1), oligo delivery (2) and incorporation (3) increases editing efficiency and genetic diversity within the mutagenized cell population (4).

MAGE has the advantage of allowing genome engineering endeavors of unparalleled complexity, and it produces up to 10^{10} unique variants within a single screen, uniquely within days [38]. Through the accelerated optimization of biosynthetic pathways (e.g. lycopene and violacein [38,44,45]) and proteins [46–49], as well as by genome-wide codon replacement to create a so-called "genomically recoded organism" (GRO) engineered to depend on synthetic amino acids [50–52], MAGE has paved the way for previously unimaginable evolutionary innovations.

The functionality of ssDNA-mediated recombineering, and thus the opportunity to perform MAGE has been described in various species besides enterobacteria, including *Staphylococcus aureus* [53]*, Lactococcus lactis* [54]*, Lactobacillus* species [55]*, Corynebacterium glutamicum* [56]*, Salmonella enterica* [57]*, Bacillus subtilis* [58]*,* and *Pseudomonas syringae* and *putida* [59,60]. However, the straightforward implementation of MAGE beyond these early systems has remained considerably limited, and so far, most of these efforts have demonstrated only minor efficiency in species besides *E. coli*. Moreover, they all require time-consuming prior optimization for each individual species.

4. The role of methyl-directed mismatch repair

Even in *E. coli*, the efficient and unbiased incorporation of mutations by MAGE, as well as extensive modifications, including the expression of the λ Red recombinase enzymes and the inactivation of the native methyl-directed mismatch repair (MMR) system need to be executed for a successful genome editing process in the host strain [40,61]. It is explained by the fact that the incorporation of mutagenizing MAGE oligonucleotides unavoidably create a mismatch, and these mismatches are inevitably recognized by the cells' endogenous MMR system. Short stretches of mismatching nucleobases, deletions or insertions of up to six consecutive nucleotides are efficiently recognized by MMR, resulting in the complete removal of the integrated modifications [62].

In *E. coli* and evolutionarily-related *Proteobacteria* these DNA lesions are recognized by MutS, and the removal of a mismatching base is initiated jointly with MutL and MutH [63,64]. Of these three proteins, the role of MutS is to locate and recognize mismatches on the newly replicated and thereby hemimethylated DNA strand. After the mismatch is recognized, MutS forms a

complex with multiple monomers of MutL. MutL acts as a linker to recruit MutH. MutH then binds to the hemimethylated strand and incises the newly synthesized, nonmethylated DNA strand. After the strand is cut, UvrD (formerly MutD) with a helicase action separates the two DNA strands, and the mismatch-containing strand is digested by exonucleases. Finally, the daughter strand is repaired by DNA polymerase and ligase. The name of Mut proteins reflects that their absence results in an elevated mutation rate [63,64].

Therefore, efficient ssDNA-mediated recombineering requires the inactivation of the mismatch repair system [61]. However, in this case background mutation rate consequently increases by nearly two orders of magnitude compared to the wild-type mutation rate. In turn, this elevated mutagenesis leads to the accumulation of undesired, off-target mutations during the course of long-term MAGE experiments (i.e. in multiple MAGE cycles). To demonstrate the magnitude of this problem it should be noted that in a recent work where MAGE was used to construct the aforementioned, so-called "genomically recoded organism", the authors recognized that off-target mutations caused reduced fitness in engineered strains [51,65,66]. Strikingly, besides the 321 mutations which were intentionally generated by MAGE, a total of 355 unwanted off-target mutations were also detected. These off-target mutations may in turn interfere with the phenotypic effects of the engineered modifications, and in extreme cases, they may mask the desired mutational effects.

Recently, we have attempted to address this issue by replacing wild-type *mutL* and *mutS* with heat-sensitive mutants of those genes, accompanied by limiting the inactivation of mismatch repair to the short period of the MAGE cycle only, executed by altering the temperature between permissive and non-permissive values during MAGE cycles [67]. Although we have managed to reduce the number of off-target mutations by 85%, the time-consuming genetic manipulation of the parental strain was still a prerequisite. This issue has been recently addressed by the so-called transient-mutator MAGE (TM-MAGE) technique, which allows the plasmid-based engineering of *E. coli* chromosomes by temporally downregulating MMR activity via the transient hypermethylation of bacterial chromosomes [68]. However, the functionality of TM-MAGE was only demonstrated in certain *E. coli* strains, therefore portability of MAGE across species remained an unsolved challenge.

5. Using genome engineering to understand microbial evolution

Methods of genome engineering enable the modification of organismal genomes in a directed and combinatorial manner. Thereby, these methods offer an unprecedented opportunity to study evolutionary processes which otherwise would not be possible with standard laboratory methods [29].

Traditional adaptive laboratory evolution (ALE) experiments have already been acknowledged as a powerful tool to observe and analyze evolutionary processes real-time, within laboratory timescales [69–72]. Bacterial populations offer an exceptionally great opportunity to perform ALE for several reasons. (I) Due to their small cell size, bacterial populations containing up to 10^8–10^{10} bacteria can be handled and iteratively cultured easily; (II) due to the modest generation time (ranging from tens of minutes to hours) of most bacterial models, ALE experiments can run for hundreds to thousands of cell generations, and also (III) due to their small genome size and the wealth of available analytical techniques, whole-genome sequence, transcriptomic, and phenotypic information can be obtained on the entire course of an evolutionary experiment. Moreover, as a practical advantage, (IV) samples taken at various time points during the experiment can be stored frozen and conveniently analyzed later [29,69,71,72].

On the other hand, even bacterial adaptive laboratory evolution experiments suffer for limitations that prevent the truly comprehensive analyses of sequence landscapes and evolutionary innovations. These limitations are primarily explained by the lack of natural variation within microbial populations, as evidenced by modest natural mutation rates and biases in natural mutagenic processes in most bacteria. These limitations, in conjunction with microbial population sizes and experimental time-frames feasible under laboratory conditions, hinder the in-depth exploration of evolutionary forces which drive complex and slowly evolving traits [15,29,73].

Genome engineering, however, offers the opportunity to overcome these limitations. Methods for multiplex, high-throughput genome engineering, including MAGE, RNA-guided endonucleases, and base-editors, enable the rapid and targeted construction of chromosomal alterations, i.e. the evolutionary steps which otherwise would not occur under laboratory conditions [26,38,74–77]. Also, genome engineering is suitable to generate diverse and unbiased mutational libraries at many different loci, thereby it facilitates the development of genetic changes that otherwise would not occur spontaneously in living cells [11,24,29].

Exploiting these capabilities has already led to a plethora of advancements in and novel insights into both biotechnology and basic science [2,78–83]. *De novo* genome constructions have shed light on genome architecture in bacteria and yeast [14,21], and top-down genome reduction has increased our understanding of the role of mobile genetic elements and prophages [84,85], as well as the feasibility of creating synthetic, minimal genomes [14,21,86]. Last but not least, genome editing and large-scale mutagenesis coupled to phenotypic screens have allowed the systematic exploration of evolutionary landscapes, the directed evolution of novel phenotypes, and the understanding of evolutionary effects

which play a role in the development of drug resistance, the evolution of metabolic pathways, and other complex cellular traits [6,27,87–90].

6. Evolution and consequences of antibiotic resistance

Microbial evolution, however, is not restricted to the laboratory only. Bacteria constantly face selection pressure, including the artificial defense lines of host organisms, including antimicrobials and small-molecule antibiotic drugs applied in therapy and agriculture. This selection pressure, coupled with the exceedingly large population size of natural bacterial populations and the ability to mutate and exchange genetic material (i.e. horizontal gene transfer), has led to the salient rise of bacterial antibiotic resistance [91–93].

Worldwide, the emergence of bacterial resistance against existing antibiotics is currently responsible for an estimated 700,000 deaths annually. According to pessimistic estimates, antimicrobial drug resistance of pathogenic bacteria will be the leading cause of death by 2050, unless novel antimicrobial strategies are developed [94]. Strikingly, 90% of major pharmaceutical companies had discontinued their antibiotic research programs by 2018, primarily because of the rapid appearance of resistant bacteria, which makes the commercial success of new antibiotics unpredictable [95–97].

Antibiotics selectively target essential constituents of microbial life, i.e. mechanisms and molecules which solely exist in bacteria and are absent in the human host [98]. This selectivity relies on targeting bacterial enzymes, ribosomal proteins, or the bacterial cell wall. However, bacteria typically utilize one or more of a few possible opportunities to evade the antibiotics' mode of action [93,99,100] (Figure 2).
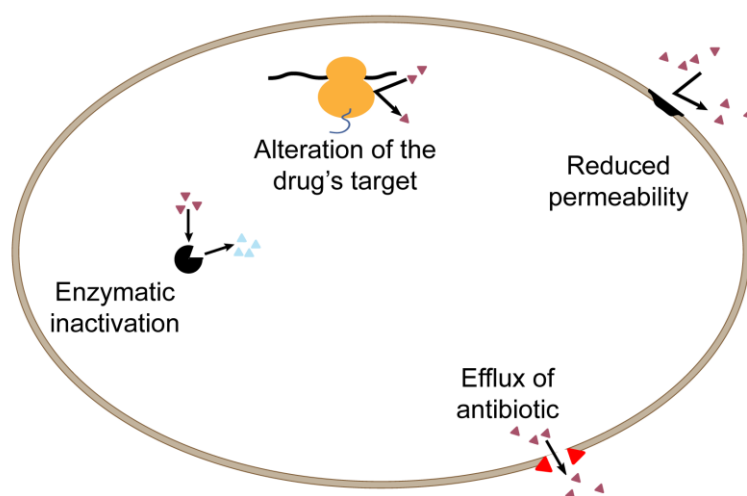
Figure 2. Schematic mechanisms of antibiotic resistance. For a detailed description of each mechanisms, see the section "Evolution and consequences of antibiotic resistance" on page 14.

Briefly, one of these resistance mechanisms relies on the inactivation of drug molecules by chemical modification or degradation (such inactivators include the widely used selection markers chloramphenicol-acetyltransferase and beta-lactamase enzymes). Second, bacteria are capable of altering or overexpressing the drug target, leading to decreased susceptibility. Such modifications explain resistance to e.g. fluoroquinolone antibiotics, achieved by mutations at GyrA and ParC, as well as include the promoter and active-site mutations developed against folate-biosynthesis blockers, such as trimethoprim. Besides these mechanisms, bacteria are also capable of keeping antibacterial drugs out of their cells efficiently, either through efflux or by altering the permeability of the cell wall. These 'counter-processes' rely either on drug-specific exporters or on multidrug efflux pumps. Alternatively, bacteria may prevent drug entry by reducing the permeability of the bacterial cell wall, either by structural changes or by the reduction of proton motive force (PMF) [101]. Moreover, high levels of drug resistance frequently require multiple complementary mechanisms, and in turn, genetic changes to evolve.

In general, antibiotic resistance can emerge as a result of mutations at endogenous chromosomal loci or may result from horizontally-transferred genetic elements. The relative importance of these genetic mechanisms in resistance evolution depends on the given antibiotic and the bacterial strain-of-interest [99,102].

7. Assessing antibiotic resistance evolution

Given the fundamental impact of antibiotic resistance on drug development and the treatment of infections caused by drug resistant pathogens, it would be imperative to accurately identify the possible mechanisms of resistance evolution at an early stage of antibiotic development [103]. However, assessing the risk of resistance development in the preclinical stage is especially challenging. Standard laboratory evolutionary techniques explore only a small fraction of the sequence space, and fail to identify exceedingly rare resistance mutations and their combinations [103–106]. The case of GSK 2251052, a novel antibiotic candidate (formerly named as epetraborole and AN3365) sheds light on the unpredictability of resistance mechanisms before clinical applications. GSK 2251052 blocks leucyl-tRNA synthetase in Gram-negative pathogens responsible for urinary tract infections, including *E. coli*. Unexpectedly, resistance to this drug candidate was identified early during its phase II clinical trial. Due to this early emergence of resistance, GlaxoSmithKline, the developer of GSK 2251052 has terminated clinical development, thereby wasting its R&D costs reaching tens of millions of USD [107].

Predicting evolutionary processes which may lead to decreased antibiotic susceptibility is a complex issue for several reasons. First, because a large number of molecular mechanisms can contribute to resistance. Second, resistance mechanisms may differ among different pathogenic bacteria. Third, the standard adaptive laboratory evolution assays that are frequently used both in basic research and in industrial laboratories to assess resistance evolution (such as fluctuation tests and serial passaging experiments) generally rely on natural mutational processes which are modest and biased in many cases [103–105,108,109]. Therefore, these methods explore only a fraction of potential real-life resistance mechanisms. For a brief visual description of these techniques see figure 3.
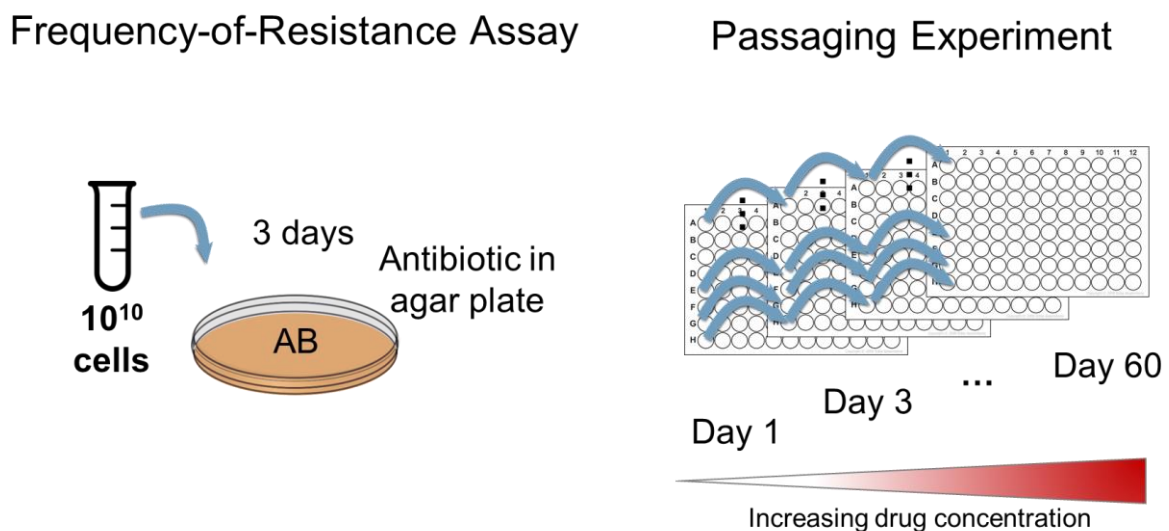
Figure 3. Traditional laboratory methods of antibiotic resistance analysis. While frequency-of-resistance (FoR) assays primarily detect single-step mutations, passaging experiments can explore multistep mutational processes that may lead to high-level antibiotic resistance.

As a good benchmark for identifying antibiotics which require only a single mutation to have their bioactivity abolished, the mutation rate and population size for predictive tests should be high enough to evaluate at least three mutations at 99.9% of nucleotide positions in the whole genetic material. This equals to assaying $2.4{\times}10^{11}$ wild-type cells, roughly 5 milliliters of a saturated *E. coli* culture, in a standard frequency-of-resistance (FoR) assay [103]. However, this assay generally detects only single-step mutational processes, when resistance against a given drug requires only a single genetic alteration to appear. Consequently, for antimicrobials that require more than a single mutation to evolve detectable resistance, the mutational space needed to be explored scales exponentially with the number of sites. In turn, for a drug that necessitates two mutations to evolve a resistant phenotype, the identification of resistance-causing mutation combinations would require screening over

$4.5×10^{22}$ wild-type *E. coli* cells in a frequency-of-resistance assay [98]. Using hypermutator strains would reduce this number to a maximum of $3.22×10^5$ (i.e. MP6-mutagenesis [110]), but this number still remains infeasible with current microbiological tests [109].

In the other hand, however, recent genome-engineering technologies are promising alternatives to explore resistance processes in a more systematic and accelerated manner. Through the generation of diverse and unbiased mutant libraries at many defined loci, genome engineering allows the generation of genetic changes which otherwise would be highly unlikely to occur under laboratory timescales. However, technologies that enable the targeted mutagenesis of multiple loci in their native genomic context, and in turn the exploration of antibiotic resistance processes, suffer from serious limitations [111].

Certain methods target the entire genome unselectively [112,113] and as a consequence, these assays result in the accumulation of undesired, off-target modifications with detrimental side effects. In other cases, the lengths of targeted regions are highly limited [74,114–117], precise adjustment of the mutation rate is unsolved [74,114], or the mutational spectrum is highly biased [74,89,110,114,116–119]. Finally, recently described techniques employing CRISPR-Cas9 nucleases and base-editors for *in vivo* mutagenesis require the presence of a protospacer adjacent motif in the vicinity of the target site, which limits the available targets and prohibit multiple rounds of mutagenesis to generate multistep evolutionary processes [26,74–76,116,120–124]. For a more detailed comparison of existing *in vivo* mutagenesis protocols, see our recent review in Current Opinion in Microbiology [24] and the comparison of *in vivo* mutagenesis methods in our article in the Proceedings of the National Academy of Sciences of the United States of America [125].

Amongst the methods available for large-scale mutagenesis and phenotyping, the aforementioned single-stranded (ss) DNA-mediated recombineering, and specifically multiplex automated genome engineering (MAGE) stands out [38,61]. However, existing ssDNA recombineering-based techniques enable randomization of very short sequences (such as promoters or neighboring residues in a protein-coding sequence) by a single oligo [46,47,126,127], which limits the explorable sequence space. The explanation for this limitation is the fact that the efficiency of ssDNA-mediated recombineering depends on the oligo's sequence identity to the target region [30,38,43] (see also Figure 10), i.e., the decrease of sequence identity for any recombineering oligo compared to its target will simultaneously increase the binding free energy (∆G), thus preventing their hybridization to the corresponding target site. Therefore, the diversification of genomic sequences longer than 30 base pairs is not feasible with a single oligonucleotide. Several strategies based on MAGE have been proposed to mutagenize the full length of individual genes in their native genomic context, including MAGE-seq and MO-MAGE [31,128]. However, as a common feature of these

methods, individual nucleotide positions are mutagenized separately by using a distinct oligonucleotide for each [31,128,129]. As a consequence, these methods demand the costly and time-consuming manufacturing of hundreds to thousands of individually designed and synthesized oligos even for a single prokaryotic gene [31,128].

To address this issue, during the eight years of my research in the Biological Research Centre of the Hungarian Academy of Sciences me and my colleagues have focused on improving ssDNA-recombineering to allow the precise and cost effective mutagenesis of multiple, long genomic segments in various species without unwanted, off-target modifications [130]. The resulting methods now enable the systematic comparison of mutational effects across different species, as well as the exploration of the phenotypic effects of a vast number of mutations in their native genetic context. Moreover, the application of these developments has contributed to a better understanding of microbial evolution and resistance profiles of antibiotics [131,132]. As a step-forward towards the predictability of antibiotic resistance, these developments have allowed us to predict the evolution of resistance against an antibiotic currently in clinical trials, and have contributed to the preclinical development of a novel class of antibiotic as well.

Objectives

Our research has focused on improving bacterial genome engineering towards precision, increased throughput, multiplexability, and a broader host range to facilitate its applications in basic and applied research. Specifically, we first aimed to develop a broad host range, plasmid-based ssDNA-recombineering system which can efficiently operate in a wide range of enteric bacteria, including those frequently used in microbial fermentations or are significant human pathogens. Next, building on this advancement, we aimed to develop an extremely high-throughput genome engineering technique that efficiently generates a vast number of genotypic alterations in bacterial genomes in their native genetic context. Finally, we examined whether ssDNA-recombineering enables an accelerated analysis of microbial mutational processes and the directed evolution of antibiotic resistance in multiple bacterial species.

To achieve these goals, our research included the following steps:
- We characterized a dominant mutator allele of the methyl-directed mismatch repair system of *E. coli*, and analyzed the phenotypic conservancy of its effect across enterobacterial species.
- We designed and constructed a broad host range, plasmid-encoded system for ssDNA-mediated genome engineering to allow simultaneous mismatch-repair control and efficient genome editing.
- We developed a method of DNA synthesis which can introduce large genetic diversity into user-defined oligonucleotide strands without mutational bias in a cost-effective way.
- We developed a method for ssDNA-mediated genome engineering which can introduce randomly distributed, random mutations and their combinations along the entire length of multiple long and continuous genomic segments simultaneously in multiple bacterial species.
- We compared the mutational effects that give rise to antibiotic resistance phenotypes across phylogenetically related bacterial strains.
- We identified evolutionary processes that can lead to antibiotic resistance to new antibiotic candidates that are currently under clinical development.

Resources and methods

Oligonucleotides

A full list of all DNA oligonucleotides used in this work are listed in Appendix 2 and 3. Oligos were ordered as standard desalting for purification from Integrated DNA Technologies or synthesized in-house at the Nucleic Acid Synthesis Laboratory of the Biological Research Centre of the Hungarian Academy of Sciences (Szeged, Hungary) according to a standard phosphoramidite-based DNA synthesis procedure. High-throughput sequencing primers for Illumina and Pacific Biosciences sequencing platforms were purified with high-performance liquid chromatography (HPLC) to avoid truncated variants that can interfere with precision and the separation of subsamples. Oligonucleotides after manufacturing, desalting, and subsequent lyophilization were suspended in 1 × TE buffer (10 mM Tris-HCl (pH 8.0); 0.1 mM ethylenediaminetetraacetic acid (EDTA)) and stored at -20 °C .

The design of ssDNA-recombineering oligos followed a general guideline [43,133]: Genome editing oligonucleotides I.) were 90 nucleobases in length, II.) had minimized secondary structure (ΔG higher than -12 kcal/mol) and III.) lacked mistargets on the target strain's genome to avoid false hybridization and off-target mutagenesis. Moreover, to perform mutagenesis at E. coli K-12 MG1655 *lacZ, malK, araB, hisB, rpsL*, and *cycA*, oligonucleotides had 2 phosphorothioate bonds at each terminus in order to evade ssDNA exonucleases within the target cells.

Synthesis of soft-randomized DIvERGE oligonucleotides

Synthesis of DIvERGE and reference oligos was performed on an ABI 3900 DNA synthesizer (Applied Biosystems Inc), according to a modified phosphoramidite chemistry-based protocol. As a solid support controlled pore glass was applied, and the following synthesis cycles were repeated: I.) Deprotection was achieved with 3% (weight/volume) trichloroacetic acid in dichloromethane. II.) Incoming phosphoramidite, dissolved in 0.055 M concentration in anhydrous acetonitrile and premixed with the other three amidites in the defined spiking ratio according to the given experiment was coupled after activation with 5-ethylthio-1H-tetrazole. III.) Capping was done with 10% (volume/volume) acetic anhydride in anhydrous tetrahydrofuran and 16% (volume/volume) N-methyl-imidazole and 10% (volume/volume) pyridine containing anhydrous tetrahydrofuran. IV.) The oxidation step was accomplished with iodine. To this aim 5 g iodine was dissolved in a liter of pyridine: water: tetrahydrofuran mixture in a 0.5:2:97.5 ratio. Oligo synthesis cycles were repeated until the 5' terminal position of the

DNA strand. Finally, DNA strands were cleaved from the solid carrier with concentrated aqueous ammonia solution. Crude oligos were then purified by high-performance liquid chromatography (HPLC). After concentration from HPLC fractions, 5'-dimethoxytrityl protecting groups were removed by using a PolyPak column (from Glen Research) according to the manufacturer's instructions. Following elution and subsequent lyophilization, purified oligos were resuspended in 1×TE buffer (10 mM Tris-HCl (pH 8.0); 0.1 mM ethylenediaminetetraacetic acid (EDTA)).

Applied bacteriological media and their compositions

Unless otherwise noted, bacterial cultures were grown in Lysogeny Broth Lennox (LB^L) medium (consisting of 10 g of tryptone, 5 g of yeast extract, 5 g of sodium chloride per 1000 mL of water). LB^L agar plates were prepared by the addition of 13.5 g agar per 1000 ml of culture medium (Sigma-Aldrich) before autoclaving for 20 minutes at 115 °C. Antibiotics that were used for either the maintenance of plasmid-constructs or the selection of mutants, were added to the LB^L broth or agar-containing media at 55 °C. Antibiotic solutions were obtained from Sigma-Aldrich, unless otherwise noted, and were prepared from powder stocks and filter sterilized before use. AraB, LacZ, and MalK enzyme-activities were assayed on supplemented MacConkey agar plates (consisting of peptone 20 g, bile salts 1.5 g, sodium chloride 5 g, agar 13.5 g, neutral red 0.03 g, and crystal violet 1.0 mg per 1000 mL of water) with either 1% of arabinose for the functional analysis of AraB, 1% of lactose for LacZ, or 1% maltose for MalK enzymatic activities. Following electroporation, terrific-broth (TB) was applied as cell recovery media (yeast extract 24 g, tryptone 12 g, K_2HPO_4 9.4 g, KH_2PO_4 2 g per 1000 mL of water). Bacterial media were heat-sterilized by autoclaving at 115 °C for 20 minutes. Minimal salt medium (MS), supplemented with 0.2% glucose and 0.1% casamino acids (Difco Laboratories) was applied for trimethoprim drug susceptibility assays due to its low folate content, that would otherwise interfere with the competitive inhibition of FolA.

ssDNA-recombineering protocol

To perform ssDNA-recombineering and iterative Multiplex Automated Genome Engineering (MAGE) cycles, individual bacterial colonies were inoculated into 2 ml LB^L medium in the presence of the corresponding antibiotic. These starter cultures were then incubated overnight in a shaking incubator at 30 °C and a continuous shaking of 250 rotation per minute (rpm). Next, these stationer phase starter cultures were diluted 100-fold in antibiotic-supplemented LB^L medium. Each MAGE cycle consisted of the following steps: Upon reaching OD$_{600}$ = 0.4-

0.6, cells were transferred to a 42 °C shaking water bath to induce λ Red protein expression for 15 minutes at 250 rpm agitation. Immediately after heat induction, cells were chilled on ice for 10 minutes. Next, cells were made electrocompetent by washing and pelleting twice in 10 ml ice-cold H2O in a refrigerated Eppendorf 5702R centrifuge (at 4000 rpm for 8 minutes). Bacterial cells were then suspended in 160 µl chilled H_2O and kept on ice until electroporation. Unless otherwise noted, 40 µl electrocompetent bacterial cell suspension was admixed with 1 µl of the corresponding 100 µM ssDNA oligonucleotide. For simultaneous allelic replacements at *lacZ, malK, araB, hisB, rpsL*, and *cycA* 1-1 µs of each oligo were admixed and electroporated. Electroporations were performed on a BTX CM-630 Exponential Decay Wave Electroporation System (from Harvard Apparatus) within 1 mm gap electroporation cuvettes Pulse conditions were 1.8 kV, 200 Ω, 25 µF. Immediately after electroporation, 1 ml TB medium was added to each cuvette and cells were then transferred to 5 ml TB media to recover and start to divide. Following 60 minutes recovery at 30 °C under constant agitation (250 rpm), 5 ml antibiotic containing LB^L medium was added to each culture. Finally, cells were allowed to reach mid-logarithmic growth phase under continuous agitation. At this point, cells were either subjected to an additional MAGE cycle or allowed to reach stationary phase and were analyzed for phenotype and genotype.

To assess the performance of ssDNA-recombineering in *E. coli* K-12 MG1655 , single recombineering cycles were performed in E. coli K-12 MG1655 + pSIM8, E. coli K-12 MG1655 + pORTMAGE2, and in E. coli K-12 MG1655 ∆mutS + pSIM8 by using oligonucleotides that introduced single base pair mismatches into *lacZ* and thereby generated premature stop codons.

For the determination of the off-target effects during long-term MAGE experiments, *E. coli* K-12 MG1655 + pSIM8, *E. coli* K-12 MG1655 + pORTMAGE1, and *E. coli* K-12 MG1655 ∆*mutS* + pSIM8 cells were subjected to oligo-mutagenesis. During every MAGE cycle, *lacZ, malK, araB, hisB, rpsL*, and *cycA* were subsequently targeted by 4 MAGE cycles for each locus. Then, allelic replacement efficiencies were measured at each locus either by colorimetric assay or allele-specific PCRs with the following primer pairs for *cycA* and *hisB*, respectively: cycAASP_f and cycAASP_r, hisBASP_f and hisBASP_r. Finally, isolates that carried the desired mutation were verified by capillary sequencing. After the identification of the correct genotype, recombineering cycles targeting the next locus were initiated. After all 6 desired allelic replacements were identified in the given clone (with the exception of *hisB* mutation in MG1655 + pSIM8 cells, which mutation was not observable even by screening 760 distinct colonies, presumably due to the high recognition efficiency of mismatch repair for the corresponding mismatch type), genomic DNA (gDNA) was isolated with a GenElute Bacterial Genomic DNA kit from Sigma-Aldrich, according to the manufacturer's protocol. These gDNA samples were subsequently subjected to whole genome sequencing.

DIvERGE cycling protocol

DIvERGE in *E. coli* K-12 MG1655, *E. coli* UPEC CFT073, *Salmonella enterica* LT2, and *Citrobacter freundii* ATCC 8090 was performed according to a modified pORTMAGE3 plasmid-based ssDNA-recombineering protocol. In our assays, *E. coli* K-12 MG1655 represented a widely used laboratory *E. coli* strain, *E. coli* UPEC CFT073 represented a human uropathogenic strain of *E. coli*, *Salmonella enterica* LT2 served as a model for human pathogenic *Salmonella* strains, while *Citrobacter freundii* ATCC 8090 is the reference and type strain for *Citrobacter freundii,* an emerging opportunistic pathogen.

In order to measure the efficiency of mutagenesis at the *asnA*::*tetR*-CAT landing pad, TETRM1 and TETRM3 soft-randomized oligos with randomization levels of 0.5%, 1%, and 2% were targeted to the landing pad on the genome of *E. coli* K-12 MG1655*, Salmonella enterica LT2, and Citrobacter freundii* ATCC 8090. Next, five consecutive cycles of DIvERGE were performed with each oligo. After the 5th mutagenesis cycle, gDNA was extracted by using GenElute Bacterial Genomic DNA kit from Sigma-Aldrich, and the DNA was subjected to PCR amplicon-based genotypic analysis on Illumina MiSeq.

To analyze the single-step mutational landscape of FolA and its promoter, a mixture of eight, *folA*-targeting, soft-randomized oligos were electroporated into pORTMAGE3 containing *E. coli* K-12 MG1655 and *E. coli* CFT073 cells in triplicates. After the addition of 1 ml TB recovery media into each electroporation cuvettes, 0.3 ml of each cell suspension was admixed, separately for all replicates and allowed to recover in 5 ml TB media at 30 °C under constant agitation for 60 minutes at which timepoint 5 additional milliliters of LB^L medium was added to each culture. The mutagenized cell populations were then allowed to grow until stationary phase at 30 °C.

Multiplex, iterative DIvERGE mutagenesis was carried out to mutagenize *folA* by performing the genomic integration of an equimolar mixture of 8, soft-randomized oligos that covered the entire target site. 0.5 - 0.5 μl of each oligo were electroporated into heat-induced, pORTMAGE3 containing electrocompetent cells. Following electroporation, cell growth and recovery phases were performed according to the general ssDNA-recombineering procedure, and the whole procedure was repeated for five consecutive times.

A single DIvERGE mutagenesis cycle was carried out in *E. coli* K-12 MG1655 to generate a *gyrA, gyrB, parE,* and *parC* combinatorial mutant library. After equimolarly admixing 130 soft-randomized oligos that covered all four loci with slight overlaps, 4 μl of this oligo pool was electroporated into *E. coli* K-12 MG1655 cells that carried the pORTMAGE3 plasmid. To increase mutant library size, electroporation was performed in 10 parallel

replicates, in each electroporating 40 µls of electrocompetent cells with 4 µl of this oligo pool. The resulted cell libraries were then pooled and was then allowed to recover in 50 ml fresh TB media for 1 hour at 30 °C under constant agitation. Following recovery, cells were diluted with 50 ml LB^L media and allowed to reach stationary phase at 30 °C while continuously agitating the library at 250 rpm.

Finally, aliquoted samples from each experiment were mixed with 50 volume/volume% of 50% glycerol and then frozen and stored at -80 °C. Besides storage from each cell library, genomic DNA (gDNA) was also isolated from 500 µl of each population with a GenElute Bacterial Genomic DNA kit from Sigma-Aldrich, according to the manufacturer's protocol. These gDNA samples were subsequently subjected to targeted amplicon sequencing.

MP6 plasmid-based *in vivo* mutagenesis

MP6 is a potent, L-arabinose inducible, broad-spectrum, plasmid-based mutagenesis system that enhances mutation rate 322 000-fold over the basal level in *E. coli*, thus surpassing the mutational efficiency of other, widely used *in vivo* and *in vitro* whole-genome mutagenesis methods [110]. In sum, the inducible expression of five mutator genes from the synthetic operon of MP6 (namely *dnaQ926, dam, seqA, emrR, ugi, and cda1*) alters the native cells proofreading, base excision repair, base selection, and mismatch repair capabilities and thereby exceedingly elevating mutation rate. For this assay, an MP6 (Addgene plasmid ID: 69669) containing *E. coli* K-12 MG1655 starter culture was grown overnight at 30 °C in LB + 25 µg/ml chloramphenicol in the presence of 25 mM glucose and diluted 1000-fold into 12 parallel samples in 1 ml LB^L media. MP6 mutagenesis was induced by adding L-arabinose at a final concentration of 200 mM. As a control, wild-type *E. coli* K-12 MG1655 population were inoculated in the same manner in 12 parallel replicates. Cultures were grown overnight at 30 °C.

Construction of pORTMAGE and pZA31tetR-mutLE32K plasmids

pORTMAGE1 (Addgene plasmid ID: 72680) and derivatives were constructed by introducing the gene encoding *E. coli* MutL (ec*mutL*) containing an E32→K mutation (ec*mutL*E32K) into the pSIM8 plasmid [57]. pSIM8 was donated by Donald L. Court (National Cancer Institute in Frederick, MD, USA). pORTMAGE1 and derivatives were constructed by introducing ec*mutL*E32K into pSIM8. To this aim, first, the native *mutL* from *E. coli* K-12 MG1655 was cloned into the λ Red operon of pSIM8, downstream of *exo* and upstream the tL terminator to yield pSIM8mutL. The correct assembly of pSIM8mutL was verified by PCR and subsequent capillary sequencing with LExoF and tL3R primers. Next, the introduction of ec*mutL*E32K was

achieved by whole-plasmid amplification mutagenesis. To this aim, PCRs in 50 μl total volume were cycled 35 times at 98 °C 10 seconds, 56 °C 30 seconds and 72 °C for 5 minutes, with a final extension time of 6 minutes at 72 °C, using mutL32F and mutL32R as PCR primers. The PCR amplicon was then treated with 1 unit of DpnI restriction enzyme, directly within the PCR buffer for 60 minutes at 37°C, and then purified and concentrated into 12 μl deionized water by using a Zymo DNA Clean-and-concentrator kit, according to the manufacturer's instruction (Zymo Research). Finally, the circularization of the plasmid was performed at 18 °C overnight with T4 DNA Ligase (Thermo Scientific). The final, ligated plasmids were then electrotransformed into E. coli K-12 MG1655 electrocompetent cells. Correct clones were verified by PCR and subsequent capillary sequencing with LExoF and tL3R primers.

pZA31tetR-mutLE32K was constructed by introducing the *mutL*E32K allele into pZA31tetR under the control of the pL$_{tetO}$ regulatory unit [134].

pORTMAGE2 (Addgene plasmid ID: 72677) was constructed by introducing a strong ribosomal binding site (5'-GAGAGGAGGTATATAC) upstream of the ec*mutL*E32K allele in pORTMAGE1 by whole-plasmid amplification mutagenesis.

Next, we constructed kanamycin and chloramphenicol resistance marker-based variants of pORTMAGE2: pORTMAGE3, and 4, respectively. The kanamycin-resistant pORTMAGE3 (Addgene plasmid ID: 72679) and the chloramphenicol-resistant pORTMAGE4 (Addgene plasmid ID: 72679) were constructed by replacing the beta-lactamase gene of pORTMAGE2 with either the Kan^R or the *cat* resistance marker by using Gibson-assembly. The plasmids were linearized by PCR amplification with pL32K frame_1, and pL32K frame_2 primer pairs and the kanamycin-phosphotransferase gene (Kan^R) or chloramphenicol-acetyltransferase (*cat*) gene was amplified using the Gibson Kan_Fw - Gibson Kan_rev and Gibson Chlo_Fw - Gibson Chlo_rev primer pairs, respectively. Assembly reactions were performed in Gibson Assembly Master Mix (New England Biolabs) at 50 °C for 60 minutes. Purified products were electroporated into *E. coli* DH5α electrocompetent cells and plated to agar plates containing the corresponding antibiotic. Plates were incubated at 30 °C overnight. Successful plasmid assemblies were then verified from the outgrowing colonies by colony PCR and capillary sequencing.

Integration of landing pad sequence into the target species

In order to measure allelic-replacement efficiencies uniformly across species, we integrated an artificial landing pad sequence into the genome of *E. coli* K-12 MG1655, *S. enterica serovar Typhimurium* LT2, and *C. freundii* ATCC 8090. This landing pad sequence was integrated into the endogenous *asnA* in every organism by utilizing dsDNA-recombineering. AsnA is one of

two asparagine synthetase enzymes in these species and is located at conserved genomic locus nearby the origin of replication in all three species. Due to the redundancy of its function in enterobacterial genomes, inactivation does not cause a fitness defect in rich media under the conditions where our assays were performed. To insert the landing pad into *asnA*, the recombineering cassette with additional 3' and 5' 50 nucleobase long homologous overhangs to the genomic target site was generated by PCR from pZA31YFPtetR, using the corresponding primer combinations Cint_F and Cint_R for *Citrobacter freundii*, Sint_F and Sint_R for *Salmonella enterica*, Eint_F and Eint_R for *E. coli*. Following PCRs, PCR products were cleaned and concentrated. Next, the insertions of the landing pad with flanking genomic homologies into the genome-of-interest were performed by dsDNA-recombineering with pORTMAGE3 plasmids. Finally, clones that successfully integrated the landing pad cassette were selected overnight based on their chloramphenicol resistance on LB[L] + 20 µg/ml chloramphenicol agar plates and integrations were sequence-verified by colony PCR with CHK_F and CHK_R primers.

Selection of drug resistance from DIvERGE libraries

To identify antibiotic resistance-causing variants, frozen cell libraries of *E. coli* K-12 MG1655 that had undergone either DIvERGE- or MP6-mutagenesis were thawed and washed three times in 1 ml Minimal Salt (MS) media. Appropriate dilutions of the 5-cycle DIvERGE population were plated onto Minimal Salt (MS) agar plates containing 4, 11, 67, and 267-times the wild-type MIC concentration of the drug. The wild-type MIC of trimethoprim is 0.75 µg/ml on MS + casamino acid agar medium for *E. coli* K-12 MG1655 and CFT073 (UPEC). To assess the single step mutational landscape of *folA* DIvERGE libraries were selected in three replicates on agar plates that contained 4-times the wild-type, agar surface-based MIC concentration of trimethoprim. Highly trimethoprim-resistant variants were selected on Minimal Salt (MS) + casamino acid agar plates containing 1000 µg/ml trimethoprim. The selection was also performed with the induced MP6 population on MS + casamino acid agar plates containing 4-times the wild-type MIC concentration of trimethoprim. Next, 800 - 1000 individual colonies were isolated for further genotype- and antibiotic susceptibility-analysis from all plates. Colonies were scraped off from agar plates in 5 ml MS medium to each plate, and from this cell-suspension, 0.5 ml was used to extract genomic DNA by using GeneElute Bacterial Genomic DNA Kit according to the manufacturer's instructions (Sigma-Aldrich).

E. coli K-12 MG1655 *gyrA, gyrB, parE, parC* libraries were selected on LB[L] agar plates containing 2-times the wild-type minimum inhibitory concentration (MIC) of ciprofloxacin (i.e., 2 ng/mL). Following incubation for three days at 30 °C, 3000 resistant clones were isolated.

Colonies were scraped off as previously, and from this cell-suspension, 0.5 ml was used to extract genomic DNA by using GeneElute Bacterial Genomic DNA Kit (Sigma-Aldrich).

Gepotidacin selection experiments were performed on Mueller Hinton II (from Sigma-Aldrich) agar plates containing 2- and 12-times the wild-type MIC concentration of gepotidacin (i.e., 140 ng/ml). Gepotidacin was obtained from MedChemExpress, China (HY-16742) and the molecular structure and purity of the drug were verified by nuclear magnetic resonance (NMR) spectroscopy. The frequency of resistant cells according to the standard frequency-of-resistance (FoR) protocol was assayed by plating 10^{10} E. coli K-12 MG1655 cells to 145 mm agar plates containing 50 ml Mueller Hinton II agar with the corresponding gepotidacin concentration. Colony counts were determined after 3 days of incubation at 30 °C. DIvERGE-generated, gepotidacin resistant variants were isolated from DIvERGE mutant libraries on Mueller Hinton II Agar (Sigma-Aldrich) plates containing 12-times the wild-type MIC concentration of gepotidacin. Gepotidacin selection experiments were performed in triplicates.

Assessing the allelic replacement efficiency of MAGE and pORTMAGE

The efficiency of ssDNA-recombineering in *E. coli, Salmonella enterica*, and *Citrobacter freundii* was assayed by high-throughput amplicon sequencing on Illumina MiSeq. High-throughput amplicon sequencing provided an unbiased and scalable method to read-out allelic composition in multiple strains without the need of a readily selectable marker gene-based allelic replacement assay (e.g., *lacZ*). To compare efficiencies of ssDNA-recombineering across strains and conditions pSIM8 containing *C. freundii* ATCC 8090 *asnA*::TET-CAT_OFF cells, pORTMAGE2 containing *E. coli* K-12 MG1655 *asnA*::TET-CAT_OFF, *S. enterica* serovar *Typhimurium asnA*::TET-CAT_OFF and pORTMAGE3 harboring *C. freundii* ATCC 8090 *asnA*::TET-CAT_OFF cells were subjected to ssDNA-recombineering according to our general ssDNA-recombineering protocol. Following a single cycle of recombineering, gDNAs were isolated from ~10^8 bacterial cells from each stationary phase culture and subjected to bulk amplicon-sequencing.

Specifically, we assessed allelic compositions at the landing pad in each strain – condition pair by the high-throughput bulk sequencing of a 436 base pair long PCR amplicon. This PCR amplicon covered the landing pad region that was targeted by oligo-recombineering. From each gDNA sample, 200 ng DNA was used as a template to amplify this target region with tetDS1 and tetDS2 PCR primers. PCRs were performed uniformly 50 μl Phusion Hot Start II High-Fidelity PCR mixture (Thermo Scientific) with Phusion High-Fidelity buffer to maximize PCR's fidelity. To increase DNA amplicon quantities, PCRs were performed in 3 × 50 μl. PCRs were cycled 26 times at 98 °C for 10 seconds, 57 °C for 30 seconds, and 72 °C for 30 seconds,

with a final extension time of 3 minutes at 72 °C. PCR products were assayed by agarose gel electrophoresis to ensure that only a single, specific PCR amplicon was generated and quantified by using a TapeStation instrument (Agilent). Next, sequencing libraries from the 436 base pair long amplicons were prepared and sequenced at SeqOmics Biotechnology Ltd. (Mórahalom, Hungary) by using NEBNext DNA Library Prep Master Mix and NEBNext Multiplex Oligos for Illumina (New England Biolabs). Sequencing libraries were prepared according to the manufacturer's instructions with minor modifications: initial fragmentation was skipped, and purifications were performed with Agencourt AMPure XP beads (Beckman Coulter). Finally, libraries were sequenced using MiSeq Reagent Kit v2 for a 250 base pair paired-end sequencing on an Illumina MiSeq sequencer.

Following sequencing, raw Illumina sequencing reads were processed by using CLC Genomics Workbench ver. 8.0.1. On the course of processing, sequencing reads were first trimmed to an error probability threshold of 5%. Next, overlapping read pairs were merged to reconstruct the 436 base pair long landing pad template. To identify mutations compared to the wild-type sequence of the landing pad, sequencing reads were then mapped to the landing pad's sequence. This mapping data was then exported from CLC Genomics Workbench and pysamstats (https://github.com/alimanfoo/pysamstats) was used to extract coverage and nucleobase composition for each DNA position of the PCR amplified landing pad region. In all samples, every nucleotide at the corresponding landing pad positions was sequenced at least 120 000-times. Finally, allelic replacement efficiencies at each targeted nucleotide position was quantified by measuring the distribution of nucleotide variants at each nucleotide position.

Library generation in *E. coli* K-12 MG1655, *Salmonella enterica serovar Typhimurium*, and *Citrobacter freundii* ATCC 8090 was assayed by targeting *asnA* with ssDNA-recombineering that carried six randomized nucleotide position. Randomization in this context meant the equimolar representation of A, C, T, and G nucleobases at a given randomized nucleotide position. To compare the ability of ssDNA-recombineering across strains and conditions, heat-induced *E. coli* K-12 MG1655 + pSIM8, *S. enterica* serovar *Typhimurium* + pSIM8, *C. freundii* ATCC 8090 + pSIM8KAN, *E. coli* K-12 MG1655 + pORTMAGE2, *S. enterica* serovar *Typhimurium* + pORTMAGE2, *C. freundii* ATCC 8090 + pORTMAGE3 cells were electroporated with the corresponding strain-specific recombineering oligos according to the general ssDNA-recombineering protocol. Following oligo integration for five recombineering cycles, gDNA from each population was extracted and amplified according to the protocol described for the landing-pad assay by using CHK_F and CHK_R primers. Next, the library composition at *asnA* in each species was assayed by Illumina high-throughput sequencing as described above.

Whole genome sequencing to assess off-target effects of ssDNA-recombineering

After 24 iterative recombineering cycles, we selected one independently edited clone from *E. coli* K-12 MG1655 + pORTMAGE, MG1655 Δ*mutS* + pSIM8, and MG1655 + pSIM8. Next, each selected clone and their corresponding parental strains (*E. coli* K-12 MG1655 wild-type and MG1655 Δ*mutS*) were inoculated into 1 ml LB^L medium and grown until stationary phase at 30 °C. Genomic DNA was isolated from 500 µl of each culture with GenElute™ Bacterial Genomic DNA kit from Sigma-Aldrich, according to the manufacturer's protocol.

To quantify off-target mutations, the genomes of these parental and the MAGE-derived clones were then whole-genome sequenced. To this aim, sequencing libraries were constructed from genomic DNA using NEBNext Fast DNA Fragmentation & Library Prep Set kit for Ion Torrent (New England Biolabs) according to manufacturer's instructions. Ion Xpress Barcode Adaptors (Life Technologies) were then ligated, and the template-fragments were size-selected by using AmPure beads (Agencourt). Adaptor-ligated fragments were then PCR amplified and cleaned by using AmPure beads and quality checked on TapeStation (Agilent). Finally, sequencing libraries were quantitated by using an Ion Library TaqMan Quantitation Kit (Life Technologies). These sequencing libraries were prepared for sequencing by using Ion OneTouch reagents according to the manufacturer's instructions (Life Technologies). Finally, template-positive sequencing beads were deposited to an Ion 318 chip and sequencing was performed with the Ion Hi-Q Sequencing Kit on an Ion Personal Genome Machine System (Life Technologies).

Following sequencing, nucleobase calling from the raw sequencing data was carried out within the Ion Torrent Suite. The tmap read mapper module of Torrent Suite was used to align sequencing reads to the *E. coli* K-12 MG1655 genome (NCBI reference sequence: U00096.3). Next, single nucleotide substitutions, small insertions, and deletions were detected compared to the reference *E. coli* genome by the Torrent Variant caller (i.e., the tvc) module. Only mutations that were represented at least on 12 sequencing reads and on at least 66% of all sequencing reads that aligned to the given reference position were voted as true mutations. Moreover, only those variants were taken into account that was supported by sequencing on both strands of the sequencing read. Mutations that were not targeted by recombineering and were detected only in the edited clones, besides being absent in the parental strains were voted as off-target mutations. For a detailed list of off-target mutations see Appendix 5.

High-throughput sequencing of soft-randomized oligos

To analyze the characteristics and mutational spectrum of soft-randomization-based oligo synthesis for DIvERGE, we synthesized 90 nucleotide long soft-randomized oligos that were

complementary to the landing pad sequence. Within these oligos, each nucleotide position was soft-randomized with up to 2% of all 3 possible mismatching nucleobases. However, as the library preparation step (i.e. the NEBNext DNA Library Prep Master Mix from New England Biolabs) of Illumina sequencing accepts only dsDNA fragments, we had to develop a PCR-free method to convert soft-randomized DNA oligonucleotides to dsDNA strands before we have been able to analyze nucleotide composition. To this aim, we made each oligo double-stranded by annealing each soft-randomized strand to their non-randomized reverse complement that was chemically synthesized under identical conditions. Therefore, we equimolarly mixed complementary oligo pairs in 50 µl of H_2O and mixed with 2.5 µl of 1 M NaCl. Samples were then heated to 95 °C for 5 minutes and slowly allowed to cool to room temperature within 120 minutes. Next, 650 ng of these annealed dsDNA oligos were 5' end phosphorylated with T4 DNA polynucleotide kinase. Ligase reactions were then cleaned with AMPure XP beads (Beckman Coulter), and the DNA was eluted in nuclease-free H_2O. Next, dsDNA oligos were dA-tailed by using a NEBNext DNA Library Prep kit. During dA tailing, end repair was excluded because we observed the additional removal of mismatching nucleobases at the end of hybridized strands that falsely generated mutation-free oligo-ends. After the purification of dA-tailed products with AMPure XP beads, 10 µls of this dA-tailed samples were ligated with sequencing adaptors. To minimize adaptor dimer formation, dA-tailed DNA:adaptor concentration was set to 1:2. Library preparation and sequencing was then performed by using NEBNext DNA Library Prep Master Mix Set for Illumina (New England Biolabs) according to the manufacturer's instructions. Finally, the sequencing of at least 10^5 oligonucleotides from each sample was achieved by using a MiSeq Reagent Kit v2 for paired-end sequencing on an Illumina MiSeq sequencer. However, it should be noted that this sequencing-library preparation protocol resulted in the removal of mismatching nucleobases from the last terminal 2 positions at both the 3' and 5' ends, due to the lack of adaptor-ligation if mismatches at those positions destabilized the ends of dsDNA oligonucleotides. Therefore these positions were excluded from further library-composition analyses.

Illumina sequencing-based analysis of *folA* and landing pad libraries

To quantify the distribution and spectrum of mutation at *folA* and at the landing pad in DIvERGE experiments, we developed a PCR amplicon deep sequencing-based mutational assay. This assay, in conjunction with subsequent Illumina sequencing and strict sequencing noise removal on the course of sequence data processing, allowed us to precisely identify mutations at the target site.

Our protocol relied on the high preciosity PCR amplification of the target locus. In order to minimize PCR mutagenesis during amplicon preparation, previously isolated gDNA

samples were subjected to the minimum number of PCR that generated enough amplicon template for sequencing reactions (i.e., between 200 – 500 ng). This was usually achievable by 18 – 20 subsequent cycles of PCR. To achieve the possible lowest error rate, PCRs were performed in 50 ul of Phusion High-Fidelity PCR Master Mix (Thermo Fisher Scientific) using the corresponding primer pairs. Next, PCR amplicons that contained the entire target locus were digested using NEBNext dsDNA Fragmentase (New England Biolabs) for 12 - 15 minutes to yield approximately 190 base pair fragments. Digested amplicons were then purified, and sequencing libraries were prepared as described previously for landing pad libraries. Finally, sequencing was done by using MiSeq v2 reagent kit for 250 base pair paired-end sequencing run on a MiSeq Illumina sequencer.

Fragmentation allowed us to decrease the error rate on the course of Paired-end Illumina MiSeq sequencing: Prior analyses demonstrated that the fidelity of Illumina MiSeq sequencing substantially decreases after the 3' 200th nucleotide in the sequencing read. Thereby, limiting the length of sequencing reads below 200 nucleobases increased the probability of correct base-calling [135]. To this aim, on the course of bioinformatic analyses of raw sequence data, based above the 200th nucleobases in each sequencing read were also excluded from analyses.

Assessing mutational profiles with Pacific Biosciences Single Molecule Real-Time sequencing

Illumina sequencing, however, is not suitable to precisely analyze the allelic composition and the combination of distant mutations at loci that are longer than ~400 base pairs. This limitation arises due to the inherent error rate of sequencing and the mediocre read lengths of current Illumina sequencing methods [135,136]. Currently, only single molecule long read sequencing methods, i.e., Pacific Biosciences Single Molecule Real-Time (SMRT)-, Oxford Nanopore, and 10X Genomics synthetic long-read sequencing are suitable to accurately assess the genotype of DNA strands that are longer than 600 nucleotides [137–139]. Therefore and based on its superior error rate, the availability of multiplex PCR amplicon sequencing methods, and the ease of sequencing library preparation, we relied on Pacific Biosciences single molecule real-time (SMRT) sequencing to assess the genotypic composition of *folA*, DNA gyrase, and topoisomerase IV variant libraries. To access Pacific Bioscences sequencing service, we built up a fee-for-service collaboration with the Norwegian Sequencing Centre (NSC UiO), a national sequencing technology platform hosted by the University of Oslo in Norway.

Specifically, to assess the genotypic composition of genomic libraries at *folA, gyrA, gyrB, parE*, and *parC* in multiple species we relied on Pacific Biosciences RSII Single Molecule

Real-Time (SMRT) circular-consensus amplicon sequencing. In order to minimize PCR-induced mutagenesis during sequencing library preparation, previously isolated gDNA samples (200 ng) served as a template for Phusion High-Fidelity PCR with the corresponding species and sample specific barcoded primer pairs (Appendix 3). These barcodes allowed us to pool and sequence up-to 80 distinct sequencing sample in a single sequencing reaction. PCR reactions were performed in 50 µl reaction volumes according to the following parameters: 98 °C initial denaturation for 3 min, 18-22 cycles of (98 °C 20 seconds; 63 °C 0.5 minutes; 72 °C 1.5 minutes), and final extension for 5 minutes at 72 °C. On the course of PCR amplicon preparation, to minimize overamplification that can cause biases in library-composition, and to avoid amplicon-chimera formation, PCR reactions were stopped at the mid-exponential phase of amplification. Finally, PCR amplicons were purified by using a Zymo DNA Clean and Concentrator kit (Zymo Research) and eluted in 30 µl 0.5× TE buffer. PCR amplicons were then mixed at an equimolar ratio and shipped to the Norwegian Sequencing Centre on dry ice. The preparation of sequencing libraries (the ligation of SMRTBell adapters) and their sequencing on Pacific Biosciences RSII SMRT cells was performed by the Norwegian Sequencing Centre (Norway) on a Pacific Biosciences RSII sequencer.

Mutational analysis of landing pad libraries and DIvERGE oligo pools

The analysis of Illumina MiSeq sequencing data was performed in collaboration with Balázs Bálint, Bálint Márk Vásárhelyi, and István Nagy (SeqOmics Biotechnology Ltd). Specifically, sequencing data were analyzed by using a bioinformatics pipeline that we developed to increase sequencing-accuracy and reduce sequencing noise. Briefly, to remove sequencing read-ends that have a higher error rate, paired-end Illumina MiSeq reads were first trimmed to 190 nucleotides. Next, trimmed reads were further trimmed based on sequencing quality and all nucleotides that had an error probability that was higher than 0.1% were excluded from follow-up analysis. Next, overlapping paired-end sequencing read pairs were merged into a single read by using CLC Genomics Workbench 9.0. Following read pre-processing, sequencing data analysis was carried out based on the sample-source of the data. When analyzing landing pad libraries following DIvERGE-mutagenesis, sequencing reads were aligned to the wild-type sequence of the landing pad by using CLC Genomics Workbench 9.0. When we assessed the sequence composition of soft-randomized DIvERGE oligos and their corresponding genomic targets after genomic integration, BWA-MEM was applied. Following alignment, it was necessary to filter out erroneous sequencing reads and alignment-artifacts. To this aim, erroneous alignments were removed with SAMtools version 0.1.19-96b5f2294a and NGSUtils version 0.5.7-e98ddfa. Next, sequencing reads that do not span the entire

TETRM oligo-target region were removed with JVarkit. Finally, the number of nucleobase alterations between mapped sequencing reads and their target region were reported with BLASTn and alterations, as compared to the wild-type target sequence, were summarized with Pysamstats version 0.24.2. In all cases, mutation frequency at every nucleotide position was calculated as the ratio of reads that contained substitutions, insertions, or deletions at the given nucleotide position. Diversified target positions within the landing pad were defined as the nucleotide positions where mutation frequency exceeded 6-times the standard deviation of the background sequencing noise. Sequencing noise for this purpose was measured in each sequencing sample at an untargeted region within the landing pad, between the target site of TETRM1 and TETRM3 oligos.

Mutation composition analysis of *folA* libraries based on Illumina sequencing

To identify mutations and assess the nucleotide composition of *folA* libraries from *E. coli* K-12 MG1655 and *E. coli* CFT073, we have developed a custom Python program-package with built-in filters for Illumina sequencing error reduction. Similarly to the previously described landing pad libraries, Illumina sequencing reads were first trimmed to 190 nucleotides to remove erroneous bases. Next, these trimmed reads were further trimmed based on sequencing quality. All nucleotides that had an error probability that was higher than 1‰ were excluded from further analysis and removed with BBduk. Next, the resulted overlapping sequencing read-pairs were merged, and sequencing reads that contained any ambiguous nucleotide (i.e., N) and reads that were shorter than 72 nucleobases were also removed. These filtered sequencing reads were subsequently mapped with BWA-MEM to their corresponding genomic targets. Finally, Pysamstats version 0.24.2 was used to measure mutation frequency and to generate a nucleotide composition table for each targeted reference position. The single-step mutational landscape of *folA* from scanning DIvERGE libraries was determined based on sequencing reads that displayed exactly and only one nucleobase substitution compared to their genomic target sequence. These reads were translated to amino acid sequences. These peptides were then compared to their corresponding coding sequence to analyze amino acid composition for each reference position. This final program was made publicly available and is now accessible at http://group.szbk.u-szeged.hu/sysbiol/EvGEn/diverge-2018-script.html under the name DIvERGE_Illumina_script.zip. All Illumina and Pacific Biosciences DNA sequencing data is available in Appendix 6.

Assessing mutation profiles in *folA* libraries based on Pacific Biosciences sequencing

Following sequencing, raw Pacific Biosciences RSII sequencing data was processed by the Norwegian Sequencing Centre to demultiplex samples based on their unique, symmetric 16 base pair long barcode sequence that we attached on the course of library. Demultiplexing and circular consensus read generation were performed by SMRT Analysis 2.3 from Pacific Biosciences. Next, circular consensus sequencing reads were imported to CLC Genomics Workbench 9.0 and mapped to their corresponding genomic target region. Reads with an ambiguous nucleotide and reads that were shorter than 80% of their wild-type genomic target or had less than 90% sequence identity to their target were discarded. This ensured the removal of erroneous sequencing reads. On the course of alignment, reads were individually mapped to their target sequence either on *Escherichia coli* K-12 MG1655 (NCBI Reference Sequence: NC_000913.3), or on *Escherichia coli* CFT073 (NCBI Reference Sequence: NC_004431.1), or on the chromosome of *Salmonella enterica* LT2 (NCBI Reference Sequence: NC_003197.1), according to their origin. Finally, nucleotide substitutions in each sequencing read were determined simultaneously with any associated amino acid alterations compared to the reference sequence. On the course of variant analysis, only mutations that had an error probability that was lower than 1% were considered. DNA gyrase allele frequency values were plotted on the crystal structure of topoisomerase protein complex from *Mycobacterium tuberculosis* (Protein Data Bank identifier: 5BS8).

Genomic integration of defined mutations

Defied mutations within the chromosomes of *E. coli* K-12 MG1655, *E. coli* UPEC CFT073, *Citrobacter freundii* ATCC 8090, and *Klebsiella pneumoniae* ATCC 10031 were constructed within the wild-type parental strains by using pORTMAGE3 recombineering according to our previously described ssDNA-recombineering protocol (see *ssDNA-recombineering protocol,* and reference [130]). To construct defined mutations, ssDNA oligonucleotides that carried the mutation-of-interest were designed and synthesized to target the replicating lagging-strand of *gyrA, parC or folA* in the corresponding bacterial strain. The corresponding oligonucleotides for mutant reconstructions and PCR-based allele confirmations are listed in Appendix 3. Mutants were generated by performing a single ssDNA-recombineering cycle with pORTMAGE3 according to the general ssDNA-recombineering protocol. Mutant clones were then plated to agar plates, and the presence of the corresponding mutations were confirmed by colony-PCR and subsequent capillary sequencing of the oligo-target region.

In vitro growth rate measurements

We measured bacterial fitness as growth rates in a rich bacterial medium (LB^L) under aerobic conditions. To investigate interferences between mutagenesis and cell viability, we measured 30, randomly chosen *E. coli* K-12 MG1655 wild-type, MP6-mutagenized *E. coli* K-12 MG1655, and DIvERGE-mutagenized *E. coli* K-12 MG1655 trimethoprim-resistant isolates in LB^L medium. Cultures of these mutants were cultured in LB^L broth and incubated at 30 °C until they reached stationary phase. Then we transferred approximately 1000 cells from each mutant into 96-well plates, containing 0.1 ml LB^L medium. Following inoculation, cells were grown at 30 °C in a Powerwave XS2 (Biotek) automated microplate spectrophotometer under continuous agitation. Growth curves were recorded by measuring the optical density of the cultures (OD$_{600\,nm}$) at every 7 minutes for 24 hours. Finally, growth rates were calculated from the obtained growth curves [140,141].

Antibiotic drug susceptibility measurements

To select variant libraries and assess the susceptibility of bacterial strains under selective conditions, we measured antibiotic susceptibilities under different growth conditions. Minimal Salt + casamino acid agar-surface-based minimum inhibitory concentrations (MICs) of trimethoprim for *E. coli* K-12 MG1655, *E. coli* CFT073, and *Salmonella enterica* LT2 were assessed by using E-test strips according to the manufacturer's protocol (bioMerieux).

Trimethoprim susceptibility was determined in Minimal Salt media + casamino acid (without thiamine). Trimethoprim resistance of individually isolated bacterial strains were quantified as the 75 % inhibitory concentration of trimethoprim (IC75) in Minimal Salt + casamino acid broth. Specifically, the IC75 value of the given bacterial isolate was calculated as the trimethoprim concentration at which the area under the growth curve of the given cell population was equal to 25% of a control cell population of the same strain that was grown without antibiotic. To obtain growth curves, bacterial isolates were grown in the presence of an increasing trimethoprim concentration gradient, according to our general protocol for in vitro growth rate measurements. All IC75 measurements were performed in triplicates.

To assess the maximal IC75 values within pooled mutant libraries, we performed a competition-based antibiotic susceptibility measurement. Compared to the drug susceptibility testing of individual genotypes, competition experiments enriched the least drug-susceptible variant from a library of different genotypes and assessed its IC75 value. To perform competition, 100 – 150 distinct bacterial colonies from the given, trimethoprim-selected library

were scraped off, and the IC75 values were determined for approximately one million cells from these pooled cell populations. Competition experiments were performed in triplicates.

Ciprofloxacin and gepotidacin MICs for wild-type *E. coli* K-12 MG1655, *Citrobacter freundii* ATCC 8090, *Klebsiella pneumoniae* ATCC 10031, and their corresponding isogenic mutants were determined by a microdilution-based MIC assay. Microdilution-based MIC assays were performed in triplicates in 96-well plates according to the general EUCAST protocol at 30 °C [142]. Ciprofloxacin susceptibilities were determined in LB^L broth, while gepotidacin MICs were assayed in cation-adjusted Mueller Hinton II broth (Sigma-Aldrich) following 18 hours of incubation.

Mutation rate measurements

We measured the phenotypic effect of the plasmid-based expression of E. coli MutL E32→K (*ecmutL* E32→K) on the mutation rates of *E. coli* K-12 MG1655, *Salmonella enterica* serovar *Typhimurium* LT2, *Citrobacter freundii* ATCC 8090, *Edwardsiella tarda* ATCC 15947, and *Escherichia hermannii* HNCMB 35034 in a rifampicin frequency-of-resistance and subsequent fluctuation assay. First, strains, harboring the anhydrotetracycline-inducible pZA31tetR-mutLE32K plasmid for *ec*MutL E32→K overexpression were inoculated into 1 ml 20 µg/ml chloramphenicol-containing LB^L broth. Strains were grown separately either in the presence or in the absence of 100 ng/ml anhydrotetracycline that induced the overexpression of *ec*MutL E32→K. After overnight growth at 30 °C, 10^4 cells from each of these starter cultures were transferred into 10 separate tubes that contained 1 ml chloramphenicol-supplemented LB^L broth with or without 100 ng/ml anhydrotetracycline, respectively. Next, these individual cultures were allowed to reach stationary phase by incubating them at 30 °C under constant agitation. Diluted samples from these cultures were then spread onto both LB^L agar plates and LB^L agar plates containing 100 µg/ml rifampicin. Agar plates were then incubated at 37 °C, and the number of colonies were determined after 1 day of incubation. These mutation frequency assays were performed in duplicates. Finally, mutation rate was calculated from the frequency of rifampicin resistant colonies in each culture by using the Ma-Sandri-Sarkar maximum-likelihood method within the FALCOR analysis tool [143] available at http://www.mitochondria.org/protocols/FALCOR.html.

Results

Characterization of a dominant negative allele of *E. coli* MutL

As the first step, we have characterized the phenotypic effect of a previously described allele of *E. coli MutL* (*ec*MutL E32→K) by Aronshtam A et. al. [144]. When expressed from a plasmid, this variant is known to induce a mutator phenotype even in the presence of the wild-type methyl-directed mismatch repair (MMR) system of *E. coli*. Therefore, by using standard cloning techniques, we have reconstructed and cloned this mutant allele into the anhydrotetracycline-inducible pZA31tetR expression vector. Next, we have measured the mutation rates in wild-type *E. coli* K-12 MG1655 cells with and without the induction of *ec*MutL E32→K overexpression. The induced overexpression of *ec*MutL E32→K resulted in an over 30-fold increase in mutation rate in wild-type *E. coli* cells as measured by rifampicin frequency-of-resistance assay and subsequent fluctuation tests [143] (Figure 4). Also, we have demonstrated that in the presence of *ec*MutL E32→K, the mutation rate approached that of an *E. coli* MG1655 strain lacking the functional MMR machinery (i.e. a Δ*mutS* mutant) [130,145]. Moreover, these findings indicate that the wild-type genomic copy of *mutL* is not able to suppress the effect of the dominant negative variant. Based on these results, we have concluded that the controlled overexpression of *ec*MutL E32→K would enable the controlled on-off switch of MMR repair, fully accomplished by a plasmid-based expression system and without the inactivation of the genomic copy of *mutL*.
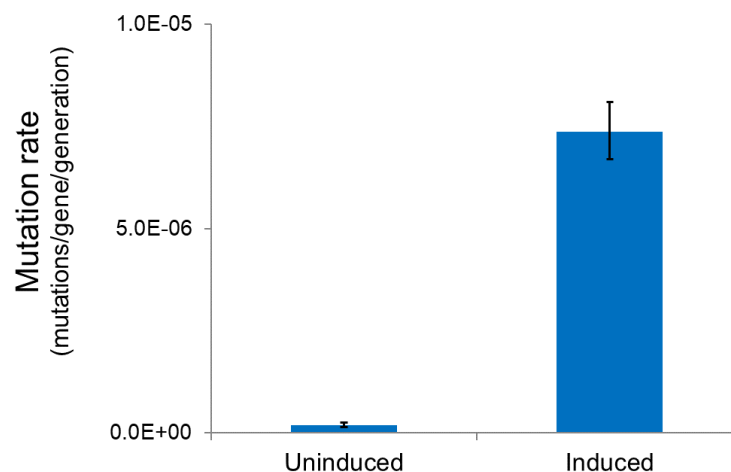
Figure 4. Mutation rate of *E. coli* K-12 MG1655 harboring the anhydrotetracycline-inducible pZA31tetR-mutLE32K plasmid which overexpresses *ec*MutL E32→K, a dominant negative mutator allele of *E. coli* MutL. Mutation rates were assessed by the

rifampicin frequency-of-resistance assay and subsequent fluctuation tests. Error bars represent 95% confidence intervals based on 20 replicates for each sample.

Construction of a plasmid-based MAGE system with mismatch repair control

We hypothesized that the plasmid-based overproduction of the dominant negative *ec*MutL E32→K allele would allow a transient and controllable switch from a non-mutator to a mutator phenotype. Moreover, as the effect of the *ec*MutL E32→K allele cannot be suppressed by the native wild-type copy of *mutL*, thus no disruption of the genomic allele would be required.

The traditional MAGE method uses a temperature-regulated expression system which is controlled by the temperature-inducible *c*I857 repressor–pL promoter system [34]. This expression platform enables an exceptionally rapid and high-level overexpression of λ Red proteins. Based on this desirable feature, we hypothesized that by constructing a synthetic bacterial operon that encodes and co-expresses *ec*MutL E32→K and the necessary genes of λ recombineering (*exo, bet, gam*), the construction of a plasmid-based MAGE system would be feasible.

To test the functionality of co-expression, we have constructed a synthetic operon that contained all three genes of λ recombineering and *mutL* E32K on a broad host-range vector. This plasmid included the pBBR1 broad host-range origin-of-replication, isolated originally from *Bordetella bronchiseptica* [146], and finally resulting in the development of pORTMAGE plasmids (Figure 5). In this synthetic operon the expression of *E. coli mutL* E32K, as well as *exo, bet*, and *gam* were under the control of the *c*I857 temperature-sensitive λ repressor-regulated pL promoter.
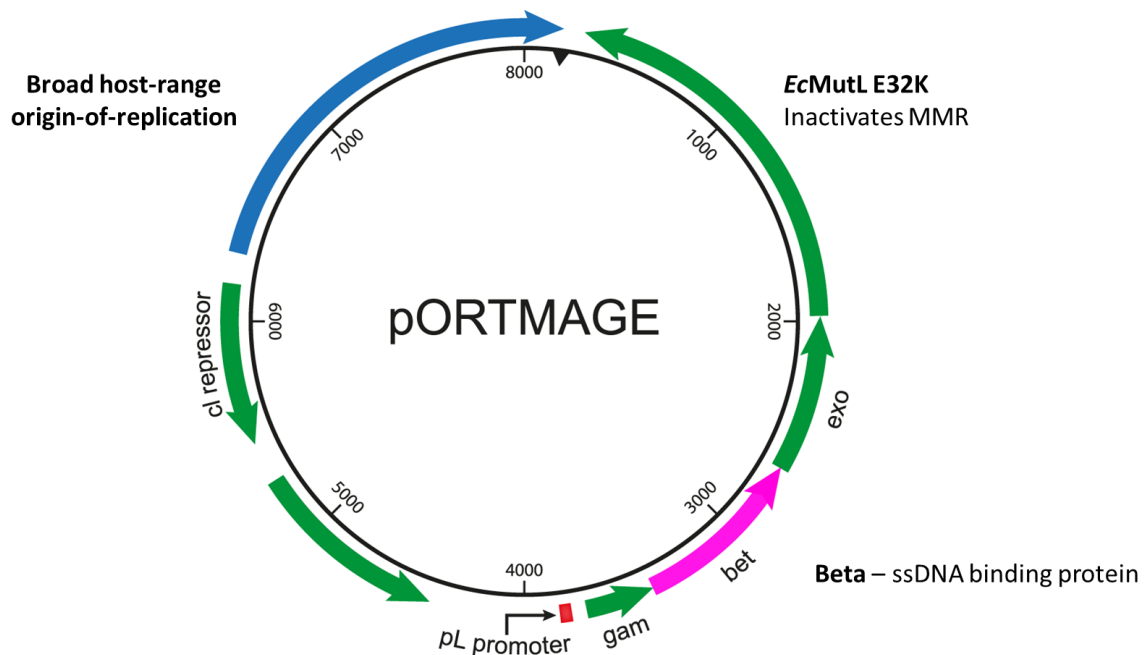
Figure 5. General structure of the pORTMAGE plasmids. Expression of *ec*MutL E32→K in conjunction with the three λ recombinase genes (*exo, bet,* and *gam*) is controlled by the *c*I857 temperature-sensitive λ repressor regulated pL promoter-based induction system.

Next, to investigate the effect of the expression of the dominant MutL allele on the efficiency of ssDNA-recombineering, we have employed a previously characterized test system in *E. coli* K-12 MG1655 [62,147]. This system relies on the genomic introduction of a diverse set of single nucleotide mismatches (A:G, T:T, A:A, G:A, C:T, G:T, G:G, and C:A, according to the given chromosomal to synthetic nucleobase mismatch) into the genomic copy of *lacZ* at specific locations. In turn, the incorporation of these mutations induces a premature stop codon. Thus, these mutations result in the premature termination of *LacZ*, so the frequency of allelic replacements can be detected easily by a colorimetric assay that visualizes the enzymatic function of LacZ. LacZ (beta-galactosidase) hydrolyzes the terminal non-reducing beta-D-galactose residues into beta-D-galactosides.

Based on single pORTMAGE cycles in *E. coli* K-12 MG1655 performed individually with each *lacZ*-targeting oligo, we have found that in all cases, pORTMAGE allowed of a highly efficient oligo incorporation, while the plasmid lacking the expression of *ec*MutL E32→K (pSIM8) produced highly biased oligo-incorporation. Moreover, the efficiency of pORTMAGE to incorporate single-nucleotide mismatches was comparable to the efficiency of the traditional MAGE protocol [38,62] (Figure 6).
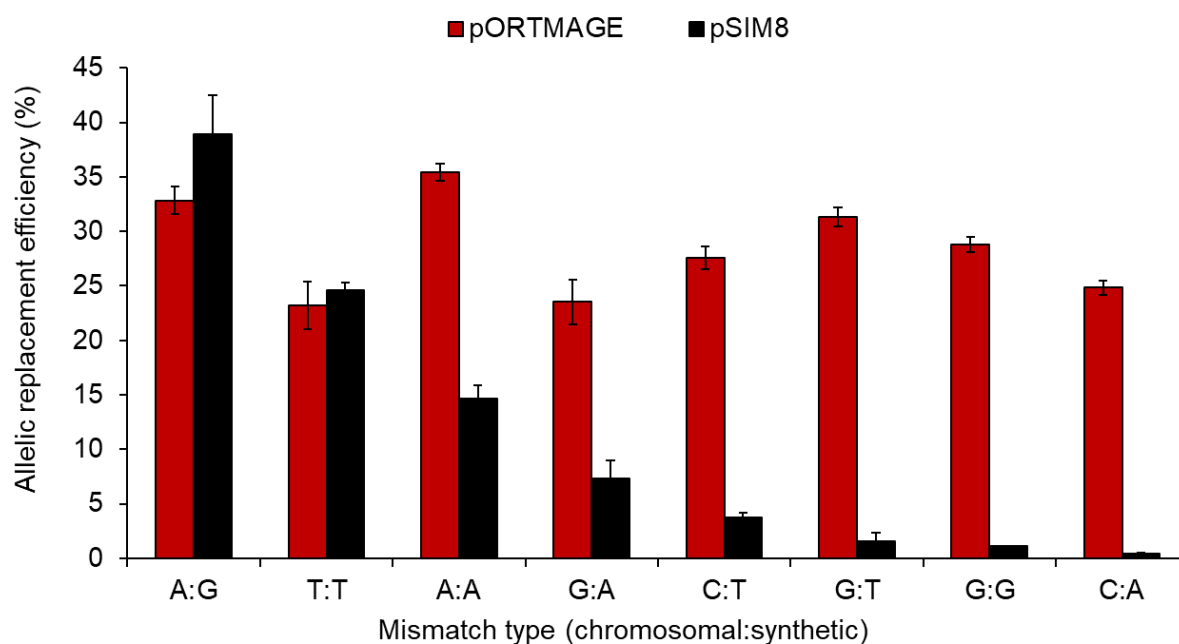
Figure 6. Allelic replacement efficiencies of oligos inducing various types of single nucleotide mismatches in the chromosome of *E. coli* K-12 MG1655 in the presence of the wild-type mismatch repair machinery (pSIM8, a plasmid that solely expresses λ Red recombinases) and *ec*MutL E32→K-controlled MMR from pORTMAGE. The frequency of allelic replacement was estimated as the number of LacZ-inactivated cells per the total number of cells on MacConkey agar plates. The values are the means of two independent measurements. Error bars represent the standard error of the mean.

Genome engineering with pORTMAGE avoids off-target mutagenesis

In each MAGE cycle, the expression of λ Red recombinases is induced by a single temporal temperature shift lasting for 15 minutes [38]. In each genome engineering cycle, decreased MMR activity is only required during the period of oligo annealing and incorporation. Therefore, by coupling λ recombinase and *ec*MutL E32→K expression on pORTMAGE to rapidly switch the cell's phenotype between mutator and non-mutator states immediately before oligo incorporation, we hypothesized that the application of pORTMAGE would minimize the time when bacterial cells display a mutator phenotype. In turn, the use of pORTMAGE would lower the time-frame when engineered bacteria are susceptible to the accumulation of off-target mutations. This would be highly advantageous compared to the traditional MAGE procedure which necessitates a permanently inactivated mismatch repair.

To investigate the performance of pORTMAGE and off-target mutagenesis simultaneously on the course of iterative MAGE cycles, we have carried out a long-term multiplex genome editing project. Three strains, and in turn, three distinct genome engineering

approaches were compared: (I) an *E. coli* K-12 MG1655 ΔMutS mutator strain carrying the control pSIM8 plasmid which solely expressed the λ Red recombinases [57] (representing the traditional MAGE method); (II) the wild-type *E. coli* K-12 MG1655 strain carrying pORTMAGE, and (III) the wild-type *E. coli* K-12 MG1655 strain carrying a control pSIM8 plasmid.

To investigate the long-term accumulation of off-target mutations, we iteratively performed 24 consecutive cycles of MAGE while targeting six different, widely distributed loci across the genome of *E. coli* K-12 MG1655. These loci were individually targeted for 4 consecutive MAGE cycles by oligos which introduced a specific type of mismatch (Table 1).

Gene	Genomic position	Mismatch	Corresponding oligo
araB	69999	A:A	araB_AA
lacZ	364878	T:T	LacZ_TT_v7
hisB	2091657	G:T	hisB_GT
rpsL	3472447	A:C	rpsL_AC
malK	4245058	C:C	MalK_CC_v1
cycA	4428025–4428026	AA:AC	cycA_AAAC

Table 1. Genomic positions of the six marker genes targeted for recombineering on the chromosome of *E. coli* K-12 MG1655 (NCBI sequence identifier: NC_000913.3). The introduced modifications are marked as nucleotide mismatches, in chromosomal to synthetic order. Table adapted from Nyerges, Á. et al. Conditional DNA repair mutants enable highly precise genome engineering. Nucleic Acids Research 42, e62–e62 (2014) [67].

Following 4 cycles of ssDNA recombineering targeting each locus, allelic replacement efficiencies were determined at all loci-of-interest either by colorimetric MacConkey agar-assays or allele-specific PCRs. Finally, as a confirmation, the clones carrying the desired alteration were also verified by capillary sequencing.

As expected, in *E. coli* K-12 MG1655 cell with λ Red recombinase expression only (i.e. pSIM8) and without the inactivation of MMR, the allelic replacement efficiency was very low in most cases. *LacZ* A652→T and *malK* C252→G are being exceptions as the corresponding mutations are poorly recognized by the cells endogenous methyl-directed mismatch repair, and thereby their incorporation is not hindered by the presence of the native mismatch repair machinery. On the contrary to what we observed in *E. coli* MG1655 cells + pSIM8, *E. coli* K-12 MG1655 with pORTMAGE-based recombineering generally displayed highly efficient allelic replacement, approaching the efficiency observed with traditional MAGE which uses permanent MMR inactivation (*E. coli* K-12 MG1655 ΔmutS + pSIM8) (Table 2).

	Allelic replacement efficiency (%)					
E. coli strain	*lacZ* A652T	*malK* C252G	*araB* T50A	*hisB* C166T	*rpsL* A128G	*cycA* AA139TG
MG1655 + pORTMAGE	54.58	61.56	39.76	22.92	33.76	22.92
MG1655 Δ*mutS* + pSIM8	51.32	60.23	62.87	39.58	38.91	41.67
MG1655 + pSIM8	45.31	51.82	1.56	< 0.1	0.72	1.04

Table 2. Allelic replacement efficiencies in *E. coli* K-12 MG1655 + pSIM8, MG1655 ΔmutS + pSIM8, and MG1655 + pORTMAGE after four consecutive ssDNA-recombineering cycles targeting each given locus. Allelic replacement efficiencies represents the ratio of cells (in %) carrying the corresponding mutations within the entire cell population that underwent mutagenesis. Allelic replacement efficiencies were determined at all loci either by colorimetric MacConkey agar-assays or allele-specific PCRs (see Methods).

Next, we have investigated the accumulation of off-target mutations. After 24 iterative recombineering cycles, we have selected one independently edited clone from *E. coli* K-12 MG1655 + pORTMAGE, MG1655 ΔmutS + pSIM8, and MG1655 + pSIM8, respectively. To quantify off-target mutagenesis, the whole genomes of the parental cell and the MAGE-derived clones were sequenced. Sequence analysis revealed that *E. coli* K-12 MG1655 + pSIM8, without mismatch repair inactivation, had accumulated only two off-target mutations. In contrast, *E. coli* K-12 MG1655 ΔmutS + pSIM8, the strain that was engineered according to the traditional MAGE protocol, had mutated at 84 non-targeted positions. This observation is in line with previous studies [52,65]. Remarkably, we have found no off-target mutations in wild-type *E. coli* K-12 MG1655 which had been engineered with pORTMAGE.

In summary, these results have demonstrated that pORTMAGE facilitates highly efficient allelic replacement, coupled with a remarkably reduced rate of off-target mutations.

pORTMAGE allows rapid genome editing in a range of bacterial species

Phylogenetic comparison of *MutL* sequences indicate that glutamic acid (E) at the 32nd position of *Ec*MutL is conserved in a wide range of species, ranging from *MutL* sequences in *Proteobacteria* to the homologous MLH1 in *Saccharomyces cerevisiae* [130]. We therefore

assumed that the dominant mutation E32→K at this amino acid residue could have a similar phenotypic effect in a broad range of bacterial species.

To investigate this concept in details, we have tested the impact of the dominant *Ec*MutL E32→K allele on mutation rates in several enterobacterial species. We have selected the human pathogen *Salmonella enterica serovar. Typhimurium*, the fish pathogen *Edwardsiella tarda*, the opportunistic pathogen *Escherichia hermanii*, *Citrobacter freundii*, and the biotechnologically relevant production-host *E. coli* BL21(DE3) as target organisms. In agreement with our hypothesis, the overexpression of *E. coli MutL* E32→K from the anhydrotetracycline-inducible plasmid pZA31tetR-mutLE32K largely increased mutation rates in all species (see Appendix 4). Also, the level of increase in mutation rate was comparable to the mismatch repair deficient variants of these strains. These results suggest that the co-expression of the dominant *MutL* allele from *E. coli* and the λ Red recombinases from pORTMAGE would have a similar effect in all phylogenetically related strains.

Therefore, to further test this assumption, we have compared the allelic replacement efficacy of pORTMAGE in *E. coli* K-12 MG1655 and its relatives*, Salmonella enterica* and *Citrobacter freundii* which diverged from *E. coli* approximately 100 - 200 million years ago [148,149]. Also, to broaden the potential applications of pORTMAGE, we have engineered three modified pORTMAGE plasmids with different antibiotic resistance markers (termed pORTMAGE3 and 4). To characterize the performance of pORTMAGE uniformly across these target species, we have constructed a landing pad sequence and have integrated it into the aspartate-ammonia ligase gene (*asnA*) of the host genome. Next, we have utilized this genomically integrated landing pad as the target sequence for recombineering (Figure 7). The application of this identical landing pad sequence at a fixed position on the bacterial chromosome nearby the origin of replication (*oriC*) has allowed us to avoid sequence-context-specific effects in recombineering. To perform recombineering at the landing pad, we have designed five, 90-nucleotide-long mutagenizing oligos that introduced all possible single nucleobase mismatches at five different positions within the landing pad.

```
G*T*AATAACGCTAAAAGTTTTAGATGTGCTTTNCTAAGTCATCGCGATGGAGCAAAAGTACATTTAGGTACACGGCCTACAGAAAAACA*G*T
G*T*AATAACGCTAAAAGTTTTAGATGTGCTTTACTAANTCATCGCGATGGAGCAAAAGTACATTTAGGTACACGGCCTACAGAAAAACA*G*T
G*T*AATAACGCTAAAAGTTTTAGATGTGCTTTACTAAGTCATCGNGATGGAGCAAAAGTACATTTAGGTACACGGCCTACAGAAAAACA*G*T
G*T*AATAACGCTAAAAGTTTTAGATGTGCTTTACTAAGTCATCGCGATGGAGCAAAAGNACATTTAGGTACACGGCCTACAGAAAAACA*G*T
G*T*AATAACGCTAAAAGTTTTAGATGTGCTTTACTAAGTCATCGCGATGGANCAAAAGTACATTTAGGTACACGGCCTACAGAAAAACA*G*T
```

Landing Pad   *cat* **(chloramphenicol acetyltransferase)**

                                        Promoter

Region subjected to deep-sequencing

Figure 7. General map of the landing pad sequence inserted into *E. coli* K-12 MG1655, *S. enterica* and *C. freundii*. The green region represents the target sequence for allelic replacement treated with a set of five oligos shown in the targeting box. For each of the five genome editing oligonucleotides, degenerate bases (Ns) are shown in red. The *cat* (chloramphenicol acetyltransferase) gene that confers resistance to chloramphenicol has allowed us to introduce the landing pad into the genome via double-stranded DNA-recombineering.

Next, allelic replacement efficiencies were measured by performing a single ssDNA-recombineering cycle using these 5 oligos admixed into a single pool. As in our previous experiments, we have compared the efficiencies characteristic of (I) the wild-type strain carrying the pSIM8 plasmid, (II) wild-type mismatch repair proficient cells carrying a pORTMAGE plasmid, and (III) the ΔmutS mutator strain carrying only pSIM8. To accurately measure the efficiency of oligo integration into the landing pad within the resulting cell library, we have developed an Illumina MiSeq-based deep-sequencing method that precisely assesses allelic composition within the bacterial population.

Allelic replacement efficiencies in the wild-type cells carrying only pSIM8 for λ Red recombinase-expression was found to vary substantially across *E. coli, S. enterica*, and *C. freundii* (Figure 8A-C). These results, in turn, suggest strain-specific variations in mismatch repair. Also, mutations introducing the same mismatch at different genomic positions frequently showed differences in their integration efficiency, indicating the DNA sequence context dependency of mismatch repair.

In the other hand, however, the allelic replacement executed by pORTMAGE was in certain cases nearly 100-fold more efficient and largely unbiased in all three species compared to the wild-type control (Figure 8A-C). In *E. coli* K-12 MG1655, allelic replacement efficiencies with pORTMAGE approached the efficiencies obtained in the corresponding ΔmutS variant. In *C. freundii*, pORTMAGE showed a similarly robust performance as in *E. coli* and *S. enterica* (Figure 8C).

A

B

C

Figure displays legend: ■ Citrobacter freundii + pSIM8 ■ Citrobacter freundii + pORTMAGE
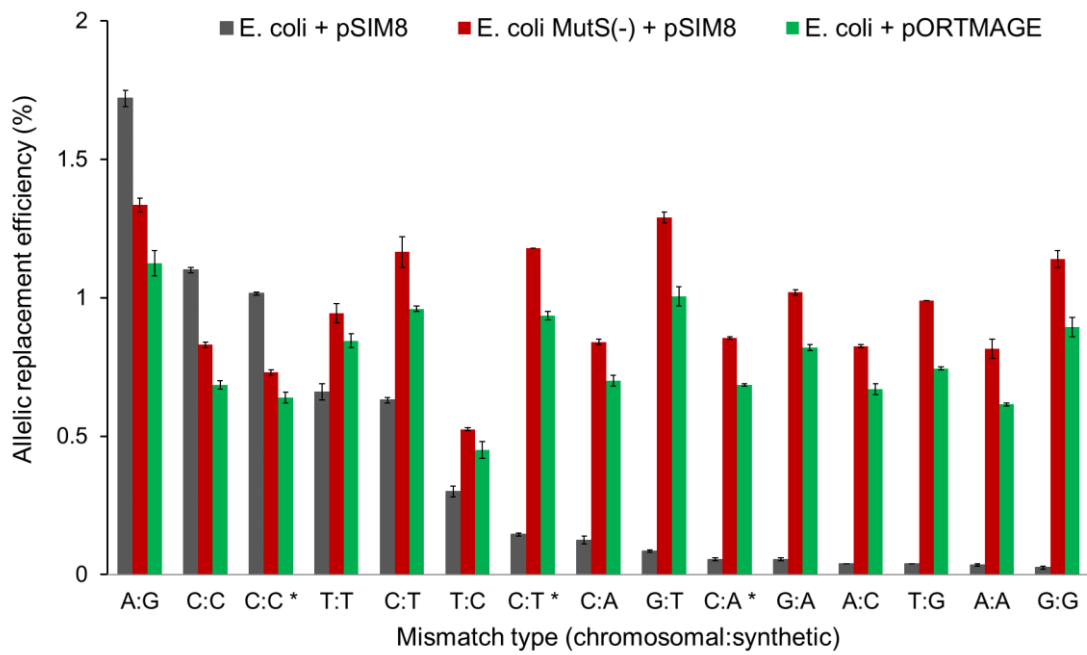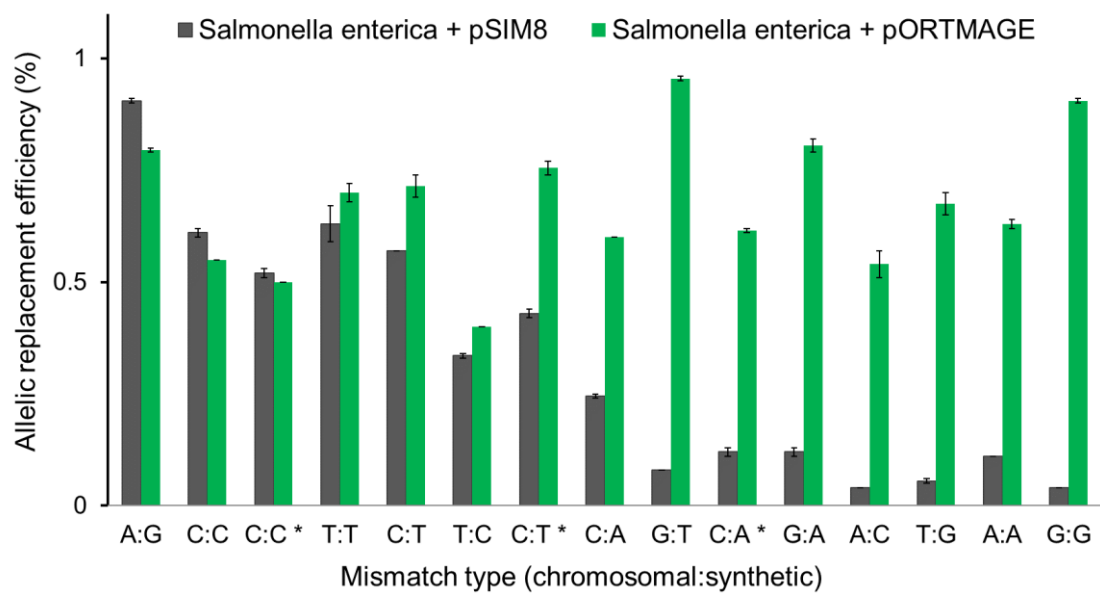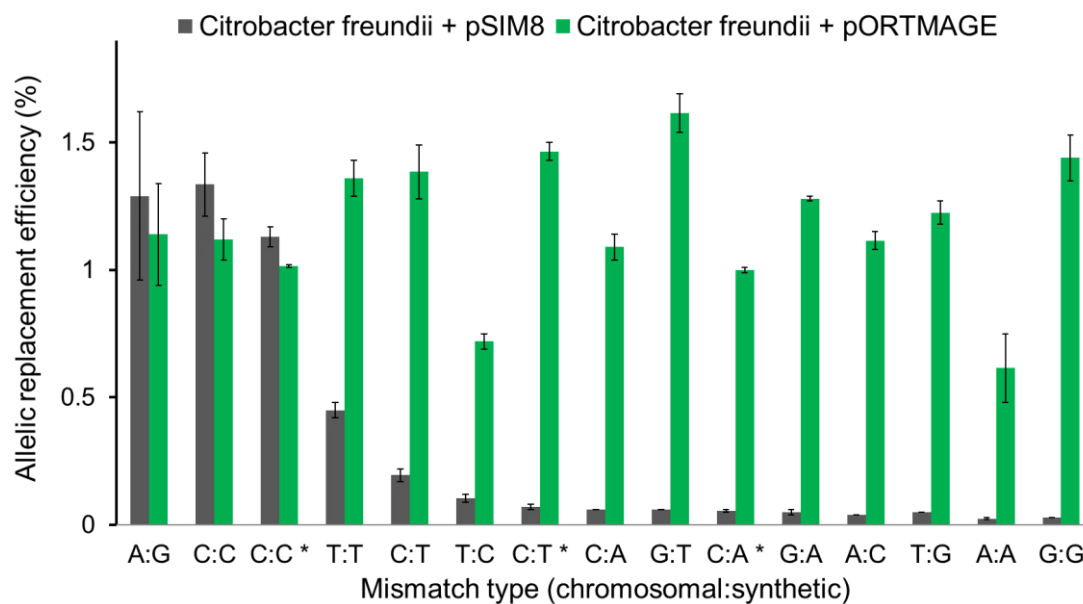
Figure 8. Allelic replacement efficiencies at the landing pad in (A) *E. coli* **K-12 MG1655, (B)** *Salmonella enterica* **serovar Typhimurium LT2, and (C)** *Citrobacter freundii* **ATCC 8090**. Figure displays the means of the results of two independent Illumina deep-sequencing assays. Error bars represent the standard error of the mean (n = 2). Star (*) denotes oligos generating the same mismatch as another oligo to demonstrate context dependency of allelic replacement. MutS(-) denotes strain carrying a deletion of *mutS*.

pORTMAGE efficiently generates mutant libraries in multiple bacterial species

To characterize pORTMAGE, we have introduced sequence diversity at a specific genetic locus in multiple bacterial species. Specifically, we have randomized six individual nucleobases within the endogenous *asnA* of three phylogenetically related species, *E. coli*, *S. enterica*, and *C. freundii*. Using organism-specific, 90-base-long oligos carrying six randomized positions, we have carried out five cycles of MAGE with and without mismatch repair control in all target species. Next, genomic DNA was isolated from the resulting cell library and the oligo target region was amplified by PCR. Finally, these PCR fragments were subjected to Illumina sequencing to analyze the allelic composition within each bacterial population. In all three species, allelic replacement efficiencies with pORTMAGE were at least an order of magnitude higher at all targeted positions than the frequencies obtained with pSIM8 (Figure 9). Additionally, pORTMAGE substantially reduced the biases in oligo-incorporation which allowed for a more uniform representation of each mutant within the population. Importantly, by using pORTMAGE, we obtained a bias-free mutant library at the target locus in all species. Moreover, sequence analysis suggested that all possible single-

step mutations and their combinations, approximately 4000 variants, were represented at a reasonable frequency within the resulting cell library.

A

E. coli K-12 MG1655

B

Salmonella enterica
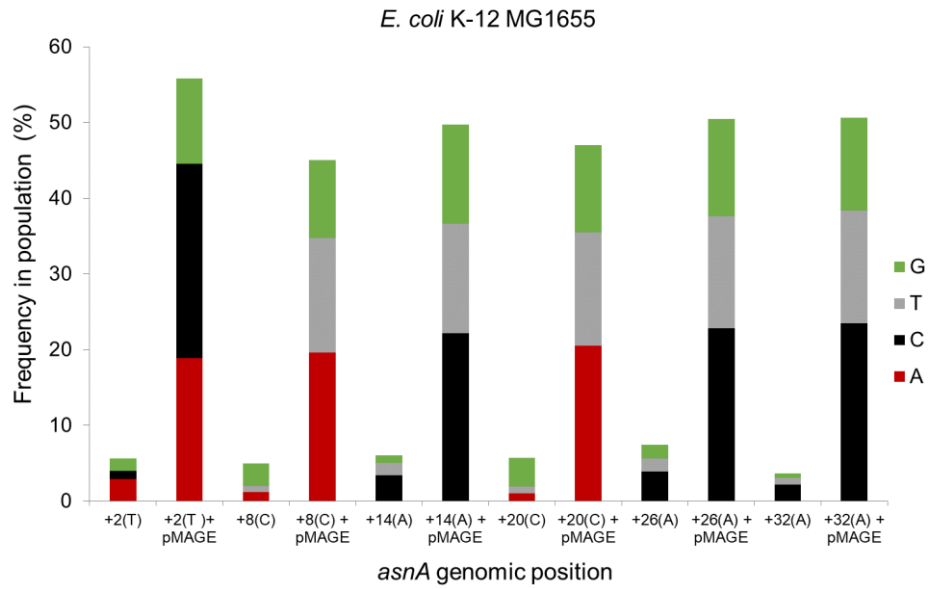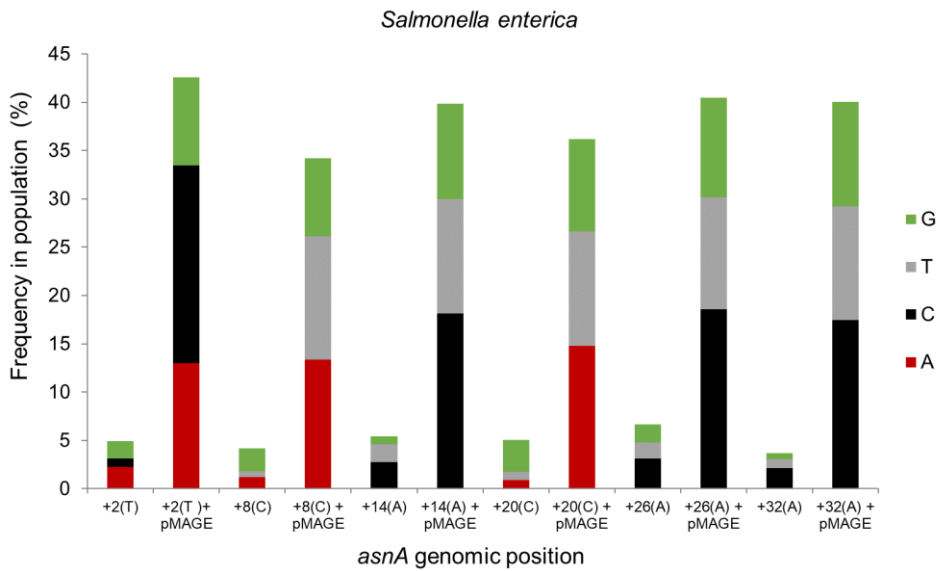
C
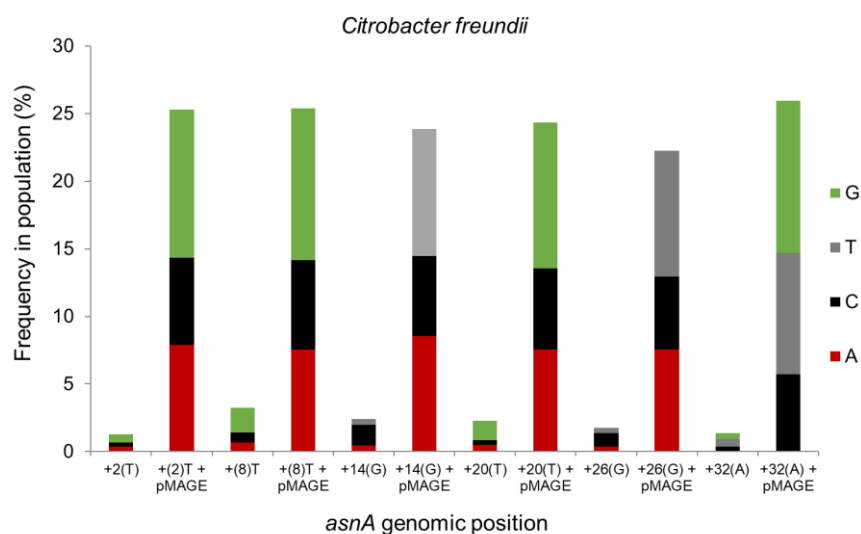
Figure 9. Mutant library generation at *asnA* in (A) *E. coli* K-12 MG1655, (B) *S. enterica* serovar *Typhimurium*, and (C) *C. freundii* ATCC 8090. The first stacked column at each position indicates mutant library generation by using pSIM8, while the second column indicates mutagenesis that were performed by pORTMAGE (indicated with "+pMAGE"). The *asnA* genomic positions of each randomized nucleotide is displayed relative to the first nucleotide of the gene. The wild-type genomic nucleotide is indicated in parentheses. Values are based on a single Illumina amplicon deep-sequencing assay for each.

These results suggest that pORTMAGE, after the transformation of a single plasmid, allows the rapid generation of large, unbiased sequence libraries carrying random mutations at desired positions in multiple enterobacterial species. In our follow-up work, this feature served as an enabling technology to scale-up oligo-recombineering in order to analyze mutational effects in a massively parallel manner within multiple bacterial species.

Development of a method for the *in vivo* targeted mutagenesis of long genomic segments

Single-stranded (ss) DNA-mediated recombineering is a highly versatile tool for multiplex bacterial genome engineering [11]. However, existing ssDNA recombineering-based techniques enable the randomization of very short sequences only (such as neighboring residues in a protein-coding sequence) [38,46,47]. This limitation is of great significance, as the efficiency of ssDNA-mediated recombineering depends on the oligo's sequence identity to its target region. Thereby increasing the number of mismatches within a single oligonucleotide exponentially decreases the efficiency of oligo-incorporation during recombineering [38,43]

(Figure 10). Therefore, diversification of genome sequences which are longer than approximately 30 base pairs is not feasible with a single oligonucleotide [38,126]. This, in turn, hinders the exploration of the combinatorial sequence space and targeting extended genomic loci [129].
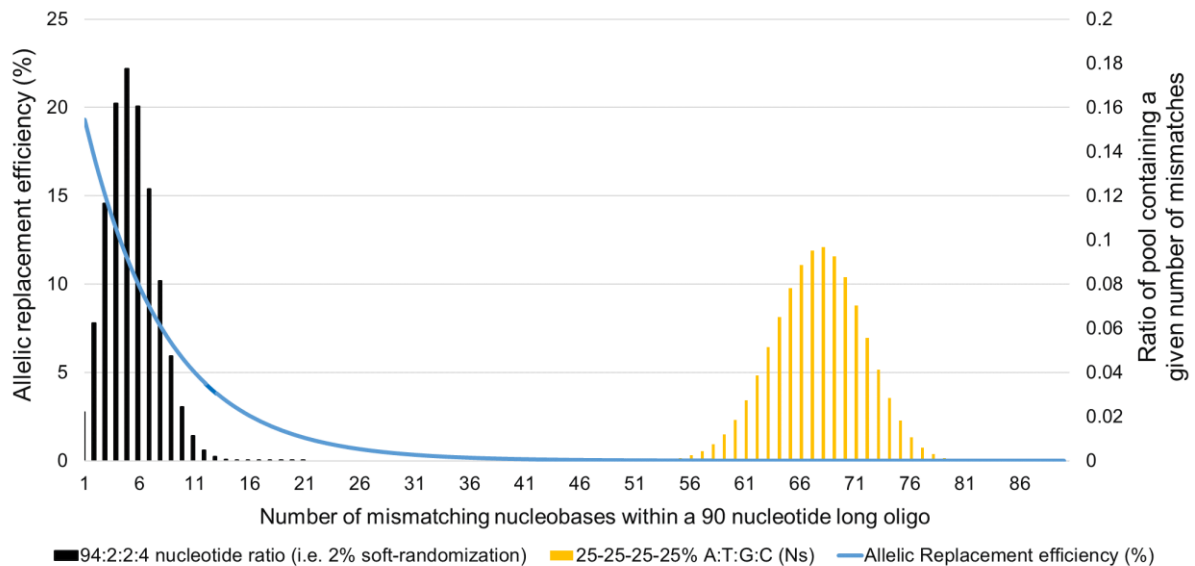
Figure 10. The relation between the level of soft-randomization of a 90-nucleobase-long oligonucleotide, the number of mismatches, and the efficiency of ssDNA-mediated genome editing. High level of randomization (i.e. 25% A : 25% T : 25% G : 25% C, marked with the letter N, the IUPAC code of degenerated nucleotide) generates a pool of sequences with limited homology to their target site. Soft-randomization (i.e. 94% of the wild-type nucleotide admixed with 2–2% of the other 3 possible nucleotides, termed as 2% soft-randomization) mainly produces highly homologous sequences compared to their target. As the number of mismatching bases is logarithmically related to allelic-replacement (AR) efficiency, increasing oligo-randomization level during DNA synthesis rapidly abolishes the efficiency of ssDNA-mediated genome editing. Figure is based on data from Wang, H. H. & Church, G. M. Meth. Enzymol. 498, 409–426 (2011) and optMAGEv0.9, available at http://arep.med.harvard.edu/optMAGE.

Although a strategy has been proposed where the individual nucleotide positions are mutagenized separately by using a distinct oligo for each nucleotide position or amino acid encoding codon (i.e. MAGE-Seq and MO-MAGE) [31,90], these protocols require hundreds to thousands of individual oligos even for a single gene or regulatory sequence. Moreover, these methods do not allow of the simultaneous exploration of epistatic interactions across mutation combinations in a bias-free manner. Consequently, their application hinders a cost-effective genome engineering and the in-depth investigation of the phenotype-to-genotype landscape.

49

Thus, a cost effective method that does not necessitate high-throughput DNA synthesis to systematically explore mutational effects and perform targeted mutagenesis is eagerly needed.

Therefore, we have developed an ssDNA-mediated, recombineering-based method which utilizes pools of partially overlapping, soft-randomized oligonucleotides and allows of up to a million-fold increase in mutation rate at multiple targets. This novel method, termed directed evolution with random genomic mutations (DIvERGE) (Figure 11), enables the exploration of vast numbers of combinatorial genetic alterations in their native context, while off-target mutagenesis is minimized [111].

To develop DIvERGE, based on the relation between the number of mismatches compared to a genomic target and the incorporation pattern of mutations from oligonucleotides with MAGE, we hypothesized that covering chromosomal segments with partially overlapping, mutagenizing oligonucleotides allows a uniform mutagenesis of the target segments. Additionally, by adjusting the randomization level in each oligonucleotide on the course of chemical DNA synthesis on a way to limit the number of mismatches compared to the target sequence would allow for an efficient integration of long mutagenizing oligos into their corresponding target site. Limiting the level of randomization during DNA synthesis would also ensure that all possible mutations and their combinations are represented in the synthesized oligo pool. Importantly, such randomized oligos can be straightforwardly synthesized using a soft-randomization-based phosphoramidite DNA synthesis protocol on most automated DNA synthesizers [150,151]. Soft-randomization-based synthesis would also enable the precise control of the rate and spectrum of mutations within each oligonucleotide. In fact, soft-randomization can be described as the chemical DNA synthesis method that introduces a small amount of nucleotide-mixture at specific variable positions of the wild-type sequence [150,151]. To generate soft-randomized oligos in our case, oligos for recombineering were manufactured with soft-randomization along their entire sequence. It thereby generated oligos with randomly positioned random mutations along their entire length. As a practical advantage, the application of soft-randomized DNA synthesis also circumvents the need for expensive massively-parallel oligonucleotide synthesis, and reduces the costs to as little as 35 EUR (~40 USD) for a single oligonucleotide library, which is two orders of magnitude lower than that for large-scale DNA synthesis methods [26,31,152].

Based on the utility of this method to introduce genetic diversity at multiple genomic loci, we termed it 'directed evolution with random genomic mutations', or DIvERGE for short (Figure 11).
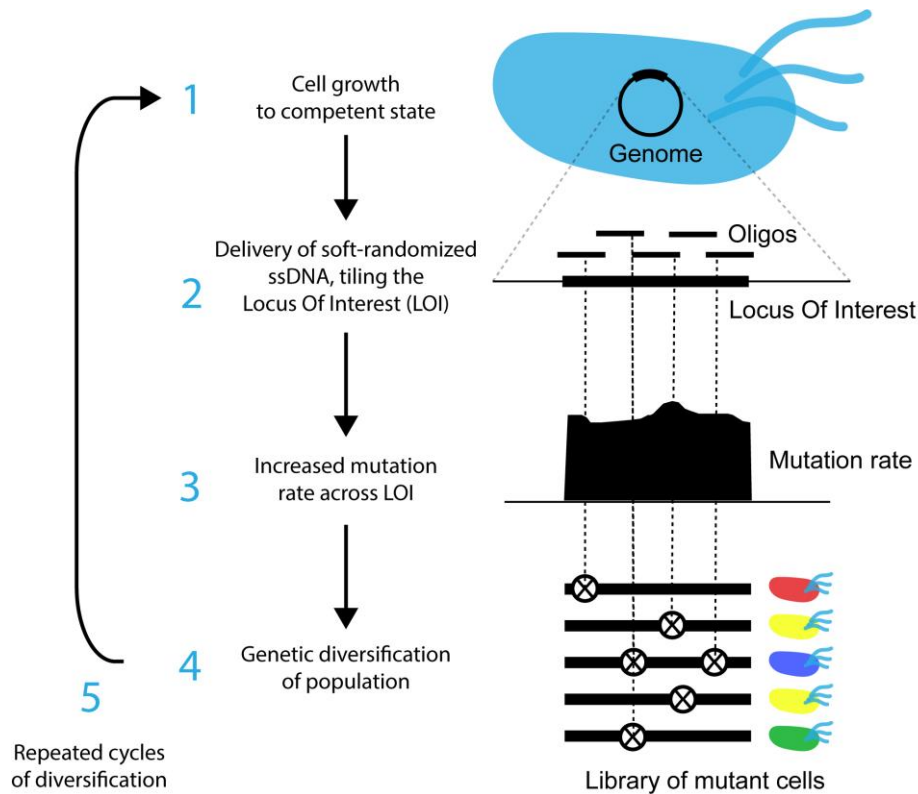
Figure 11. Schematic overview of directed evolution with random genomic mutations (DIvERGE). Soft-randomized DNA oligo synthesis precisely controls the rate and spectrum of mutations in partially overlapping oligos. These oligos fully cover the locus of interest (LOI) and induce mutagenesis at this target site after incorporation with pORTMAGE. By building on the cyclic workflow of pORTMAGE, DIvERGE proceeds via cell growth (1), oligo delivery and incorporation (2) and subsequent mutagenesis (3) that leads to a high genetic diversity at the target locus (4). Iterative repetition of this workflow (5) highly elevates genetic diversity within the mutagenized cell population. Figure is adapted from Nyerges, Á. et al. (2018) Directed evolution of multiple genomic loci allows the prediction of antibiotic resistance. Proceedings of the National Academy of Sciences, 115, E5726–E5735 [125].

Uniform and adjustable mutagenesis of selected genomic targets by soft-randomized oligos

To test our hypothesis, we first tested whether the soft-randomized synthesis of DNA oligos can keep mutation rate and the mutational spectrum uniform along a defined sequence. Therefore, we synthesized 90-nucleotide-long soft-randomized oligos which were complementary to the previously designed landing pad [130]. Within these oligos, each nucleotide position was soft-randomized with 0 to 2% of all 3 possible mismatching nucleobases. This ratio of soft-randomization was defined as the fraction of mismatching nucleobases at each nucleotide position.

51

Next, the nucleotide composition within the resulting oligo was assayed by Illumina sequencing. In line with our expectations, soft-randomization induced a balanced distribution of mutations compared to the wild-type sequence of the landing pad, along the entire length of the oligonucleotide (Figure 12).
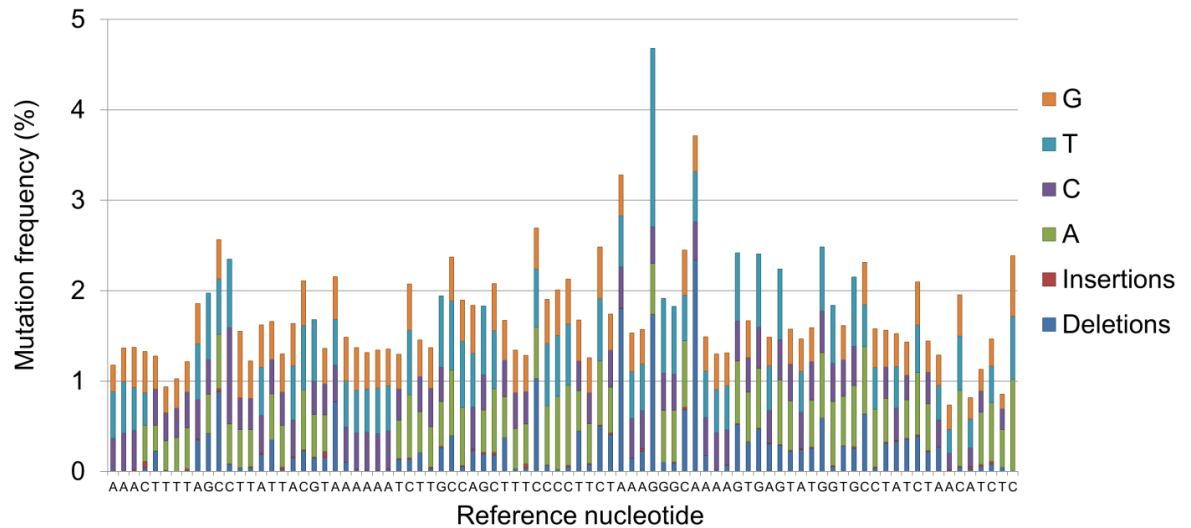
Figure 12. Mutation frequency along a soft-randomized, 90-nucleotide-long oligonucleotide, TETRM3_05. The oligonucleotide was synthesized at each nucleotide position along its entire length of sequence by using a 0.5% nucleoside-phosphoramidite soft-randomization (i.e. 98.5% of the wild-type nucleotide admixed with the other 3 possible nucleotides, 0.5% of each). Mutation frequencies are based on Illumina MiSeq sequencing of 5×10^4 individual oligonucleotide strands.

Next, we examined the genomic incorporation of soft-randomized ssDNA oligos at the landing pad sequence by pORTMAGE (Figure 13). Using two soft-randomized oligos, we have simultaneously targeted two 90-basepair-long regions within the landing pad, and performed five iterative pORTMAGE genome editing cycles with an equimolar mixture of both oligos (TETRM1 and TETRM3). Mutation frequency and the distribution of mutations at the genomic target were determined by Illumina sequencing. Sequence analysis of the two genomic loci revealed that soft-randomized oligos had successfully mutagenized their targets. Moreover, mutagenesis from a single 90-nucleotide-long oligo was extended to 78 nucleobases which is over 2.5-times longer than the maximum range of mutagenesis demonstrated by previous ssDNA-recombineering methods [38,126] (Figure 13). The decrease of recombineering efficiency at the oligo-termini is possibly due to the destabilizing effects of mismatches at these regions that subsequently prevents integration of mutations. This hypothesis is also supported by the detected mutation-integration pattern along the length of MAGE-oligos [40].
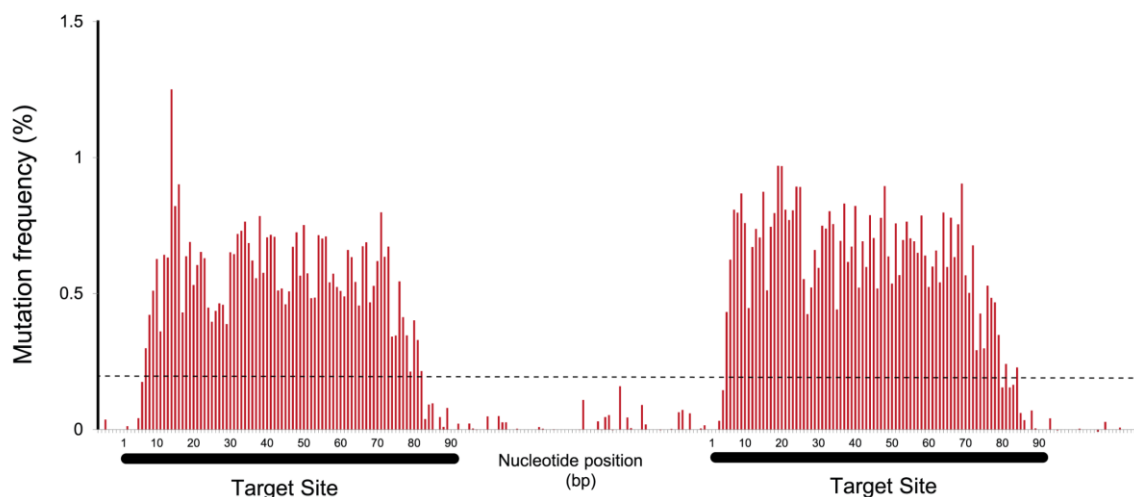
Figure 13. Genomic mutation frequency of DIvERGE in *E. coli* K-12 MG1655 after 5 cycles of mutagenesis. Two 90-basepair-long genomic regions separated by an intermediate region of 70 base pairs were targeted for mutagenesis with two pools of 2% soft-randomized oligos (TETRM1_2 and TETRM3_2). The target sites are indicated as black lines beneath the x axis. A cut-off value of 0.2 % (dashed line) was used as a threshold to qualify diversified positions. Mutation frequencies are based on Illumina amplicon deep sequencing.

Reassuringly, no major bias in mutational spectrum was detected along the target region, and the frequency of each individual substitution type fell between 13.8 and 22.4% in the genomic mutant library (Table 3). Moreover, the extent of sequence diversity of the population was tunable by the nucleotide admixing rate of the oligo pool used for soft-randomization based DNA synthesis (Figure 14).

Mutational bias indicator	Oligonucleotide		Genomic	
	Frequency (%)	Standard deviation	Frequency (%)	Standard deviation
A→G, T→C	13.8	0.8	13.8	0.4
G→A, C→T	22.4	1.5	22.2	1.8
A→T, T→A	15	1.1	17.1	1.4
A→C, T→G	12.9	0.5	13.8	0.2
G→C, C→G	14.2	0.4	14.8	1.1
G→T, C→A	21.7	0.2	18.2	0.6

Table 3. Mutagenic spectrum indicators of DIvERGE mutagenesis. Table shows the spectrum of substitutions in a 0.5% soft-randomized oligo, TETRM1_05, after

phosphoramidite-based DNA synthesis, and the resultant spectrum of mutations at the genomic target after five cycles of DIvERGE mutagenesis in *E. coli* K-12 MG1655. Frequencies and standard deviations are calculated from Illumina sequencing-based mutational composition analysis.
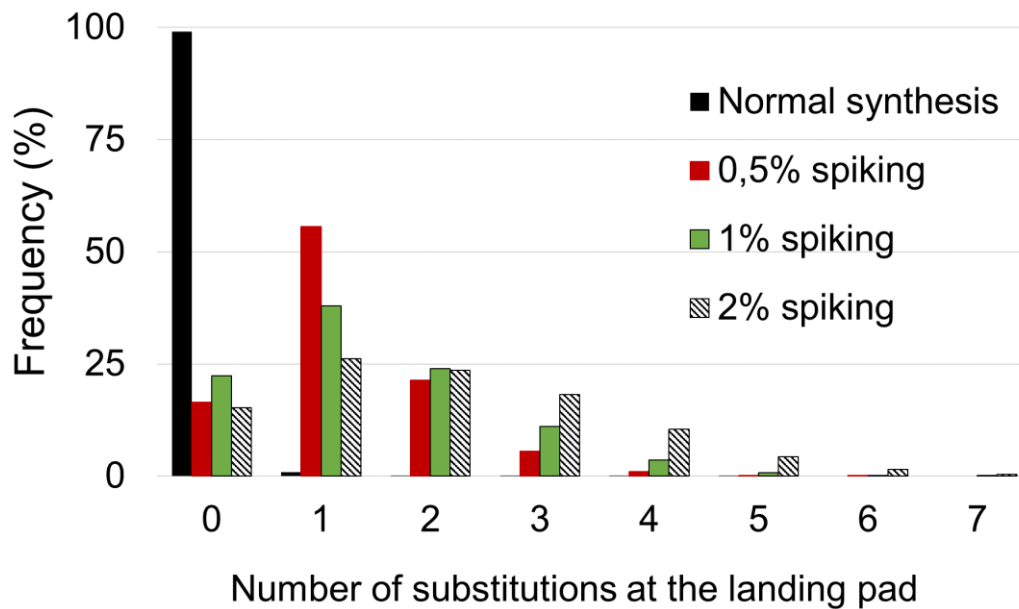
Figure 14. Frequency of substitutions within the landing pad in the course of DIvERGE mutagenesis as the function of nucleotide admixing ratio (termed 'spiking') during DNA soft-randomization. Samples represent the allelic composition at the landing pad in *E. coli* K-12 MG1655 after 5 iterative DIvERGE mutagenesis cycles with oligos of the corresponding soft-randomization (i.e. 'spiking') ratio [151]. 'Normal synthesis' represents mutagenesis with a non-randomized oligo. Values are based on Illumina amplicon deep sequencing.

Performing 5 DIvERGE cycles with medium-level soft-randomization, i.e. introducing 2-2-2% of admixed mismatching nucleobases within the oligo sequence at every position, resulted in an over one million-fold increase in the mutation rate at the oligo target compared to the background level. This corresponds to an increase from 2.2×10^{-10} to 2.4×10^{-4} mutations per nucleotide per generation, as measured by Illumina sequencing of the landing pad, and compared to the wild-type mutation rate of *E. coli* K-12 MG1655 [130]. Importantly, the mutation rate of non-targeted regions, which was measured at the non-targeted regions of the landing pad, remained low (Figure 13). Overall, DIvERGE enabled an efficient control of sequence diversification at the target loci of interest by the recombineering-based integration of soft-randomized ssDNA oligos.

Soft-randomized oligos efficiently mutagenize multiple bacterial species

Based on the broad host-range functionality of pORTMAGE [125], we hypothesized that DIvERGE could also be applicable to distant relatives of *E. coli* as well. To test our hypothesis, again we selected *Salmonella enterica serovar. Typhimurium* LT2 and *Citrobacter freundii* ATCC 8090 as models. To characterize the efficiency of DIvERGE in a uniform manner across these species, we utilized our previously established landing pad assay. Similarly to our prior tests to characterize pORTMAGE, our landing pad system allowed us to assess the performance of DIvERGE by using the same set of oligos and the same protocol as in *E. coli* K-12 MG1655.

To test the functionality of DIvERGE in both strains we relied on our general pORTMAGE protocol and integrated two soft-randomized oligos (TETRM03 and TETRM03) into the genomic landing pad of *Salmonella* and *Citrobacter*. As expected, the iterative integration of these oligos efficiently induced mutagenesis in both species (Figure 15) Moreover, we were able to achieve a mutation rate of at least 10^5-times higher at the target sequence in both species compared to their corresponding wild-type mutation rates (Table 4).

A
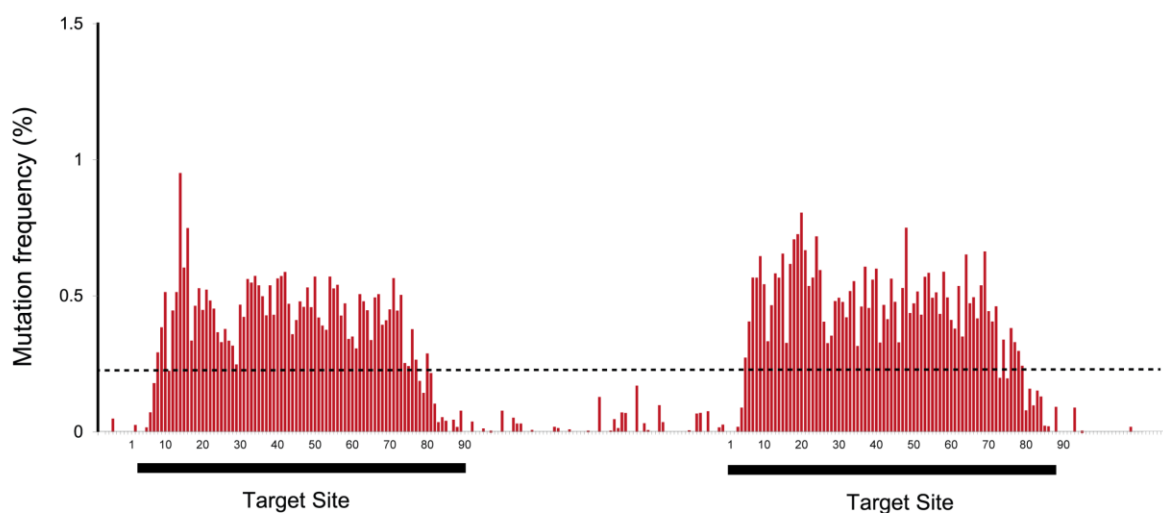
B
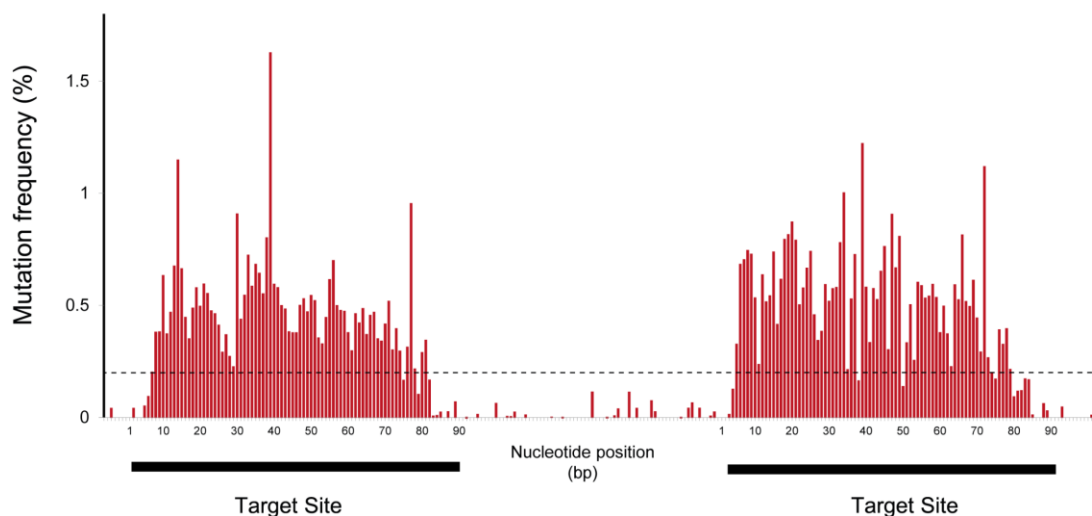
Figure 15. DIvERGE is applicable to multiple bacterial species. Figure shows the increase in genomic mutation frequency, measured as the frequency of mutations occurring at a given nucleotide position after 5 consecutive DIvERGE cycles in **(A)** *Salmonella enterica serovar. Typhimurium* LT2 and **(B)** *C. freundii* ATCC 8090. Two 90-basepair-long genomic regions, separated by an intermediate region of 70 base pairs, were targeted for mutagenesis with two pools of 2% soft-randomized oligos (TETRM1_2 and TETRM3_2). The target sites of oligos are indicated as black lines beneath the x axis. The dashed line indicates a cut-off value of 0.2% which served as a qualification marker of diversified positions. Mutation frequencies are based on Illumina amplicon deep sequencing.

Strain	Wild-type mutation rate (mutation/locus/generation)	DIvERGE mutation rate (mutation/locus/generation)	Fold increase in mutation rate induced by DIvERGE
E. coli **K-12 MG1655**	1.10×10^{-8}	1.22×10^{-2}	1.11×10^{6}
Salmonella enterica **LT2**	6.37×10^{-8}	0.92×10^{-2}	1.44×10^{5}
Citrobacter freundii **ATCC 8090**	1.13×10^{-8}	1×10^{-2}	8.9×10^{5}

Table 4. Locus-specific elevation of mutation rates in *Salmonella enterica*, *E. coli* K-12 MG1655, and *C. freundii* ATCC 8090 by DIvERGE mutagenesis at the landing pad.

56

DIvERGE can mutagenize extended genomic regions

Based on the observed incorporation pattern of soft-randomized oligos at the landing pad, we hypothesized that covering long genomic segments with multiple, partially overlapping soft-randomized oligos would enable the uniform mutagenesis of extended targets. Therefore, we increased our genomic target site to 560 base pairs by designing and synthesizing multiple overlapping oligos where each oligo overlapped with the adjacent one, and thereby covered the entire target (as shown in figure 11). For the target loci we chose *folA*, which encodes the enzyme dihydrofolate reductase (FolA, or alternatively DHFR), and its promoter region.

The enzyme encoded by *folA* provides the major dihydrofolate reductase activity in the tetrahydrofolate biosynthetic pathway. FolA catalyzes the reduction of dihydrofolate to tetrahydrofolate via hydride transfer from NADPH to the pteridine ring, using NADPH as a reducing cofactor [153,154]. *E. coli* FolA is a monomeric protein containing 159 amino acids. Tetrahydrofolate is an essential intermediate in the biosynthesis of proteins and nucleic acids, and therefore FolA is essential for cell division and growth. Thus, FolA is a frequent drug target: numerous approved drugs widely used in therapy are characterized by this mechanism of action, and have antitumor, antibacterial or antimalarial properties [154–156] (e.g. the antibiotic trimethoprim).

E. coli K-12 MG1655 FolA
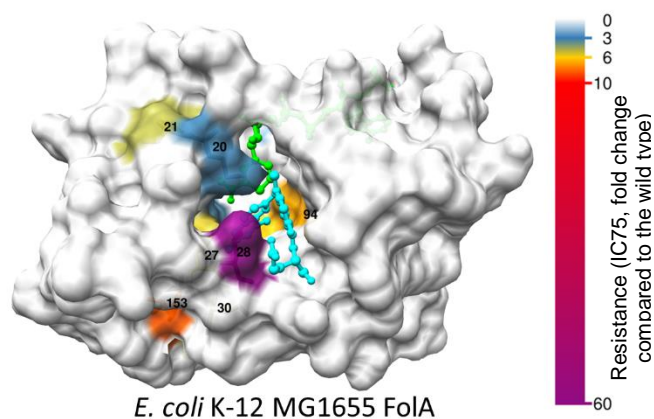
Figure 16. Dihydrofolate reductase (FolA) of *E. coli* (Protein Data Bank ID: 1RH3) and the distribution of trimethoprim resistance-conferring amino acid positions on the surface FolA. Reported mutations conferring trimethoprim resistance are located at the active site of the enzyme and around the binding site for the NADPH cofactor (green). The figure is based on trimethoprim IC75 values of *folA* mutations as were reported in *Nyerges, A* et al., 2018, PNAS.

Studying the evolutionary processes of *folA* has profound clinical importance, not only because dihydrofolate reductase is the target of the widely used antimicrobial drug trimethoprim, but also because it serves as a target for developing novel antibiotics (e.g. iclaprim [157]). Trimethoprim is also routinely used in therapy to treat *E. coli* infections [158]. Concerning *E. coli*, prior studies have demonstrated that under prolonged trimethoprim selection the evolution of antibiotic resistance develops predominantly through mutations at *folA* (Figure 16) or is caused by horizontally transferred, trimethoprim-insensitive *folA* variants [158–163]. Based on its role in trimethoprim resistance and the fact that *folA* mutants can easily be selected on antibiotic-containing agar plates, bacterial dihydrofolate reductases has been extensively studied to understand the evolution of antibiotic resistance and involved evolutionary mechanisms [164–166].

Therefore, in order to investigate the utility of DIvERGE to rapidly evaluate trimethoprim resistance-conferring mutations of *folA* we have performed mutagenesis on the entire resistance-determinant locus of *E. coli*. As prior studies had demonstrated that trimethoprim resistance frequently results from mutations in both the *folA* promoter [161] and the protein-coding region of the gene [162,165,167], we have decided to mutagenize both regions in *E. coli* K-12 MG1655 by the simultaneous use of eight overlapping soft-randomized oligos. Based on the previously observed incorporation pattern of mutations induced by soft-randomized DIvERGE oligos (Figure 13 and 15), we have applied 18-nucleotide-long overlaps between adjacent oligos to ensure uniform mutagenesis at all nucleotide positions. Overall, one DIvERGE oligo targeted the promoter region, while seven targeted the 480-basepair-long protein-coding sequence of *folA*.

Using an equimolar mixture of these overlapping soft-randomized oligos, we first generated *folA* variant libraries with single point mutations at the target sequence. To achieve that we electroporated each of the soft-randomized oligos separately into pORTMAGE-containing *E. coli* K-12 MG1655 cells. Resulting mutant libraries were then subjected to trimethoprim selection pressure on agar plates containing 3 µg/ml trimethoprim, and the resistant colonies were collected. Subsequently, the genotypes of pooled resistant colonies were determined by Illumina amplicon deep sequencing at the *folA* locus. Variants with more than one mutation were excluded from this sequence data analysis. Thus, we solely focused on the single-step adaptive mutational landscape of *folA* and its promoter sequence. We have found that 81% of the identified point mutations reside in the protein-coding region, primarily localized at the active site, as well as at the NADPH binding site of FolA as reported in literature [163], while the rest is located at the promoter region. DIvERGE has induced 17 previously described [160–163,165,166] mutations that are already known to be linked to trimethoprim resistance. Besides, we have also revealed at least 7 new substitutions. Next, we have

individually reconstructed each of these mutations in *E. coli* MG1655 and confirmed their resistance phenotype (Figure 23).

Consecutive DIvERGE and selection cycles rapidly evolve high-level antibiotic resistance

High-level resistance to trimethoprim generally demands multiple mutations at *folA* and its promoter. These mutations jointly act to up-regulate gene expression and/or decrease the drug molecule's binding to its target site [156].

To investigate the ability of DIvERGE to model, and thus to enable the analysis of multistep evolutionary processes, we have subjected *E. coli* K-12 MG1655 cells to consecutive cycles of DIvERGE mutagenesis. Using eight soft-randomized oligos, five rounds of oligo-integrations were carried out with pORTMAGE, simultaneously targeting all nucleotide positions in the *folA* promoter and in the protein-coding region. Sequence analysis of the resulting mutant libraries revealed that the iterative integration of these eight oligos has successfully randomized the entire target (Figure 17 A), and has successfully generated higher-order mutational combinations as well (Figure 17 B).

A
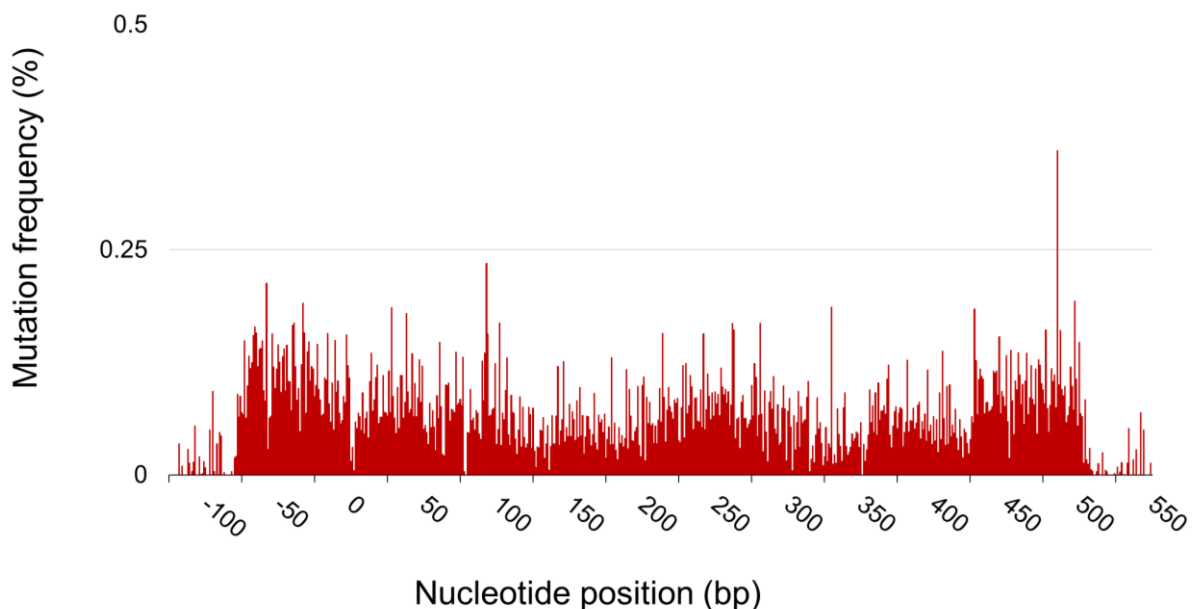
B
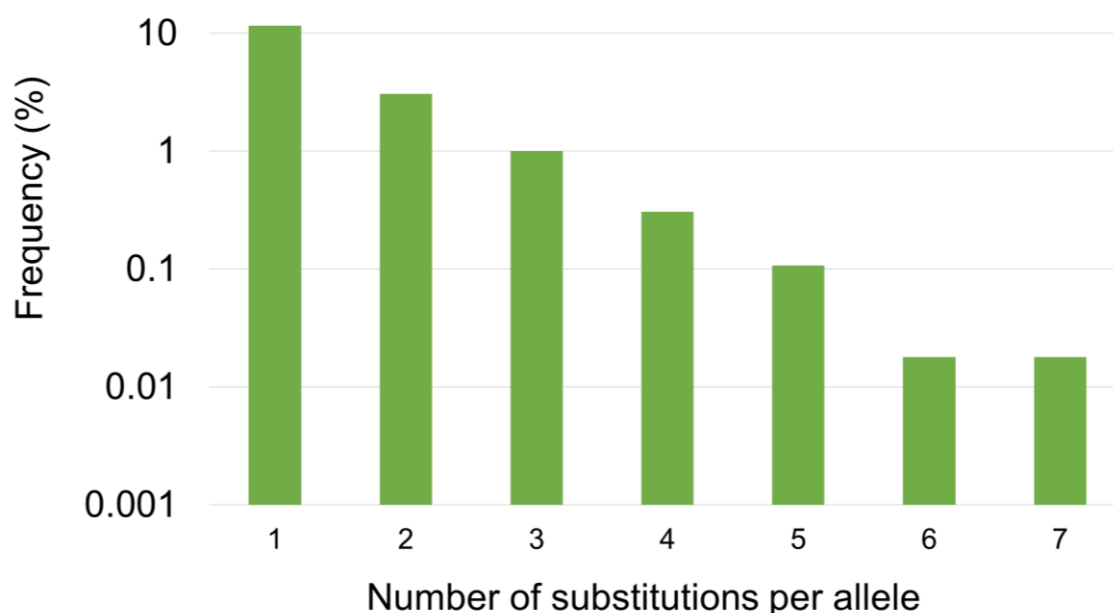
Figure 17. DIvERGE mutagenesis along the full length of the drug target of trimethoprim. (A) Mutation frequency at *E. coli* K-12 MG1655 *folA* after five consecutive cycles of DIvERGE. Positions 0 and 480 refer to the first nucleotide position of the start and the last position of the stop codon of *folA*, respectively. Mutation frequency is defined as the background-normalized frequency of substitutions occurring at a given nucleotide position, analyzed by Pacific Biosciences Single Molecule Real-Time sequencing. **(B)** Naïve library composition at *folA* after five cycles of DIvERGE mutagenesis targeting the *E. coli* MG1655 K-12 *folA* locus. The figure is based on Pacific Biosciences Single Molecule Real-Time sequencing reads, showing the frequency of each allele with the corresponding number of nucleobase substitutions within the oligo-target region.

After five cycles of DIvERGE mutagenesis, we identified *folA* mutants characterized with an up to 895.7-fold increase in the 75% inhibitory concentration (IC75) of trimethoprim compared to the wild-type parental strain (Table 5). IC75 is a generally accepted measure of quantifying bacteriostatic antibiotic activity, referring to the drug concentration that inhibits bacterial growth by 75% compared to the drug-free condition, thus we have chosen IC75 to characterize trimethoprim resistance in our experiments [166].

Strain	*folA* regulatory mutation(s)	*FolA* mutation(s)	Trimethoprim IC75 value (μg/ml)	Fold change of IC75 compared to wild-type
E. coli K-12 Strain 1	C-58T	A26T, L28R, P39R	1254	895.7
E. coli K-12 Strain 2	C-58T, T-74A	P21P, L28R, N147D	447.5	319.6
E. coli K-12 Strain 3	C-58T	L28R	610	492.8
E. coli K-12 Strain 4	C-43T, C-58T	A26D, L28R, H45R	794	567.1

Table 5. Susceptibility of individually selected *E. coli* K-12 MG1655 *folA* mutants to trimethoprim after five cycles of DIvERGE mutagenesis. Data represent the 75% inhibitory concentrations (IC75) based on the average of three independent measurements. Trimethoprim-specific IC75 of the wild-type equals to 1.4 μg/ml.

Whether DIvERGE can be applied to multi-round directed evolution using only a single set of soft-randomized oligos generated at the beginning of an experiment is a significant issue regarding the potential application of DIvERGE in adaptive laboratory evolution experiments. One may argue that oligos designed to target the wild-type *folA*, may revert mutations which had accumulated at an earlier stage of laboratory evolution [90]. To address this question, we focused on a *folA* variant which was generated in our 5-cycle DIvERGE experiment and contained three mutations within *folA*. Next, we carried out five additional DIvERGE cycles on this mutant with the soft-randomized oligo pool designed to target the wild-type sequence, and subsequently we sequenced the resulting library. Reassuringly, sequence randomization was successful along the whole length of the target sequence, and no substantial decrease in the level of nucleotide variation was observed (Figure 18). The only exceptions were the nucleotides which were directly adjacent to the three pre-existing mutations. Importantly, compared to the parental allele, novel genotypes could be selected which displayed an extremely high level of trimethoprim resistance. Selected genotypes showed an over 3900-fold increase in the relative IC75 value of trimethoprim as compared to wild-type *E. coli* K-12 MG1655 (Table 6).
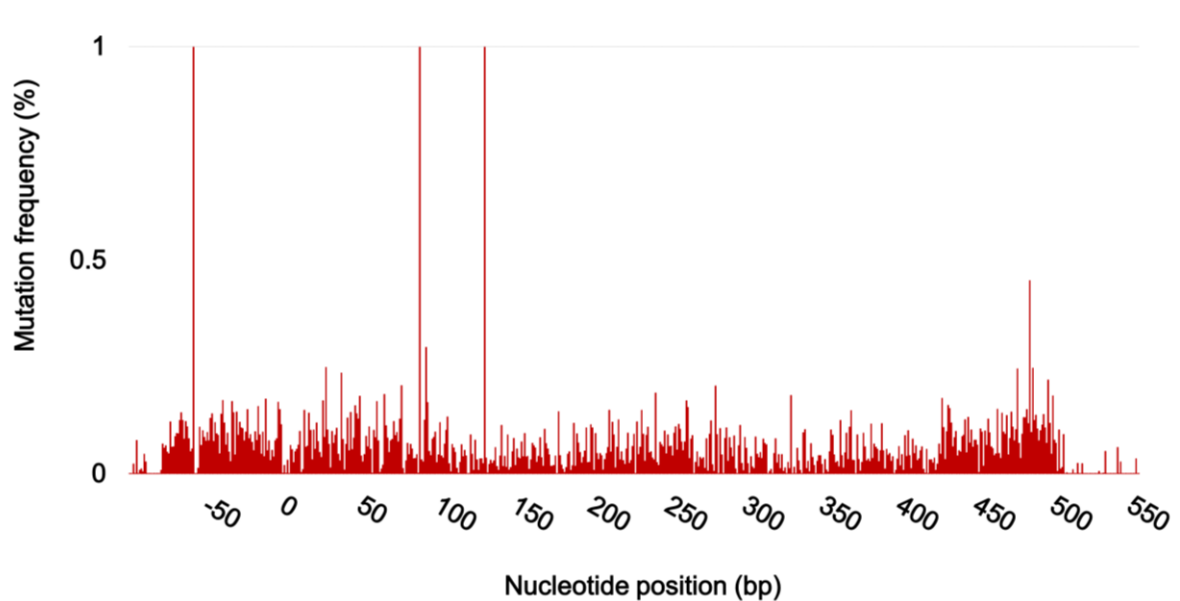
Figure 18. Library composition of an *E. coli* K-12 MG1655 *folA* variant, differing in three substitutions at the nucleotide positions 58, 90, and 132 as compared to the wild-type sequence, after five additional cycles of DIvERGE mutagenesis using an oligo pool designed for the wild-type sequence. Allelic composition was determined by Illumina sequencing.

Strain ID	folA regulatory mutation	Amino acid change (or same-sense SNP)	IC75 (-fold change compared to wild-type)
8	C-58A	A26S, L28R, W30C, (C132G)	>3900
11	C-58A	A26T, L28R, W30C	>3900
12	C-58A	A26T, L28R, (C132G), M88L	2140
23	C-58A	A26T, L28R, (C132G)	1430

Table 6. Susceptibility of individually selected *E. coli* K-12 MG1655 *folA* mutants to trimethoprim after five plus five cycles of DIvERGE mutagenesis. Multi-round DIvERGE-generated *folA* alleles selected at agar plates treated with 1000 μg/ml trimethoprim displayed an extremely high level of trimethoprim resistance and contained additional mutation combinations compared to the parental variant resulting from the first five cycles of DIvERGE. The parental variant generated after the first 5 cycle of DIvERGE contained C-58A; as well as W30C and/or C132G as same-sense mutations at *folA*. Data represent the 75% inhibitory concentrations (IC75) based on the average of three independent measurements.

Overall, DIvERGE was capable of generating a diverse set of trimethoprim-resistant variants while simultaneously retaining mutations that were introduced prior to the second round of mutagenesis-selection cycles. As we employed the same oligo pool during the whole course of the experiment, this result indicates that there is no need for a new set of soft-randomized oligos after each round of mutagenesis. Therefore, iterative DIvERGE

mutagenesis-selection cycles can be run rapidly, without interruptions, and can be potentially scaled up towards many parallel allelic variants.

Taken together, these results indicate that DIvERGE rapidly generates higher-order mutational combinations that induce high-level antibiotic resistance; a result that would be otherwise inaccessible via single mutational steps. In fact, it should be noted that adaptive laboratory evolution experiments relying on natural mutagenesis frequently require over one week to achieve a similar level of trimethoprim resistance [162,165,167], indicating that standard methods for adaptive laboratory evolution are outperformed by DIvERGE. Accordingly, when the main targets of bacterial selection are known, DIvERGE is a highly useful tool to accelerate laboratory evolution.

DIvERGE outperforms a state-of-the-art method for whole genome mutagenesis

We next compared the performance of DIvERGE to an established *in vivo* mutagenesis method called MP6 mutagenesis [110]. MP6 is an inducible plasmid-based system which can upregulate genomic mutation rate up to 322 000-fold, and thereby surpasses the mutational efficiency of other widely used *in vivo* methods. We compared three key aspects of mutagenesis: (I) the level of resistance achieved in the mutagenized populations, (II) the spectra of mutations and their combinations, and finally (III) the extent and consequences of off-target mutagenesis. As MP6 mutagenesis was developed for use in *E. coli*, we compared DIvERGE and MP6 in *E. coli* K-12 MG1655.

Bacterial populations were subjected to MP6-based whole-genome mutagenesis and then to five consecutive cycles of *folA*-targeted DIvERGE. Experimental conditions, including the estimated population size and time-frame, were comparable in the two protocols. Following mutagenesis, the resulting cell libraries were exposed to mild, medium and high levels of trimethoprim stress. The fraction of consequent resistant cells was determined and the resistant clones, approximately 1000 resistant mutants for each trimethoprim concentration, were sequenced by Pacific Biosciences Single Molecule Real-time sequencing to read out the genotypic information for *folA*. Library composition analysis revealed that, compared to MP6, DIvERGE resulted in an 83-fold higher fraction of resistant clones under mild trimethoprim stress. Moreover, the level of resistance reached in DIvERGE-treated populations was 37.6-fold higher than in the MP6-treated population (Figure 19).
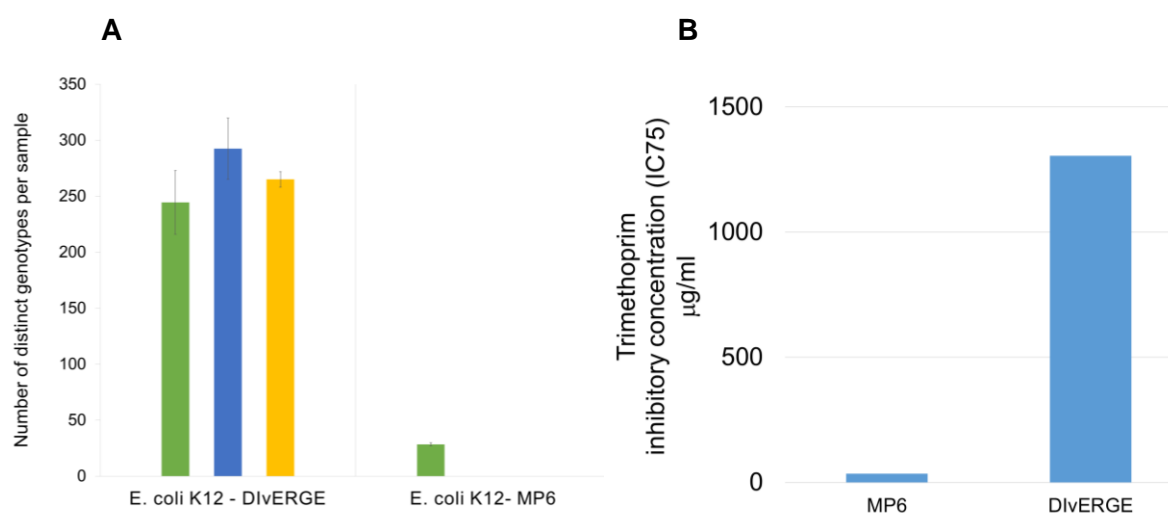
Figure 19. DIvERGE promotes the *in vivo* evolution of antibiotic resistance compared to MP6 mutagenesis. (A) Genotypic analysis of 1000 trimethoprim-resistant variants, generated by MP6 and DIvERGE mutagenesis in *E. coli* K-12 MG1655. Libraries were selected at trimethoprim concentrations 4- (green), 67- (blue), and 267-fold (yellow) higher compared to the IC75 of trimethoprim in wild-type *E. coli* K-12 MG1655. Sequence analysis of each individual variant was performed by amplicon sequencing of the *folA* target site on a Pacific Biosciences RSII instrument in circular-consensus sequencing (CCS) mode. Library composition analyses were performed in duplicates and error bars are calculated as the standard error of the mean for the two replicates. Bar diagrams indicate various levels of trimethoprim stress including trimethoprim concentrations of 4- (green), 67- (blue), and 267-fold (yellow) higher compared to the IC75 of trimethoprim in wild-type *E. coli*. **(B)** The maximal IC75 values of trimethoprim in *E. coli* K-12 MG1655 populations mutagenized with 5 DIvERGE cycles targeting the *folA* locus compared to MP6 mutagenesis.

Sequence analyses of the resistant clones also revealed that under low trimethoprim stress, 89.5 % of the MP6 generated variants contained a single mutation only. By contrast, under the same circumstances, DIvERGE frequently resulted in variants with multiple mutations of different combinations (Figure 20).
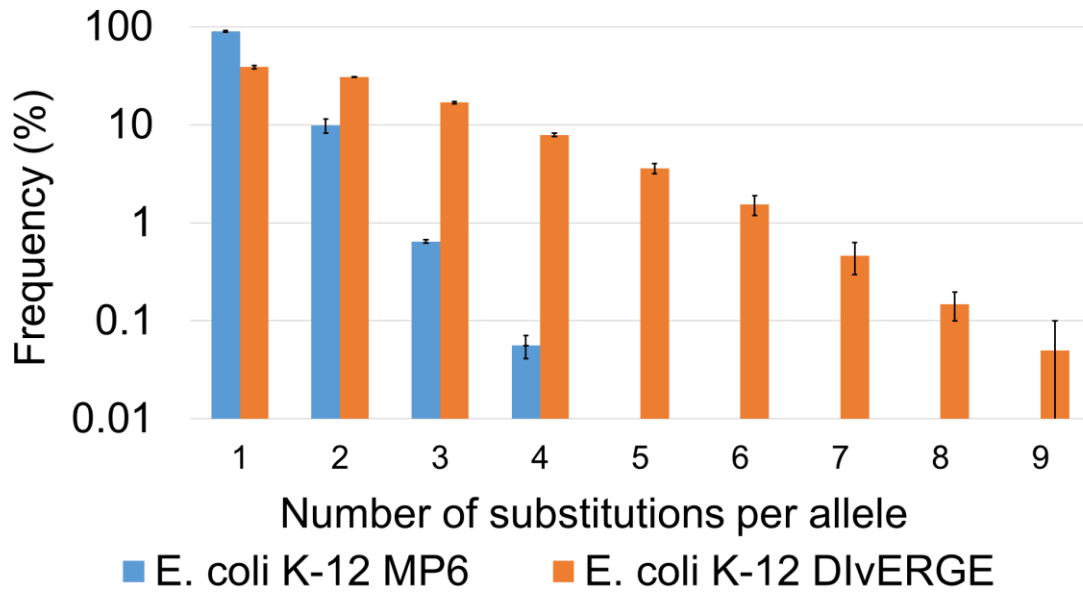
Figure 20. Distribution of the number of substitutions within MP6 and DIVERGE-generated trimethoprim resistant *folA* variants (*n* = 1000) in *E. coli* K-12 MG1655. Error bars represent the standard error of the mean of two replicates.

Finally, as DIvERGE utilizes the pORTMAGE system for an efficient allelic replacement, we expected that off-target mutations would be minimized in this procedure. To investigate the extent of off-target mutagenesis, we measured the fraction of rifampicin-resistant cells in MP6 and DIvERGE-mutagenized cell populations as a measure of off-target mutagenesis. Importantly, the rifampicin resistance-conferring locus (*rpoB* [168]) was not targeted by DIvERGE. As expected, the frequency of rifampicin-resistant cells increased significantly in the MP6-treated cell populations, causing an over 1000-fold increase in the frequency of rifampicin-resistant cells. In contrast, DIvERGE-mutagenized cells showed no significant increase in the frequency of rifampicin-resistant cells compared to the untreated wild-type population (Figure 21). These findings confirm our earlier results on the shortage of undesired mutations in non-targeted regions of the landing pad sequence (Figure 13).
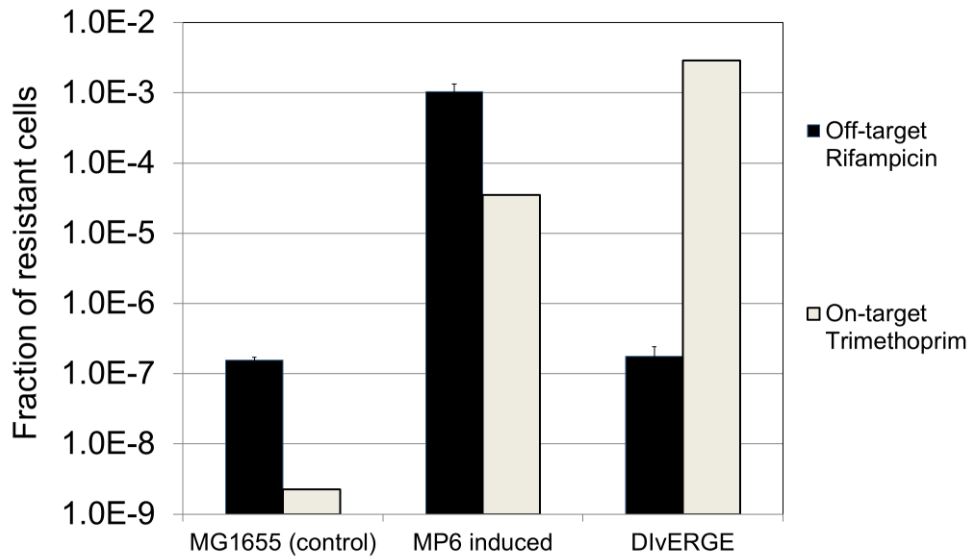
Figure 21. Off-target effects of mutagenesis as indicated by the fraction of rifampicin-resistant cells (representing off-target mutagenesis) and trimethoprim-resistant cells (targeted by DIvERGE) in DIvERGE and MP6-treated *E. coli* K-12 MG1655 populations compared to the wild-type *E. coli* K-12 MG1655 control. Error bars denote the standard error of the mean for 12 biological replicates.

As another important issue, the accumulation of undesired, off-target mutations may interfere with the phenotypic effects of the engineered target modifications. Indeed, resistant clones derived from MP6 mutagenesis showed a 19.8% reduction in growth rate (fitness) in the absence of trimethoprim stress (Figure 22). Of note, this finding does not reflect a potential fitness cost of the MP6 plasmid, as MP6 was removed prior to these fitness measurements. In contrast, no significant fitness decline was observed in DIvERGE-generated, trimethoprim-resistant clones (Figure 22), indicating the lack of fitness-decreasing off-target mutations. Thus, these results suggest again that DIvERGE could be a useful and practical tool to selectively and efficiently target predefined genomic loci, while minimizing off-target effects.
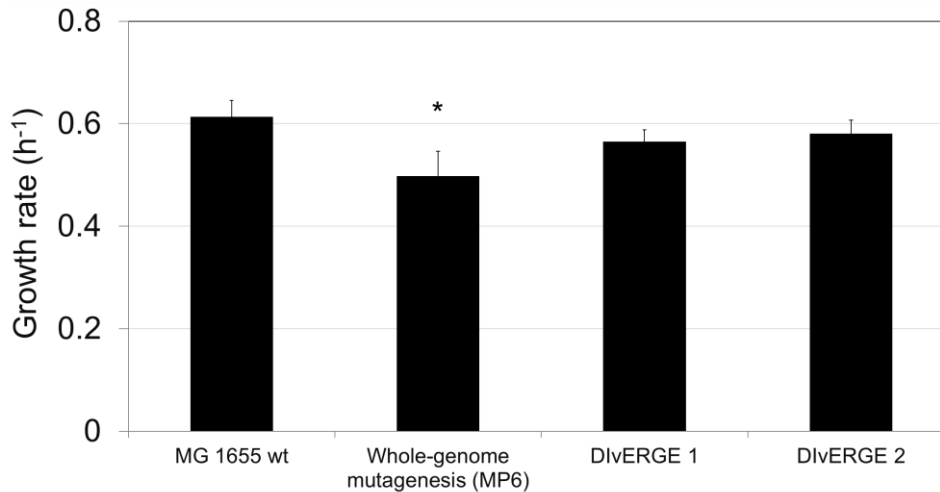
Figure 22. Fitness effects of the different mutagenesis strategies. Growth rate measurements were performed on MP6-generated variants (representing whole-genome mutagenesis) selected using 3 μg/ml trimethoprim, as well as on two, 3 μg/ml (DIvERGE 1) and 50 μg/ml (DIvERGE 2) trimethoprim-selected DIvERGE-generated variants, analyzing 30 individual isolates for each. T-tests were conducted for each strain compared to the wild-type *E. coli* K-12 MG1655 control (MG1655 wt). Star indicates significance of $p < 0.05$.

Differences in the evolution of antibiotic resistance between closely related bacterial strains

Concerning the predictability of antibiotic resistance, it is an important question whether and how closely related bacterial strains evolve resistance to specific antibiotics, and whether these processes differ between close relatives on the phylogenetic tree. This is an issue of special importance, because most of our knowledge about drug resistance evolution is originating from laboratory experiments which were performed on non-pathogenic, laboratory model strains. DIvERGE allowed us to investigate this issue for the first time, due to its exceedingly high mutagenesis rate and the ability to systematically and comprehensively compare mutational effects of entire resistance determinants.

To evaluate resistance evolution in two closely related bacterial strains, we have targeted *folA* and its promoter sequence in a uropathogenic *E. coli* (UPEC) CFT073 strain, a close relative of *E. coli* K-12 MG1655. Importantly, trimethoprim is frequently used against uropathogenic *E. coli* infections in the clinical practice, as this antibiotic is excreted to urine by the kidneys, thus reaches high concentrations at the infection's site [154,158,169]. As the protocol we employed to mutagenize UPEC was quantitatively the same as the one we had previously applied in the case of *E. coli* K-12 MG1655, the mutagenesis of *folA* gave an opportunity to compare single-step resistance processes in both strains. As for *E. coli* K-12

MG1655, we performed a single round of DIvERGE mutagenesis to generate cell libraries with single point mutations at *folA*. Next, as for *E. coli* K-12 MG1655, approximately 3000 resistant clones were selected at mild (3 µg/ml) trimethoprim stress, and were subsequently sequenced using Illumina MiSeq.

Comparison of adaptive mutations in *E. coli* CFT073 has revealed novel resistance mutations compared to the non-pathogenic strain, *E. coli* K-12 MG1655. Despite a *folA* sequence similarity of 99% and almost the same mutation frequently between *E. coli* K-12 MG1655 and the UPEC strain, the conferred relative resistance levels frequently differed between the two strains (Figure 23). Most notably, one of the identified mutations, *FolA* Ala7→Ser displayed resistance phenotype only in the uropathogenic isolate. This mutational effect is especially interesting, as this alanine amino acid in the gene product protein directly interacts with trimethoprim during the drug's mode of action [156]. One explanation for this strain-specific effect could be the importance of 5' mRNA secondary structure on translation [170] and in turn, FolA expression. Thereby, due to slight differences in the sequence-context, the expression of the drug-target can differ between the two isolates.

These results highlight that mutational effects may differ even between closely related strains of the same species, and in turn, the analysis of evolutionary processes in genotypes beyond the given strain of interest may be misleading in some cases. This finding indicates that for reliable information mutational processes should be directly analyzed in the specific strain of interest. DIvERGE, however, offers a practical method to investigate evolutionary processes directly in these strains of interest, due to its broad host-range functionality.
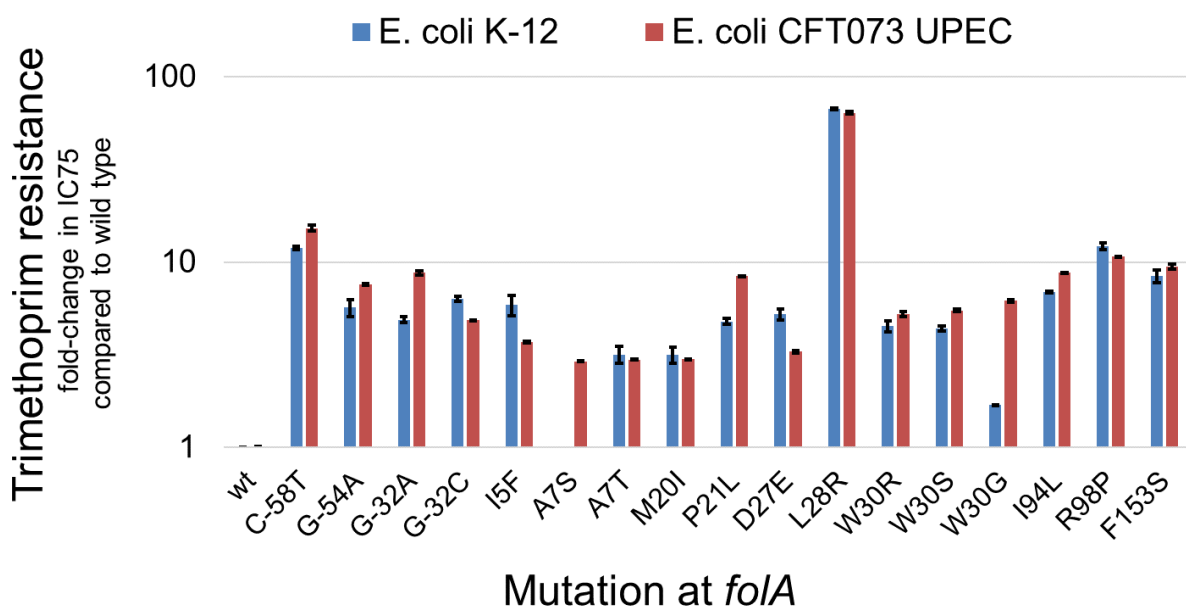
Figure 23. Comparison of relative susceptibility to trimethoprim associated with selected *folA* alleles in *E. coli* K-12 MG1655 (*E. coli* K-12, in blue) and CFT073, a uropathogenic *E. coli* strain (UPEC, in red). Data represent fold-change in IC75 compared

to the corresponding wild-type strain. Results are based on the average of three independent measurements. Error bars denote standard deviation based on three replicates.

Mutagenesis along the full length of multiple drug targets to predict resistance

As indicated by our results using the FolA-inhibitor trimethoprim, DIvERGE is a versatile tool to explore resistance-conferring mutations at a predefined drug target. As DIvERGE has allowed of an exceptionally deep and rapid analysis of trimethoprim resistance, we have anticipated that it could be utilized to evaluate resistance-conferring mutations induced by other antibiotics of various classes as well. From a practical point of view this is especially relevant, as DIvERGE could serve as a tool to forecast resistance evolution to antibiotic candidates at an early stage of development by predicting potential resistance mutations. Therefore, we next investigated whether DIvERGE can mutagenize longer genomic regions with a comparable resolution, and thereby, whether resistance processes for other drug classes can be evaluated successfully. Importantly, other methods for *in vivo* mutagenesis do not enable an exhaustive combinatorial mutagenesis of genomic loci longer than a few hundred base pairs.

For our experiments, we have chosen DNA gyrase and topoisomerase IV inhibitor antibiotics. Studying resistance mechanisms against existing and novel DNA gyrase and topoisomerase IV inhibitors is of special importance, especially because 25% of all antibiotic drug candidates currently in clinical trials (11 out of the 44) target these bacterial proteins [95,171]. In fact, topoisomerase-targeting antibacterial drugs, namely the quinolone group, were introduced into the therapy in 1962 [171,172]. Since then they have become the most widely used synthetic antibiotics, and are also extensively studied as a potential source of novel antibiotic development. In Gram-negative bacteria, such as *E. coli*, the primary target of quinolone drugs is the enzyme DNA gyrase [172]. However, quinolone antibiotics also tend to have affinity to topoisomerase IV [173] which is a bacterial enzyme homologous to DNA gyrase. As the test compound, we have chosen ciprofloxacin, a fluoroquinolone drug widely used in clinical practice, and already known to be concerned by resistance [174,175]. Furthermore, as ciprofloxacin resistance poses a significant clinical challenge [91,176], ciprofloxacin seemed to be a rational target to compare the resistance processes discovered by DIvERGE in laboratory strains to real-life resistance processes resulting in actual, clinically-significant resistance mutants.

To seek mutations at the genomic targets that may influence the target-binding of ciprofloxacin, we mutagenized *gyrA, gyrB and parC, parE* loci along the full lengths of their protein coding sequences in *E. coli* K-12 MG1655. GyrA and GyrB proteins constitute a DNA

gyrase heterodimer, while ParC and ParE form topoisomerase IV. These 4 loci altogether are more than 16-times longer than *folA* mutagenized in our previous tests. To induce mutagenesis, we scaled up DIvERGE and performed a single round of ssDNA-recombineering in *E. coli* K-12 MG1655 by using a mixture of 130 soft-randomized oligos, covering a total of 9503 base pairs. The resulting mutant libraries were then subjected to ciprofloxacin stress using a ciprofloxacin concentration of two-fold higher than the MIC of ciprofloxacin (i.e. 3 ng/ml) in the wild-type *E. coli* K-12 MG1655. Next, by exploiting the precision and throughput of Pacific Biosciences Single Molecule Real-Time sequencing, the genotypes at all four loci in 3000 resistant clones were determined. Sequence analysis indicated that the majority of the identified mutants carried single mutations only. Mutations were detected in *gyrA* and *gyrB*, but were not present in *parC* and *parE*. Detected genotypes were dominated by mutations at Ser 83 and Asp 87, and their combinations, known to confer ciprofloxacin resistance in clinical practice [174,177]. Most notably, the analysis revealed a novel 46-aminoacid-long region of GyrB, which was found to be mutated in more than 22% of the analyzed mutants. To understand the role of this region on GyrB, we subsequently performed protein structure analyses which demonstrated that this region is in the close proximity of GyrA, and forms the binding site of fluoroquinolone antibiotics [178,179] (Figure 24).
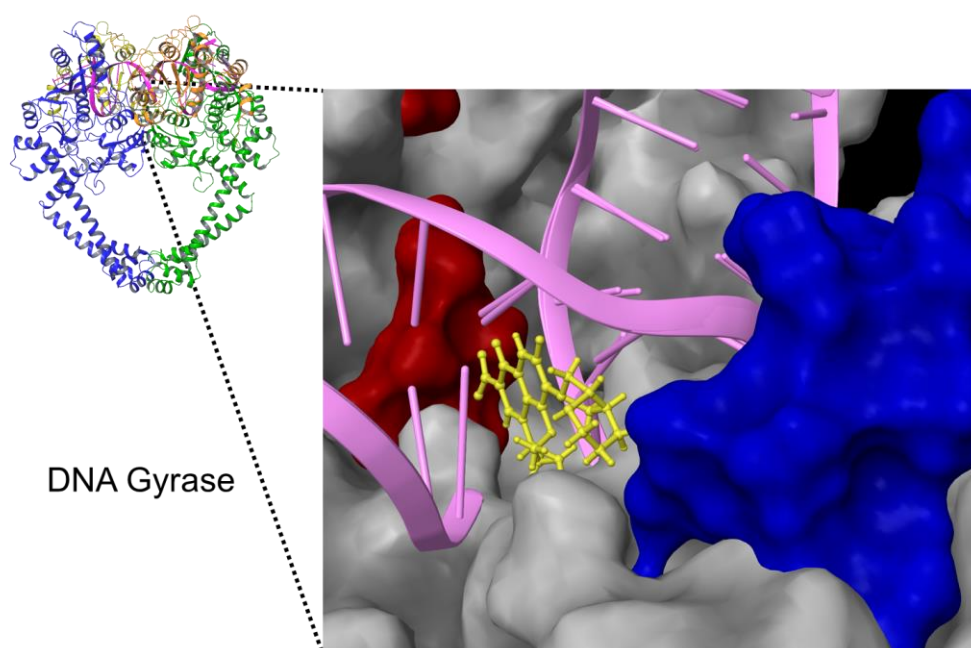
Figure 24. Map of DIvERGE-induced mutations conferring ciprofloxacin resistance at DNA gyrase protein complex, the primary target of ciprofloxacin in *E. coli* K-12 MG1655. Figure shows the detected mutational hot-spots based on simultaneous Pacific Biosciences Single Molecule Real-Time sequencing of the *gyrA* (red) and *gyrB* (blue) loci (Protein Data Bank ID: 5BS8). Furthermore, the fluoroquinolone drug molecule is illustrated in yellow and

the double strand DNA is shown in magenta. Mutated amino acid positions with a mutation frequency of over 0.5% were qualified as resistance-conferring mutations and were plotted on the crystal structure of the protein complex from *Mycobacterium tuberculosis* (Protein Data Bank identifier: 5BS8).

Also, it should be noted that the lack of resistance-conferring mutations at ParC and ParE is in-line with previous observations [175]. Mutations accounting for the first step of fluoroquinolone resistance are generally localized to the primary drug target, and mutations in ParC and ParE rely on epistatic interactions with mutations in GyrA [180]. Considering that for fluoroquinolone compounds, the interaction with either target enzyme is sufficient to block bacterial growth, the level of susceptibility of the wild-type bacterium is determined by the more sensitive target of these two [175,181]. In Gram-negative bacteria, mutations accounting for the first step of fluoroquinolone resistance are located at DNA gyrase, the drugs' primary target, while mutations concerning the less sensitive secondary target (i.e. topoisomerase IV) do not induce resistance on their own. This phenomenon is explained by the dominance of the interaction between the drug and its primary target [172]. Thus, in the hierarchical acquisition of high-level fluoroquinolone resistance, the secondary target's level of sensitivity sets a limit to the resistance level conferred by mutations in the primary target. As secondary mutations at the secondary target (topoisomerase IV) may occur only once the primary target is mutated, these secondary mutations are essential to abolish the bioactivity of fluoroquinolone antibiotics. To investigate this co-evolution of high-level fluoroquinolone resistance involving both protein targets, we are planning to further investigate this topic, and, based on our experimental findings with iterative DIvERGE-mutagenesis and trimethoprim selection cycles, we expect that DIvERGE will be able to explore multi-step resistance acquisition as well.

Analysis of resistance to an antibiotic currently in human clinical trials

Findings from our experiments with these two distinct classes of antibiotics indicate that DIvERGE-induced mutagenesis with subsequent antibiotic selection and mutant-genotyping can identify resistance mechanisms in a high-throughput manner. Moreover, it is capable of highlighting the binding site(s) of antibiotics. Therefore, we have investigated whether DIvERGE is applicable to identifying or predicting resistance to antibiotics with unexplored resistance mechanisms, or even in cases where the exact mode-of-action is to be clarified.

As a first step into this direction, we focused on gepotidacin (GSK2140944), an antibiotic candidate currently in human Phase II clinical trials (ClinicalTrials.gov registry

71

number: NCT02045797 and NCT02294682) [182–184]. It selectively inhibits bacterial DNA gyrase and topoisomerase IV via a unique mechanism of action, different from that of any other approved antibiotics [185]. A recent study has failed to identify resistant mutants to this drug in *Neisseria gonorrhoeae* [186], but this finding may only reflect the limitations of standard microbial assays for the detection of resistance processes. Thus, no literature data was available on the mutations that may cause resistance to gepotidacin in Gram-negative bacteria. To investigate this issue in details, we have attempted to generate resistance-conferring variants by exposing as many as ten billion wild-type *E. coli* K-12 MG1655 cells to gepotidacin stress according to standard frequency-of-resistance assays [187]. No resistant variants were observable after 72 hours. By contrast, when we subjected the four potential target gene loci of gepotidacin (*gyrA, gyrB and parC, parE*) to DIvERGE mutagenesis, and selected the resulting cell library on agar plates containing gepotidacin, we identified resistant clones in three days. All of these DIvERGE-generated *E. coli* variants showed a similar, ~560-fold decrease in gepotidacin susceptibility compared to the wild-type parental strain (Table 7). Sequence analyses of three independently isolated clones showed that the combination of only two specific mutations (GyrA Asp82→Asn and ParC Asp79→Asn) induced a high level of gepotidacin resistance. Moreover, the simultaneous introduction of the same mutation combination into the genomes of *E. coli* CFT073 (UPEC), *Citrobacter freundii* ATCC 8090, and *Klebsiella pneumoniae* ATCC 10031 revealed that these mutations together highly decrease susceptibility to gepotidacin in human pathogenic isolates as well.

Strain	Gepotidacin MIC (µg/ml)	Fold change in MIC compared to the corresponding wild-type
Escherichia coli CFT073	>150	>750
Citrobacter freundii ATCC 8090	>150	>330
Klebsiella pneumoniae ATCC 10031	125	2080

Table 7. Gepotidacin-susceptibility of human pathogenic bacteria containing the mutation combination of GyrA Asp82→Asn and ParC Asp79→Asn. The minimum inhibitory concentrations (MICs) of gepotidacin in the wild-type *E. coli* CFT073, *C. freundii* ATCC 8090, and *K. pneumoniae* ATCC 10031 equal to 200, 450, and 60 ng/mL, respectively.

These findings indicate the applicability of DIvERGE to explore rare combinations of resistance-causing mutations which would otherwise remain undetected with standard methods of laboratory evolution. We anticipate that our new method will be a useful tool for rapid resistance screening both at the early stages of drug design and during lead optimization of novel antibiotic candidates. Also, we hope that it could offer a practical support to identify novel molecular entities prone to resistance, and thus it could successfully direct lead optimization to mitigate drug-to-target interactions that are promoting resistance.

Discussion

Methods of bacterial genome engineering offer an unprecedented opportunity to systematically interrogate phenotype-to-genotype relationships. However, available tools for bacterial genome engineering suffer from limitations. Previous methods have been optimized for a few laboratory model strains only (such as *E. coli* K-12 MG1655, the "work-horse" strain of laboratory and biotechnological research) or demand the extensive modification of the host genome, labor-intensive cloning steps, or DNA synthesis prior to genome editing. Moreover, most methods lead to the accumulation of undesired, off-target modifications which, in extreme cases, may outnumber the desired edits and may mask the effect of intentional engineering. These issues have serious implications on the widespread applicability of genome engineering in both basic and applied research [24,29,130].

Our research work has specifically addressed these issues in consecutive projects.

Dominant mismatch repair control allows precise genome editing

As our first step, we aimed to extend the most high-throughput microbial genome engineering method (multiplex automated genome engineering, shortly MAGE [38]) towards straightforward applicability in multiple bacterial species. To achieve that, we first characterized a dominant mutation in MutL, a key protein [144] of the methyl-directed mismatch repair system, and utilized its dominant effect to precisely control mismatch-repair processes in target cells. We have demonstrated that this MutL E32→K mutant of *E. coli* rapidly abolishes mismatch repair activity even in the presence of the wild-type protein. By integrating this MutL E32→K variant into the MAGE workflow, we have developed a plasmid-based system, termed pORTMAGE, for genome engineering, and we have successfully demonstrated its applicability for high-throughput genome editing at multiple loci [130]. Importantly, whole-genome sequencing has revealed that the pORTMAGE-modified strains lack off-target mutations, which is a significant improvement compared to prior MAGE-based methods.

Conserved functionality across various strains and species

Exploiting the highly conserved nature of bacterial methyl-directed mismatch repair, the application of the dominant MutL E32→K provided a unique solution for the interspecies portability in MAGE. The dominant mutator phenotype of MutL E32→K was conserved across a diverse set of enterobacteria [130]. Also, follow-up studies after the introduction of

pORTMAGE pointed out that the very same mutation confers a dominant effect in *Vibrio cholerae*, thereby permitting efficient recombineering in this strain as well [188,189].

Finally, by placing the entire synthetic operon encoding all necessary elements of MAGE and mismatch repair control to a vector of broad host-range, we have successfully adapted ssDNA-recombineering to a wide range of biotechnologically and clinically relevant enterobacteria. In turn, pORTMAGE has allowed of the rapid generation of specific mutations, and thus the development of large, unbiased mutant libraries mutagenized at desired positions in these species.

The spectrum of species in which the functionality of pORTMAGE has been demonstrated so far enlists multiple strains including *Escherichia, Salmonella, Citrobacter,* and *Klebsiella* genera [125,130]. Moreover, since the introduction of pORTMAGE, other research laboratories and our follow-up collaborations have demonstrated that the rationale behind pORTMAGE is generally applicable. Until now, pORTMAGE has been implemented in more than 100 laboratories worldwide, and its utilization has contributed to studies that focus on clinical and biotechnological issues. Moreover, through to the discovery of novel recombinases which allow efficient genome editing in phylogenetically more distant species, we have also adapted ssDNA-recombineering to the biotechnologically important species *Pseudomonas putida* [190].

Systematic analysis of phenotype-to-genotype associations

The systematic investigation of genotype-to-phenotype associations for complex traits, however, still remained a significant challenge at that point, partly because the evolution of such evolutionary innovations frequently require the acquisition of multiple, rare mutations at the same time [69,71]. In turn, the availability of genome engineering tools which enable the targeted combinatorial mutagenesis of multiple loci is an inherent prerequisite for these goals to be met. However, current *in vivo* genome engineering and mutagenesis methods suffer from serious limitations in this regard. In most cases, the length and/or the number of targeted loci are severely limited. As a consequence, the attainable throughput and achievable library sizes are generally moderate. Also, certain methods target the entire genome unselectively, and in turn, lead to the accumulation of undesired, off-target modifications with detrimental side effects. Last but not least, most methods are limited to single microbial species or strains which prevents evolutionary investigations beyond model species [125,24].

Previous attempts to address these shortcomings rely on CRISPR/Cas-mediated genome engineering and CRISPR/Cas-guided base-editors to achieve high specificity [26,74,114,116]. Although they represent a major step forward, these techniques still produce

limited library sizes, frequently due to toxicity, and have limitations in terms of target sequence length and the number of simultaneous targets [26]. Moreover, they result in considerable bias regarding the types of mutations generated, and are all hindered by the availability of a protospacer adjacent motif (PAM) required for Cas recognition.

Directed evolution with random genomic mutations

To address the aforementioned shortcomings, we have advanced pORTMAGE-based recombineering to allow the systematic multiplex mutagenesis of long genomic segments. By building on the efficiency, multiplexability and throughput of multiplex automated genome engineering (MAGE) and the cross-species portability of pORTMAGE, as well as by combining these with a method for mutagenic chemical DNA synthesis, we have managed to advance recombineering to meet specific expected criteria. Thus, this new method termed directed evolution with random genomic mutations or DIvERGE for short (I) targets multiple, user-defined genomic regions, (II) has a broad and controllable mutagenesis spectrum for each nucleotide position, (III) allows of up to a million-fold increase in mutation rate at the target sequence, (IV) enables multiple rounds of mutagenesis and selection in a fast and continuous way, (V) is applicable to a wide range of enterobacterial species without the need for prior genomic modification(s), (VI) avoids off-target mutagenesis, and (VII) is also cost-effective, as it relies on soft-randomized oligos which can easily be manufactured at a modest cost. In summary, DIvERGE offers a versatile solution for high-precision directed evolution at multiple loci in their native genomic context. As a significant advantage, it utilizes soft-randomized ssDNA oligos, coupled to pORTMAGE-based ssDNA-recombineering, to allow for the in-depth exploration of the sequence space [132,125].

Soft-randomized oligos randomize extended targets

First, we have demonstrated that DIvERGE can mutagenize multiple, distinct genomic segments at nucleotide level precision, without affecting non-targeted regions (Figure 13 and 15). A unique application of soft-randomized oligos has enabled us to extend the target sequence undergoing mutagenesis up to 87% of the length of an entire oligonucleotide, by using only a single oligo. Next, the partially overlapping design of such oligos, as well as the coverage of entire genes and their regulatory regions permitted rapid protein engineering through random mutagenesis (Figure 17). Importantly, DIvERGE has the advantage to place no constraint on the target sequence regarding the availability of a protospacer adjacent motif, which is a strict requirement for CRISPR/Cas-based mutagenesis techniques.

Precise control of mutagenesis

Second, modifications of the parameters for soft-randomized oligo synthesis have made it possible to execute an unbiased introduction of mutations at each targeted nucleotide position (Table 3), resulting in a more comprehensive generation of combinatorial mutants. This is a significant advantage of DIvERGE over techniques where the mutational spectrum is biased, such as those relying on CRISPR/Cas-directed base-editors or *E. coli* DNA polymerase I, or other recombineering-based methods [115,116,119]. Besides the types of introduced sequence alterations, the rate of mutagenesis can also be precisely adjusted by controlling the parameters and composition of soft-randomization of the oligos during synthesis, as well as by the number of iterative DIvERGE cycles. Thereby, mutation rate can be upregulated to achieve an increase of up to a million-fold compared to wild-type mutation rates at the targeted loci, a range exceeding that of most *in vivo* methods.

Iterative cycles of recombineering accelerate laboratory evolution

Third, we have demonstrated that DIvERGE can be performed iteratively using the same oligo pools which are designed at the beginning of a given experiment. Thereby it permits multiple rounds of directed evolution (consisting of iterative mutagenesis and selection steps), which has been demonstrated to facilitate the rapid attainment of bacterial variants highly resistant to trimethoprim in our case. By performing multi-round directed evolution of *folA*, we have successfully demonstrated the rapid generation of variants containing up to 10 mutations. This feature is particularly important, as it facilitates the combination of independently generated mutations whose co-occurrence would normally be highly unlikely under laboratory conditions, and would require time-consuming laboratory evolution protocols [71,162,165,166]. Thereby, DIvERGE accelerates the laboratory evolution of slowly evolving traits.

Broad-host mutagenesis identifies strain-specific mutational effects

Fourth, we have demonstrated the portability of DIvERGE by mutagenizing multiple enterobacterial species, including biotechnologically and clinically relevant organisms. DIvERGE has enabled us to explore the *in vivo* evolution of drug resistance in pathogen bacteria, in a much faster and more comprehensive manner compared to prior techniques. Using DIvERGE, we have identified numerous previously undetected resistance-conferring mutations. Moreover, we have also demonstrated that phenotypic effects of certain

77

trimethoprim resistance-associated mutations vary considerably across phylogenetically related strains. Despite the 99% sequence similarity of *folA* between *E. coli* K-12 MG1655 and the uropathogenic strain *E. coli* CFT073, the relative resistance level induced by the very same mutation differed between the two strains. One of the identified mutations, FolA Ala7→Ser was associated with a decreased trimethoprim susceptibility only in the uropathogenic isolate. Although the reason for this strain-specific effect is yet to be explored, it might be explained by the different mRNA stability-effects of the same mutation within the two strains. These results also have implications on the utility of previous laboratory evolutionary experiments of drug resistance which were performed in non-pathogenic, laboratory model strains. Also, these results have shed light on the possibility on strain-specific antibiotic resistance processes.

High-throughput recombineering is well suited to analyze antibiotic resistance evolution

Together, pORTMAGE and DIvERGE are paving the way towards the high-throughput *in vivo* exploration of fitness landscapes of endogenous genes or gene networks in multiple species. In a direct clinically relevant application, these methods have allowed us to explore the evolutionary routes which induce target-specific antibiotic resistance in multiple bacterial species. As the new method executes an exceptionally high mutation rate at the drug's target, and also lacks mutational biases, using DIvERGE we have comprehensively generated resistance-conferring mutations at specific drug targets. We have successfully demonstrated that DIvERGE is capable of rapidly identifying mutations which contribute to the development of resistance against three distinct antibiotic classes. Utilizing DIVERGE we have identified several previously described mutations conferring resistance against trimethoprim and fluoroquinolone antibiotics: the predominant *folA* and *gyrA* mutations evolved by DIvERGE have already been observed to arise in *E. coli* under laboratory settings, and have also been reported in the clinical practice in patients treated with trimethoprim or fluoroquinolone medications [174,181]. Moreover, using DIvERGE we have also revealed several previously unknown mutations conferring trimethoprim resistance.

By exploiting the ability of DIvERGE to explore the combinatorial mutational space at long genomic segments, we have identified a combination of two specific mutations at *gyrA* and *parC* that might lead to clinically significant resistance against a novel antibiotic that is under clinical development (gepotidacin, GSK2140944) in the near future. This higher sensitivity of DIvERGE has also contributed to the detection of resistance-mutations to antibiotics which have previously been described as resistance-proof based on conventional techniques (i.e.

gepotidacin). Importantly, as DIvERGE can mutagenize a large fraction of genes that are targets of antibiotics currently being developed or clinically utilized, it could be a useful tool for the massively parallel discovery of mutations conferring resistance against multiple classes of antimicrobial agents, and it could be executed in a faster way and with higher resolution compared to alternative methods.

Conclusion

To address the shortcomings of currently available genome editing and *in vivo* directed evolution techniques, we have developed a plasmid-based method for broad host-range genome engineering (pORTMAGE), and based on pORTMAGE, a method for *in vivo* directed evolution. This new method, termed DIvERGE (directed evolution with random genomic mutations) allows the systematic multiplex mutagenesis of long genomic segments. DIvERGE has numerous advantages over the alternative techniques, including (I) the possibility to target multiple, user-defined genomic regions; (II) it has a broad and controllable mutagenesis spectrum for each nucleotide position; (III) it allows of up to a million-fold increase in mutation rate at the target sequence; (IV) it enables multiple rounds of mutagenesis and selection in a fast and continuous manner; (V) it is applicable to a wide range of enterobacterial species without the need for prior genomic modification(s); (VI) it avoids off-target mutagenesis, and (VII) it is also cost-effective as it relies on soft-randomized oligos which can easily be manufactured at a modest cost. In summary, DIvERGE offers a versatile solution for high-precision directed evolution at multiple loci in their native genomic context. Due to these favorable characteristics, DIvERGE is especially well-suited to study bacterial evolution leading to antibiotic resistance.

The standard protocols currently used by pharmaceutical companies to predict the speed at which antibiotic resistance would emerge against a new drug candidate are frequently inadequate and give inaccurate information because of their limited throughput to explore the sequence space. This is a significant problem, as antibiotics developers are frequently plagued by the waste of considerable efforts and funding as rapid resistance evolves to new antimicrobial agents. Clearly, high-throughput mutagenesis platforms are required to identify antibiotics with the promise of extended clinical efficacy, and these should be utilized from the very early stage of drug development. DIvERGE implements a significant improvement in this direction. Although the genes for targeted mutagenesis need to be defined in advance, once these genes of interest are known, DIvERGE allows of a reliable exploration and prediction of resistance processes in clinically relevant pathogens. Coupled to subsequent fitness and virulence assays, DIvERGE offers an opportunity to identify antimicrobial agents with a

potential of extended clinical efficacy, even at an early stage of drug development. Regarding this direction, our experiments have already demonstrated the feasibility of resistance analyses for novel antimicrobial agents [125,191].

In the future, we hope that our DIvERGE strategy would be adaptable to other organisms, including yeasts and mammalian cells, in which ssDNA-mediated recombineering would be expected to function efficiently. As a first step in this direction, we have already developed a modified version of DIvERGE which is applicable to the eukaryotic species *Saccharomyces cerevisiae* [132].

Personal contribution

On the course of projects that are described in this PhD dissertation, Akos Nyerges designed, coordinated and supervised experiments, interpreted experimental data, and conceived the concepts behind pORTMAGE and DIvERGE. He performed the construction of pORTMAGE plasmids and evaluated its performance in *E. coli*, constructed and integrated the landing pad sequence into *E. coli* K-12 MG1655, *S. enterica,* and *C. freundii*, and performed the characterization of recombineering-efficiency at these landing pads. On the course of DIvERGE-experiments, he performed the preparation of sequencing samples for Illumina and Pacific Biosciences Single Molecule Real-Time sequencing. DNA sequencing and data analysis protocols for Illumina DNA sequencing were developed jointly with Péter Bihari, Balázs Bálint, Bálint Márk Vásárhelyi, and István Nagy from Seqomics Ltd, while Pacific Biosciences Single Molecule Real-Time sequencing projects were performed jointly with Ave Tooming-Klunderud, Eszter Ari, and Balázs Bálint. DIvERGE oligonucleotides and other DNA oligonucleotides that were utilized in these studies were manufactured by Györgyi Ferenc at the DNA Synthesis Laboratory of the Biological Research Centre, or were ordered from commercial vendors.

References

1. Szybalski W: In vivo and in vitro initiation of transcription. *Adv Exp Med Biol* 1974, 44:23–24.

2. Cameron DE, Bashor CJ, Collins JJ: A brief history of synthetic biology. *Nat Rev Micro* 2014, advance online publication.

3. www.parts.igem.org

4. Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, Leavell MD, Tai A, Main A, Eng D, et al.: High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* 2013, 496:528–532.

5. Hubbard BP, Badran AH, Zuris JA, Guilinger JP, Davis KM, Chen L, Tsai SQ, Sander JD, Joung JK, Liu DR: Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nat Meth* 2015, 12:939–942.

6. Dickinson BC, Packer MS, Badran AH, Liu DR: A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. *Nat Commun* 2014, 5:5352.

7. Butterfield GL, Lajoie MJ, Gustafson HH, Sellers DL, Nattermann U, Ellis D, Bale JB, Ke S, Lenz GH, Yehdego A, et al.: Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature* 2017, 552:415–420.

8. Burns JR, Lamarre B, Pyne ALB, Noble JE, Ryadnov MG: DNA Origami Inside-Out Viruses. *ACS Synth Biol* 2018, 7:767–773.

9. Farzadfard F, Lu TK: Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science* 2014, 346:1256272.

10. Kalhor R, Mali P, Church GM: Rapidly evolving homing CRISPR barcodes. *Nature Methods* 2017, 14:195–200.

11. Esvelt KM, Wang HH: Genome-scale engineering for systems and synthetic biology. *Molecular Systems Biology* 2013, 9.

12. Porcar M, Danchin A, Lorenzo V de, Santos VA dos, Krasnogor N, Rasmussen S, Moya A: The ten grand challenges of synthetic life. *Syst Synth Biol* 2011, 5:1–9.

13. Halweg-Edwards AL, Grau WC, Winkler JD, Garst AD, Gill RT: The emergence of commodity-scale genetic manipulation. *Current Opinion in Chemical Biology* 2015, 28:150–155.

14. Hutchison CA, Chuang R-Y, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, Gill J, Kannan K, Karas BJ, Ma L, et al.: Design and synthesis of a minimal bacterial genome. *Science* 2016, 351:aad6253.

15. Haimovich AD, Muir P, Isaacs FJ: Genomes by design. *Nat Rev Genet* 2015, 16:501–516.

16. Cello J, Paul AV, Wimmer E: Chemical Synthesis of Poliovirus cDNA: Generation of Infectious Virus in the Absence of Natural Template. *Science* 2002, 297:1016–1018.

17. Tumpey TM, Basler CF, Aguilar PV, Zeng H, Solórzano A, Swayne DE, Cox NJ, Katz JM, Taubenberger JK, Palese P, et al.: Characterization of the Reconstructed 1918 Spanish Influenza Pandemic Virus. *Science* 2005, 310:77–80.

18. Noyce RS, Lederman S, Evans DH: Construction of an infectious horsepox virus vaccine from chemically synthesized DNA fragments. *PLOS ONE* 2018, 13:e0188453.

19. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang R-Y, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, et al.: Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* 2010, 329:52–56.

20. Lartigue C, Glass JI, Alperovich N, Pieper R, Parmar PP, Hutchison CA, Smith HO, Venter JC: Genome Transplantation in Bacteria: Changing One Species to Another. *Science* 2007, 317:632–638.

21. Richardson SM, Mitchell LA, Stracquadanio G, Yang K, Dymond JS, DiCarlo JE, Lee D, Huang CLV, Chandrasegaran S, Cai Y, et al.: Design of a synthetic yeast genome. *Science* 2017, 355:1040–1044.

22. Ostrov N, Landon M, Guell M, Kuznetsov G, Teramoto J, Cervantes N, Zhou M, Singh K, Napolitano MG, Moosburner M, et al.: Design, synthesis, and testing toward a 57-codon genome. *Science* 2016, 353:819–822.

23. VogelOct. 3 G, 2018, Am 5:00: 'Revolution based on evolution' honored with chemistry Nobel. *Science | AAAS* 2018,

24. Csörgő B, Nyerges Á, Pósfai G, Fehér T: System-level genome editing in microbes. *Current Opinion in Microbiology* 2016, 33:113–122.

25. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al.: Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* 2013, 339:819–823.

26. Garst AD, Bassalo MC, Pines G, Lynch SA, Halweg-Edwards AL, Liu R, Liang L, Wang Z, Zeitoun R, Alexander WG, et al.: Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat Biotech* 2017, 35:48–55.

27. Packer MS, Liu DR: Methods for the directed evolution of proteins. *Nat Rev Genet* 2015, 16:379–394.

28. Kelwick R, MacDonald JT, Webb AJ, Freemont P: Developments in the tools and methodologies of synthetic biology. *Front Bioeng Biotechnol* 2014, 2:60.

29. Pál C, Papp B, Pósfai G: The dawn of evolutionary genome engineering. *Nat Rev Genet* 2014, advance online publication.

30. Gallagher RR, Li Z, Lewis AO, Isaacs FJ: Rapid editing and evolution of bacterial genomes using libraries of synthetic DNA. *Nat Protocols* 2014, 9:2301–2316.

31. Bonde MT, Kosuri S, Genee HJ, Sarup-Lytzen K, Church GM, Sommer MOA, Wang HH: Direct Mutagenesis of Thousands of Genomic Targets using Microarray-derived Oligonucleotides. *ACS Synth Biol* 2014, doi:10.1021/sb5001565.

32. Oppenheim AB, Rattray AJ, Bubunenko M, Thomason LC, Court DL: In vivo recombineering of bacteriophage λ by PCR fragments and single-strand oligonucleotides. *Virology* 2004, 319:185–189.

33. Swaminathan S, Ellis HM, Waters LS, Yu D, Lee E-C, Court DL, Sharan SK: Rapid engineering of bacterial artificial chromosomes using oligonucleotides. *Genesis* 2001, 29:14–21.

34. Yu D, Ellis HM, Lee E-C, Jenkins NA, Copeland NG, Court DL: An efficient recombination system for chromosome engineering in Escherichia coli. *PNAS* 2000, 97:5978–5983.

35. Rodrigues AL, Trachtmann N, Becker J, Lohanatha AF, Blotenberg J, Bolten CJ, Korneli C, de Souza Lima AO, Porto LM, Sprenger GA, et al.: Systems metabolic engineering of Escherichia coli for production of the antitumor drugs violacein and deoxyviolacein. *Metabolic Engineering* 2013, 20:29–41.

36. Posfai G, Kolisnychenko V, Bereczki Z, Blattner FR: Markerless gene replacement in Escherichia coli stimulated by a double-strand break in the chromosome. *Nucleic Acids Research* 1999, 27:4409–4415.

37. Ellis HM, Yu D, DiTizio T, Court DL: High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *PNAS* 2001, 98:6742–6746.

38. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM: Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 2009, 460:894–898.

39. Sergueev K, Court D, Reaves L, Austin S: E. coli Cell-cycle Regulation by Bacteriophage Lambda. *Journal of Molecular Biology* 2002, 324:297–307.

40. Sawitzke JA, Costantino N, Li X-T, Thomason LC, Bubunenko M, Court C, Court DL: Probing cellular processes with oligo-mediated recombination and using the knowledge gained to optimize recombineering. *J Mol Biol* 2011, 407:45–59.

41. Datta S, Costantino N, Zhou X, Court DL: Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *PNAS* 2008, 105:1626–1631.

42. Yu D, Sawitzke JA, Ellis H, Court DL: Recombineering with overlapping single-stranded DNA oligonucleotides: Testing a recombination intermediate. *PNAS* 2003, 100:7207–7212.

43. Wang HH, Church GM: Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering. *Meth Enzymol* 2011, 498:409–426.

44. Oesterle S, Gerngross D, Schmitt S, Roberts TM, Panke S: Efficient engineering of chromosomal ribosome binding site libraries in mismatch repair proficient Escherichia coli. *Scientific Reports* 2017, 7:12327.

45. Jeschek M, Gerngross D, Panke S: Rationally reduced libraries for combinatorial pathway optimization minimizing experimental effort. *Nature Communications* 2016, 7:11163.

46. Diner EJ, Hayes CS: Recombineering reveals a diverse collection of ribosomal proteins L4 and L22 that confer resistance to macrolide antibiotics. *J Mol Biol* 2009, 386:300–315.

47. Amiram M, Haimovich AD, Fan C, Wang Y-S, Aerni H-R, Ntai I, Moonan DW, Ma NJ, Rovner AJ, Hong SH, et al.: Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids. *Nat Biotech* 2015, 33:1272–1279.

48. Higgins SA, Ounkap S, Savage DF: Rapid and Programmable Protein Mutagenesis Using Plasmid Recombineering. *ACS Synth Biol* 2017, doi:10.1021/acssynbio.7b00112.

49. Pu Y, Dong C, Hang B, Huang L, Cai J, Xu Z: Novel approach for the evolution of pyrroloquinoline quinone glucose dehydrogenase by multiplex-site in situ engineering. *Process Biochemistry* [date unknown], doi:10.1016/j.procbio.2016.08.003.

50. Lajoie MJ, Kosuri S, Mosberg JA, Gregg CJ, Zhang D, Church GM: Probing the Limits of Genetic Recoding in Essential Genes. *Science* 2013, 342:361–363.

51. Lajoie MJ, Rovner AJ, Goodman DB, Aerni H-R, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH, Carr PA, Mosberg JA, et al.: Genomically Recoded Organisms Expand Biological Functions. *Science* 2013, 342:357–360.

52. Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen AC, Gianoulis TA, Goodman DB, Reppas NB, et al.: Precise Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement. *Science* 2011, 333:348–353.

53. Penewit K, Holmes EA, McLean K, Ren M, Waalkes A, Salipante SJ: Efficient and Scalable Precision Genome Editing in Staphylococcus aureus through Conditional Recombineering and CRISPR/Cas9-Mediated Counterselection. *mBio* 2018, 9:e00067-18.

54.   Pijkeren J-P van, Britton RA: High efficiency recombineering in lactic acid bacteria. *Nucl Acids Res* 2012, 40:e76–e76.

55.   van Pijkeren J-P, Neoh KM, Sirias D, Findley AS, Britton RA: Exploring optimization parameters to increase ssDNA recombineering in Lactococcus lactis and Lactobacillus reuteri. *Bioengineered* 2012, 3:209–217.

56.   Binder S, Siedler S, Marienhagen J, Bott M, Eggeling L: Recombineering in Corynebacterium glutamicum combined with optical nanosensors: a general strategy for fast producer strain generation. *Nucl Acids Res* 2013, 41:6360–6369.

57.   Datta S, Costantino N, Court DL: A set of recombineering plasmids for gram-negative bacteria. *Gene* 2006, 379:109–115.

58.   Sun Z, Deng A, Hu T, Wu J, Sun Q, Bai H, Zhang G, Wen T: A high-efficiency recombineering system with PCR-based ssDNA in Bacillus subtilis mediated by the native phage recombinase GP35. *Appl Microbiol Biotechnol* 2015, 99:5151–5162.

59.   Bao Z, Cartinhour S, Swingle B: Substrate and Target Sequence Length Influence RecTEPsy Recombineering Efficiency in Pseudomonas syringae. *PLoS ONE* 2012, 7:e50617.

60.   Aparicio T, Jensen SI, Nielsen AT, de Lorenzo V, Martínez-García E: The Ssr protein (T1E_1405) from Pseudomonas putida DOT-T1E enables oligonucleotide-based recombineering in platform strain P. putida EM42. *Biotechnology Journal* 2016, doi:10.1002/biot.201600317.

61.   Costantino N, Court DL: Enhanced levels of λ Red-mediated recombinants in mismatch repair mutants. *PNAS* 2003, 100:15748–15753.

62.   Wang HH, Xu G, Vonner AJ, Church G: Modified bases enable high-efficiency oligonucleotide-mediated allelic replacement via mismatch repair evasion. *Nucleic Acids Research* 2011, 39:7336–7347.

63.   Junop MS, Yang W, Funchain P, Clendenin W, Miller JH: In vitro and in vivo studies of MutS, MutL and MutH mutants: correlation of mismatch repair and DNA recombination. *DNA Repair* 2003, 2:387–405.

64.   Putnam CD: Evolution of the methyl directed mismatch repair system in Escherichia coli. *DNA Repair* 2016, 38:32–41.

65.   Napolitano MG, Landon M, Gregg CJ, Lajoie MJ, Govindarajan L, Mosberg JA, Kuznetsov G, Goodman DB, Vargas-Rodriguez O, Isaacs FJ, et al.: Emergent rules for codon choice elucidated by editing rare arginine codons in Escherichia coli. *PNAS* 2016, 113:E5588–E5597.

66.   Wannier TM, Kunjapur AM, Rice DP, McDonald MJ, Desai MM, Church GM: Adaptive evolution of genomically recoded Escherichia coli. *PNAS* 2018, doi:10.1073/pnas.1715530115.

67. Nyerges Á, Csörgő B, Nagy I, Latinovics D, Szamecz B, Pósfai G, Pál C: Conditional DNA repair mutants enable highly precise genome engineering. *Nucleic Acids Res* 2014, 42:e62–e62.

68. Lennen RM, Nilsson Wallin AI, Pedersen M, Bonde M, Luo H, Herrgård MJ, Sommer MOA: Transient overexpression of DNA adenine methylase enables efficient and mobile genome engineering with reduced off-target effects. *Nucl Acids Res* 2015, doi:10.1093/nar/gkv1090.

69. Barrick JE, Lenski RE: Genome dynamics during experimental evolution. *Nat Rev Genet* 2013, 14:827–839.

70. Papp B, Notebaart RA, Pál C: Systems-biology approaches for predicting genomic evolution. *Nat Rev Genet* 2011, 12:591–602.

71. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF: Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature* 2009, 461:1243–1247.

72. Elena SF, Lenski RE: Microbial genetics: Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Reviews Genetics* 2003, 4:457–469.

73. Szappanos B, Fritzemeier J, Csörgő B, Lázár V, Lu X, Fekete G, Bálint B, Herczeg R, Nagy I, Notebaart RA, et al.: Adaptive evolution of complex innovations through stepwise metabolic niche expansion. *Nat Commun* 2016, 7:11607.

74. Ma Y, Zhang J, Yin W, Zhang Z, Song Y, Chang X: Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nature Methods* 2016, 13.

75. Sadhu MJ, Bloom JS, Day L, Siegel JJ, Kosuri S, Kruglyak L: Highly parallel genome variant engineering with CRISPR–Cas9. *Nature Genetics* 2018, 50:510–514.

76. Abbott TR, Qi LS: Evolution at the Cutting Edge: CRISPR-Mediated Directed Evolution. *Molecular Cell* 2018, 72:402–403.

77. Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M, Janizek JD, Huang X, Starita LM, Shendure J: Accurate functional classification of thousands of BRCA1 variants with saturation genome editing. *bioRxiv* 2018, doi:10.1101/294520.

78. d'Oelsnitz S, Ellington A: Continuous directed evolution for strain and protein engineering. *Current Opinion in Biotechnology* 2018, 53:158–163.

79. Bloom JD, Arnold FH: In the light of directed evolution: Pathways of adaptive protein evolution. *PNAS* 2009, 106:9995–10000.

80. Boyle NR, Gill RT: Tools for genome-wide strain design and construction. *Current Opinion in Biotechnology* 2012, 23:666–671.

81. Kosuri S, Church GM: Large-scale de novo DNA synthesis: technologies and applications. *Nat Meth* 2014, 11:499–507.

82. Dai Z, Liu Y, Guo J, Huang L, Zhang X: Yeast synthetic biology for high-value metabolites. *FEMS Yeast Res* 2015, 15:1–11.

83. Liu R, Bassalo MC, Zeitoun RI, Gill RT: Genome scale engineering techniques for metabolic engineering. *Metabolic Engineering* 2015, 32:143–154.

84. Umenhoffer K, Draskovits G, Nyerges Á, Karcagi I, Bogos B, Tímár E, Csörgő B, Herczeg R, Nagy I, Fehér T, et al.: Genome-Wide Abolishment of Mobile Genetic Elements Using Genome Shuffling and CRISPR/Cas-Assisted MAGE Allows the Efficient Stabilization of a Bacterial Chassis. *ACS Synth Biol* 2017, 6:1471–1483.

85. Umenhoffer K, Fehér T, Balikó G, Ayaydin F, Pósfai J, Blattner FR, Pósfai G: Research Reduced evolvability of Escherichia coli MDS42, an IS-less cellular chassis for molecular and synthetic biology applications. *Microbial cell factories* 2010, 9.

86. Vickers CE: The minimal genome comes of age. *Nat Biotech* 2016, 34:623–624.

87. Firnberg E, Labonte JW, Gray JJ, Ostermeier M: A Comprehensive, High-Resolution Map of a Gene's Fitness Landscape. *Mol Biol Evol* 2014, 31:1581–1592.

88. Steinberg B, Ostermeier M: Environmental changes bridge evolutionary valleys. *Science Advances* 2016, 2:e1500921.

89. Ravikumar A, Arzumanyan GA, Obadi MKA, Javanpour AA, Liu CC: Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell* 2018, 0.

90. Kelsic ED, Chung H, Cohen N, Park J, Wang HH, Kishony R: RNA Structural Determinants of Optimal Codons Revealed by MAGE-Seq. *cels* 2016, 3:563-571.e6.

91. Organization WH: *Critically important antimicrobials for human medicine: ranking of antimicrobial agents for risk management of antimicrobial resistance due to non-human use.* World Health Organization; 2017.

92. Cassini A, Högberg LD, Plachouras D, Quattrocchi A, Hoxha A, Simonsen GS, Colomb-Cotinat M, Kretzschmar ME, Devleesschauwer B, Cecchini M, et al.: Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *The Lancet Infectious Diseases* 2018, 0.

93. Hughes D, Andersson DI: Evolutionary Trajectories to Antibiotic Resistance. *Annual Review of Microbiology* 2017, 71:579–596.

94. Jim O'Neill: The Review on Antimicrobial Resistance

95. The Pew Charitable Trusts: Antibiotics Currently in Clinical Development

96. Jackson N, Czaplewski L, Piddock LJV: Discovery and development of new antibacterial drugs: learning from experience? *J Antimicrob Chemother*, doi:10.1093/jac/dky019.

97. DRIVE-AB Final Report 2018.

98. Silver LL: Multi-targeting by monotherapeutic antibacterials. *Nat Rev Drug Discov* 2007, 6:41–55.

99. Crofts TS, Gasparrini AJ, Dantas G: Next-generation approaches to understand and combat the antibiotic resistome. *Nature Reviews Microbiology* 2017, 15:422.

100. Hughes D, Andersson DI: Evolutionary consequences of drug resistance: shared principles across diverse targets and organisms. *Nat Rev Genet* 2015, 16:459–471.

101. Pál C, Papp B, Lázár V: Collateral sensitivity of antibiotic-resistant microbes. *Trends in Microbiology* [date unknown], doi:10.1016/j.tim.2015.02.009.

102. Hughes D, Andersson DI: Evolutionary consequences of drug resistance: shared principles across diverse targets and organisms. *Nat Rev Genet* 2015, 16:459–471.

103. Bell G, MacLean C: The Search for 'Evolution-Proof' Antibiotics. *Trends in Microbiology* 2017, 0.

104. Martínez JL, Baquero F, Andersson DI: Predicting antibiotic resistance. *Nat Rev Micro* 2007, 5:958–965.

105. Martínez JL, Baquero F, Andersson DI: Beyond serial passages: new methods for predicting the emergence of resistance to novel antibiotics. *Current Opinion in Pharmacology* 2011, 11:439–445.

106. Sommer MOA, Munck C, Toft-Kehler RV, Andersson DI: Prediction of antibiotic resistance: time for a new preclinical paradigm? *Nat Rev Micro* 2017, advance online publication.

107. O'Dwyer K, Spivak AT, Ingraham K, Min S, Holmes DJ, Jakielaszek C, Rittenhouse S, Kwan AL, Livi GP, Sathe G, et al.: Bacterial Resistance to Leucyl-tRNA Synthetase Inhibitor GSK2251052 Develops during Treatment of Complicated Urinary Tract Infections. *Antimicrob Agents Chemother* 2015, 59:289–298.

108. Hughes D, Karlén A: Discovery and preclinical development of new antibiotics. *Ups J Med Sci* 2014, 119:162–169.

109. O'Neill AJ, Chopra I: Use of Mutator Strains for Characterization of Novel Antimicrobial Agents. *Antimicrob Agents Chemother* 2001, 45:1599–1600.

110. Badran AH, Liu DR: Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nat Commun* 2015, 6:8425.

111. Nyerges Á, Csörgő B, Draskovits G, Kintses B, Szili P, Ferenc G, Révész T, Ari E, Nagy I, Bálint B, et al.: Directed evolution of multiple genomic loci allows the prediction of antibiotic resistance. *PNAS* 2018, doi:10.1073/pnas.1801646115.

112. Wong TS, Roccatano D, Zacharias M, Schwaneberg U: A statistical analysis of random mutagenesis methods used for directed protein evolution. *J Mol Biol* 2006, 355:858–871.

113. Tee KL, Wong TS: Polishing the craft of genetic diversity creation in directed evolution. *Biotechnology Advances* 2013, 31:1707–1721.

114. Hess GT, Frésard L, Han K, Lee CH, Li A, Cimprich KA, Montgomery SB, Bassik MC: Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat Meth* 2016, advance online publication.

115. Ravikumar A, Arzumanyan GA, Obadi MKA, Javanpour AA, Liu CC: Scalable continuous evolution of genes at mutation rates above genomic error thresholds. *bioRxiv* 2018, doi:10.1101/313338.

116. Halperin SO, Tou CJ, Wong EB, Modavi C, Schaffer DV, Dueber JE: CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature* 2018, doi:10.1038/s41586-018-0384-8.

117. Finney-Manchester SP, Maheshri N: Harnessing mutagenic homologous recombination for targeted mutagenesis in vivo by TaGTEAM. *Nucl Acids Res* 2013, 41:e99–e99.

118. Kim YB, Komor AC, Levy JM, Packer MS, Zhao KT, Liu DR: Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotech* 2017, advance online publication.

119. Camps M, Naukkarinen J, Johnson BP, Loeb LA: Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. *PNAS* 2003, 100:9727–9732.

120. Roy KR, Smith JD, Vonesch SC, Lin G, Tu CS, Lederer AR, Chu A, Suresh S, Nguyen M, Horecka J, et al.: Multiplexed precision genome editing with trackable genomic barcodes in yeast. *Nature Biotechnology* 2018, doi:10.1038/nbt.4137.

121. Guo X, Chavez A, Tung A, Chan Y, Kaas C, Yin Y, Cecchi R, Garnier SL, Kelsic ED, Schubert M, et al.: High-throughput creation and functional profiling of DNA sequence variant libraries using CRISPR–Cas9 in yeast. *Nature Biotechnology* 2018, doi:10.1038/nbt.4147.

122. Ma L, Boucher JI, Paulsen J, Matuszewski S, Eide CA, Ou J, Eickelberg G, Press RD, Zhu LJ, Druker BJ, et al.: CRISPR-Cas9–mediated saturated mutagenesis screen predicts clinical drug resistance with improved accuracy. *PNAS* 2017, doi:10.1073/pnas.1708268114.

123. Donovan KF, Hegde M, Sullender M, Vaimberg EW, Johannessen CM, Root DE, Doench JG: Creation of Novel Protein Variants with CRISPR/Cas9-Mediated

Mutagenesis: Turning a Screening By-Product into a Discovery Tool. *PLOS ONE* 2017, 12:e0170445.

124. Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, Tabata M, Mochizuki M, Miyabe A, Araki M, Hara KY, et al.: Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 2016, doi:10.1126/science.aaf8729.

125. Nyerges Á, Csörgő B, Draskovits G, Kintses B, Szili P, Ferenc G, Révész T, Ari E, Nagy I, Bálint B, et al.: Directed evolution of multiple genomic loci allows the prediction of antibiotic resistance. *PNAS* 2018, 115:E5726–E5735.

126. Wang HH, Kim H, Cong L, Jeong J, Bang D, Church GM: Genome-scale promoter engineering by coselection MAGE. *Nature Methods* 2012, 9:591–593.

127. Sandoval NR, Kim JYH, Glebes TY, Reeder PJ, Aucoin HR, Warner JR, Gill RT: Strategy for directing combinatorial genome engineering in Escherichia coli. *Proceedings of the National Academy of Sciences* 2012, 109:10540–10545.

128. Kelsic ED, Chung H, Cohen N, Park J, Wang HH, Kishony R: RNA Structural Determinants of Optimal Codons Revealed by MAGE-Seq. *cels* 2016, 3:563-571.e6.

129. Nordwald EM, Garst A, Gill RT, Kaar JL: Accelerated protein engineering for chemical biotechnology via homologous recombination. *Current Opinion in Biotechnology* 2013, 24:1017–1022.

130. Nyerges Á, Csörgő B, Nagy I, Bálint B, Bihari P, Lázár V, Apjok G, Umenhoffer K, Bogos B, Pósfai G, et al.: A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *PNAS* 2016, doi:10.1073/pnas.1520040113.

131. Nyerges Á, Csörgő B, Draskovits G, Kintses B, Szili P, Ferenc G, Révész T, Ari E, Nagy I, Bálint B, et al.: Directed evolution of multiple genomic loci allows the prediction of antibiotic resistance. *PNAS* 2018, 115:E5726–E5735.

132. Nyerges AJ, Pal C, Csorgo B, Kintses B: Mutagenizing Intracellular Nucleic Acids. 2018,

133. Gallagher RR, Patel JR, Interiano AL, Rovner AJ, Isaacs FJ: Multilayered genetic safeguards limit growth of microorganisms to defined environments. *Nucl Acids Res* 2015, doi:10.1093/nar/gku1378.

134. Lutz R, Bujard H: Independent and Tight Regulation of Transcriptional Units in Escherichia Coli Via the LacR/O, the TetR/O and AraC/I1-I2 Regulatory Elements. *Nucl Acids Res* 1997, 25:1203–1210.

135. Schirmer M, D'Amore R, Ijaz UZ, Hall N, Quince C: Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics* 2016, 17:125.

136. Chen L, Liu P, Evans TC, Ettwiller LM: DNA damage is a major cause of sequencing errors, directly confounding variant identification. *bioRxiv* 2016, doi:10.1101/070334.

137. Rhoads A, Au KF: PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics* [date unknown], doi:10.1016/j.gpb.2015.08.002.

138. Gabrieli T, Sharim H, Fridman D, Arbib N, Michaeli Y, Ebenstein Y: Selective nanopore sequencing of human BRCA1 by Cas9-assisted targeting of chromosome segments (CATCH). *Nucleic Acids Res* [date unknown], doi:10.1093/nar/gky411.

139. Hebert PD, Braukmann TW, Prosser SW, Ratnasingham S, deWaard JR, Ivanova NV, Janzen DH, Hallwachs W, Naik S, Sones JE, et al.: A Sequel to Sanger: Amplicon Sequencing That Scales. *bioRxiv* 2017, doi:10.1101/191619.

140. Warringer J, Blomberg A: Automated screening in environmental arrays allows analysis of quantitative phenotypic profiles in Saccharomyces cerevisiae. *Yeast* 2003, 20:53–67.

141. Karcagi I, Draskovits G, Umenhoffer K, Fekete G, Kovács K, Méhi O, Balikó G, Szappanos B, Györfy Z, Fehér T, et al.: Indispensability of Horizontally Transferred Genes and Its Impact on Bacterial Genome Streamlining. *Mol Biol Evol* 2016, 33:1257–1269.

142. ISO 20776-1:2006 - Clinical laboratory testing and in vitro diagnostic test systems -- Susceptibility testing of infectious agents and evaluation of performance of antimicrobial susceptibility test devices -- Part 1: Reference method for testing the in vitro activity of antimicrobial agents against rapidly growing aerobic bacteria involved in infectious diseases. *ISO* [date unknown],

143. Hall BM, Ma C-X, Liang P, Singh KK: Fluctuation AnaLysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics* 2009, 25:1564–1565.

144. Aronshtam A, Marinus MG: Dominant Negative Mutator Mutations in the mutL Gene of Escherichia Coli. *Nucl Acids Res* 1996, 24:2498–2504.

145. Nyerges Á, Csörgő B, Nagy I, Latinovics D, Szamecz B, Pósfai G, Pál C: Conditional DNA repair mutants enable highly precise genome engineering. *Nucl Acids Res* 2014, doi:10.1093/nar/gku105.

146. Antoine R, Locht C: Isolation and molecular characterization of a novel broad-host-range plasmid from Bordetella bronchiseptica with sequence similarities to plasmids from Gram-positive organisms. *Molecular Microbiology* 1992, 6:1785–1799.

147. Nyerges Á, Csörgő B, Nagy I, Latinovics D, Szamecz B, Pósfai G, Pál C: Conditional DNA repair mutants enable highly precise genome engineering. *Nucl Acids Res* 2014, 42:e62–e62.

148. Eyre-Walker A, Bulmer M: Synonymous Substitution Rates in Enterobacteria. *Genetics* 1995, 140:1407–1412.

149. Baumler DJ, Ma B, Reed JL, Perna NT: Inferring ancient metabolism using ancestral core metabolic models of enterobacteria. *BMC Systems Biology* 2013, 7:46.

150. Arunachalam TS, Wichert C, Appel B, Müller S: Mixed oligonucleotides for random mutagenesis: best way of making them. *Org Biomol Chem* 2012, 10:4641–4650.

151. Hermes JD, Parekh SM, Blacklow SC, Koster H, Knowles JR: A reliable method for random mutagenesis: the generation of mutant libraries using spiked oligodeoxyribonucleotide primers. *Gene* 1989, 84:143–151.

152. Kosuri S, Eroshenko N, LeProust EM, Super M, Way J, Li JB, Church GM: Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nature Biotechnology* 2010, 28:1295–1299.

153. Bennett BC, Wan Q, Ahmad MF, Langan P, Dealwis CG: X-ray structure of the ternary MTX·NADPH complex of the anthrax dihydrofolate reductase: A pharmacophore for dual-site inhibitor design. *Journal of Structural Biology* 2009, 166:162–171.

154. Eliopoulos GM, Huovinen P: Resistance to Trimethoprim-Sulfamethoxazole. *Clin Infect Dis* 2001, 32:1608–1614.

155. Rajagopalan PTR, Zhang Z, McCourt L, Dwyer M, Benkovic SJ, Hammes GG: Interaction of dihydrofolate reductase with methotrexate: Ensemble and single-molecule kinetics. *PNAS* 2002, 99:13481–13486.

156. Volpato JP, Pelletier JN: Mutational 'hot-spots' in mammalian, bacterial and protozoal dihydrofolate reductases associated with antifolate resistance: Sequence and structural comparison. *Drug Resistance Updates* 2009, 12:28–41.

157. Sader HS, Fritsche TR, Jones RN: Potency and Bactericidal Activity of Iclaprim against Recent Clinical Gram-Positive Isolates. *Antimicrob Agents Chemother* 2009, 53:2171–2175.

158. Brolund A, Sundqvist M, Kahlmeter G, Grape M: Molecular Characterisation of Trimethoprim Resistance in Escherichia coli and Klebsiella pneumoniae during a Two Year Intervention on Trimethoprim Use. *PLoS ONE* 2010, 5:e9233.

159. Al-Hasan MN, Lahr BD, Eckel-Passow JE, Baddour LM: Antimicrobial resistance trends of Escherichia coli bloodstream isolates: a population-based study, 1998–2007. *J Antimicrob Chemother* 2009, 64:169–174.

160. Flensburg J, Sköld O: Regulatory changes in the formation of chromosomal dihydrofolate reductase causing resistance to trimethoprim. *J Bacteriol* 1984, 159:184–190.

161. Flensburg J, Sköld O: Massive overproduction of dihydrofolate reductase in bacteria as a response to the use of trimethoprim. *European Journal of Biochemistry* 1987, 162:473–476.

162. Toprak E, Veres A, Michel J-B, Chait R, Hartl DL, Kishony R: Evolutionary paths to antibiotic resistance under dynamically sustained drug stress. *Nat Genet* 2011, 44:101–105.

163. Watson M, Liu J-W, Ollis D: Directed evolution of trimethoprim resistance in Escherichia coli. *FEBS Journal* 2007, 274:2661–2671.

164. Toprak E, Veres A, Michel J-B, Chait R, Hartl DL, Kishony R: Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nature Genetics* 2011, 44:101–105.

165. Baym M, Lieberman TD, Kelsic ED, Chait R, Gross R, Yelin I, Kishony R: Spatiotemporal microbial evolution on antibiotic landscapes. *Science* 2016, 353:1147–1151.

166. Palmer AC, Toprak E, Baym M, Kim S, Veres A, Bershtein S, Kishony R: Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes. *Nat Commun* 2015, 6:7385.

167. Lázár V, Nagy I, Spohn R, Csörgő B, Györkei Á, Nyerges Á, Horváth B, Vörös A, Busa-Fekete R, Hrtyan M, et al.: Genome-wide analysis captures the determinants of the antibiotic cross-resistance interaction network. *Nat Commun* 2014, 5.

168. Jin DJ, Gross CA: Mapping and sequencing of mutations in the Escherichia colirpoB gene that lead to rifampicin resistance. *Journal of Molecular Biology* 1988, 202:45–58.

169. Wisell KT, Kahlmeter G, Giske CG: Trimethoprim and enterococci in urinary tract infections: new perspectives on an old issue. *J Antimicrob Chemother* 2008, 62:35–40.

170. Cambray G, Guimaraes JC, Arkin AP: Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in Escherichia coli. *Nature Biotechnology* 2018, 36:1005–1015.

171. Collin F, Karkare S, Maxwell A: Exploiting bacterial DNA gyrase as a drug target: current state and perspectives. *Appl Microbiol Biotechnol* 2011, 92:479–497.

172. Hooper DC, Jacoby GA: Topoisomerase Inhibitors: Fluoroquinolone Mechanisms of Action and Resistance. *Cold Spring Harb Perspect Med* 2016, 6:a025320.

173. Khodursky AB, Zechiedrich EL, Cozzarelli NR: Topoisomerase IV is a target of quinolones in Escherichia coli. *PNAS* 1995, 92:11801–11805.

174. Piddock LJV: Mechanisms of Fluoroquinolone Resistance: An Update 1994–1998. *Drugs* 1999, 58:11–18.

175. Hooper DC, Jacoby GA: Mechanisms of drug resistance: quinolone resistance. *Ann N Y Acad Sci* 2015, 1354:12–31.

176. Prioritization of pathogens to guide discovery, research and development of new antibiotics for drug resistant bacterial infections, including tuberculosis. [date unknown],

177. Morgan-Linnell SK, Boyd LB, Steffen D, Zechiedrich L: Mechanisms Accounting for Fluoroquinolone Resistance in Escherichia coli Clinical Isolates. *Antimicrob Agents Chemother* 2009, 53:235–241.

178. Heddle J, Maxwell A: Quinolone-Binding Pocket of DNA Gyrase: Role of GyrB. *Antimicrob Agents Chemother* 2002, 46:1805–1815.

179. Blanc-Potard A-B, Labesse G, Figueroa-Bossi N, Bossi L: Mutation at the "Exit Gate" of the Salmonella Gyrase A Subunit Suppresses a Defect in the Gyrase B Subunit. *J Bacteriol* 2005, 187:6841–6844.

180. Moon DC, Seol SY, Gurung M, Jin JS, Choi CH, Kim J, Lee YC, Cho DT, Lee JC: Emergence of a new mutation and its accumulation in the topoisomerase IV gene confers high levels of resistance to fluoroquinolones in Escherichia coli isolates. *International Journal of Antimicrobial Agents* 2010, 35:76–79.

181. Marcusson LL, Frimodt-Møller N, Hughes D: Interplay in the Selection of Fluoroquinolone Resistance and Bacterial Fitness. *PLOS Pathogens* 2009, 5:e1000541.

182. O'Riordan W, Tiffany C, Scangarella-Oman N, Perry C, Hossain M, Ashton T, Dumont E: The Efficacy, Safety, and Tolerability of Gepotidacin (GSK2140944) in the Treatment of Patients with Suspected or Confirmed Gram-Positive Acute Bacterial Skin and Skin Structure Infections. *Antimicrobial Agents and Chemotherapy* 2017, doi:10.1128/AAC.02095-16.

183. Taylor SN, Morris DH, Avery AK, Workowski KA, Batteiger BE, Tiffany CA, Perry CR, Raychaudhuri A, Scangarella-Oman NE, Hossain M, et al.: Gepotidacin for the Treatment of Uncomplicated Urogenital Gonorrhea: A Phase 2, Randomized, Dose-Ranging, Single-Oral Dose Evaluation. *Clin Infect Dis* [date unknown], doi:10.1093/cid/ciy145.

184. Jacobsson S, Golparian D, Scangarella-Oman N, Unemo M: In vitro activity of the novel triazaacenaphthylene gepotidacin (GSK2140944) against MDR Neisseria gonorrhoeae. *J Antimicrob Chemother* [date unknown], doi:10.1093/jac/dky162.

185. Biedenbach DJ, Bouchillon SK, Hackel M, Miller LA, Scangarella-Oman NE, Jakielaszek C, Sahm DF: In Vitro Activity of Gepotidacin, a Novel Triazaacenaphthylene Bacterial Topoisomerase Inhibitor, against a Broad Spectrum of Bacterial Pathogens. *Antimicrob Agents Chemother* 2016, 60:1918–1923.

186. Farrell DJ, Sader HS, Rhomberg PR, Scangarella-Oman NE, Flamm RK: In Vitro Activity of Gepotidacin (GSK2140944) against Neisseria gonorrhoeae. *Antimicrob Agents Chemother* 2017, 61:e02047-16.

187. Savage VJ, Charrier C, Salisbury A-M, Moyo E, Forward H, Chaffer-Malam N, Metzger R, Huxley A, Kirk R, Uosis-Martin M, et al.: Biological profiling of novel tricyclic inhibitors of bacterial DNA gyrase and topoisomerase IV. *J Antimicrob Chemother* 2016, doi:10.1093/jac/dkw061.

188. Dalia TN, Yoon SH, Galli E, Barre F-X, Waters CM, Dalia AB: Enhancing Multiplex Genome Editing by Natural Transformation (MuGENT) via inactivation of ssDNA exonucleases. *bioRxiv* 2017, doi:10.1101/127308.

189. Dalia TN, Hayes CA, Stolyar S, Marx CJ, McKinlay JB, Dalia AB: Multiplex Genome Editing by Natural Transformation (MuGENT) for Synthetic Biology in Vibrio natriegens. *ACS Synth Biol* 2017, 6:1650–1655.

190. Ricaurte DE, Martínez-García E, Nyerges Á, Pál C, de Lorenzo V, Aparicio T: A standardized workflow for surveying recombinases expands bacterial genome-editing capabilities. *Microb Biotechnol*, doi:10.1111/1751-7915.12846.

191. Durcik M, Lovison D, Skok Ž, Cruz CD, Tammela P, Tomašič T, Tiz DB, Draskovits G, Nyerges Á, Pál C, et al.: New N-phenylpyrrolamide DNA gyrase B inhibitors: Optimisation of efficacy and antibacterial activity. *European Journal of Medicinal Chemistry*, doi:10.1016/j.ejmech.2018.05.011.

List of Publications and Patents

Number of peer-reviewed scientific publications: 12 (+ 4 pre-prints)

Number of patent applications: 2

Number of citations (independent citations): 137 (117)

H-index: 7

Total impact factor: 106.526

MTMT identification number: 10043320

First-author publications:

1*.   **Nyerges, Á.**, Csörgő, B., Draskovits, G., Kintses, B., Szili, P., Ferenc, G., Révész, T., Ari, E., Nagy, I., Bálint, B., Vásárhelyi, B.M., Bihari, P., Számel, M., Balogh, D., Papp, H., Kalapis, D., Papp, B., Pál, C., 2018. Directed evolution of multiple genomic loci allows the prediction of antibiotic resistance. PNAS 115, E5726–E5735. https://doi.org/10.1073/pnas.1801646115 IF: 9.504

2*.   **Nyerges, Á.**, Csörgő, B., Nagy, I., Bálint, B., Bihari, P., Lázár, V., Apjok, G., Umenhoffer, K., Bogos, B., Pósfai, G., Pál, C., 2016. A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. PNAS 201520040. https://doi.org/10.1073/pnas.1520040113 IF: 9.661

* These two publications served as the basis of the current PhD dissertation.

3.   **Nyerges, Á.**, Csörgő, B., Nagy, I., Latinovics, D., Szamecz, B., Pósfai, G., Pál, C., 2014. Conditional DNA repair mutants enable highly precise genome engineering. Nucleic Acids Res 42, e62–e62. https://doi.org/10.1093/nar/gku105. IF: 9.112

4.   **Nyerges, A.**, Balint, B., Cseklye, J., Nagy, I., Pal, C., Feher, T., 2018. CRISPR-interference based modulation of mobile genetic elements in bacteria. bioRxiv 428029. https://doi.org/10.1101/428029 . IF: N.A.

Patent applications:

5.   **Nyerges Akos J**, Pal C, Csorgo B, Kintses B (2017) Mutagenizing Intracellular Nucleic Acids,   PCT/EP2017/082574

https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2018108987

6.   Tihomir Tomašič, Lucija Peterlin Mašič, **Akos Nyerges**, Csaba Pal, Danijel Kikelj, et. al. LU100918, Luxembourgian Patent Application (2018) New class of DNA gyrase B and/or topoisomerase IV inhibitors with activity against Gram-positive and Gram-negative bacteria

Co-authored publications:

7.   Szili, P., Draskovits, G., Revesz, T., Bogar, F., Balogh, D., Martinek, T., Daruka, L., Spohn, R., Vasarhelyi, B.M., Czikkely, M., Kintses, B., Grezal, G., Ferenc, G., Pal, C.*, **Nyerges, A***., 2018. Antibiotic usage promotes the evolution of resistance against gepotidacin, a novel multi-targeting drug. bioRxiv 495630. https://doi.org/10.1101/495630 IF: N.A.

8.   Ricaurte, D.E., Martínez-García, E., **Nyerges, Á.**, Pál, C., Lorenzo, V. de, Aparicio, T., 2018. A standardized workflow for surveying recombinases expands bacterial genome-editing capabilities. Microbial Biotechnology 11, 176–188. https://doi.org/10.1111/1751-7915.12846. IF: 3.33

9.   Kintses, B., Méhi, O.K., Ari, E., Számel, M., Györkei, Á., Jangir, P.K., Nagy, I., Pál, F., Fekete, G., Tengölics, R., **Nyerges, Á.**, Likó, I., Bálint, B., Vásárhelyi, B.M., Bustamante, M., Papp, B., Pál, C. (2018) Phylogenetic barriers to horizontal transfer of antimicrobial peptide resistance genes in the human gut microbiota. Nature Microbiology, 10.1038/s41564-018-0313-5 IF: 14.174

10.   Csörgő, B., **Nyerges, Á.**, Pósfai, G., Fehér, T., 2016. System-level genome editing in microbes. Current Opinion in Microbiology 33, 113–122. https://doi.org/10.1016/j.mib.2016.07.005. IF: 6.635

11.   Lázár, V., Martins, A., Spohn, R., Daruka, L., Grézal, G., Fekete, G., Számel, M., Jangir, P.K., Kintses, B., Csörgő, B., **Nyerges, Á.**, Györkei, Á., Kincses, A., Dér, A., Walter, F.R., Deli, M.A., Urbán, E., Hegedűs, Z., Olajos, G., Méhi, O., Bálint, B., Nagy, I., Martinek, T.A., Papp, B., Pál, C., 2018. Antibiotic-resistant bacteria show widespread collateral sensitivity to antimicrobial peptides. Nature Microbiology 3, 718–731. https://doi.org/10.1038/s41564-018-0164-0. IF: 14.174

12.   Kintses, B., Méhi, O.K., Ari, E., Számel, M., Györkei, Á., Jangir, P.K., Nagy, I., Pál, F., Fekete, G., Tengölics, R., **Nyerges, Á.**, Likó, I., Bálint, B., Vásárhelyi, B.M., Bustamante, M., Papp, B., Pál, C., 2018. Phylogenetic barriers to horizontal transfer of antimicrobial peptide

resistance genes in the human gut microbiota. bioRxiv 385831. https://doi.org/10.1101/385831. IF: N.A.

13.   Durcik, M., Lovison, D., Skok, Ž., Cruz, C.D., Tammela, P., Tomašič, T., Tiz, D.B., Draskovits, G., **Nyerges, Á.**, Pál, C., Ilaš, J., Mašič, L.P., Kikelj, D., Zidar, N., n.d. New N-phenylpyrrolamide DNA gyrase B inhibitors: Optimisation of efficacy and antibacterial activity. European Journal of Medicinal Chemistry, 2018. https://doi.org/10.1016/j.ejmech.2018.05.011. IF: 4.816

14.   Guzman, G.I., Sandberg, T.E., LaCroix, R.A., **Nyerges, A.**, Papp, H., Raad, M. de, King, Z.A., Northen, T.R., Notebaart, R.A., Pal, C., Palsson, B.O., Papp, B., Feist, A.M., 2018. Enzyme promiscuity shapes evolutionary innovation and optimization. bioRxiv 310946. https://doi.org/10.1101/310946. IF: N.A.

15.   Umenhoffer, K., Draskovits, G., **Nyerges, Á.**, Karcagi, I., Bogos, B., Tímár, E., Csörgő, B., Herczeg, R., Nagy, I., Fehér, T., Pál, C., Pósfai, G., 2017. Genome-Wide Abolishment of Mobile Genetic Elements Using Genome Shuffling and CRISPR/Cas-Assisted MAGE Allows the Efficient Stabilization of a Bacterial Chassis. ACS Synth. Biol. 6, 1471–1483. https://doi.org/10.1021/acssynbio.6b00378. IF: 5.382

16.   Bódi, Z., Farkas, Z., Nevozhay, D., Kalapis, D., Lázár, V., Csörgő, B., **Nyerges, Á.**, Szamecz, B., Fekete, G., Papp, B., Araújo, H., Oliveira, J.L., Moura, G., Santos, M.A.S., Jr, T.S., Balázsi, G., Pál, C., 2017. Phenotypic heterogeneity promotes adaptive evolution. PLOS Biology 15, e2000644. https://doi.org/10.1371/journal.pbio.2000644. IF: 9.163

17.   Lázár, V., Nagy, I., Spohn, R., Csörgő, B., Györkei, Á., **Nyerges, Á.**, Horváth, B., Vörös, A., Busa-Fekete, R., Hrtyan, M., Bogos, B., Méhi, O., Fekete, G., Szappanos, B., Kégl, B., Papp, B., Pál, C., 2014. Genome-wide analysis captures the determinants of the antibiotic cross-resistance interaction network. Nat Commun 5. https://doi.org/10.1038/ncomms5352. IF: 11.47

18.   Méhi, O., Bogos, B., Csörgő, B., Pál, F., **Nyerges, Á.**, Papp, B., Pál, C., 2014. Perturbation of Iron Homeostasis Promotes the Evolution of Antibiotic Resistance. Mol Biol Evol msu223. https://doi.org/10.1093/molbev/msu223. IF: 9.105

Összefoglaló

Az evolúciós folyamatok vizsgálatára és a genotípus fenotípusra gyakorolt hatásának tanulmányozására a nagy áteresztőképességű genommérnökség páratlan lehetőséget kínál. Az örökítő anyag nagyszámú, tervezett változatát, mutációját létrehozva és hatását hasonló áteresztőképességgel vizsgálva a komplex, több mutációt igénylő evolúciós folyamatok is laboratóriumi időlépték alatt vizsgálhatóvá válnak. Mindezen evolúciós lépések azonban hagyományos laboratóriumi evolúciós módszerekkel nem, vagy csak igen hosszú idő alatt válnának tanulmányozhatóvá. Mindezek ellenére a genommérnökség eszköztára súlyos hiányosságoktól szenved, mely egyes biológiai kérdések megválaszolását hátráltatja. Ennek oka, hogy az elérhető módszerek java mindösszesen néhány laboratóriumi modellszervezetre (például a széles körben alkalmazott *Escherichia coli* K-12 bélbaktériumra) optimalizált, vagy kiterjedt genommódosítást, DNS szintézist, és időigényes klónozási lépéseket igényel – ezáltal praktikusságukat és áteresztőképességüket csökkentve. Továbbá az elérhető módszerek java nem kívánt, háttér-mutációk felhalmozódásához vezet, mely egyes esetekben akár a célzott genommódosítás hatást is elfedheti.

Kutatásunk ennek következtében a genommérnöki módszerek hátrányainak leküzdésére irányult. Ennek eléréséhez, munkánk során a jelenleg elérhető legnagyobb áteresztőképességű genommérnöki módszert, az egyes szálú DNS-rekombináción alapuló multiplex génmérnökség (MAGE) módszerét fejlesztettük tovább számos baktérium fajban történő alkalmazásra. Ezen túl pedig a nem kívánt háttérmutációk megjelenését is nagyságrendekkel csökkentettük, a módszer precíz alkalmazása érdekében. Ezen új eljárás, melyet pORTMAGE-nek neveztünk el, precíz és könnyen alkalmazható eszközt kínál számos bélbaktérium genomjának gyors és célzott módosítására. Ezt követően a pORTMAGE módszerét alapul véve olyan eljárást fejlesztettünk ki mely elsőként teszi lehetővé kiterjedt genomi szakaszok átfogó, célzott *in vivo* mutagenezisét – és ezáltal az evolúciós folyamatok korábbiaknál gyorsabb tanulmányozását.

Első célunk elérése érdekében, azaz, hogy a káros háttér-mutációk számának minimalizálása mellett egy széles gazdaspecifitású genommérnöki módszert fejlesszünk ki, egy általánosan alkalmazható, plazmid alapú eljárást dolgoztunk ki. Ennek megvalósítását az *Escherichia coli* metiláció-függő DNS hibajavítását végző egyik fehérje, a MutL domináns mutátor mutánsa tette lehetővé. A MutL a metiláció által irányított hibás bázispárosodás kijavítás („methyl-directed mismatch repair", MMR) útvonalának tagja, mely a MutHLS komplex részeként a MutH endonukleáz DNS-hibához való vonzásában vesz részt. Vizsgálataink során megfigyeltük, hogy *Escherichia coli* MutL fehérjéjének a E32→K mutációt hordozó változata a sejt mutátor állapotát indukálja, azaz a MutHLS komplex működését

meggátolja, vad típusú változatának jelenléte esetén is. Ezen domináns hatást biztosító allél lehetővé tette a bakteriális metiláció által irányított hibás bázispárosodás kijavítás szabályozását a genom módosítása nélkül, mindösszesen egyetlen, plazmid alapú fehérjetermelési rendszer sejtbe juttatásával. Mindezt kihasználva, a genommódosítást és DNS hibajavítás szabályozását együttesen megvalósító, plazmid alapú genommódosítási rendszert hoztunk létre, melyet pORTMAGE-nek neveztünk el.

A pORTMAGE egyszerre képes szabályozottan a bakteriális genommódosításhoz szükséges rekombinációs fehérjék és a domináns MutL E32→K variáns termelésére, melyet hőmérséklet-szabályozható módon a λ bakteriofágból származó cl857 represszor – pL promóter-alapú fehérjetermelési rendszer révén valósít meg. A pORTMAGE rendszere számos hasonlóságot mutat MAGE korábbi módszerével, így a korábbi módosított MAGE-módszerekben is átalakítás nélkül alkalmazható. Azonban míg a MAGE tradicionális módszere csupán limitált számú bakteriális törzsre volt alkalmazható melyekben előzetesen a DNS hibajavítás metiláció-függő útvonalát el kellett távolítani, addig a pORTMAGE alkalmazása egyetlen plazmid sejtbe juttatásával azonnal lehetővé tette a hatékony genommódosítást. A pORTMAGE alapú genommódosítás során a rekombinációt katalizáló fehérjék és a domináns MutL allél termelése egyetlen rövid hősokk segítségével indukálható, mely egyúttal a sejtek MutHLS rendszerének inaktiválásához is vezet. Ezáltal a genommódosítást végző DNS oligonukleotidok beépülésének idejére a sejtek metiláció által irányított hibás bázispárosodás kijavítási rendszere kikapcsolt állapotban marad, így egyformán hatékony mutációbeépítést tesz lehetővé legyen szó bármilyen genomi módosítás típusról. Ezen felül azonban a domináns mutátor MutL allél szabályozható termelése további előnyt is kínált. Segítségével a sejtek mutátor állapota a genommódosítás teljes folyamata során rövid időperiódusára vált korlátozhatóvá, ezáltal csökkentve a sejtosztódások számát melyet a sejt háttérmutációk felhalmozódására érzékeny állapotban tölt. Kísérleteink ezen elmélet helyességét igazolták. Ismételt pORTMAGE-genommódosítási ciklusokat követően a sejtek teljes genomját szekvenciaanalízisnek alávetve a módszer segítségével kapott sejt-változatok nem mutattak háttérmutációkat a bakteriális genomon. Ezzel szemben azonban a tradicionális módszerrel módosított sejtek több mint 80 nemkívánt mutációt halmoztak fel.

Végezetül pedig, a MutHLS rendszer erős konzerváltságát kihasználva a domináns MutL E32→K allél felhasználása a pORTMAGE általános alkalmazását is lehetővé tette. Ennek oka, hogy az *Escherichia coli*-ból eredő MutL E32→K allél számos bélbaktériumban hasonló, domináns DNS hibajavításra gyakorolt hatást mutatott, ezáltal pedig a pORTMAGE hatékony működését lehetővé tette az *Escherichia, Salmonella, Citrobacter, Klebsiella* nemzettség számos fajában is, melyek jelentős biotechnológiai és klinikai mikrobiológiai szereppel bírnak.

Ezt követően, hogy lehetővé tegyük a mutációk hatásának szisztematikus analízisét kiterjedt genomi régiók esetén is, az egyes szálú DNS-rekombináción alapuló multiplex génmérnökséget módszerét tovább fejlesztve olyan eljárást dolgoztunk ki mely képes kiterjedt genomi szakaszok egyenletes *in vivo* mutagenezisére. A jelenlegi *in vivo* mutagenezis módszerek erre nem adnak lehetőséget. Ennek oka, hogy az elérhető módszerek mutációs rátája vagy limitált, vagy épp nem specifikus, ezáltal nemkívánt háttér-mutációkat generálnak. Ezen túl egyes módszerek esetén pedig a szakasz, melyre mutáció vihető be erősen limitált. Ennek következtében pedig ezen módszerek nem teszik lehetővé a szekvenciatér hatékony feltérképezését, amely azonban elengedhetetlen lenne az evolúciós folyamatok hatékony vizsgálata érdekében. Ezen probléma megoldása érdekében a pORTMAGE alapú oligo-rekombináció módszerét olyan DNS szintézis eljárással kapcsoltuk össze, mely képes a rekombinációban felhasznált DNS oligonukleotidok teljes hosszán random elhelyezkedésű és típusú mutációkat létrehozni. Mindezt egy módosított foszforamidit-kémia alapú DNS szintézis eljárással sikerült elérnünk, mely képes volt a genommérnökség során felhasznált oligonukleotidok teljes hosszán a lehetséges nukleotid-szubsztitúciók egyenletes létrehozására.

Ezen új eljárás, melyet DIvERGE-nek (mely a 'directed evolution with random genomic mutations' megnevezésből képzett mozaikszó) neveztünk el, képes a pORTMAGE előnyös tulajdonságainak kihasználása mellett (precizitás, gyorsaság, pontosság, széles gazdaspecifitás) nagyszámú mutáció és mutáció-kombináció bakteriális genomba építésére. Kutatásaink során sikerrel igazoltuk, hogy a DIvERGE módszere során alkalmazott mutagén oligonukleotidok nukleotid pontossággal képesek genomi célpontjuk mutagenezisére, akár egymilliószoros mutációs ráta emelkedést elérve a sejtek természetes mutációs folyamataihoz képest. A DIvERGE fejlesztése során bizonyítottuk, hogy ezen mutagén oligonukleotidok részlegesen átfedő alkalmazásával és genomba építésével teljes gének, illetve azok szabályázásáért felelős szekvenciák is egyidejűleg mutagenizálhatók. Mindezt kihasználva pedig sikeresen értünk el egyenletes mutáció-bevitelt több mint 9500 bázispárnyi, négy teljes bakteriális gént (*gyrA, gyrB, parE, parC*) kódoló genomi szekvencián. Ezt követően pedig az egyenletes mutagenezis és igen magas mutációs ráta segítségével sikerrel demonstráltuk, hogy a DIvERGE módszere kiválóan alkalmas bakteriális jellegek irányított evolúciójára, olyan esetekben is, mely korábbi laboratóriumi evolúciós stratégiák felhasználásával nem vagy csak igen hosszú idő alatt vált volna megvalósíthatóvá.

Ennek egy praktikus példájaként DIvERGE segítségével sikeresen térképeztük fel a bakteriális antibiotikum rezisztencia kialakulása mögött meghúzódó genomi mutációs folyamatokat, több antibiotikum család esetén is. A DIvERGE segítségével megvalósított célzott mutagenezis laboratóriumi körülmények között sokszorosára gyorsította az antibiotikum-célfehérje mutációi által megjelenő rezisztencia kialakulását, ezáltal pedig

átfogóan vizsgálhattunk olyan antibiotikum rezisztencia folyamatok megjelenését is, melyek a tradicionális mikrobiológiai módszerekkel felfedezetlenek maradtak: nagyáteresztőképességű mutagenezis segítségével mindösszesen néhány nap időtartam alatt nagyszámú trimetoprim és fluoroquinolon-antibiotikum ellenes mutációt is sikerült felfednünk, mely közül számos a klinikai gyakorlatban is előfordult. Továbbá kihasználva a pORTMAGE széles gazda-specifitását, ezen antibiotikum rezisztenciafolyamatok hatását rokon bakteriális törzsek között is összehasonlítottuk. Az azonos körülmények között összehasonlított rezisztencia folyamatok tanulmányozásából kiderült, hogy egyes mutációk hatása az antibiotikum toleranciára az anyatörzs genotípusától függően jelentős eltéréseket mutathat, továbbá törzs-függő rezisztenciafolyamatokat is feltártunk.

Végezetül pedig, a DIvERGE egyik legfontosabb gyakorlati előnyeként, sikerrel igazoltuk, hogy egy, ismert célponttal rendelkező, fejlesztés alatt álló antibiotikum (gepotidacin) feltételezett támadáspontjait mutagenizálva az ellene megjelenő rezisztenciafolyamatok még a klinikai alkalmazása előtt felfedhetők. Így feltételezéseink szerint a DIvERGE alkalmazása mind az alapkutatásban – az evolúciós folyamatok korábbiaknál átfogóbb vizsgálata révén - mind az alkalmazott kutatásban - a hatékonyabb irányított evolúciós stratégiák lehetővé tétele által - jelentős előnyökkel szolgál majd.

Summary

Methods for large-scale manipulation of organismal genomes offer an unprecedented opportunity to understand phenotype-to-genotype relationships, and thereby systematically map fitness landscapes. In turn, microbial genome engineering is a promising and powerful tool to analyze evolutionary processes that would otherwise require exceedingly time-consuming laboratory experiments or would even be infeasible with current techniques. However, the available genome engineering methods suffer from limitations that prevent the widespread application of microbial genome engineering. These methods are either optimized for a few laboratory strains (such as *Escherichia coli* K-12) or demand labor-intensive cloning steps, DNA synthesis and/or extensive genome manipulation prior to genome editing. Moreover, most methods lead to the accumulation of undesired, off-target mutations whose effects may mask the results of intentional engineering.

Here we present significant advancements at these challenging fronts. As our first step, we have improved the currently available, most high-throughput and multiplexable genome engineering method, ssDNA-based recombineering (recombination-based genetic engineering), to achieve broad host-range functionality and to reduce off-target effects to negligible levels. This new method, termed pORTMAGE offers an all-in-one tool for enterobacterial genome engineering. Next, we have further advanced genome engineering to allow the systematic mutagenesis of extended genomic segments.

To achieve our first aim, i.e. to overcome prior limitations of genome engineering associated with off-target mutagenesis and the lack of interspecies portability, we have developed a generalized, plasmid-based method for ssDNA-recombineering. The tool enabling these advancements was a dominant-negative mutator allele of *mutL* encoding an indispensable component of the methyl-directed mismatch repair system of *E. coli*. MutL is a component of the MutHLS protein machinery which is responsible for methyl-directed mismatch removal and acts by recruiting MutH endonuclease to the site of DNA replication error. Importantly, we have noticed that the mutator effect of a particular *MutL* variant, MutL E32→K, cannot be complemented by the native, genomically-expressed MutL protein when the dominant protein is overexpressed from an extra-chromosomal vector. Based on this ability, we have decided to incorporate this dominant allele into the workflow of ssDNA-based recombineering to abolish mismatch repair activity when it is required during the process of genome editing. To achieve that, we have constructed a set of plasmids (termed pORTMAGE) expressing the λ Red recombinase enzymes, as well as the dominant-negative mutator allele of MutL, all under the control of the *c*I857 temperature-sensitive λ repressor-pL expression

system. During genome engineering, the expression of this synthetic operon is induced by a brief, temporal temperature shift, which in turn produces the necessary recombinases and the dominant-negative mutator allele of MutL which suppresses mismatch repair for the brief period of oligo-mediated genome editing. Thereby pORTMAGE has allowed efficient genome engineering and unbiased oligo incorporation, with a performance comparable to that of prior ssDNA-based recombineering protocols. Also, as pORTMAGE allows a rapid switch between mutator and non-mutator phenotypes, pORTMAGE has minimized the length of time when the bacterial population is susceptible to the accumulation of off-target mutations (i.e. it limited the time of mutator state). In line with this characteristic, whole-genome sequencing has revealed that the pORTMAGE-modified strain lacks observable off-target mutagenesis, which is a significant improvement compared to prior ssDNA-recombineering methods.

Finally, the phylogenetically highly conserved nature of MutL also offered a solution for interspecies portability. The expression of *E. coli* MutL E32→K was found to suppress mismatch repair even in relatives of *E. coli*. Thus, pORTMAGE has allowed of an efficient genome engineering and mutant library generation in numerous biotechnologically and clinically relevant bacterial species, including ones from the *Escherichia*, *Salmonella*, *Citrobacter*, and *Klebsiella* genera.

Next, to enable the systematic analyses of fitness landscapes and mutational effects, we have advanced ssDNA-recombineering to mutagenize extended genomic loci. The rationale behind this step is that the throughput and library sizes attainable with existing *in vivo* long-range mutagenesis methods are generally either moderate or lead to the accumulation of undesired, off-target modifications with detrimental side effects. Thus, these methods do not allow to explore of sequence space at extended genomic loci, which would be important for evolutionary investigations. Therefore, we further advanced our pORTMAGE-based recombineering in a way to utilize pools of partially overlapping, soft-randomized ssDNA oligonucleotides, and as a consequence, to achieve an increase of an up to a million-fold in mutation rate at defined genomic targets.

This novel method, termed directed evolution with random genomic mutations (DIvERGE), has enabled us to explore a vast number of combinatorial genetic alterations in their native genetic context. By using DIvERGE, we have demonstrated that the application of soft-randomized oligos during ssDNA-recombineering can mutagenize multiple genomic loci with nucleotide precision and without affecting off-target sites. Thereby, genomic mutation rates can be upregulated to achieve up to a million-fold increase compared to the wild-type mutation rate of the target loci. Next, by partially overlapping such oligos and covering entire genes and their regulatory regions, we have performed rapid directed evolution at genomic

loci consisting of over 9500 base pairs. Furthermore, we have demonstrated that DIvERGE can be performed iteratively, using the same oligo pools designed at the beginning of a given experiment. Thereby, DIvERGE has permitted multiple rounds of mutagenesis and selection, facilitating the rapid attainment of highly drug-resistant bacterial variants which would otherwise require time-consuming laboratory evolution protocols. Thus, DIvERGE is capable for accelerating the laboratory evolution of slowly evolving phenotypes. Exploiting the multi-species functionality of pORTMAGE, we have also demonstrated that DIvERGE functions efficiently in multiple species, as confirmed by our experiments of mutagenizing biotechnologically and clinically relevant enterobacteria. Thus, DIvERGE has enabled us to rapidly explore the *in vivo* evolution of drug resistance directly in human pathogens. During sequence analyses, we have identified numerous, previously undetected resistance-conferring mutations. Moreover, we have demonstrated that phenotypic effects of certain antibiotic resistance-associated mutations vary considerably across closely related bacterial strains. In a direct clinically relevant application, our new method has allowed us to explore the evolutionary routes promoting antibiotic resistance.

As a highly relevant practical achievement, we have demonstrated that DIvERGE is capable for rapidly identifying mutations that contribute to the evolution of resistance to three distinct antibiotic classes, and we have executed these studies in a much faster way and with a higher throughput compared to alternative methods. By utilizing DIVERGE we have identified previously described mutations responsible for resistance against trimethoprim and fluoroquinolone antibiotics: the predominant FolA and GyrA mutations evolved by DIvERGE have already been observed to arise in *E. coli* under laboratory settings, and have also been reported in the clinical practice in patients treated with trimethoprim or fluoroquinolone medications. Moreover, using DIvERGE we have also revealed several previously unknown mutations conferring trimethoprim resistance, which induce resistance to ciprofloxacin as well, and we have revealed major differences in mutational effects across related strains. By exploiting the ability of DIvERGE to explore the combinatorial mutational space at long genomic segments, we have identified a combination of two specific mutations that might lead to resistance against a novel antibiotic that is under clinical development (gepotidacin).

In the future, we hope that our DIvERGE strategy will be adaptable to other organisms, including yeasts and mammalian cells, in which ssDNA-mediated recombineering would be expected to function efficiently. As a first step in this direction, we have already developed a modified version of DIvERGE which is applicable to the eukaryotic species *Saccharomyces cerevisiae*.

Appendix
Appendix 1. Bacterial strains used in our research

Strains used in this study		
Strain	**Genotype**	**Note**
Escherichia coli K-12 MG1655	Wild-type	
Escherichia coli BL21(DE3)	Wild-type	
Salmonella enterica serovar *Typhimurium* LT2	Wild-type	from S. Datta et al. / Gene 379 (2006) 109–115
Salmonella enterica serovar *Typhimurium* TS616	derived from *Salmonella enterica* serovar *Typhimurium* LT2	from S. Datta et al. / Gene 379 (2006) 109–115
Citrobacter freundii ATCC 8090	Wild-type	from Microbiologics
Escherichia hermannii HNCMB 35034	Wild-type	from the Hungarian National Collection of Medical Bacteria (HNCMB)
Edwardsiella tarda ATCC 15947	Wild-type	from the Hungarian National Collection of Medical Bacteria (HNCMB)
Escherichia coli ΔmutS	ΔmutS	*mutS* permanently deleted, from T. Fehér et al. / Mol Biol Evol (2012) 29(10):3153-3159
Escherichia coli DH5α	Wild-type	

Appendix 2. List of oligonucleotides for ssDNA-recombineering

Oligo name	Sequence	Description
LacZ_M9_v5	5'- A*T*GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGAGAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATC*C*C	Oligonucleotide for the introduction of C:A mismatch to lacZ gene
LacZ_GG_v3	5'- A*T*AACAATTTCACACAGGAAACAGCTATGACCATGATTACGGATTGACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGG*C*G	Oligonucleotide for the introduction of G:G mismatch to lacZ gene
LacZ_AA_v1	5'- G*G*AAACAGCTATGACCATGATTACGGATTCACTGGCCGTCGTTTAACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAAT*C*G	Oligonucleotide for the introduction of A:A mismatch to lacZ gene
MAGELacZ	5'- G*G*AAACAGCTATGACCATGATTACGGATTCACTGGCCGTCGTTTGACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAAT*C*G	Oligonucleotide for the introduction of A:G mismatch to lacZ gene
LacZ_TT_v7	5'- T*G*TGGCGGATGAGCGGCATTTTCCGTGACGTCTCGTTGCTGCATTAACCGACTACACAAATCAGCGATTTCCATGTTGCCACTCGCTT*T*A	Oligonucleotide for the introduction of T:T mismatch to lacZ gene
LacZ_GA_v4	5'- A*T*AACAATTTCACACAGGAAACAGCTATGACCATGATTACGGATTAACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGG*C*G	Oligonucleotide for the introduction of G:A mismatch to lacZ gene
LacZ_GT_v2	5'- G*G*AAACAGCTATGACCATGATTACGGATTCACTGGCCGTCGTTTTATAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAAT*C*G	Oligonucleotide for the introduction of G:T mismatch to lacZ gene
LacZ_CT_v6	5'- A*T*GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGTAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATC*C*C	Oligonucleotide for the introduction of C:T mismatch to lacZ gene
MalK_CC_v1	5'- C*C*AAATGACATGTTTTCTGCTACTGACAGGTGGGGATAGAGCGGCCTAAGACTGAAACACCATACCAACGCCGCGTTCTGCTGGCGGAG*T*G	Oligonucleotide for the introduction of C:C mismatch to malK gene
araB_AA	5'- A*T*GGCGATTGCAATTGGCCTCGATTTTGGCAGTGATTCTGTGCGAGCTTAGGCGGTGGACTGCGCTACCGGTGAAGAGATCGCCACCA*G*C	Oligonucleotide for the introduction of A:A mismatch to araB gene
cycA_AAAC	5'- A*A*TGATCATATAAACGAAAATGATCGACGGCCCGGCAAGGCTAATCGTTCAGCCAGACCCCATAAACAACCCCGTACCAATGGCACCG*C*C	Oligonucleotide for the introduction of AA:AC double mismatch to cycA gene
hisB_GT	5'- C*T*GAAGCTGCAAAAAGCGGGCTACAAGCTGGTGATGATCACTAATTAGGATGGTCTTGGAACACAAAGTTTCCCACAGGCGGATTTCG*A*T	Oligonucleotide for the introduction of G:T mismatch to hisB gene
rpsL_AC	5'- G*T*CAGACGAACACGGCATACTTTACGCAGCGCGGAGTTCGGTTTTCTAGGAGTGGTAGTATATACACGAGTACATACGCCACGTTTTT*G*C	Oligonucleotide for the introduction of A:C mismatch to rpsL gene
ECASN	5'- G*A*CGAGAAAGTGAGATTTCACGAAGCTAATTNGACGTNGTTTGNCAATGNAAGCGNTTTTCNTTTTTTATACTCCTGCGTCCTGTTG*C*T	Oligonucleotide for asnA mutagenesis
SEASN	5'- G*C*TGGCGAGAGAAATGTGATTTCACGAAGCTAATTNGACGTNGTTTGNCAATGNAAGCGNTTTTCNTTTTTTTACTCCTGCGTCCTGT*T*G	Oligonucleotide for asnA mutagenesis
CFASN	5'- C*T*GGCGAGAAAAATGAGATTTTACGAAGCTAATTNGACGTNGTTTGNCAATGNAAGCGNTTTTCNTTATGTGTACTCCTGTGTCCTGT*T*G	Oligonucleotide for asnA mutagenesis
TET_DT	5'- G*T*AATAACGCTAAAAGTTTTAGATGTGCTTTNCTAAGTCATCGCGATGGAGCAAAAGTACATTTAGGTACACGGCCTACAGAAAAACA*G*T	TET landing-pad targeting oligonucleotide for Illumina deep-sequencing based MAGE performance assay
TET_DC	5'- G*T*AATAACGCTAAAAGTTTTAGATGTGCTTTACTAANTCATCGCGATGGAGCAAAAGTACATTTAGGTACACGGCCTACAGAAAAACA*G*T	TET landing-pad targeting oligonucleotide for Illumina deep-sequencing based MAGE performance assay
TET_DG	5'- G*T*AATAACGCTAAAAGTTTTAGATGTGCTTTACTAAGTCATCGNGATGGAGCAAAAGTACATTTAGGTACACGGCCTACAGAAAAACA*G*T	TET landing-pad targeting oligonucleotide for Illumina deep-sequencing based MAGE performance assay
TET_DA	5'- G*T*AATAACGCTAAAAGTTTTAGATGTGCTTTACTAAGTCATCGCGATGGAGCAAAAGNACATTTAGGTACACGGCCTACAGAAAAACA*G*T	TET landing-pad targeting oligonucleotide for Illumina deep-sequencing based MAGE performance assay
TET_DC2	5'- G*T*AATAACGCTAAAAGTTTTAGATGTGCTTTACTAAGTCATCGCGATGGANCAAAAGTACATTTAGGTACACGGCCTACAGAAAAACA*G*T	TET landing-pad targeting oligonucleotide for Illumina deep-sequencing based MAGE performance assay
* represents phosphorothioate bonds		

DIvERGE Oligonucleotides

| Oligo ID	Sequence (5'-3')	Description
TETRM1_01	TTAGATAAAAGTAAAGTGATTAACAGCGCATTAGAGCTGCTTAATGAGGTCGGAATCGAAAGGGTTTAACAACCCGTAAACTCGCCCAGAAG	Spiked-oligonucleotide with 0,1% spiking, targeting the Landing pad
TETRM3_01	TTGAGATGTTAGATAGGCACCATACTCACTTTTGCCCTTTAGAAGGGGAAAGCTGGCAAGATTTTTTACGTAATAACGCTAAAAGTTTTA	Spiked-oligonucleotide with 0,1% spiking, targeting the Landing pad
TETRM1_05	TTAGATAAAAGTAAAGTGATTAACAGCGCATTAGAGCTGCTTAATGAGGTCGGAATCGAAAGGGTTTAACAACCCGTAAACTCGCCCAGAAG	Spiked-oligonucleotide with 0,5% spiking, targeting the Landing pad
TETRM3_05	TTGAGATGTTAGATAGGCACCATACTCACTTTTGCCCTTTAGAAGGGGAAAGCTGGCAAGATTTTTTACGTAATAACGCTAAAAGTTTTA	Spiked-oligonucleotide with 0,5% spiking, targeting the Landing pad
TETRM2_1	TTAGATAAAAGTAAAGTGATTAACAGCGCATTAGAGCTGCTTAATGAGGTCGGAATCGAAAGGGTTTAACAACCCGTAAACTCGCCCAGAAG	Spiked-oligonucleotide with 1% spiking, targeting the Landing pad
TETRM3_1	TTGAGATGTTAGATAGGCACCATACTCACTTTTGCCCTTTAGAAGGGGAAAGCTGGCAAGATTTTTTACGTAATAACGCTAAAAGTTTTA	Spiked-oligonucleotide with 1% spiking, targeting the Landing pad
TETRM2_2	TTAGATAAAAGTAAAGTGATTAACAGCGCATTAGAGCTGCTTAATGAGGTCGGAATCGAAAGGGTTTAACAACCCGTAAACTCGCCCAGAAG	Spiked-oligonucleotide with 2% spiking, targeting the Landing pad
TETRM3_2	TTGAGATGTTAGATAGGCACCATACTCACTTTTGCCCTTTAGAAGGGGAAAGCTGGCAAGATTTTTTACGTAATAACGCTAAAAGTTTTA	Spiked-oligonucleotide with 2% spiking, targeting the Landing pad
TETRM1.rv	CTTCTGGGCGAGTTTACGGGTTGTTAAACCTTCGATTCCGACCTCATTAAGCAGCTCTAATGCGCTGTTAATCACTTTACTTTTATCTAA	Reverse complement of TETRM1
TETRM3rv	TAAAACTTTTAGCGTTATTACGTAAAAATCTTGCCAGCTTTCCCCTTCTAAAGGGCAAAGTGAGTATGGTGCCTATCTAACATCTCAA	Reverse complement of TETRM3
TETRM1	TTAGATAAAAGTAAAGTGATTAACAGCGCATTAGAGCTGCTTAATGAGGTCGGAATCGAAAGGGTTTAACAACCCGTAAACTCGCCCAGAAG	Reference-oligonucleotide without spiking, targeting the Landing pad
TETRM3	TTGAGATGTTAGATAGGCACCATACTCACTTTTGCCCTTTAGAAGGGGAAAGCTGGCAAGATTTTTTACGTAATAACGCTAAAAGTTTTA	Reference-oligonucleotide without spiking, targeting the Landing pad
folORM1_05	ATCGGCAGGCAGGTTCCACGGCATGGCGTTTTCCATGCCGATAACGCGATCTACCGCTAACGCCGCAATCAGACTGATCATTGAGATTTC	Spiked-oligonucleotide with 0,5% spiking, targeting the folA locus of E. coli K12 MG1655
folORM2_05	ACCGATTGATTCCCAGGTATGGCGGCCCATAATCACGGGTTTATTAAGGTGTTGCGTTTAAACCAGGCGAGATCGGCAGGCAGGTTCCA	Spiked-oligonucleotide with 0,5% spiking, targeting the folA locus of E. coli K12 MG1655
folORM3_05	CACCCACGTTACGCGATCGTCCGTACCCGGTTGACTGCTGAGGATAATATTTTTGCGTCCTGGCAACGGACGACCGATTGATTCCCAGGT	Spiked-oligonucleotide with 0,5% spiking, targeting the folA locus of E. coli K12 MG1655
folORM4_05	AACGCGACCGCCGCCAATCACCATGATTTCTGGTACGTCACCACACGCCGCGATGGCTTCATCCACCGACTTCACCCACGTTACGCGATC	Spiked-oligonucleotide with 0,5% spiking, targeting the folA locus of E. coli K12 MG1655
folORM5_05	GGTGTCGCCTTCCACTTCTGCGTCGATATGCGTCAGATACAGTTTTTGCGCTTTGGCAAGAACTGTTCATAAACGCGACCGCCGCCAAT	Spiked-oligonucleotide with 0,5% spiking, targeting the folA locus of E. coli K12 MG1655
folORM6_05	GTTCTGCGCATCAGCATCGTGGAATTCGCTGAATACCGATTCCCCAGTCATCCGGCTCGTAATCCGGGAAATGGGTGTCGCCTTCCACTTC	Spiked-oligonucleotide with 0,5% spiking, targeting the folA locus of E. coli K12 MG1655
folORM7_05	GCATCCGGCGCTAGCCGTAAATTCTATACAAAATTACCGCCGCTCCAGAATCTCAAAGCAATAGCTGTGAGAGTTCTGCGCATCAGCATC	Spiked-oligonucleotide with 0,5% spiking, targeting the folA locus of E. coli K12 MG1655
folPRM1_05	ACTGATCATTGAGANTTTCCCGATAAAAAAAATTGTCGCCACTATACGTAAAGCGTAAACCGTCGTCGACTGGTGCGAGGATGATGTTGAG	Spiked-oligonucleotide with 0,5% spiking, targeting the folA locus of E. coli K12 MG1655
CfolORM1	ATCGGCAGGCAGGTTCCACGGCATGGCGTTTTCCATGCCGATAACGCGATCTACCGCTAACGCCGCAATCAGACTGATCATTGAGATTTC	Spiked-oligonucleotide with 0,5% spiking, targeting the folA locus of E. coli CFT073
CfolORM2	ACCGATGGATTCCCAGGTATGGCGGCCCATAATCACGGGTTTATTTAAAGTGTTGCGTTTAAACCAGGCGAGATCGGCAGGCAGGTTCCA	Spiked-oligonucleotide with 0,5% spiking, targeting the folA locus of E. coli CFT073
CfolORM3	CACCCACGTTACGCGATCGTCCGTACCCGGTTGACTGCTGAGGATAATATTTTTGCGTCCTGGCAACGGACGACCGATTGAATTCCCAGGT	Spiked-oligonucleotide with 0,5% spiking, targeting the folA locus of E. coli CFT073
CfolORM4	AACGCGACCGCCGCCAATCACCATGATTTCTGGTACGTCACCACACGCCGCAATGGCTTCATCCACCGACTTCACCCACGTTACGCGATC	Spiked-oligonucleotide with 0,5% spiking, targeting the folA locus of E. coli CFT073
CfolORM5	GGTGTCGCCTTCCACTTCCCGCGTCGATATGCGTCAGATACAGTTTTTGCGCTTTCGGCAGGAACTGCTCATAAACGCGACCGCCGCCAAT	Spiked-oligonucleotide with 0,5% spiking, targeting the folA locus of E. coli CFT073
CfolORM6	GTTCTGCGCATCAGCATCGTGGAATTCGCTGAATACCGATTCCCAGTCATCCGGCTCGTAATCCGGGAAATGGGTGTCGCCTTCCACTTC	Spiked-oligonucleotide with 0,5% spiking, targeting the folA locus of E. coli CFT073
CfolORM7	GTATCAGTGATGCCGGAATTCTAATATACAAAATTACCGCCGCTCCACGAATCTCAAAGCAATAGCTGTGAGAGTTCTGCGCATCGGCATC	Spiked-oligonucleotide with 0,5% spiking, targeting the folA locus of E. coli CFT073
CfolPRM1	ACTGATCATTGAGANTTTCCCGATAAAAAAAATTGTCGCCACTATACGTAAAGCGTAAACCGTCGTCGACTGGTGCGAGGATGATGTTGAG	Spiked-oligonucleotide with 0,5% spiking, targeting the folA locus of E. coli CFT073
MgyrA.PRM_0.5	gtcgctcatctaaccgctatccctctactgtatcccgaattcaaagcgtgcaaattataacacagccgcgcagtttgaggtaaacctata	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM1_0.5	CGACATCGCATAATCCAGATAGGAGCTCTTCAGCTCTTCCTCAATGTTGACCGGTGTAATTTCTCTCGCAAGGTCGCTCATCTAACCGCT	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM2_0.5	GGCGTAAAGTACGCGACGGTGTACCGGCTTCAGGCCATCTCGGACATCTGGCAGCGCACGGCCAACAATGACCGACATCGCATAATCCAG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM3_0.5	ACCGATTACGTCACCAACGACACGGGCAGATTTTTTATAGGCTTTGTTCCAGTCATTGCCTAGTACGTTCATGGCGTAAAGTACGCGACG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM4_0.5	ACGCAGCGAGAATGGCTGCGCCATGCGGACGATCGTGTCATAGACCGCCGAGTCACCATGGGATGTATTTACCGATTACGTCACCAAC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM5_0.5	CGTATAACGCATTGCCGGCGATAGGCTCGCCGTCGATAGAACCGAAGTTACCCTGACGTCTACCAGCATATAACGCAGCGAGAATGGCTG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM6_0.5	ATCAACGAAATCGACCGTCTCTTTTCGAGATCGGCCATCAGTTCATGGGCAATTTCGCCAGACGGATTTCCGTATAACGCATTGCCGC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM7_0.5	AGAACCGTTCACCAGCAGGTTAGGAATTTTGGTTGGCATGACGTCCGGAATTTTTTCCGTGCCGTCATAGTTATCAACGAAATCGACCGT	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM8_0.5	CAGACAACCGTTGATGACTTCCGTCAGGTTGTGCGGCGGGATGTTGGTTGCCATACTACGGCGATACCGGAAGAACCGTTCACCAGCAG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM9_0.5	CGTCGGGAAGTCCGGCCCGGGATGTGTTCCATCAGCCCTTCAATGCTGATGTCTTCATCATCAATATACGCCAGACAACCGTTGATGAC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM10_0.5	GCGGATATACACCTTGCCGCGACCGGTACGGTAAGCTTCTTCAATACCGCGACGACCGTTAATGATTGCCGCGCGTCGGGAAGTCCGGCCC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM11_0.5	CTGATACGGAATTTCGTGGACGATAATGGTTTCACGACCGGTTTTGGCGTCAACTTCACTTCTGCGCGAGCGCGGATATACACCTTGCC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM12_0.5	CGCGCTGATGCCTTCCACGCCGTTTTTCTTTTACCAGTTCCGCAATCTTCTCGATCAGGCGCGTTTGTTTACCTGATACGGAATTTCGTG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM13_0.5	AACTTCACCGACCGCATCGCGTTTCACTTCAATCACGATGCGCATACCGTCTTTGTCAGACTCGTCACGCAGCGCGTGATGCCTTCCAC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM14_0.5	ATGGTGCAATGCCACCATGTTGATACCGAAAGAAACCTGCAACTGGGTCTGGGAGTAGAGGTTGTTGAGCACAACTTCACCGACCGCATC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM15_0.5	GGTCACCACTTCACGGCGGTGACGAACAAACGCCGCGATGATGTCTTTCAGGTTCATGATCTTCGGCTGACCATGGTGCAATGCCACCAT	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM16_0.5	CAGCGCCCACGGCTAATGCTTCAAGGATATGAGCACGATCGCGAGCTTTACGCAGTTCGAAAATAGTACGACGGGTCACCACTTCACGGCG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM17_0.5	AACCAGCGCAGTTTTCGCTTCTGCAGGCGTCGGCGCATGACGGATCAGTTCGATGAGTCGGGTCGATGTTCGCCAGCGCCACGGCTAATGC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM18_0.5	TTCCGGACGCGCAGCATCGTCGCCAGCAGTTCGAGCATCGCGGCAACGTTGCCCCAGCTGCCACGGATTAGCAACCAGCGCAGTTTTCGC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM19_0.5	CAGAATCGCCTGAGCTTGCTGTTCGGTCAGGTAGTACAGACCATCACGCACGCCGAACTCTGGCTCCAGCCATTCCGGACGCGCAGCATCG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM20_0.5	ATCCAGCAGCTCTTTGTATTCGTCGAGCAGTTTTTCGTGCTCAAGACCGGTCAGTTTCTGCAAACGCAGATCCAGAATCGCCTGAGCTTG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM21_0.5	CTCCAGCTCTTCACGGATCACTTCCATCAGACGATCGGCGCTACCAAGAATAGCAGCACAGTTCGGGATCTGATCCAGCAGCTCTTTGTA	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM22_0.5	TTCCAGGTTGATGTCTGCGCTTGTCGGCGGTGATTTCAGTACGACGTTTGTCACCGAACTGTTCACGAACGACTCCAGCTCTTCACGGAT	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM23_0.5	AGAAAGCGGCTGATACTTAACGTAGCCCTGGTGAGAGAGCGTCACGACCACATCTTCCTGGGTGATCAGATCTTCCAGGTTGATGTCTGC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM24_0.5	GTCGATAAAGTCTTCTTCTTTAATACGTGCGGCAGATTTACCTTTCCCGCCACGACGCTGCGCTTCGTATTCAGAAAGCGGCTGATACTT	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM25_0.5	TTTCATCGAATAGACGCGACCACGGCTGGAGAAGCACAGAATATGGTCGTGAGTGTTCGCCACCAGCAGTCGGTCGATAAAGTCTTCTTC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM26_0.5	CTGCTCAGCGGCAGCAGGTTGACGATCGGACGACCGCGCGCGCACGAGTGGCTTCCGGCAACTGATAAACTTTCATCGAATAGACGCG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM27_0.5	AGCCGGTCGCCATGAGAGTTTCACGCCTTCTTCAAACTCGGTCACTGGCAGGATCGCAGTGATACGTTCGTCCTGCTCCAGCGGCAGCAG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM28_0.5	TTTGATCGCCACTTTACCGGCGGTACGCAGACGGTTGAACTCGGTGAGGACAGTTTTCTTCACGGTACCGTTAGCGGTCGCCATGAGAGC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM29_0.5	AGCGGAGAACAGCATTACTTCGTCTTCGCCGCTGGTCAGGTCAACGCCGATCAGCTCATCGCCGTCAACCAGTTTGATCGCCACTTTACC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM30_0.5	ACCGCGAACACCGGTGGTGTTGCAGCCCATCGCACGGACAGAAGACTCTTTAGCCTTGCAGTCGGCGAGAACACGCATTAC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM31_0.5	TGCGGTGAGGATTGCGCATCGCCACGAGGACAGATCAGACAGAGAGACTTTATCGCCTTCACCTAACCTTTGACCGGCGAACACCGGTGGT	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM32_0.5	AACCCCTTTCGTCGCACGCGACTTGGTTGGGTATTCCGCCACTGCAGGTTACGGTAACCGTTTTGCGTGAGGTTGCGCATCGCC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM33_0.5	GATCTGGTCGCAGTCATCTACCTGTACCGCGCCAACAACTAAACCGTTACGTTCAGATGGATACCCGTTTTACGGTCGATAACCCCTTTCGTCGCACG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM34_0.5	GGTGTTACGGCCCACGATGCTGATTTCCGAAACGCGAGTACGTACCAGCGTACCGGCATCGGTGATCATCATGATCTGGTCGCAGTCATC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM35_0.5	AACCGGTTTCAGCAACACGTTGCAGACCCACTTACGTTTTCATCTTCCGCAGTACGGATGTACACGCCCTGGTCCGGTCCACATGGTGTTACGGCCCACGAT	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM36_0.5	GTCCACTTCCGGAGCGATTTCATCGTCCCCTTCCCGCGGCACTGCCGTCGATGGTATCCAGATCTTCCTCGTCAACCGGTTCAGCAACACG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrA.ORM37_0.5	ATTCAAACAAGGGAGATAGCTCCCTTTTGGCATGAAGAGTAAAATTATTCTTCTTCTGGCTCGTCGTCAACGTCCACTTCCGGAGCGAT	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrA locus of E. coli K-12 MG1655
MgyrB.ORM1_0.5	TTACGCACCGCATCCAGCCCTTTCAGGACTTTGATACTGAAGGAGTCATAAGAATTCGACATCAACGTTTTCTGCTTTTTATACTTGGG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM2_0.5	TCTACCACCTCGAATACCATGTGGTGCAGACCCGGTGCCGTCATCCGTGTCGCCGATATACATACCCGGGCGCTTACGCACCGCCATCCAGC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM3_0.5	ACAGAGTTATCGGCGTGAATGGTGACGATAATTCTTTACAGTGACCCGCGAGCGCTTCGTCGATAGCGTTATCTACCACCTCGAATACC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM4_0.5	ACTTCCGCCGCCGATACGCCCTCTTCCGGGTGAATACCGGTCGGAATGCCGCGCCCGTCATCCTGTACAGAGACAGAGTTATCGGCGTGA	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM5_0.5	CCGTGCAGACCGCCGGACACTTTATAGGAGTTATCGTCAAATTTACCGCCTGCGTCGAGAACGGTCATGATCACCTTCCGCCGCCGATACG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM6_0.5	TGAATTTTACCCTCGCGCTGGATAACGCAGCTCCAGTTTTTGCGACAGGGCCTTGCCCGGCGGCAATCACCGTGCAGACCGCCGGACACG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM7_0.5	GTGCCGGTTTTTTCAGTCTGCCGGTAACGCCAGGGCGACGGGGGCCTGGGTACAGCCACCGCTGTTGGTAGATTACCCTCGCGC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM8_0.5	TTCGCCAGAATTTCATATTCGAACTCGGTCACATTGGTGAAGGTTTCGAGGCTGGGCAGAGCGCACCATCAGGTCGAGGTGCCGGTTTTTCAGTC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM9_0.5	TCTTTGCCGTCGCGCTTGTCGCGCAGACGAATGGAAACGCCGGAGTTGAGGAACGGCAACTCACGCAGACGTTTCGCCAGAATTTCATAT	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM10_0.5	TGGATCGGCGTTTTGTTCTTGTTCAGATATCAACGACCGCCTTGATGCCGCCTTCATAGTGGAAGTTGGTCTGTTGCGGTGCGGTTG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM11_0.5	CCATCGTTCCACTGCAACGCCACTTCGAGCGCCAATACCGTCTTTTCAGTGGAGAAGTAGAAGATATTCGGGTGGATCGGCGTTTTGTTC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM12_0.5	AAGCCTGCCAGGTAGATACCGCCGTCACGCTGCCGAATGTTGTTGGTAAAGCAGTGATGTTTTCCTGGAAGCCATCGTTCCACTGCAAC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM13_0.5	CTGACTTTGGCTTTTTGCTGTAGCCTTCTTTGTCATCGTAGGCGTTCAGGGTACGGGTCATCGCCGCACGAAGCTGCCAGGTAGATA	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM14_0.5	GAGAATTTCGGGTCCGGCACTTTCACGGAAACGACCGCAATCAGGCCTTCACGCGCCATCGTCACCGGTGGCGCTGACTTTGGCTTTTTTG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM15_0.5	AGCAGTTCGTTCATCTGCTGTTCAACCGCCGATTTCGGCGTCCTTCAGAGAACTCAAGTTTCTTTGGTCGGATGAGATTTCGGGTCCGGC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM16_0.5	GCACGGGCAGCATCGATAATTTTGCCAACCACGATTTTCGGCGTCGGTTGGGTTTTCCAGCAGGATATTCTGCCAGCAGTTCGTTCATCTGC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM17_0.5	CCCGGCAGGCCCGCTAAGTCGAGCGCACCTTTACGGCGGGTCATTTCACGCGCGCACGACGCCGTTCACGGGCACGGGCAGCATCGATA	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM18_0.5	CCCCGCGGAGTCCCCTTCCACCAGGTACAGTTCGGAAAGCGCCGGATCGGCGTTCCTGGCAGTCTGACCAGTTTGCCCGGCAGGCCCGCTAAG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM19_0.5	ACGTTGAGGATTTTACCCTTCAGCGGCAGAATCGCCTGGTTCTTGCAGTTCAGCCCTGCTTCGCAGAACCCGCCCGGAAGTCCCCTTCC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM20_0.5	CCACAGCCCAACGCGCGGTGAATAAGCGTGCGCACTTCCTGAGAACGAGACGATCTTATCGAAGCGCACGTTCTGCAACGTTGAGGATTTTACCC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM21_0.5	ACGTCCGCATCGGTCATGATGATTGACTGCTGTGATAACGCAGTTTGTCCGGGTTGTACTGGTCAGCACCGATACCACGCCAAGCGCGGTG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM22_0.5	CCGCGTTCAACGATTTCCGGCATCTGACGATAGAAGAAGGTCAACAGCAGCGTACGAATGTGCGAGCCGTCGACGTCCGCATCGGTCATG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM23_0.5	TCGTCTTTAATGTACTGTCGTCGCTTGCCTTTCTTCACTTTTGTACAGCGGCGGCTGACTGGAGATGTGACCGCGTTCAACGATTTCC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM24_0.5	GGTGCACTGGCGTTGGTGTCGACGTTCGCCGTCGCCACGATAGAGAATCTGGTACTGATCCATCGCTTCGTCTTTAATGTACTGT	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM25_0.5	ATACGATTGATCATTTTCTGCGTCGCGTTGTACTCAGATACCAGTTTCTCTAACGCTTCGCCAGCCAATGCACCGGTGCACTGGCGTTGGTG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM26_0.5	GAAAGGTCAGCTTCCGTCAACGTCGGCTGATAGATAAGCTCTTTCAGCATTGCTTTCAGGACGGCTCCATACGATTGATCATTTTC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM27_0.5	CTGCCGTGCTGTTCTTTGTCGTTCAGTTCGCTGACCAGCGCGTTCACCCAGCGGGTAACGGTCTGCTCATCAGAAAGGTCAGCTTCCGTC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM28_0.5	TGGGTACGCACGCGAACAATCGGCTGAAGGTTTTGCTGAGCATCAAACTTCCACTGGCTGCCGTGTTCCTGCCGTGCTGTTCTTTG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM29_0.5	AGCGTGCAGATACGACCGATATTTCGCCACGGATCTGCTGATCCAGCGGCGATAGTCCAACCGTGGGTACGCACGCGAACAATCGGCTGAACA	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM30_0.5	GCTACCGGCTGACGCGCGTCGCCACGTTCGATAAACGCATCTTCTTCCAACAAGCCACGCCAGTTTCTCACCAACGGTGCAGATACGACCGA	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM31_0.5	CCTTTATAACGCTGGATGGAGAGGCCGCGACGGCACTCTTTCACCAGACGTTCAGAAGGCGATCGTGTGTAACCGGCTACCGGCTGACG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM32_0.5	ACGCGCAGCATACGACGACTTTCCGGGTCCATAGTGGTTTCCCCACAGCTGTTCCGGGGTTCATCTCGCCCAGACCTTTATAACGCTGGATG	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM33_0.5	CGGCGCGGTTCAACGGCGTCGCCCATCAGCGTGGTGAACAACTGGTCGGCAGCAATCGCATCTTTAACGGTAACGCCGCCGATACGACGA	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MgyrB.ORM34_0.5	GCACGCTCGACATGGTTAGCGCCATTAAATATCGATATTCGCCGCTTTCAGGGCGTTCGTTCCTCAATAAACGCAGGCGCGGTTCAACGGC	Spiked-oligonucleotide with 0,5% spiking, targeting the gyrB locus of E. coli K-12 MG1655
MparC.ORM1_0.5	GTAGTTTAAGTAGGCGTTTTCCGTAAATTCATGTAGCGCAAGGCGCTCTGCCATCATCCTGCTCATTAATTCTGATTCCTCAACTTATTG	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM2_0.5	AATGCGGCGCTGAACAGGTTTCAGACCATCACCAATAAACGGCAACGCACGGTCCATGATCACGTATGGAGTAGTTTAAGTAGGCGTT	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM3_0.5	GTCACCGACGGTACGGGCCGATTTTTAAATTTGGGCGCTGGCATTCAGGCCCAGTTCAGACATCGCATCACAGATGGCGGCGCTGAACAGG	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM4_0.5	GAACGGTTGCGCCATCAGGAACCATCGCTTCATAACAGGCGGCTATCGCCGTGTCGGGATGGTATTTTACCCAGTACGATGGCGTTCCAGG	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM5_0.5	TGCCGCGAACGATTTCGGATCGTCCGGCGCGCCCAGTTCCCCTGACCATCAACCAGCGGATAACGGTAAGAGAACGGTTGCGCCATCAG	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM6_0.5	AGCCGTCCCCTGCCCCAGCTCGCTCAATAGCAGCTCGGAATATTTCGACAACCGGGATTCGGTGTAACGCATTGCCGCGAACGATTTCGG	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM7_0.5	CAAATCTTTGCCAGACGGGCAGGTAGCATTTTCGGCTCCTGCAAAGTGCCGTCGAAGTTTGGCACCCAGTCAGCGTCCCCTGCCCCCAG	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM8_0.5	AGCCACTTCACGCAGGTTATTCGGCTGGAATATCGGTCGCCATGCCGACGGCAATACCGGTGCCTGTTCAGACAAAATGTTTGGCAGACG	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM9_0.5	CGGCCCCTGCACGATATCCAGCAGTGATCGAGCGTGGTTTTCGGCTGGTCGATTAATGCGATTGGCCGCCTGCAGGCTAGCCACGAGGTT	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM10_0.5	ACCACGTCCGTTCTCGTAGATTTTTACGGATCTCGGCGCGCGCGAAGTGATAATTTTCCGCTTCAGTCGAATCCGGCCCCTGCACGATATC	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM11_0.5	AACCTGATGCGGCAATGCGGCTGATAACCACCCGCCGCCATCTTCTTTTCTTCCACACCGACGGCGGACAACGCAGTTCTTCGGTGGA	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM12_0.5	AGATCGTCAACCATCGGCAGCTTTTTGTTGCCGATTTTGCGGCAATTTGCTCCAGTACGCGGCGCGCACCTGAAACCTGATGCGGCAATGC	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM13_0.5	TCCATATCCACGGGTTTGGAACGCGGCACATCAACCAGCGGGTTCGGGTTCTCGTGGTCAGATTCATCGCGCAGAATATCAACCATCGGC	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM14_0.5	CCGATCATATTAAGGTTAATACGATAGCTCTTTTCCAGATCGGTGGTAGCGAAGAGGTGGTTCATTCACCACCTGATCCATATCCACGGGTT	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM15_0.5	GTATCGCGGCGGAACACCAGCCATTCGGGAGAGGATTTCCACGATGAGCAGTTTTTCACCGCCGGACGACCATCCAGACCGATCATATTAAGGTTA	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM16_0.5	ACCAGCAAACCTTCGAGGATATGCAGGCGCTTGAGGACTTTCTGGCCCGATGATGATTTCAGTCGACGTCGAAGTAGGTATCGCGGCGGAAC	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM17_0.5	GACATCAGCCGCCGGTTTCGGTTCATCTTCATTACGGCATGCTCAATCACTTCGTCGATATTGGAAAACGCCACCAGCACCAACCTTCGAGG	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM18_0.5	TCCAGTTTGGCAGATGACGCAGTTTCAGTTCGAGGATCGCTTCCGCCTGGGTTTCCGTAAGGCCAAACCGCGACATCAGCCGCCGGTTTC	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM19_0.5	GAAGCCAAAATGCCCCTGCAACTGGTCGCGCTCTTTTCCAATGCCTCGCTGCTGCGAATCTTCATCTCTTCCAGTTTGGCAGATGACGCA	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM20_0.5	CGACGATCGTCACCGTAGGCTTGCGCGTCTGCCTGCAGTTCTTTCTTCAGCAGGTTATTCATTTTTCCCTTGCACCGGGCAACAAAGCCTACCTAG	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM21_0.5	GTGACAGGTTCAGACGGCAGGCACATGTCGTGCTGCTCATCGCTTTCGGCTTCTTGCCGTTCACCGGCGACGATGTCGACCCTAG	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM22_0.5	TAATTCAGGCCCGGCGCGTCGATATCATGGCCTTTTAGCGCTGCGTACCCAGCCCATCTGCGAGCACAATGGTGACAGGTTCAGACGGCAG	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM23_0.5	GTGGAATCAACAAACACTACCGGTTGGTTGCTCTTACCTTTCACCGCCGCTTTGAAGCTATCACCCGCTTTATAATTCAGGCCCGGCGCG	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM24_0.5	TTGCCGGTGAGCGGCTCGCCCTGACCAGCGCGCAGCGGAACGGTATCGGGTCAATGGCATAGCTACGACCGGGTGAATCACGTGGAATCAAC	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM25_0.5	GCCATCAGCAGTTTCGATTCCAGAGCGTTTCCTCCTCATGGCGCAGCACCACGGCGCCCAGGCGGCAACGTTTATTTGCCGGTGAGCGGCTC	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM26_0.5	AAAGCCTTACCTGCACGGTTTAGCGCCATCCGGTTTAAAGGTGCAGACGAACGATGAGCCATGCCGATCGGTGTCGAGTGCATCAGCAGTTTCGA	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM27_0.5	GCCAGCAGCATATCGGAAGCATCTTCAATCACCACCGCGGCGAACATGGTAGAGTCCCATTTTCCGATAGGGTAACACTTCAGCACGGT	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM28_0.5	TTGCCTTTGCCCCTTCGCAACAGCTGCGGCAGATCAACTTACCGGGAACATCAACATACGGCGCCTGCCTGAGTGATTGCCAGCAGCATATCGGAA	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM29_0.5	GCCAGCGACGTACAATTGCGGCAGACCATCTTCCTCCACGCGTGCAGGCTTGCTTTTGGGCAGTTGAATACTTTGTTGCCTTTGCCCCTTCG	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM30_0.5	ACTTTCTGTAACTCTTCCGGGGCGCAGTTAATTTTGCGTTTCCCAACATGAATGGTCAGCGTTACGGGGATGATGCCAGACGTACAATTGC	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM31_0.5	GAGTCGATGCATCCACAGATCGATACTGCCAAACCGCGCATCAACGTACCGGCGGCGTTCGCATCAGGTAGACTTTCTGTAACTCTTCC	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparC.ORM32_0.5	TGCAAGCGGGAGGAAACAGCGCCCTCCCCGGCATTATTACTCTTCGCTCGATCAGCCGCTGCTGACGAACTGGAATGATGAGTCGATGCATCCACAGA	Spiked-oligonucleotide with 0,5% spiking, targeting the parC locus of E. coli K-12 MG1655
MparE.ORM1_0.5	CGGCGAACCGGCTCAAGCCCCGGTGAGTACCTCAATGGCATCAGCGTTATAAGTTTGGCGTCGTGTCATGGTTTAAGTTAGTAATTCGAGTTGATCG	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM2_0.5	TCCACACTGTTATCAATGACTTCTTGCCCCAAATGGTTAGGGCGAGTGGTATCAGTATACATCCCCGGACGGCGGGCGAACCGGCTCAAGC	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM3_0.5	TCAATAACTTCTAACGACTGGTCAGCATGTAAAATAACGTCCACGACGTTTTGCGTGACCCGCCAGTGCTTCATCCACACTGTTATCAATG	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM4_0.5	CGGCAAAGAATCAGTTCAACCGCACGGGTACACCCTCTTCTTCCGGGTGAAATATCACCCGGCATCCGTCAATAACTTCTAACGACTTAG	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM5_0.5	AAGGCGATGTATAAACCTGACCATCGCGGCGACAGTTTAACTTCTACGCGCTTCGACAGGGTGTTACCAGCGAACGCCCACAGCGCCATGC	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM6_0.5	ACAGTGGTACCATATTGCCGTTACCCGCGCCAGGATGCCCTGGGCTGTCAAGGGGTATAAACCCCGCCTGCCAACGGCATGCAAAGGCGATGTATAAACC	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM7_0.5	TTCAGCACATTGCGTCAGGCGTGAAACAGAAATCGCGGCTGTCAAAGAAGGTTTCATCCGGCACAGGACGGCAAACGACACAGTGGTACCATATTAAAC	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM8_0.5	CAGCGTTGTTGCTGATTGTGATCTCATCTTTAAAGTGATCTGGCCGTTTACAGGTGTTGCAGCCATGCGTTGTTCAGCACATTGCGTCAGG	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM9_0.5	TCACGCCTTCGGGAGCATCGCCGACAGACCATTTTACTTTCAACGCCAGGGCACTACTTCAGCACGTCCTTGATGACCAGCATGCGTCAGG	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM10_0.5	TGGTTATCTGCGCCAGCGTCGCGAGACCATTTCTGCCTTCCGGCCATGGAGTCAGATCGCCTGATAGCAGCAGTCGCGTGCAGCGTC	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM11_0.5	TCACCGCCTTCCGGCAGCCCACAGTAGCGCCAGTTCCACAGCTTCACCAGCGAAAGCGAGAGGTTTTTTCCGCGAAATGGTTATCTGCGCGTC	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM12_0.5	TGACGCAGACCATTAACGATGGTACCGCCTGCATCGTTGGGATAAGGTTAGTCGCCGCTGGCTAACGGTACTTCTCACCGCCTTCCGGCAG	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM13_0.5	GCCGACAGCTTTACCGCGCGCGACGAATATTGCGGTTCATACAGAACTCACGCCATCGCGTCCAACAGCCGGTACTGACGCAGACCATTAACA	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM14_0.5	GTCTGCCCGCGGCAACTGGGCAATGTCTGCAGTTTTTACTGACAGCGCATCGGTCAGCCGTGCCAGGCCCTGCGTGCGCCGGCCGACAGC	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM15_0.5	CACAGGATAAAGGCATCTTTCACCACGACGCAGATTTTACTGACAGCACATCAGCACGGCGCAGATCTCCAGCTTTACGTCTGCCCGCGGC	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM16_0.5	CCGCATACGCGGCTGGGCGCTGGAAATCGCCATCTCCGCCAGGCAGTTCAGCCGCTGAACGTTCTGGTTCACCCCACAGGATAAAGGCATCT	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM17_0.5	GTACAATCAGCCAGTTTGCCAGCGGCGAGTCACCATCGCGGCAGGATTTCCAGCTTTTGGCCCGCGCAGGTCACGCATACGCGGCTGGGCG	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM18_0.5	GCCTGCTTTGGCAGATCCGCGTGCGGAGTCACCTTCCACAAGGAACAGCTCGGTACGGTTAAGGTCCTGCGGTACGAATCAAGCCAGTTTG	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM19_0.5	TCGGAAGAGACTTCCAAGGTTAAGGATCTTCACTTTCAGTGGCTGCATGATCGCCGTTGATTGAACTGCCTGGGAGTCGTTGGCCTGCTTTG	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM20_0.5	AGATCGTCGCTGTCAGGATGATACCGATCCGTTCCGGTTGATCAGTCGTGCACTTCCTGCGGCAGGTCGGCTGCTGGATCGGAAGAGACTTCCA	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM21_0.5	AGCGTGGCAATGTGCAGACCATCAGATCCGCCATCCGCGAGGATACAGATTTTGCCATAACGAAGCTGGCTCAGATCGTCGCTGTCAGGA	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM22_0.5	GGTGGCACTCGCTGCTTGCGCAGATACGCGCCGGAAATTTTACGAACGAAATTTTACGGTAGAGCGTGGCAATGTGCAGACCATCAGA	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM23_0.5	TCAAGTACGGCCTCTTTCTTCTTCCGTCGCCGGTATAAACCTTCTGGTTATCAGCAATACGGTAGAGCCGTATGGTGGCACTCGCTGCTTGC	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM24_0.5	TGCATCGGGTTCATTTCCCCCAGACCTTTAAAGACTGCTACCAGACTGGGCTTTCGTTTTGGCTCGATTGTCATTCATTCAAGTACGGCCTCTTTC	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM25_0.5	TGATCGTCTTCATCATCCGATATGTCAACAGCATCCCAGTGGTTTGCTGTTTCGGCTGGCCAATTGCATCGGGTTCATTTCCCC	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM26_0.5	TCTTGCAACCAGTTGCGGCATCTTCCGAGCGTTTCTTCGCCAGCAGACGCATATCATCGCCGTCCAGTACGCTGATCGTCTTCATCATC	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655
MparE.ORM27_0.5	CTTGTTTGCCCCGGCCATCCTGACCGGGCAATGTCTTTCCTTTTAAACCTCAATCTCCGCCATGTCGCCTTTCTCTTGCAACCAGTTGCGG	Spiked-oligonucleotide with 0,5% spiking, targeting the parE locus of E. coli K-12 MG1655

Appendix 3. List of PCR primers used in this study

	Oligonucleotides	
Oligo name	Sequence	Description
mutL2	5P-TCATCGCCATTTCACTCATCTTTCAGGGCTTTTATC	Primers for cloning *mutL* into pSIM8
mutL32F	5P-GTGAAAAACAGCCTCGATGC	Primers for introducing dominant E32K mutation into *mutL*
mutL32R	5P-TAGTTCTTTGACTACCGACGCAGG	Primers for introducing dominant E32K mutation into *mutL*
LExoF	5'-AAATGCCTGGTACTTTGCCAA	Primers for sequencing mutL on pORTMAGE
tL3R	5'-ATAACAGAAAGGCCGGGAA	Primers for sequencing mutL on pORTMAGE
pZAF	5'- GCTTAATTAGCTGAGTCTAGAGGC	Primers for amplification of pZA31 plasmid backbone for mutLE32K cloning
pZAR	5'- CCTTTCTCCTCTTTAATGAATTCGGT	Primers for amplification of pZA31 plasmid backbone for mutLE32K cloning
mutLRBS	5P-GAGAGGAGGTATATACATGCCAATTCAGGTCTTACCGCC	Primer for the introduction of a strong RBS upstream mutL on pORTMAGE1
pL32K frame_1	5'-CGGTACCCACTCTTCCTTTTTCAATAT	Primer for amplification of pORTMAGE1 from flanking region of Ap resistance cassette
pL32K frame_2	5'-CGGATCCTAACTGTCAGACCAAGTTTA	Primer for amplification of pORTMAGE1 from flanking region of Ap resistance cassette
Gibson Kan_Fw	5'-AAAGGAAGAGTGGTACCGAATATGTATCCGCTCATGAG	Primer for amplification of Km resistance cassette with homologous sequence to pORTMAGE for Gibson assembly
Gibson Kan_rev	5'-GGTCTGACAGTTAGGATCCG ATGAAGTTTTGACGGTATCG	Primer for amplification of Km resistance cassette with homologous sequence to pORTMAGE for Gibson assembly
Gibson Chlo_Fw	5'-AAAGGAAGAGTGGTACCG CGCAGAATAAATAAATCCTG	Primer for amplification of Cm resistance cassette with homologous sequence to pORTMAGE for Gibson assembly
Gibson Chlo_rev	5'-GGTCTGACAGTTAGGATCCG AGGCACCAATAACTGCCTTA	Primer for amplification of Cm resistance cassette with homologous sequence to pORTMAGE for Gibson assembly
ARCK1	5'-CTGAATTGCTATGTTTAGTGAGT	Checking primer for antibiotic resistance marker switch on pORTMAGE
ARCK2	5'-CACAGTACCCAATGATCCCA	Checking primer for antibiotic resistance marker switch on pORTMAGE
tetDS1	5'-AAGGCTAATTGATTTTCGAGA	Deep-sequencing primers for the landing-pad
tetDS2	5'-CATAAATTTGAGAGAAGAGTT	Deep-sequencing primers for the landing-pad
Cint_F	5'- ATGAAAACCGCTTACATTGCCAAACAACGTCAAATTAGCTTCGTAAAATCTCATTTTTCTTGTAGCAC CTGAAGTCAGC	Primer used for landing pad incorporation into Citrobacter freundii
Cint_R	5'- TTACAGCAGAGAAGCGACGCTTTCACGTACCTGGGCCGGCCATACGCCACACTGAACCTGGTAAGT TAGCGCGAATTGTC	Primer used for landing pad incorporation into Citrobacter freundii
Sint_F	5'- ATGAAAACCGCTTACATTGCCAAACAACGTCAAATTAGCTTCGTGAAATCACATTTCTCTTGTAGCAC CTGAAGTCAGC	Primer used for landing pad incorporation into Salmonella enterica LT2
Sint_R	5'- TTATAAAATAGCAGGAATGCTTTCGCGAACCTGCGCGGGCCATACGCCGCACTGTACCTGGTAAGTT AGCGCGAATTGTC	Primer used for landing pad incorporation into Salmonella enterica LT2
Eint_F	5'- ATGAAAACCGCTTACATTGCCAAACAACGTCAAATTAGCTTCGTGAAATCTCACTTTTCTCTGTAGCA CCTGAAGTCAGC	Primer used for landing pad incorporation into Escherichia coli K-12 MG1655
Eint_R	5'- TTACAGCAGAGAAGGGACGCTCTCGCGAACAGCAGCTGGCCATACTCCACACTGAACCTGGTAAGT TAGCGCGAATTGTC	Primer used for landing pad incorporation into Escherichia coli K-12 MG1655
CHK_F	5'- CAGATTGTCGATCAGATAATTTTCCAT	Primer to test for landing pad integration
CHK_R	5'- CATGCGCATTTTCGATCATATCTG	Primer to test for landing pad integration

Sequencing Data ID (.fastq in Single Molecule Real_Time Sequencing Data)	Sample	Amplification Primers	Barcode (5'-3')	Amplification Primer 1. (5'-3')	Amplification Primer 2. (5'-3')
SMRT_3	E. coli DIvERGE TRM5-1 (Sample 1)	EfolAF3-EfolAR3	>0017_Forward CATAGCGACTATCCGTG	CATAGCGACTATCCGTGTCATTGTAATGCGGCGAGTCCA	CACGATAGTCGCTATGAAGCGGCGGCGTCTTAAACA
SMRT_4	E. coli DIvERGE TRM5-2 (Sample 2)	EfolAF4-EfolAR4	>0019_Forward CGCATCTGTGCATGCA	CGCATCTGTGCATGCATCATTGTAATGCGGCGAGTCCA	TGCATGCACAGATGCGAAGCGGCGGCGTCTTAAACA
SMRT_5	E. coli DIvERGE TRM50-1 (Sample 1)	EfolAF5-EfolAR5	>0021_Forward GTACACGCTGTGACTA	GTACACGCTGTGACTATCATTGTAATGCGGCGAGTCCA	TAGTCACAGCGTGTACAAGCGGCGGCGTCTTAAACA
SMRT_6	E. coli DIvERGE TRM50-2 (Sample 2)	EfolAF6-EfolAR6	>0029_Forward GCTCGACTGTGTGAGAGA	GCTCGACTGTGTGAGAGATCATTGTAATGCGGCGAGTCCA	TCTCTCACAGTCGAGCAAGCGGCGGCGTCTTAAACA
SMRT_7	E. coli DIvERGE TRM200-1 (Sample 1)	EfolAF7-EfolAR7	>0032_Forward TATCTCTGTAGAGTCT	TATCTCTGTAGAGTCTTCATTGTAATGCGGCGAGTCCA	AGACTCTACAGAGATAAGCGGCGGCGTCTTAAACA
SMRT_8	E. coli DIvERGE TRM200-2 (Sample 2)	EfolAF8-EfolAR8	>0034_Forward ACTCTCGCTCTGTAGA	ACTCTCGCTCTGTAGATCATTGTAATGCGGCGAGTCCA	TCTACAGAGCGAGAGTAAGCGGCGGCGTCTTAAACA
SMRT_9	S. enterica DIvERGE TRM5-1 (Sample 1)	SfolAF11-SfolAR11	>0040_Forward CAGTGAGAGCGCGATA	CAGTGAGAGCGCGATAACCCGAGGTCAAACCGTCAATC	TATCGCGCTCTCACTGCGCTACGCTTATCAGGCCTACA
SMRT_10	S. enterica DIvERGE TRM5-2 (Sample 2)	SfolAF12-SfolAR12	>0041_Forward GTACATATGCGTCTGT	GTACATATGCGTCTGTACCCGAGGTCAAACCGTCAATC	ACAGACGCATATGTACCGCTACGCTTATCAGGCCTACA
SMRT_11	S. enterica DIvERGE TRM50-1 (Sample 1)	SfolAF17-SfolAR17	>0054_Forward GTGTGAGATATATATC	GCTCGTCGCGACACACCCGAGGTCAAACCGTCAATC	TGTGCGCGCGACGAGCCGCTACGCTTATCAGGCCTACA
SMRT_12	S. enterica DIvERGE TRM50-2 (Sample 2)	SfolAF14-SfolAR14	>0048_Forward TCACACTCTAGAGCGA	TCACACTCTAGAGCGAACCCGAGGTCAAACCGTCAATC	TCGCTCTAGAGTGTGACGCTACGCTTATCAGGCCTACA
SMRT_13	S. enterica DIvERGE TRM200-1 (Sample 1)	SfolAF15-SfolAR15	>0051_Forward ACACACAGACTGTGAG	ACACACAGACTGTGAGACCCGAGGTCAAACCGTCAATC	CTCACAGTCTGTGTGTCGCTACGCTTATCAGGCCTACA
SMRT_14	S. enterica DIvERGE TRM200-2 (Sample 2)	SfolAF16-SfolAR16	>0052_Forward GCAGACTCTCACACGC	GCAGACTCTCACACGCACCCGAGGTCAAACCGTCAATC	GCGTGTGAGAGTCTGCCGCTACGCTTATCAGGCCTACA
SMRT_15	E. coli CFT073 library after 5 DIvERGE cycles, selected on 3 ug/ml TRM, sample 1	folCDS11F - folCDS11R	>0070_Forward CTGCGCAGTACGTGCA	CTGCGCAGTACGTGCAGTGGACTCGCCAGCAGAATAT	TGCACGTACTGCGCAGTGCAGTGATCGTCGTGCTCCT
SMRT_16	E. coli CFT073 library after 5 DIvERGE cycles, selected on 3 ug/ml TRM, sample 2	folCDS12F - folCDS12R	>0075_Forward GAGATACGCTGCAGTC	GAGATACGCTGCAGTCGTGGACTCGCCAGCAGAATAT	GACTGCAGCGTATCTCTGCAGTCATGATCTCGTGCTCCT
SMRT_17	E. coli CFT073 library after 5 DIvERGE cycles, selected on 50 ug/ml TRM sample 1	folCDS13F - folCDS13R	>0082_Forward ATGCTCACTACTACAT	ATGCTCACTACTACATGTGGACTCGCCAGCAGAATAT	ATGTAGTAGTGAGCATTGCAGTCATGATCTCGTGCTCCT
SMRT_18	E. coli CFT073 library after 5 DIvERGE cycles, selected on 50 ug/ml TRM sample 2	folCDS14F - folCDS14R	>0009_Forward CTGCGTGCTCTACGAC	CTGCGTGCTCTACGACGTGGACTCGCCAGCAGAATAT	GTCGTAGAGCACGCAGTGCAGTCATGATCTCGTGCTCCT
SMRT_19	E. coli CFT073 library after 5 DIvERGE cycles, selected on 200 ug/ml TRM	folCDS15F - folCDS15R	>0002_Forward CTATACATGACTCTGC	CTATACATGACTCTGCGTGGACTCGCCAGCAGAATAT	GCAGAGTCATGTATAGTGCAGTCATGATCTCGTGCTCCT
SMRT_20	E. coli MG1655 folA naive library after 5+5 DIvERGE cycles	EfolAF4-EfolAR4	>0019_Forward CGCATCTGTGCATGCA	CGCATCTGTGCATGCATCATTGTAATGCGGCGAGTCCA	TGCATGCACAGATGCGAAGCGGCGGCGTCTTAAACA
SMRT_21	E. coli MG1655 folA naive library after 5+5 DIvERGE cycles, selected on 1000 ug/ml TRM	EfolAF7-EfolAR7	>0032_Forward TATCTCTGTAGAGTCT	TATCTCTGTAGAGTCTTCATTGTAATGCGGCGAGTCCA	AGACTCTACAGAGATAAGCGGCGGCGTCTTAAACA
SMRT_22	E. coli. MG1655 folA naive library after 5 DIvERGE cycles	EfolAF2-EfolAR2	>0010_Forward GCGCGATACGATGACT	GCGCGATACGATGACTTCATTGTAATGCGGCGAGTCCA	AGTCATCGTATCGCGCAAGCGGCGGCGTCTTAAACA
SMRT_23	E. coli MG1655 wild type folA control	EfolAF1-EfolAR1	>0001_Forward TCAGACGATGCCGTCAT	TCAGACGATGCCGTCATCATTGTAATGCGGCGAGTCCA	ATGACGCATCGTCTGAAAGCGGCGGCGTCTTAAACA
SMRT_24	E. coli CFT073 wild type folA control	folCDS16F-folCDS16R	>0003_Forward TACTAGAGTAGCACTC	TACTAGAGTAGCACTCGTGGACTCGCCAGCAGAATAT	GAGTGCTACTCTAGTATGCAGTCATGATCTCGTGCTCCT
SMRT_31	E. coli MG1655 gyrA-parE-parC library, selected on 2 fold of the wild-type MIC of Ciprofloxacin	F170GA2F - R182GA2R	>0070_Forward CTGCGCAGTACGTGCA // >0082_Reverse ATGTAGTAGTGAGCAT	GTCGAGTTGATCCATGGTCTCTGCGCAGTACGTGCAGGCACTTCTACTCCGTAATTGG	TGACACAGACGCATAGGATCATGTAGTAGTGAGCATAGCGGAAAGCAGTGCTATTG
SMRT_32	E. coli MG1655 gyrA-parE-parC library, selected on 2 fold of the wild-type MIC of Ciprofloxacin	F170GB2F - R182GB2R	>0070_Forward CTGCGCAGTACGTGCA // >0082_Reverse ATGTAGTAGTGAGCAT	GTCGAGTTGATCGCTCTCTGCGCAGTACGTGCAGAATATGCTGAGGTGCTGGAAC	TGACACAGACGCATAGGATCATGTAGTAGTGAGCATCATTGGCTGCGGACTAATTGCGGATTGC
SMRT_33	E. coli MG1655 gyrA-parE-parC library, selected on 2 fold of the wild-type MIC of Ciprofloxacin	F170PCF - R182PCR	>0070_Forward CTGCGCAGTACGTGCA // >0082_Reverse ATGTAGTAGTGAGCAT	GTCGAGTTGATCCATGGTCTCTGCGCAGTACGTGCAGCACCGCTGATTCCTATCTAC	TGACACAGACGCATAGGATCATGTAGTAGTGAGCATGTTCGATGCTGTCACCATGTC
SMRT_34	E. coli MG1655 gyrA-parE-parC library, selected on 2 fold of the wild-type MIC of Ciprofloxacin	F170PEF - R182PER	>0070_Forward CTGCGCAGTACGTGCA // >0082_Reverse ATGTAGTAGTGAGCAT	GTCGAGTTGATCCATGGTCTCTGCGCAGTACGTGCACATCACCCTGACGTTGAGTG	TGACACAGACGCATAGGATCATGTAGTAGTGAGCATCTGGAACGCACCAATGGAAG

Oligo ID	Sequence (5'-3')	Description
TET_DS3	AAACATATCCATGAAATCCCG	Landing pad Illumina deep-sequencing primer
TET_DS4	AATGTACTTTTGCTCCATCG	Landing pad Illumina deep-sequencing primer
folAF	TCATTGTAATGCGGCGAGTCCA	E. coli K12 MG1655 folA Illumina deep-sequencing primer
folAR	AAGCGGCGGCGTCTTAAACA	E. coli K12 MG1655 folA Illumina deep-sequencing primer
SFOL1	ACCCGAGGTCAAACCGTCAATC	Salmonella enterica LT2 folA Illumina deep-sequencing primer
SFD2	CGCTACGCTTATCAGGCCTACA	Salmonella enterica LT2 folA Illumina deep-sequencing primer
FOLCDS1	GTGGACTCGCCAGCAGAATAT	E. coli CFT073 UPEC folA Illumina deep-sequencing primer
folACR	TGCAGTCATGATCTCGTGCTCCT	E. coli CFT073 UPEC folA Illumina deep-sequencing primer
MgyaA_G288D	ACTCGTCACGCAGCGCGCTGATGCTTCCACGCGTTTTTCTTTTACCAGTTCCGCAATCTTCTCGATCAG	pORTMAGE allele-reconstruction oligo for Escherichia coli K12 MG1655 GyrA G288D
MGYRAS1	CTGCCGTCGATGGTATCCAGAT	E. coli K-12 gyrA sequencing primer 1.
MGYRAS2	TCGTTGGCGGGAAAGGTAAATCT	E. coli K-12 gyrA sequencing primer 1.
MGYRAS3	CGCATGACGGATCAGTTCGATG	E. coli K-12 gyrA sequencing primer 2.
MGYRAS4	CGGTCAGGGTAACTTCGGTTCT	E. coli K-12 gyrA sequencing primer 2.
UnivGA1	GCGAGAGAAATTACACCGGTCAACATT	gyrA sequencing primer 1.
UnivGA3	CATCGCGTTTCACTTCAATCACGAT	gyrA sequencing primer 2.
CFT073Pc1	GCTACATGAATTACGGAAAACGCCTACTT	Escherichia coli O6:K2:H1 CFT073 (UPEC) parC sequencing primer 1.
CFT073Pc3	TCACCTGATCCATATCCACGCGGTT	Escherichia coli O6:K2:H1 CFT073 (UPEC) parC sequencing primer 2.
8090Pc1	TTGAGGAACCACGATTAATGAGCGATA	Citrobacter freundii ATCC 8090 parC sequencing primer 1.
8090Pc3	AGGTGGTTCATCACGCTGCTCCATA	Citrobacter freundii ATCC 8090 parC sequencing primer 2.
gyrAF_ILM	TTCGTGGTCTACGTTATGGTTTAC	E. coli K-12 gyrA sequencing primer with MgyrAS3, for QRDR
folA2F	GATGAACCGGAAACGAAACCC	E. coli K-12 folA sequencing primer, with folAR
MfolA_A7T	GATCTACCGCTAACGtCGCAATCAGACTGATCATTGAGATTTCCCGATAAAAAAAATTGTCGCCACTATA	pORTMAGE allele-reconstruction oligo for Escherichia coli K12 MG1655 folA
MfolA_A7S	CGATAACGCGATCTACCGCTAACGaCGCAATCAGACTGATCATTGAGATTTCCCGATAAAAAAAATTGTC	pORTMAGE allele-reconstruction oligo for Escherichia coli K12 MG1655 folA
MfolA_M20I	GTTTATTTAAGGTGTTGCGTTTAAACCAGGCGAGATCGGCAGGCAGGTTCCACGGaATGGCGTTTTCCAT	pORTMAGE allele-reconstruction oligo for Escherichia coli K12 MG1655 folA
MfolA_M20V	TATTTTAAGGTGTTGCGTTTAAACCAGGCGAGATCGGCAGGCAGGTTCCACGGCAcGGCGTTTTCCATGCC	pORTMAGE allele-reconstruction oligo for Escherichia coli K12 MG1655 folA
MfolA_P21T	GTTTATTTAAGGTGTTGCGTTTAAACCAGGCGAGATCGGCAGGCAGGTTCCACGGCATGGCGTTTTCCAT	pORTMAGE allele-reconstruction oligo for Escherichia coli K12 MG1655 folA
MfolA_W30S	TAATCACGGGTTTATTTAAGGTGTTGCGTTTAAACgAGGCGAGATCGGCAGGCAGGTTCCACGGCATGGC	pORTMAGE allele-reconstruction oligo for Escherichia coli K12 MG1655 folA
MfolA_G97S	GTTTTTGCGCTTTTGGCAAGAACTGTTCATAAACGCGACtGCCGCCAATCACCATGATTTCTGGTACGTC	pORTMAGE allele-reconstruction oligo for Escherichia coli K12 MG1655 folA
MfolA_G-32C	GAGATTTCCCGATAAAAAAAATTGTCGCCAGTATACGTAAAGCGTAAACCGTCGTCGACTGGTGCGAGGA	pORTMAGE allele-reconstruction oligo for Escherichia coli K12 MG1655 folA
CfolA_W30S	TAATCACGGGTTTATTTAAAGTGTTGCGTTTAAACgAGGCGAGATCGGCAGGCAGGTTCCACGGCATGGC	pORTMAGE allele-reconstruction oligo for Escherichia coli O6:K2:H1 CFT073 (UPEC) folA
CfolA_G-32A	ATAAAAAAAATTGTCGCCATTATACGTAAAGCGTAAACCGTCGTCGACTGGTGCGAGGATGATGTTGAGG	pORTMAGE allele-reconstruction oligo for Escherichia coli O6:K2:H1 CFT073 (UPEC) folA
CfolA_C-58T	ATAAAAAAAATTGTCGCCACTATACGTAAAGCGTAAACCGTCGTCAACTGGTGCGAGGATGATGTTGAGG	pORTMAGE allele-reconstruction oligo for Escherichia coli O6:K2:H1 CFT073 (UPEC) folA
CfolA_G-54A	ATAAAAAAAATTGTCGCCACTATACGTAAAGCGTAAACCGTGTCGACTGGTGCGAGGATGATGTTGAGG	pORTMAGE allele-reconstruction oligo for Escherichia coli O6:K2:H1 CFT073 (UPEC) folA
CfolA_G-32C	ATAAAAAAAATTGTCGCCAGTATACGTAAAGCGTAAACCGTCGTCGACTGGTGCGAGGATGATGTTGAGG	pORTMAGE allele-reconstruction oligo for Escherichia coli O6:K2:H1 CFT073 (UPEC) folA
CFT_folA_I5F	ATCTACCGCTAACGCCGCAAaCAGACTGATCATTGAGATTTCCCGATAAAAAAAATTGTCGCCACTATAC	pORTMAGE allele-reconstruction oligo for Escherichia coli O6:K2:H1 CFT073 (UPEC) folA
CFT_folA_D27E	CACGGGTTTATTTAAAGTGTTGCGTTTAAACCAGGCGAGTCGGCAGGCAGGTTCCACGGCATGGCGTTT	pORTMAGE allele-reconstruction oligo for Escherichia coli O6:K2:H1 CFT073 (UPEC) folA
CFT_folA_L28R	GGGTTTATTTAAAGTGTTGCGTTTAAACCAGGCGcGATCGGCAGGCAGGTTCCACGGCATGGCGTTTTCC	pORTMAGE allele-reconstruction oligo for Escherichia coli O6:K2:H1 CFT073 (UPEC) folA
CFT_folA_I94L	GCTTTCGGCAGGAACTGCTCATAAACGCGACCGCCGCCAAgCACCATGATTTCTGGTACGTCACCACACG	pORTMAGE allele-reconstruction oligo for Escherichia coli O6:K2:H1 CFT073 (UPEC) folA
CFT_folA_R98P	TTTTTGCGCTTTCGGCAGGAACTGCTCATAAACGgGACCGCCGCCAATCACCATGATTTCTGGTACGTCA	pORTMAGE allele-reconstruction oligo for Escherichia coli O6:K2:H1 CFT073 (UPEC) folA
CFT_folA_A7S	TTTTTCCATGCCGATAACGCGATCTACCGCTAACGaCGCAATCAGACTGATCATTGAGATTTCCCGATAA	pORTMAGE allele-reconstruction oligo for Escherichia coli O6:K2:H1 CFT073 (UPEC) folA
CFT_folA_A7T	TTTTTCCATGCCGATAACGCGATCTACCGCTAACGtCGCAATCAGACTGATCATTGAGATTTCCCGATAAA	pORTMAGE allele-reconstruction oligo for Escherichia coli O6:K2:H1 CFT073 (UPEC) folA
CFT_folA_M20I	TAAACCAGGCGAGATCGGCAGGCAGGTTCCACGGaATGGCGTTTTCCATGCCGATAACGCGATCTACCGC	pORTMAGE allele-reconstruction oligo for Escherichia coli O6:K2:H1 CFT073 (UPEC) folA
CFT_folA_P21L	TTTAAACCAGGCGAGATCGGCAGGCAGGTTCCACaGCATGGCGTTTTCCATGCCGATAACGCGATCTACC	pORTMAGE allele-reconstruction oligo for Escherichia coli O6:K2:H1 CFT073 (UPEC) folA
CFT_folA_W30R	ATCACGGGTTTATTTAAAGTGTTGCGTTTAAACcgGGCGAGATCGGCAGGCAGGTTCCACGGCATGGCGT	pORTMAGE allele-reconstruction oligo for Escherichia coli O6:K2:H1 CFT073 (UPEC) folA
CFT_folA_W30G	ATCACGGGTTTATTTAAAGTGTTGCGTTTAAACcGGCGAGATCGGCAGGCAGGTTCCACGGCATGGCGT	pORTMAGE allele-reconstruction oligo for Escherichia coli O6:K2:H1 CFT073 (UPEC) folA
CFT_folA_F153S	TAATATACAAAATTACCGCCGCTCCAGAATCTCAgAGCAATAGCTGTGAGAGTTCTGCGCATCGGCATCG	pORTMAGE allele-reconstruction oligo for Escherichia coli O6:K2:H1 CFT073 (UPEC) folA
MG_GyrA_S83L	CGAGAATGGCTGCGCCATGCGGACGATCGTGTCATAGACCGCCAAGTCACCATGGGGATGGTATTTACCGGATTACGTCACCAACGACACG	pORTMAGE allele-reconstruction oligo for Escherichia coli K-12 MG1655 GyrA
MG_GyrA_S83L_D87N	CGCCATGCGGACGATCGTGTtATAGACCGCCAaGTCACCATGGGGATGGTATTTACCGGATTACGTCAACGACACGGGGCAGATTTTTT	pORTMAGE allele-reconstruction oligo for Escherichia coli K-12 MG1655 GyrA
KPGAD82N	GCACGATGGTGTCGTATACCGCGGAGTTGCCGTGCGGGTGGTATTTACCGGATTACGTCACCAACGACACGGGGCTGATTTTTTATAGGCTT	pORTMAGE allele-reconstruction oligo for Klebsiella pneumoniae subsp. pneumoniae
KPPCD79N	GCACCATCGCTTCATAGCAGGCGCTGTTGCCGTGCGGGTGATATTTACCCAACACGTCGCCGACGGTGCGGGCGGACTTTTTGAATTTCG	pORTMAGE allele-reconstruction oligo for Klebsiella pneumoniae subsp. pneumoniae
CFGA82N	GACGTAATCAGGTAAATACCACCCTCATGGTaatACCGCCGTTTACGACACCATTGTTCGTATGGCGCAGCCATTCTCCTTGCGTTACATG	pORTMAGE allele-reconstruction oligo for Citrobacter freundii ATCC 8090
CFPC79N	AAATTTAAAAAATCCGCCCGTACCGTCGGTGACGTGCTGGGTAAATACCATCCGCACGGCaacAGCGCCTGCTATGAAGCGATGGTGCTG	pORTMAGE allele-reconstruction oligo for Citrobacter freundii ATCC 8090
CFTGA82N	ACGATCGTGTCATAAACCGCCGAgttACCATGGGGATGGTATTTACCGGATTACGTCACCAACGACACGGGCAGATTTTTATAGGCTTTG	pORTMAGE allele-reconstruction oligo for Escherichia coli O6:K2:H1 CFT073 (UPEC)
CFTPC79N	ATAACGGTAAGAGAACGGCTGCGCCATCAGGACCATCGCTTCATAACAGGCACTattGCCCGTGCGGATGGTATTTACCCAGTACGTCACC	pORTMAGE allele-reconstruction oligo for Escherichia coli O6:K2:H1 CFT073 (UPEC)

Salmonella enterica

Mutation rate (mutations/bp/generation)

3.0E-05 / 2.0E-05 / 1.0E-05 / 0.0E+00

Uninduced — Induced

Escherichia hermanii

Mutation rate (mutations/bp/generation)

4.0E-06 / 3.0E-06 / 2.0E-06 / 1.0E-06 / 0.0E+00

Uninduced — Induced

Citrobacter freundii

Mutation rate (mutations/gene/generation)

6.0E-06 / 4.0E-06 / 2.0E-06 / 0.0E+00

Uninduced — Induced

Escherichia coli K-12 MG1655

Mutation rate (mutations/gene/generation)

1.0E-05 / 5.0E-06 / 0.0E+00

Uninduced — Induced

E. coli BL21(DE3)

Mutation rate (mutations/gene/generation)

2.5E-06 / 2.0E-06 / 1.5E-06 / 1.0E-06 / 5.0E-07 / 0.0E+00

uninduced — induced

Edwardsiella tarda

Mutation rate (mutations/gene/generation)

4.5E-06 / 4.0E-06 / 3.5E-06 / 3.0E-06 / 2.5E-06 / 2.0E-06 / 1.5E-06 / 1.0E-06 / 5.0E-07 / 0.0E+00

uninduced — induced

Appendix 5. Off-target mutations identified by whole-genome sequencing

Parental strain: *E. coli* K-12 MG1655 + pORTMAGE

Reference Position	Type	Reference	Allele	Count	Coverage	Frequency	Overlapping annotations	Coding region change	Amino acid change
69999	SNV	A	T	49	49	100	Gene: araB, CDS: araB	AAC73174.1:c.50T>A	AAC73174.1:p.Leu17*
365654	SNV	T	A	49	49	100	Gene: lacZ, CDS: lacZ	AAC73447.1:c.652A>T	AAC73447.1:p.Lys218*
2093633	SNV	C	T	66	66	100	Gene: hisB, CDS: hisB	AAC75083.2:c.166C>T	AAC75083.2:p.Gln56*
3474425	SNV	T	C	59	59	100	Gene: rpsL, CDS: rpsL	AAC76367.1:c.128A>G	AAC76367.1:p.Lys43Arg
4247035	SNV	C	G	37	37	100	Gene: malK, CDS: malK	AAC77005.1:c.252C>G	AAC77005.1:p.Tyr84*
4296381	Insertion	-	CG	17	17	100	Repeat region: RIP321 (repetitive extragenic palindromic) element;contains 11 REP sequences and 4 IHF sites		
4430002	MNV	AA	TG	68	68	100	Gene: cycA, CDS: cycA	AAC77165.1:c.139_140delAAin	AAC77165.1:p.Lys47*
				#					
Targeted mutations:				6					
Mutation from parental strain:				1					
Off-target mutations:				0					

Parental strain: *E. coli* K-12 MG1655 + pSIM8

Reference Position	Type	Reference	Allele	Count	Coverage	Frequency	Overlapping annotations	Coding region change	Amino acid change
69999	SNV	A	T	82	83	99	Gene: araB, CDS: araB	AAC73174.1:c.50T>A	AAC73174.1:p.Leu17*
365654	SNV	T	A	76	76	100	Gene: lacZ, CDS: lacZ	AAC73447.1:c.652A>T	AAC73447.1:p.Lys218*
703548	SNV	G	C	72	72	100	Gene: nagB, CDS: nagB	AAC73273.1:c.67C>G	AAC73273.1:p.Arg23Gly
2093633	Deletion	CAGGA	-	77	77	100	Gene: hisB, CDS: hisB	AAC75083.2:c.166_170delCAGGA	AAC75083.2:p.Gln56fs
2093639	Replacement	GGTCT	A	77	77	100	Gene: hisB, CDS: hisB	AAC75083.2:c.172_176delGGTCTinsA	AAC75083.2:p.Gly58fs
3904366	SNV	G	T	26	27	96			
3474425	SNV	T	C	70	73	96	Gene: rpsL, CDS: rpsL	AAC76367.1:c.128A>G	AAC76367.1:p.Lys43Arg
4247035	SNV	C	G	84	84	100	Gene: malK, CDS: malK	AAC77005.1:c.252C>G	AAC77005.1:p.Tyr84*
4296381	Insertion	-	CG	50	50	100	Repeat region: RIP321 (repetitive extragenic palindromic) element;contains 11 REP sequences and 4 IHF sites		
4430002	MNV	AA	TG	84	84	100	Gene: cycA, CDS: cycA	AAC77165.1:c.139_140delAAinsTG	AAC77165.1:p.Lys47*

				#					
Targeted mutations:				5					
Mutation from parental strain:				1					
Off-target mutations:				2					

Parental strain: *E. coli* K-12 MG1655 ΔmutS + pSIM8

Reference Position	Type	Reference	Allele	Count	Coverage	Frequency	Overlapping annotations	Coding region change	Amino acid change
26396	SNV	A	G	51	51	100	Gene: ispH, CDS: ispH	AAC73140.1:c.120A>G	
37584	SNV	C	T	52	53	98	Gene: caiC, CDS: caiC	AAC73148.2:c.241G>A	AAC73148.2:p.Gly81Arg
47501	SNV	C	T	58	61	95	Gene: keff, CDS: keff	AAC73157.1:c.356C>T	AAC73157.1:p.Pro119Leu
69999	SNV	A	T	69	69	100	Gene: araB, CDS: araB	AAC73174.1:c.50T>A	AAC73174.1:p.Leu17*
131653	SNV	C	T	50	53	94	Gene: acnB, CDS: acnB	AAC73229.1:c.39C>T	
139086	SNV	C	T	82	84	98	Gene: gcd, CDS: gcd	AAC73235.1:c.2140G>A	AAC73235.1:p.Gly714Arg
252969	SNV	C	T	70	70	100	Gene: yafP, CDS: yafP	AAC73333.1:c.261C>T	
253413	SNV	T	C	38	48	79			
278118	SNV	C	T	28	29	97	Gene: afuC, CDS: afuC, Misc. feature: cryptic prophage CP4-6	AAC73365.2:c.685G>A	AAC73365.2:p.Ala229Thr
345829	SNV	A	G	53	53	100	Gene: yahN, CDS: yahN	AAC73431.1:c.509T>C	AAC73431.1:p.Val170Ala
347605	SNV	G	A	49	49	100	Gene: prpR, CDS: prpR	AAC73433.1:c.839C>T	AAC73433.1:p.Ala280Val
365654	SNV	T	A	51	51	100	Gene: lacZ, CDS: lacZ	AAC73447.1:c.652A>T	AAC73447.1:p.Lys218*
407729	SNV	A	G	83	85	98	Gene: aroM, CDS: aroM	AAC73493.1:c.302A>G	AAC73493.1:p.Asn101Ser
415870	SNV	G	A	74	75	99	Gene: sbcD, CDS: sbcD	AAC73501.1:c.1083C>T	
542805	SNV	G	A	48	48	100	Gene: glxK, CDS: glxK	AAC73616.1:c.918G>A	
552660	SNV	A	G	37	37	100	Gene: purE, CDS: purE	AAC73625.1:c.441T>C	
636153	SNV	T	C	70	70	100	Gene: ybdN, CDS: ybdN	AAC73703.1:c.417A>G	
705964	Deletion	T	-	41	46	89			
899485	SNV	C	T	50	54	93	Gene: rlmC, CDS: rlmC	AAC73946.1:c.968C>T	AAC73946.1:p.Thr323Met
987523	SNV	A	G	57	57	100			
1015009	SNV	C	T	51	51	100	Gene: pqiC, CDS: pqiC	AAC74038.2:c.114C>T	
1037429	SNV	G	A	75	79	95	Gene: hyaF, CDS: hyaF	AAC74062.1:c.681G>A	
1053197	SNV	A	G	52	52	100	Gene: yccM, CDS: yccM	AAC74077.1:c.166T>C	AAC74077.1:p.Tyr56His
1082600	SNV	A	G	65	65	100	Gene: efeO, CDS: efeO	AAC74103.1:c.358A>G	AAC74103.1:p.Thr120Ala
1105797	SNV	A	G	65	65	100			
1224114	SNV	T	C	57	59	97			
1225089	SNV	T	C	68	68	100	Gene: minD, CDS: minD	AAC74259.1:c.273A>G	
1485344	SNV	C	T	48	48	100	Gene: hrpA, CDS: hrpA	AAC74495.2:c.2284C>T	AAC74495.2:p.Arg762Cys
1575474	SNV	C	T	91	95	96	Gene: yddB, CDS: yddB	AAC74568.1:c.2146G>A	AAC74568.1:p.Glu716Lys
1595306	SNV	C	T	58	58	100	Gene: yneO, CDS: yneO	yneO:c.2682G>A	
1603448	SNV	T	C	56	56	100	Gene: lsrC, CDS: lsrC	AAC74587.1:c.430T>C	
1660661	SNV	A	G	79	79	100	Gene: ynfF, CDS: ynfF	AAC74660.4:c.106A>G	AAC74660.4:p.Thr36Ala
1781659	SNV	G	A	59	62	95	Gene: ydiS, CDS: ydiS	AAC74769.1:c.265G>A	AAC74769.1:p.Asp89Asn
1872236	SNV	C	T	81	82	99	Gene: yeaI, CDS: yeaI	AAC74856.2:c.196C>T	AAC74856.2:p.Arg66Cys
1902259	SNV	G	A	74	75	99	Gene: manX, CDS: manX	AAC74887.1:c.212G>A	AAC74887.1:p.Gly71Asp
1904680	SNV	G	A	45	45	100	Gene: manZ, CDS: manZ	AAC74889.2:c.78G>A	
1938164	SNV	A	G	61	62	98	Gene: pykA, CDS: pykA	AAC74924.1:c.516A>G	
1980655	SNV	G	A	71	71	100	Gene: otsA, CDS: otsA	AAC74966.1:c.958C>T	AAC74966.1:p.Arg320*
2071002	SNV	G	A	57	57	100	Gene: yeeP, CDS: yeeP, Misc. feature: cryptic prophage CP4-44	yeeP:c.343G>A	yeeP:p.Ala115Thr
2093633	SNV	C	T	47	47	100	Gene: hisB, CDS: hisB	AAC75083.2:c.166C>T	AAC75083.2:p.Gln56*
2208450	SNV	G	A	45	45	100	Gene: yehP, CDS: yehP	AAC75182.1:c.484G>A	AAC75182.1:p.Val162Ile
2283699	SNV	C	T	66	67	99	Gene: yejK, CDS: yejK	AAC75247.1:c.249G>A	
2319619	SNV	A	G	76	77	99	Gene: rcsC, CDS: rcsC	AAC75278.2:c.258T>C	
2326914	SNV	G	A	47	47	100	Gene: atoB, CDS: atoB	AAC75284.1:c.806G>A	AAC75284.1:p.Gly269Asp
2386124	SNV	A	G	57	57	100	Gene: yfbL, CDS: yfbL	AAC75331.2:c.265A>G	AAC75331.2:p.Thr89Ala
2638541	SNV	G	A	45	48	94	Gene: glrR, CDS: glrR	AAC75607.1:c.263C>T	AAC75607.1:p.Ser88Phe
2753329	SNV	C	T	62	62	100	Gene: recN, CDS: recN	AAT48145.1:c.1535C>T	AAT48145.1:p.Ala512Val
2818145	SNV	T	C	12	13	92			
2818145	SNV	T	C	11	13	85			
2818158	SNV	T	A	10	10	100			
2818162	SNV	A	T	10	10	100			
2819391	SNV	T	C	66	67	99	Gene: alaS, CDS: alaS	AAC75739.1:c.2621A>G	AAC75739.1:p.Lys874Arg
2836760	SNV	G	A	39	40	98	Gene: hypF, CDS: hypF	AAC75754.1:c.666C>T	
2838000	SNV	T	C	74	74	100	Gene: hydN, CDS: hydN	AAC75755.1:c.106A>G	AAC75755.1:p.Thr36Ala
2867032	SNV	C	T	55	56	98	Gene: rpoS, CDS: rpoS	AAC75783.1:c.520G>A	AAC75783.1:p.Glu174Lys
2883989	SNV	C	T	71	71	100	Gene: casA, CDS: casA	AAC75802.1:c.150G>A	
2940575	SNV	G	A	51	51	100	Gene: rlmM, CDS: rlmM	AAC75848.1:c.669C>T	
2956746	SNV	T	C	49	50	98	Gene: ptrA, CDS: ptrA	AAC75860.1:c.2139A>G	
2975948	SNV	G	A	61	61	100	Gene: aas, CDS: aas	AAC75875.1:c.67C>T	AAC75875.1:p.Gln23*
3053403	SNV	G	A	46	46	100	Gene: ubiH, CDS: ubiH	AAC75945.1:c.116C>T	AAC75945.1:p.Ser39Leu
3100610	SNV	T	C	77	77	100	Gene: ansB, CDS: ansB	AAC75994.1:c.119A>G	AAC75994.1:p.Asp40Gly
3132771	SNV	T	C	38	45	84	Gene: yghR, CDS: yghR	AAC76020.1:c.442A>G	AAC76020.1:p.Ile148Val
3155620	SNV	T	C	37	37	100	Gene: ychD, CDS: ychD	AAC76047.1:c.266T>C	AAC76047.1:p.Leu89Pro
3222157	SNV	G	A	71	71	100	Gene: ebgR, CDS: ebgR	AAC76110.1:c.692G>A	AAC76110.1:p.Arg231Gln
3252467	SNV	T	C	59	62	95	Gene: ygaH, CDS: ygaH	AAC76138.1:c.164T>C	AAC76138.1:p.Leu55Pro
3263556	SNV	G	A	82	83	99	Gene: tdcD, CDS: tdcD	AAC76150.2:c.105C>T	
3299463	SNV	C	T	35	35	100	Gene: yhbO, CDS: yhbO	AAC76187.2:c.490C>T	AAC76187.2:p.Arg164Cys
3304164	SNV	T	C	40	42	95	Gene: yhbW, CDS: yhbW	AAC76194.1:c.717T>C	
3342102	Insertion	-	G	74	84	88	Gene: kdsD, CDS: kdsD	AAC76229.1:c.836_837insG	AAC76229.1:p.Pro279fs
3350200	SNV	T	C	63	63	100	Gene: elbB, CDS: elbB	AAC76241.2:c.260A>G	AAC76241.2:p.Asp87Gly
3474425	SNV	T	C	54	54	100	Gene: rpsL, CDS: rpsL	AAC76367.1:c.128A>G	AAC76367.1:p.Lys43Arg
3542923	SNV	G	A	50	57	88	Gene: feoC, CDS: feoC	AAC76435.1:c.19G>A	AAC76435.1:p.Glu6Lys
3546156	SNV	G	A	47	48	98	Gene: nfuA, CDS: nfuA	AAC76439.1:c.533G>A	AAC76439.1:p.Arg178His
3586147	SNV	T	C	59	60	98	Gene: ggt, CDS: ggt	AAC76472.1:c.677A>G	AAC76472.1:p.Asn226Ser
3742817	SNV	T	C	50	51	98	Gene: ylaK, CDS: ylaK	AAC76599.1:c.85T>C	AAC76599.1:p.Cys29Arg
3760852	SNV	G	A	57	57	100	Gene: selA, CDS: selA	AAC76615.1:c.398C>T	AAC76615.1:p.Ala133Val
3778841	SNV	T	C	53	53	100	Gene: lldP, CDS: lldP	AAC76627.1:c.1443T>C	
3900055	SNV	G	A	101	101	100	Gene: yieL, CDS: yieL	AAC76742.3:c.523C>T	AAC76742.3:p.Pro175Ser
4175885	SNV	A	G	61	61	100			
4182512	SNV	A	G	60	60	100	Gene: rpoB, CDS: rpoB	AAC76961.1:c.1268A>G	AAC76961.1:p.Asp423Gly
4232708	SNV	C	T	74	77	96	Gene: lysC, CDS: lysC	AAC76994.1:c.52G>A	AAC76994.1:p.Glu17Lys
4247035	SNV	C	G	39	39	100	Gene: malK, CDS: malK	AAC77005.1:c.252C>G	AAC77005.1:p.Tyr84*
4296381	Insertion	-	CG	30	30	100	Repeat region: RIP321 (repetitive extragenic palindromic) element;contains 11 REP sequences and 4 IHF sites		
4326999	SNV	A	G	41	41	100			
4411889	SNV	C	T	67	67	100	Gene: yjfK, CDS: yjfK	AAC77140.1:c.588C>T	
4430002	MNV	AA	TG	82	82	100	Gene: cycA, CDS: cycA	AAC77165.1:c.139_140delAAin	AAC77165.1:p.Lys47*
4479257	SNV	T	C	61	61	100	Gene: yjgM, CDS: yjgM	AAT48245.1:c.281A>G	AAT48245.1:p.Tyr94Cys
4529218	SNV	C	T	52	52	100	Gene: sgcC, CDS: sgcC	AAC77260.1:c.1026G>A	
4563609	SNV	T	C	78	84	93			
4567871	SNV	A	G	68	68	100	Gene: mdtM, CDS: mdtM	AAC77293.1:c.649T>C	AAC77293.1:p.Phe217Leu
4610455	SNV	T	C	50	50	100	Gene: prfC, CDS: prfC	AAC77328.1:c.1042T>C	AAC77328.1:p.Phe348Leu

				#					
Targeted mutations:				6					
Mutation from parental strain:				1					
Off-target mutations:				84					

Appendix 6. Data deposition

Sequencing data reported in this thesis have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive at https://www.ncbi.nlm.nih.gov/sra/SRP144255 (Accession number: SRP144255)