Title: A two-step PCR assembly for construction of gene variants across large mutational distances

Authors:  
Shreya Routh†, Anamika Acharyya†, Riddhiman Dhar*

Affiliation:  
Department of Biotechnology, Indian Institute of Technology Kharagpur, Kharagpur-721302, West Bengal, India

* Correspondence: riddhiman.dhar@iitkgp.ac.in
† These authors contributed equally

Abstract

Construction of empirical fitness landscapes has transformed our understanding of genotype-phenotype relationships across genes. However, most empirical fitness landscapes have been constrained to the local genotype neighborhood of a gene primarily due to our limited ability to systematically construct genotypes that differ by a large number of mutations. Although a few methods have been proposed in the literature, these techniques are complex owing to several steps of construction or contain a large number of amplification cycles that increase chances of non-specific mutations. A few other described methods require amplification of the whole vector thereby increasing the chances of vector backbone mutations that can have unintended consequences for study of fitness landscapes. Thus, this has substantially constrained us from traversing large mutational distances in the genotype network, thereby limiting our understanding of the interactions between multiple mutations and the role these interactions play in evolution of novel phenotypes. In the current work, we present a simple but powerful approach that allows us to systematically and accurately construct gene variants at large mutational distances. Our approach relies on building up small fragments containing targeted mutations in the first step followed by PCR assembly of these fragments into the complete gene fragment. We demonstrate the utility of our approach by constructing variants that differ by up to 11 mutations in a model gene. Our work thus provides an accurate method for construction of multi-mutant variants of genes and therefore, will transform the studies of empirical fitness landscapes by enabling exploration of genotypes that are far away from a starting genotype.

Keywords: PCR assembly, genotype, mutagenesis, multi-mutant variants, fitness landscape
Introduction

Empirical fitness landscapes are key to a better understanding of the principles of genotype-phenotype mapping in biological systems (Poelwijk et al. 2007; Hayden et al. 2011; Crona et al. 2013; de Visser and Krug 2014; Mira et al. 2015; Schmiedel et al. 2019; Zheng et al. 2019). Empirical fitness landscapes have greatly advanced our knowledge of the functional impact of clinically observed mutations on antibiotic resistance genes (Weinreich et al. 2006; Goulart et al. 2013; Hietpas et al. 2013; Chevereau et al. 2015; Palmer et al. 2015; Rodrigues et al. 2016; Salverda et al. 2017; Palmer et al. 2018; Li et al. 2019; Das et al. 2020), impact of disease mutations on protein stability and aggregation (Spiller et al. 1999; Shimotohno et al. 2001; Bloom et al. 2005; Jacquier et al. 2013; Figliuzzi et al. 2016; Bolognesi et al. 2019; Faber et al. 2019; Li et al. 2019; Bertram and Masel 2020) and study of splice variants of a gene (Keren et al. 2010; Ke et al. 2011; Barbosa-Morais et al. 2012; Merkin et al. 2012; Gamazon and Stranger 2014; Lee and Rio 2015; Julien et al. 2016). In addition, fitness landscapes have provided great insights into molecular evolution of proteins (Canale et al. 2018) and RNA molecules (Smith 1970; Macken and Perelson 1989; Wagner 2008; Li et al. 2016; Aguilar-Rodríguez et al. 2018; Domingo et al. 2018; Pressman et al. 2019). However, majority of the empirical fitness landscapes have been limited to the local genotype neighbourhood and to mostly one, two or three mutant variants of a gene (Weinreich et al. 2006; Hietpas et al. 2011; Boucher et al. 2014; Firnberg et al. 2014; Melamed et al. 2015; Sarkisyan et al. 2016; Steinberg and Ostermeier 2016; Aguilar-Rodríguez et al. 2017; Salverda et al. 2017; Starr and Thornton 2017; Bajić et al. 2018; Bendixsen et al. 2019; Flynn et al. 2020) with exceptions of small genes such as tRNA genes (Li et al. 2016; Domingo et al. 2018). Although local neighbourhoods provide important insights into changes in fitness, they fail to capture full evolutionary trajectories occurring over deep evolutionary times. Uncovering the phenotypes of genotypes at large mutational distances can provide unprecedented insights into the interactions between large number of mutations (Hietpas et al. 2013; Melamed et al. 2015; Palmer et al. 2015; Diss and Landry 2016; Filteau et al. 2016; Diss and Lehner 2018; Domingo et al. 2018), their impact on fitness (Firmberg et al. 2014; Bank et al. 2015; Steinberg and Ostermeier 2016; Bendixsen et al. 2017; Bajić et al. 2018; Ferreriet al. 2018; Bendixsen et al. 2019; Gonzalez et al. 2019; Mehlhoff and Ostermeier 2020) and into evolution of novel phenotypes (Landry and Rifkin 2010; Starr and Thornton 2017; Aguilar-Rodríguez et al. 2018; Yang et al. 2019; Zheng et al. 2019).
Construction of fitness landscapes (Romero and Arnold 2009) rely on our ability to systematically construct variants of a genotype. There are several approaches that have been employed to construct multi-mutant variants of genes. First, site-directed mutagenesis can be used to introduce targeted mutations in a gene using methods such as Overlap-extension PCR (Ho et al. 1989). However, this method is limited in its ability to introduce more than one or two mutations in a gene (Ho et al. 1989). Thus, introduction of a large number of mutations using this technique requires a stepwise introduction of one mutation at a time. In fact, such an approach has been described by Wäneskog et al. (Wäneskog and Bjerling 2014), where the authors introduced 13 mutations in a gene. However, their method required six PCR steps and some of these steps required many amplification cycles. This made the whole assembly process complex, time-consuming and at the same time increased the chances of introduction of unintended mutations due to the large number of amplification cycles. Similarly, another method by Hejlesen and Füchtbauer (2020) utilized prolonged overlap-extension PCR and required 55 amplification cycles. Other studies have tried different approaches (Kadkhodaei et al. 2016, Young and Dong 2004), however they required special primer design with long overlaps between fragments or required special complex PCR steps.

Another set of methods for construction of multi-mutant variants that have been described in the literature are either based on or bear similarity to QuikChange site-directed mutagenesis protocol (Agilent Technologies). These methods utilize single-primer amplification reactions on the whole vector carrying the gene of interest and generate linearized fragments (Wang and Malcolm, 1999; Liu H and Naismith, 2008; Edelheit et al., 2009; Trehan et al., 2016; Kuo et al., 2017; Zeng et al., 2018). In the next step, the parental template DNA molecules are digested and the linearized plasmid fragments carrying mutations are transformed into bacteria to generate recombinant clones carrying targeted mutations. One of the earliest methods in this regard has been described by Wang and Malcolm (1999) where they introduced up to nine mutations in a gene fragment. Several variations of this method have been proposed subsequently (Liu H and Naismith, 2008; Edelheit et al., 2009; Trehan et al., 2016; Kuo et al., 2017; Zeng et al., 2018). Notably, Zeng et al. (2018) described one such method and introduced up to 15 mutations in a gene. However, all these methods rely on amplifying the whole vector and thus, risk accumulating mutations in the vector backbone. This can profoundly influence the
outcomes of selection experiments that are part of studies on empirical fitness landscapes. In this regard, one study suggested a variation of this method that required amplification of only a part of the plasmid but the process utilized a substantially larger number of cycles (Hallak et al., 2017).

Another recent method for construction of genotype variants utilizes doped oligos (Hietpas et al. 2011; Hietpas et al. 2012; Hietpas et al. 2013) which essentially uses a random mutagenesis method (Bershtein et al. 2006) with specific probabilities assigned to each type of mutation (Fowler et al. 2014; Cárcamo et al. 2017; Diss and Lehner 2018; Li and Zhang 2018). Although this method has been successfully applied to construct fitness landscapes of protein-coding (Hietpas et al. 2011; Hietpas et al. 2012) as well as tRNA genes (Fowler et al. 2014; Li et al. 2016; Li and Zhang 2018), this method is inherently limited in its capability to construct gene variants with a large number of mutations. Increasing the mutation rate in this method can lead to gene variants that differ by a large number of mutations but in turn will reduce the total number of variants sampled in the local genotype neighbourhood (Cárcamo et al. 2017). Further, gene assembly techniques from a large number of oligos can also be employed for construction of gene variants containing a large number of mutations (Ge and Rudolph 1997; Gibson 2011; Hsieh and Vaisvila 2013). Using these methods, one can combine mutated fragments with wild-type fragments and can get a mutant library with genotypes in the local as well as far-away neighbourhoods. However, these methods require a large number of amplification cycles which increases the likelihood of unwanted mutations in the gene construct.

We hereby describe a simple yet powerful and accurate two-step gene assembly method that enables us to systematically construct genotypes differing by a large number of mutations. Our method utilizes normal amplification primers and low number of amplification cycles, thus ensuring a quick and efficient gene assembly process with extremely low probability of introduction of unintended mutations. We demonstrate the capability of our method by constructing an 11-mutant variant of a model gene. Further, our method can also be adapted to combine wild-type and mutated fragments and can allow construction of gene variants in the immediate genotype neighbourhood as well as at far-away distances. Thus, we believe that our method will substantially boost the capabilities of researchers to include genotypes across large mutational distances in the study of
empirical fitness landscapes. This will facilitate developing a deeper understanding of the principles of genotype-phenotype mapping and molecular evolution.

**Materials and Methods**

**Template DNA and primers**

TEM1 β-lactamase gene on pUC19 plasmid in *E. coli* cells was chosen as the model gene system for introducing mutations at 11 different amino acid positions. The numbering scheme for residues of TEM1 gene were obtained from Bush *et al.*, 2010 (Bush and Jacoby 2010).

**First PCR amplification**

Systematic mutagenesis of TEM1 gene for 11 different amino acid positions was performed by PCR using Q5 DNA polymerase (NEB). The reaction was set up as shown in Table 1.

| Reagent                  | Amount/reaction |
|--------------------------|-----------------|
| Q5 PCR buffer            | 10 µl           |
| 10 mM dNTP               | 1 µl            |
| Q5 DNA polymerase        | 0.5 µl          |
| 10 µM forward primer     | 2.5 µl          |
| 10 µM reverse primer     | 2.5 µl          |
| Template                 | 1 µl            |
| Molecular Grade water    | 32.5 µl         |
| **Total**                | **50 µl**       |

The PCR program was set as follows: (1) initial denaturation, 30 seconds at 98°C; (2) 20 cycles of denaturation for 10 seconds at 98°C, annealing at 60°C for 30 seconds, extension at 72°C for 30 seconds; (3) final extension at 72°C for 2 minutes.

The fragment sizes varied according to the amino acid position. Our desired amino acid positions were 21, 39, 69, 104, 164, 182, 238, 240, 244, 265, and 275 of the TEM1 protein to introduce mutations as these mutations have been reported from clinical isolates of TEM1 gene. More specifically, the mutations L21F, Q39K, M69L, M69I, M69V, E104K,
R164S, R164C, R164H, M182T, G238S, E240K, R244H, R244C, R244S, T265M, R275L, and R275Q, have been deemed most prevalent across TEM1 mutants. The PCR was performed according to the procedure briefed above with primers as follows for amplifying individual fragments (Table 2).

**Table 2 – Fragments and their sizes.** The primer combinations used to generate the fragments are shown inside the parentheses.

| Serial No. | Fragments generated by primer combinations | Sizes (bp) |
|------------|------------------------------------------|------------|
| Fragment 1 | Promoter region start to amino acid residue 21 (TEM1_for+L21F_rev) | 161 |
| Fragment 2 | Residue 21 to 39 (L21F_for+Q39K_rev) | 81 |
| Fragment 3 | Residue 39 to 69 (Q39K_for+M69I_rev) | 116 |
| Fragment 4 | Residue 69 to 104 (M69I_for+E104K_rev) | 131 |
| Fragment 5 | Residue 104 to 164 (E104K_for+R164C/S_rev) | 204 |
| Fragment 6 | Residue 164 to 182 (R164C/S_for+M182T_rev) | 80 |
| Fragment 7 | Residue 182 to 244 (M182T_for+Mut238240244_rev) | 209 |
| Fragment 8 | Residue 244 to 265 (Mut238240244_for+T265M_rev) | 101 |
| Fragment 9 | Residue 265 to 275 (T265M_for+R275L/Q_rev) | 61 |
| Fragment 10 | Residue 275 to end of TEM1 gene segment (R275L/Q_for+TEM1_rev) | 61 |

**Purification and quantification of first PCR products**

After the first step of PCR, the quantity and quality of PCR products were checked in 2% Agarose gel. PCR products were then digested with 0.5 µl *DpnI* (to remove methylated DNA of plasmid template to prevent its interference in the second PCR step) and 2 µl ExoSAP (to hydrolyse excess primers and nucleotides) at 37°C for 1 hour. The enzymes were then inactivated at 80°C for 20 minutes. Next, the fragments were purified by QIAGEN
MinElute spin column as per manufacturer’s protocol and were eluted in 15µl Molecular Grade water. The purified products were checked in 2% agarose gel. The concentrations of purified PCR products were then measured by Qubit Broad Range assay (Invitrogen). Molar Mass of each PCR fragment was determined using Sequence Manipulation Suite (SMS) which calculated the number of moles present per microlitre of solution.

**Second PCR for Assembly**

Equal concentration of each purified fragment (0.5picomole or 1 picomole) from the first PCR amplification were taken as templates for the next round of reaction. First, a reaction was set up according to the Table 1 in a total reaction volume of 90µl but without adding primers. Thermal cycling conditions were set as follows - 1 cycle at 98ºC for 10 seconds; 10 cycles at 98ºC for 10 seconds, at 55ºC for 30 seconds and at 72ºC at 30 seconds; and a final extension at 72ºC for 10 minutes.

In the next step, the terminal primers (TEM1_for and TEM1_rev) were added (5 µl each and total reaction volume of 100µl). The final amplification program was done as follows - 1 cycle at 98ºC for 2 minutes; 15 cycles at 98ºC for 10 seconds, at 55ºC for 30 seconds and at 72ºC for 30 seconds; and a final extension at 72ºC for 10 minutes. The final amplified products were checked on 1% agarose gel. In absence of unspecific bands on gel, the products were purified using QIAGEN PCR purification kit following manufacturer’s protocol. In case of unspecific bands visible on gel, band representing assembled gene product was cut from the gel and was purified using QIAGEN Gel extraction kit.

**Cloning and Sanger sequencing**

The purified or extracted gene product was cloned into the plasmid pUA67 (Dhar et al. 2014) as cloning vector. The vector and insert were digested using high-fidelity restriction enzymes *Eco*RI and *Hind*III at 37ºC for 16 hours followed by inactivation of the enzymes at 80ºC for 20min. The digested vector was dephosphorylated by adding Quick CIP at 37ºC for 10min (and heat inactivating at 80ºC for 10min) to avoid self-ligation. The digested vector and insert were analysed on 1% agarose gel and purified by QIAquick gel extraction kit. The purified products were then ligated using T4 DNA ligase at 16ºC for 16 hours followed by heat inactivation of the enzyme at 65ºC for 10min. The ligated products were then transformed into chemically competent *E. coli* DH5α cells using calcium chloride and the transformant colonies were selected on LB plates supplemented with 100µg/ml.
Kanamycin. The colonies were then screened by colony PCR to check for the presence of TEM-1 gene (Fig.2C). The mutated TEM-1 gene sequence was finally confirmed from the selected colonies by Sanger sequencing.

**Results and Discussion**

Our method consisted of two PCR amplification steps (Fig. 1). In the first PCR step, we constructed individual fragments containing targeted mutations (Fig. 1). We designed the gene fragments in such a way that primers for amplifying each fragment contained the desired mutations (Fig. 1). Thus, after the first amplification we obtained gene fragments of variable lengths that contained mutations. In the second step, we assembled these fragments into the whole gene in a single PCR (Fig. 1).

![Figure 1: An outline of the two-step PCR method.](https://mc.manuscriptcentral.com/bmp)

The first PCR step yields ten DNA fragments of various sizes each containing their respective targeted mutations. The primers were designed to contain the mutations and had 12bp overlap with neighbouring primers to enable fragment assembly in the second and final step of the two-step PCR.

We used the gene TEM-1 beta lactamase for demonstration of the Proof-of-concept. We aimed to construct TEM-1 variants containing up to 11 mutations. For TEM-1 gene, we targeted the following amino acid mutations as these have been observed in clinical samples very frequently - L21F, Q39K, M69I, E104K, R164C/S, M182T, G238S, E240K, R244C/S, T265M, and R275L/Q. To introduce these mutations, we divided the TEM-1 gene into 10 fragments and designed forward and reverse primers for amplification of each of these fragments. The first fragment contained the TEM-1 promoter region to the L21 amino
acid residue. Thus, the reverse primer contained the L21F amino acid mutation. We also ensured that the targeted mutations in the primers were succeeded by at least 12 nucleotides at the 3’ end of the primer to enable efficient PCR amplification (Table 3). The second fragment contained the region starting from L21 residue and ending at Q39 residue. The forward primer contained the L21F mutation and the reverse primer contained Q39K mutation. Similarly, we designed primers for amplifications of other fragments (Table 3). When two mutations were too close to each other for making a fragment by PCR, we constructed these mutations in a single fragment using primers containing both mutations. Further, we designed the primers in such a way that the adjoining fragments had 12 base-pair overlap for efficient assembly in the next step (Fig. 1, Table 3).

Table 3 – Sequences of primers used with the mutated bases shown in bold.
Degenerate bases: W= A or T; D= A or G or T; H= A or C or T, Mut238240244_for and Mut238240244_rev primers denote mutations of three amino acid residues, namely, G238S, E240K, and R244C/S

| Primer name | Primer Sequence 5’ to 3’ |
|-------------|--------------------------|
| TEM1_for    | ACGGAATTCCGCGGAACCCCTATTTGTTTATTTTTC |
| TEM1_rev    | ACGGAAGCTTCCAAATGCTTATATCGTACGGCAG |
| L21F_for    | GCGGCATTTTGCTTTCTGTATTGTGT |
| L21F_rev    | AGCAAAAACAGGAAAAGCAAAATGGCAG |
| Q39K_for    | GATGCTGAAGATATAGCTATGAAGTCA |
| Q39K_rev    | TCGTGACCCAACCTATTATCTTCAGCAT |
| M69l_for    | CGTTTTCAATTGATCAAGCACTTTTTTA |
| M69l_rev    | TTTAAAGTGCTGATCATTTGGAAAAC |
| E104K_for   | AATGACTTTGTTAAAGTCACTCAAAGT |
| E104K_rev   | GACTGGTGAGTAGTTAAGAACAAAGCT |
| R164C/S_for | ACTCGCCTTTGATWGGTTGGGAAACGGGAG |
| R164C/S_rev | CTCCGGTTCCCAACWATCAAGGCGAGT |
| M182T_for   | CGTGACACCACGAACGCCCTTATGCAAT |
| M182T_rev   | CATGCTACAGGCGTCGGGCTGTCACAG |
| Mut238240244_for | AAATCTGGAGCCAAGTAAGCGTGGTGCTHGCGGATCAT |
| Mut238240244_rev | TGCAATGATACCGCDAGACCCACGCTTACTTGCTCCAGATT |
| T265M_for   | GTAGTTATCTACATGACGGGGAGTCAG |
| T265M_rev   | CTGACTCCCGTCACTGATGAAATAC |
| R275L/Q_for | ACTATGGATGAACTDAAATAGACAGATCGC |
| R275L/Q_rev | AGCGATCTGCTTATTHGACATCAATG |
We performed the first amplification for 20 cycles using a high-fidelity DNA polymerase (see Methods). This resulted in 10 gene fragments with sizes of 161, 81, 116, 131, 204, 80, 209, 101, 61 and 61 bps respectively (Fig. 2A). We then digested these PCR products with DpnI to remove template DNA and with ExoSAP to hydrolyse excess primers and nucleotides. We then column purified the treated DNA fragments and quantified the concentrations of the purified DNA fragments (in moles/µl) by Qubit Broad Range Assay. Next, we mixed 0.5 picomole and 1 picomole of each fragment in a PCR reaction and performed thermal cycling for 10 cycles without addition of any primers. This enabled annealing of overlapping regions of neighbouring fragments and subsequently allowed gap filling. We then added the terminal primers and performed 15 cycles of PCR amplification. This resulted in assembly of the complete gene of size ~ 1 kb (Fig. 2B).

**Figure 2: Analysis of PCR amplified fragments by agarose gel electrophoresis.** (A) Lanes 1 and 12 - GeneRuler ULR DNA Ladder (Thermo Scientific). Lanes 2-11 - Amplified products of each of the ten fragments carrying the intended mutations after the first PCR (B) Lane 1 - GeneRuler 1kb ladder (Thermo Scientific); Lanes 2 and 3 - Final joined fragments after the second PCR. (C) Lane 1 - GeneRuler 1kb ladder; Lanes 2-10 - PCR amplified products confirming the presence of the whole construct from transformant E. coli colonies.
To confirm the accuracy of our method, we digested the whole gene fragment with restriction enzymes, ligated with appropriately digested vector and transformed into competent *E. coli* cells. We confirmed transformation of the gene fragment by colony PCR (Fig. 2C) and verified the introduced mutations in the TEM-1 gene by Sanger sequencing (Fig. 3). We observed introduction of the targeted mutations at specific sites in the TEM-1 gene (Fig. 3).

**Figure 3 - Confirmation of mutagenesis by Sanger sequencing.** Multiple sequence alignment of the mutated TEM1 gene sequences from three clones show presence of the desired mutations at the targeted sites.

Further, we observed that two factors are critical for accurate and reproducible reconstruction of mutant variants. The first one is the use of clean and purified PCR product obtained from the first PCR for the second amplification step. Second, use of equimolar amounts of products generated by the first PCR in the second assembly step.

We compared different aspects of our method with that of available methods in literature. Our method used relatively small number of amplification cycles and thus had one of the lowest error rates amongst all methods (Table 4). Methods that used even smaller number of amplification cycles required amplification of the vector along with the gene of interest (Wang and Malcolm, 1999; Liu H and Naismith, 2008; Edelheit et al., 2009; Trehan et al.,...
2016; Kuo et al., 2017; Zeng et al., 2018; Table 4). This meant a risk of introducing unwanted mutations in the plasmid backbone which could impact plasmid copy number and antibiotic selection. Changes in plasmid copy number could have unintended critical influence on selection experiments deployed for studies of empirical fitness landscapes. Further, some of these methods generated nicked plasmid as the amplification product which could not be utilized as template in the subsequent amplification cycles (Wang and Malcolm, 1999; Zheng et al., 2004; Liu and Naismith, 2008; Edelheit et al., 2009). This led to use of increased starting template DNA and could lead to formation of hybrid hemi-methylated DNA after amplification that could resist enzymatic digestion (Vovis and Lacks, 1977). This increased the risk of parental DNA carry-over and contamination with the mutants (Liu and Naismith, 2008). Further, some of the published methods required quite complex primer design. For example, Zeng et al. (2018) required four primers for each of the mutants that increased complexity and cost. Some of the other methods required large overlaps that led to long primers again making the process complex and costly.

Thus, our method provides a balanced approach for construction of multi-mutant variants in all aspects compared to the published methods. Our method uses simple molecular biology tools and requires only two PCR amplification steps for construction in contrast to many steps adopted by earlier methods (Wäneskog and Bjerling 2014; Kadkhodaei et al. 2016). Our method is also robust as it can assemble genes from DNA fragments of sizes ranging from 60 to 210 base pairs and do not require any complicated primer design or long primers. Further, our method uses a total of 35 cycles across the two steps and hence, have low chance of introduction of unintended mutations or indels in the gene fragment. Finally, our method does not require amplification of vector and utilize a very small amount of starting template DNA. This also reduces the chance of parental wild-type DNA contamination with mutants and avoids occurrence of unintended mutations in the plasmid backbone. However, the efficiency of our method for assembling larger DNA fragments remains to be tested.
Table 4 – Comparison between the method proposed in this work and the related published methods in terms of the demonstrated capability to introduce multiple mutations in a gene fragment, number of amplification cycles, estimated error rates, risk of plasmid backbone mutations, risk of parental plasmid carry-over and design of oligos. The method proposed here has among the lowest error rates, has low chance of plasmid backbone mutations and has low risk of parental plasmid carry-over.

| Methods | This Method | Wang and Malcom, Biotechnique s (1999) | Hejlesen and Fuchtbauer, BioTechniqu es (2020) | Zeng et al., Scientific Reports (2018) | Kuo et al., Biol Proc Online (2017) | Hallak et al., PLoS ONE (2017) | Kadkhodaei et al., Scientific Reports (2016) | Trehan et al., Scientific Reports (2018) | Wäneskog and Bjerling, Analytical Biochemistry (2014) | Edelheit et al., BMC Biotechnology (2009) | Liu and Naismith; BMC Biotechnology (2008) |
|----------------|-------------|---------------------------------------|-----------------------------------------------|-----------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| Maximum no. of mutations introduced | 11 | 9 | 3 | 15 | 2 | 9 | 2 | 13 | 3 | 2 |
| Amplification Product | Gene of Interest | Whole Plasmid (nicked) | Gene of Interest | Whole plasmid | Part of Plasmid | Gene of Interest | Whole Plasmid (nicked) | Gene of Interest | Whole plasmid (nicked) | Whole plasmid (nicked) |
| No. of PCR amplification Cycles | 1<sup>st</sup> PCR: 20 2<sup>nd</sup> PCR: 15 Total: 35 cycles | 1<sup>st</sup> PCR: 1/3/10 2<sup>nd</sup> PCR: 12-16 Total: 13-26 cycles | 1<sup>st</sup> PCR: 25 2<sup>nd</sup> PCR: 30 Total: 55 cycles | 1<sup>st</sup> PCR: 30 2<sup>nd</sup> PCR: 30 Total: 60 cycles | 14 cycles | 3 step PCR (25 cycles in each step) Total: 75 cycles | 1<sup>st</sup> PCR: 18 2<sup>nd</sup> PCR: 15/25 Total: 33-43 cycles | 30 cycles | 1<sup>st</sup> PCR: 15 2-4<sup>th</sup> PCR: 10×3=30 Total: 45 cycles | 30 cycles |
| Estimated Error Rate | 0.7% (for a 1kb gene) | 0.8-1.5% (for a 3 kb plasmid) | 1.1% (for a 1kb gene) | 3.5% (for a 3 kb plasmid) | 0.8% (for a 3 kb plasmid) | 1.75% (with starter DNA size of 532 bp) | 0.7-0.9% (for a 1kb gene) | 1.8% (for a 3 kb plasmid) | 0.9% (for a 1kb gene) | 1.8% (for a 3 kb plasmid) | 0.7% (for a 3 kb plasmid) |
| Chance of plasmid backbone mutations | No | Yes | No | Yes | Yes | Yes | No | Yes | No | Yes | Yes |
| Size of oligos | 27-42bp (Overlap 12 bp) | 34-73bp | 26-33bp | 25bp (Overlap 6-10 bp) | 44-46bp (Overlap 19-22bp) | 22 bp | 50-78 bp (Overlap 50 bp) | 23 bp (Overlap 17bp) | 31-120 bp | 33-57bp | 39-51bp |
| Template Amount | 0.5 ng | 50-200 ng | 2-5ng | 50-500ng | 10ng | 0.1 ng | 10-50 ng | 100 ng | 100-250 ng | 500ng | 10ng |
| Risk of template carry over | Low | High | Low | Low | Low | Low | Low | Low | Low | High | Low |
| Number and size of fragments | 10; 60-210bp | N/A | 3; 70-3000bp | Variable | N/A | N/A | 9; 304-2191bp | N/A | 13; 63-1800bp | N/A | N/A |

https://mc.manuscriptcentral.com/bmp
Taken together, our work describes a powerful tool for constructions of genotypes at large mutational distances (Fig. 4). Our method can also be adapted to explore genotypes at any mutational distance from the starting genotype by choosing the number of fragments carrying mutations. In addition, one can also mix wild-type and mutated fragments during the second step of assembly, thus enabling construction of a library containing genotypes at local and far-away neighbourhoods. This can eventually help us to systematically reconstruct the long evolutionary paths of proteins and RNA molecules and can transform our understanding of the principles of molecular evolution.

Figure 4 - A schematic genotype map encompassing large mutational distances. The network shows nodes representing genotypes with different phenotypes denoted by different colours. Edges connect nodes (genotypes) that can be reached by one mutation. The distant genotypes can be reached after 10 or more mutations from the starting genotype.
References

Aguilar-Rodríguez J, Payne JL, Wagner A. 2017. A thousand empirical adaptive landscapes and their navigability. *Nat. Ecol. Evol.* 1:45.

Aguilar-Rodríguez J, Peal L, Stella M, Wagner A, Payne JL. 2018. The architecture of an empirical genotype-phenotype map. *Evolution* 72:1242–1260.

Bajić D, Vila JCC, Blount ZD, Sánchez A. 2018. On the deformability of an empirical fitness landscape by microbial evolution. *Proc. Natl. Acad. Sci. U. S. A.* 115:11286–11291.

Bank C, Hietpas RT, Jensen JD, Bolon DNA. 2015. A systematic survey of an intragenic epistatic landscape. *Mol. Biol. Evol.* 32:229–238.

Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Guerousssov S, Lee LJ, Slobođeniuc V, Kutter C, Watt S, Colak R, et al. 2012. The evolutionary landscape of alternative splicing in vertebrate species. *Science* 338:1587–1593.

Bendixsen DP, Collet J, Østman B, Hayden EJ. 2019. Genotype network intersections promote evolutionary innovation. *PLoS Biol.* 17:e3000300.

Bendixsen DP, Østman B, Hayden EJ. 2017. Negative Epistasis in Experimental RNA Fitness Landscapes. *J. Mol. Evol.* 85:159–168.

Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS. 2006. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* 444:929–932.

Bertram J, Masel J. 2020. Evolution Rapidly Optimizes Stability and Aggregation in Lattice Proteins Despite Pervasive Landscape Valleys and Mazes. *Genetics* 214:1047–1057.

Bloom JD, Silberg JJ, Wilke CO, Drummond DA, Adami C, Arnold FH. 2005. Thermodynamic prediction of protein neutrality. *Proc. Natl. Acad. Sci. U. S. A.* 102:606–611.

Bolognesi B, Faure AJ, Seuma M, Schmiedel JM, Tagartlia GG, Lehner B. 2019. The mutational landscape of a prion-like domain. *Nat. Commun.* 10:4162.

Boucher JI, Cote P, Flynn J, Jiang L, Laban A, Mishra P, Roscoe BP, Bolon DNA. 2014. Viewing protein fitness landscapes through a next-gen lens. *Genetics* 198:461–471.

Bush K, Jacoby GA. 2010. Updated functional classification of beta-lactamases. *Antimicrob. Agents Chemother.* 54:969–976.

Canale AS, Cote-Hammarlof PA, Flynn JM, Bolon DN. 2018. Evolutionary mechanisms studied through protein fitness landscapes. *Curr. Opin. Struct. Biol.* 48:141–148.

Cárcamo E, Roldán-Salgado A, Osuna J, Bello-Sanmartín I, Yañez JA, Saab-Rincón G, Viadú H, Gaytán P. 2017. Spiked Genes: A Method to Introduce Random Point Nucleotide Mutations Evenly throughout an Entire Gene Using a Complete Set of Spiked Oligonucleotides for the Assembly. *ACS Omega* 2:3183–3191.

Chevereau G, Dravecká M, Batur T, Guvenek A, Ayhan DH, Toprak E, Bollenbach T. 2015. Quantifying the Determinants of Evolutionary Dynamics Leading to Drug Resistance. *PLoS Biol.* 13:e1002299.

Cróna K, Greene D, Barlow M. 2013. The peaks and geometry of fitness landscapes. *J. Theor. Biol.* 317:1–10.

Das SG, Direito SOL, Waclaw B, Allen RJ, Krug J. 2020. Predictable properties of fitness landscapes induced by adaptational tradeoffs. *Elife* 9: e55155.

Dhar R, Bergmiller T, Wagner A. 2014. Increased gene dosage plays a predominant role in the initial stages of evolution of duplicate TEM-1 beta lactamase genes. *Evolution* 68:1775–1791.

Diss G, Landry CR. 2016. Combining the Dihydrofolate Reductase Protein-Fragment Complementation Assay with Gene Deletions to Establish Genotype-to-Phenotype Maps of Protein Complexes and Interaction Networks. *Cold Spring Harb. Protoc.* 2016.

Diss G, Lehner B. 2018. The genetic landscape of a physical interaction. *Elife* 7: e32472.

Domínguez J, Diss G, Lehner B. 2018. Pairwise and higher-order genetic interactions during the evolution of a tRNA. *Nature* 558:117–121.

Edelheit O, Hanukoglu A, Hanukoglu I. 2009. Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC Biotechnol.* 9:61.

Faber MS, Wrenbeck EE, Azouz LR, Steiner PJ, Whitehead TA. 2019. Impact of In Vivo Protein Folding Probability on Local Fitness Landscapes. *Mol. Biol. Evol.* 36:2764–2777.

Ferretti L, Weinreich D, Tajima F, Achaz G. 2018. Evolutionary constraints in fitness landscapes. *Heredity* 121:466–481.

Filizetti M, Jacquieri H, Schug A, Tenailleon O, Weigt M. 2016. Coevolutionary Landscape Inference and the Context-Dependence of Mutations in Beta-Lactamase TEM-1. *Mol. Biol. Evol.* 33:268–280.
perturbations of protein interactomes reveal their mechanisms of regulation, robustness and insights into genotype-phenotype maps. Brief. Funct. Genomics 15:130–137.

Finnberg E, Labonte JW, Gray JJ, Ostermeier M. 2014. A comprehensive, high-resolution map of a gene’s fitness landscape. Mol. Biol. Evol. 31:1581–1592.

Flynn JM, Rossouw A, Cote-Hammarlof P, Fragata I, Mayor D, Hollins C, Bank C, Bolon DNA. 2020. Comprehensive fitness maps of Hsp90 show widespread environmental dependence. Elife 9: e53810.

Fowler DM, Stephany JJ, Fields S. 2014. Measuring the activity of protein variants on a large scale using deep mutational scanning. Nat. Protoc. 9:2267–2284.

Gamazon ER, Stranger BE. 2014. Genomics of alternative splicing: evolution, development and pathophysiology. Hum. Genet. 133:679–687.

Ge L, Rudolph P. 1997. Simultaneous introduction of multiple mutations using overlap extension PCR. Biotechniques 22:28–30.

Gibson DG. 2011. Enzymatic Assembly of Overlapping DNA Fragments. Methods in Enzymology 498: 349–361.

Gonzalez CE, Roberts P, Ostermeier M. 2019. Fitness Effects of Single Amino Acid Insertions and Deletions in TEM-1 β-Lactamase. J. Mol. Biol. 431:2320–2330.

Goulart CP, Mahmudi M, Crona KA, Jacobs SD, Kallmann M, Hall BG, Greene DC, Barlow M. 2013. Designing Antibiotic Cycling Strategies by Determining and Understanding Local Adaptive Landscapes. PLoS One 8:e65640

Hallak LK, Berger K, Kaspar R, Kwilas AR, Montanaro F, Peeples ME. 2017. Efficient method for site-directed mutagenesis in large plasmids without subcloning. PLoS One 12:e0177788.

Hayden EJ, Ferrada E, Wagner A. 2011. Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. Nature 474:92–95.

Hejlesen R, Füchtbauer EM. 2020. Multiple site-directed mutagenesis via simple cloning by prolonged overlap extension. Biotechniques 68:345-348.

Hietpas R, Roscoe B, Jiang L, Bolon DNA. 2012. Fitness analyses of all possible point mutations for regions of genes in yeast. Nat. Protoc. 7:1382–1396.

Hietpas Ryan T, Bank C, Jensen JD, Bolon DNA. 2013. Shifting fitness landscapes in response to altered environments. Evolution 67:3512–3522.

Kadkhodaei S, Memari HR, Abbasiliasi S, Rezaei MA, Movahedi A, Shun TJ, Ariff A Bin. 2016. Multiple overlap extension PCR (MOE-PCR): an effective technical shortcut to high throughput synthetic biology. RSC Adv. 6:66682–66694.

Kuo TY, Tsai CC, Fu HW. 2017. Enhanced Mutant Screening in One-step PCR-based Multiple Site-directed Plasmid Mutagenesis by Introduction of Silent Restriction Sites for Structural and Functional Study of Proteins. Biol Procoud Online 19:12.

Landry CR, Rifkin SA. 2010. Chromatin regulators shape the genotype-phenotype map. Mol. Syst. Biol. 6:434.

Lee Y, Rio DC. 2015. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu. Rev. Biochem. 84:291–323.

Li C, Qian W, Maclean CJ, Zhang J. 2016. The fitness landscape of a tRNA gene. Science. 352:837–840.
Li C, Zhang J. 2018. Multi-environment fitness landscapes of a tRNA gene. *Nat. Ecol. Evol.* 2:1025–1032.

Li X, Lalici J, Baeza-Centurion P, Dhar R, Lehner B. 2019. Changes in gene expression predictably shift and switch genetic interactions. *Nat. Commun.* 10:3886.

Liu H, Naismith JH. 2008. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* 8:91.

Macken CA, Perelson AS. 1989. Protein evolution on rugged landscapes. *Proc. Natl. Acad. Sci. U. S. A.* 86:6191–6195.

Mehlhoff JD, Ostermeier M. 2020. Biological fitness landscapes by deep mutational scanning. *Methods in Enzymology* 643: 203–224.

Melamed D, Young DL, Miller CR, Fields S. 2015. Combining natural sequence variation with high throughput mutational data to reveal protein interaction sites. *PLoS Genet.* 11:e1004918.

Merkin J, Russell C, Chen P, Burge CB. 2012. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* 338:1593–1599.

Mira PM, Meza JC, Nandipati A, Barlow M. 2015. Adaptive landscapes of resistance genes change as antibiotic concentrations change. *Mol. Biol. Evol.* 32:2707–2715.

Palmer AC, Chait R, Kishony R. 2018. Nonoptimal Gene Expression Creates Latent Potential for Antibiotic Resistance. *Mol. Biol. Evol.* 35:2669–2684.

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

Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ. 2007. Empirical fitness landscapes reveal accessible evolutionary options. *Nature* 445:383–386.

Pressman AD, Liu Z, Janzen E, Blanco C, Müller UF, Joyce GF, Pascal R, Chen IA. 2019. Mapping a Systematic Ribozyme Fitness Landscape Reveals a Frustrated Evolutionary Network for Self-Aminoacylating RNA. *J. Am. Chem. Soc.* 141:6213–6223.

Rodrigues J V., Bershtein S, Li A, Lozovsky ER, Hartl DL, Shakhnovich EI. 2016. Biophysical principles predict fitness landscapes of drug resistance. *Proc. Natl. Acad. Sci. U. S. A.* 113:E1470-8.

Romero PA, Arnold FH. 2009. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* 10:866–876.

Salverda MLM, Koomen J, Koopmanschap B, Zwart MP, de Visser JAGM. 2017. Adaptive benefits from small mutation supplies in an antibiotic resistance enzyme. *Proc. Natl. Acad. Sci. U. S. A.* 114:12773–12778.

Sarkisyan KS, Bolotin DA, Meer M V., Usmanova DR, Mishin AS, Sharonov G V., Ivankov DN, Bozhanova NG, Baranov MS, Soylemez O, et al. 2016. Local fitness landscape of the green fluorescent protein. *Nature* 533:397–401.

Schmiedel JM, Carey LB, Lehner B. 2019. Empirical mean-noise fitness landscapes reveal the fitness impact of gene expression noise. *Nat. Commun.* 10:3180.

Shimotohno A, Oue S, Yano T, Kuramitsu S, Kagamiyama H. 2001. Demonstration of the importance and usefulness of manipulating non-active-site residues in protein design. *J. Biochem.* 129:943–948.

Smith JM. 1970. Natural selection and the concept of a protein space. *Nature* 225:563–564.

Spiller B, Gershenson A, Arnold FH, Stevens RC. 1999. A structural view of evolutionary divergence. *Proc. Natl. Acad. Sci. U. S. A.* 96:12305–12310.

Starr TN, Thornton JW. 2017. Exploring protein sequence-function landscapes. *Nat. Biotechnol.* 35:125–126.

Steinberg B, Ostermeier M. 2016. Shifting Fitness and Epistatic Landscapes Reflect Trade-offs along an Evolutionary Pathway. *J. Mol. Biol.* 428:2730–2743.

Trehan A, Kielbus M, Czapinski J, Stepulak A, Huhntiani I, Rivero-Müller A. 2016. REPLACR-mutagenesis, a one-step method for site-directed mutagenesis by recombineering. *Sci Rep.* 6:19121.

de Visser JAGM, Krug J. 2014. Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.* 15:480–490.

Vovis GF, Lacks S. 1977. Complementary action of restriction enzymes endo R-DpnI and Endo R-DpnII on bacteriophage f1 DNA. *J Mol Biol.* 115:525-38.

Wagner A. 2008. Robustness and evolvability: a paradox resolved. *Proc. R. Soc. B Biol. Sci.* 275:91–100.

Wäneskog M, Bjerling P. 2014. Multi-fragment site-directed mutagenic overlap extension polymerase chain reaction as a competitive alternative to the enzymatic assembly method. *Anal. Biochem.* 444:32–37.
Wang W, Malcolm BA. 1999. Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site-Directed Mutagenesis. *Biotechniques* 26:680-2.

Weinreich DM, Delaney NF, Depristo MA, Hartl DL. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312:111–114.

Yang G, Anderson DW, Baier F, Dohmen E, Hong N, Carr PD, Kamerlin SCL, Jackson CJ, Bornberg-Bauer E, Tokuriki N. 2019. Higher-order epistasis shapes the fitness landscape of a xenobiotic-degrading enzyme. *Nat. Chem. Biol.* 15:1120–1128.

Young L, Dong Q. 2004. Two-step total gene synthesis method. *Nucleic Acids Res.* 32:e59.

Zeng F, Zhang S, Hao Z, Duan S, Meng Y, Li P, Dong J, Lin Y. 2018. Efficient strategy for introducing large and multiple changes in plasmid DNA. *Sci Rep.* 8:1714.

Zheng J, Payne JL, Wagner A. 2019. Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks. *Science* 365:347–353.

Zheng L, Baumann U, Reymond JL. 2004. An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.* 32:e115.
Data Availability Statement
No new data were generated or analysed in support of this research. The sequencing data generated by Sanger sequencing to confirm the clones are shown in the article.

Author contributions
RD conceived the study, SR and AA performed all experiments, SR, AA and RD wrote the manuscript. All authors read and approved the manuscript.

Acknowledgements
Work in the RD lab was funded by an ISIRD grant, IIT Kharagpur and an Early career research (ECR) grant (ECR/2017/002328), Science and Engineering Research Board (SERB), India.

Competing interests
The authors declare no competing interests.
Target Mutations

1st PCR

2nd PCR (Multi-fragment assembly)

Primers

Template DNA

Fragment 1

Fragment 2

Fragment 3

Fragment 10

https://mc.manuscriptcentral.com/bmp
10 or more mutations

Starting genotype

Distant genotypes

10 or more mutations