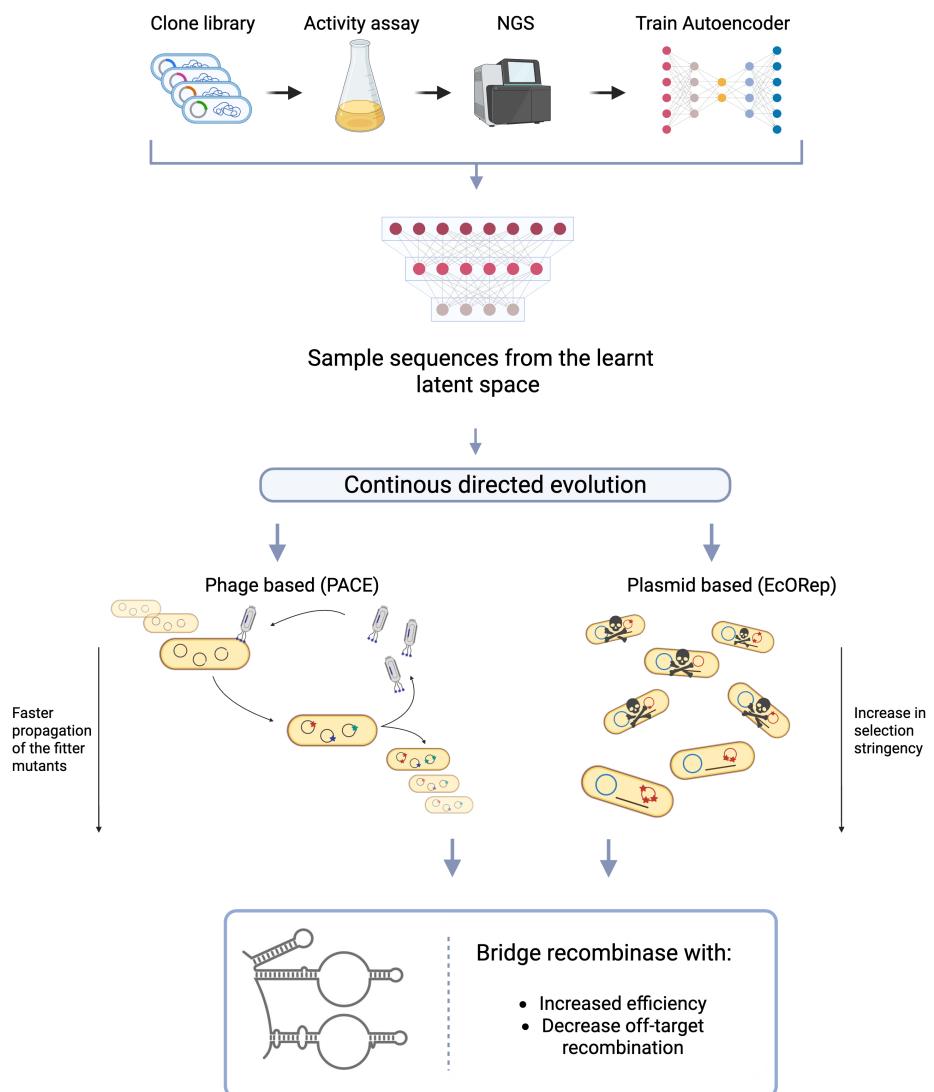


# Directed Evolution Platform for High-Efficiency, Large-Scale Genome Editing using Bridge Recombinases

October 1<sup>st</sup> 2025

Daria Belous<sup>1,\*</sup>, David Binder<sup>1,\*</sup>, Sofia Filippini<sup>1,\*</sup>, Saskia Gerecke<sup>2,\*</sup>, David Jewanski<sup>1,\*</sup>, Etienne Mathier<sup>1,\*</sup>, Philip Nitsch<sup>1,\*</sup>, Elia Weber<sup>1,\*</sup>, Annabelle Winzer<sup>1,\*</sup>, Lorenz Widmer<sup>1, 3, †</sup>, Kian Bigovic Villi<sup>1</sup>, Michael Bohl<sup>1</sup>, Michael Nash<sup>1, 3</sup>

e-mail: <sup>†</sup> [lorenz.widmer@unibas.ch](mailto:lorenz.widmer@unibas.ch)



<sup>1</sup> ETH Zurich (Swiss Federal Institute of Technology Zurich), Zurich, Switzerland

<sup>2</sup> University of Zurich, Zurich, Switzerland

<sup>3</sup> University of Basel, Basel, Switzerland

\* The authors contributed equally

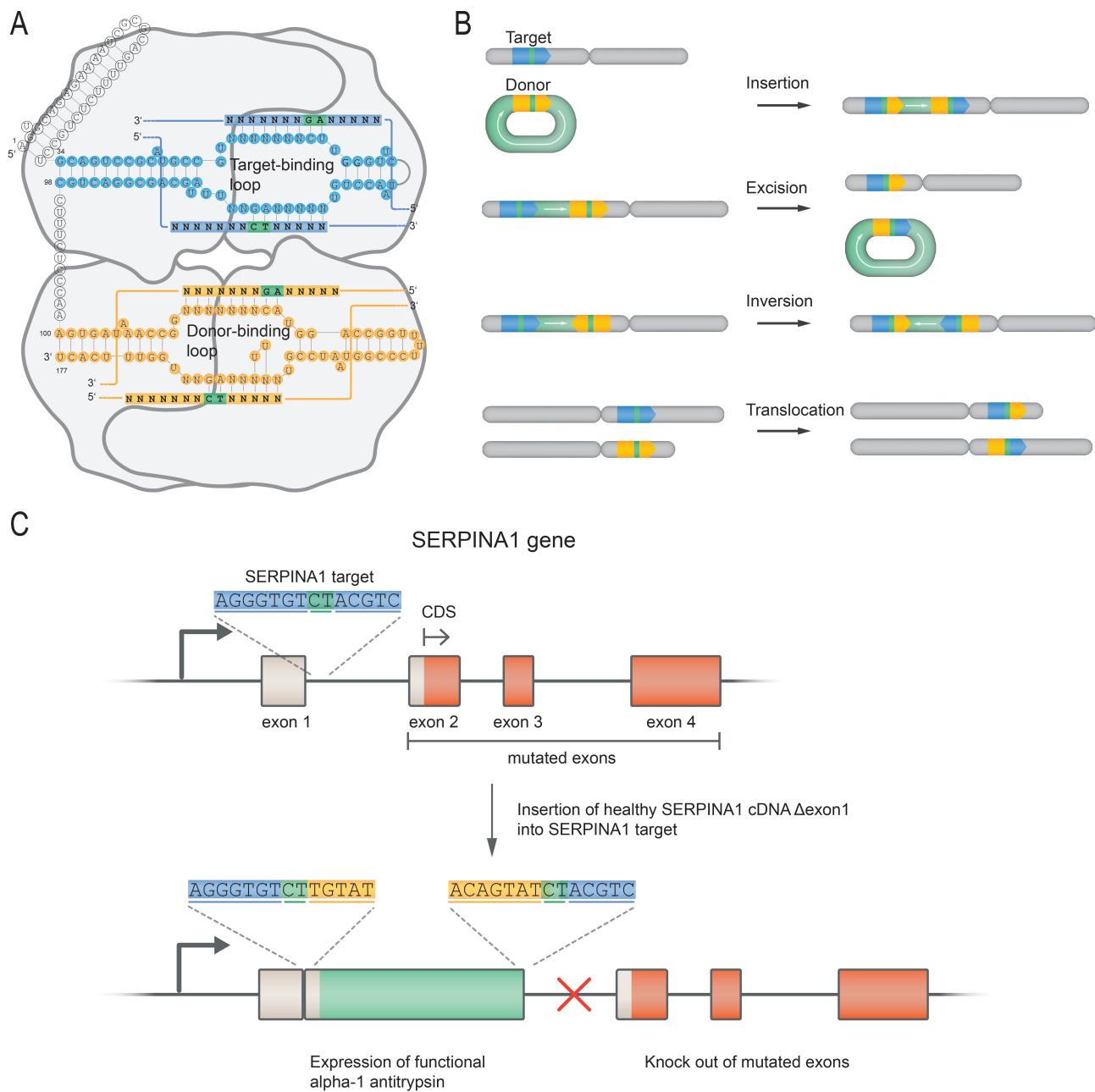
## 1 Abstract

Bridge recombinases are RNA-guided DNA recombinases that catalyse site-specific insertions, inversions, and excisions through a programmable bridge RNA (bRNA) which recognizes both donor and target DNA sequences. Their ability to manipulate large DNA fragments in a targeted manner makes them a powerful genome engineering tool. The applicability is currently limited by low catalytic activity and specificity. Here, we establish a workflow that enables broad computational sampling of the fitness landscape through DML, followed by two evolutionary logics compatible with continuous directed evolution platforms, namely EcORep and PACE, to efficiently converge on local fitness optima. As a proof of principle, we aim to evolve an efficient bridge recombinase for the treatment of  $\alpha$ -1 antitrypsin deficiency, enabling replacement of the defective SERPINA1 gene with a healthy copy. Although optimization is ongoing, our results provide evidence that our platform could be used to increase the efficiency of bridge recombinases.

## 2 Introduction

The ability to edit genomic DNA represents a major advance in molecular biology, with applications spanning basic research, medicine, agriculture, and biotechnology. Programmable nucleases [1], [2], base editors [3], [4], and prime editors [5], [6] can disrupt, install, or correct sequence under 200 bp and have entered the clinic as one-time treatments with more than 60 trials to date [7]. However, insertion of large, gene-sized DNA ( $\geq 1$  kb) in many cell types remains difficult. Current approaches rely on nuclease-induced double-stranded breaks (DSBs) and homology-directed repair [8], [9], which yield low insertion efficiency and perform poorly in post-mitotic cells [10], [11]. Moreover, DSBs can cause unintended insertions and deletions [12], [13], complex chromosomal rearrangements [14]–[16], or p53 activation [17]. Lastly, many genetic disorders arise from numerous distinct mutations, requiring variant-specific designs and making allele-specific therapies applicable only to subsets of patients. Thus, the field is in need of genomic tools capable of delivering kilobase-scale editing.

In 2024, *Durrant et al.* reported the discovery of the IS110 family of transposons in bacteria and archaea, which mobilize within host genomes via an RNA-guided ribonucleoprotein complex composed of a recombinase enzyme and a programmable, structured non-coding RNA termed bridge RNA (bRNA) [18]. The bRNA contains two internal loops that mediate sequence-specific recombination between donor and target DNA through RNA–DNA base pairing. The target-binding loop (TBL) encodes nucleotides that base pair with the genomic target site, whereas the donor-binding loop (DBL) encodes nucleotides that base pair with the donor sequence [18]–[20] (Figure 1A). Depending on donor–target orientation and location, recombination can result in insertions, inversions, or excisions (Figure 1B). More recently, the ISCro4 bridge recombinase ortholog has been shown to mediate genome editing in human cells, albeit with a maximum of 20% insertion efficiency and approximately 80% specificity [21].



**Figure 1: Large gene insertion with programmable bRNA.** **A** Schematic of the ribonucleoprotein complex (hnRNP + bRNA) bound to target and donor DNA. Positions annotated “N” in the target- and donor-binding loops are reprogrammable to specify any target–donor pair. [21] **B** Bridge recombinases can catalyze genomic insertions, inversions, excisions, and translocations depending on the orientation and placement of target and donor sequences in *E. coli* and mammalian cells. [18], [21] **C** Therapeutic concept: a compact, programmable, DSB-free bridge recombinase restores  $\alpha$ 1-antitrypsin by inserting a healthy cDNA copy of SERPINA1 at its endogenous site.

The bridge recombinase technology has the potential to become a foundation for next-generation gene therapies due to its DSB-independent mechanism, compact size, and inherent programmability. However, the previously mentioned limited activity and specificity remain challenges for therapeutic application. As a proof of concept, we focus on  $\alpha$ -1 antitrypsin deficiency (AATD), a polyallelic monogenic disorder caused by many different mutations in the SERPINA1 gene. AATD leads to insufficient levels of functional  $\alpha$ 1-antitrypsin, resulting in

progressive lung damage, while accumulation of misfolded protein in hepatocytes causes liver disease [22]. Current therapies are limited to symptomatic management or, in sever cases, liver transplantation [23]. To provide a curative therapy applicable to all patients, a strategy that restores functional  $\alpha$ 1-antitrypsin independent of the causative mutation is needed. Hepatocytes are the primary clinical target as these are the main  $\alpha$ 1-antitrypsin producing cells regardless of the disease mechanism [23]. Moreover, they are readily targetable by lipid nanoparticles [24], a leading therapeutic delivery platform. Together, these features make AATD an ideal model for demonstrating the therapeutic potential of bridge recombinase-mediated gene insertion.

Here, we sought to develop an evolutionary framework to enhance the efficiency of bridge recombinases. Our approach combines computational sampling of the fitness landscape through deep mutational learning (DML) with two evolution logics that are compatible with continuous evolution strategies and enable rapid convergence toward local optima: *E. coli* orthogonal replicon (EcORep) [25] and phage-assisted continuous evolution (PACE) [26].

### 3 Results

#### 3.1 Sample functional bridge recombinase variants using DML

To efficiently explore the sequence space of functional bridge recombinases, we screened an unbiased and high-distance mutational library in an inversion-based efficiency assay (Supplementary Text 6.1 Figure 11). The resulting NGS dataset will serve to train an autoencoder model aimed at learning a meaningful latent space representation of the fitness landscape. Sampling from this latent space will generate functional sequences that broadly cover the fitness landscape. These sequences provide diverse starting points that can subsequently be refined toward local optima using the described directed evolution logics. A more detailed description of this sub-project can be found in the supplementary section.

#### 3.2 Establishing a plasmid-based selection system for evolution of of bridge recombinases in *E. coli*

##### 3.2.1 Evolutionary strategy and plasmid design

We designed an evolution logic that allows for the selection of improved bridge recombinases in *E. coli* (Figure 2). All components necessary for selection are contained on a single "selection plasmid" while the bridge recombinase is located on a different plasmid or orthogonal replicon. This system is adaptable to several continuous DE methods such as MutaT7 [27], T7-ORACLE[25] and *E. coli* orthogonal replication system (EcORep) [28]. Selection of variants with higher activity occurs as follows: the selection plasmid carries two bRNAs under the control of inducible promoters (Figure 2 A) and a cassette encoding two antibiotic resistance genes facing in opposite directions. This cassette can be inverted via bRNA-mediated recombination. Only one of the resistance genes is expressed in each of the orientations of the cassette. Therefore, it is possible to select for successful inversion of the cassette using antibiotics (2). This allows for the enrichment of recombinase variants with increased activity over time. Expression of bRNA A can be induced by the addition of anhydrotetracycline (aTc) and results in inversion of the cassette from its initial orientation (gentamycin (Gm) → kanamycin (Kan)). Expression of bRNA B is induced using N-(3-Hydroxytetradecanoyl)-DL-homoserine lactone (OHC14) and mediates the reverse process (Kan → Gm).

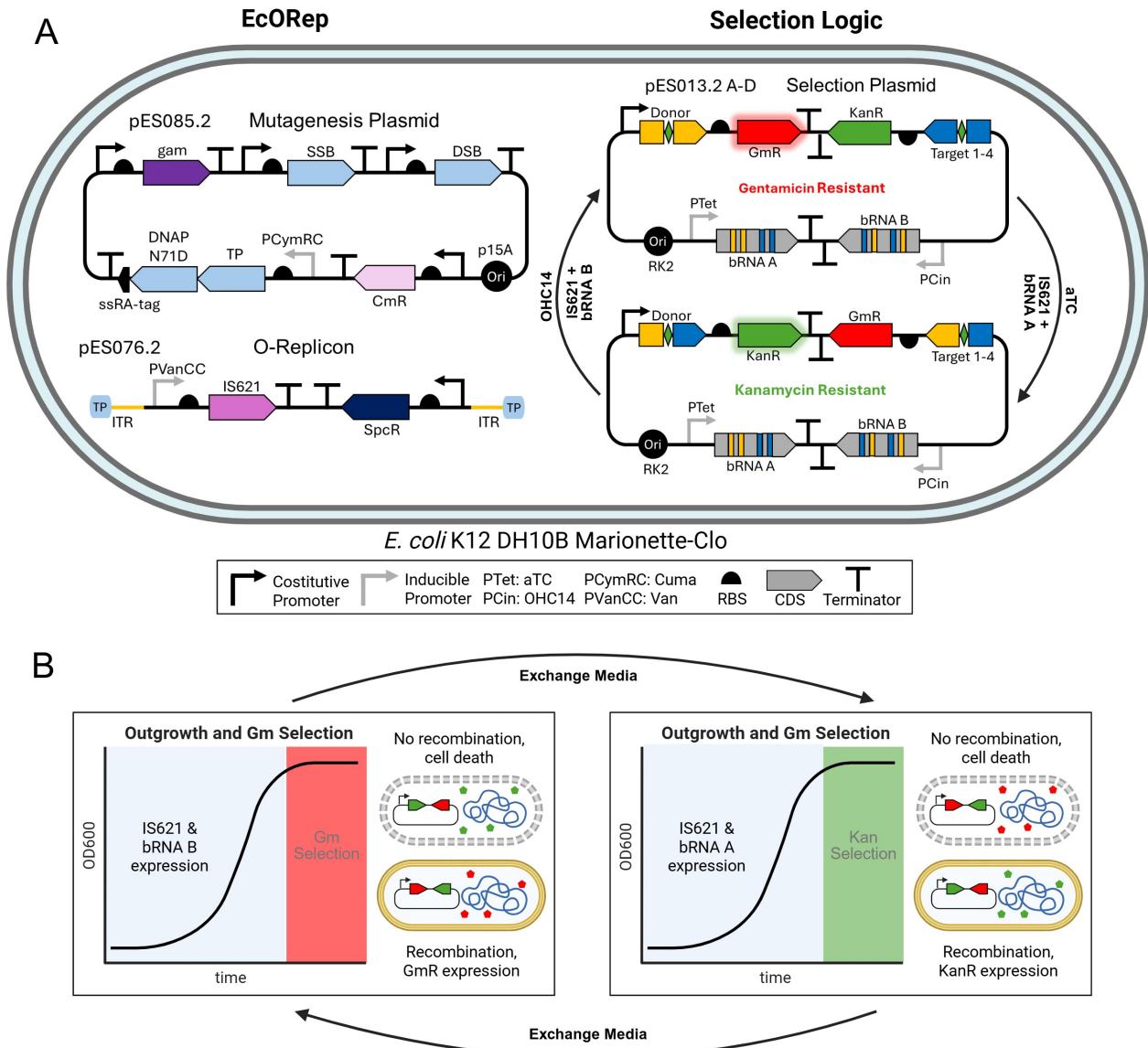
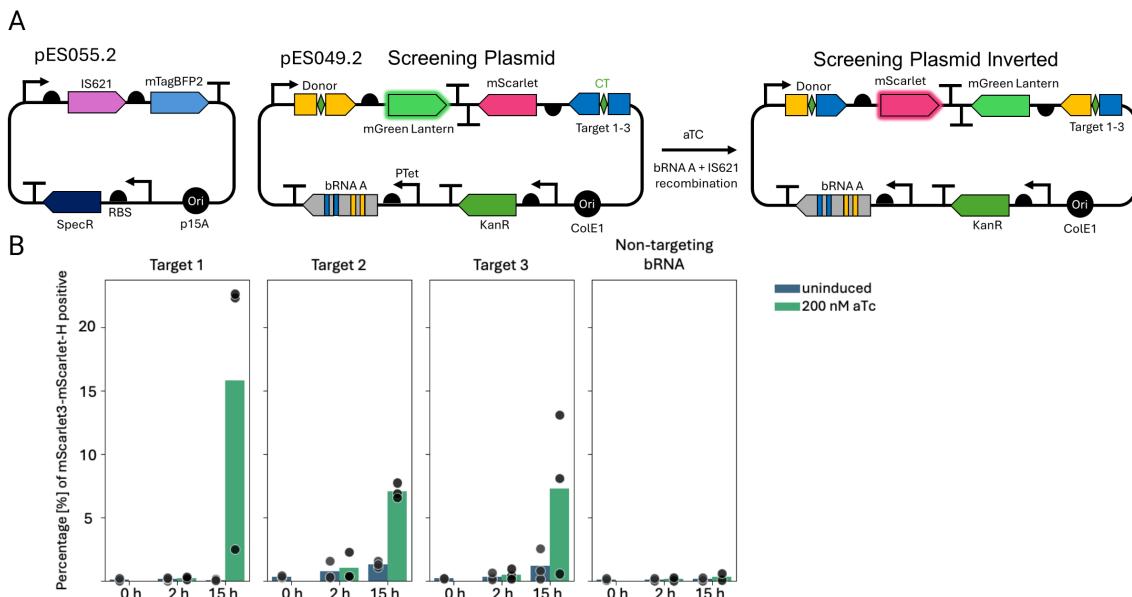


Figure 2: *DE of IS621 using EcORep*. **A** depicts the mutagenesis plasmid, the O-Replicon and the selection plasmid in an *E. coli* cell. The mutagenesis plasmid carries the genes necessary for replicating the O-Replicon. The O-Replicon carries IS621 under the control of the inducible PVanCC promoter. The selection plasmid contains a cassette with two antibiotic resistances, Gm resistance (*GmR*) and Kan resistance (*KanR*) facing in opposite directions. Depending on the orientation of the cassette, either *GmR* or *KanR* is transcribed. The cassette is flanked by donor and target sequences facing each other. This allows for IS621 + bRNA mediated inversion. **B** depicts the selection process. Cells are grown in LB liquid culture. IS621 expression is induced using vanillic acid (Van). Shortly before stationary phase, bRNA expression is induced by addition of aTc for bRNA A or OHC14 for bRNA B. Presence of IS621 + bRNA promotes recombination. Selection is performed by regrowing cells in media containing Kan or Gm. Cells that have successfully recombined express the correct AB resistance and survive, cells that did not recombine die. This process promotes the survival of bacteria carrying more active IS621 variants.

EcORep is a continuous DE method that allows DE of large DNA segments in *E. coli* [28]. EcORep uses genes from PRD1 phage to replicate a linear orthogonal replicon (O-Replicon) with an increased mutation rate[28]. We attempted to adapt EcORep as specified in Supplementary 6.2 and use it for the evolution of IS621. IS621

is encoded on the O-replacon, which allows the generation of many variants in continuous culture. Expression of IS621 is controlled by the inducible PVanCC promoter. Cells that carry more active variants of IS621 are more likely to recombine during the time interval between bRNA induction and antibiotic (AB) addition. This results in enrichment of bacteria carrying more active recombinase variants.

### 3.2.2 Target sequence selection

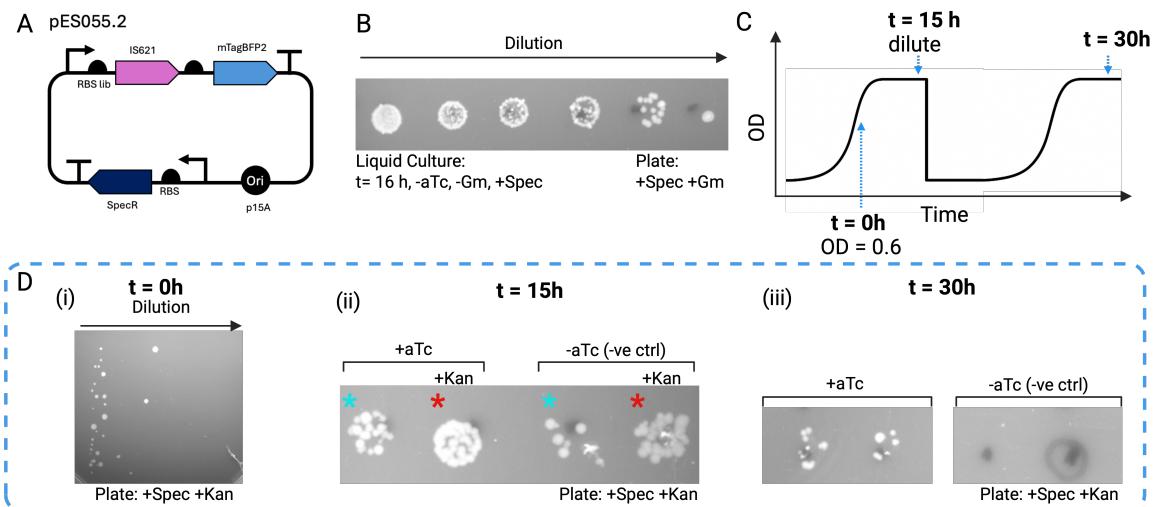


**Figure 3: Flow cytometry analysis and sequencing validation of IS621 recombination activity between the wild-type donor and the targets identified within the SERPINA1 locus.** **A** Schematic representation of the plasmids used in the flow cytometry analysis. pES049.2 contains the bRNA targeting the respective target sequence and the two fluorescent reporter genes, mGreenLantern and mScarlet3, facing in opposite directions. These are flanked by donor and target sequences to allow for IS621 + bRNA mediated flipping. pES055.2 contains IS621 followed by mTagBFP2 to facilitate the gating of cells bearing both plasmids. Induction with aTc leads to expression of the bRNA and therefore recombination. **B** Percentage of mScarlet3-positive cells over the number of cells expressing both IS621 recombinase and the fluorescent reporter plasmid. Marionette-Clo cells carrying plasmids pES049.2 and pES055.2 were either uninduced (blue) or induced (green) with 200 nM aTc for the indicated times before the analysis. Bars indicate mean values of three replicates, with individual data points shown. Each panel represents a different target sequence (1, 2 or 3). The non-targeting bRNA condition corresponds to cells carrying pES055.2 and pES049.2 bearing bRNA 1 and target sequence 2, serving as a negative control.

Suitable targets for bridge recombinase mediated AATD treatment were selected as follows. As IS621 activity is target and donor sequence dependant [18], we identified three candidate target sequences (target 1-3) that show high similarity to the IS621 wild-type target sequence and minimal similarity to any sequence in the *E. coli* genome to reduce off targets (methods 5.3). The sequence is inserted between exon 1 and exon 2 within the SERPINA1 locus, as nearly all disease causing mutations occur downstream of exon 1 (Figure 1 C). To assess the recombination efficiency of selected targets, we expressed two plasmids in the *E. coli* Marionette-Clo strain: one encoding the IS621 recombinase and another, the "screening plasmid", carrying the respective bRNA under the control of the PTet promoter along with two fluorescent reporter genes oriented in opposite

directions (Figure 3A). Natively, the plasmid expresses *mGreenlantern* under the control of a constitutively active promoter; after inversion, *mScarlet* is placed downstream of the promoter and expressed. *mGreenlantern* is no longer expressed (Figure 3 A). Flow cytometry can be used to measure the recombination efficiency for each target site by measuring the percentage of mScarlet3-positive cells. Inversion was induced by addition of aTc which triggers bRNA A expression. After 15 h of induction targets 1-3 showed 6 % to 16 % of mScarlet3-positive cells, whereas, the non-targeted target-sequence showed < 1 % positive cells (Figure 3B), confirming target-specific inversion events. Similarly, we observed <1 % positive cells in the uninduced control in all conditions, except target 2 and 3, where we observed 2% positive cells at the 15 h time point. This is indicative of effective promoter repression in the absence of the inducer. Furthermore, we were only able to observe a change in fluorescent protein expression after 15 hours of induction, as there were <1 % mScarlet positive cells 2 hours after induction across all conditions. Lastly, flipping events were confirmed by Oxford Nanopore sequencing. Target 1 showed the highest activity and was selected as the primary target sequence from this point forward.

### 3.2.3 Validation of the selection logic

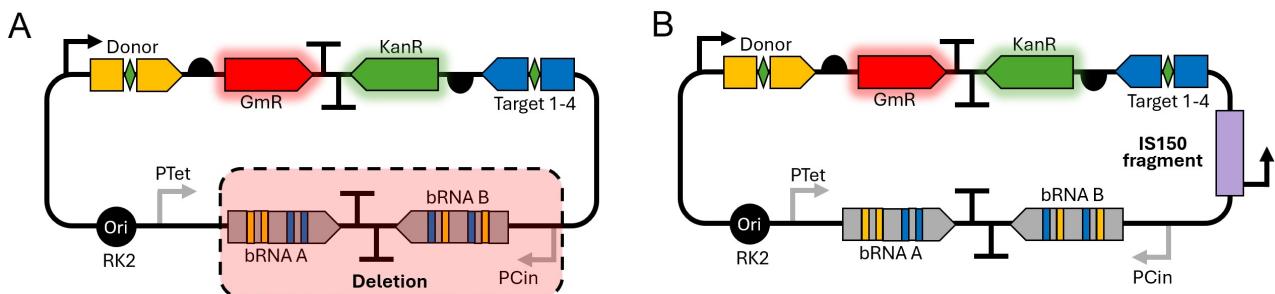


**Figure 4: Phenotypic selection for successful inversion by IS621 following induction of bRNA expression with aTc.** **A** Schematic representation of plasmid pES055.2 containing the IS621 gene under control of a constitutively active promoter and the SpecR selection marker. **B** Serial dilution droplet plating of the strain carrying pES055.2 and pES013.2A (see Figure 2) after 15 h growth in absence of Gm on Gm/Spec plates. **C** Schematic representation of the induction and growth conditions of bacteria containing the selection plasmid and constitutively expressing IS621. The cells were induced with aTc at an OD<sub>600</sub> of 0.6 either in the absence or presence of Kan. The cultures were diluted in fresh media after 15 h and grown for an additional 15 h (methods). **D** Bacterial growth on spectinomycin/Kan agar plates of cells after 0 h, 15 h and 30 hours after induction and antibiotic selection. (i) Serial dilution of the bacterial culture before induction, (ii) induced samples with their respective negative controls, sample and control indicated with coloured asterisks (No selection with Kan in liquid cultures (cyan), with Kan in liquid culture (red)). Cultures grown with Kan were terminated after 15 h, (iii) induced samples grown in absence of Kan for 30 h plated as droplets in duplicate. *Black markings on plates indicate the position of bacterial droplet when plated. All plates were imaged with the coomassie setting on a gel-doc XR.*

To identify optimal induction and selection conditions for detecting IS621-mediated inversions, we screened nearly 100 conditions, with the most promising ones discussed below. In this screening assay, we used a simplified model system where we co-transformed the selection plasmid (pES013.2A) with a plasmid constitutively expressing IS621 (pES055.2 Figure 4) as in section 3.2.2. In this system, inversion is induced by the addition of aTc to the media leading to the expression of the bRNA A. We first demonstrated that the bacterial population maintains the selection plasmid in the absence of AB selection by Gm or Kan in liquid culture over the course of 15 h (Figure 4 A). This confirmed that recombination can be induced in absence of AB selection, leading us to devise an induction scheme 4.

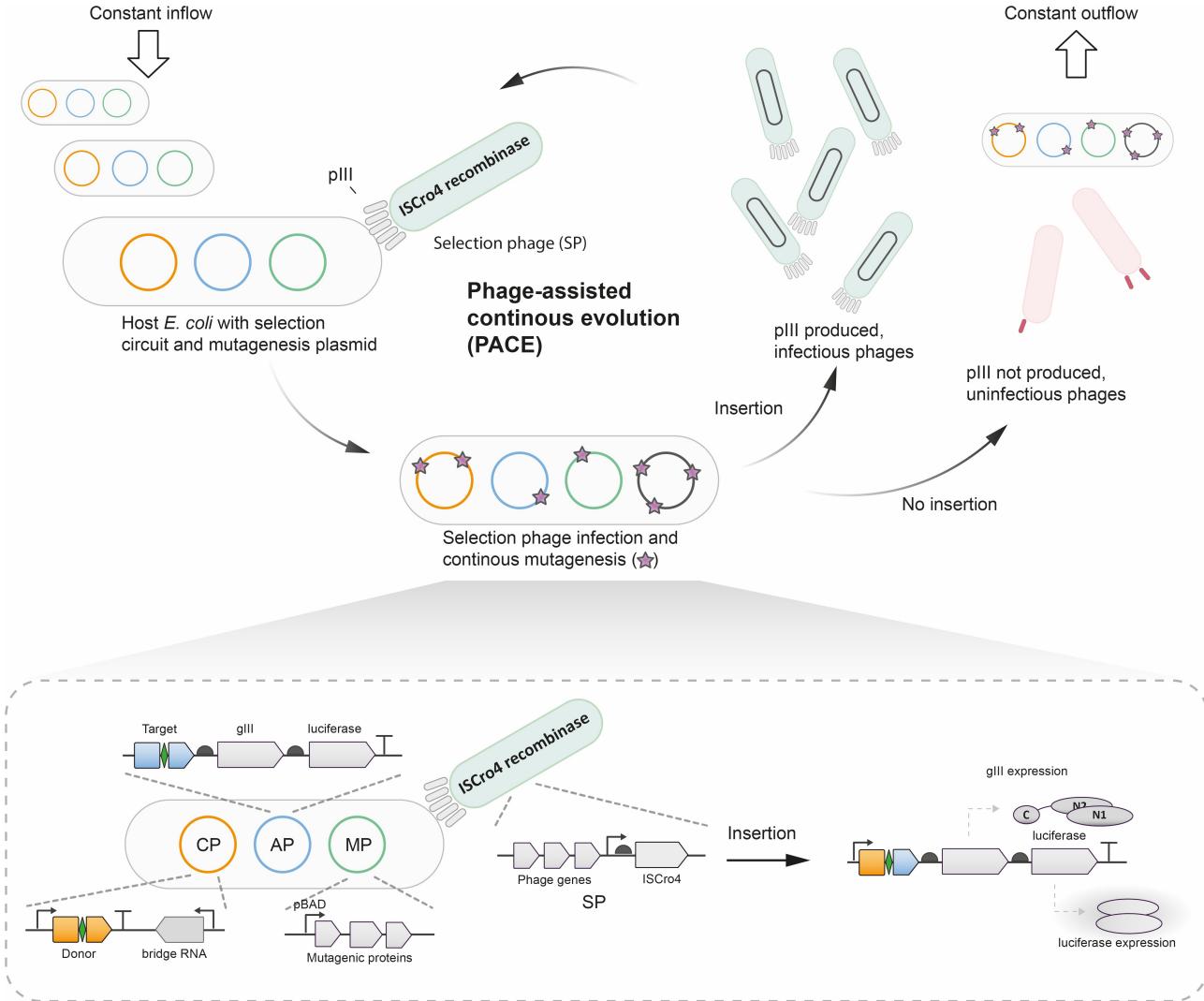
The bacterial population was induced in late exponential phase ( $OD_{600} = 0.6$ ). We tested two conditions: One culture was grown without AB, and another was grown with Kan to select for successful inversion of the AB cassette. Controls where bRNA expression was not induced were included for each condition. After 15 h the bacteria which were grown in the absence of Kan were diluted in fresh media and grown for an additional 15 hours, allowing for a longer induction period while reducing cell death. Before induction ( $t=0$ ) we observed some bacterial growth on Kan plates, indicating some background resistance of the native bacterial population (Figure 4 D (i)). After 15 h, the bacterial cultures induced with aTc showed more growth than their uninduced control. However, we observed significant growth in the uninduced conditions, particularly in the culture grown in the presence of Kan. Interestingly, after 30h, uninduced cultures showed no growth on Kan plates, whereas induced cultures grew (4 D). Following this assay monoclonal cultures were grown and their DNA sequenced. While the sequencing data confirmed inversion of the AB cassette, it also detected escape mutations, in some of the colonies grown in the presence of Kan.

Two escape mutations were identified: large deletions as well as insertions of genomic host DNA (5). The deletion placed transcription of *KanR* under the control of the pTet promoter, which otherwise controls bRNA A expression, enabling cell survival without inversion of the cassette. The second escape mutation we observed was a partial insertion of the transposable element IS150 from IS3 family [29]. This fragment contains a promoter which leads to the expression of *KanR* and cell survival.



**Figure 5: Schematic of escape mutants.** In the unmodified selection plasmid, the resistance cassette is under the control of only one promoter. Gene deletion (A), off-target recombination or insertion of an exogenous gene sequence (B) may lead to re-localization of a second promoter downstream of the cassette. This can lead to the formation of escape mutants. Escape mutants express *GmR* and *KanR* simultaneously and thus no longer need to flip the cassette for survival during selection. During our selection experiments we found two escape mutants. One showed a large deletion. Both bRNAs were absent from the plasmid and the *KanR* gene was under the control of the pTet promoter. The other escape mutants carried a fragment of IS150. The IS150 fragment contains an additional promoter and supports *KanR* expression.

### 3.3 System for phage-assisted evolution of bridge recombinases



**Figure 6: Overview of phage-assisted continuous evolution (PACE) for bridge recombinase ISCr04:** Selection phage (SP) carries the evolving ISCr04 recombinase but lacks *gIII*, which encodes the essential coat protein pIII. The host *E. coli* harbours three plasmids: (i) an accessory plasmid (AP) containing the target site plus *gIII* and a luciferase reporter that are promoterless; (ii) a complementary plasmid (CP) expressing the bRNA and donor sequence under an upstream promoter; and (iii) a mutagenesis plasmid (MP) [26]. Upon successful insertion, the CP promoter is positioned immediately upstream of *gIII* (and luciferase), which drives pIII expression and enables replication of SPs encoding active recombinase variants. PACE is conducted in a fixed-volume “lagoon” with continuous dilution by fresh host cells, so only SPs that replicate faster than the dilution rate persist and evolve; ongoing mutation from the MP diversifies progeny for subsequent rounds of selection.

Although ISCr04 is highly active in human cells, its activity in *E. coli* has not been documented. To test whether ISCr04 can catalyse promoter-repositioning insertions (modelling the insertion needed to restore a healthy *SERPINA1* copy) in bacteria when paired with designed bRNAs, we constructed an sfGFP reporter that mirrors the PACE selection logic (Figure 7). In this system, sfGFP is carried on a low-copy plasmid with the recombination target site (either ISCr04 WT target, SERPINA1 target 1-3 (3.2.2) or ISCr04 WT donor sequence) upstream and no promoter - analogous to *gIII* on the AP - and is co-transformed with a CP encoding

the bRNA and a promoter–donor cassette, together with an arabinose-inducible ISCrO4 expression plasmid.

Here we outline the rationale, design and preliminary testing of a insertion based phage-assisted continuous evolutions (PACE) [26] logic for evolving bridge recombinases on the example of the ISCrO4 bridge recombinase.

The selection logic leverages the essential phage gene *gIII*, which encodes the coat protein pIII required for infectivity and replication. In our system (see Figure 6A), the selection phage (SP) carries the evolving ISCrO4 recombinase but does not contain *gIII*; as a consequence, SPs cannot replicate unless *gIII* is provided in trans by the host. To couple *gIII* expression tightly to successful recombination, the accessory plasmid (AP) carried by the host encodes *gIII* (and a luciferase reporter) without a promoter, and places the appropriate recombination target sequence upstream of these genes. A separate complementary plasmid (CP) encodes the bRNA and the donor sequence, and includes a promoter upstream of the donor. The intended recombination event moves that promoter into the correct position in front of *gIII* (and luciferase, when present). Thus, without recombination, there is no *gIII* expression and the SP cannot propagate; with successful recombination, the promoter is repositioned to drive pIII production, enabling replication of SPs that encode functional (and ideally improved) ISCrO4 variants. In continuous culture, this design should favour variants that recombine efficiently at the programmed site, because only those variants gain access to pIII and can outcompete the dilution imposed by the lagoon flow.

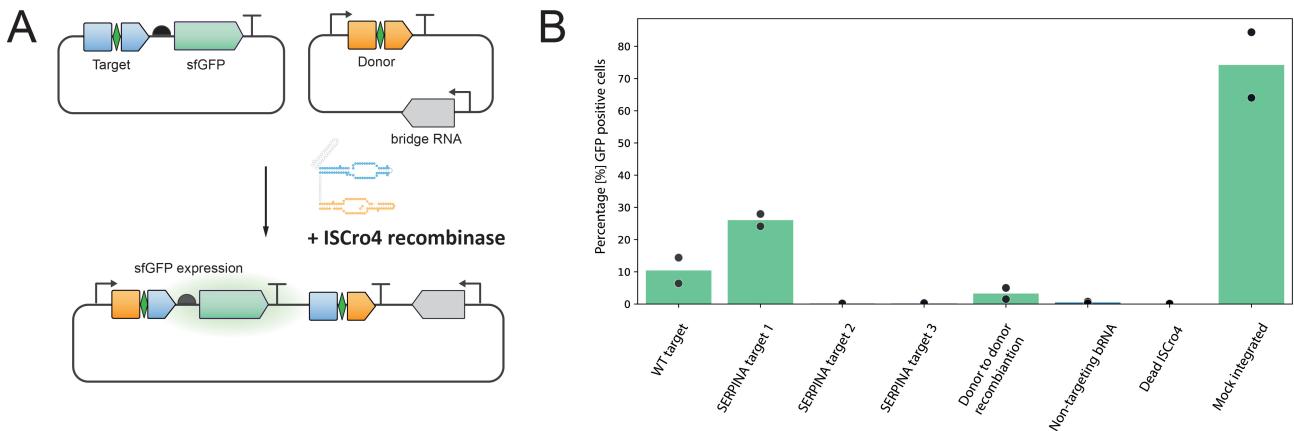


Figure 7: *sfGFP* activity assay for ISCrO4 in *E. coli*. **A** Schematic of sfGFP activity assay. A low-copy plasmid carries sfGFP with its target site upstream and no promoter (AP-analog). The high-copy CP encodes the bRNA and a promoter–donor cassette. ISCrO4 is expressed from an arabinose-inducible plasmid. Recombination repositions the CP promoter directly upstream of sfGFP, leading to sfGFP expression. **B** Flow-cytometry readout after ~20 h post arabinose induction at 37°C showing %GFP-positive cells

With this setup, we observed that wild-type ISCrO4 catalyses insertion in *E. coli* in a recombinase- and bRNA-dependent manner: sfGFP activation was not detected with catalytically dead ISCrO4 or with non-targeting bRNA controls. Moreover, bRNA reprogrammed to SERPINA target 1 exhibits activity above the wild-type target after 20 h after arabinose induction. While the precise percentage of GFP-positive cells depends on construct and growth conditions, these results establish that (i) ISCrO4 is active in *E. coli* and (ii) activity can be redirected by designed bRNAs to non-native targets. Together, the sfGFP data support the feasibility of using the same promoter-repositioning logic to control *gIII* during selection. Although ISCrO4 catalyses the intended donor–target insertion, prior work has also reported donor–donor recombination, a major potential source of

off-target events [21]. To directly assess donor–donor activity, we modified the AP by replacing the upstream target site with a copy of the donor sequence. In line with the previously published data [21], ISCro4 exhibited donor–donor recombination.

To quantify the transcriptional gain produced by promoter repositioning and to validate that the post-recombination state strongly expresses *gIII*, we designed a mock-integrated plasmid that mimics the recombination product: the promoter is pre-installed upstream of *gIII*. On the same transcript downstream of *gIII*, a luciferase is expressed to enable a luminescent readout proportional to transcription through the *gIII*–lux cassette. When comparing strains carrying the pre-recombination configuration versus the mock-integrated (post-recombination) construct, we observed a ~10,000-fold (four-orders-of-magnitude) increase in luminescence for the mock-integrated state (Figure 8), indicating a working transcription unit upon site-specific insertion.

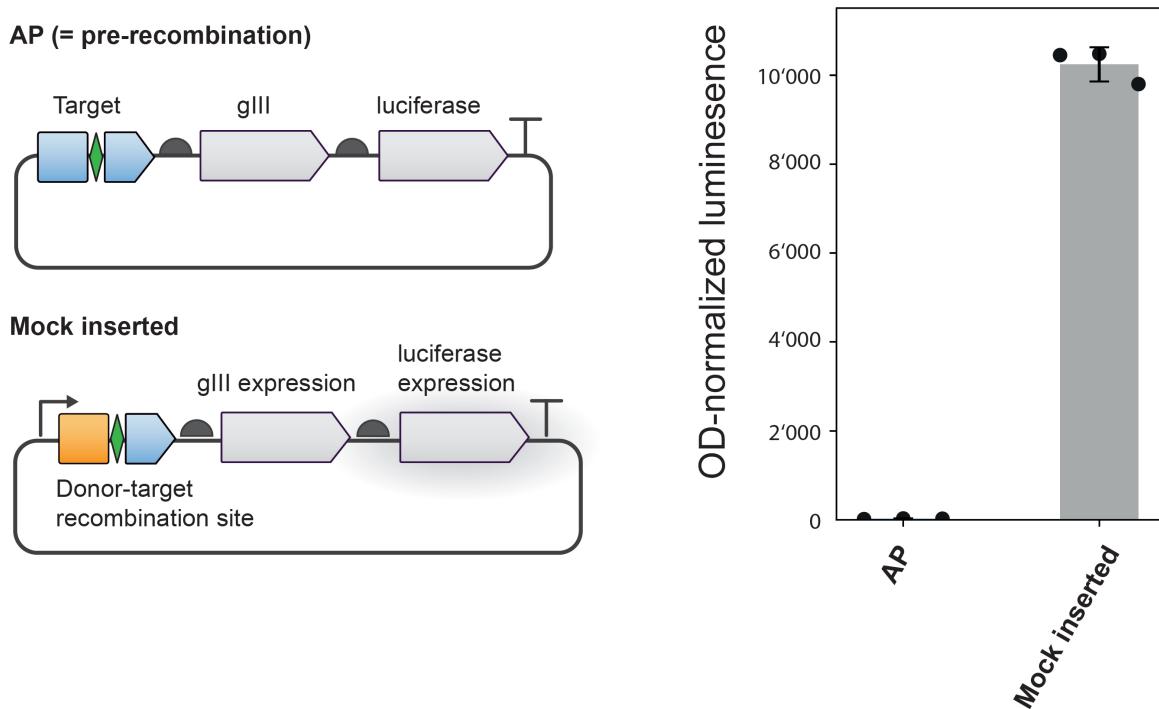


Figure 8: *Validation of promoter repositioning*: The mock inserted plasmid mimics the post-recombination product which leads to constitutive *gIII* expression. Additionally, a luciferase gene placed downstream on the same transcript provides a luminescent readout proportional to transcription to *gIII*. Comparing the pre-recombination state and the mock inserted states reveals a ~ 10<sup>4</sup>-fold luminescence increase in the mock-integrated condition, confirming strong *gIII* transcription upon promoter repositioning in *E. coli* after 16–18 hours incubation at 37°C.

Having established activity and a transcriptional readout, we next asked whether the full selection logic supports phage propagation. We infected *E. coli* carrying the AP and CP with the ISCro4 selection phage (SP-ISCro4) and incubated cultures overnight (Figure 9A). As an initial test, we used WT ISCro4 with its cognate bRNA and quantified titres the following day. Under these conditions, no significant titre increase was observed relative to input phage titre and negative controls (non-targeting bRNA). Nonetheless, junction-specific PCR spanning the donor–target recombination site yielded the expected product, confirming that recombination occurred despite the lack of phage amplification.

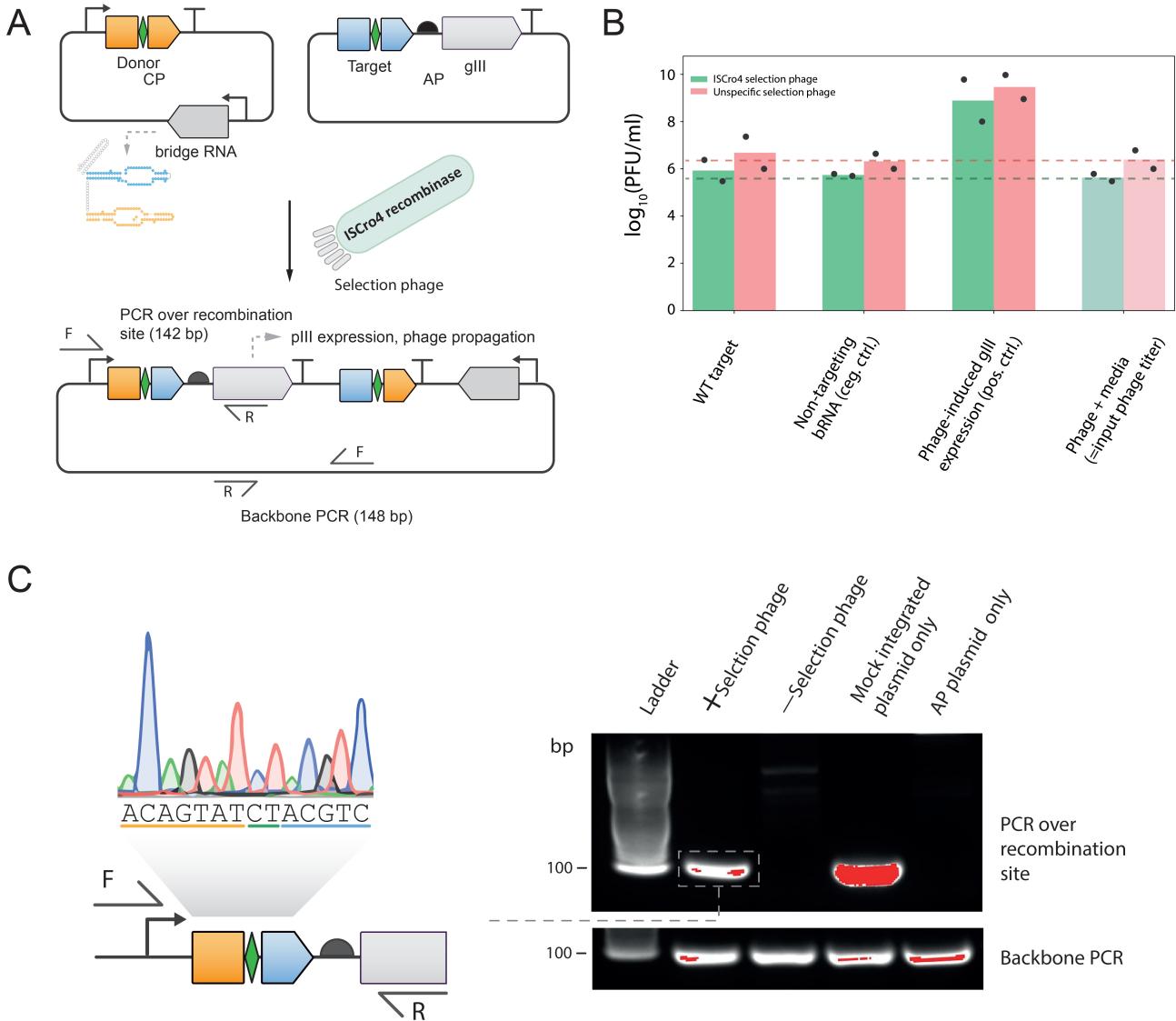


Figure 9: *Overnight phage propagation assay with WT ISCr4*. **A** Setup: SP-ISCr4 (no *gIII*) infects hosts bearing AP (promoterless *gIII* + target) and CP (bRNA + promoter-donor); recombination should reposition the promoter to drive *gIII*. **B** Result: After overnight infection with SP-ISCr4, phage titers show no significant increase versus input phage titer and non-targeting bRNA negative controls; likewise for the unspecific selection phage expressing *rpoZ*. Positive control shows that SP-ISCr4 is capable of propagation when supplied with pIII in trans. **C** Validation: Junction PCR detects the expected donor-target product, confirming recombination despite the lack of phage amplification.

## 4 Discussion

### 4.1 Plasmid-based selection system for evolution of IS621 in *E. coli*

Our project shows significant progress made towards developing an evolutionary platform for bridge recombinases. We were able to induce and select for inversion catalysed by IS621 in the forward direction (Gm → Kan). As a next step we aim to demonstrate inversion in the reverse direction (Kan → Gm) following induction of the bRNA. Successful implementation of this step would enable continuous directed evolution. This could in-

volve growing the bacteria in a continuous culture where the media is exchanged slowly overtime to contain different concentrations of the inducers and antibiotics used for selection. Alternately, non-continuous cultures can be used where the bacteria are subjected to iterative cycles of induction and selection. The applicability of our system is currently limited by a number of escape mutations that allow cells to grow without the need for IS621-catalysed recombination. To counteract this, we have devised an experimental strategy which aims to reduce the number of possible escape mutants. The incorporation of CcdA/CcdB toxin/antitoxin system into the selection plasmid as shown in figure 10 should greatly reduce the number of escape mutants.

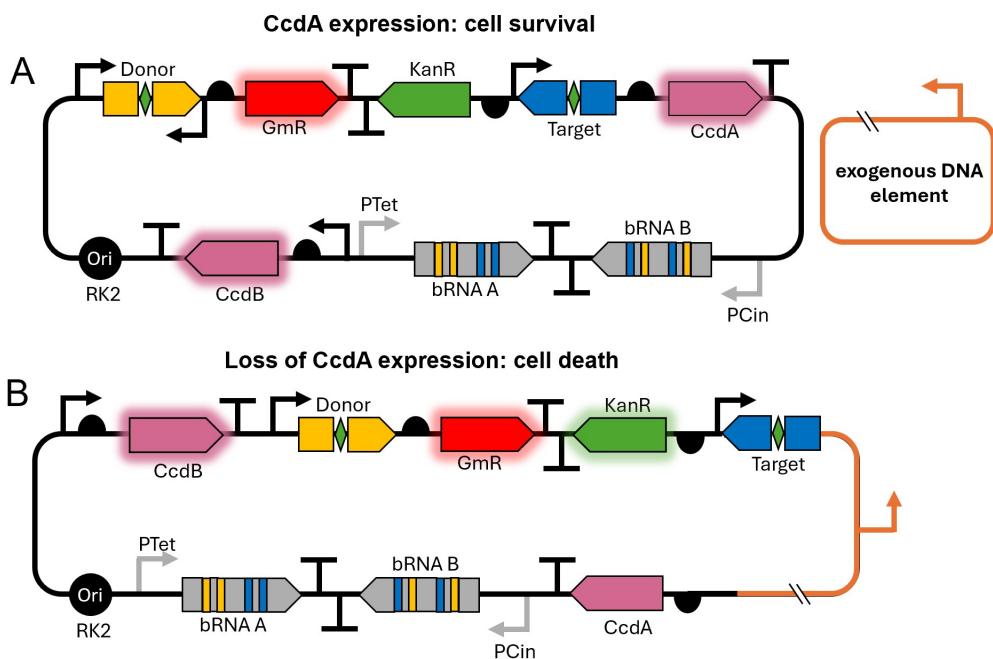


Figure 10: *Introduction of the CcdA/CcdB toxin/antitoxin system in the selection plasmid to reduce the number of possible escape mutants.* **A.** Native configuration of the selection plasmid with the integrated CcdA/CcdB system. Expression of CcdA suppresses CcdB toxicity. **B.** Relocalization of the second promoter downstream of the cassette leads to loss of CcdA expression and cell death.

We attempted to apply our evolutionary logic in combination with EcoRep. This system allows for an orthogonally increased mutation rate on the target gene *in vivo*. [25]. Unfortunately, we have not been able to test our selection strategy with the EcoRep system within the time frame of our project. However, the selection logic is not directly linked to the method of gene diversification. Consequently, the system should work in combination with EcoRep given that both systems have been shown to work independently. The selection could also be employed in combination with any other DE method as long as selection takes place *in vivo*. Furthermore, the strategy requires only simple laboratory equipment.

#### 4.2 PACE based selection system for evolution of ISCro4

Although we confirmed that ISCro4 is active in *E. coli* and that correct insertion drives strong gIII transcription, we saw no titer increase under initial conditions, despite junction-PCR confirming recombination. Two factors likely explain this: (i) host burden from AP+CP plus ISCro4 expression and the recombination process itself, which can slow growth, impair F-pilus biogenesis, and limit phage production/release; and (ii) kinetics: the rates of recombination of WT ISCro4 and subsequent pIII accumulation may be too slow to support measurable

propagation.

To overcome these barriers, we will begin with neutral drift of the selection phage and phage-assisted non-continuous evolution (PANCE) under permissive conditions to accumulate diversity. Once amplification is detectable, we will incrementally raise stringency and transition toward continuous selection. Additionally DML can be employed to select a higher efficiency starting variant to be used in combination with the PACE logic.

In this project we have been able illustrate two alternative approaches to evolve bridge recombinases. Having access to a toolbox capable of doing so is an important milestone to effectively expanding the genomic engineering toolbox.

## 5 Methods

### 5.1 Strains, plasmids, and growth media

#### 5.1.1 Generation and validation of sEM6

The strain was generated using lambda red recombineering using pSIJ8 [30] to delete the csg operon from Marionette-Clo (sAJM.1504) [31]. The csg operon was deleted to prevent biofilm formation [32]. The deletion was done using lambda red recombineering as described here [30]. Fresh cultures of Marionette-Clo were grown to an OD<sub>600</sub> of 0.3 and then induced with 15 mM L-arabinose (final concentration) for an hour. Subsequently the cells were made electrocompetent. The linear dsDNA (see Table 2) was electroporated in to the cells. Then they were recovered and plated. Six colonies were picked the following day and liquid cultures were set up. Finally the flippase was induced in theese liquid cultures with 50 mM L-rhamnose (final concentration) and grown at 38°C to cure pSIJ8. The deletion was validated using colony PCR . The cultures from the single colonies are designated S1-S6 (see Figure 13).

#### 5.1.2 Cloning strains

*E. coli* NEB 10-beta (New England BioLabs, USA) was used for all routine cloning. The plasmids with inducible promoters were transformed into sEM6 (Further details can be found in section 5.1.1.).

For PACE subproject *E. coli* strain S2060 (Addgene #105064) was used. Unless stated otherwise, cultures were grown in LB or 2xYT (16 g/L Tryphone, 10 g/L Yeast extract, 5 g/L NaOH) at 37 °C with shaking at 220 rpm.

To achieve the higher plasmid yield, the cultures were grown in LB25-Media (10 g/L Tryphone, 24 g/L Yeast extract, 5 g/L NaCl, 1mM NaOH) for plasmid amplification. LB-Media (1% Tryphone, 0.5% Yeast extract, 1% NaCl, 1mM NaOH) was used for flow cytometry assay and for the phenotypic selection. LB+1% agar plates, supplemented with corresponding antibiotic, were used for cloning and selection.

Antibiotics (Sigma-Aldrich) were used at: streptomycin (50 µg/mL), spectinomycin (60 µg/mL), chloramphenicol (25 µg/mL), carbenicillin (50 µg/mL), kanamycin (25 µg/mL), and ampicillin (100 µg/mL).

The recovery after transformation was done in SOC medium (0.5% Yeast Extract, 2% Tryptone; 10 mM NaCl,

2.5 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM MgSO<sub>4</sub>, 20 mM Glucose).

### 5.1.3 Cloning

Restriction enzymes PaqCI, BsaI, and BbsI were purchased from New England BioLabs, USA. As a reaction buffer NEBridge® Golden Gate Assembly (New England BioLabs, USA) was used for Golden Gate reactions. All promoters, terminators, Ribosome Binding Sites (RBSs), Stops, and tags were provided by the MoClo kit (Addgene Kit #1000000044). All further DNA fragments were obtained codon-optimized from Twist Bioscience or Integrated DNA Technology (IDT).

Inserts and vector backbones were PCR-amplified with Phusion™ High-Fidelity DNA Polymerase (NEB) to append assembly overhangs. The full list of Oligos used for PCR can be found in chapter 1. Post-PCR, reactions were treated with DpnI to remove template DNA, purified by gel extraction (GeneJET PCR Purification Kit, Thermo Fisher Scientific), and assembled per manufacturer's protocols.

Wild-type ISCr4 and IS621 protein sequences and cognate bRNA sequences were obtained from ISfinder. bRNAs for SERPINA targets 1–3 were designed using the enhanced single bRNA architecture (TBL4+DBL3) [21] following published reprogramming rules for the IS621 bridge recombinase.[18], [20]

A combination of Golden Gate (EcoFlex MoClo kit)[33] and Gibson Assembly Master Mix (New England BioLabs) was used to assemble the final plasmids. To increase the cloning efficiency, the golden gate assembly was carried out with the following protocol: 20 min at 37 °C, (5 min at 37 °C, 5 min at 16 °C) x 60 cycles, 20 min at 65 °C, 10 min at 80 °C.

The final cloning step of pES013.2A was performed with Gibson Assembly due to the size and complexity of the plasmid. The DNA vectors were digested with PaqCI (rCutSmart™ Buffer, New England BioLabs, USA; 2 hours incubation at 37 °C), gel purified with agarose gel electrophoresis (1-2 % agarose; ReadySub-Cell GT Cell, BioRad, USA; 1 kb DNA Ladder, New England BioLabs, USA), and ligated with T4 Ligase (New England BioLabs, USA) with the following protocol: 10 min at 37°C, (5 min at 37°C, 5 min at 16°C) x 30 cycles, 10 min at 80°C.

Minipreps were performed using GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) or Zippy Plasmid Miniprep Kit (Zymo Research), and DNA was eluted in Milli-Q water. A complete list of plasmids, selection phages, and primers is provided in Table 1.

The plasmid sequences were confirmed with full plasmid sequencing by Microsynth, Germany.

### 5.1.4 Transformation

For routine transformations commercially available competent NEB 10-beta and NEB 5-alpha cells were purchased from New England BioLabs. The heat-shock competent sEM6 cells were prepared in-house with Mix 'n Go kit (Zymo Research, USA). Before use competent cells were stored at –80 °C.

Single plasmid transformations were performed by adding 1-5 µL of assembly reaction to 20-50 µL of competent

cells, incubated on ice for 30 min, followed by heat shock at 42°C for 30 s and recovery in SOC medium at 37°C, 750 rpm for 60 min. The culture was then spread on LB agar plates containing the appropriate antibiotic and incubated at 37°C overnight. For co-transformation of multiple plasmids, 50 ng of the two accessory plasmids were mixed in an equimolar ratio and added to 25 µL of competent cells. The cells were then incubated on ice for 30 min, heat shocked at 42°C for 30 s, and recovered in SOC medium at 37°C, 750 rpm for 60 min. Cells were then spread on LB agar plates containing the appropriate antibiotics and incubated at 37°C overnight.

## 5.2 Preparation of electrocompetent *E. coli*

To prepare the electrocompetent cells, 100 µL of the overnight cultures of sEM 6 *E. coli* were subcultured into 10 ml of LB media (initial OD of tilde 0.05) and grown at 37 °C, 250 rpm. until the OD<sub>600</sub>=0.3 was reached. Cultures were induced with 15 mM L-arabinose for 45 min and later centrifuged (4,400xg, 4 °C, 5 min). Cells were washed 4 times by resuspending the pellets in 10 ml of 10% glycerol in deionised water and centrifugation (6,000xg, 4 °C, 3.5 min) and pellets were resuspended in 2 ml 10% glycerol. After the washing, pellets were resuspended in 5 mL of 10% glycerol and stored at -80 °C until use.

## 5.3 SERPINA1 target sequence selection

The SERPINA1 gene sequence was extracted from NCBI gene databases (GCF 000001405.40). All 14 bp sequences with a CT core of intron 1 (position 2,135-7,460) were extracted. Sequences with a direct match or one mismatch to any sequence in the *E. coli* K12 MG1655 genome (GCA 000005845) were removed.

Target sequences with a log read abundance above two in supplementary Tablethree, sheet "Fig. 4E Enriched Nucleotides", from Durrant *et al.*[18], were classified as WT target sequences and used to calculate a position weight matrix. The first 11 bp of all extracted sequences were scored by the position weight matrix and sequences with a score > -17 were selected. Finally, Target A (position 2406), Target B (position 2518), and Target C (position 7094) were selected.

## 5.4 Inducers

Bacteria were induced with the following compounds: N-(3-Hydroxytetradecanoyl)-DL-homoserine lactone (OHC14) from Sigma Aldrich and anhydrotetracycline (aTc) from Bioss.

## 5.5 Flow Cytometry

*E. coli* Marionette-Clo strain was transformed with the plasmids of interest and spread on 1% agar LB-plates and incubated overnight at 37°C. Single colonies were inoculated in 5 mL of LB medium, supplemented with antibiotics and grown at 37°C, 220 rpm overnight. The overnight cultures were then diluted 1:100 in 5mL LB + antibiotics and incubated at 37°C, 220 rpm until they reached OD<sub>600</sub>=0.1. 0-hour time point was collected just before treatment started. Then, 200 nM aTc were added to the culture and samples were collected at 2 and 15 hours post treatment. Samples were diluted in PBS and analyzed with Attune NxT flow cytometer. Samples were run at a rate of 25 µL/min. Measurements were acquired using the following instrument settings: FSC voltage of 240 V, SSC voltage of 240 V, blue laser (488 nm) voltage of 360 V, violet laser (405 nm) voltage of 400 V and yellow laser (561 nm) voltage of 420 V. The events were gated by forward scatter height (1,500-15,000),

forward scatter area (700-10,000), and side scatter area (2,000-80,000) to identify singlets. IS621-expressing cells were further gated using the mTagBFP2 marker, with positive cells defined by fluorescence intensity values greater than those of a mTagBFP2-negative control (>1,661). Finally, cells were gated for higher expression of either mGreenLantern (>11,130) or mScarlet3 (>441). For each sample, the percentage of mScarlet3-positive cells (threshold: 441) was calculated. All gating, threshold and quantification procedures described above were implemented using custom Python scripts.

## 5.6 Plating Assay

pES013.2 and pES055.2 were co-transformed into *E. coli* sEM6 strain, spread onto LB agar containing appropriate antibiotics and incubated overnight at 37°C. Single colonies were inoculated in 5 mL of LB medium containing appropriate antibiotics and grown at 37°C, 220 rpm overnight. The overnight cultures were diluted 1:100 in 5 mL LB, supplemented with spectinomycin and gentamicin until OD<sub>600</sub> = 0.6. The cultures were centrifuged at 3000 xg for 7 min. The pellet was resuspended in fresh culture medium (LB and Spec) and aTc (10 µM/ml) and Kan (µg/ml) were added as described in results 3.2.3. After 15 h a 2 ml 96-deep-well plate was prepared with the various induction (0 or 200 nM/ml aTC) /antibiotic conditions (0-200 µg/ml Kan) and inoculated with the 15 h bacterial cultures (1:100 dilution). The plate was sealed with Breathe Easier sealing membrane for multiwell plates (Merck, Germany) and grown at 37°C, 220 rpm. Conditions were plated after 0, 15 and 30 h on Spec/Kan, Spec/Gm and Spec/Gm/Kan LB agar plates.

## 5.7 GFP Reporter Assay for ISCro4

S2060 cells were co-transformed with: (i) an sfGFP accessory plasmid encoding the genomic target upstream of an sfGFP CDS; (ii) a complementary plasmid (CP) encoding the promoter and donor sequence and the bRNA; and (iii) an arabinose-inducible plasmid expressing wild-type ISCro4. Single colonies were inoculated into selective medium and grown overnight. Cultures were back-diluted 1:100 and grown at 37 °C, 220 rpm to OD<sub>600</sub> = 0.4–0.6. ISCro4 expression was induced with 10 mM L-arabinose. After 20 h, samples were analyzed using an Attune NxT flow cytometer. Data were acquired at a flow rate of 25 µL/min and 30,000 events were analysed per sample. Measurements were made using a FSC voltage of 240 V, a SSC voltage of 240 V, and a blue laser (488 nm) voltage of 480 V. Mean GFP fluorescence (population-wide, GFP<sup>+</sup> and GFP<sup>-</sup>) was quantified using custom Python scripts.

## 5.8 Luminescence Assay

S2060 cells were transformed with either AP+CP or a mock-integrated control. Single colonies were used to start overnight cultures that were back-diluted 1:100 and grown for 20 h. Aliquots (200 µL) were transferred into black-walled, clear-bottom 96-well plates. Luminescence and absorbance at 600 nm were recorded on a plate reader (Tecan). Luminescence values were normalized to OD<sub>600</sub>.

## 5.9 Determination of Phage Titer by Plaque Assay

LB agar plates (diameter 96 mm, height 16 mm; containing ampicillin) were used as bottom agar. For top agar, 4 mL molten LB agar (kept at 50 °C) was mixed with 0.6 mL S2208 cells at OD<sub>600</sub> = 0.4–0.6, then poured onto pre-warmed plates. Ten-fold serial dilutions of phage were prepared, and 10 µL of each dilution was spotted

onto the top agar. Plates were incubated overnight at 37 °C and plaques were counted. Titers were calculated as:

$$\text{PFU/mL} = \# \text{plaques} \times \text{dilution factor} \times 100.$$

### 5.10 Determination of Phage Titer by qPCR

Phage samples were clarified by centrifugation (8,000 rpm, 2 min) and the supernatant was used directly as template. Reactions (10 µL) contained 5 µL LUNA qPCR Master Mix (NEB), 0.5 µM forward primer (5-CACCGTTCATCTGTCCTCTT-3), 0.5 µM reverse primer (5-CGACCTGCTCCATGTTACTTAG-3), and 2 µL sample. Titers were computed from Ct values against a plasmid/phage DNA standard curve.

### 5.11 Phage production

SplitC (Addgene #138523) and SplitD (Addgene #138521) plasmids were assembled with a PCR amplicon containing wild-type IS<sub>C</sub>re4 plus an RBS, using SphI Golden Gate with 5 overhangs AGT/GGC. Cycling: 30 cycles of 37 °C for 1 min and 16 °C for 1 min, followed by 60 °C for 5 min. Two microliters of assembly were mixed with 98 µL KCM solution (100 mM KCl, 30 mM CaCl<sub>2</sub>, 50 mM MgCl<sub>2</sub>) and added to 100 µL competent S2208 cells. After 10 min on ice, cells were heat-shocked for 1 min 15 s at 42 °C, then recovered by transferring the entire sample to 8 mL LB (no antibiotics) and incubating 18 h at 37 °C, 500 rpm on a microplate shaker. Phage plaques were assayed, a single plaque was picked for propagation, and supernatants were 0.22 µm filter-sterilized. Insert sequences were verified by Sanger sequencing (Microsynth AG) using primer “M13 gVI rev” (5-GAAGGAAACCGAGGAAACGC-3).

### 5.12 Phage Propagation Assay

The design of selection logic plasmids (AP/CP) was guided by prior work increasing CRISPR-associated transposon insertion via PACE and activity assays of the *IS621* bridge recombinase ortholog.[18], [34] S2060 cells harboring AP and CP were inoculated into 2xYT, grown overnight, and back-diluted 1:100 into 1 mL 2xYT in 96-well deep-well plates (Axygen). At OD<sub>600</sub>=0.4–0.6, selection phage was added to a target concentration of 10<sup>6</sup> PFU/mL. Cultures were incubated 16–18 h at 37 °C with shaking at 800 rpm on a microplate shaker. For junction-PCR verification, aliquots were normalized to OD<sub>600</sub>=0.05 and used directly as template (2 µL in 50 µL reaction). 0.5 µM each primer was used. For the junction-PCR oES510\_junction\_qPCR\_F and oES511\_junction\_qPCR\_R were used. For backbone PCR oES512\_BB\_qPCR\_F and oES513\_BB\_qPCR\_R were used. PCRs were carried out with Phusion™ High-Fidelity DNA Polymerase (NEB) using the following program: 10 min at 98°C, (15 s at 98°C, 30 s at 62°C, 15 s at 72°C) x 35 cycles, 10 min at 72°C, on hold at 10°C. Amplicons were resolved on 1% agarose, imaged, excised, and purified for Sanger sequencing. For titer determination, cultures were clarified (10 min, 4,400 ×g), and supernatants were collected. Replicate 1 titers were measured by qPCR; Replicate 2 by plaque assay.

## REFERENCES

- [1] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity,” *Science*, vol. 337, no. 6096, pp. 816–821, Aug. 2012, Epub 2012 Jun 28; Publisher: American Association for the Advancement of Science. doi: 10.1126/science.1225829. (visited on 09/27/2025).
- [2] L. Cong, F. A. Ran, D. Cox, *et al.*, “Multiplex genome engineering using CRISPR/Cas systems,” *Science*, vol. 339, no. 6121, pp. 819–823, Feb. 2013, Epub 2013 Jan 3; Publisher: American Association for the Advancement of Science. doi: 10.1126/science.1231143. (visited on 09/27/2025).
- [3] A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, and D. R. Liu, “Programmable editing of a target base in genomic dna without double-stranded dna cleavage,” *Nature*, vol. 533, no. 7603, pp. 420–424, May 2016, ISSN: 1476-4687. doi: 10.1038/nature17946.
- [4] N. M. Gaudelli, A. C. Komor, H. A. Rees, *et al.*, “Programmable base editing of a t to g in genomic dna without dna cleavage,” *Nature*, vol. 551, no. 7681, pp. 464–471, Nov. 2017. doi: 10.1038/nature24644.
- [5] A. V. Anzalone, P. B. Randolph, J. R. Davis, *et al.*, “Search-and-replace genome editing without double-strand breaks or donor DNA,” *Nature*, vol. 576, no. 7785, pp. 149–157, Dec. 2019, Epub 2019 Oct 21; Publisher: Nature Publishing Group. doi: 10.1038/s41586-019-1711-4.
- [6] P. J. Chen and D. R. Liu, “Prime editing for precise and highly versatile genome manipulation,” *Nature Reviews Genetics*, vol. 24, no. 3, pp. 161–177, Mar. 2023, Epub 2022 Nov 7; Publisher: Nature Portfolio. doi: 10.1038/s41576-022-00541-1. (visited on 09/27/2025).
- [7] U.S. National Library of Medicine, *Clinicaltrials.gov*, National Institutes of Health; Website, Sep. 2025. [Online]. Available: <https://clinicaltrials.gov/> (visited on 09/27/2025).
- [8] N. Rudin, E. Sugarman, and J. E. Haber, “Genetic and physical analysis of double-strand break repair and recombination in *Saccharomyces cerevisiae*,” *Genetics*, vol. 122, no. 3, pp. 519–534, Jul. 1989. doi: 10.1093/genetics/122.3.519.
- [9] P. Rouet, F. Smih, and M. Jasin, “Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells,” *Proceedings of the National Academy of Sciences*, vol. 91, no. 13, pp. 6064–6068, Jul. 1994. doi: 10.1073/pnas.91.13.6064.
- [10] S. Vaidyanathan, R. Baik, L. Chen, *et al.*, “Targeted replacement of full-length CFTR in human airway stem cells by CRISPR-Cas9 for pan-mutation correction in the endogenous locus,” *Molecular Therapy*, vol. 30, no. 1, pp. 223–237, Jan. 2022, Epub 2021 Mar 29. doi: 10.1016/j.ymthe.2021.03.023.
- [11] S. S. De Ravin, J. Brault, R. J. Meis, *et al.*, “Enhanced homology-directed repair for highly efficient gene editing in hematopoietic stem/progenitor cells,” *Blood*, vol. 137, no. 19, pp. 2598–2608, May 2021. doi: 10.1182/blood.2020008503.
- [12] W.-D. Heyer, K. T. Ehmsen, and J. Liu, “Regulation of homologous recombination in eukaryotes,” *Annual Review of Genetics*, vol. 44, pp. 113–139, Dec. 2010. doi: 10.1146/annurev-genet-051710-150955.
- [13] M. E. Moynahan and M. Jasin, “Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis,” *Nature Reviews Molecular Cell Biology*, vol. 11, no. 3, pp. 196–207, Mar. 2010. doi: 10.1038/nrm2851.

- [14] M. Kosicki, K. Tomberg, and A. Bradley, “Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements,” *Nature Biotechnology*, vol. 36, no. 8, pp. 765–771, Sep. 2018, Epub 2018 Jul 16. doi: 10.1038/nbt.4192.
- [15] G. Cullot, J. Boutin, J. Toutain, *et al.*, “CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations,” *Nature Communications*, vol. 10, no. 1, p. 1136, Mar. 2019. doi: 10.1038/s41467-019-09006-2.
- [16] M. L. Leibowitz, S. Papathanasiou, P. A. Doerfler, *et al.*, “Chromothripsis as an on-target consequence of CRISPR-Cas9 genome editing,” *Nature Genetics*, vol. 53, no. 6, pp. 895–905, Jun. 2021, Epub 2021 Apr 12. doi: 10.1038/s41588-021-00838-7.
- [17] E. Haapaniemi, S. Botla, J. Persson, B. Schmierer, and J. Taipale, “CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response,” *Nature Medicine*, vol. 24, no. 7, pp. 927–930, Jul. 2018, Epub 2018 Jun 11. doi: 10.1038/s41591-018-0049-z.
- [18] M. G. Durrant, N. T. Perry, J. J. Pai, *et al.*, “Bridge RNAs direct programmable recombination of target and donor DNA,” *Nature*, vol. 630, no. 8018, pp. 984–993, Jun. 2024, Epub 2024 Jun 26. doi: 10.1038/s41586-024-07552-4.
- [19] R. Siddiquee, C. H. Pong, R. M. Hall, and S. F. Ataide, “A programmable seekRNA guides target selection by IS1111 and IS110 type insertion sequences,” *Nature Communications*, vol. 15, no. 1, p. 5235, Jun. 2024, Published 2024 Jun 19. doi: 10.1038/s41467-024-49474-9.
- [20] M. Hiraizumi, N. T. Perry, M. G. Durrant, *et al.*, “Structural mechanism of bridge RNA-guided recombination,” *Nature*, vol. 630, no. 8018, pp. 994–1002, Jun. 2024, Epub 2024 Jun 26. doi: 10.1038/s41586-024-07570-2.
- [21] N. T. Perry, L. J. Bartie, D. Katrekar, *et al.*, “Megabase-scale human genome rearrangement with programmable bridge recombinases,” *Science*, eadz0276, Sep. 2025, Online ahead of print, 2025 Sep 25. doi: 10.1126/science.adz0276.
- [22] C. M. Greene, S. J. Marciniak, J. Teckman, *et al.*, “A1-antitrypsin deficiency,” *Nature Reviews Disease Primers*, vol. 2, no. 16051, p. 16 051, Jul. 2016, Publisher: Springer Nature. doi: 10.1038/nrdp.2016.51. (visited on 09/24/2025).
- [23] J.-F. Mornex, J. Traclet, O. Guillaud, *et al.*, “Alpha1-antitrypsin deficiency: An updated review,” *Presse Médicale*, vol. 52, no. 104170, p. 104 170, Jul. 2023, Publisher: Elsevier Masson SAS. doi: 10.1016/j.lpm.2023.104170. (visited on 09/24/2025).
- [24] M. Hosseini-Kharat, K. E. Bremmell, and C. A. Prestidge, “Why do lipid nanoparticles target the liver? understanding of biodistribution and liver-specific tropism,” *Molecular Therapy: Methods & Clinical Development*, vol. 33, no. 101436, p. 101 436, Mar. 2025, Publisher: Cell Press. doi: 10.1016/j.omtm.2025.101436. (visited on 09/24/2025).
- [25] C. S. Diercks, P. Sondermann, C. Rong, *et al.*, “An orthogonal T7 replisome for continuous hypermutation and accelerated evolution in *E. coli*,” *Science*, vol. 389, no. 6760, pp. 618–622, Aug. 2025, Publisher: American Association for the Advancement of Science. doi: 10.1126/science.adp9583. (visited on 09/19/2025).
- [26] S. M. Miller, T. Wang, and D. R. Liu, “Phage-assisted continuous and non-continuous evolution,” *Nature Protocols*, vol. 15, no. 12, pp. 4101–4127, 2020. doi: 10.1038/s41596-020-00410-3.

- [27] A. A. Mengiste, J. L. McDonald, M. T. Nguyen Tran, *et al.*, “MutaT7GDE: A Single Chimera for the Targeted, Balanced, Efficient, and Processive Installation of All Possible Transition Mutations In Vivo,” *ACS Synthetic Biology*, vol. 13, no. 9, pp. 2693–2701, Sep. 2024, Publisher: American Chemical Society. doi: 10.1021/acssynbio.4c00316. (visited on 04/30/2025).
- [28] R. Tian, F. B. H. Rehm, D. Czernecki, *et al.*, “Establishing a synthetic orthogonal replication system enables accelerated evolution in *E. coli*,” *Science*, vol. 383, no. 6681, pp. 421–426, Jan. 2024, Publisher: American Association for the Advancement of Science. doi: 10.1126/science.adk1281. (visited on 05/21/2025).
- [29] M. Haas and B. Rak, “Escherichia coli insertion sequence is150: Transposition via circular and linear intermediates,” *Journal of Bacteriology*, vol. 184, no. 21, pp. 5833–5841, 2002. doi: 10.1128/JB.184.21.5833-5841.2002.
- [30] S. I. Jensen, R. M. Lennen, M. J. Herrgård, and A. T. Nielsen, “Seven gene deletions in seven days: Fast generation of *escherichia coli* strains tolerant to acetate and osmotic stress,” *Scientific Reports*, vol. 5, no. 17874, 2016. doi: 10.1038/srep17874.
- [31] A. J. Meyer, T. H. Segall-Shapiro, E. Glassey, J. Zhang, and C. A. Voigt, “Escherichia coli ‘Marionette’ strains with 12 highly optimized small-molecule sensors,” in, *Nature Chemical Biology*, vol. 15, no. 2, pp. 196–204, Feb. 2019, Publisher: Nature Publishing Group, ISSN: 1552-4469. doi: 10.1038/s41589-018-0168-3. (visited on 04/12/2025).
- [32] C.-H. Yan, F.-H. Chen, Y.-L. Yang, *et al.*, “The transcription factor CsgD contributes to engineered *escherichia coli* resistance by regulating biofilm formation and stress responses,” *International Journal of Molecular Sciences*, vol. 24, no. 18, p. 13681, Sep. 5, 2023, ISSN: 1422-0067. doi: 10.3390/ijms241813681. (visited on 09/29/2025).
- [33] S. J. Moore, H.-E. Lai, R. J. R. Kelwick, *et al.*, “EcoFlex: A Multifunctional MoClo Kit for *E. coli* Synthetic Biology,” *ACS Synthetic Biology*, vol. 5, no. 10, pp. 1059–1069, Oct. 2016, Publisher: American Chemical Society. doi: 10.1021/acssynbio.6b00031. (visited on 04/13/2025).
- [34] I. P. Witte, G. D. Lampe, S. Eitzinger, *et al.*, “Programmable gene insertion in human cells with a laboratory-evolved CRISPR-associated transposase,” *Science*, vol. 388, no. 6748, eadtl5199, May 2025, Epub 2025 May 15. doi: 10.1126/science.adt5199.
- [35] X. ; Zhang, Y. ; Li, X. ; Wang, *et al.*, “Multi-source interactive stair attention for remote sensing image captioning,” *Remote Sensing 2023, Vol. 15, Page 579*, vol. 15, p. 579, 3 Jan. 2023, ISSN: 2072-4292. doi: 10.3390/RS15030579.
- [36] L. Frei, B. Gao, J. Han, *et al.*, “Deep mutational learning for the selection of therapeutic antibodies resistant to the evolution of omicron variants of sars-cov-2,” *Nature Biomedical Engineering*, vol. 9, no. 4, pp. 552–565, Apr. 2025, ISSN: 2157-846X. doi: 10.1038/s41551-025-01353-4.
- [37] F.-Z. Li, J. Yang, K. E. Johnston, E. Gürsoy, Y. Yue, and F. H. Arnold, “Evaluation of machine learning-assisted directed evolution across diverse combinatorial landscapes,” *Cell Systems*, vol. 16, no. 9, p. 103187, Sep. 2025. doi: 10.1016/j.cels.2025.103187.
- [38] F.-Z. Li, J. Yang, K. E. Johnston, E. Gürsoy, Y. Yue, and F. H. Arnold, “Evaluation of machine learning-assisted directed evolution across diverse combinatorial landscapes,” *Cell Systems*, vol. 16, no. 9, p. 101387, Sep. 2025. doi: 10.1016/j.cels.2025.101387.

- [39] L. Sellés Vidal, M. Isalan, J. T. Heap, and R. Ledesma-Amaro, “A primer to directed evolution: Current methodologies and future directions,” *RSC Chemical Biology*, vol. 4, no. 4, pp. 271–291, 2023. doi: 10.1039/D2CB00231K.

## 6 Supplementary Materials

### 6.1 Side project: Deep Mutational Learning of bridge recombinases

#### 6.1.1 Introduction

DE offers a powerful strategy to increase the fitness of proteins [39]. While most approaches are restricted to finding optima within local regions of the fitness landscape, machine learning offers an opportunity to broadly explore the fitness landscape and capture epistatic effects [38]. Combining ML-based exploration to sample functional sequence space with highly efficient DE methods to refine and optimize variants toward local optima represents a powerful and complementary approach for protein engineering.

Previous work has performed a deep mutational scan of the ISCr04 recombinase and demonstrated that the C-terminal region of the TnP domain is especially tolerant to mutations [18]. Building on this insight, we set the goal of designing and screening an unbiased library that samples a distribution of mutations across the majority of the TnP domain rather than being restricted to single amino acid substitutions. The library is evaluated in an inversion-based functional assay, where successful recombination events are encoded in a DNA sequence motif. Therefore, the variant and its inversion activity are both encoded on the DNA sequence and can be identified NGS.

The resulting labelled dataset can be used to train machine learning models that capture the underlying fitness landscape, including epistatic effects extending beyond single mutations, hence the term Deep Mutational Learning (DML) [36]. Through this approach, we aim to develop a computational tool for broadly sampling functional sequences from the fitness landscape. These sequences can then be further optimized towards their local optima using the evolution logics described above.

#### 6.1.2 Results

We aim to couple an unbiased ISCr04 variant library (genotype) with its corresponding functional states (phenotype) directly at the DNA level through our assay design. This strategy generates a genotype–phenotype–linked dataset via NGS. On this basis, a machine learning model can be trained to capture epistatic interactions among multiple mutations and to broadly explore the underlying fitness landscape. From the model’s latent space, we can then sample variants predicted to exhibit enhanced performance (Figure 11A).

Our sequence-based functional assay encodes the ISCr04 variant library, the corresponding ISCr04 bRNA, target Sequence, and the IS621 wild-type donor sequence on a single plasmid, pES071.2 (Figure 11B). The bRNA recognizes both the target and donor sequences, which are essential for mediating inversion. The donor sequence is located in close proximity to the ISCr04 library variants. This spatial arrangement ensures that the presence or absence of an inversion event can be directly linked to the sequence of the corresponding variant. This design further enables direct assessment of both genotype and phenotype with high-quality short-read sequencing.

A mutagenesis library covering the majority of the TnP domain of ISCr04 (amino acids 226 - 326) was con-

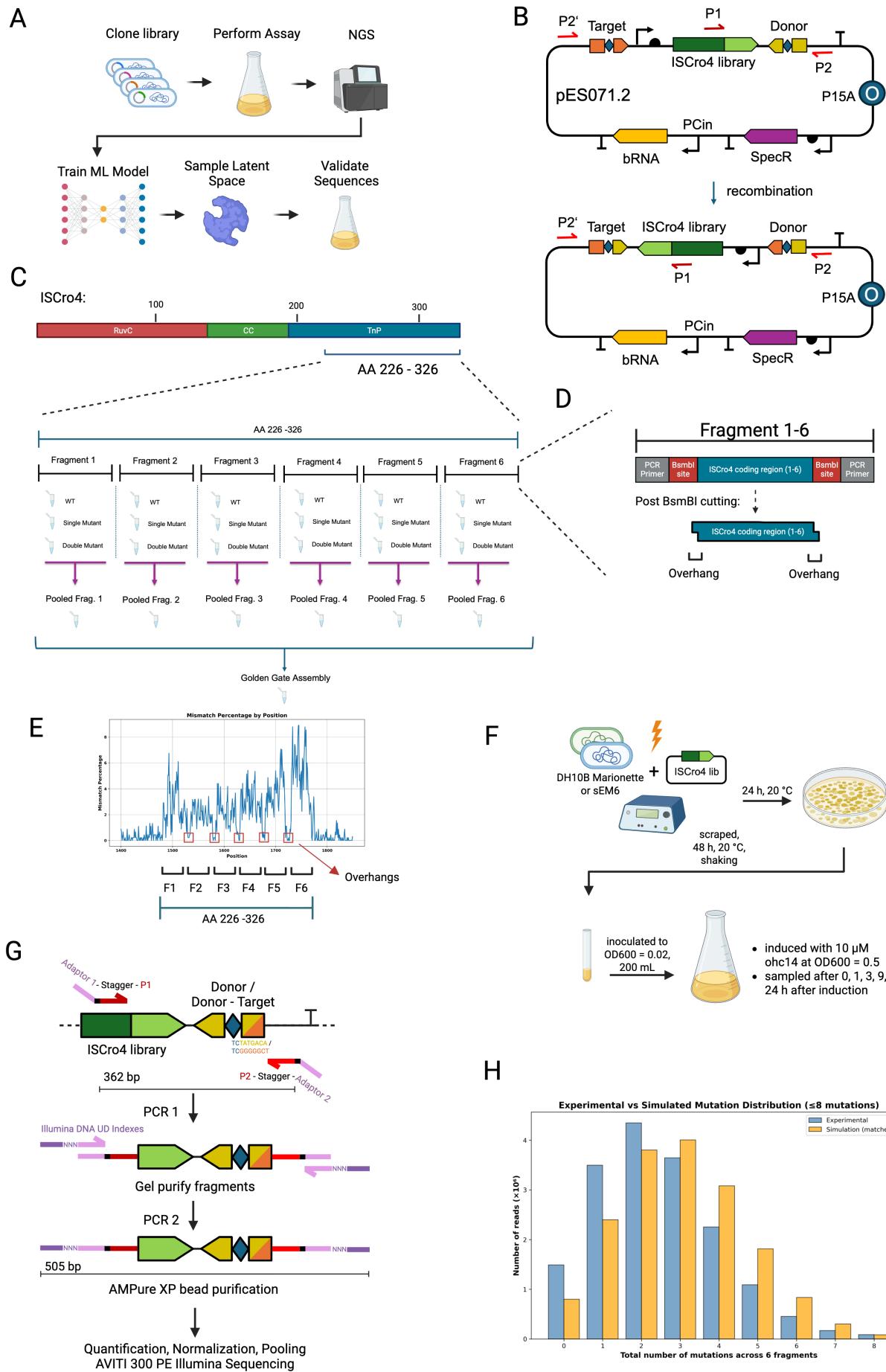
structed. The mutagenized region was divided into six fragments, and for each fragment three oligo pools were synthesized: one containing only wild type sequences, one containing all possible single-point mutations, and one containing all possible double-point mutations (Figure 11 C). These pools were PCR amplified and then mixed at a ratio of 60% wild type, 30% single mutant, and 10% double mutant to yield a final pool per fragment. The resulting six pooled fragments were then combined in a Golden Gate cloning reaction to assemble the library (Figure 11 D). This library assembly approach and mixing ratio enables a broad distribution of mutations across all variants, thereby increasing the likelihood of capturing epistatic effects with the ML model [37].

Successful assembly of the mutagenesis library was confirmed by long-read Oxford Nanopore sequencing (Figure 11E). The mismatch percentage relative to the wild type sequence was plotted across mutagenized region of the ISCr04. As expected, the regions targeted for mutagenesis displayed a marked increase in mismatch frequency (position 1480 - 1780), confirming the introduction of mutations into the library. In contrast, the cut site overhangs, where no mutations were introduced, showed much lower mismatch percentages, consistent with the preservation of the wild type sequence. This pattern validates that the designed variant space was specifically incorporated at the intended positions while non-targeted regions remained largely unaltered.

Next, we transformed *E. coli* Marionette-Clo and sEM6 strains with pES071.2, scraped and cultured the bacteria (Figure 11F). After induction with OHC14 the bRNA is transcribed and recombination can occur. The ISCr04-bRNA complex inverts the DNA sequence flanked by the target and donor sequences right after their CT core. This includes the ISCr04 gene.

After execution of the functional assay, a NGS library preparation was performed to recover the screened variants and their functional states. Therefore, primers with Nextera Transposase Adaptors anneal to the plasmid to amplify the library fragment for subsequent sequencing (as indicated in Figure 11G). Index Primer 1 (P1) anneals in the ISCr04 gene right before the start of the library fragments. Index Primer 2 anneals twice on the plasmid, before the target sequence (P2') and after the donor sequence (P2) on the reverse stand (as indicated in Figure 11H). The PCR results in a 362 bp long product, regardless of whether the plasmid sequence was the original, P1 - P2, or inverted, P1 - P2' (Figure 11 H). The PCR products of the original or inverted plasmids differ by the first 7 bp before the CT core in the donor or target region, as these nucleotides are not affected by the inversion. Therefore, these PCR products serve as binary markers, encoded on the DNA sequence, indicating whether an inversion has occurred. The PCR product is amplified a second time using Illumina DNA / RNA UD Indexes, gel and bead purified and submitted for Illumina sequencing. A representative fragment analyser quality control trace is shown in Figure 11I, with the main peak corresponding to the expected fragment length and the appropriate concentration for flow cell loading.

Sequencing across all submitted samples yielded approximately 300 million reads, confirming successful library preparation. Analysis of the mutation distribution across all reads for the uninduced sample revealed that the majority of assemblies contained two to three mutations, with fewer variants carrying four or more mutations (Figure 11 H). Only a small fraction of reads contained zero or more than six mutations. This pattern closely matches the expected distribution derived from *in silico* simulations of the fragment pooling ratios (60% wild type, 30% single mutant, 10% double mutant). When comparing experimental sequencing data to the simulated distribution, both showed a peak around two to three mutations per variant, confirming that the designed mixing ratios were faithfully represented in the assembled library. The remaining samples are currently being processed at the Genomics Facility Basel and will be analyzed to generate the training dataset for machine learning.



**Figure 11: A genotype-phenotype coupled IS<sub>C</sub>ro4 library for Deep Mutational Learning.** **A** Workflow of Deep Mutational Learning. **B** pES071.2 plasmid scheme including the IS<sub>C</sub>ro4 library, target and Donor sequence, and bRNA elements before and after successful inversion. **C** The mutagenized region (AA 226–326) in the TnP domain of IS<sub>C</sub>ro4 was divided into six fragments; for each fragment, wild type (WT), single-mutant, and double-mutant variants were synthesized and pooled to yield one pool per fragment. The library was assembled by Golden Gate into pES071.2, combining the six pooled fragments with the constant part of IS622. **D** Each fragment was flanked by BsmBI recognition sites and PCR primer binding sites to enable amplification and cloning. **E** Per-position mismatch rate after cloning and long-read sequencing (Oxford Nanopore Technologies). **F** Assay scheme describing the transformation and cultivation of the Marionette DH10a and SEM6 strain. **G** Logic of the Illumina Sequencing library preparation using defined binding sites in on the plasmid, P1 & P2, Nextera Transposase Adapters, and Illumina DNA / RNA UD Indexes. **H** Experimental (NGS) versus simulated distribution of the total number of mutations across the six fragments after library assembly.

### 6.1.3 Discussion / Outlook

In this project, we established a method to link the genotype of IS<sub>C</sub>ro4 variants with their functional phenotype through a sequence-based inversion assay. Our approach builds on previous deep mutational scanning efforts, which demonstrated mutational tolerance in the C-terminal region of the TnP domain but remained limited to single amino acid substitutions [18].

The library design, combining wild type, single-mutant, and double-mutant pools at defined ratios for each fragment, resulted in a broad distribution of variants, with the majority containing two to three mutations. Oxford Nanopore sequencing revealed higher mismatch percentages in Fragments 1, 5, and 6, which is due to a greater incorporation of double-mutant fragments at these positions. Overall, the balance between diversity and functionality is important: while higher diversity increases the probability of capturing epistatic effects, retaining a sufficient proportion of functional sequences ensures that machine learning models receive meaningful training data. Importantly, the library was designed to generate an optimal dataset for training machine learning models rather than to directly identify the most promising IS<sub>C</sub>ro4 variants. We consider this distinction crucial for the successful application of ML-guided DE. While the assay design enables high-throughput screening of diverse libraries, the current implementation is limited to the TnP domain, reflecting read-length constraints of short-read sequencing technologies.

The NGS results demonstrate that our library assembly strategy successfully generated the intended distribution of variants, with experimental sequencing closely matching *in silico* simulations. The observed enrichment for two to three mutations per variant confirms that the fragment pooling approach reliably controls mutational load while maintaining sufficient diversity. Although only the uninduced sample was sequenced, these findings validate the underlying concept of library assembly and provide a solid foundation for subsequent functional screening and machine learning model training once the remaining samples are available.

The choice of machine learning architecture will then need to be evaluated. In the optimal scenario, an encoder-decoder model could be trained to learn a meaningful latent space representation of IS<sub>C</sub>ro4 variants. Sampling from such a latent space would in turn enable the generation of functional sequences that cover different areas of the fitness landscape. These sampled variants can be evolved to their local fitness optima. In the event that screening the unbiased library does not provide sufficient information to train a model capable

of learning a meaningful latent space, a second, biased library can be designed and screened. Such a library could be guided by the mutational tolerance patterns revealed by the previous DMS [18], thereby enriching the dataset with functionally relevant variants.

By coupling genotype and phenotype in a pooled, sequencing-based assay, this framework enables the high-throughput generation of training data and lays the foundation for ML-guided DE of bridge recombinases.

#### 6.1.4 Methods

##### Library Design and Assembly

Each fragment was flanked by primer-binding sites for amplification and *BsmBI* restriction sites for cloning. Primers were designed using the Primer3Web tool (<https://primer3.ut.ee/>). Overhang sequences were optimized using the New England Biolabs ligase fidelity viewer (<https://ligasefidelity.neb.com/viewset/run.cgi>). Amino acids that overlapped with the restriction enzyme cut site were excluded from mutagenesis.

Fragments were designed as synthetic single-stranded oligodeoxynucleotides (ssODNs) and ordered from IDT Technologies as oPools (<https://eu.idtdna.com/page>). For each of the six fragments, three separate pools were synthesized: one containing the wild type sequence, one containing all possible single-point mutations, and one containing all possible double-point mutations, resulting in a total of 18 pools. Mutations were introduced through NNK codons.

PCR amplification was performed using the Phusion™ High-Fidelity DNA Polymerase (NEB #M0530) according to the manufacturer's protocol, with 28 amplification cycles. The pooled fragments were then assembled using the NEBridge® Golden Gate Assembly Kit (*BsmBI*-v2) following the supplied protocol.

##### Transformation and cultivation of *E. coli* during the Efficiency Assay

Electrocompetent *E. coli* DH10B Marionette-Clo and sEM6 cells were thawed on ice, 5 µl of the Golden Gate Assembly was mixed with 100 µl of each cell strain and incubated for one minute. The mixtures were transferred to an electroporation cuvette, loaded into the BIO-RAD MicroPulser Electroporator and electroporated using the setting "Ec1" (~1.4 Volt). The transformed cells were recovered for one hour in SOC medium. The bacteria were then plated on LB plates containing 120 µg/ml of Spectinomycin and incubated at 37 °C for 24 hours. Colonies were counted and scraped from the plate to inoculate 2 ml LB cultures with 120 µg/ml of Spectinomycin. Successful cloning was verified by long-read Oxford Nanopore Technology sequencing (Service from Microsynth AG Switzerland). The liquid cultures were incubated at 20°C shaking for 48 hours.

200 ml of LB culture + 120 µg/ml spectinomycin was inoculated by each of the 2 ml culture to OD<sub>600</sub>=0.01 and incubated at 37 °C, 220 rpm until they reached OD<sub>600</sub>=0.5. 50 mL serving as 0-hour time point were collected and frozen. Then, cultures were induced with 10µM N-(3-Hydroxytetradecanoyl)-DL-homoserine lactone (OHC14, 51481, Sigma-Aldrich). Samples were collected after 1 h, 3 h, 9 h, and 24 h using the following calculation:  $V_{sample} = \frac{50ml * 0.5}{OD600_{sample}}$ . All samples were thawed simultaneously and the plasmids were purified using the Zippy Plasmid Miniprep Kit (Zymo Research, D4019).

## Illumina Sequencing Library Preparation

PCR amplification was performed with Q5™ High-Fidelity 2X Master Mix (M0492, NEB). Notably, the reaction mix contained a final primer concentration of 2  $\mu$ M and about 5 ng of template plasmid. The thermocycler conditions were kept at manufacturer's instructions, the melting temperature was calculated using the NEB Tm Calculator (<https://tmcalculator.neb.com/>), and the annealing and extension time was set to 5 and 8 seconds, respectively. The amplification was run for 20 cycles. The PCR products were loaded on a 1.5 % agarose gel and DNA was separated by gel electrophoresis. DNA of the expected size was cut out and gel purified using the Zymoclean Gel DNA Recovery kit (D4007).

Cleaned PCR products were used for the second PCR reaction. Again, Q5™ High-Fidelity 2X Master Mix was used with Illumina™ DNA / RNA UD Indexes set A (UDP0017 - UDP0026). Cycler conditions were used according to the Illumina DNA Prep recommendations. The final PCR product was purified using AMPure XP bead and following manufacturer's instructions. Finally, samples were diluted to 10 nM using Qubit™ Fluorometric Quantification.

Samples were submitted to the Genomics Facility Basel for Element Biosciences Aviti 300 bp paired end sequencing.

### 6.2 Directed Evolution of IS621 using EcORep

Since we were unable to obtain the EcORep strain, we set out to engineer an adapted version of EcORep. Our goal was to create a system that carries all the genes necessary for stable replication of the O-Replicon on a single plasmid. The original system works as follows: A synthetic replication operon containing the PRD1 genes TP (pVIII), DNAP (pI), SSB (pXII) and DSB (pXIX) is integrated in the *E. coli* strain DH10B. The operon is under the control of the IPTG-responsive Ptac promoter. To establish the O-Replicon, the strain is first transformed with an additional "helper plasmid" (pFR160GB) expressing SSB, DSB and Gam. The O-Replicon is then electroporated into this strain. The helper plasmid can then be cured from the strain. To expand and mutagenize the O-Replicon, the genomically integrated cassette is repressed using a plasmid (pRT19) carrying dCas9 and gRNA targeted to the IPTG promoter. An additional plasmid expressing the PRD1 genes including a mutagenic version of DNAP (N71D), is expressed from another plasmid (pRT4-2). In our simplified design (pES084.2, 17), all genes necessary for replication of the O-Replicon are carried on a low copy plasmid with the p15A ori. The TP and mutagenic DNAP-N71D are under the control of the PCymRC promoter optimized by Meyer *et al.*[31] This inducible promoter shows a great dynamic range and low leakiness. Expression in the induced state is 870 fold higher. Furthermore, we tried to optimize the expression levels of SSB, DSB, and Gam by creating a promoter and RBS library for each of the genes. This plasmid-borne system can easily be transformed. The reduced number of genetic components should also decrease the metabolic burden of carrying the system.

#### 6.2.1 Assembly of Plasmids for the EcORep Subproject

All plasmids were assembled using golden gate cloning. An adapted version of EcoFlex [33] was used. Restriction enzymes and Bridge Ligase MM were purchased form NEB.

### 6.2.2 Images of plates used for selection validation

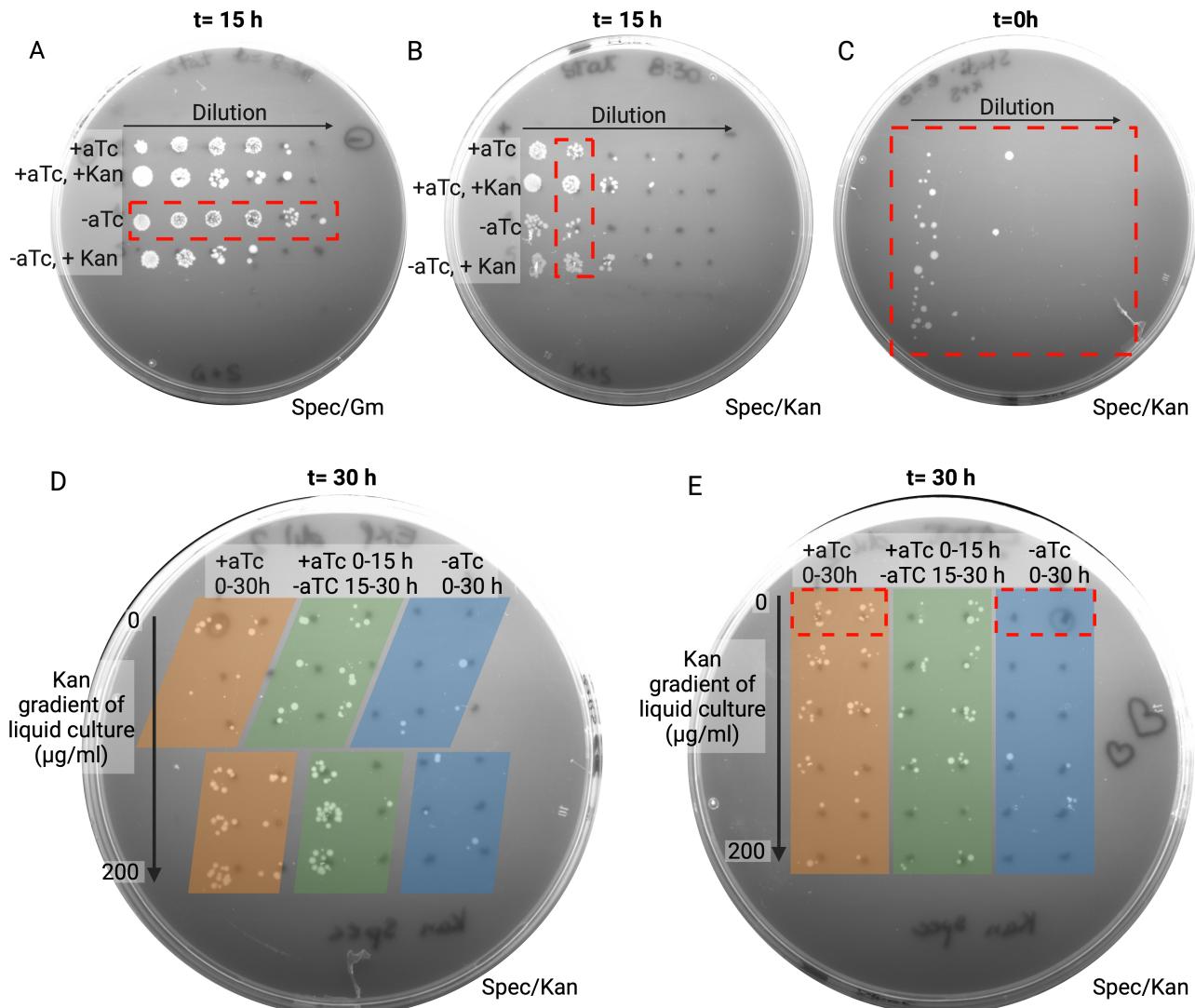


Figure 12: pES013.2A and pES055.2 co-transformed SEM6 bacteria plated on double antibiotic plates after growth in liquid culture. **A** Droplet plating of a serial dilution of bacteria on Spec/Gm plate after 15 h of growth in four induction and selection conditions. **B** Droplet plating of a serial dilution of bacteria on Spec/Gm plate after 15 h of growth in four induction and selection conditions. **C** Serial dilution on a Spec/Kan plate before induction with aTc. **D** Growth of 33 different liquid culture conditions 30 h after induction at  $\text{OD}_{600} = 0.1$  plated as droplets on a Spec/Kan plate. **E** Growth of 33 different liquid culture conditions 30 h after induction at  $\text{OD}_{600} = 0.6$  plated as droplets on a Spec/Kan plate. Apart from **D** all cultures were induced at  $\text{OD}_{600} = 0.6$ , **D** was induced at  $\text{OD}_{600} = 0.1$ . Red boxes indicate cut outs of the plates shown in Figure 4. Orange, green and blue boxes indicate that the cultures have the same induction conditions. Orange induced for 30 h, green for the first 15 h and blue uninduced.

### 6.3 Generation and validation of sEM6

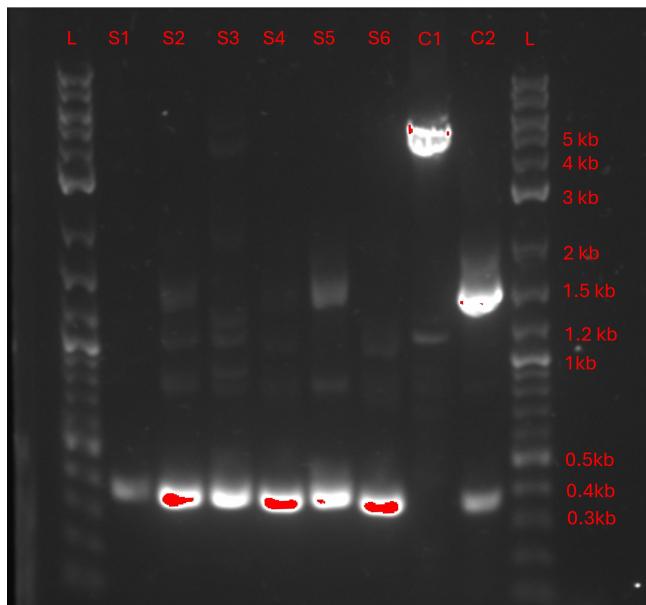


Figure 13: 1% agarose gel of a colony PCR done with the csg primers (see Table2). S1-S6 as described in methods. Lanes L are ladders made with NEB 1kb plus. C1 is Marionette-Clo. C2 is a sample before the induction of the fliptase. The amplicon size for the csg primer pair in Marionette-Clo is 4.8 kb. This matches well with lane C1. Validating that the primers work. The amplicon size for Marionette-Clo with the integrated dsDNA substrate is 1.5 kb. This is seen in C2, thus confirming the successful integration of the dsDNA substrate. After the resistance cassette is flipped out, the amplicon is expected to be roughly 400 bp. This can be observed in S1-S6. Meaning that the deletion has successfully worked.

### 6.4 Plasmids designed within the project

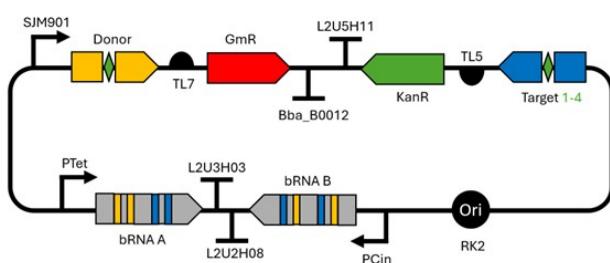


Figure 14: *Schematic of pES013.2A-D*. The illustration above shows the selection plasmid. There are four versions of each target and bRNA.

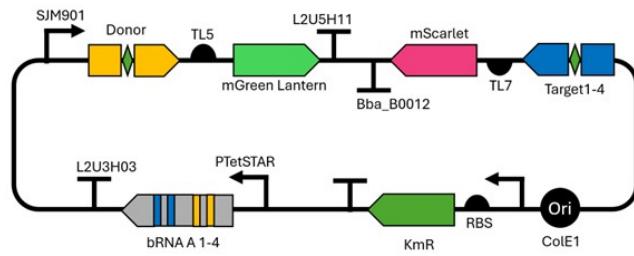


Figure 15: *Schematic of pES049.2A-D.* The illustration above shows the plasmid used to detect IS621 activity in the flow cytometry based activity assay.

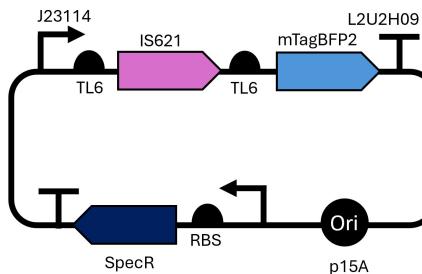


Figure 16: *Schematic of pES055.2.* The illustration above shows the plasmid used for expression of IS621 during the floe cytometry based and plating based activity assay.

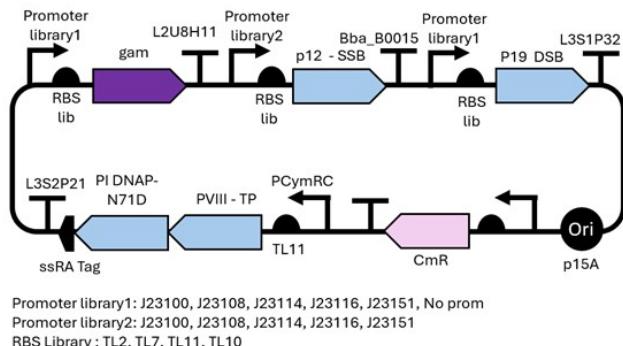


Figure 17: *Schematic of pES084.2.* The illustration above shows the plasmid necessary for replication of the O-Replicon

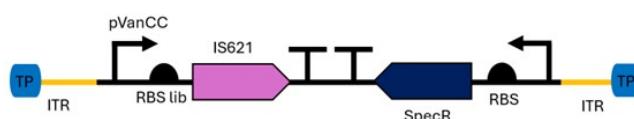


Figure 18: *Schematic of pES067.2.* The illustration above shows O-Replicon carrying IS621

## 6.5 Primers used in the project

All primers were obtained from Microsynth AG, Germany.

Oligo ID	Description	Sequence
ES_4F	Inversion of a TU1rb site in backbone plasmid	GCT TCT AGT CAC CTG CAT ATA AGG AAC ATG AGA CCT TGC TAC ACG CCT GC
ES_4R	Inversion of a TU1rb site in backbone plasmid	CGC CAC CAC CTG CAT ATC AGT CTA TTG AGA CCG ATC AGG TGC GTG G
ES_1F	PCR amplification of KanR gene from a Twist fragment	CACCTGCaCTaatctggctctGTCGTTATTAGAAAAATTCCATCCAGCA-GACG
ES_1R	PCR amplification of KanR gene from a Twist fragment	CACCTGCCATtctaaGGTCTCtCATATGATTGAACAGGATGGCCTG
ES_5F	Inversion of a TU1rc site in backbone plasmid	GCT TCT AGT CAC CTG CAT ATA CTG AAC ATG AGA CCT TGC TAC ACG CCT GC
ES_5R	Inversion of a TU1rc site in backbone plasmid	CGC CAC CAC CTG CAT ATT CCT CTA TTG AGA CCG ATC AGG TGC GTG G
ES_8F	PCR amplification of SJM901-Donor sequence	aggctcaCTATTTACAGCTAGCTCAGTCCTAGGTATAATGCTAGCG
ES_8R	PCR amplification of SJM901-Donor sequence	tggctctAAAGTATACAAGATACTGTACGCTAGCATTATACCTAGGACTG
RK2Ori_F	qPCR primer	GGG CTA TCG CTT TTC CGT
ES_9F	PCR amplification of target 1	AGG TCT CTC TAT AGG GTG TCT AGG TCA C
ES_9R	PCR amplification of target 1	TGGTCTCtAAAGTGACCTAGACACCCCTATAAGaG
ES_10F	PCR amplification of target 2	AGG TCT CTC TAT TCG GGG GCT CCT GTA C
ES_10R	PCR amplification of target 2	TGGTCTCtAAAGTACAGGAGCCCCGAATAG
ES_11F	PCR amplification of target 3	AGG TCT CTC TAT GCA GGA CCT AGA TGA C
ES_11R	PCR amplification of target 3	TGGTCTCtAAGTCATCTAGGTCCCTGCATAG
ES_12F	PCR amplification of target 4	AGG TCT CTC TAT ATA ATC CCT AGG CAA CTT AGA G
ES_12R	PCR amplification of target 4	TGGTCTCtAAGTTGCCTAGGGATTATATAG
SSB MoClo F	PCR amplification of SSB gene from a Twist fragment	AAC GTC TCA ATC TAG GTC TCA CAT ATG GAA ATC GTA AGC AAG CTG ACT CTG AAA ACC
SSB MoClo R	PCR amplification of SSB gene from a Twist fragment	TTC GTC TCG CTA ATG GTC TCT GTC GCC GTG GTG CGC TGT TTA GGC TTC G
DSB MoClo F	PCR amplification of DSB gene from a Twist fragment	AAC GTC TCA ATC TAG GTC TCA CAT ATG GAA AAG CAA ACC GAA AAT ACC CGC CCT G
DSB MoClo R	PCR amplification of DSB gene from a Twist fragment	TTC GTC TCG CTA ATG GTC TCT GTC GCC GCA AGG GAA TCA TGC GGC CTT C

Table 1: Primers used in the project

Oligo ID	Description	Sequence
Gam MoClo F	PCR amplification of Gam gene from a Twist fragment	AAC GTC TCA ATC TAG GTC TCA CAT ATG GAT ATT AAT ACT GAA ACT GAG
Gam MoClo R	PCR amplification of Gam gene from a Twist fragment	TTC GTC TCG CTA ATG GTC TCT GTC GTT TTG TGT TTG CCC TTT CGA TTT ATA CCT CAC
H-Arginase MoClo F	PCR amplification of H-Arginase gene from a Twist fragment	AAC GTC TCA ATC TAG GTC TCA AGT GCA AAA AGC CGT ACC ATT GGT ATT ATT GG
H-Arginase MoClo R	PCR amplification of H-Arginase gene from a Twist fragment	TTC GTC TCG CTA AAG GTC TCT CGG TTT TCG GAG GGT TCA GAT AAT CAA TCG G
CcdB MoClo F	PCR amplification of CcdB gene from a Twist fragment	AAC GTC TCA ATC TAG GTC TCA AGT GGT ATG CAG TTT AAG GTT TAC ACC TAT AAA AGA GAG
CcdB MoClo R	PCR amplification of CcdB gene from a Twist fragment	TTC GTC TCG CTA AAG GTC TCT CGG TTA TTC CCC AGA ACA TCA GGT TAA TGG C-
pES_13F	PCR amplification of target 1-SJM902	AGGTCTCtCTATAGGGTGTCTAGGTCGCTAGCATAATCCCTAGG
pES_13R	PCR amplification of target 2-SJM902	TGGTCTCtAAGTTTACAGCTAGCTCAGTCTAGGGATTATGCTAGC-GACC
pES_14F	PCR amplification of target 2-SJM902	AGGTCTCtCTATTGGGGCTCTGTGCTAGCATAATCCCTAGG
pES_14R	PCR amplification of target 2-SJM902	TGGTCTCtAAGTTTACAGCTAGCTCAGTCTAGGGATTATGCTAG-CACAGG
DML_Lib_RV	Library fragment amplification	AGC ATG GAG TCA CAC AGT GC
csg_fwd	Forward primer used in the colony PCR	AAAAACAGTGGCCTGTGGCTGGCGG
csg_rev	Reverse primer used in the colony PCR	AAATGGATCACGCCCTTGACCG
oES567_Recombination_primer_F	To PCR amplify over the insertion junction	CAGCTAACACCACGTGTCCTATC
oES568_Recombination_primer_R	To PCR amplify over the insertion junction	GTTACCAGAGTCGGCCAAGGAAC
oES550_rpoZ_SP_IS621_ins_R	IS621 ins PCR: Cloning IS622 and IS621 into rpoZ SP BB (second SP cloning strategy)	GTCAGTCAGGTCTCaaTTACGCCGCTACCGGATTATGCCG
oES558_SP_cloning_test_F	test SP cloning success	cgtggcgcgttgttttg
oES559_SP_cloning_test_R	test SP cloning success	CAGCTTAGCTTGCTAACGTCGATG
oES560_DtoD_SDm_gIII_F	pES504 for DML cloning	GATAACAGTATCTTGTATTATCCCAAGGAGGAAAAAAAAAAT-Gaaaaattatttcgc

Table 1: Primers used in the project

Oligo ID	Description	Sequence
pES_15F	PCR amplification of target 3-SJM902	AGGTCTCtCTATGCAGGACCTAGATGGCTAGCATAATCCCTAGGACT-GAGC
pES_15R	PCR amplification of target 3-SJM902	TGGTCTCtAAGTTTACAGCTAGCTCAGTCCTAGGGATTATGCTAGCC
pES_16F	PCR amplification of target 4-SJM902	AGGTCTCtCTATATAATCCCTAGGCAGCTAGCATAATCCCTAGG
pES_16R	PCR amplification of target 4-SJM902	TGGTCTCtAAGTTTACAGCTAGCTCAGTCCTAGGGATTATGCTAGGCTAG
pES_17F	Illumina primer target 2	GGTCTCtCTATCTAACGCTCAGACAGTGGTGCTCG
pES_17R	Illumina primer target 2	TGGTCTCtAAACACAGGAGCCCCGAGCACCACTGTCTGAGG
pES_18F	Illumina primer target 1	GGTCTCtCTATCTAACGCTCAGACAGTGGTGCTCG
pES_18R	Illumina primer target 1	TGGTCTCtAACAGACCTAGACACCCTGCACCACTGTCTGAGG
pES_19F	Illumina primer target 3	GGTCTCtCTATCTAACGCTCAGACAGTGGTGCGCAGG
pES_19R	Illumina primer target 3	TGGTCTCtAACACATCTAGGCCTGCGCACCACTGTCTGAGG
pES_20F	Illumina primer target 4	CCTCAGACAGTGGTGCTAAATCC
pES_20R	Illumina primer target 4	TGGTCTCtAACATGCCTAGGGATTATGCACCACTGTCTGAGG
pES_21F	No-RBS-No-CDS	aggctcaACTTAGCAACAAACGATCGTACGCACCATCCATTGTCG
pES_21R	No-RBS-No-CDS	tggctctGTCGAATCCGAGCGTCACCACGACAATGGATGGTGCCTACG
pES_22F	No-CDS	aggctcaCATATAGCAACAAACGATCGTACGCACCATCCATTGTCG
pES_22R_SalI	No-CDS	tggctctGTCGAATCCGAGCGTCACGACAATGGATGGTGCCTACG
ES_23F	Filler 1	aacgtctcaTCACctgcataatAGGAACATGCATAGATTATCAGAGTGTGAGC
ES_23R	Filler 1	tttgtctcgACCAccctgcataatCTAAGAGCTCACACTCTGATAATCTATGC
ES_24F	Gibson-O-Replicon-Amp	AAG AAG ACT AAT CTT CAC CTG CAT ATA TCT CCC TTT GTG C
ES_24R	Gibson-O-Replicon-Amp	ATG AAG ACT ACT AAT CAC CTG CAT ATC CTT AAC ACT CAA C
ES_25F	Gibson-O-Replicon-Backbone	ccacggagtgactaattctgtaaccattctgaggacaactctaggcattactgg
ES_25R	Gibson-O-Replicon-Backbone	cttttattttgttatccccaaaggcgcaatgtttactactttcccgaaatc
FW_IS622_Reconst.	Sequencing primer	CGA TAT GCA CAC CCA GGA AC
oES564_DML_rec_BB_amp2_R	pES504 for DML cloning	aatatgttttcgtTtcagccaatccctggtagttcac
oES565_DML_rec_ins_amp3_F	pES504 for DML cloning	acggcgataactcaactctgacCATAATGGAACAGGAGCTGCATTCATCG-GCATCG
oES523_pES512_luxAB_R	LuxAB reporter without gIII	taagcaaaaagtccaaattcatTTTTTTCCCTCCTACTGTGACGTAGGCCTG
oES524_pES512_luxAB_MI_F	LuxAB reporter without gIII, mock integrated	aaaaagttccaaattcatTTTTTTCCCTCCTACTGTGACGTAGACTGT-TATCC
oES561_DML_rec_BB_amp1_F	pES504 for DML cloning	gattggctgaAacgaaaaacatattctcaataaaccctttaggaaatagg
oES562_DML_rec_BB_amp1_R	pES504 for DML cloning	TATGgtcagacttgagttacgcgtcc
oES563_DML_rec_BB_amp2_F	pES504 for DML cloning	CGACacttaataacggcactcctcagcc

Table 1: Primers used in the project

Oligo ID	Description	Sequence
oES501_pIII_luxAB_insert_Gib_F	pES507 and pES508 cloning (pIII/lux AB insert)	atgaaaaaaatttattcgcaattccttagttgtcc
oES502_pIII_luxAB_insert_Gib_R	pES507 and pES508 cloning (pIII/lux AB insert)	ttaggtatattccgtggtaactttaatattatcatcaac
oES566_DML_rec_ins_amp3_R	pES504 for DML cloning	gttaattaagtGTCGTTAGTTCAATTATGCCAGTTGGACG-GTAAGTCGC
oES525_cPCR_Kan_F	Colony PCR primers	ATGCATTTCTTCCAGACTTGTCAACAG
oES526_cPCR_Kan_R	Colony PCR primers	CGATTCCCTGTTGTAATTGTCCTTTAACAGC
oES527_cPCR_Spec_F	Colony PCR primers	acaacgtaaagcacatcgtcgc
oES528_cPCR_Spec_R	Colony PCR primers	cgttgccctggtaggtcag
oES503_MP6_BB_BbsI_F	pES504 cloning	CGACGGTCTCaCGACacttaacggcactcgcagcc
oES504_MP6_BB_BbsI_R	pES504 cloning	GTCAGTCAGGTCTCaTATGtcagacttgatgtacgcgtcc
oES505_pES501+2_ins_PaqCl_F	pES505 cloning from pES501	GTCAGTCTCACCTGCTataATCTCCCTTATCAGGCCTACGT-CACAGTAAG
oES506_pES501_ins_PaqCl_R	pES505 cloning from pES501	GTCAGTCTGTCAGTCTCACCTGCTataCTAATATAAACGCAGAAAGGC-CCACCC
oES507_pES502_ins_PaqCl_F	pES506 cloning from pES502	GTCAGTCTCACCTGCTataATCTCACAGCTAACACCACGTCGTCC
oES508_pES505+6_BB_Gib_F	pES507 and pES508 cloning (BB amplification for pIII/lux AB insert)	ttaagaagtaccacggaatataccaaCGACTCGATCACACTGG
oES509_pES505+6_BB_Gib_R	pES507 and pES508 cloning (BB amplification for pIII/lux AB insert)	taaaggaaatttgcgaaaataatttttcatTTTTTTTCCCTACTGTGACGTAG
oES510_junction_qPCR_F	qPCR over junction	GGATAACTTACGGGCATGCATAAG
oES511_junction_qPCR_R	qPCR over junction	cagcggagtggaaatagaaggaaac
oES512_BB_qPCR_F	qPCR of backbone	cggcggatccatagcgtaag
oES513_BB_qPCR_R	qPCR of backbone	cgacattgtcgatcttgcgt
oES514_NT_TBL_F	cloning pES509, CP with non-targeting bRNA (negative control)	ACAGGTTATTAGCCTGctggtagcaACGGCATGCGGAC
oES515_NT_TBL_R	cloning pES509, CP with non-targeting bRNA (negative control)	CTAATAACCTGTtccgactATTATGCAGCGGACCGCCGTT
oES516_S241A_F	IS622 S241A mutant	TTATGAGcgGGGTCGAGCGTCAATCGCGCCAGTC
oES517_S241A_R	IS622 S241A mutant	CGACCCcgcCTCATACGGCGTGGTGTCAAGACCGCGAAAG
oES518_SP_IS622_ins_F	IS622 PCRed from pES504 for ES_SP1	cgagaatgtttcttagtaaggaggaaaatgATGGAACAGGAGCTGCATTCATCGG
oES519_SP_IS622_ins_R	IS622 PCRed from pES504 for ES_SP1	ttaaccaggcttcagccaTTAGGCCGCTACAGGATTATGACGGGAAG
oES520_SP_IS621_ins_F	IS621 PCRed from pES511 for ES_SP2	cgagaatgtttcttagtaaggaggaaaatgATGGAACATGAACCTCATTATATCG-GTATCGACACCG
oES521_SP_IS621_ins_R	IS621 PCRed from pES511 for ES_SP2	ttaaccaggcttcagccaTTACGCCGCTACCGGATTATGCCGTG
oES522_pES512_luxAB_F	LuxAB reporter without gIII	atgaaatttggaaacttttgttatcacatccaacctcc

Table 1: Primers used in the project

Oligo ID	Description	Sequence
oES529_cPCR_Chlo_F	Colony PCR primers	ccgacatggaaaggccatcaaaac
oES530_cPCR_Chlo_R	Colony PCR primers	ggtagttccaccagtttattaaacgtg
oES531_SP_BbsI_F	for SP cloning – integrate recombinases with SapI overhang into pBP Bbs1 LacZ lvl0 plasmid	aacatatgGAAGACTtatctcgagaatgtcttcgttaag
oES532_SP_BbsI_R	for SP cloning – integrate recombinases with SapI overhang into pBP Bbs1 LacZ lvl0 plasmid	cgcattgcGAAGACTactaatttaaccaggcttcag
oES533_target_Gibson_CP_BB_F	BB PCR: Cloning SERPINA1 target TBLs into CP	CTGCTAGCATTATACCTAGGACTGAGCTAGCTGTCAAAGGCTAG
oES534_target_Gibson_CP_BB_R	Insert PCR: Cloning BREAT1-4 into CP by Gibson cloning	CAACATGCTTCGGCATGGCGAATGGGAC
oES535_enh_bRNAs_SERPINA_TBLs_F	Insert PCR: Cloning SERPINA1 target 1–4 TBL into CP	ttagacagcttagctcgttgcgtataatgcgtac
oES536_enh_bRNAs_SERPINA_TBLs_R	Insert PCR: Cloning SERPINA1 target 1 TBL into CP	gtccccattcgccatgcgaage
oES537_SDM_GFP_AP_target1_F	SDM PCR: Cloning SERPINA1 target 1 into GFP AP	CCCTTAGGGTGTCTAGGTACAGTAAGGAGGAAAAAAATGCG-TAAAGG
oES538_SDM_GFP_AP_target1_R	SDM PCR: Cloning SERPINA1 target 1 into GFP/gIII AP	TGTGACCTAGACACCCTAACGGAGATATAAGtgagacctctagaagc
oES539_SDM_GFP_AP_target2_F	SDM PCR: Cloning SERPINA1 target 2 into GFP AP	CCCTTCGGGGCTCTGTACAGTAAGGAGGAAAAAAATGCG-TAAAGG
oES540_SDM_GFP_AP_target2_R	SDM PCR: Cloning SERPINA1 target 2 into GFP/gIII AP	TGTACAGGAGCCCCGAAAGGGAGATATAAGtgagacctctagaagc
oES541_SDM_GFP_AP_target3_F	SDM PCR: Cloning SERPINA1 target 3 into GFP AP	CCCTGCAGGACCTAGATGACAGTAAGGAGGAAAAAAATGCG-TAAAGG
oES542_SDM_GFP_AP_target3_R	SDM PCR: Cloning SERPINA1 target 3 into GFP/gIII AP	TGTCATCTAGGTCTGCAAGGGAGATATAAGtgagacctctagaagc
oES543_SDM_GFP_AP_target4_F	SDM PCR: Cloning SERPINA1 target 4 into GFP AP	CCCTTATAATCCCTAGGCAACAGTAAGGAGGAAAAAAATGCG-TAAAGG
oES544_SDM_GFP_AP_target4_R	SDM PCR: Cloning SERPINA1 target 4 into GFP/gIII AP	TGTTGCCTAGGGATTATAAGGGAGATATAAGtgagacctctagaagc
oES545_rpoZ_SP_BB_F	BB PCR: Cloning IS622 and IS621 into rpoZ SP BB (second SP cloning strategy)	acttataaggctcATAAtctagaaggatttcaacatggcttagcacag
oES546_rpoZ_SP_BB_R	BB PCR: Cloning IS622 and IS621 into rpoZ SP BB (second SP cloning strategy)	tacatattGGTCTCCcattttccttactccaaaaaaaaggctccaaaggag
oES547_rpoZ_SP_IS622_ins_F	IS622 ins PCR: Cloning IS622 and IS621 into rpoZ SP BB (second SP cloning strategy)	GTCAGTCAGGTCTCaATGGAACAGGAGCTGCATTCATCGG
oES548_rpoZ_SP_IS622_ins_R	IS622 ins PCR: Cloning IS622 and IS621 into rpoZ SP BB (second SP cloning strategy)	GTCAGTCAGGTCTCaATTAGGCCGCTACAGGATTATGACGGGAAG
oES549_rpoZ_SP_IS621_ins_F	IS621 ins PCR: Cloning IS622 and IS621 into rpoZ SP BB (second SP cloning strategy)	GTCAGTCAGGTCTCaATGGAACATGAACCTCATTATATCGGTATCGA-CACCG

Table 1: Primers used in the project

Oligo ID	Description	Sequence
oES551_DtoD_GFP_AP_F	AP for Donor to donor GFP assay	GATAAACAGTATCTTGATTATCCCAAGGAGGAAAAAAAATGCG-TAAAGG
oES552_DtoD_GFP_AP_R	AP for Donor to donor GFP assay	GGGATAATACAAGATACTGTTATCAGATATAAGtgagaccctagaagcg
oES553_enh_bRNAs_SERPINA_TBLs_2_F	Insert PCR: Cloning BREAT1-4 into CP by Gibson cloning	GCTCAGTCCTAGGTATAATGCTAGCAGTCAGGGAGAACCGGCCAG
oES554_SDM_gIII_AP_target1_F	SDM PCR: Cloning SERPINA1 target 1 into gIII AP	CCCTTAGGGTGTCTAGGTACAGTAAGGAGGAAAAAAAAT-Gaaaaatttttcgc
oES555_SDM_gIII_AP_target2_F	SDM PCR: Cloning SERPINA1 target 2 into gIII AP	CCCTTCGGGGCTCCTGTACAGTAAGGAGGAAAAAAAAtgaaaaat-tatttcgc
oES556_SDM_gIII_AP_target3_F	SDM PCR: Cloning SERPINA1 target 3 into gIII AP	CCCTTGCAAGGACCTAGATGACAGTAAGGAGGAAAAAAAAT-Gaaaaatttttcgc
oES557_SDM_gIII_AP_target4_F	SDM PCR: Cloning SERPINA1 target 4 into gIII AP	CCCTTATAATCCCTAGGCAACAGTAAGGAGGAAAAAAAAT-Gaaaaatttttcgc

Table 1: Primers used in the project

Table 2: DNA substrate information for sEM6 strain