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, for the purpose of creating 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.

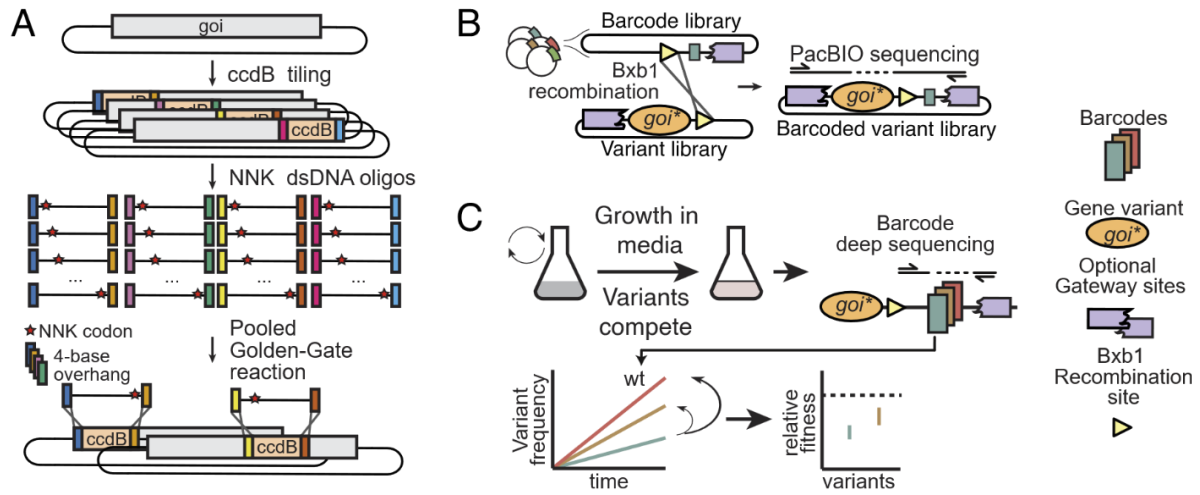


Figure 1: (A) The gene of interest (goi) is split into non-overlapping tiled regions with unique inward-facing BsaI sites. The native sequence is temporarily replaced with a *ccdB* negative selection cassette, and then swapped for the desired dsDNA mutagenic oligo containing NNK codons at each position 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 µl 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 µl 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 (oligo pools from IDT, which use the Ultramer synthesis platform). They were
104 then resuspended in 10 mM Tris, 0.1 mM EDTA pH 8.0 to 10 µM. Conversion of the single-stranded
105 oligonucleotide to double-stranded DNA and preparation of the *E. coli* library was performed as in
106 previously with small modifications (which will be described in their respective sections)(21). In
107 some trials, gBlocks would substitute for these Ultramers, with specific introduced mutations, for a
108 site-directed mutagenesis approach (three to four mutations in different tiles can be ordered as a
109 single 500 bp gBlock) in a 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 µl of
113 purified dsDNA oligonucleotide in a 20 µl 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 µl of 10x Cutsmart buffer, 20 µl of 100 mM ATP, 20 µl of 1 M
116 DTT, 60 µl of propylene diol, 100 mg of PEG 8000 and water to 400 ul. Our 20x enzyme mix was: 10 µl
117 T4 ligase (2000 U/µl), 30 µl BsaI-HFv2 (20 U/µl). 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 Bxb1-integrase fusion

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

134 Purification of Bxb1 integrase

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

150 Yeast transformations

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

164 **Sequencing library preparation**

165 gDNA extraction

166 After growing the transformed yeast pools to saturation, the genomic DNA was extracted using the
167 following gDNA extraction protocol. 1-3 mL was spun down and the supernatant was removed. The
168 cell pellet was resuspended in 100 µl of yeast lysis extraction buffer (5 mg/mL Zymolyase 20T (100
169 U/mL), 100 mM Sodium Phosphate buffer pH 7.4 (0.5 M Na2HPO4 and 0.5 M NaH2PO4), 10 mM
170 EDTA, 0.5% SB3-14, 200 µg/mL Rnase A, 1 M Sorbitol, 20 mM DTT, stored at -20 °C) and placed at
171 37°C for 30 minutes or more until complete lysis. After lysis, 400 µl of lysis/binding buffer (100 mM
172 MES pH 5, 4.125 M Guanidine thiocyanate, 25% isopropanol, 10 mM EDTA) was added to the tube,
173 and vortexed until all precipitates were dissolved. The tube was spun down for 30 seconds if unlysed
174 cells remained. The supernatant of the tube was passed onto a standard miniprep silica column and
175 spun for 30 seconds to pass the supernatant through the column. 1x wash with 400 µl Wash buffer 1
176 (10% GuSCN, 25% Isopropanol, 10 mM EDTA) followed by 1x wash with 600 µl 10 mM Tris/80%
177 Ethanol was performed and spun for 30 seconds after each wash. The column can then be dried by
178 spinning for 3 minutes at maximum speed. The gDNA was then eluted with 50 µl of elution buffer (10

179 mM Tris-HCl, pH 8.5), where the expected DNA concentration was around 20-30 ng/µl for a 1 mL
180 culture. Success of the extraction was verified by agarose gel electrophoresis.

181 PCR and sequencing

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

187 **Barcode association bioinformatics**

188 Read extraction of genes and barcode sequences

189 Reads were de-multiplexed from inline indexes and barcodes were analyzed as in a previous
190 study(21). To extract the gene from the reads, 20 bp anchors corresponding to the promoter and
191 terminators were used, allowing 2 mismatches per anchor. The barcode was similarly extracted,
192 removing fixed bases included in the barcode to prevent BsaI restriction enzyme sites.

193 Barcode clustering

194 Sequenced barcodes contain a mixture of barcodes with and without sequencing errors. Assuming
195 that most sequencing reads have no base calling mistakes on the barcodes, we can perform
196 clustering and error correction to obtain a final set of barcodes. This was performed essentially as in
197 (21), by sorting barcodes by their counts, and error correcting barcodes starting from least common
198 to most common, using a threshold of 2 Hamming distance. Finally, barcodes that were observed
199 fewer than 10 times in the whole sequencing library were removed from further analyses.

200 Mutation association

201 To associate mutations within the gene to a barcode, we tabulated for each barcoded read the list of
202 mutations detected. Each read can contain, in principle, four different types of mutations: 1) the
203 correct codon that was mutated by Tiled-Region Exchange mutagenesis, 2) an oligo synthesis
204 mistake that is associated with the mutated codon, 3) a sequencing error due to basecalling, and 4)
205 a sequencing error due to PCR chimeras. Because oligo synthesis errors will occur at the same
206 frequency as the desired mutated codon (forming a mutation set), we take advantage of the Apriori
207 algorithm (28) to identify the most common set of mutations for each barcode at the nucleotide level.
208 Given a total list of possible mutations associated with a barcode, the Apriori algorithm can return,
209 for each possible set of mutations, the most common to least common sets. Sequencing errors
210 occur usually as singletons and can be filtered out by a simple frequency threshold. However, PCR
211 chimeras occur during the amplification reaction and can therefore appear in several reads. To
212 investigate the effect of PCR chimeras in our analysis, we first cumulated the frequency of the most
213 common mutation set and the second most common mutation set for each barcode (assuming that
214 the most common mutation set is the real mutation associated with a specific barcode). We found
215 that the most common mutation set was usually found at over 30% of the reads, while the second
216 most common mutation set was usually below 10% of the reads. Thus, implementing a mutation set
217 threshold of 30% and a minimum read count of 10, both accelerates the Apriori algorithm and likely
218 returns true barcode-mutation associations.

If two mutations were found at over 30% of reads, this would indicate barcodes that are shared between different mutations. These were discarded from our analysis, as well as barcodes that were associated with an indel.

RESULTS AND DISCUSSIONS

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

There is no shortage of approaches for systematically mutagenizing genes of interest on a plasmid. The conceptually simplest approach to do this is by performing standard site-directed mutagenesis (SDM) at every desired position, however this is extremely laborious(11). Higher throughput methods have used mutagenic oligonucleotides and a polymerase, or error-inducing enzymes (such as error-prone PCR(29) or fusion of T7 RNA polymerase with AID(30)). All these approaches, however, require fine-tuning the Poisson rate of introduced mutations and cannot guarantee that all constructs contain a single mutation (and not zero or more than one). In contrast, the EMPIRIC (Extremely Methodical and Parallel Investigation of Randomized Individual Codons) method is an oligonucleotide-based approach that clones the designed oligonucleotides directly with the use of a ligase (17) and thus falls into a different class of mutagenesis techniques that can guarantee the final product. In EMPIRIC, plasmids receiving the oligonucleotide cassettes have a Self-Encoded Removal Fragment (SERF) flanked by inverted BsaI restriction sites(17). As such, plasmids can be treated enzymatically with BsaI to ligate oligo fragments flanked by sticky ends that have been designed to ligate at the desired position in the target vector. Where the EMPIRIC method falls short is that it only mutates a single region of a gene. However, it is trivial to imagine how the EMPIRIC approach can be parallelized to mutagenize a complete gene.

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

To multiplex EMPIRIC and decrease unmutated sequences in the library, we made two major modifications to the protocol: 1) we include the toxic *ccdB* gene in the SERF, and 2) we use tiles that are positioned such that they contain unique and optimized overhangs that can be used to clone synthetic mutagenizing oligos in a single reaction. The addition of *ccdB* in the SERF virtually guarantees that only cells that have undergone a successful tile-exchange will be viable(31), thus decreasing the frequency of clones that maintain the WT, unmutated sequence. The unique optimized overhangs enable the entire gene to be mutated in a single reaction vessel as these minimize off-target ligations of oligonucleotides. Finally, to aid in downstream phenotyping of mutants, we introduce an approach to add unique barcodes to the mutagenesis libraries using the Bxb1 recombinase.

A brief overview of the workflow of T-REx is as follows: parallel cloning of SERFs in a gene of interest on a plasmid, converting designed oligonucleotide pools to dsDNA, one-pot cloning of all the oligonucleotides with a plasmid pool of SERF-containing genes (Fig. 1A), followed by one-pot fusion

263 with a unique barcode (Fig. 1B). The final library can be introduced in a model system of choice for
264 further phenotyping (Fig. 1C).

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

266 One critical component of T-REx is the initial insertion of SERFs. Simultaneous cloning of
267 oligonucleotides that potentially encode different tile sequences requires that their ligation and
268 exchange with corresponding SERF is highly specific. Here, a trade-off between the number of tiles
269 and the length of the cloning oligonucleotide must be considered, as longer oligonucleotides are
270 more costly and may have a higher rate of synthesis errors (Ultramers from IDT are expected to yield
271 full length oligos at 50% of the unpurified products at 140 bp (32), while having more tiles increases
272 the complexity of the reaction (notwithstanding the increased number of SERFs that must be cloned
273 in parallel).

274 To explore these constraints, we first verified whether the length of mutagenic oligonucleotides is a
275 limiting factor in T-REx mutagenesis. We cloned SERFs inside regions of different lengths into the
276 eforRed pink-producing chromoprotein(33), and performed a single oligonucleotide exchange,
277 counting resulting colonies for cloning efficiency and yield. A successful reaction yields pink
278 colonies, while oligonucleotide synthesis errors (which are frequently indels) would yield a white
279 colony. Across a variety of overhangs and a large range of oligonucleotide sizes (from <60 bp to 175
280 bp), using standard desalted oligos or using Ultramers from IDT (for longer oligos), we found no
281 prohibiting differences in correct assembly or colony counts after *E. coli* transformation (Fig. S1).
282 Others have also found similar rates of full-length synthesis with Ultramers from IDT (34) and as will
283 be described later, even tiles of 120 bp (40 amino acids) that require oligonucleotides of about 180
284 bp can be mutated successfully with minimal errors (~5% indel rate, and ~5% wrong base
285 synthesized). Therefore, contrary to our original assumption, the length of the oligonucleotides or the
286 synthesis platform is not a major limiting factor for EMPIRIC and the requirement for semi-accurate
287 synthesis on both 5' and 3' ends of the purchased oligo is sufficient to ensure high-efficiency cloning.
288 While shorter oligos may be more cost-effective and accessible, they do not strongly influence the
289 reaction efficiency, and thus the trade-off for a comprehensive mutagenesis library is simply the cost
290 of oligonucleotides and the total number of SERFs. In a practical sense, the span of a tile can
291 therefore be about 40 amino acids.

292 To reduce cost, we further explored the number of bases needed for cleavage close to the end of DNA
293 fragments for the BsaI restriction endonuclease. According to NEB, BsaI can cleave fragments when
294 the recognition sequence is 1 bp from the end of a double-stranded fragment, but they recommend
295 6 base pair for Golden Gate assembly(35). We thus varied the number of bases from 0-20 bp past
296 one of the two BsaI site in the oligo (one site is necessarily longer due to the need for a primer during
297 conversion of ssDNA oligos to dsDNA, see Methods) and tested the efficiency of the reaction using
298 the same colorimetric assay as previously discussed except that we mutagenized a plasmid
299 containing the amilCP blue-producing chromoprotein. In our hands, all extensions, including 0, 1, 2,
300 3, 4, 5, 6, 7, 8, and 17 bp, supported a robust assembly with high colony counts (Fig. 2A and Fig. S2
301 for similar experiment using the amilOrange chromoprotein).

302 Upon testing many different overhangs, we encountered one case where an increase in white
303 colonies was observed and traced this to a case where overhangs differing by one base could
304 promote re-circularization of the plasmid without the oligonucleotide. While this is a common
305 occurrence in some restriction-endonuclease cloning workflows, it was not expected to occur based
306 on the measured fidelity of T4 DNA ligase (5-GGAA ligated to 3-CCCT was seen approximately 0.1%
307 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.

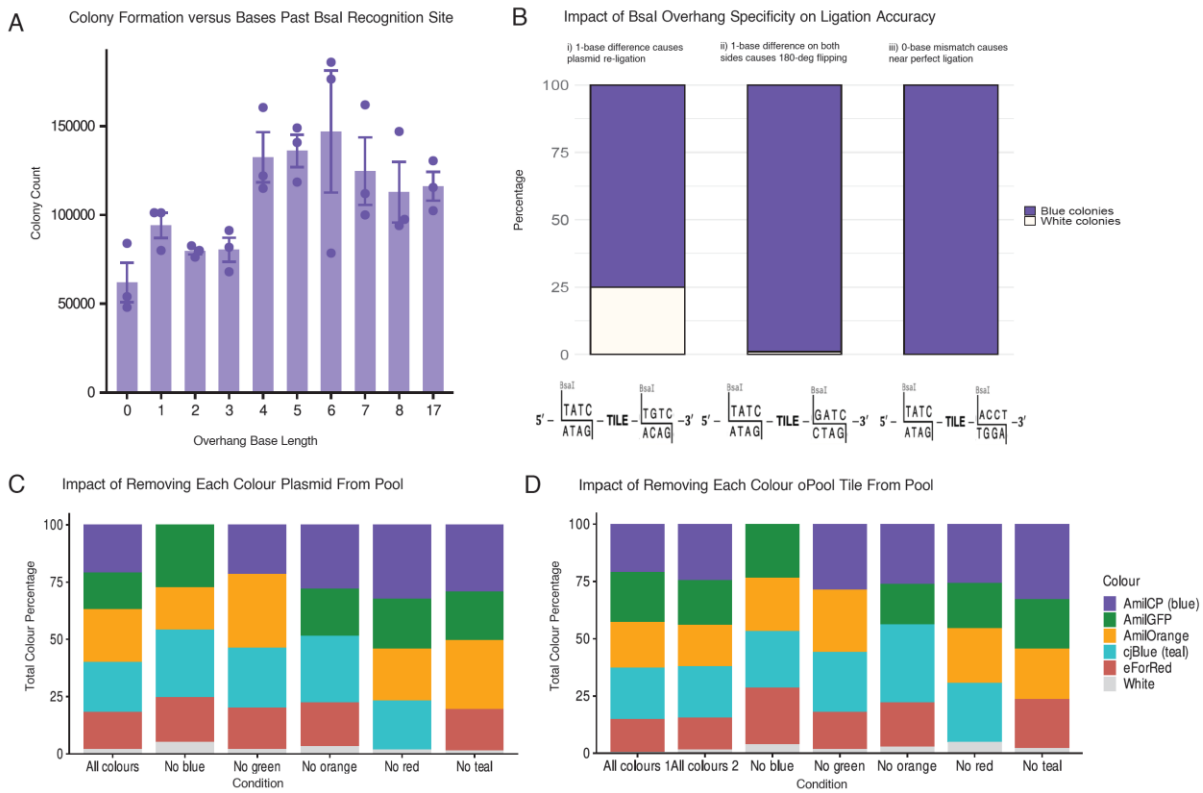


Figure 2: (A) Barplot depicting colony formation vs the number of overhang bases past the BsaI recognition sites, with three replicates per base length. **(B)** Stacked barplots indicating the overhang specificity and its impact on ligation accuracy. **(C-D)** Stacked barplots indicating colony counts grouped by chromoprotein identity (five chromoproteins each represented by a different colour). **(C)** depicts the difference between all colours included in the pool, versus omitting each colour plasmid individually from the pool, while **(D)** shows the colony counts of the pool when omitting each oPool oligo individually from the pool of oligos.

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 oligo pools, are returned in a simple output.

While it has been routine in synthetic biology labs to remove BsaI sites from genes for cloning purposes(36), 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(32). 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 coloured colonies. The results of this experiment are shown in Fig. 2C-D, 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 (Fig. 2C) (and thus having too many 'repair oligos') or omitting one swapping oligonucleotide (Fig. 2D) (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. S3, 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 effectively. Previously, it has been shown that short, unique DNA barcodes can be linked to variations of interest and used to phenotype libraries effectively with short-read sequencing(37).

In many deep-mutational scanning protocols, DNA barcodes are incorporated into the mutagenized library using PCR, by incorporating random nucleotides in the primers and cloning of the amplified product (38), or by direct ligation (39). Due to the large number of random bases in these designs, this approach virtually guarantees that barcodes do not associate with more than one gene fragment, enabling high fidelity in the following analyses. However, if the PCR product that is amplified with the barcodes is the mutagenized library, then the final libraries may contain PCR chimeras and contain

364 more than one variant. To remedy this, mutagenized libraries can be isolated by restriction enzymes
365 and ligated on a barcoded PCR product of the desired plasmid backbone.

366 Here, we chose a different strategy where barcode libraries can be sequenced and characterized
367 beforehand, effectively generating subsamples of the random nucleotide space. These barcodes can
368 be sequenced at high depth, aiding in future long-read mutation association sequencing. This library,
369 however, must be fused, or linked, to the mutagenized genes. To do this, we designed our
370 mutagenesis to be performed on a plasmid containing a Bxb1 attB(25) (bacterial attachment)
371 integration site and ampicillin resistance. On another plasmid, a barcode library is cloned on an attP-
372 containing (phage attachment) plasmid(25) with an R6K origin of replication with kanamycin
373 resistance(24). Bxb1 serine-integrase can thus fuse both plasmids *in vitro* on the conserved "GA"
374 dinucleotide sequence within the attachment sites, yielding a barcoded deep-mutation scan library
375 after transformation into standard DH5a cells and selecting on both kanamycin and ampicillin. The
376 linkage between the mutagenized variant and the unique barcode can be determined by using any of
377 the long-read sequencing technologies available and the use of Bxb1 integrase enables flexibility for
378 downstream users that may wish to use the Gateway(40) system to clone their library onto different
379 backbones while preserving this linkage. Thus, this approach allows the unique tagging,
380 quantification, and identification of each variant in the sequencing analysis from complex
381 multiplexed mixtures.

382 In our workflow, we have optimized this Bxb1 fusion reaction through several parameters and
383 assessed these reactions on an agarose gel and by transformation. We first tested a range of
384 temperatures, 20 °C, 25 °C, 30 °C, and 37 °C, where we noticed that the most complete fusion reaction
385 occurred at 37 °C (Fig. 3A). We also tested various incubation times, at 37 °C, by periodically plating
386 the reaction every 15-30 minutes and found that the number of colonies obtained was generally
387 sufficient after two hours, which roughly corresponded to results from the agarose gel
388 electrophoresis (Fig. S4). We also tested various salt concentration and additives (Spermidine,
389 Bovine Serum Albumin (BSA), propylene glycol, polyethylene glycol (PEG), etc.) to the fusion buffer,
390 where good fusion results occurred in the presence of at least 75 mM NaCl (Fig. 3B), 50 mM KCl and
391 incorporation of 5% PEG 8000 (Fig. 3C and Fig. S5). Finally, to confirm fusion of both plasmids,
392 restriction enzyme digestion was used to show a successful reaction (Fig. 3D).

393 One drawback of our approach is that the barcode diversity is necessarily much lower than barcodes
394 added by PCR or by ligation. Indeed, this diversity is limited exactly by the number of barcodes found
395 in the barcoded R6K plasmid pools. However, several lines of evidence suggest that this
396 disadvantage can be circumvented by simply having a modestly large number of barcodes. First, a
397 barcoded R6K plasmid pool from our barcoding construction protocol usually contains about
398 250,000 barcodes as established from colony counts, far exceeding the number of variants found in
399 most deep-mutational scan studies. For example, a gene of 1000 amino acids will have 32 000
400 possible single-NNK mutants, allowing about ~8 barcodes per variant as biological replicates. We
401 believe having fewer biological replicates (such as ~8) is preferential to enable high-quality
402 phenotyping as 1000 reads per barcode may be required to estimate barcode frequencies. Second,
403 even if barcodes get associated with two different mutations, this should occur relatively infrequently
404 as we will discuss in the next section. To assess this more generally, we reanalyzed the dataset from
405 (21), which used the same barcode library to insert random barcodes into several yeast strains that
406 contained a fixed known DNA barcode. Analyzing the total set of barcodes from the libraries yielded

188478 barcodes (from the expected ~250000 from colony counts). This library was inserted into two starting yeast strains, with the first having 41207 barcodes and the second yeast strain having 28013 barcodes. Between the two yeast populations, the number of shared barcodes was 5227, which was close to the expectation of 6119. Thus, the number of barcodes that will be associated with the same mutations is directly controlled by the barcode library size, which can be built to be sufficiently large for most deep-mutational scanning studies. Finally, if this overlap is found to be too high, it is trivial to obtain more barcodes by constructing more barcoded R6K plasmid pools.

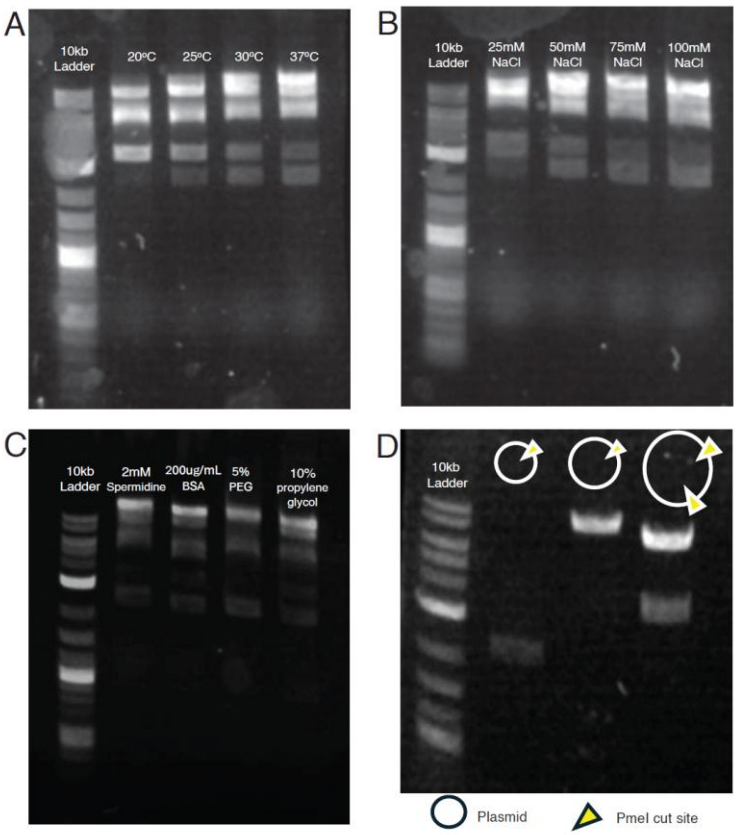


Figure 3: Optimization of Bxb1 integrase-mediated fusion between a linear R6K (oriγ) recipient vector and a donor fragment. A complete fusion reaction yields four diagnostic bands corresponding to the linear R6K backbone, the donor fragment, and the two recombination products and is consistent across 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 1 hour. **(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 1 hour. **(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 1 hour. **(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(41). The

library was produced in two technical replicates and was integrated into the yeast genome at the benign *ho* locus, which contains loss-of-function mutations in the laboratory yeast strain(42). We then sequenced the barcoded libraries 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. In total, 14,476 barcode-mutation association pairs were obtained (7104 from replicate 1, and 7372 from replicate 2). As mentioned previously, only a small number of these barcodes were associated with oligo synthesis errors despite pooled oligo lengths of 140 bp to 179 bp: 588 contained a frameshift (4%) and 425 (3%) barcodes were associated with two mutations, either by oligo synthesis error or through fusing of the same barcode to two variants. To verify whether our barcode library was diverse enough, we found 269 barcodes present in both libraries that were associated with different mutations (close to the expectation of 259). Thus, the diversity of our R6K barcode library of ~250,000 barcodes was relatively high enough to ensure only a small number of unusable barcodes.

We found on average 4.35 barcodes per NNK codon, with the median being 3. However, 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 13,463 barcodes-mutation pair 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 a single identifiable mutation or an indel were excluded from this analysis (1013 barcodes removed). **(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 backgrounds, 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(43). 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 with 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. Finally, Our FKBP1A library cost about $600 CAD for the oligos required to produce it, while a pre-made variant libraries from Twist Bioscience was quoted at $5400 USD. Thus, the cost of making libraries in-house is also relatively competitive compared to commercial products.

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

511 diversified gene and promoter libraries for isolating gain-of-function mutations. *Front*
512 *Bioeng Biotechnol*, **11**, 1202388.

513 5. Weiss,G.A., Watanabe,C.K., Zhong,A., Goddard,A. and Sidhu,S.S. (2000) Rapid mapping of
514 protein functional epitopes by combinatorial alanine scanning. *Proc Natl Acad Sci U S A*,
515 **97**, 8950–8954.

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

518 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.,
519 Rubin,A.F., Starita,L.M., *et al.* (2023) An Atlas of Variant Effects to understand the genome at
520 nucleotide resolution. *Genome Biol*, **24**, 147.

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

523 9. Rubin,A.F., Gelman,H., Lucas,N., Bajjalieh,S.M., Papenfuss,A.T., Speed,T.P. and Fowler,D.M.
524 (2017) A statistical framework for analyzing deep mutational scanning data. *Genome*
525 *Biology*, **18**, 150.

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

528 11. Watanabe,S., Ito,M. and Kigawa,T. (2021) DiRect: Site-directed mutagenesis method for protein
529 engineering by rational design. *Biochemical and Biophysical Research Communications*,
530 **551**, 107–113.

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

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

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

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

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

542 17. Hietpas,R.T., Jensen,J.D. and Bolon,D.N.A. (2011) Experimental illumination of a fitness
543 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. Odland,T. (2025) tommyod/Efficient-Apriori.

29. 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.

30. 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.

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

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

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

581 34. Filges,S., Mouhanna,P. and Ståhlberg,A. (2021) Digital Quantification of Chemical
582 Oligonucleotide Synthesis Errors. *Clin Chem*, **67**, 1384–1394.

583 35. Cleavage Close to the End of DNA Fragments | NEB.

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

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

588 38. Frank,F., Keen,M.M., Rao,A., Bassit,L., Liu,X., Bowers,H.B., Patel,A.B., Cato,M.L., Sullivan,J.A.,
589 Greenleaf,M., *et al.* (2022) Deep mutational scanning identifies SARS-CoV-2 Nucleocapsid
590 escape mutations of currently available rapid antigen tests. *Cell*, **185**, 3603-3616.e13.

591 39. Fowler,D.M., Stephany,J.J. and Fields,S. (2014) Measuring the activity of protein variants on a
592 large scale using deep mutational scanning. *Nat Protoc*, **9**, 2267–2284.

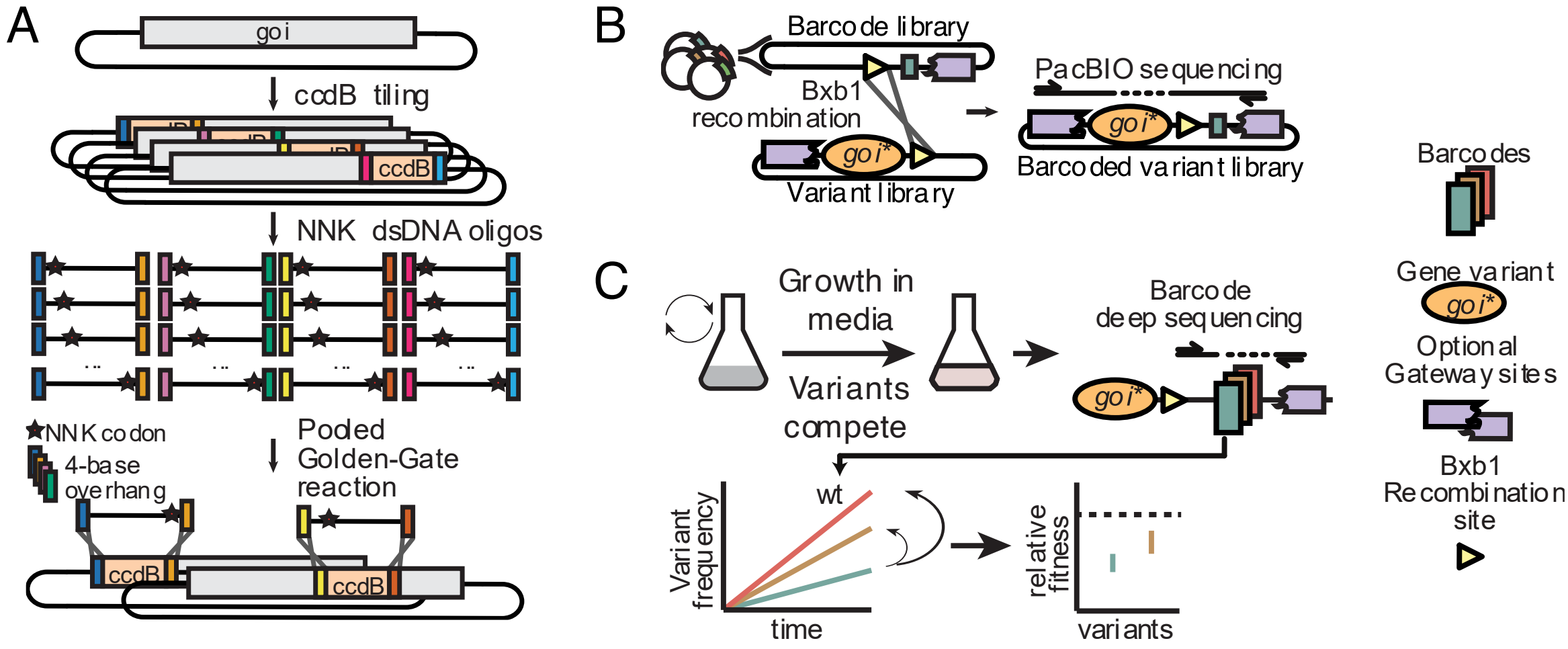
593 40. Reece-Hoyes,J.S. and Walhout,A.J.M. (2018) Gateway Recombinational Cloning. *Cold Spring*
594 *Harb Protoc*, **2018**.

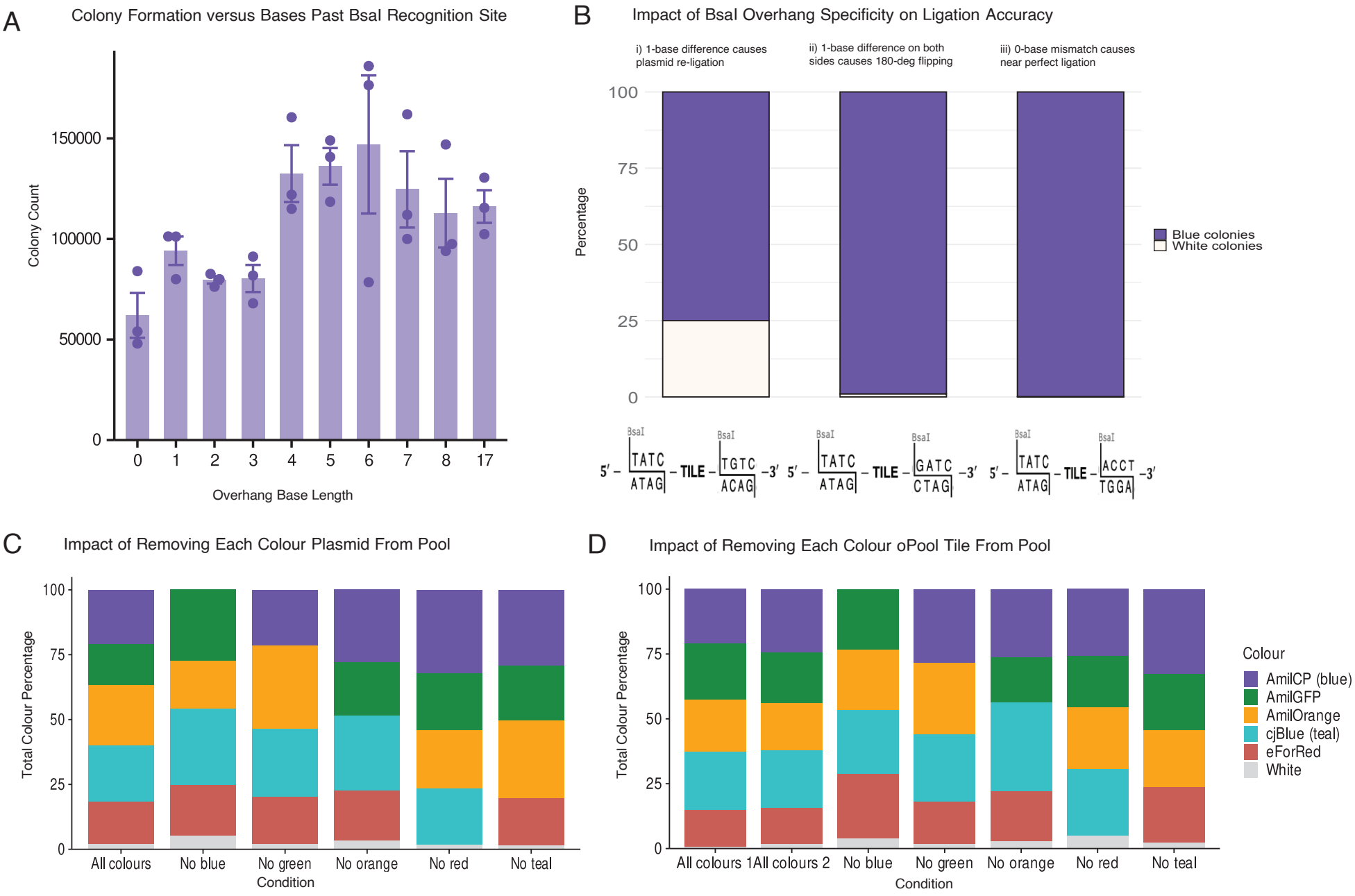
595 41. Sabers,C.J., Martin,M.M., Brunn,G.J., Williams,J.M., Dumont,F.J., Wiederrecht,G. and
596 Abraham,R.T. (1995) Isolation of a protein target of the FKBP12-rapamycin complex in
597 mammalian cells. *J Biol Chem*, **270**, 815–822.

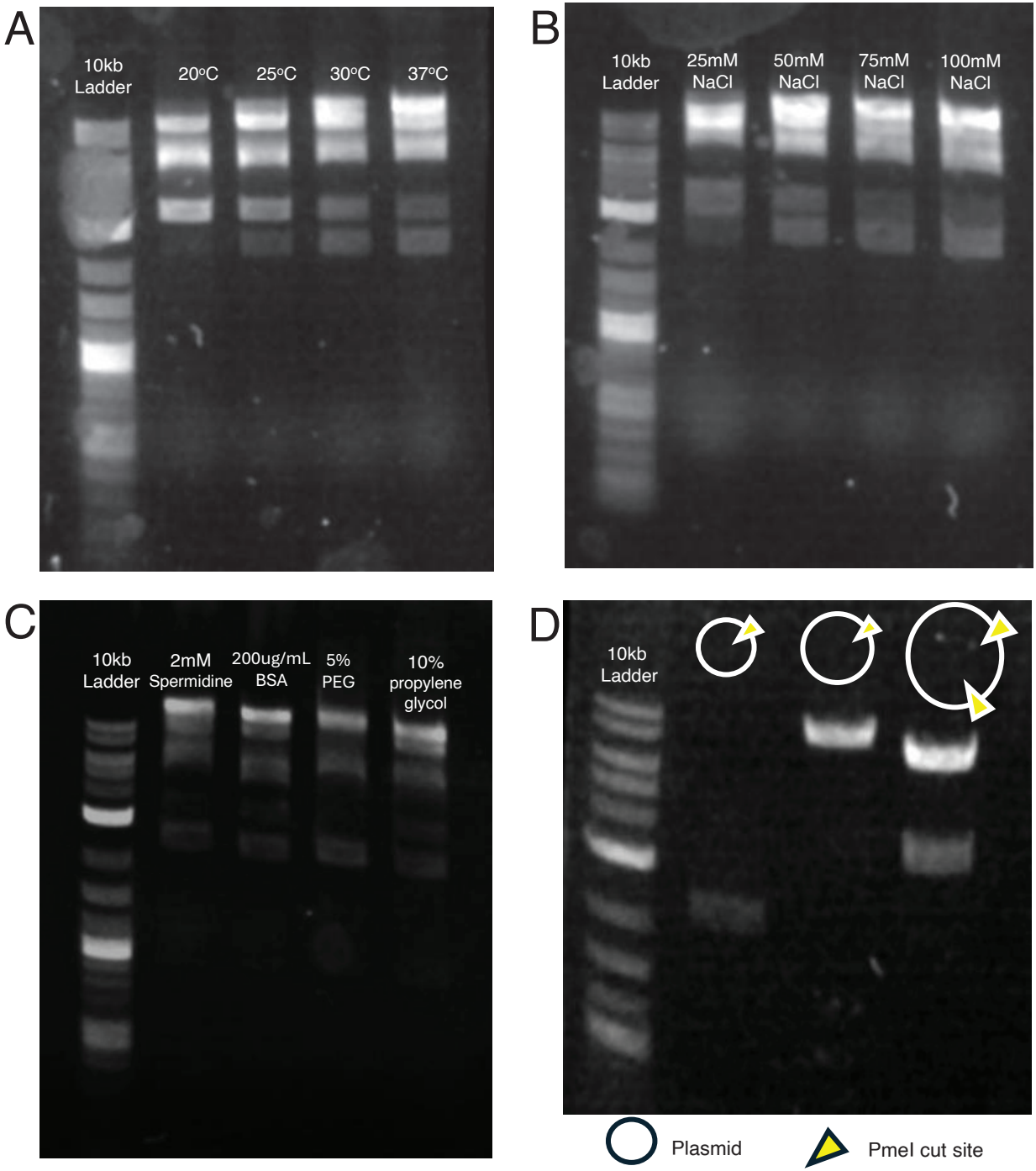
598 42. SGD Project (2011) Saccharomyces Genome Database.

599 43. Jann,J., Gagnon-Arsenault,I., Pageau,A., Dubé,A.K., Fijarczyk,A., Durand,R. and Landry,C.R.
600 (2025) Making deep mutational scanning accessible: a cost-efficient approach to construct
601 barcoded libraries for genes of any length. 10.1101/2025.05.29.656836.

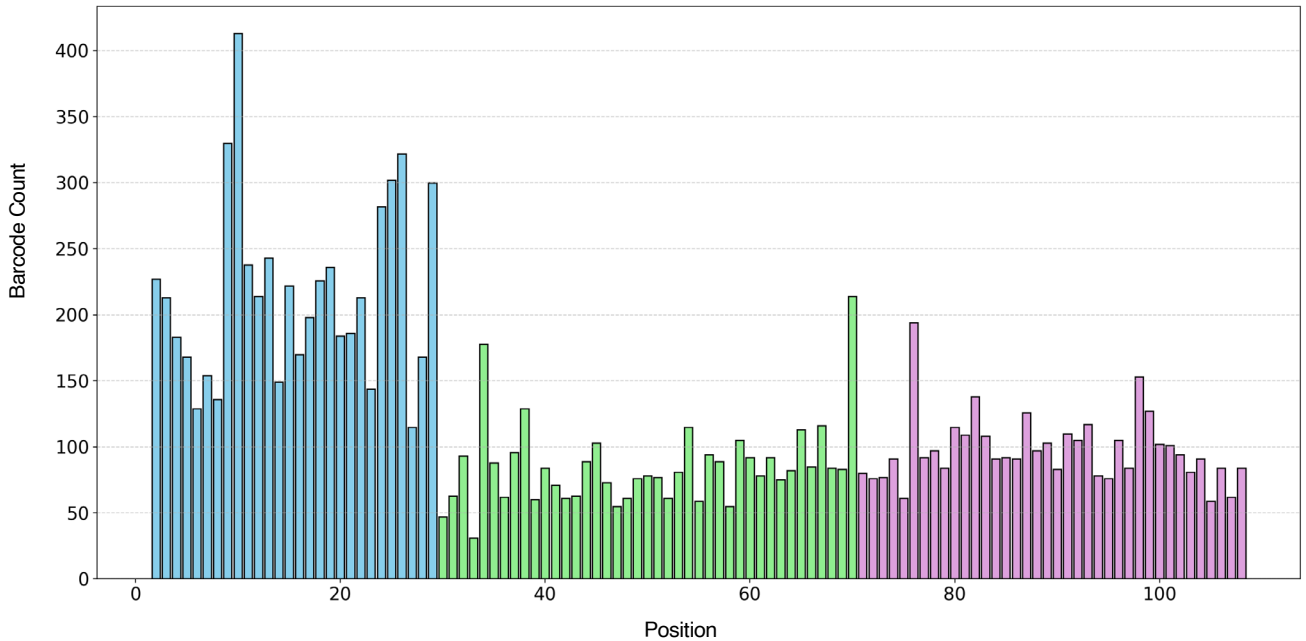
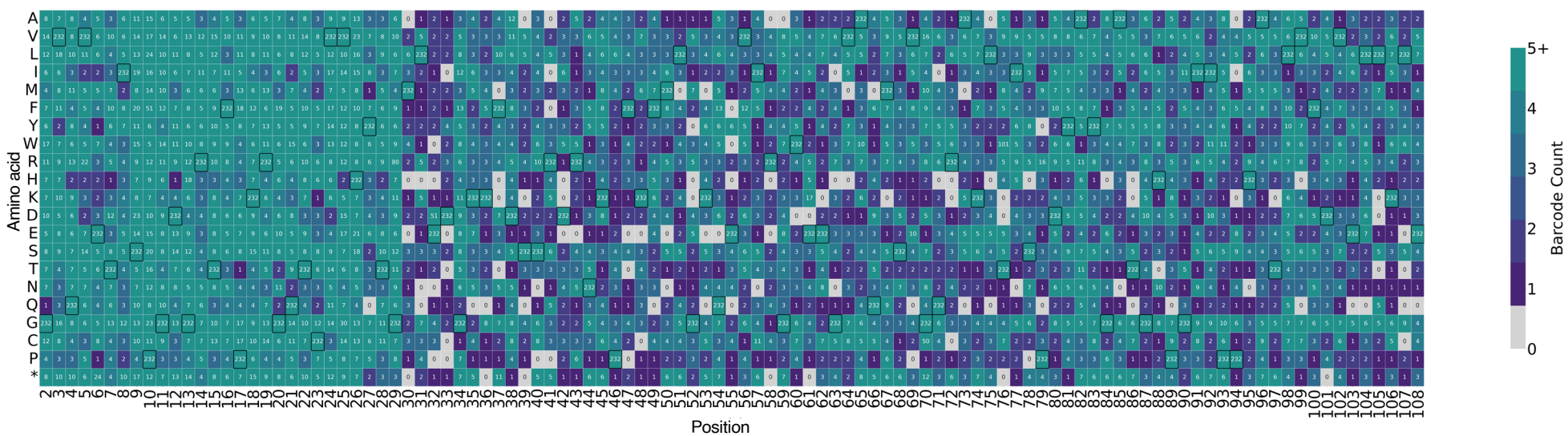
602


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 versus Bases Past BsaI Recognition Site
Colony Count
150000
100000
50000
0
0 1 2 3 4 5 6 7 8 17
Overhang Base Length
B
Impact of BsaI Overhang Specificity on Ligation Accuracy
i) 1-base difference causes plasmid re-ligation
ii) 1-base difference on both sides causes 180-deg flipping
iii) 0-base mismatch causes near perfect ligation
Percentage
100
75
50
25
0
Blue colonies
White colonies
BsaI
5′ – TATC ATAG – TILE – TGTC ACAG –3′
5′ – TATC ATAG – TILE – GATC CTAG –3′
5′ – TATC ATAG – TILE – ACCT TGGA –3′
C
Impact of Removing Each Colour Plasmid From Pool
Total Colour Percentage
100
75
50
25
0
All colours
No blue
No green
No orange
No red
No teal
Condition
D
Impact of Removing Each Colour oPool Tile From Pool
Total Colour Percentage
100
75
50
25
0
All colours 1
All colours 2
No blue
No green
No orange
No red
No teal
Condition
Colour
AmilCP (blue)
AmilGFP
AmilOrange
cjBlue (teal)
eForRed
White


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



B