Deep-mutational scanning libraries using tiled-region exchange mutagenesis

Kortni Kindree[1,*], Claire A. Chochinov[1,*], Keerath Bhachu[1,*], Yunyi Cheng[2] , Amelia Caron[1], Molly McDonald[3], Zaynab Mamai[3], Alex N. Nguyen Ba[1,3,&]

[1]Department of Cell and Systems Biology, University of Toronto, Toronto ON Canada, M5S 3G5
[2]Department of Computer Science, University of Toronto, Toronto ON Canada, M5S 2E4
[3]Department of Biology, University of Toronto at Mississauga, Mississauga ON Canada, L5L 1C6

[*]These authors have contributed equally to this work
[&]alex.nguyenba@utoronto.ca

Running head: Tiled-region exchange mutagenesis
Keywords: Deep-mutational scanning, cloning

ABSTRACT

The analysis of gene function frequently requires the generation of mutants. Deep-mutational scanning (DMS) has emerged as a powerful tool to decipher important functional residues within genes and proteins. However, methods for performing DMS tend to be complex or laborious. Here, we introduce Tiled-Region Exchange (T-REx) Mutagenesis, which is a multiplexed modification of the EMPIRIC mutagenesis approach. Self-encoded removal fragments are cloned in parallel in non-overlapping gene locations and pooled. In a one-pot reaction, oligonucleotides are then swapped with their corresponding self-encoded removal fragments in bulk using a single Golden Gate reaction. To aid in downstream phenotyping, the library is then fused with unique DNA barcodes using the Bxb1 recombinase. We demonstrate this approach and its optimizations, to show that it is both easy to perform and efficient. This method offers simple and expedient means to create comprehensive mutagenesis libraries.

ARTICLE SUMMARY

Researchers in the field of molecular genetics frequently investigate protein function via mutations. Here, we present a rapid deep-mutational scanning methodology called Tiled-Region Exchange Mutagenesis that can comprehensively create mutant libraries. The approach is a multiplexed version of the established EMPIRIC mutagenesis approach and retains its simplicity. We show optimizations of the in vitro reactions to reduce cost and improve robustness. This method will be useful to researchers seeking to perform deep-mutational scanning on any gene of interest.

INTRODUCTION

In the field of genomics and proteomics, assessing the functional impacts of mutations has been a crucial area of research. Efficient strategies for site-directed mutagenesis have allowed improved understanding of catalytic enzymes(1), binding interfaces(2), and regulatory elements(3). Standard methods for site-directed mutagenesis frequently assess mutations at specific sites to specific residues and therefore only offer a small glimpse of possible mutations on a gene of interest(4). In proteins, a less biased method is alanine scanning(5), which aims to decipher the functional contribution of each native amino acid in a protein. More recently, alanine scanning has been supplanted by deep-mutational scanning (DMS) where the activity of each possible amino acid is compared at each position in a protein(6). These screens offer an extremely rich view of the functional and evolutionary constraints on proteins as amino acids of similar biochemical properties can be leveraged to obtain deeper insight on functional residues(6, 7). However, one major challenge with obtaining these improved mutagenesis screens is that deep-mutational scanning is far less accessible to current molecular biology labs due to cost and difficulty of production(8). This limitation has precluded the genome-scale analysis of variants.

To capture a complete set of comprehensive variants for a gene of interest, the DMS approach must be unbiased and provide a saturated library pool(9, 10). Ideally, every possible amino acid change across the target gene or region should be present, encompassing all variants identified through evolutionary comparisons or variant screening and large-scale sequencing efforts. Moreover, each variant created in the pool should be equally represented before having undergone any phenotypic assays. This approach must also have the capability of being performed in bulk, using high-throughput screening techniques, while being scalable to any desired gene size.

Several methods have been previously developed for generating variant libraries to assess mutations in bulk including parallel site-directed mutagenesis (SDM)(11), error-prone PCR (ep-PCR), Saturated Programmable Integration of Elements (SPINE)(12), and various oligonucleotide-based approaches (Kunkel(13), PFunkel(14), PALS(15), POPcode(16), and EMPIRIC(17)). These oligonucleotide-based approaches have become more popular in recent years due to the rapidly decreasing cost of oligo pool synthesis. While these methods all have their advantages, to our knowledge no existing method is both easy to perform and generates all possible single amino acid variants or missense mutations in a gene, while simultaneously ensuring that each plasmid copy contains only one single amino acid variant, and that the wildtype sequences are minimized in this library.

To address these current limitations, we have developed a simple, novel DMS approach that multiplexes the EMPIRIC methodology, called Tiled-Region Exchange (T-REx) Mutagenesis (Fig. 1). T-REx aims to achieve a high mutational efficiency while multiplexing the workflow within a streamlined one-pot reaction, such to create a library of evenly represented single amino acid mutations along a gene of interest. This approach also aims to increase throughput, while decreasing the off-target mutation rate, to improve scalability for variant effect mapping and functional interpretation. Here, we outline our novel approach, its advantages, and implications as an integral DMS approach that generates a variant library enriched in mutations which may inform biomedical research and be relevant to disease.

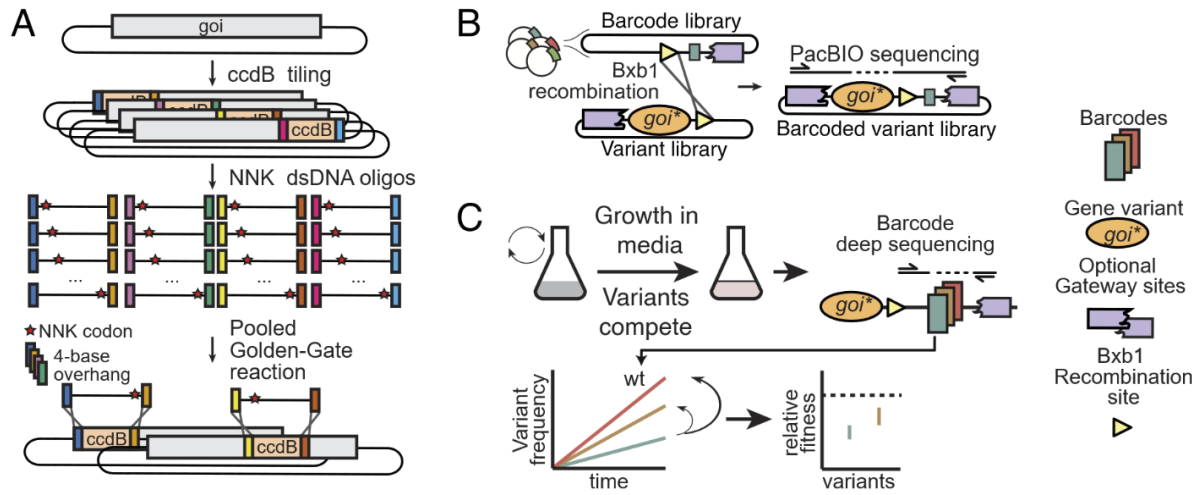


Figure 1: (A) The gene of interest (goi) is split into non-overlapping tiled regions with unique inward-facing BsaI sites. The native sequence is temporarily replaced with a *ccdB* negative selection cassette, and then swapped for the desired dsDNA mutagenic oligo containing NNK codons at each position in a Golden Gate reaction with 4-base overhangs. **(B)** The variant library and barcode plasmid library are fused in a Bxb1 recombinase reaction and can be sequenced to associate the mutation to their unique barcode. **(C)** To complete a selection assay on the variant library, variants can compete for growth and their relative fitness can be obtained by sequencing the abundance of each barcode before and after selection.

MATERIALS AND METHODS

In silico data-assisted design

To multiplex the EMPIRIC mutagenesis method, self-encoded removal fragments (SERFs) are cloned in non-overlapping locations (or tiles) in a gene in parallel, and then subsequently pooled followed by exchange with desired fragments using simultaneous Golden Gate assembly in a one pot reaction, where T4 ligase and BsaI restriction endonuclease work in tandem for scarless ligation reactions over non-palindromic overhangs. Reaction fidelity is ensured by the specific overhangs used in this Golden Gate reaction, but minimizing off-target ligations is still desirable by choosing tile locations that produce overhangs with minimal crosstalk.

To select these tile locations, we took inspiration from the Data-optimized Assembly Design process that was developed for simultaneous gene assembly of 52 fragments(18, 19). We begin by randomly choosing tile boundary locations and calculating an objective function, which scores the probability of on-target assembly, the probability of off-target assembly, the presence of palindromes, the presence of vector-only relegation, and the tile-size variance. Tile positions are then randomly moved in a local region and positions that maximize this objective function are maintained. After a few iterations, we obtain sets of tile locations with compatible overhangs that can be assembled in one-pot. A script implementing this algorithm can be obtained at: https://github.com/annb-lab/TRex.

Library generation and cloning

92 *ccdB* entry vector cloning

93 Tiled *ccdB*-containing plasmids were generated with parallel PCR reactions, where the *Escherichia*
94 *coli* plasmid containing a gene of interest was first amplified to exclude the tile region. Each of these
95 PCR reactions were performed separately, so that there was an amplified fragment for each tile. The
96 primers used for these PCRs contained a complementary overhang to the *ccdB* PCR fragment, which
97 were then cloned with Gibson Assembly(20) in a 10 ul reaction by combining equal amounts of the
98 *ccdB* fragment and backbone PCR fragment. Reactions were then incubated at 50 °C for 1 hour on
99 the thermocycler. 1 ul of this assembled product was transformed into *ccdB* Survival 2 competent
100 cells (Invitrogen), which allow for the otherwise toxic *ccdB* fragment to be propagated.

101 Oligo extension

102 Oligos were ordered from IDT (Integrated DNA Technologies) as standard desalted oligos, as
103 Ultramers, or as oPools (which use the Ultramer synthesis platform). They were then resuspended
104 in 10 mM Tris, 0.1 mM EDTA pH 8.0 to 10 μM. Conversion of the single-stranded oligonucleotide to
105 double-stranded DNA and preparation of the *E. coli* library was performed as in previously with small
106 modifications (which will be described in their respective sections)(21). In some trials, gBlocks would
107 substitute for these Ultramers, with specific introduced mutations, for a site-directed mutagenesis
108 approach (three to four mutations in different tiles can be ordered as a single 500 bp gBlock) in a
109 more cost effective and simple manner.

110 NEBridge and transformation

111 Golden Gate assembly(22) was performed using a cycling method with a modified reaction buffer
112 that shows higher cloning efficiency. We usually used about 500 ng of *ccdB*-entry vector, and 1 ul of
113 purified dsDNA oligonucleotide in a 20 ul reaction. We used the following cycling protocol: 1) 37 °C
114 for 1 minute, 2) 16 °C for 1 minute, 3) GOTO 1, 60 times, 4) 50 °C for 5 minutes, 5) 12 °C for infinite
115 time. We used a 5x assembly buffer: 200 ul of 10x Cutsmart buffer, 20 ul of 100 mM ATP, 20 ul of 1 M
116 DTT, 60 ul of propylene diol, 100 mg of PEG 8000 and water to 400 ul. Our 20x enzyme mix was: 10 ul
117 T4 ligase (2000 U/ul), 30 ul BsaI-HFv2 (20 U/ul). The product of the cycling reaction can be directly
118 transformed into homemade chemical competent cells(23), and we routinely obtain 10^5-10^6 cfu.
119 from this transformation.

120

121 Bxb1-integrase fusion

122 To barcode the deep-mutational scanning library, barcodes were cloned in a plasmid containing
123 kanamycin resistance and an R6K origin of replication, which requires a strain containing the Pir
124 element for propagation(24). Barcodes were cloned as in previously, except using strains BAN004
125 (DH10B, *uidA*::pir116) and BAN005 (ccdB Survival 2, *uidA*::pir116) for plasmid propagation. The
126 plasmid also contained a Bxb1 attP site, while the variant library, containing ampicillin resistance
127 and a pUC origin of replication, contained a Bxb1 attB site(25). A Bxb1 fusion reaction was prepared
128 as follows: 400 ng of the plasmid containing the mutagenized library, 200 ng of the barcode library, 2
129 ul of 5x Bxb1 buffer (250 mM Tris pH 8, 250 mM KCl, 375 mM NaCl, 5 mM EDTA, 500 ug/mL BSA, 25%
130 PEG 8000), 1 ul of Bxb1 integrase enzyme (at approximately 0.5 mg/mL) and topped with water to 10
131 ul. The reaction was then incubated at 37°C for 2 hours in a thermocycler. Gel electrophoresis
132 verification of the fusion reaction indicated that the reaction is essentially complete within 2 hours
133 at 37 °C. The reaction was then transformed directly into chemical competent *E. coli* cells and
134 recovered in LB+Amp+Kan to select for cells that received the fused plasmid only.

135 Purification of Bxb1 integrase

136 Purification of the Bxb1 integrase was performed using a Bxb1 integrase construct that was C-
137 terminally tagged with a 6xHis. This construct was expressed in BL21(DE3) *E. coli* cells that were
138 grown in ZYM505 medium(26). The cultures were induced with IPTG at 0.25 mM at mid-log phase
139 (OD_{600} ~ =0.4–0.8) and incubated overnight at 18°C with shaking. Cells were then harvested by
140 centrifugation and resuspended in lysis buffer (20 mM Tris (pH 8), 1 M NaCl and 5% (v/v) glycerol).
141 After cell lysis by sonication and centrifugation to remove cell debris, the clarified lysate was loaded
142 into a nickel-nitrilotriacetic acid (Ni-NTA) resin column, which was pre-equilibrated with lysis buffer
143 + 10mM Imidazole. The column was then washed with the same lysis buffer + 10mM Imidazole, to
144 remove excess non-specific binders and outcompeting weakly bound proteins, before eluting the
145 6xHis-tagged Bxb1 integrase with lysis buffer + 250 mM Imidazole. The eluted protein then underwent
146 a buffer exchange to completely remove the imidazole and concentrate the protein to 1 mg/ml. The
147 purified protein was then diluted with equal volumes of glycerol (final 50% v/v glycerol), reaching a
148 final concentration of ~8.5 uM and stored at –20 °C. We did not remove the 6xHis-tag from the Bxb1
149 integrase protein since the tag is small (~0.8 kDa) and has not been seen to interfere with the robust,
150 high-yielding Bxb1 reaction efficiency.

151 Yeast transformations

152 After barcoding the library, the purified plasmid pools can be transfected into a model organism of
153 choice. We used a standard Lithium Acetate (LiOAc) and 50% PEG protocol to transform our libraries
154 into yeast(27). An overnight culture of the desired yeast strain was grown to saturation in YPD (2%
155 peptone, 1% yeast extract, 2% (w/v) glucose), and 150-300 ul was transferred into 5 ml of fresh YPD
156 and grown for 4.5-5 hours at 30 °C until an optimal OD600 of 0.4-0.6 was reached. The cells were
157 then pelleted, washed, and the following were added on top of the cells: 240 ul of 50% PEG 3350, 36
158 ul of 1 M Lithium Acetate, 50 ul denatured Salmon Sperm DNA (2 mg/mL, boiled 5 min, snap cooled
159 on ice), 10-20 ul of PmeI digested mutagenized barcoded library. The PmeI digestion was done to
160 release the mutagenized fragment with regions of homology for integration in the yeast genome. It is
161 specific to our experiment and other means of transfection are possible. The reaction was then
162 vortexed until homogenized and left to incubate in a 42 °C heat bath for 1 hour. The cells were then
163 pelleted, the supernatant was removed, and the pellet was resuspended in 1 ml of ddH_2O before
164 being plated on standard dropout plates and incubated for 2-3 days at 30 °C.

165 **Sequencing library preparation**

166 gDNA extraction

167 After growing the transformed yeast pools to saturation, the genomic DNA was extracted using the
168 following gDNA extraction protocol. 1-3 mL was spun down and the supernatant was removed. The
169 cell pellet was resuspended in 100 ul of yeast lysis extraction buffer (5 mg/mL Zymolyase 20T (100
170 U/mL), 100 mM Sodium Phosphate buffer pH 7.4 (0.5 M Na2HPO4 and 0.5 M NaH2PO4), 10 mM
171 EDTA, 0.5% SB3-14, 200 ug/mL Rnase A, 1 M Sorbitol, 20 mM DTT, stored at -20 °C) and placed at
172 37°C for 30 minutes or more until complete lysis. After lysis, 400 ul of lysis/binding buffer (100 mM
173 MES pH 5, 4.125 M Guanidine thiocyanate, 25% isopropanol, 10 mM EDTA) was added to the tube,
174 and vortexed until all precipitates were dissolved. The tube was spun down for 30 seconds if unlysed
175 cells remained. The supernatant of the tube was passed onto a standard miniprep silica column and
176 spun for 30 seconds to pass the supernatant through the column. 1x wash with 400 ul Wash buffer 1
177 (10% GuSCN, 25% Isopropanol, 10 mM EDTA) followed by 1x wash with 600 ul 10 mM Tris/80%
178 Ethanol was performed and spun for 30 seconds after each wash. The column can then be dried by

179 spinning for 3 minutes at maximum speed. The gDNA was then eluted with 50 ul of elution buffer (10
180 mM Tris-HCl, pH 8.5), where the expected DNA concentration was around 20-30 ng/ul for a 1 mL
181 culture. Success of the extraction was verified by agarose gel electrophoresis.

182 PCR and sequencing

183 Unique barcodes linked with mutated codons can be sequenced using long-read sequencing or with
184 Illumina in the case of short libraries. This can be done with PCR (necessary for libraries integrated
185 inside model organisms), or by restriction digest followed by adapter ligations according to the
186 recommendations of the sequencing platforms. Our FKBP1a library was sequenced on a MiSeq v2
187 500-cycle kit using PCR protocols as previously described in a previous study(21).

188

189 **RESULTS AND DISCUSSIONS**

190 **Efficient and comprehensive deep-mutational scanning libraries using tiled Golden-Gate**
191 **assembly**

192 There is no shortage of approaches for systematically mutagenizing genes of interest on a plasmid.
193 The conceptually simplest approach to do this is by performing standard site-directed mutagenesis
194 (SDM) at every desired position, however this is extremely laborious(11). Higher throughput methods
195 have used mutagenic oligonucleotides and a polymerase, or error-inducing enzymes (such as error-
196 prone PCR(28) or fusion of T7 RNA polymerase with AID(29)). All these approaches, however, require
197 fine-tuning the Poisson rate of introduced mutations and cannot guarantee that all constructs
198 contain a single mutation (and not zero or more than one). In contrast, the EMPIRIC (Extremely
199 Methodical and Parallel Investigation of Randomized Individual Codons) method is an
200 oligonucleotide-based approach that clones the designed oligonucleotides directly with the use of a
201 ligase (17) and thus falls into a different class of mutagenesis techniques that can guarantee the final
202 product. In EMPIRIC, plasmids receiving the oligonucleotide cassettes have a Self-Encoded Removal
203 Fragment (SERF) flanked by inverted BsaI restriction sites(17). As such, plasmids can be treated
204 enzymatically with BsaI to ligate oligo fragments flanked by sticky ends that have been designed to
205 ligate at the desired position in the target vector. Where the EMPIRIC method falls short is that it only
206 mutates a single region of a gene. However, it is trivial to imagine how the EMPIRIC approach can be
207 parallelized to mutagenize a complete gene.

208 In light of this, the theoretical ideal approach would generate one and only one mutation, with high
209 mutational efficiency in a one-pot reaction. This would allow the production of comprehensive
210 libraries of all possible missense variants in a gene of interest. To this end, we developed Tiled-Region
211 Exchange (T-REx) mutagenesis, which was greatly inspired by the EMPIRIC approach. Briefly, the
212 approach is a *multiplexed* version of EMPIRIC where all the SERFs (referred to as 'tiles') are replaced
213 with their corresponding oligonucleotides in a single reaction. The following four objectives were
214 prioritized prior to the development of this method: ease of use, generation of all possible single-
215 mutant variants, production of one and only one variant per clone, and minimization of unmutated
216 sequences in the library.

217 To multiplex EMPIRIC and decrease unmutated sequences in the library, we made two major
218 modifications to the protocol: 1) we include the toxic *ccdB* gene in the SERF, and 2) we use tiles that
219 are positioned such that they contain unique and optimized overhangs that can be used to clone
220 synthetic mutagenizing oligos in a single reaction. The addition of *ccdB* in the SERF virtually
221 guarantees that only cells that have undergone a successful tile-exchange will be viable(30), thus

222 decreasing the frequency of clones that maintain the WT, unmutated sequence. The unique
223 optimized overhangs enable the entire gene to be mutated in a single reaction vessel as these
224 minimize off-target ligations of oligonucleotides. Finally, to aid in downstream phenotyping of
225 mutants, we introduce an approach to add unique barcodes to the mutagenesis libraries using the
226 Bxb1 recombinase.

227 A brief overview of the workflow of T-REx is as follows: parallel cloning of SERFs in a gene of interest
228 on a plasmid, converting designed oligonucleotide pools to dsDNA, one-pot cloning of all the
229 oligonucleotides with a plasmid pool of SERF-containing genes (Fig. 1A), followed by one-pot fusion
230 with a unique barcode (Fig. 1B). The final library can be introduced in a model system of choice for
231 further phenotyping (Fig. 1C).

232 **Automated and optimized design of tile locations**

233 One critical component of T-REx is the initial insertion of SERFs. Simultaneous cloning of
234 oligonucleotides that potentially encode different tile sequences requires that their ligation and
235 exchange with corresponding SERF is highly specific. Here, a trade-off between the number of tiles
236 and the length of the cloning oligonucleotide must be considered, as longer oligonucleotides are
237 more costly and may have a higher rate of synthesis errors, while having more tiles increases the
238 complexity of the reaction (notwithstanding the increased number of SERFs that must be cloned in
239 parallel).

240 To explore these constraints, we first verified whether the length of mutagenic oligonucleotides is a
241 limiting factor in T-REx mutagenesis. We cloned SERFs inside regions of different lengths into the
242 eforRed pink-producing chromoprotein(31), and performed a single oligonucleotide exchange,
243 counting resulting colonies for cloning efficiency and yield. A successful reaction yields pink
244 colonies, while oligonucleotide synthesis errors (which are frequently indels) would yield a white
245 colony. Across a variety of overhangs and a large range of oligonucleotide sizes (from <60 bp to 200
246 bp), using standard desalted oligos or using Ultramers from IDT (for longer oligos), we found no
247 drastic difference in correct assembly or colony counts after *E. coli* transformation (results not
248 shown). Therefore, contrary to our original assumption, the length of the oligonucleotides or the
249 synthesis platform is not a limiting factor for EMPIRIC and the requirement for semi-accurate
250 synthesis on both 5' and 3' ends of the purchased oligo is sufficient to ensure high-efficiency cloning.
251 While shorter oligos may be more cost-effective and accessible, they do not influence the reaction
252 efficiency, and thus the trade-off for a comprehensive mutagenesis library is simply the cost of
253 oligonucleotides and the total number of SERFs. In a practical sense, the span of a tile can therefore
254 be about 50 amino acids.

255 To reduce cost, we further explored the number of bases needed for cleavage close to the end of DNA
256 fragments for the BsaI restriction endonuclease. According to NEB, BsaI can cleave fragments when
257 the recognition sequence is 1 bp from the end of a double-stranded fragment, but they recommend
258 6 base pair for Golden Gate assembly(32). We thus varied the number of bases from 0-20 bp past
259 one of the two BsaI site in the oligo (one site is necessarily longer due to the need for a primer during
260 conversion of ssDNA oligos to dsDNA, see Methods) and tested the efficiency of the reaction using
261 the same colorimetric assay as previously discussed except that we mutagenized a plasmid
262 containing the amilCP violet-producing chromoprotein. In our hands, 0 and 1 base lengths failed to
263 produce many colonies, while extensions of 2 bases or more, including 3, 4, 5, 6, 7, 8, 16, and 20 bp,
264 supported a robust assembly at high colony counts (Fig. 2A). We observed no notable difference in
265 colony counts past 2 base extensions, suggesting this as a minimum threshold to support the
266 reaction.

Upon testing many different overhangs, we encountered one case where an increase in white colonies was observed and traced this to a case where overhangs differing by one base could promote re-circularization of the plasmid without the oligonucleotide. While this is a common occurrence in some restriction-endonuclease cloning workflows, it was not expected to occur based on the measured fidelity of T4 DNA ligase (5-GGAA ligated to 3-CCCT was seen approximately 0.1% of the time in NEB's screen(18)). Despite this, we observed about 25% white colonies in this reaction (Fig. 2B). Thus, care must be taken when designing tile locations to minimize these off-target ligations.

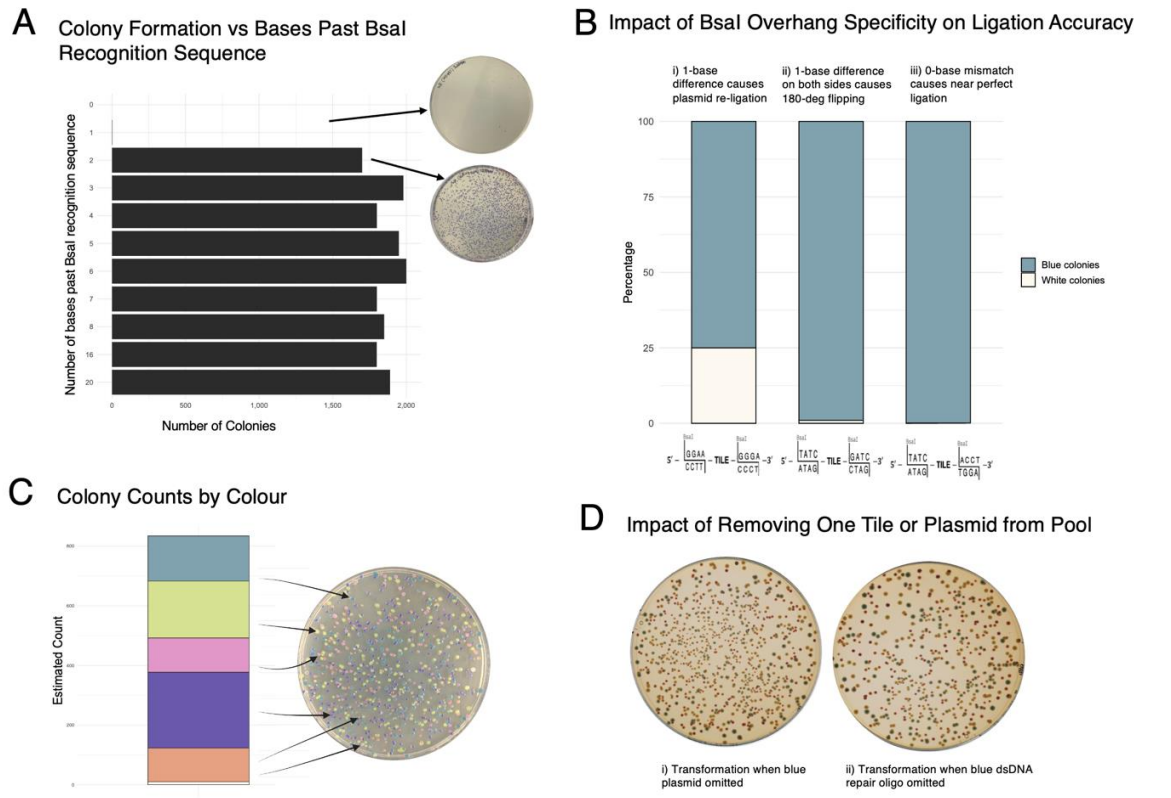


Figure 2: (A) Barplot depicting colony formation vs number of bases past the BsaI recognition sequence with plate photographs indicating threshold number for reaction success. **(B)** Stacked barplots indicating the overhang specificity and its impact on ligation accuracy. **(C)** Stacked bar plot indicating colony counts grouped by chromoprotein identity (five chromoproteins, each represented by a different colour) indicating variability in ligation efficiency. The white colonies are incomplete reactions, represented at a very small frequency. **(D)** Representative agar plates demonstrating i) the impact on the transformation of removing the blue plasmid backbone from the pool of plasmids and ii) of removing the blue repair oligo from the reaction. No blue colonies were observed on either plate and there was no increase in white colonies.

Our previous results suggest that T-REx could be used as a comprehensive mutagenesis technique that can mutagenize whole genes with minimal constraints. Thus, to aid in developing this methodology, we developed a script that can automate the cloning process and choose optimal tile locations, inspiring ourselves from data-optimized assembly design. The efficiency and fidelity of T4 DNA ligation on all possible 4 bp overhangs was previously measured by NEB(18), and serves as a guide for optimized tile locations. The number of tiles is chosen by the user (taking into account the total cost of assembly), and all oligonucleotides, including mutagenic oligopools, are returned in a simple output.

While it has been routine in synthetic biology labs to remove BsaI sites from genes for cloning purposes(33), it is also necessary for deep-mutagenesis to consider that sets of random nucleotides can generate a *de novo* BsaI restriction sequence within the mutagenized region. In our lab, we usually perform degenerate NNK mutagenesis (though NNN or NNS is also possible) to randomly generate all 20 amino acids at a single site, while reducing redundancy and the number of stop codons. In certain instances (for example NNK CTC), a spurious BsaI site can be produced. In these cases, the script automatically attempts to further mutagenize surrounding bases to preserve the coding amino acid (for example NNK CTC to NNK CTT) or will resort to NNS mutagenesis. One advantage of data-optimized assembly design is that it can easily incorporate further constraints in the design, and so other restriction endonuclease sites can also be automatically screened and removed in a similar manner.

Tiled-Region Exchange mutagenesis can be performed as a one-pot reaction

One strength of T-REx is that, by choosing unique overhangs where off-target ligations have been minimized, the reaction might be performed in one-pot. Though not necessarily required for successful and comprehensive mutagenesis, this may improve throughput and cost as long as the reaction remains specific. For example, the list price for IDT oligo pools in small scale is 5 cents per base, while it is 1 cent per base at very high scale(34). To showcase the specificity of this one-pot reaction, we cloned SERFs into 5 different chromoproteins. Single-stranded oligos that 'repair' these SERFs were ordered and pooled prior to conversion to dsDNA, and a one-pot Tiled-Region Exchange mutagenesis reaction was performed. The tile locations were chosen such that incorrect assembly (an oligo swapped with the wrong SERF) would yield an unpigmented colony. Under ideal conditions, we would expect approximately equal cloning efficiencies for all five chromoproteins and thus observe an equal proportion of colored colonies. The results of this experiment are shown in Fig. 2C, suggesting that the reaction is highly efficient and specific, depicting <1% of colonies having been incorrectly ligated and appearing white, with each colour appearing at a similar frequency.

To further highlight the specificity of this reaction, we performed the same experiment but this time omitting either one plasmid entirely (and thus having too many 'repair oligos') or omitting one swapping oligonucleotide (and thus having one plasmid with a SERF that cannot be successfully cloned). In neither of these cases did we observe a detrimental effect. As shown in Fig. 2D, all 4 of the other chromoproteins were properly ligated with their exchange oligos and there was no excess of white colonies. As expected, no colonies produced the omitted intact chromoprotein.

Barcodes can be attached to the mutagenic library using recombinase-mediated fusion

In the EMPIRIC approach, the mutagenized fragment can be sequenced directly for phenotyping purposes using standard short-read sequencing. However, when mutagenizing whole-genes in one-pot reactions, this approach cannot be guaranteed to sequence the intended tile and long-read sequencing typically do not offer the throughput required for phenotyping thousands of variants

321 effectively. Previously, it has been shown that short, unique DNA barcodes can be linked to variations
322 of interest and used to phenotype libraries effectively with short-read sequencing(35).

323 To do this, we designed our mutagenesis to be performed on a plasmid containing a Bxb1 attB(25)
324 (bacterial attachment) integration site and ampicillin resistance. On another plasmid, a barcode
325 library is cloned on an attP-containing (phage attachment) plasmid(25) with an R6K origin of
326 replication with kanamycin resistance(24). Bxb1 serine-integrase can thus fuse both plasmids *in vitro*
327 on the conserved "GA" dinucleotide sequence within the attachment sites, yielding a barcoded
328 deep-mutation scan library after transformation into standard DH5a cells and selecting on both
329 kanamycin and ampicillin. The linkage between the mutagenized variant and the unique barcode can
330 be determined by using any of the long-read sequencing technologies available and the use of Bxb1
331 integrase enables flexibility for downstream users that may wish to use the Gateway(36) system to
332 clone their library onto different backbones while preserving this linkage. Thus, this approach allows
333 the unique tagging, quantification, and identification of each variant in the sequencing analysis from
334 complex multiplexed mixtures.

335 In our workflow, we have optimized this Bxb1 fusion reaction through several parameters and
336 assessed these reactions on an agarose gel. We first tested a range of temperatures, 20°C, 25°C, 30°C,
337 and 37°C, where we noticed that the most complete fusion reaction occurred at 37°C (Fig. 3A). We
338 also tested various incubation times, at 37°C, by inactivating the reaction every 15-30 minutes until
339 complete fusion (results not shown). Fusion of plasmids by Bxb1 was essentially complete after two
340 hours. We also tested various salt concentration and additives (Spermidine, Bovine Serum Albumin
341 (BSA), propylene glycol, polyethylene glycol (PEG), etc.) to the fusion buffer, where the best fusion
342 results occurred in the presence of 75 mM NaCl (Fig. 3B), 50 mM KCl and incorporation of 5% PEG
343 8000 (Fig. 3C).

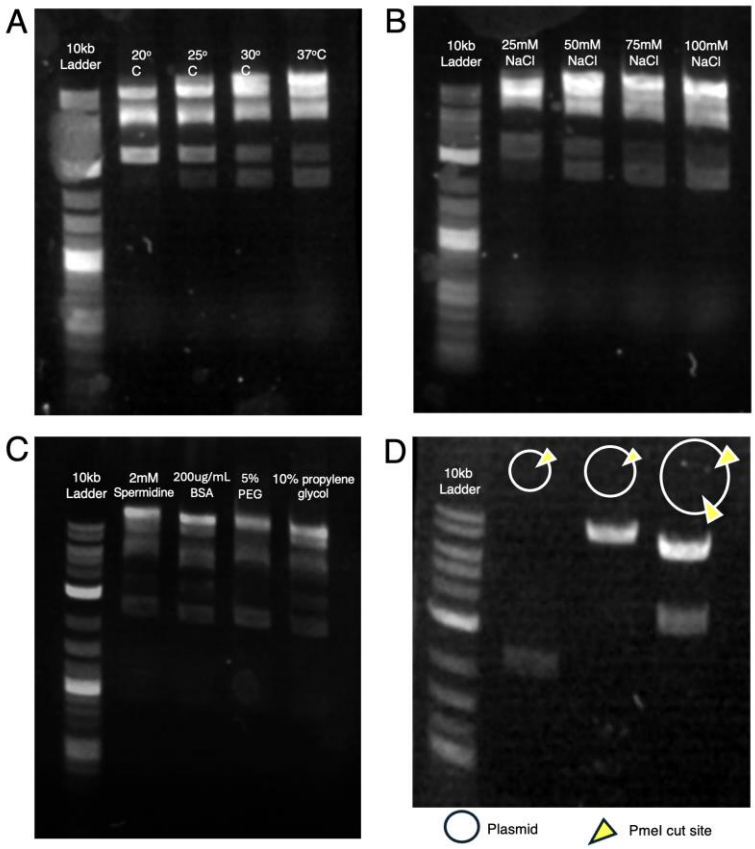


Figure 3: Optimization of Bxb1 integrase-mediated fusion between a linear R6K (oriγ) recipient vector and a donor fragment. A complete fusion reaction yields four diagnostic bands corresponding to the linear R6K backbone, the donor fragment, and the two recombination products and is consistent across each reaction. **(A)** Agarose gel of temperature series. Lane 1: DNA size ladder. Lanes 2-5: fusion reactions performed at 20 °C, 25 °C, 30 °C, and 37 °C, respectively for 2 hours. **(B)** Additive series. Lane 1: DNA ladder. Lanes 2-5: reactions supplemented with 2 mM spermidine, 200 µg/mL BSA, 5% PEG, and 10% propylene glycol, respectively. Reaction at 37 °C for 2 hours. **(C)** Salt tolerance series. Lane 1: DNA ladder. Lanes 2-5: reactions supplemented with 25 mM, 50 mM, 75 mM, and 100 mM NaCl, respectively. Reaction at 37 °C for 2 hours. **(D)** Validation of plasmid fusions using PmeI restriction endonuclease which confirms correct linearization of fragments. **Lane 1:** 10kb DNA ladder; **Lane 2:** PmeI digested barcode plasmid before Bxb1 fusion reaction; **Lane 3:** PmeI digested mutagenized plasmid before Bxb1 fusion reaction; **Lane 4:** PmeI digested fusion construct of barcode plasmid and mutagenized plasmid after Bxb1 fusion reaction. Band sizes are consistent with PmeI restriction site locations.

Deep-mutational scanning libraries produced by tiled-region exchange mutagenesis are comprehensive

To confirm that Tiled-Region Exchange mutagenesis can be used for the production of comprehensive single-variant libraries, we produced a mutant library for the human FKBP1a gene, whose protein product binds to TOR in the presence of rapamycin to inhibit cell growth(37). The

library was integrated into the yeast genome at the benign *ho* locus, which contains loss-of-function mutations in the laboratory yeast strain(38). We then sequenced the barcoded library after yeast integration and sought to obtain several quality metrics that could confirm successful mutagenesis. The biological insights gained from performing comprehensive variant scanning on FKBP1a after selection will be described elsewhere.

We first verified whether NNK mutagenesis ordered as oPools had a nucleotide bias, which would result in a skewed amino acid representation. In standard desalted oligos, IDT offers 'hand-mixing' to ensure equal base representation, but this cannot be done at the scale of oPools. We found a modest synthesis bias with 30%G: 29%T: 24%A: 17%C. With NNK mutagenesis, this translated to an overrepresentation of glycine amino acids, and a decrease in histidine and glutamines. While proline was also reduced compared to expectations, there are more codons that code for proline than histidine, and as such the absence of prolines in the final mutagenesis library was not very pronounced. To rectify this bias, we suggest that oligos can be synthesized as both Cricks and Watson strand (so as to have MNN as well), however this will double the cost of making such mutagenesis libraries and it may be more cost-effective to transform the libraries at higher multiplicity if possible.

Finally, to show that the library can produce variants at every position, we cumulated sequencing reads with barcodes and their consensus mutations. We found on average 4.35 barcodes per mutants, with the median being 3. Although we observed a slight mutational bias in the first tile (Fig. 4A), presumably due to uneven plasmid mixing or exchange efficiency, our mutational coverage was 94.66% and a heatmap indicating positions with the number of barcodes representing the mutation is shown in Fig. 4B. Obtaining about 100 colony-forming units per mutational position appears sufficient to ensure high coverage of all variants. These results suggest that T-REx mutagenesis can be used for the production of comprehensive deep-mutational scanning libraries.

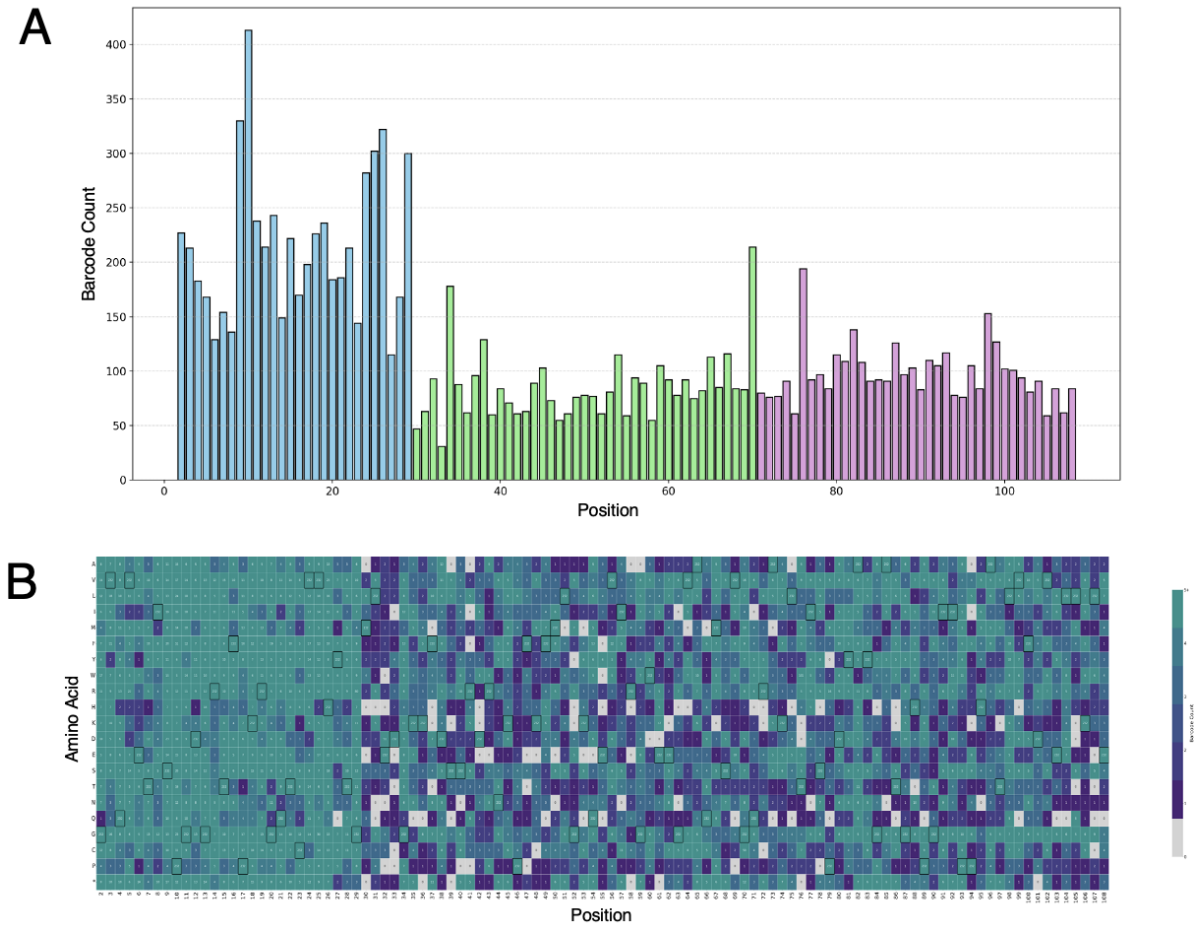


Figure 4: (A) Bar plot showing the number of barcodes per position along FKBP1A. The gene was divided into three tiles: Tile 1 (positions 2-29), Tile 2 (positions 30-70), and Tile 3 (positions 71-108) and a total of 13,062 barcodes were detected, distributed as follows: 6,279 in Tile 1, 3,743 in Tile 2, and 3,950 in Tile 3. Barcodes corresponding to variants without an identifiable mutation were excluded from this analysis. **(B)** Heatmap of barcode frequency per mutation. Wild-type positions are outlined in black boxes.

CONCLUSION

Tiled-Region Exchange (T-REx) mutagenesis combines the simplicity of the EMPIRIC approach and the throughput of other one-pot mutagenesis techniques. One caveat, however, is that the initial *ccdB* cloning stage must be done in parallel and can be laborious for large genes or for more systematic studies. Further, while T-REx can essentially guarantee one mutation per plasmid within the pool, it is often desirable to introduce multiple mutations per fragment. With T-REx, multiple variants can be introduced by designing oligonucleotides containing several mutations, however the user is constrained to a small region within a tile. In principle, this can be rectified by assembling several oligonucleotides simultaneously (in a three- or four-piece Golden Gate assembly), however

this can only produce combinatorial libraries. In cases where single variants must be observed under different gene background, it may be preferable to combine T-REx with error-prone PCR or with other mutagenic techniques.

During the writing of this manuscript, another study showcased a similar assembly technique for deep-mutational scanning(39). Our approach differs in a few ways, namely by using *ccdB* during one of the reaction steps, and by allowing pooling of the complete reaction during assembly. In contrast, one particular strength of that assembly method is that barcodes are programmed during mutagenesis, which avoids the need for long-read sequencing and can even bypass sequencing to associate barcodes to mutations. Nevertheless, as deep-mutational scanning becomes more accessible, there will be a rise of new methodologies that increase efficiency, cost, and ease of use. New developments in this area will enable combining the strength of different established methodologies to ultimately give flexibility to researchers in this field.

Despite the limitations of Tiled-Region Exchange mutagenesis, our lab has leveraged the relatively straightforward methodology to train undergraduate students with minimal molecular biology experience, and we anticipate that its ease of use will be useful for labs seeking to produce routine comprehensive variant libraries without much trial and error.

DATA AVAILABILITY

Strains and plasmids are available upon request. Raw sequencing data has been submitted to Sequence Read Archive. BioProject ID: PRJNA1316630. Code is available on https://github.com/annb-lab/TRex.

ACKNOWLEDGMENTS

KK, CC, KB, and ANNB acknowledge funding from the Canadian Institute of Health Research (CIHR: Canada Graduate Scholarship-Master's, OGB-185738). ANNB also acknowledges funding from NSERC (RGPIN-2021-02716), Connaught, and the OVPRI at UTM. We also acknowledge Fritz Roth for input during the development of this method. Finally, we acknowledge The Centre for Applied Genomics sequencing center for technical help during the sequencing of the libraries.

REFERENCES

1. Plapp,B.V. (1995) [4] Site-directed mutagenesis: A tool for studying enzyme catalysis. In *Methods in Enzymology*, Enzyme Kinetics and Mechanism Part D: Developments in Enzyme Dynamics. Academic Press, Vol. 249, pp. 91–119.

2. Casipit,C.L., Tal,R., Wittman,V., Chavaillaz,P.-A., Arbuthnott,K., Weidanz,J.A., Jiao,J.-A. and Wong,H.C. (1998) Improving the binding affinity of an antibody using molecular modeling and site-directed mutagenesis. *Protein Science*, **7**, 1671–1680.

3. Pattanaik,S., Werkman,J.R., Kong,Q. and Yuan,L. (2010) Site-Directed Mutagenesis and Saturation Mutagenesis for the Functional Study of Transcription Factors Involved in Plant

Secondary Metabolite Biosynthesis. In Fett-Neto,A.G. (ed), *Plant Secondary Metabolism Engineering: Methods and Applications*. Humana Press, Totowa, NJ, pp. 47–57.

4. Huttanus,H.M., Triola,E.-K.H., Velasquez-Guzman,J.C., Shin,S.-M., Granja-Travez,R.S., Singh,A., Dale,T. and Jha,R.K. (2023) Targeted mutagenesis and high-throughput screening of diversified gene and promoter libraries for isolating gain-of-function mutations. *Front Bioeng Biotechnol*, **11**, 1202388.

5. Weiss,G.A., Watanabe,C.K., Zhong,A., Goddard,A. and Sidhu,S.S. (2000) Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc Natl Acad Sci U S A*, **97**, 8950–8954.

6. Fowler,D.M. and Fields,S. (2014) Deep mutational scanning: a new style of protein science. *Nat. Methods*, **11**, 801–807.

7. Fowler,D.M., Adams,D.J., Gloyn,A.L., Hahn,W.C., Marks,D.S., Muffley,L.A., Neal,J.T., Roth,F.P., Rubin,A.F., Starita,L.M., *et al.* (2023) An Atlas of Variant Effects to understand the genome at nucleotide resolution. *Genome Biol*, **24**, 147.

8. Wei,H. and Li,X. (2023) Deep mutational scanning: A versatile tool in systematically mapping genotypes to phenotypes. *Front Genet*, **14**, 1087267.

9. Rubin,A.F., Gelman,H., Lucas,N., Bajjalieh,S.M., Papenfuss,A.T., Speed,T.P. and Fowler,D.M. (2017) A statistical framework for analyzing deep mutational scanning data. *Genome Biology*, **18**, 150.

10. Coyote-Maestas,W., Nedrud,D., Okorafor,S., He,Y. and Schmidt,D. (2020) Targeted insertional mutagenesis libraries for deep domain insertion profiling. *Nucleic Acids Res*, **48**, e11.

11. Watanabe,S., Ito,M. and Kigawa,T. (2021) DiRect: Site-directed mutagenesis method for protein engineering by rational design. *Biochemical and Biophysical Research Communications*, **551**, 107–113.

12. Coyote-Maestas,W., Nedrud,D., Okorafor,S., He,Y. and Schmidt,D. (2020) Targeted insertional mutagenesis libraries for deep domain insertion profiling. *Nucleic Acids Res*, **48**, e11.

13. Kunkel,T.A. (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc Natl Acad Sci U S A*, **82**, 488–492.

14. Firnberg,E. and Ostermeier,M. (2012) PFunkel: efficient, expansive, user-defined mutagenesis. *PLoS One*, **7**, e52031.

15. Kitzman,J.O., Starita,L.M., Lo,R.S., Fields,S. and Shendure,J. (2015) Massively parallel single-amino-acid mutagenesis. *Nat Methods*, **12**, 203–206, 4 p following 206.

16. Weile,J., Sun,S., Cote,A.G., Knapp,J., Verby,M., Mellor,J.C., Wu,Y., Pons,C., Wong,C., van Lieshout,N., *et al.* (2017) A framework for exhaustively mapping functional missense variants. *Mol. Syst. Biol.*, **13**, 957.

17. Hietpas,R.T., Jensen,J.D. and Bolon,D.N.A. (2011) Experimental illumination of a fitness landscape. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 7896–7901.

18. Potapov,V., Ong,J.L., Kucera,R.B., Langhorst,B.W., Bilotti,K., Pryor,J.M., Cantor,E.J., Canton,B., Knight,T.F., Evans,T.C., *et al.* (2018) Comprehensive Profiling of Four Base Overhang Ligation Fidelity by T4 DNA Ligase and Application to DNA Assembly. *ACS Synth Biol*, **7**, 2665–2674.

19. Pryor,J.M., Potapov,V., Kucera,R.B., Bilotti,K., Cantor,E.J. and Lohman,G.J.S. (2020) Enabling one-pot Golden Gate assemblies of unprecedented complexity using data-optimized assembly design. *PLoS One*, **15**, e0238592.

20. Gibson,D.G., Young,L., Chuang,R.-Y., Venter,J.C., Hutchison,C.A.,3rd and Smith,H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*, **6**, 343–345.

21. Nguyen Ba,A.N., Cvijović,I., Rojas Echenique,J.I., Lawrence,K.R., Rego-Costa,A., Liu,X., Levy,S.F. and Desai,M.M. (2019) High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature*, **575**, 494–499.

22. Engler,C., Gruetzner,R., Kandzia,R. and Marillonnet,S. (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS ONE*, **4**, e5553.

23. Inoue,H., Nojima,H. and Okayama,H. (1990) High efficiency transformation of Escherichia coli with plasmids. *Gene*, **96**, 23–28.

24. Metcalf,W.W., Jiang,W. and Wanner,B.L. (1994) Use of the rep technique for allele replacement to construct new Escherichia coli hosts for maintenance of R6K gamma origin plasmids at different copy numbers. *Gene*, **138**, 1–7.

25. Singh,S., Ghosh,P. and Hatfull,G.F. (2013) Attachment site selection and identity in Bxb1 serine integrase-mediated site-specific recombination. *PLoS Genet*, **9**, e1003490.

26. Studier,F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.*, **41**, 207–234.

27. Gietz,R.D. (2014) Yeast transformation by the LiAc/SS carrier DNA/PEG method. *Methods Mol Biol*, **1205**, 1–12.

28. Ossa-Hernández,N., Marins,L.F. and Almeida,D.V. (2024) Combination of error-prone PCR (epPCR) and Circular Polymerase Extension Cloning (CPEC) for improving the coverage of random mutagenesis libraries. *Sci Rep*, **14**, 15874.

29. Ali,M., Khan,A., Nursimulu,T.V. and Shin,J.A. (2025) Unlocking genetic potential: harnessing phage for targeted mutagenesis in phage-assisted evolution. *Nucleic Acids Res*, **53**, gkaf746.

30. Bernard,P. (1996) Positive selection of recombinant DNA by CcdB. *Biotechniques*, **21**, 320–323.

31. Liljeruhm,J., Funk,S.K., Tietscher,S., Edlund,A.D., Jamal,S., Wistrand-Yuen,P., Dyrhage,K., Gynnå,A., Ivermark,K., Lövgren,J., *et al.* (2018) Engineering a palette of eukaryotic chromoproteins for bacterial synthetic biology. *J Biol Eng*, **12**, 8.

32. Cleavage Close to the End of DNA Fragments | NEB.

33. Marillonnet,S. and Grützner,R. (2020) Synthetic DNA Assembly Using Golden Gate Cloning and the Hierarchical Modular Cloning Pipeline. *Curr Protoc Mol Biol*, **130**, e115.

34. oPools Oligo Pools | IDT *Integrated DNA Technologies*.

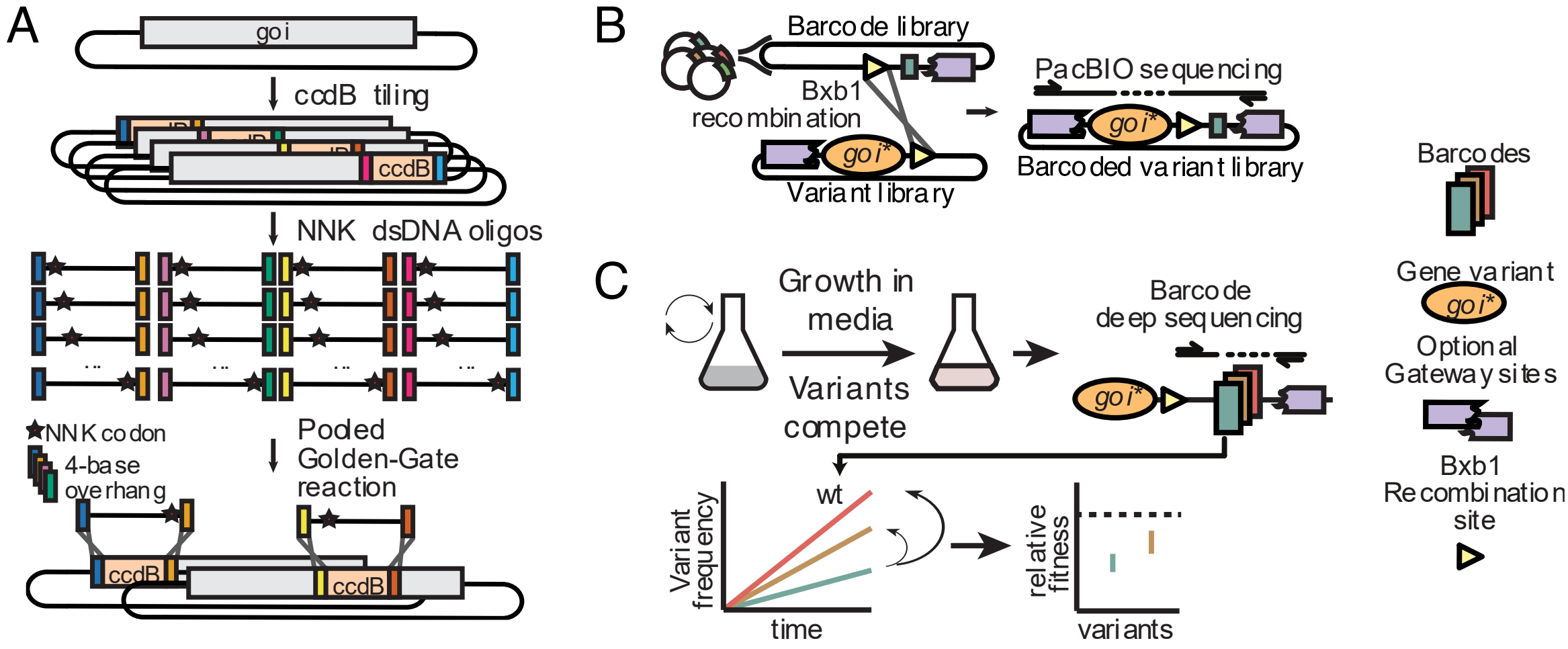
35. Chochinov,C.A. and Nguyen Ba,A.N. (2022) Bulk-Fitness Measurements Using Barcode Sequencing Analysis in Yeast. *Methods Mol Biol*, **2477**, 399–415.

36. Reece-Hoyes,J.S. and Walhout,A.J.M. (2018) Gateway Recombinational Cloning. *Cold Spring Harb Protoc*, **2018**.

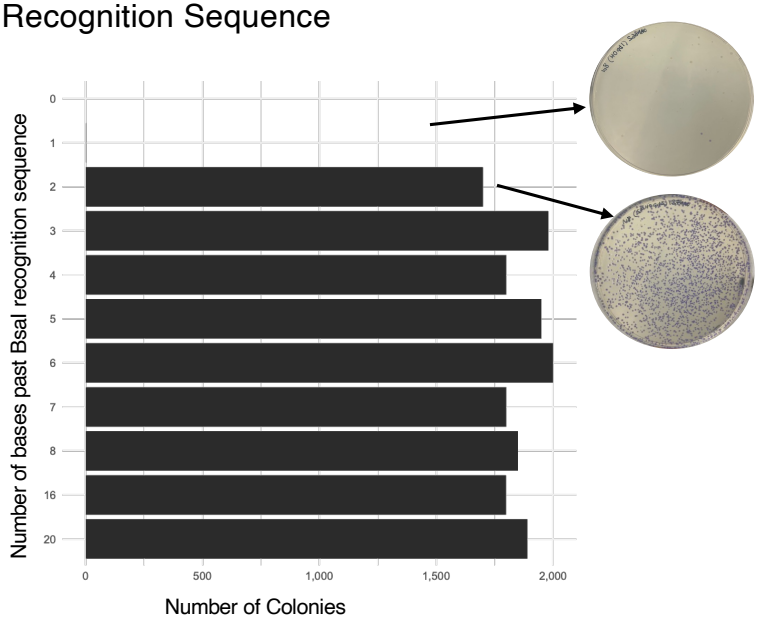
37. Sabers,C.J., Martin,M.M., Brunn,G.J., Williams,J.M., Dumont,F.J., Wiederrecht,G. and Abraham,R.T. (1995) Isolation of a protein target of the FKBP12-rapamycin complex in mammalian cells. *J Biol Chem*, **270**, 815–822.

38. SGD Project (2011) Saccharomyces Genome Database.

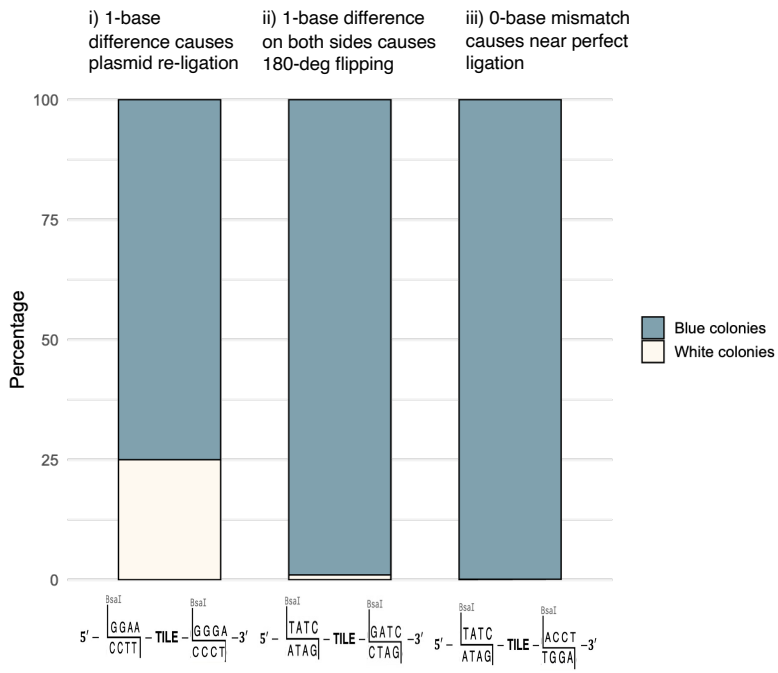
39. Jann,J., Gagnon-Arsenault,I., Pageau,A., Dubé,A.K., Fijarczyk,A., Durand,R. and Landry,C.R. (2025) Making deep mutational scanning accessible: a cost-efficient approach to construct barcoded libraries for genes of any length. 10.1101/2025.05.29.656836.


A
goi
ccdB tiling
ccdB
NNK dsDNA oligos
NNK codon
4-base overhang
Pooled Golden-Gate reaction
ccdB
ccdB
B
Barcode library
Bxb1 recombination
goi*
Variant library
PacBIO sequencing
goi*
Barcoded variant library
Barcodes
Gene variant
goi*
Optional Gateway sites
Bxb1 Recombination site
C
Growth in media
Variants compete
Barcode deep sequencing
goi*
Variant frequency
wt
time
relative fitness
variants

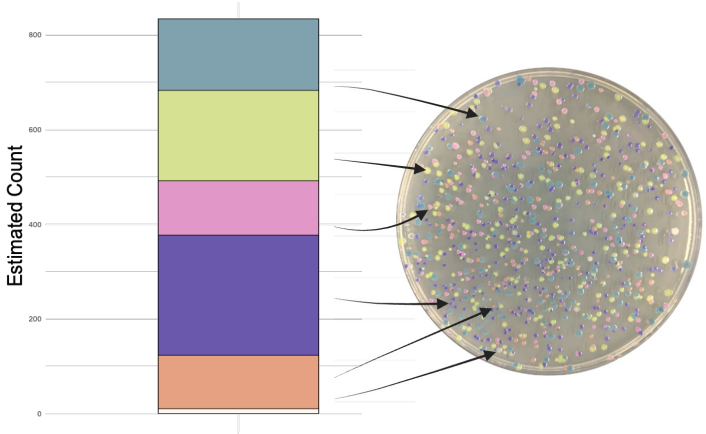
A Colony Formation vs Bases Past BsaI Recognition Sequence



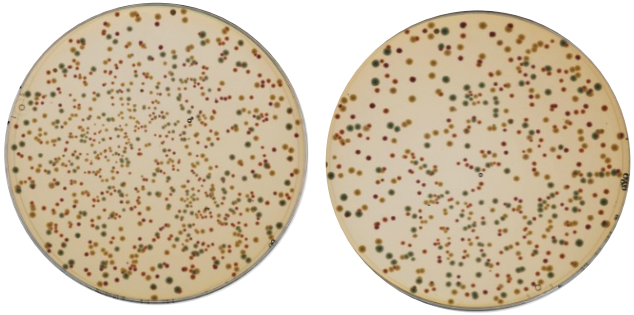
B Impact of BsaI Overhang Specificity on Ligation Accuracy



C Colony Counts by Colour

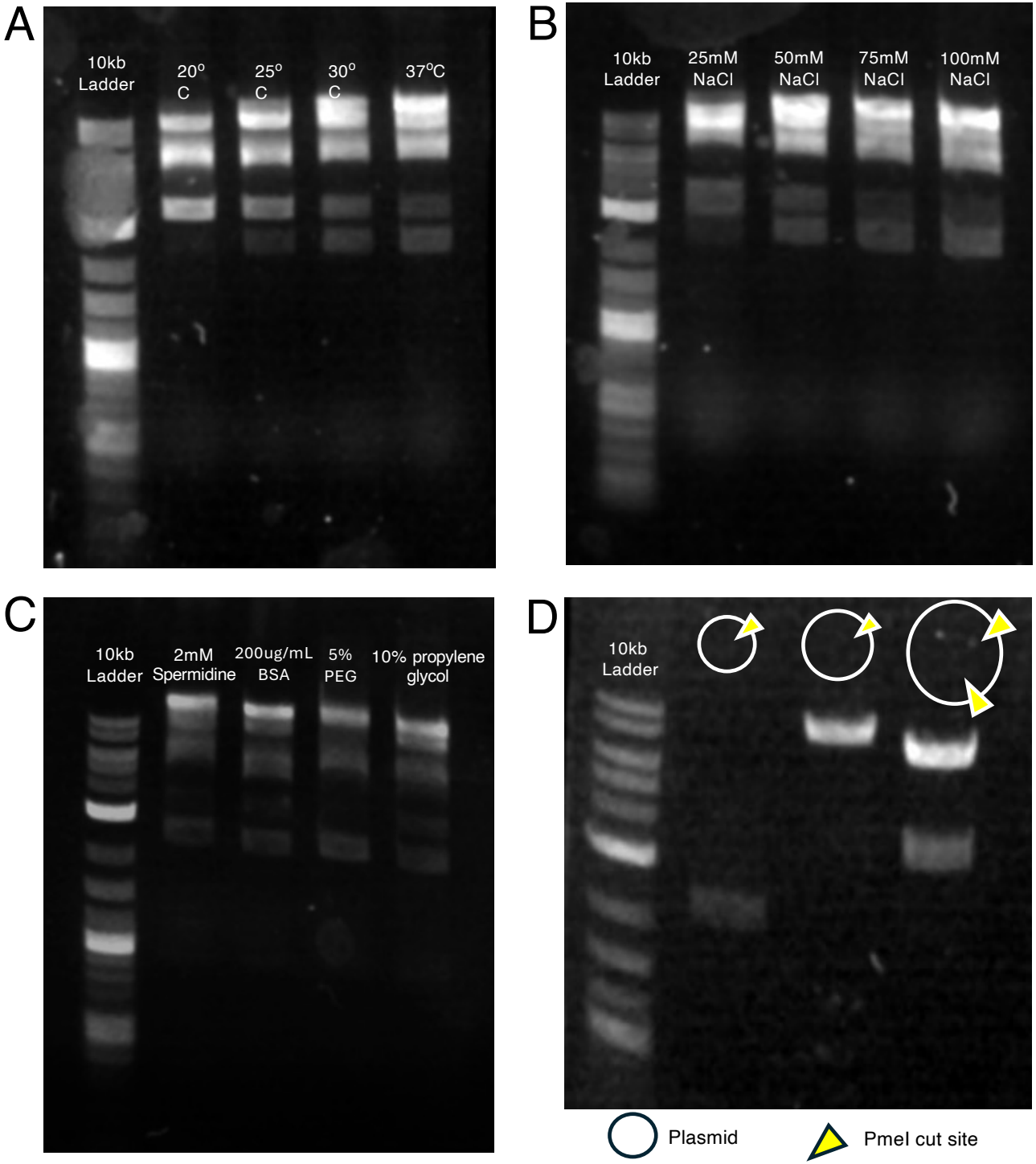


D Impact of Removing One Tile or Plasmid from Pool

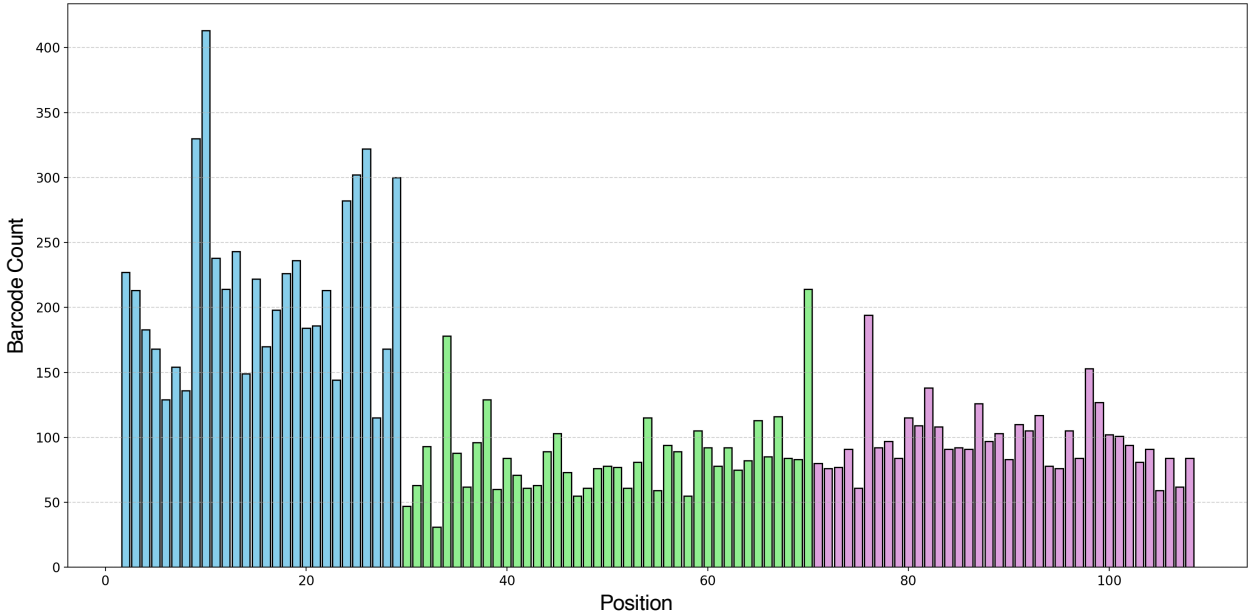


i) Transformation when blue plasmid omitted

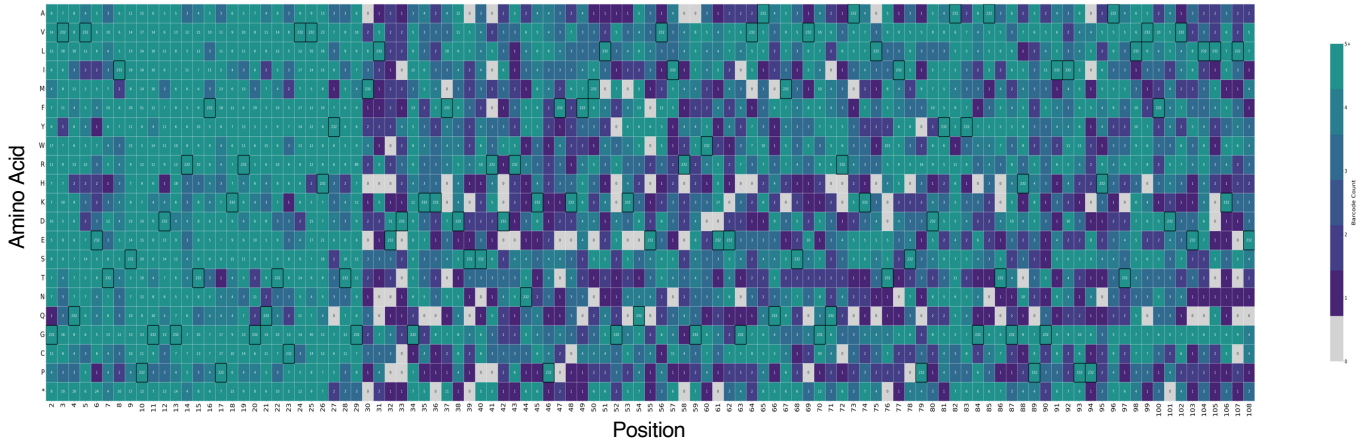
ii) Transformation when blue dsDNA repair oligo omitted


A
10kb Ladder
20° C
25° C
30° C
37°C
B
10kb Ladder
25mM NaCl
50mM NaCl
75mM NaCl
100mM NaCl
C
10kb Ladder
2mM Spermidine
200ug/mL BSA
5% PEG
10% propylene glycol
D
10kb Ladder
Plasmid
PmeI cut site

A


Barcode Count
Position
0
50
100
150
200
250
300
350
400
0
20
40
60
80
100

B


Amino Acid
Position
Barcode Count