MANUSCRIPT TITLE

MOSAIC: a highly efficient, one-step recombineering approach to plasmid editing and diversification

AUTHORS

Marijn van den Brink[1,†], Timotheus Y. Althuis[2,†], Christophe Danelon[1,3,*] and Nico J. Claassens[2,*]

[1] Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, 2629 HZ Delft, The Netherlands

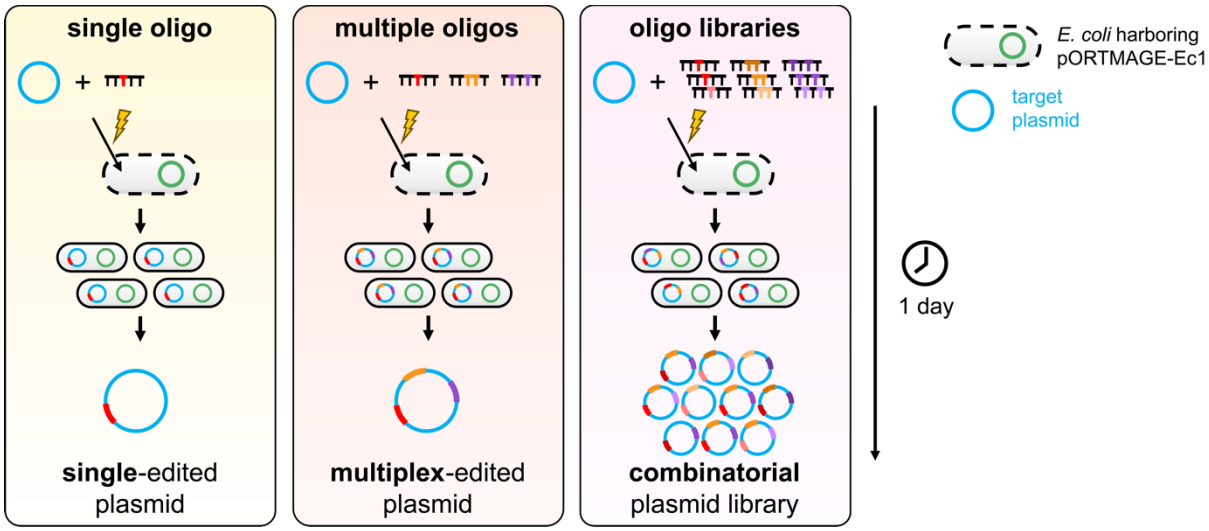
[2] Laboratory of Microbiology, Wageningen University, Stippeneng 4, 6708 WE Wageningen, The Netherlands

[3] Toulouse Biotechnology Institute (TBI), Université de Toulouse, CNRS, INRAE, INSA, 31077 Toulouse, France

* To whom correspondence should be addressed. Email: nico.claassens@wur.nl. Correspondence may also be addressed to Christophe Danelon. Email: danelon@insa-toulouse.fr.

† Joint Authors

GRAPHICAL ABSTRACT



ABSTRACT

The editing of plasmids and construction of plasmid libraries is paramount to the engineering of desired functionalities in synthetic biology. Typically, plasmids with targeted mutations are produced through time- and resource-consuming DNA amplification and/or cloning steps. In this study, we establish MOSAIC, a highly efficient protocol for the editing of plasmids and generation of combinatorial plasmid libraries. This one-step protocol employs efficient single-stranded DNA annealing proteins (SSAP) to incorporate (libraries of) DNA oligos harboring the desired mutations into a target plasmid in *E. coli*. MOSAIC can be used to modify virtually any plasmid and is integrated with a validation pipeline based on Nanopore sequencing. In addition to up to 90% single-target plasmid editing efficiency, MOSAIC is demonstrated to enable the generation of a combinatorial plasmid library spanning four different target regions on a plasmid, in a single transformation. We anticipate that MOSAIC will provide researchers with a simple, rapid and resource-effective method to edit plasmids or generate large, diverse plasmid libraries for a wide range of *in vivo* or *in vitro* applications in molecular and synthetic biology.

Keywords: plasmid libraries; recombineering; MAGE; multiplex; combinatorial libraries; Nanopore sequencing

INTRODUCTION

The development of tools that enable the construction of DNA parts and variants thereof are driving the field of synthetic biology. With the continuing advances in modelling and machine-learning, our predictive capabilities and *a priori* design of functional genetic systems and proteins are rapidly improving (1–5). Still, *in silico* designed genetic parts and proteins often do not behave as expected in the complex genetic and molecular contexts of cells or cell-free expression systems. As a result, strain and protein engineering efforts often benefit from the exploration of wide solution spaces provided by (semi-)rational design and/or computational tools (6, 7).

Plasmid libraries offer an efficient means to test a range of designs *in vivo* or *in vitro*. Generally, mutant plasmids are constructed from DNA fragments amplified by polymerase chain reaction (PCR) with degenerate primers to introduce the desired variation at specific locations. Subsequently, the produced DNA fragments are (re)assembled into a plasmid *in vitro* or *in vivo* by enzymatic assembly (8, 9). For example, site-directed mutagenesis methods based on high-fidelity polymerases and mutagenic primers are widely used (e.g., QuikChange or Q5 site-directed mutagenesis). However, these methods can only diversify one region at a time. Alternative approaches, including restriction enzyme-based and homology-based assembly methods (e.g., Golden Gate, Gibson Assembly, In-Fusion Cloning or Ligase Cycling Reaction) can create plasmids from multiple DNA parts harboring mutations, but are limited in their efficiency and flexibility for the generation of large combinatorial libraries (10–13). Specifically, the number of correctly assembled clones decreases as the number of assembly parts increases. Therefore, researchers need to upscale their experimental efforts or rely on laboratory automation to obtain sufficient numbers of clones to generate larger libraries. These challenges are exacerbated for combinatorial libraries with multiple target sites (multiplexing), large plasmids resulting in low transformation efficiencies, or plasmids containing repetitive DNA. Altogether, there is a need for efficient and flexible strategies to generate large, multiplex plasmid libraries.

Recombineering is a widely used tool to introduce targeted and scarless modifications in bacterial genomes (14). This approach relies on single-stranded DNA (ssDNA) or double-stranded DNA (dsDNA) fragments containing desired mutations, which are introduced into the organism, usually by electroporation. These DNA fragments are then incorporated into replicating chromosomes using phage-derived ssDNA-annealing proteins (SSAP). This system is extensively employed for genome engineering, especially in *Escherichia coli*, to make large insertions and deletions using dsDNA recombineering and small edits using ssDNA (15, 16). Multiplex automated genome engineering (MAGE) builds on the latter by introducing mutations to many genomic loci at the same time using iterative, automated or manual, editing cycles (17).

Whilst underutilized, ssDNA-mediated recombineering has also been employed to modify plasmids in *E. coli* (18–23). Initially, plasmid recombineering efficiencies of 5-10% were observed for single point mutations with the phage λ-derived SSAP Recβ and two sequential transformations of *E. coli,* first with the target plasmid and then with the mutagenic ssDNA oligos (19). Later, co-electroporation of an optimized ratio of mutagenic ssDNA and the target plasmid yielded editing efficiencies of 20-30% (20). This was further improved to 60% when combined with a co-selection strategy, in which a restriction site on the plasmid is

simultaneously mutated, whereafter unmodified variants are eliminated by restriction digestion. Higher efficiencies have also been obtained by combining recombineering with counterselection of non-mutated variants by a CRISPR-Cas nuclease (21). However, this approach requires additional, time-consuming cloning steps and complicates the experimental design and setup. To the best of our knowledge, only a few studies have applied recombineering-based approaches to produce diversified and multiplex plasmid libraries (20, 23). Presumably, the low efficiency of plasmid recombineering and laborious methods relying on co- or counterselection explain the limited application range thus far.

Recently, a systematic screen of phage SSAPs in *E. coli* identified CspRecT, which has a two-fold higher genomic recombineering efficiency than the commonly used Recβ (24). This prompted us to develop MOSAIC: a multiplex one-step SSAP-mediated plasmid diversification protocol. Its name is derived from *mosaicism*, a phenomenon where mutations give rise to distinct genetic compositions within an organism or a cell population. In this study, we show that MOSAIC's high plasmid editing efficiency enables the generation of large (10^4 variants) combinatorial plasmid libraries in a single transformation. Furthermore, MOSAIC employs a validation methodology based on Nanopore long-read sequencing, which quantifies the frequency of (multiplex) library variants directly from the plasmid library sample. We believe that the easy experimental and sequence validation protocols of MOSAIC will facilitate plasmid diversification and expand its range of applications throughout many laboratories.

MATERIAL AND METHODS

Reagents

Chemicals were purchased from Sigma-Aldrich, unless stated otherwise. M-Toluic acid was dissolved in ethanol at a concentration of 1 M and stored at -20 °C. Plasmids were isolated from bacterial cells using PureYield Plasmid Miniprep System (Promega) or QIAprep Spin Miniprep Kit (Qiagen). DNA concentrations were measured using NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific), DS-11 FX spectrophotometer (DeNovix) or Qubit 4 Fluorometer (Invitrogen). Electroporation was performed with 0.1-cm gap Gene Pulser/MicroPulser electroporation cuvettes (Bio-Rad Laboratories). Different electroporators were used which all performed robustly for the MOSAIC protocol: the Eppendorf Eporator electroporator (Eppendorf) (1.8 kV) and the ECM 630B electroporator (BTX) (1.8 kV, 200 Ω, 25 µF).

Biological resources

E. coli strains used for the recombineering experiments were *E. coli* K-12 MG1655 (Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures GmbH, Germany) and *E. coli* K-12 DH5α. The bacteria were grown in Lysogeny Broth (LB) medium or on LB agar plates containing antibiotics (kanamycin, ampicillin, apramycin) at a concentration of 50 µg/mL unless indicated otherwise. The plasmids used in this study are listed in **Table 1**. pORTMAGE-Ec1 was a gift from the lab of George Church (Addgene plasmid #138474; http://n2t.net/addgene:138474; RRID:Addgene_138474). pUC19 was acquired from New England Biolabs. pSEVAb plasmids were cloned according to the method reported earlier (25).

Plasmid G555 was constructed by subcloning of the construct containing genes *plsB*, *plsC*, *cdsA* and *pssA* (amplified by primers 1285 ChD and 1286 ChD from plasmid G363) into the backbone of plasmid G340 (amplified by primers 1287 ChD and 1288 ChD) via restriction enzyme digestion (NcoI/XhoI) and ligation. Primers 1285 ChD-1288 ChD are listed in **Supplementary Table 1**. G340 was constructed as described before (paper to be published soon on BiorXiv). G363 was assembled using a stepwise Golden Gate ligation of six PCR fragments containing independent transcriptional cassettes. First, *plsB* and *plsC*, *cdsA* and *pssA*, *tp* and *dnap* and Phi29 origins were ligated. Then, these three fragments and the pTU1 backbone (Addgene #72934) were ligated to form G363.

Table 1. List of plasmids used in this study.

Plasmid name	Addgene number	Origin of replication	Antibiotic resistance marker	Target region in plasmid for MOSAIC
pORTMAGE-Ec1	#138474	RSF1010	kanamycin	-
pUC19	#50005	pUC	ampicillin	*lacZ*
pSEVAb827	#217500	RK2	apramycin	*sfGFP*
pSEVAb837	#217501	pBBR1	apramycin	*sfGFP*
pSEVAb847	#217502	pRO1600/ColE1	apramycin	*sfGFP*
pSEVAb867	#217503	p15A	apramycin	*sfGFP*
pSEVAb887	#217504	pUC	apramycin	*sfGFP*
pSEVAb897	#217505	pBR322/ROP	apramycin	*sfGFP*
G555	#216483	pUC	ampicillin	RBSs of *plsB*, *plsC*, *cdsA* and *pssA*

Recombineering oligos and library design

Mutagenic ssDNA oligos of 89-91 nucleotides were designed to anneal with at least 30 nucleotides at both ends to the target-DNA. The ssDNA oligos are listed in **Supplementary Table 1**. The oligos were modified with two phosphorothioate bonds at the 5′ end. The ssDNA oligos were synthesized and purified by desalting by Sigma-Aldrich (oligos BG31272 and BG31273) or synthesized and purified by HPLC by ELLA Biotech GmbH (Germany) (all other oligos). The oligos were diluted in Milli-Q water to a concentration of 100 µM and stored at -20 °C.

RBS variants were designed using the RBS Library Calculator in the "Optimize Expression Levels" mode (https://salislab.net/software/design_rbs_library_calculator) with the following input parameters: the host organism was *Escherichia coli*; target minimum and maximum

translation initiation rates were 1 and 1,000,000, respectively; the genomic RBS sequence was the mRNA sequence from the 5′ end until the start codon (1).

Plasmid recombineering with the MOSAIC protocol

E. coli cells harboring pORTMAGE-Ec1 were streaked from glycerol stocks on a kanamycin-supplemented LB agar plate and grown overnight at 37 °C. The day before the MOSAIC experiment, an individual colony was picked and grown overnight in LB medium supplemented with kanamycin in a shaking incubator at 37 °C and 180-250 rpm. The following day, the overnight culture was diluted 1:100 in LB supplemented with kanamycin in a 50-mL falcon tube and incubated at 37 °C and 180-250 rpm. At an OD_{600} of 0.2-0.3, expression of the pORTMAGE-Ec1 machinery was induced by supplementing the culture with m-toluic acid to a final concentration of 1 mM. Following induction, the cells were incubated for an additional 45 minutes before being placed on ice for 1 hour. To make the cells electrocompetent, the culture was pelleted by centrifugation at 3200 rcf and 4 °C for 10 minutes. Next, the supernatant was carefully decanted before the cells were resuspended in 1 mL of ice-cold Milli-Q water containing 10% glycerol (v/v) and transferred to a 1.5-mL Eppendorf tube. The cells were washed another two to three times. Following the last wash step, the cells were resuspended in 250 µL of ice-cold Milli-Q water per 10 mL of initial culture. Next, 40 µL of cell suspension, 1 ng of target plasmid and 1 µL of 100 µM ssDNA oligos were combined in a 1.5-mL Eppendorf tube. For multi-target MOSAIC reactions, the oligos of interest were pre-mixed at equimolar concentrations and added to the cells to a final concentration of 2.5 µM per oligo or degenerate set of oligos. Next, 40 µL of the cell-DNA mixture were transferred to a 1-mm gap electroporation cuvette and electroporated at 1.8 kV. Immediately after electroporation, 960 µL of pre-warmed LB were added to the cell suspension and the cells were allowed to recover for 1 hour at 37 °C and 180-250 rpm. Following recovery, single-target transformants were transferred to a 50-mL falcon tube, supplemented with 4 mL of LB containing the appropriate antibiotic and incubated overnight at 37 °C and 180-250 rpm. The next day, the plasmids were isolated from the cells. For multi-target MOSAIC transformations, the recovered cells were plated on large (15 cm diameter) selective agar plates and incubated at 37 °C overnight. The following day, the colonies were counted by hand, whereafter the colonies were scrapped off the plate for plasmid isolation. To quantify the number of DNA variants present in single colonies for multi-target MOSAIC, single colonies were picked and grown overnight in ampicillin-supplemented LB (100 µg/mL ampicillin) for plasmid isolation. All plasmids were eluted from the plasmid purification columns using Milli-Q water. The DNA purity and concentration were validated by spectrophotometry and fluorometry.

Genomic recombineering control experiment

When recombineering was performed on the *E. coli* genome, electroporation of the cells was followed by 1 hour of incubation in 1 mL of LB and, subsequently, 2 hours of incubation in 6 mL of kanamycin-supplemented medium, whereafter the cells were plated on LB agar plates containing kanamycin, 100 µM isopropyl-β-D-thiogalactopyranoside (IPTG) and 100 µg/mL 5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside (X-gal) (Thermo Fisher Scientific). After incubation overnight at 37 °C, the fraction of white colonies relative to the total number of colonies was counted and used as a measure for the genomic recombineering efficiency.

Nanopore sequencing and analysis

Samples for Nanopore sequencing were prepared by diluting the mix of target plasmid DNA and pORTMAGE-Ec1 DNA in Milli-Q water to a concentration of 30-40 ng/µL as quantified by Qubit. Nanopore sequencing was performed by Plasmidsaurus (Oregon, US). To extract the sequencing reads that map to the target plasmid, the raw reads were filtered based on size (the plasmid size plus and minus 100 bp) and mapped to the wild-type DNA sequence of the target plasmid using the *Filter FASTQ reads by quality score and length* (26) and *Map with minimap2* (27) tools, respectively (accessed in the Galaxy web platform (https://usegalaxy.org)) (28). To minimize the numbers of insertions/deletions rather than mismatches during mapping, the minimap2 alignment parameters gap open penalty for deletions and insertions were increased from 4 (default) to 16 and from 24 (default) to 48, respectively, for the MOSAIC experiments incorporating 18 nt-insertions and deletions in pUC19 and diversifying the four RBSs in plasmid G555. For the 18-nt mismatch samples, these were increased to 32 and 72, respectively. For the 18-nt insertion samples, the reads were mapped to the designed modified DNA sequence instead of the wild-type sequence. If the reads from multiple Nanopore sequencing runs of the same sample were used for analysis, the FASTQ datasets were first merged using *Concatenate datasets tail-to-head* tool in the Galaxy web platform.

In-house developed R scripts were run in Rstudio (Version 1.1.456) to determine the editing efficiency (available at https://dx.doi.org/10.4121/4464ab86-9214-49b3-a808-10ca655385a6). The sequential steps in the analysis pipeline are illustrated in **Supplementary Figure 1**. In short, the DNA sequences from the target loci were extracted from the mapped reads and filtered based on the per-base quality scores recorded in the FASTQ files; the target sequences that contained at least one base with a score lower than 50 were excluded from the analysis. Then, the target sequences were identified as the wild type or as successfully mutated based on 100% similarity. The fraction of mutated sequences relative to the total number of target sequences was used to determine the editing efficiency. If the plasmid was modified in multiple loci, the number of mutated target loci was also counted per plasmid. To that end, an additional filtering step was applied to remove all reads that did not span the full sequence from the first to the last target region.

Statistics

The editing efficiencies described in the main text were averaged over at least three biological replicates. The mean and standard deviation are given where indicated.

RESULTS AND DISCUSSION

CspRecT-mediated recombineering achieves ~85% single-locus plasmid editing efficiency

We investigated the efficiency of plasmid recombineering using SSAP CspRecT expressed from plasmid pORTMAGE-Ec1 (**Figure 1A**). The plasmid recombineering efficiency was first tested with the high-copy number plasmid pUC19. Two types of ssDNA oligos were designed to introduce a single-nucleotide deletion or a three-nucleotide mismatch in the *lacZ* gene on pUC19. Because this vector replicates unidirectionally in a DNA sequence-controlled manner (29–31), we designed the oligo sequences such that they target the lagging strand during plasmid replication, as this is believed to lead to the highest efficiency during recombineering (14, 20, 32). The ssDNA oligos (2.5 μM) were co-electroporated with 1 ng of pUC19 plasmid into electrocompetent *E. coli* MG1655 expressing CspRecT and the dominant negative *E. coli* MutL mutant (EcMutL^{E32K}) for temporal repression of mismatch repair. Usually, deleterious mutations in *lacZ* can be quantified using blue-white screening on an LB agar plate with X-gal. However, as plasmid recombineering leads to mixed plasmid populations in single colonies, we determined the editing efficiencies by DNA sequencing. Hence, after overnight growth, the plasmids were isolated, and the editing efficiency was quantified from raw Nanopore sequencing reads. The Nanopore sequencing analysis pipeline is shown in **Supplementary Figure 1**. Only reads with a quality threshold ≥50 in the target region were used to reduce the incorrect detection of mutations in the population to < 1% (**Supplementary Figure 2**).

Remarkably, the editing efficiency after a single round of MOSAIC was 88% and 83% for the single-nucleotide deletion and three-nucleotide mismatch, respectively (**Figure 1B**). As expected, the use of complementary oligos that bind the leading strand in the replication fork resulted in lower editing efficiencies (74% and 54% for the single-nucleotide deletion and three-nucleotide mismatch, respectively) (**Supplementary Figure 3A**). The addition of both the leading and lagging strand-targeting oligos did not further improve the editing efficiency. The observed editing efficiencies were more than double those previously reported for plasmid editing with the Redβ SSAP (20). This coincides with the previously observed two-fold increase in genomic editing efficiency with CspRecT versus Redβ (24), highlighting the large impact of SSAP on the editing efficiency for both genomic and plasmid recombineering.

Importantly, when the oligos were electroporated into bacteria already harboring the target plasmid (i.e., "2-step electroporation"), the fraction of edited plasmids was very low (<7%) (**Figure 1B**). So, co-electroporation of the target plasmid and ssDNA is key to reach high plasmid editing efficiencies. This is in agreement with a previous study using Redβ for plasmid recombineering (20). The large effect of co-electroporation is likely explained by the fact that recombineering is most effective during plasmid replication. When a single plasmid or a low number of plasmids enters the cell, the plasmid(s) may rapidly be replicated many times to reach the copy number at which the plasmid is maintained in the cells.

Overall, plasmid recombineering resulted in much higher efficiencies than recombineering on the *E. coli* genome, whose highest reported efficiency is ~50% but in our hands reached only 14% based on a blue-white screening (**Supplementary Figure 3B**) (20). We also tested the MOSAIC protocol in *E. coli* DH5α, which is routinely used for transformation and cloning purposes. However, the observed plasmid editing efficiency in this strain was lower than in *E.*

coli MG1655, and hence the latter strain was used for the remainder of this study (**Supplementary Figure 3C**).

Higher-copy plasmids are edited more efficiently than low-copy plasmids

As we hypothesized that the high editing efficiency was coupled to a high plasmid replication rate, we anticipated that higher-copy plasmids with comparatively higher replication rates after electroporation would be edited more efficiently than lower-copy plasmids. To test this, we applied MOSAIC to a series of pSEVAb vectors that differed only in their origins of replication and, consequently, the copy number at which they are maintained in *E. coli* (25). The selected origins of replications were RK2 (low-copy number), pBBR1, p15A, and pBR322/ROP (medium-copy number), and pRO1600/ColE1 and pUC (high-copy number) (33, 34) (**Supplementary Figure 4)**. We designed oligos to incorporate a deletion or a three-nucleotide mismatch in the gene encoding *sfGFP* present in all plasmids. We identified the plasmid leading and lagging strands based on the known class B theta replication mechanism of ColE1 and ColE1-like origins (pRO1600/ColE1, pUC, pBR322 and p15A) (29–31, 35, 36), and similarly for the class A theta replication mechanism of the RK2-plasmid origin (31, 37, 38). Based on this, we tested the oligos targeting the lagging strand assuming this would lead to the highest recombineering efficiency (**Figure 1C**). To the best of our knowledge, the precise replication mechanism of pBBR1-derived plasmids is still unknown. Therefore, we tested both the forward and reverse oligos for this plasmid, which performed equally well (**Supplementary Figure 3A**). As hypothesized, the vector with the low-copy RK2 origin of replication was edited with the lowest efficiency, on average 30%, followed by 58% for the pBBR1 origin of replication (**Figure 1C**). Surprisingly, the four other plasmids that we evaluated were all modified with 70-90% efficiency, with pBR322/ROP being most efficient with 89% and 86% for a single-nucleotide deletion and a three-nucleotide mismatch, respectively. As such, it appears that beyond a certain copy number, additional replication events no longer increase the efficiency at which a single oligo is incorporated. Such a threshold might be due to a saturation effect, possibly caused by exceeding a time window during which the recombineering machinery and/or oligos are sufficiently active, and may represent an upper boundary for MOSAIC.

Large insertions and deletions are incorporated with high efficiency

To probe the potential broad applicability of MOSAIC, we investigated if a high editing efficiency could still be obtained with a substantially larger number of mutations per oligo. If so, larger regions, such as regulatory sequences (e.g., promoters, RBSs and operator sites), could readily be inserted, deleted, replaced or diversified. Hence, we designed three ssDNA oligos that incorporate an 18-nt wide insertion, deletion or mismatch into the *lacZ* gene on pUC19. The efficiency for the insertion and deletion was 65-70%, whilst the efficiency for substituting 18 nucleotides was lower (30%). Nevertheless, these efficiencies are 5-fold higher than the genomic recombineering efficiency for an 18-nt mismatch and sufficient for many of the aforementioned applications (**Figure 1D**) (24). Altogether, these results demonstrate that MOSAIC is a powerful method to edit and diversify both small and larger regions of plasmid DNA.

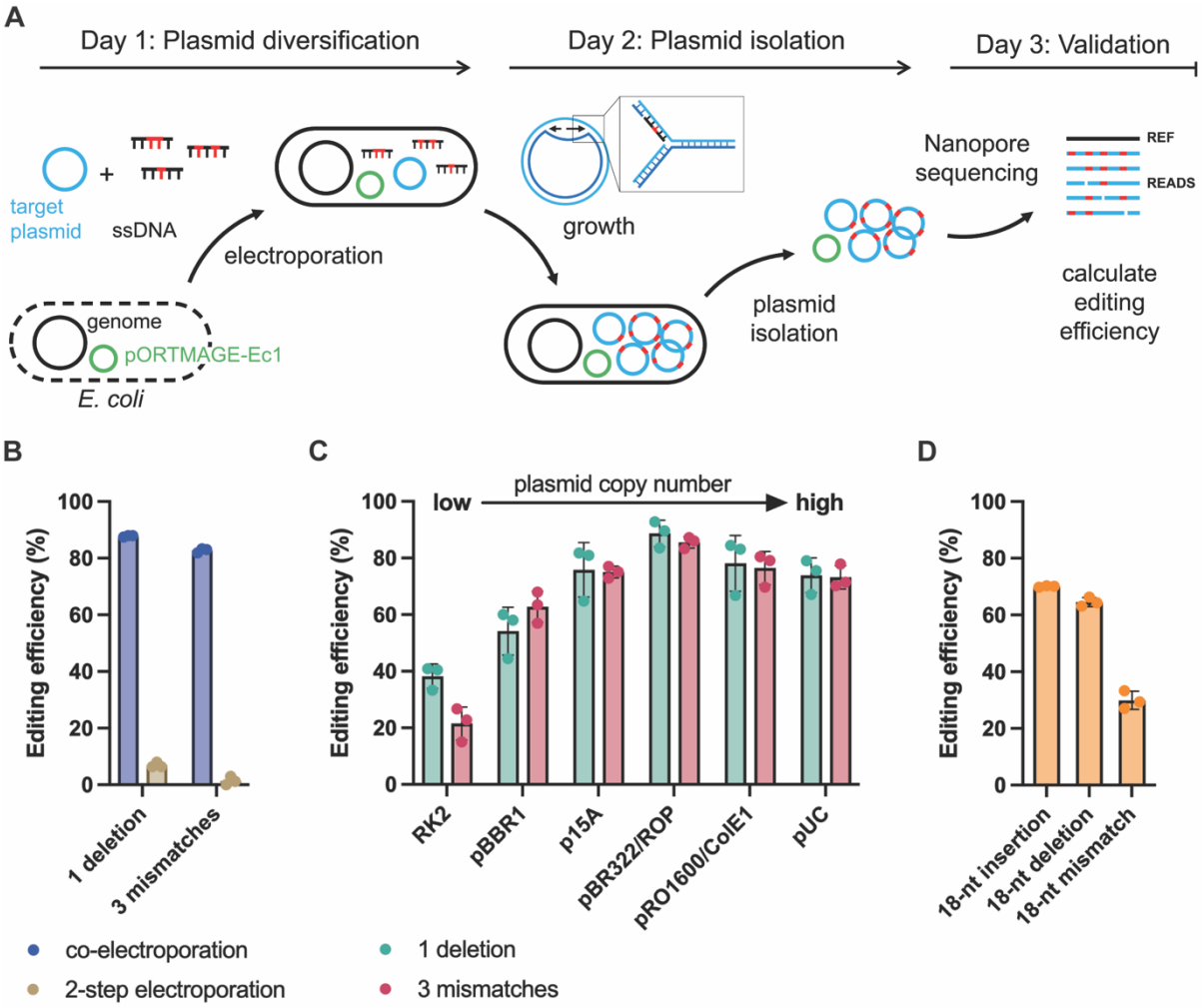


Figure 1. Plasmid editing efficiency of MOSAIC. A) Schematic representation of MOSAIC's protocol for plasmid editing using ssDNA recombineering. The expression of SSAP CspRecT and the dominant negative mutant EcMutL^{E32K} is induced in *E. coli* cells harboring pORTMAGE-Ec1. The cells are made electrocompetent and transformed with the target plasmid and mutagenic ssDNA oligos. During plasmid replication, the oligos anneal to one of the two DNA strands in the replication fork with the help of CspRecT, introducing mutations into the plasmid sequence. Plasmids are then isolated from the cells, and the editing efficiency is calculated from raw Nanopore sequencing reads. **B)** The effect of co- or 2-step electroporation on the incorporation of a single-nucleotide deletion or a three-nucleotide mismatch in the *lacZ* gene on the high-copy number plasmid pUC19. **C)** Plasmid editing efficiency for the incorporation of a single-nucleotide deletion or a three-nucleotide mismatch in the gene encoding sfGFP on the pSEVAb plasmids with different origins of replication. **D)** DNA editing efficiency for the incorporation of an 18-nt long insertion, deletion or mismatch in the *lacZ* gene on the high-copy number plasmid pUC19.

One round of MOSAIC yields a large multiplex plasmid library

Next, we investigated if we could apply MOSAIC to create a combinatorial plasmid library of 10^4 variants in a single electroporation step. As a proof-of-principle we diversified the ribosome binding site (RBS) of four genes encoding a phospholipid synthesis pathway on the

pUC19-derived plasmid G555 (**Figure 2A**) (39). To modulate translation, RBS calculator was used to design RBS variants with a wide range of predicted translation initial rates (1). The resulting RBS variants contained up to 7 mismatches relative to the wild-type DNA and yielded a total library of 13 x 12 x 9 x 9 = 12,636 DNA variants (**Figure 2B**). Because two of the four target sites were identical, three degenerate oligo libraries targeting four different sites on the plasmid were sufficient to produce the combinatorial library. In a single electroporation reaction, a mix of the three degenerate oligo libraries and the plasmid G555 were transformed into *E. coli* MG1655 cells expressing CspRecT and EcMutL^{E32K}. Following incubation overnight, the plasmids were isolated and sequenced.

Nanopore sequencing is specifically useful for combinatorial libraries where the different target loci on the plasmid are thousands of base pairs apart from each other and need to be linked to identify the full genotype of the plasmid variant. Using the analysis pipeline outlined previously (**Supplementary Figure 1**), the different RBS variants in the Nanopore reads were counted and compared to those designed by RBS calculator. Of the sequenced library plasmids, 21% had all four RBS sequences mutated (**Figure 2C**). The editing efficiency averaged over all four target regions was 51 ± 4% (mean ± standard deviation, n=4) (**Supplementary Figure 5A**). Moreover, the sequencing data suggest an unbiased RBS diversification, as all designed RBS variants were represented in the generated library (**Supplementary Figure 5B**). As only a small fraction of the total library was sequenced, the library coverage could not be exactly determined from the sequencing data. Still, adequate library coverage is plausible as >90% of the sequenced plasmids with at least one modified locus corresponded to unique plasmid variants (n=4 independent transformations).

To gather a large, if not complete, fraction of the designed library variants, it is important that the plasmids are isolated from a substantial number of colonies. After a single transformation, we obtained $5 \pm 3 \times 10^3$ colonies (mean ± standard deviation, n=7 biological replicates). In contrast to assembly-based cloning methods, where cells typically harbor only a single plasmid variant, colonies may contain multiple DNA variants in MOSAIC, as mutations are incorporated in the plasmids after uptake by the cells. To estimate the number of DNA variants per colony, we sequenced the plasmids isolated from single colonies. The number of DNA variants per colony was 6 ± 4 (n=6 colonies) (**Supplementary Table 2**). With these numbers, we estimated that per electroporation, we could generate generated (5×10^3 colonies) × (6 DNA variants per colony) = 3×10^4 DNA variants. As approximately 21% of the plasmid variants had all target sites modified, we estimated that 6×10^3 of the 12,636 designed DNA variants could have been generated, which corresponds to a library coverage of 50%. While acknowledging that duplicate variants could introduce some variability in the estimated library coverage, these results demonstrate that MOSAIC enables the generation and validation of large, diversified plasmid libraries in a single transformation.

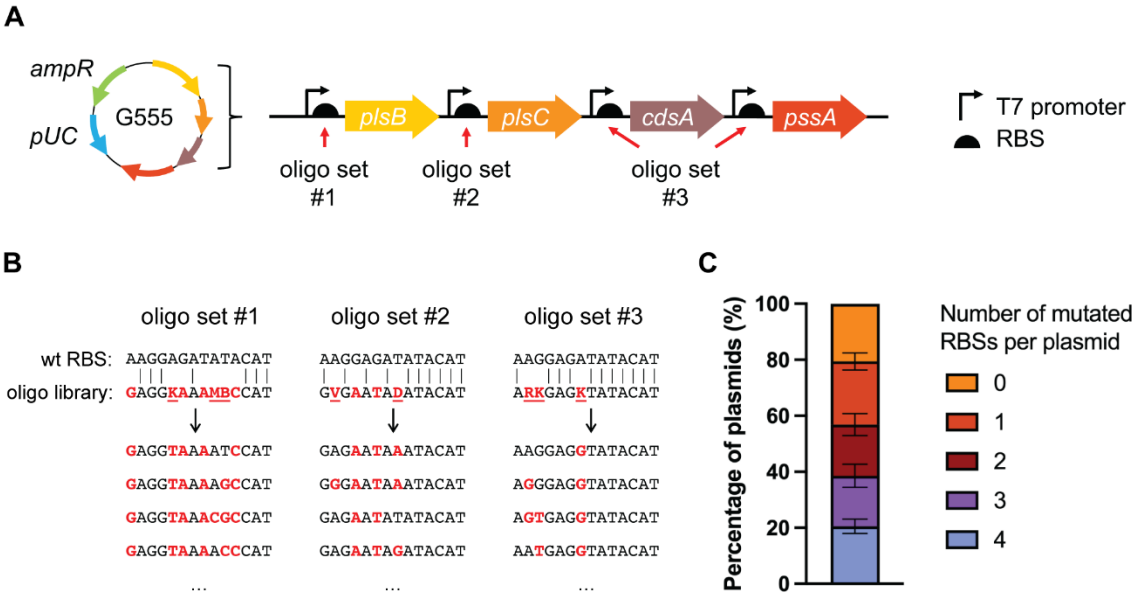


Figure 2. Construction of a multiplex plasmid library with four diversified RBSs using MOSAIC. A) pUC-derived target plasmid G555 containing four genes from the Kennedy phospholipid biosynthesis pathway (*plsB*, *plsC*, *cdsA* and *pssA*). Three ssDNA oligo sets of degenerate sequences (9-13 variants per set) targeted four regions on plasmid G555. **B)** The three sets of ssDNA oligos were designed using RBS calculator (1) to target the RBSs of four genes. Letters in red indicate changes relative to the wild-type DNA. Underlined letters indicate degenerate nucleotides. **C)** Number of mutated RBSs per plasmid (n=4). The four data sets contained reads that spanned over all target regions and consisted of 50, 35, 29 and 20 Nanopore reads.

MOSAIC is a versatile tool for plasmid editing and diversification with applications *in vivo* and in cell-free systems

In contrast to state-of-the-art plasmid editing and diversification methods, MOSAIC does not require PCR or plasmid assembly from fragments. The only requirements for MOSAIC are fast-to-order mutagenic oligos and a publicly available *E. coli* strain harboring pORTMAGE-Ec1. A simple co-electroporation of the target plasmid and oligos is sufficient to perform the desired mutagenesis. This makes the protocol especially powerful to edit plasmids that are hard to clone due to high GC-content regions, repetitive sequences, or large sizes. Single-variant plasmids can be easily isolated by re-transformation, while plasmid libraries can be used directly.

The simplicity of the protocol lends itself to a plethora of *in vivo* and cell-free synthetic biology applications ranging from protein engineering to the optimization of natural or synthetic metabolic pathways (13, 40–42). More specifically, MOSAIC enables the rapid prototyping of pathway or enzyme variants when coupled to high-throughput phenotypic screening or growth-coupled selection approaches to isolate well-performing variants. These isolated variants could be rapidly subjected to further rounds of diversification using MOSAIC during subsequent design-build-test-learn cycles.

The presence of full and partial wild-type plasmid variants alongside the mutant DNA variants in combinatorial MOSAIC libraries is currently unavoidable. As such, increasing the number of target sites or the variability at each site may hinder library coverage. The fraction of fully mutated library variants could be improved by employing iterative rounds of MOSAIC or CRISPR-Cas-based counterselection (21). Another promising strategy is to restrain the library size by computational design. Excitingly, recent efforts to preselect library variants from a larger pool using machine learning prior to wet lab characterization showed promising results in generating small but smart libraries to accelerate the evolutionary optimization (43). All in all, we anticipate that MOSAIC, combined with the continuing advances in computationally aided design of genetic and protein libraries, will enable the rapid exploration of biological solution spaces throughout many labs and research projects.

DATA AVAILABILITY

The data underlying this article are available in the 4TU.ResearchData repository at https://dx.doi.org/10.4121/4464ab86-9214-49b3-a808-10ca655385a6.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

We thank Ana Restrepo Sierra for her contribution to the conceptualization and design of the RBS library. We also thank Suzan Yilmaz for assistance in setting up the recombineering experiments. We are thankful to Akos Nyerges for providing advice on recombineering and the use of pORTMAGE. Lastly, we thank Haroun Bensaadi for performing preliminary recombineering experiments. MvdB, CD and NJC acknowledge financial support from The Netherlands Organization of Scientific Research (NWO/OCW) Gravitation program Building A Synthetic Cell (BaSyC) (024.003.019). TYA and NJC acknowledge funding from an NWO-XL grant (OCENW.XL21.XL21.007). CD acknowledges funding from ANR pour CPJ (ANR-2023-004). NJC also acknowledges support from an NWO-Veni Fellowship (VI.Veni.192.156).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Salis,H.M., Mirsky,E.A. and Voigt,C.A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol*, **27**, 946–950.

2. LaFleur,T.L., Hossain,A. and Salis,H.M. (2022) Automated model-predictive design of synthetic promoters to control transcriptional profiles in bacteria. *Nat Commun*, **13**, 5159.

3. Leman,J.K., Weitzner,B.D., Lewis,S.M., Adolf-Bryfogle,J., Alam,N., Alford,R.F., Aprahamian,M., Baker,D., Barlow,K.A., Barth,P., *et al.* (2020) Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat Methods*, **17**, 665–680.

4. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Žídek,A., Potapenko,A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.

5. Baek,M., DiMaio,F., Anishchenko,I., Dauparas,J., Ovchinnikov,S., Lee,G.R., Wang,J., Cong,Q., Kinch,L.N., Schaeffer,R.D., *et al.* (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, **373**, 871–876.

6. Yilmaz,S., Nyerges,A., van der Oost,J., Church,G.M. and Claassens,N.J. (2022) Towards next-generation cell factories by rational genome-scale engineering. *Nat Catal*, **5**, 751–765.

7. Simon,A.J., d'Oelsnitz,S. and Ellington,A.D. (2019) Synthetic evolution. *Nat Biotechnol*, **37**, 730–743.

8. Currin,A., Parker,S., Robinson,C.J., Takano,E., Scrutton,N.S. and Breitling,R. (2021) The evolving art of creating genetic diversity: From directed evolution to synthetic biology. *Biotechnology Advances*, **50**, 107762.

9. Alejaldre,L., Pelletier,J.N. and Quaglia,D. (2021) Methods for enzyme library creation: Which one will you choose? *BioEssays*, **43**, 2100052.

10. Engler,C., Kandzia,R. and Marillonnet,S. (2008) A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLOS ONE*, **3**, e3647.

11. Gibson,D.G., Young,L., Chuang,R.-Y., Venter,J.C., Hutchison,C.A. and Smith,H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*, **6**, 343–345.

12. Kok,S. de, Stanton,L.H., Slaby,T., Durot,M., Holmes,V.F., Patel,K.G., Platt,D., Shapland,E.B., Serber,Z., Dean,J., *et al.* (2014) Rapid and reliable DNA assembly via ligase cycling reaction. *ACS Synth. Biol.*, **3**, 97–106.

13. Naseri,G. and Koffas,M.A.G. (2020) Application of combinatorial optimization strategies in synthetic biology. *Nat Commun*, **11**, 2446.

14. Wannier,T.M., Ciaccia,P.N., Ellington,A.D., Filsinger,G.T., Isaacs,F.J., Javanmardi,K., Jones,M.A., Kunjapur,A.M., Nyerges,A., Pal,C., *et al.* (2021) Recombineering and MAGE. *Nat Rev Methods Primers*, **1**, 1–24.

15. Datsenko,K.A. and Wanner,B.L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences*, **97**, 6640–6645.

16. Wang,X., Zheng,W., Zhou,H., Tu,Q., Tang,Y.-J., Stewart,A.F., Zhang,Y. and Bian,X. (2022) Improved dsDNA recombineering enables versatile multiplex genome engineering of kilobase-scale sequences in diverse bacteria. *Nucleic Acids Research*, **50**, e15–e15.

17. Wang,H.H., Isaacs,F.J., Carr,P.A., Sun,Z.Z., Xu,G., Forest,C.R. and Church,G.M. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, **460**, 894–898.

18. Swaminathan,S., Ellis,H.M., Waters,L.S., Yu,D., Lee,E.C., Court,D.L. and Sharan,S.K. (2001) Rapid engineering of bacterial artificial chromosomes using oligonucleotides. *Genesis*, **29**, 14–21.

19. Thomason,L.C., Costantino,N., Shaw,D.V. and Court,D.L. (2007) Multicopy plasmid modification with phage λ Red recombineering. *Plasmid*, **58**, 148–158.

20. Li,Y., Gu,Q., Lin,Z., Wang,Z., Chen,T. and Zhao,X. (2013) Multiplex iterative plasmid engineering for combinatorial optimization of metabolic pathways and diversification of protein coding sequences. *ACS Synth. Biol.*, **2**, 651–661.

21. Geng,Y., Yan,H., Li,P., Ren,G., Guo,X., Yin,P., Zhang,L., Qian,Z., Zhao,Z. and Sun,Y.-C. (2019) A highly efficient in vivo plasmid editing tool based on CRISPR-Cas12a and phage λ Red recombineering. *Journal of Genetics and Genomics*, **46**, 455–458.

22. Zhang,J.J. and Moore,B.S. (2020) Site-directed mutagenesis of large biosynthetic gene clusters via oligonucleotide recombineering and CRISPR/Cas9 targeting. *ACS Synth. Biol.*, **9**, 1917–1922.

23. Hueso-Gil,A., Nyerges,Á., Pál,C., Calles,B. and de Lorenzo,V. (2020) Multiple-site diversification of regulatory sequences enables interspecies operability of genetic devices. *ACS Synth. Biol.*, **9**, 104–114.

24. Wannier,T.M., Nyerges,A., Kuchwara,H.M., Czikkely,M., Balogh,D., Filsinger,G.T., Borders,N.C., Gregg,C.J., Lajoie,M.J., Rios,X., *et al.* (2020) Improved bacterial recombineering by parallelized protein discovery. *Proc Natl Acad Sci U S A*, **117**, 13689–13698.

25. Damalas,S.G., Batianis,C., Martin-Pascual,M., de Lorenzo,V. and Martins dos Santos,V.A.P. (2020) SEVA 3.1: enabling interoperability of DNA assembly among the SEVA, BioBricks and Type IIS restriction enzyme standards. *Microbial Biotechnology*, **13**, 1793–1806.

26. Blankenberg,D., Gordon,A., Von Kuster,G., Coraor,N., Taylor,J., Nekrutenko,A., and the Galaxy Team (2010) Manipulation of FASTQ data with Galaxy. *Bioinformatics*, **26**, 1783–1785.

27. Li,H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**, 3094–3100.

28. The Galaxy Community (2022) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Research*, **50**, W345–W351.

29. Tomizawa,J.-I., Sakakibara,Y. and Kakefuda,T. (1974) Replication of colicin E1 plasmid DNA in cell extracts. Origin and direction of replication. *Proc Natl Acad Sci U S A*, **71**, 2260–2264.

30. Selzer,G., Som,T., Itoh,T. and Tomizawa,J. (1983) The origin of replication of plasmid p15A and comparative studies on the nucleotide sequences around the origin of related plasmids. *Cell*, **32**, 119–129.

31. Lilly,J. and Camps,M. (2015) Mechanisms of theta plasmid replication. *Microbiology Spectrum*, **3**, PLAS-0029-2014.

32. Gallagher,R.R., Li,Z., Lewis,A.O. and Isaacs,F.J. (2014) Rapid editing and evolution of bacterial genomes using libraries of synthetic DNA. *Nat Protoc*, **9**, 2301–2316.

33. Jahn,M., Vorpahl,C., Hübschmann,T., Harms,H. and Müller,S. (2016) Copy number variability of expression plasmids determined by cell sorting and Droplet Digital PCR. *Microbial Cell Factories*, **15**, 211.

34. Silva-Rocha,R., Martínez-García,E., Calles,B., Chavarría,M., Arce-Rodríguez,A., de Las Heras,A., Páez-Espino,A.D., Durante-Rodríguez,G., Kim,J., Nikel,P.I., *et al.* (2013) The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res*, **41**, D666-675.

35. Kuzminov,A., Schabtach,E. and Stahl,F.W. (1997) Study of plasmid replication in Escherichia coli with a combination of 2D gel electrophoresis and electron microscopy11Edited by J. Karn. *Journal of Molecular Biology*, **268**, 1–7.

36. Doran,K.S., Konieczny,I. and Helinski,D.R. (1998) Replication origin of the broad host range plasmid RK2: positioning of various motifs is critical for initiation of replication. *Journal of Biological Chemistry*, **273**, 8447–8453.

37. Meyer,R.J. and Helinski,D.R. (1977) Unidirectional replication of the P-group plasmid RK2. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, **478**, 109–113.

38. Stalker,D.M., Thomas,C.M. and Helinski,D.R. (1981) Nucleotide sequence of the region of the origin of replication of the broad host range plasmid RK2. *Mol Gen Genet*, **181**, 8–12.

39. Blanken,D., Foschepoth,D., Serrão,A.C. and Danelon,C. (2020) Genetically controlled membrane synthesis in liposomes. *Nat Commun*, **11**, 4317.

40. Gargiulo,S. and Soumillion,P. (2021) Directed evolution for enzyme development in biocatalysis. *Current Opinion in Chemical Biology*, **61**, 107–113.

41. Li,Z., Deng,Y. and Yang,G.-Y. (2023) Growth-coupled high throughput selection for directed enzyme evolution. *Biotechnology Advances*, **68**, 108238.

42. Orsi,E., Claassens,N.J., Nikel,P.I. and Lindner,S.N. (2021) Growth-coupled selection of synthetic modules to accelerate cell factory development. *Nat Commun*, **12**, 5295.

43. Marchal,D.G., Schulz,L., Schuster,I., Ivanovska,J., Paczia,N., Prinz,S., Zarzycki,J. and Erb,T.J. (2023) Machine Learning-Supported Enzyme Engineering toward Improved CO2-Fixation of Glycolyl-CoA Carboxylase. *ACS Synth. Biol.*, **12**, 3521–3530.