Making deep mutational scanning accessible: a cost-efficient approach to construct barcoded libraries for genes of any length

Jessica Jann[1-6], Isabelle Gagnon-Arsenault[1-6], Alicia Pageau[1-6], Alexandre K Dubé[1-6], Anna Fijarczyk[1-6], Romain Durand[1-6], Christian R. Landry*[1-6]

Affiliations:

[1]Département de biochimie, de microbiologie et de bio-informatique, Faculté des sciences et de génie, Université Laval, G1V 0A6, Québec, Canada

[2]Institut de biologie intégrative et des systèmes, Université Laval, G1V 0A6, Québec, Canada

[3]PROTEO, Le regroupement québécois de recherche sur la fonction, l'ingénierie et les applications des protéines, Université Laval, G1V 0A6, Québec, Canada

[4]Centre de recherche sur les données massives, Université Laval, G1V 0A6, Québec, Canada

[5]CRI, Centre de recherche en infectiologie, Université Laval, G1V 0A6, Québec, Canada

[6]IID, Institut intelligence et données, Université Laval, G1V 0A6, Québec, Canada

*Correspondence to Christian.landry@bio.ulaval.ca

Abstract

Recent developments in DNA synthesis and sequencing have allowed the construction of comprehensive gene variant libraries and their functional analysis. Achieving high-replication and thorough mutation characterization remains technically and financially challenging for long genes. Here, we developed an efficient, affordable and scalable library construction approach that relies on low-cost DNA synthesis and standard cloning technologies, which will increase accessibility to systematic mutational studies and help advance the field of protein science.

Main text

Understanding how genetic variation alters protein function is a central goal in genetics, molecular biology, and evolutionary and biomedical research. Over the past decade, Deep Mutational Scanning (DMS) has emerged as a powerful technology to systematically evaluate the functional consequences of all possible single amino acid substitutions in a protein within a single experiment [1,2]. By coupling a comprehensive variant library, selection assays, and high-throughput sequencing, DMS allows the systematic quantification and mapping of genotype–phenotype relationships at high resolution and coverage [1,2].

DMS involves four major steps: (i) the construction of a comprehensive variant library, typically introducing all possible single amino acid substitutions into a target gene (or specific locus); (ii) the delivery of this library into a biological system via vectors (such as plasmids) or genomic integration; (iii) the application of a selective pressure (e.g., drug exposure, receptor activation, sorting based protein stability); and (iv) the quantification of variant frequencies before and after selection using deep sequencing, allowing the estimation of functional effects for each substitution. Connecting mutations to phenotypes lead to the construction of mutation-phenotype landscapes that provide insights into diverse biological processes, such as structure-function relationships [3], disease-associated mutations [4], and mechanisms of adaptation [5].

The robustness and completeness of a DMS experiment rely largely on the quality and design of the variant library. Constructing a systematic variant library offers several advantages over random or non-systematic approaches (e.g., error-prone PCR) [6,7]. It ensures exhaustive mutational coverage and removes biases introduced by targeted or region-focused mutagenesis. Furthermore, vector-based variant libraries offer flexibility in downstream applications: they can be easily stored and reused in different genetic backgrounds or biological

systems [8], coupled with fluorescent tags or reporters [9], and they allow flexible expression under constitutive or inducible promoters [10]. Vector-based libraries can be used not only for transformation but also as templates for transferring mutations into genomic loci using CRISPR-Cas9 or homologous recombination [11,12]. Examples of vector-based DMS applications include signaling-specific variants of MC4R in obesity [13], immune escape mutations in the SARS-CoV-2 Spike protein [14], and pathogenicity classifications for CRX and PAX6 in inherited eye disorders [15]. In microbes, recent efforts have used DMS to map resistance-conferring mutations [5,16,17].

While DMS is a powerful tool for high-throughput functional analysis, generating reproducible and high-coverage data remains challenging. Incorporating internal replicates, such as multiple unique codons or DNA barcodes per amino acid variant, reduces technical noise, improves the precision of effect estimates, and enhances sensitivity and reproducibility [18–20]. The use of plasmid-based libraries further facilitates linking each variant to unique short DNA barcodes, enabling cost-effective tracking of mutation frequencies across diverse conditions using short-read sequencing. Similar barcoding strategies have previously been used in CRISPR-based pooled screens to enable high-throughput variant tracking [21].

Although the cost of high-throughput sequencing has steadily declined over the past 20 years, the large size of many target genes makes the cost of such experiments prohibitive. Gene size limits DMS scalability due to the complexity of the variant library and the associated synthesis and sequencing costs, including for linking variants to DNA barcodes. To overcome these challenges, we developed an efficient and cost-effective two-step cloning strategy for plasmid-based DMS libraries that facilitates the study of large genes using DNA-barcoded synthetic oligonucleotide variants.

Our approach involves fragmenting a gene of interest (*Your Favorite Gene: YFG*) into subregions of up to 150 base pairs (bp), in accordance with the synthesis length limits of single-stranded oligonucleotide pools (oPools). oPools are composed of five parts (Figure 1A): (i) ~40 bp of homology upstream of the *YFG* fragment of interest, (ii) a *YFG* fragment sequence containing a single NNK codon substitution, (iii) BsaI restriction sites, (iv) a 30 bp DNA barcode alternating random dinucleotides (NN) with codon position-specific sequences, and (v) a conserved i7 primer binding site (PBS_i7) present in all oPools.

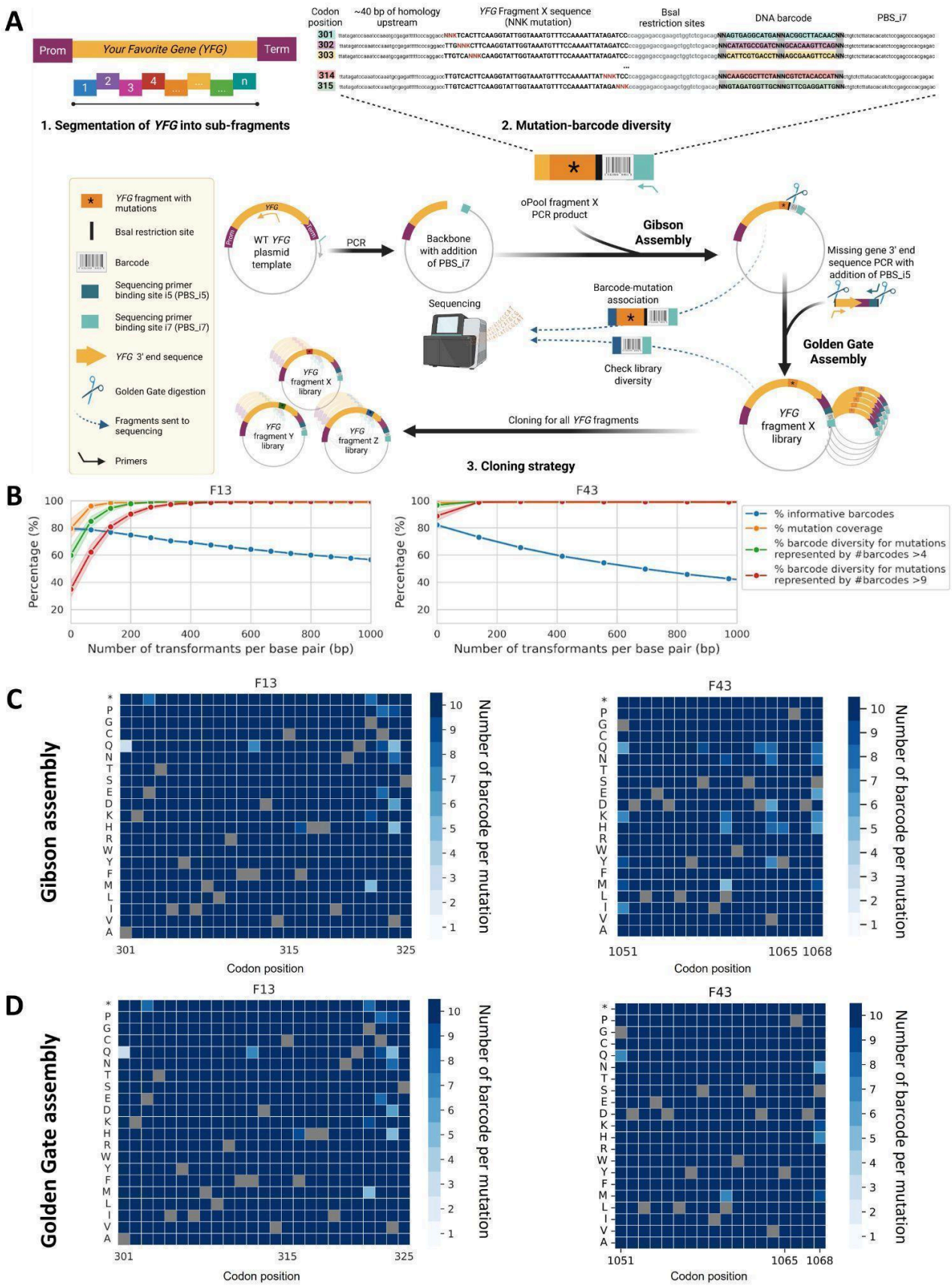

A
Prom
Your Favorite Gene (YFG)
Term
1. Segmentation of YFG into sub-fragments
Codon position
~40 bp of homology upstream
YFG Fragment X sequence (NNK mutation)
BsaI restriction sites
DNA barcode
PBS_i7
2. Mutation-barcode diversity
YFG fragment with mutations
BsaI restriction site
Barcode
Sequencing primer binding site i5 (PBS_i5)
Sequencing primer binding site i7 (PBS_i7)
YFG 3' end sequence
Golden Gate digestion
Fragments sent to sequencing
Primers
WT YFG plasmid template
PCR
Backbone with addition of PBS_i7
oPool fragment X PCR product
Gibson Assembly
Missing gene 3' end sequence PCR with addition of PBS_i5
Sequencing
Barcode-mutation association
Check library diversity
Golden Gate Assembly
YFG fragment X library
YFG fragment Y library
YFG fragment Z library
Cloning for all YFG fragments
3. Cloning strategy
B
F13
F43
Percentage (%)
Number of transformants per base pair (bp)
% informative barcodes
% mutation coverage
% barcode diversity for mutations represented by #barcodes >4
% barcode diversity for mutations represented by #barcodes >9
C
Gibson assembly
Codon position
Number of barcode per mutation
D
Golden Gate assembly

Figure 1 - A) Strategy to generate a DMS plasmid library for *Your Favorite Gene* (*YFG*) using short, degenerate libraries. 1. Segmentation of YFG into sub-fragments, each fragment corresponding to a DNA region to be synthesized. The same approach can be applied to promoter and terminator regions, if desired. **2.** Example of a pool of degenerate oligonucleotides (oPool) derived from one *YFG* fragment associated with DNA barcodes. Each oPool contains: (i) ~40 bp of homology upstream of the *YFG* fragment of interest, (ii) the YFG fragment sequence with a single NNK codon, (iii) BsaI cloning sites, (iv) a DNA barcode composed of codon-position specific regions and six degenerate nucleotides (N), and (v) a conserved i7 primer binding site (PBS_i7) present in all oPools and used for rapid and efficient sequencing library preparation. Current oligonucleotide synthesis technologies allow for a total of nine degenerate positions per fragment: three are used for the degenerate codon (NNK), and six for the barcode. A complete list of all oPool sequences and their detailed composition is provided in Supplementary Table 1. **3.** Protocol for constructing *YFG* DMS plasmid library from oPools using two cloning steps that maintain the physical barcode-mutation association. The libraries of oPools are cloned into the plasmid template by Gibson cloning. Following this step, for each fragment, a necessary short-read sequencing using PBS_i5 (included in the 5' sequencing primer) and PBS_i7 is performed to associate each barcode with its corresponding mutation and to assess both barcode diversity per mutation and mutation coverage for the whole fragment. The ultimate step consists in Golden Gate cloning of the missing 3' gene fragment between the degenerate fragment and the barcode. An additional short-read sequencing step of the barcodes can be performed to make sure that coverage and diversity have been maintained. Figure created in BioRender. Barff, T. (2025) https://BioRender.com/1vfl2on. **B) Optimisation of cloning steps.** Cumulative percentage of informative barcodes (those associated with a single coding mutation and not with the wild-type (WT)), mutation coverage (percentage of mutations represented by at least one informative barcode) and barcode diversity (percentage of mutations represented by informative #barcodes >4 or >9) as a function of the number of transformants recovered after Gibson cloning for two fragments (F13 and F43) of a long gene coding for a transcription factor. Data shown here are derived from the combined results of multiple independent small-scale transformations, and results are normalized to the number of transformants per base pair to facilitate comparison across gene fragments of different lengths. Means and confidence intervals obtained for 100 draws of randomly picked transformants. **Library quality control after (C) Gibson and (D) Golden Gate assemblies for *PDR1* F13 and F43**. Heatmaps show barcode diversity for each possible amino acid substitution at each codon position. For each fragment, a total of 25,000 and 100,000 transformants were recovered

and analyzed after Gibson and Golden Gate assemblies, respectively. Covered mutations are mutations represented by #barcode >0, while barcode diversity represented by #barcode >4 or >9 per mutation demonstrates an increasing number of replicates for a same mutation. Grey squares represent WT amino acids. In the heatmaps, the number of barcodes per mutation is clipped at 10. For a complete visualization of barcode diversity data, untrimmed heat maps are available (Supplementary Figures 1 and 2).

The integration of a barcode facilitates variant identification while allowing each codon and thus each amino acid substitution to be represented by multiple distinct barcodes, thereby providing internal replicates for downstream analysis. The barcodes and their corresponding variants remain physically linked on the same DNA molecule (plasmid), ensuring accurate tracking throughout the experiment.

This strategy utilizes the cost-efficient oPools, which are ~100-fold cheaper per base than double-stranded DNA synthesis (<$0.001 vs. ~$0.05 per base) [22,23]. In addition, low-cost short-read sequencing is used (i) to associate random barcodes to their corresponding coding sequence variants and (ii) to track mutations in downstream experiments via the barcodes. This approach offers a cost-effective alternative to methods that require long-read sequencing to variant identification and tracking [24,25].

The cloning strategy is divided into two steps: a Gibson assembly, followed by a Golden Gate assembly (Figure 1A). In the first step, the plasmid backbone is generated via PCR from a yeast expression vector containing the wild-type (WT) *YFG*. Gibson assembly allows inserting the oPools (variant fragment + barcode + PBS_i7) into the backbone. These constructs then serve as backbone plasmids for the second cloning step.

In the second step, the missing *YFG* 3′ end sequence is amplified, including the C-terminal coding region, terminator, and PBS_i5. Golden Gate assembly enables the one-pot assembly of this missing *YFG* 3′ end sequence into the Gibson-assembled plasmids, using BsaI digestion

and vector-insert ligation cycles (see *Methods* section). This strategy results in one distinct plasmid library per *YFG* fragment, each containing a full-length *YFG* sequence carrying a single mutation (in the targeted fragment) specifically linked to a DNA barcode flanked by the PBS_i5 and PBS_i7.

Each library undergoes quality control by high-throughput sequencing at both cloning steps (Figure 1):

- After Gibson assembly, a necessary short-read sequencing (mutated fragment + barcode) is used to (i) establish barcode–mutation associations and (ii) assess both mutation coverage and barcode diversity per mutation.
- After Golden Gate assembly, only the DNA barcodes are sequenced (facultative) to confirm that coverage and diversity have been maintained.

Our method is unique but shares some key elements with other methods previously developed, while being significantly more affordable. For instance, Jones et al. [26] (and slightly modified by Howard et al. [27]) also employed a two-step approach to link barcodes to variants, thereby overcoming the need for long-read sequencing. Their approach relies on pools of individual oligonucleotides with the specific sequences to be screened, i.e., without degenerate positions. The variants are linked to a random DNA barcode (15N) following their synthesis using a degenerate oligonucleotide, and this occurs before the missing gene fragment is inserted. There are major differences with our approach. The gene variant synthesis relies on more expensive gene synthesis technology. Second, the approach utilizes a random barcode, which increases the complexity of the library and potentially increases the fraction of low-complexity barcodes that can cause issues during downstream sequencing. Our design enables a stratified library with a codon-specific barcode, which helps reduce sequence complexity and allows easily tracking position and fragments easily for quality control. This is made feasible because the barcodes and mutated gene fragments are located on the same DNA molecules. Another

approach that allows for the use of affordable oPools is SUNi mutagenesis [28]. This protocol employs a completely different method for mutagenesis, using degenerate pools of oligonucleotides that anneal to single-stranded plasmids to introduce variants. One noticeable difference between this approach and others is the proportion of mutants successfully recovered, which is lower than what we achieve, as the authors report 77% and 69% of mutation coverage for two independent libraries. In addition, this protocol does not use DNA barcodes that would bypass the need for tile-sequencing, particularly for long genes, although it could arguably be adapted to do so. Finally, DIMPLE [29] relies on oligonucleotides generated using DNA microarrays, which likely improves coverage compared to SUNi, but at a significantly higher synthesis cost. One similarity between this approach and ours is the use of Golden Gate cloning, specifically for assembling the mutated gene fragments synthesized with full-length gene backbones. The advantage of DIMPLE is the introduction of other types of mutations, such as insertions and deletions (indels). We did not consider them in our current library, but we have successfully used oPools for making indel libraries [30], so our protocol could also include such mutations. In addition to synthesis cost, one significant difference between our protocol and DIMPLE is that it does not allow for rapid and affordable library barcoding during gene fragment synthesis.

To validate the feasibility and the strength of our approach, DMS libraries for all 43 fragments of a long gene coding for a transcription factor were constructed and characterized (Figure 1B, C and D, and Supplementary Figures 1 and 2). Even if the information on all fragments are provided, we focus here on fragments F13 and F43 because they cover a wide range and distances from the DNA barcodes and F13 contains mutations of interest. The performance of a DMS plasmid library depends on four key parameters: Informative barcodes (i.e., barcodes associated with only one coding mutation and not with the WT), which ensure data usability and limit sequencing costs of non informative barcodes; Mutation coverage (i.e., percentage of

mutations represented by at least one informative barcode); Barcode diversity (i.e., percentage of mutations represented by informative #barcodes >4 or >9), which provides internal replicates; Variant uniformity (i.e., how evenly individual variants are represented within the library, ensuring balanced abundance and accurate quantitative comparisons).

These parameters are largely dependent on the number of recovered transformants after each cloning step. To better estimate this relationship, we combined sequencing data obtained from multiple independent small-scale transformations and normalized the results to the number of transformants per base pair, allowing direct comparison across fragments of different lengths. A random sampling of resulting distributions indicates that after Gibson assembly, ~350 transformants per base pair are needed to optimize sufficient mutation coverage, barcode diversity, and informative barcode (Figure 1B). This level of recovery yields 300-350 informative barcodes covering all possible mutations, with over 99.8% of mutations represented by #barcodes >4 and over 97.0% by #barcodes >9 (Figure 1B and C). Altogether, the findings demonstrate excellent mutation coverage and barcode diversity, two essential features for the robustness and reproducibility of DMS data.

As each cloning step inevitably results in some diversity loss, the number of transformants recovered after Golden Gate assembly was increased fourfold (~1400 transformants per bp). Although a small subset of mutations showed lower representation, our cloning strategy ensured over 99.8% of mutation coverage and preserved high barcode diversity: over 93.6% of mutations were still covered by #barcodes >4, and over 68.6% by #barcodes >9 (Figure 1D).

The uniformity of variant representation was assessed after both the Gibson and the Golden Gate steps using the Gini coefficient, an index of distribution bias ranging from 0 (perfect uniformity) to 1 (maximum bias) [31]. We observed that the Golden Gate assemblies showed a slight increase in the Gini coefficient compared to the Gibson assemblies (Supplementary

Figures 3 and 4), reflecting a minimal bottleneck related to multi-fragment cloning. Overall, uniformity remained high across all 43 fragments (Gini coefficients consistently <0.5).

To demonstrate the feasibility and utility of our library construction strategy, we focused on antifungal resistance, an increasingly pressing issue due to the rising incidence of invasive fungal infections and the limited number of effective treatment options [32]. Drug resistance mutations significantly compromise the efficacy of antifungal therapies, underscoring the urgent need for innovative approaches to investigate and counteract resistance mechanisms [33]. Among current antifungal drugs, azoles are widely used in clinical settings, including posaconazole, a broad-spectrum triazole [34–37]. The transcription factor Pleiotropic Drug Resistance 1 (Pdr1) plays a key role in mediating multidrug resistance [38]. Even if it regulates many downstream pathways, *PDR1* itself is not essential for yeast viability. However, mutations in *PDR1* lead to overexpression of downstream targets, including ABC efflux pump genes like *CDR1, CDR2,* and *SNQ2*, conferring broad-spectrum drug resistance in both the model yeast *Saccharomyces cerevisiae* and the human pathogen *Nakaseomyces glabratus* (*Candida glabrata*) [39,40]. Although numerous studies have demonstrated the central role of *PDR1*, the precise identity and functional impact of resistance-conferring mutations remain largely unknown. Thus, our cloning strategy targeted *PDR1*, a large protein of 1,068 amino acids, whose comprehensive mutagenesis would have been financially and technically challenging using conventional methods. To demonstrate the accuracy of our method, we focused on fragment 13 (F13; residues 301–325) of Pdr1. This fragment combines both unknown and well-characterized regions associated with drug resistance [38], including previously identified and experimentally validated mutations [40–44].

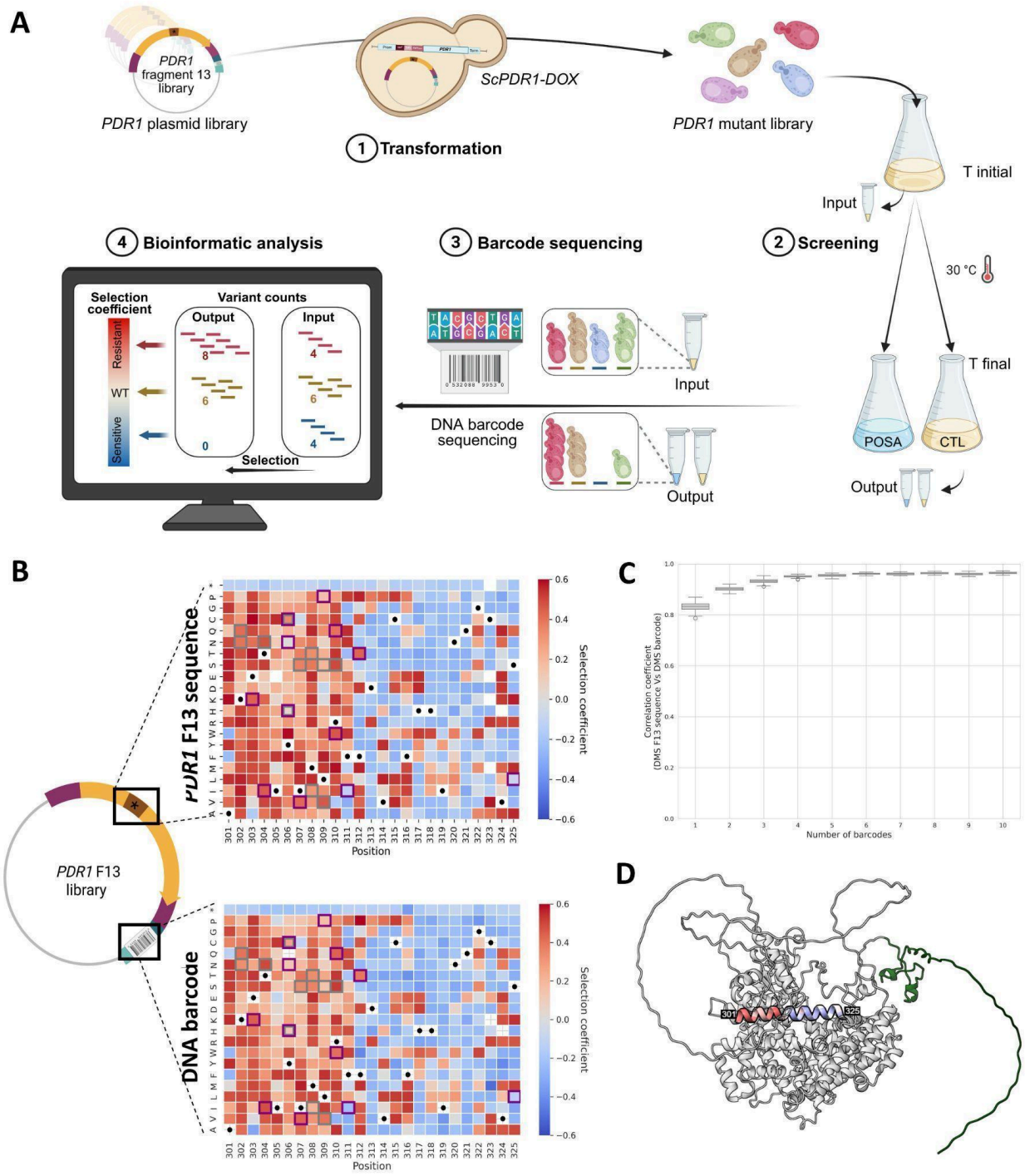


Figure 2 - Demonstration of the high-accuracy mutation profile obtained through barcode sequencing of a DMS plasmid library built from oPools. A) Pooled competition workflow using our barcoded DMS plasmid library targeting a drug resistance hotspot. 1. The *S. cerevisiae PDR1* variant library for F13 (residues 301-325) was transformed into a *S. cerevisiae* strain with a doxycycline-repressible endogenous *PDR1* expression (*ScPDR1-DOX*). **2.**

Transformed strain expressing the variant library was grown in the presence of posaconazole (POSA, 0.40 µg/mL) or under drug-free control (CTL) conditions. **3.** High-throughput short-read sequencing of DNA barcodes. **4.** Selection coefficient analysis to identify resistant variants. Figure created in BioRender. Barff, T. (2025) https://BioRender.com/4wpf522. **B) Azole resistance landscape of *PDR1* F13.** Heatmaps of selection coefficients for the 525 single amino acid variants screened, based on either direct sequencing of the mutated F13 region (top) or DNA-barcode inference (bottom). Squares in purple (*N. glabrata*) and grey (*S. cerevisiae* or *S. paradoxus*) highlight azole resistance mutations previously reported [38]. Positive selection coefficients (>0, red) indicate increased fitness relative to the WT, while negative values (<0, blue) indicate reduced fitness. Black dots mark WT amino acids, and asterisks represent stop codons (as internal controls). Each value is the median of synonymous codons per amino acid. White cells indicate missing data. **C) Distribution of correlation coefficients between the selection coefficients measured with direct sequencing of the mutated F13 coding region and DNA-barcode inference**, as a function of the number of barcodes per variant considered. Each boxplot summarizes 100 random subsamplings per variant, performed using a standardized resampling procedure to normalize barcode counts across variants. **D) Pdr1 tertiary structure with DMS azole resistance landscape on F13 region.** Predicted tertiary structure of *S. cerevisiae* Pdr1 (AlphaFold 3 prediction, pTM = 0.72) shown in grey with its Zn_2Cys_6 DNA binding domain (residues 1-78) in green. The median selection coefficients from DMS are mapped onto the F13 (residues 301-325).

The *PDR1* F13 variant library was used to perform high-throughput functional screening of all possible amino acid substitutions, linking genotype to antifungal resistance phenotypes (Figure 2). The DMS experiment consisted of four main steps (Figure 2A): (i) generation of the *PDR1* variant library by transforming the plasmid library into a *S. cerevisiae* strain with doxycycline-repressible genomic *PDR1* expression (*ScPDR1-DOX*), (ii) competition assay with posaconazole, (iii) high-throughput sequencing, and (iv) fitness analysis of each variant.

In this system, the endogenous *PDR1* locus is placed under the control of a doxycycline-repressible promoter, allowing conditional regulation of its genomic expression. Although *PDR1* is not essential for yeast viability, its activity strongly influences azole resistance phenotypes. The design allows conditional inhibition of genomic *PDR1* expression and

functional complementation with the variant libraries generated in this study, as it's shown in the Supplementary Figure 5. With this system in place, we next evaluated the performance of barcode-based sequencing for accurate quantification of variant fitness.

In principle, only short-read sequencing of DNA barcodes would be required. Here, we also performed full-length sequencing of the F13 fragment to compare the accuracy of the selection coefficients estimated from the barcode and from the mutated fragment directly, as is routinely done for small genes [45–47] or through tile-seq for longer genes [5,48]. The estimates from the barcodes or from the F13 coding region were strongly correlated (r = 0.97, all available barcodes, Supplementary Figure 6). This high level of concordance was also observed with a much smaller number of barcodes (four barcodes per mutation (r = 0.91)), confirming the robustness of our cloning and barcoding strategy (Figure 2C). Only a few discrepancies between barcode- and sequence-based estimates were observed (8 out of 525 mutations), most of which involving mutations supported by few barcodes (≤ 5) and showing neutral-to-sensitive fitness effects. By analysing selection coefficients across all amino acid substitutions, we identified two distinct regions within fragment F13 with different contributions to resistance (Figure 2B and D). Residues 301–312 showed frequent gain-of-resistance mutations and residues 313–324 appeared to play a limited role in resistance, with fewer substitutions conferring resistance. These results closely match resistance-conferring mutations reported in the literature for Pdr1 in both *S. cerevisiae* and the closely related species *N. glabratus* (Figure 2B; e.g., T304N, I307S, M308I, R310S, etc.) [38], supporting the power of our cloning strategy for mapping resistance landscapes.

In summary, we developed an efficient, scalable, and cost-effective two-step cloning strategy for the construction of high-quality, barcoded DMS plasmid libraries for genes of any length. By combining affordable oPools, DNA barcodes, and optimized Gibson and Golden Gate assemblies, while using short-read sequencing to minimize financial costs, our method ensures

high mutation coverage and barcode diversity that provide robust internal replicates, key parameters for accurate and reproducible mutational effects on fitness or resistance. This approach reduces both technical and financial barriers, facilitating its broader application across diverse gene sizes, biological systems, and experimental contexts. By supporting comprehensive genotype-phenotype analyses, our strategy strengthens functional genomics and mechanistic insights, and enhances the accessibility of these technologies across laboratories.

Methods

General information for plasmid and strain constructions and media

All cloning procedures were performed using *Escherichia coli* strain MC1061. *Saccharomyces cerevisiae* strain R1158, described in [49], served as the background for generating the doxycycline-repressible *S. cerevisiae PDR1* strain (*ScPDR1*-DOX, details below).

All antimicrobials were prepared as 1,000× stock solutions: ampicillin (AMP), doxycycline (DOX), geneticin (G418), and nourseothricin (NAT) were dissolved in sterile water, filtered, and stored at −20 °C; posaconazole (POSA) and itraconazole (ITR) were dissolved in dimethyl sulfoxide (DMSO) and also stored at −20 °C.

For bacterial liquid cultures, *E. coli* was grown in LB medium (0.5% yeast extract, 1% tryptone, 1% NaCl) at 37 °C with shaking at 250 rpm. For solid cultures, cells were grown on 2YT agar plates (1% yeast extract, 1.6% tryptone, 0.5% NaCl, 0.2% glucose, 2% agar) and incubated at 37 °C. When required, AMP was added at a final concentration of 100 µg/mL for plasmid selection. *S. cerevisiae* was cultured in YPD medium (1% yeast extract, 2% tryptone, 2% glucose) at 30 °C with shaking at 250 rpm. For solid cultures, 2% agar was added to YPD. Yeast transformants with pRS31N plasmids were selected on YPD plates supplemented with 100 µg/mL G418 and 100 µg/mL NAT.

All PCRs were performed using KAPA High-Fidelity HotStart polymerase unless otherwise specified. Strains, reagents, plasmids, computational tools and additional resources are listed in Supplementary Table 2. Oligos are listed in Supplementary Table 3, and PCR conditions, in Supplementary Table 4.

All transformations were performed using *Escherichia coli* NEB® 5-alpha Competent Cells (High Efficiency) following the manufacturer's recommended heat-shock protocol. Briefly, 2–5 µL of

assembly reaction (Gibson or Golden Gate; see below) was added to 50 µL of chemically competent cells, followed by incubation on ice for 30 min, a 30-second heat shock at 42 °C, and immediate cooling on ice for 5 min. Cells were then recovered in 450 µL of SOC medium (0.5% Yeast Extract, 2% Vegetable Peptone, 0.5% Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose) and incubated at 37 °C for 60 min with shaking (250 rpm) before plating on selective agar plates. The complete transformation protocol is available from New England Biolabs [50].

The integrity of all assembled and mutagenized plasmids was verified either by Sanger sequencing (Plateforme de séquençage et de génotypage des génomes, Centre de recherche du centre hospitalier de Québec–Université Laval (CRCHUL), Canada) or by whole plasmid sequencing using Oxford Nanopore Technology with custom analysis and annotation (Plasmidsaurus, USA or Flow Genomics, Canada).

Plasmid library design

pRS31N-ScPDR1 WT

S. cerevisiae PDR1 (*PDR1*) was expressed from a yeast centromeric plasmid pRS31N-*ScPDR1* that was constructed as follows. First, a plasmid containing the *PDR1* sequence was created, including its promoter and terminator. The plasmid pRS31N containing AMP and NAT resistance markers was doubly digested with 20 U of both HindIII-HF and XbaI restriction enzymes. The insert, a fragment containing *S. cerevisiae PDR1* native promoter (834 bp), coding sequence (3,207 bp) and terminator (345 bp), was amplified from the pMoBY-*ScPDR1* plasmid with primers adding homology to the pRS31N plasmid cloning site (PCR 1). The resulting PCR product corresponding to the insert was incubated for 1h30 at 37 °C with 10 U of DpnI enzyme to remove parental DNA. Subsequently, the plasmid and insert were purified on magnetic beads and assembled by Gibson assembly to obtain pRS31N-*ScPDR1*.

As a BsaI restriction site is used in the construction of the *PDR1* deep mutational scanning (DMS) plasmid library, we had to remove three BsaI restriction sites in pRS31N-*ScPDR1*. The three sites were mutated by site-directed mutagenesis (site 1) or by homologous recombination with a short double-stranded DNA fragment (gBlock) corresponding to *PDR1* coding sequence between positions 2238 and 2720 without BsaI restriction sites (sites 2 and 3). The BsaI site 1 was removed by performing site-directed mutagenesis based on the QuickChange Site-Directed Mutagenesis System (Stratagene, La Jolla, CA). We amplified the pRS31N-*ScPDR1* plasmid using a pair of primers containing the desired mutation at the center (PCR 2). The PCR product was then incubated for 1h30 at 37 °C with 10 U of DpnI enzyme to remove parental DNA, and pRS31N-*ScPDR1* mutated plasmids (pRS31N-*ScPDR1* minus site 1) were retrieved directly by transformation in bacteria. For sites 2 and 3, the plasmid pRS31N-Sc*PDR1* minus site 1 was first PCR amplified (PCR 3). The resulting PCR product was incubated for 1h30 at 37 °C with 10 U of DpnI enzyme to remove parental DNA and then, purified on magnetic beads. Finally, the plasmid and gBlock insert were assembled by Gibson assembly to obtain pRS31N-*ScPDR1* WT. For clarity, throughout this study, pRS31N-*ScPDR1* WT refers to the plasmid after the BsaI site replacements, which served as the template for constructing the whole *PDR1* DMS plasmid library.

oPools

The DMS library was generated using only the *PDR1* coding sequence (CDS). Due to the limited length of synthetic DNA and to enable barcode-mutation association with short-read sequencing, the 3,204 bp *PDR1* gene was divided into 43 fragments of 75 bp each, except the last one which is of 54 bp. Fragment 1 goes from codon 1 to 25, fragment 2 goes from residue 26 to 50, and so on, up to the last fragment 43 (the smallest) going from residue 1050 to 1068. For each fragment, a library was built using affordable pools of degenerate oligonucleotides (oPools). oPools are composed of five parts designed as follows (Figure 1A):

I. ~40 bp of homology upstream of the *PDR1* fragment of interest;

II. *PDR1* sequence of the fragment containing a single NNK codon (32 codons, including a stop codon);

III. BsaI module (two BsaI sites separated by CCGAAGCT to avoid self-dimerization); and

IV. 30 bp DNA barcode with six random N and two fixed sequences (12 bp each) for each codon position. This design facilitates the identification of the mutated codon while allowing multiple distinct barcodes to represent the same mutation. As a result, each unique mutation is linked to several barcodes, providing internal replicates for downstream DMS analyses. oPools currently only allow for nine degenerate positions per DNA fragment. Three are used for the codon and six for the barcode;

V. the sequencing PBS_i7 site (CTGTCTCTTATACACATCTCCGAGCCCACGAGAC), which is common to all oPools and used for rapid and efficient short-read sequencing library preparation.

To generate the barcodes, we use the following Python script: github.com/lzamparo/DNAbarcodes, which generates all possible 12-nucleotide barcode sequences. From these, 30,000 barcodes with a GC content between 40 and 60% and a hamming distance of 3 were selected. We then removed barcodes in which BsaI was present to prevent any problem in the Golden Gate assembly step of our protocol. We also removed barcodes with two identical bases at the beginning or at the end to prevent stretches of the same bases. We split our filtered barcodes based on their GC content. We generated all individual 30 bp barcodes for the oPools by combining two 12 bp barcodes with six degenerate nucleotides, NN+barcode1+NN+barcode2+NN. We made sure the GC content of the sum of the two barcodes would be ~50%.

All oPool sequences were then generated using a custom Jupyter notebook (opool_generator_20240212.ipynb). We cut the *PDR1* coding sequence into 75 bp fragments (25 codons per fragment) and we identified the 40-90 bp before and the 4 bp after each fragment. For each fragment, we generated 25 variant oligonucleotides in which a single codon was replaced by an NNK degenerate codon. Each oPool contains 25 different oligonucleotide sequences, each targeting one specific codon. The final sequences were generated by combining the 40-90 bp before the fragment, the NNK sequence, the 4 bp after the fragment, the BsaI module (GGAGACCgaagCTGGTCTCGACAG), one 30 bp barcode and the PBS_i7 (CTGTCTCTTATACACATCTCCGAGCCCACGAGAC).

All oPools are flanked by homology sequences at both the 5′ end (40 bp upstream of the target fragment within the *PDR1* sequence) and the 3′ end (PBS_i7 sequence, facilitating rapid sequencing library preparation), for a total of 186 to 258 bp per oPool. The complete list of oPool sequences used in this study is provided in Supplementary Table 1.

Because cloning applications require double-stranded DNA inserts, oPools were amplified using two PCR reactions. The first PCR consisted of a single cycle using only the reverse primer (PCR4). Amplicons were purified using magnetic beads at a bead:DNA ratio of 1.8 to enrich for full-length oligonucleotides and to minimize the incorporation of truncated sequences (linked to oPool incomplete synthesis) into the final product. Then, 2 µL of the purified product was used as template for a second PCR, involving both forward and reverse primers and limited to five cycles to reduce PCR-related errors (PCR5). Each oPool amplification was performed in four replicates, which were pooled after the second PCR to maximize the diversity of variants carried into downstream cloning steps. This approach is intended to improve the overall quality of the oPools, ensuring sufficient insert quantity and diversity for subsequent steps of DMS library construction.

Plasmid library construction

Cloning strategy of the DMS plasmid library (Figure 1)

The cloning strategy was divided into two steps: Gibson assembly followed by Golden Gate assembly.

In the first step, the plasmid backbone was amplified from pRS31N-*ScPDR1* WT (PCR 6). The amplified region includes homology arms corresponding to the sequence upstream of the targeted fragment and extends to the plasmid region downstream of the terminator (i7 region). The resulting linear PCR product was incubated for 1h30 at 37 °C with 10 U of DpnI enzyme to remove parental plasmid DNA and then, purified using magnetic beads. The oPools are amplified as described in the "*oPools"* section. Backbone and insert fragments were assembled using Gibson assembly protocol [51]. To reach a target of ~25,000 transformants per fragment, necessary to ensure good barcode diversity and mutation coverage (Figure 1), two parallel Gibson reactions (10 µL each) followed by ~5 transformations (3 µL of Gibson reaction per transformation) using NEB 5-alpha Competent *E. coli* cells were performed for each fragment. Following transformations, 5 mL of 2YT medium was added directly onto each agar plate, and colonies were scraped using a sterile glass spreader. All transformants corresponding to a given fragment were pooled together. Aliquots were prepared, each corresponding to an optical density (OD_{600}) of 20 per fragment, and stored at -80 °C without (cell pellet) or with (glycerol stock) 15% glycerol. Plasmid libraries were extracted and purified from one cell pellet aliquot per fragment. These purified plasmids served as backbones for the subsequent cloning step.

In the second step, the *PDR1* missing 3′ end region, including the 3′ end coding sequence, terminator, and PBS_i5 (Figure 1A), was amplified (PCR 7). BsaI recognition sites were added to both ends of the amplicon to generate non-palindromic overhangs compatible with those in the Gibson-assembled plasmids. The resulting PCR product was incubated for 1h30 at 37 °C

with 10 U of DpnI enzyme to remove parental DNA and then, purified on magnetic beads. The previously prepared plasmid libraries served as vectors, and the *PDR1* missing 3′ region acted as inserts for Golden Gate assembly, a one-pot cloning method alternating between BsaI digestion and ligation (PCR 8). To ensure sufficient barcode diversity and adequate mutation coverage (Figure 1), approximately 100,000 transformants per fragment were needed. To achieve this, two parallel Golden Gate reactions (20 µL each) and ~10 transformations (2.5 µL of Golden Gate reaction per transformation) were performed for each fragment using NEB 5-alpha Competent *E. coli* cells, according to the manufacturer's instructions. After transformation, 5 mL of 2YT medium was added directly to each agar plate, and colonies were scraped using a sterile glass spreader. All transformants corresponding to a given fragment were pooled together. Aliquots were prepared, each corresponding to an optical density (OD_{600}) of 80 per fragment, and stored at -80 °C without (cell pellet) or with (glycerol stock) 15% glycerol. Plasmid libraries were extracted and purified from one cell pellet aliquot per fragment. The resulting plasmid preparations corresponded to the final *PDR1* DMS plasmid libraries, with a distinct library for each *PDR1* fragment.

Plasmid library quality control by sequencing

At each cloning step, library quality was assessed by high-throughput sequencing, focusing on parameters such as mutation coverage and barcode diversity per mutation (Figure 1 - C and D).

Sequencing library preparation

To perform sequencing-based quality control after Gibson assembly, we started with 20 ng of plasmid library obtained from bacteria minipreps. Sequencing libraries were prepared via three PCRs. The first two PCRs (PCR9 and PCR10) amplified the fragment of interest along with the downstream DNA barcode, while simultaneously adding the PBS_i5 adapter sequence at the 5′ end. The forward primer included the PBS_i5 and 20 bp upstream of the targeted region, while

the reverse primer corresponded only to the PBS_i7. The amplicons were purified, diluted 1:1,000 and used as a template for a third PCR (PCR11), adding the appropriate Illumina i5 and i7 sample indexes for multiplexed sequencing. For each sample, the last PCR was performed in triplicate to maintain the diversity of the sequenced sample. All amplicons from the same sample were pooled and purified using magnetic beads. All samples were sent for high-throughput sequencing (sample description in Supplementary Table 5). Although this approach ensures robust amplification of the library and sufficient sequencing depth, it does not directly account for potential amplicon sequencing biases. Such biases can occur when random DNA molecules are over-amplified compared to others, which can cause significant issues for downstream quantitative analyses. Incorporating Unique Molecular Identifiers (UMIs) during PCR when preparing sequencing libraries would help mitigate the possible occurrence of such biases.

Following Golden Gate assembly, only the DNA barcode region was sequenced, in order to verify that barcode diversity and mutation coverage were maintained. Because the PBS_i5 and PBS_i7 were already present on each side of the DNA barcode, sequencing libraries after Golden Gate assembly were generated in a single PCR (PCR12), which added the appropriate Illumina i5 and i7 sample indexes. Like after Gibson assembly, this PCR was also performed in triplicate per sample to prevent sampling bottlenecks. All replicate amplicons were pooled and purified on magnetic beads prior to sequencing. All samples were then sent for high-throughput sequencing (sample description in Supplementary Table 6).

Sequencing platform

High-throughput sequencing was performed in paired-end 100 bp or 150 bp at the Centre de recherche du CHU de Québec – Université Laval sequencing platform (CHUL, Canada) using an Illumina NovaSeq 6000 S4 system. Libraries were sequenced at an average depth of ~250 reads per expected variant to assess barcode diversity and mutation coverage.

Sequencing read processing

To analyse quality control after Gibson assembly, first, low-quality reads were removed with *fastp* using default parameters -q 15 (the minimum phred score quality value that a base is qualified equal to 15) and -u 40 (40% of bases allowed to be unqualified). Reads were then merged with *bbmerge* from *bbmap*. All subsequent processing was done using custom Python scripts available at https://github.com/Landrylab/Jann_et_al_2025. To remove potential sequencing errors, the following reads were discarded: reads with incorrect length (reads including insertions or deletions), reads with more than one mutated codon within the *PDR1* coding region (75 bp), reads with unexpected mutations within barcode sequences, and reads with missing nucleotides within the *PDR1* or barcode sequence. After filtering, more than 60% and 80% of all reads were retained for F13 and F43, respectively.

Mutations with coverage of less than two reads were discarded. To measure the impact of combining results from multiple transformations (16 reactions, each with 5,000 (F13) or 7,500 (F43) transformants), we calculated the cumulative % of different metrics across all transformations sorted in a randomized manner. Means and confidence intervals were obtained for 100 random draws of transformation reactions. Metrics included: (i) informative barcodes, (ii) mutation coverage, (iii) barcode diversity: #barcodes >4 per mutation, and (iv) barcode diversity: #barcodes >9 per mutation. To standardize data, all metrics have been converted to the number of transformants/bp.

As done following the Gibson assembly, raw reads obtained after Golden Gate assembly product sequencing were filtered using *fastp* (default parameters), and overlapping paired-end reads were merged using *bbmerge.* Reads with incorrect length, unexpected barcode sequences, or missing nucleotides were removed, along with single-count reads. Barcode-mutation associations were performed by merging barcode reads from the Golden

Gate assembly with the reference association generated after the Gibson assembly. To evaluate the library quality after the Golden Gate step, we re-calculated key metrics: informative barcodes, mutation coverage, and barcode diversity.

To evaluate the quality and representation of variants in the constructed libraries, we quantified their uniformity using two established metrics. First, the Gini coefficient, an index of distribution bias ranging from 0 (perfect uniformity) to 1 (maximum bias) [31], was calculated to assess the equal representation of variants after each cloning step (Gibson and Golden Gate). Second, we used the uniformity score proposed by [28], defined as the logarithmic difference between the 90th and 10th percentiles of the read count for each variant per fragment, where a small value indicates greater uniformity. These analyses were performed on the plasmid sequencing data to confirm a balanced distribution of variants before functional selection (Supplementary Figures 3 and 4; Supplementary Table 7).

The presence of all expected mutations, with excellent DNA barcode diversity per mutation was confirmed in libraries F13 and F43 (after Gibson assembly: all mutations are covered by #barcodes >4 and after Golden Gate assembly: ~93.6-100% of mutations are covered by #barcodes >4) (Figure 1C and D). A small number of mutations, specifically those at residues 322 to 324, were detected at lower read counts. Nevertheless, they still showed high mutation coverage (98.3%) and acceptable barcode diversity, with 71.7% of mutations with #barcodes > 4 (Figure 1D), which is the threshold required to have over 90% accuracy in selection coefficients (Figure 2C).

Although the main text presents in detail only the characterization of the mutant libraries for two representative *PDR1* fragments (F13 and F43), we also carried out, in parallel, the construction of a comprehensive plasmid mutant library covering the entire *PDR1* coding sequence (total of

43 libraries, one for each fragment of 25 aa). This extended library was generated using the same two-step cloning strategy described above.

Analysis of the complete 43-fragment library confirmed strong performance across all quality metrics. Nearly all expected mutations (~98,4%) were recovered, with only a few absent positions likely attributable to synthesis dropout in the oPools. Barcode diversity per mutation remained high throughout all libraries: after Gibson assembly, ~97,5% mutations were represented by #barcodes >4, and after Golden Gate assembly this remained true for ~95.4% of positions (Supplementary Figures 7 and 8; Supplementary Table 8). Uniformity analysis further indicated that variant abundance across the library (Gini coefficient ~0.32) was well balanced and acceptable for downstream DMS experiments (Supplementary Table 8).

We additionally quantified the representation of non-programmed variants and found that ~10% of reads corresponded to the WT sequence and ~10% to double-mutants (Supplementary Table 8). These species can either be excluded during downstream fitness estimation or leveraged for dedicated analyses of epistatic interactions, depending on the experimental aims.

To support reproducibility and provide a broader overview of the experimental workflow, the results and quality metrics for each cloning stage of the full-length *PDR1* library are presented in Supplementary Table 8 and Supplementary Figures 7 and 8. In addition, we provide a detailed estimated experimental timeline for the construction of the full 43-fragment *PDR1* mutant library in Supplementary Table 9. This resource is intended to serve as a practical reference for researchers aiming to apply or adapt such an approach in future DMS projects.

Quality control of final plasmid libraries

To assess the accuracy of the cloning strategy and evaluate potential PCR-induced mutations, we performed Whole Plasmid Sequencing of six plasmids derived from independent fragments assembly by Golden Gate. No unintended mutations were present in the *PDR1* promoter, coding sequence, or terminator regions. This quality-control step validates the reliability of our cloning approach. Although pre-sequencing of all PCR-amplified fragments prior to Golden Gate assembly could further ensure accuracy, it was not performed here due to cost and scalability considerations.

Variant library construction

Because *PDR1* is not essential for *S. cerevisiae* viability, we replaced its endogenous promoter with a doxycycline-repressible promoter (tetO7). It allows inhibition of genomic *PDR1* expression and complementation by plasmid-encoded variant libraries. We chose this approach after observing low transformation efficiency when attempting to transform a yeast strain harboring a complete *PDR1* deletion.

Yeast strain construction (Supplementary Figure 5A)

The *ScPDR1*-DOX strain was generated based on the Yeast Tet-Promoters Hughes Collection (yTHC) [52]. The endogenous promoter of *PDR1* (spanning 50 bp upstream of the ATG codon; chrVIII:121683–121733) was replaced with a tetracycline-regulatable promoter. The KANMX-tetO7 cassette was amplified from plasmid pKB33 (PCR13) and introduced into competent *S. cerevisiae* R1158 cells using an adapted lithium acetate transformation protocol [53]. Transformants were selected on YPD medium containing 200 µg/mL G418.

Validation of genomic PDR1 expression control in ScPDR1-DOX (Supplementary Figure 5B)

To validate the control of genomic *PDR1* expression in the *ScPDR1*-DOX strain, *PDR1* expression levels were quantified in the presence and absence of doxycycline (DOX) in the culture medium. For this purpose, the gene coding for a monomeric enhanced Green Fluorescent Protein (*mEGFP*) was fused to the *PDR1* coding sequence in the genome of *ScPDR1*-DOX strain. The mEGFP sequence along with an hygromycin resistance module was amplified from plasmid pFA-*mEGFP-HPHNT1* (see construction below) using primers introducing homology arms targeting the 3′ end of the *PDR1* CDS and its terminator (PCR 14). Amplicon was treated with 20 U of DpnI for 1 h at 37 °C, purified using magnetic beads, and used to transform *ScPDR1*-DOX competent cells using the standard lithium acetate transformation protocol [53].

Yeast strains (*ScPDR1*-GFP-DOX and *ScPDR1*-DOX) were cultured overnight at 30 °C with shaking in synthetic complete (SC) medium in 24 deep-well plates. The SC medium contained 0.17% yeast nitrogen base without amino acids and ammonium sulfate, 0.1% monosodium glutamate, 2% glucose and amino acids drop-out without tryptophan. Saturated cultures were diluted to an OD_{600} of 0.15 in fresh medium with or without DOX (10 µg/mL), and grown again under agitation at 30 °C until reaching an OD_{600} of 0.5. Cells were then diluted to ~500 cells/µL in 0.2 µm filtered distilled water for flow cytometry analysis. GFP fluorescence was measured using a Guava easyCyte flow cytometer (MilliporeSigma) with excitation at 642 nm (GRN-B channel, 525/30 filter). A preliminary gating step (FSC-H (Forward Scatter Height) <15000 and FSC-A (Forward Scatter Area) <15000) was applied to select the main population of morphologically normal cells and to exclude non-representative events. A total of 5,000 events were recorded per replicate. Three independent biological replicates were analyzed per condition. Fluorescence signals were processed using a custom Python script and all the data (including control lacking GFP) are shown in Supplementary Figure 5B.

In order to generate *ScPDR1*-GFP-DOX yeast strain, we first had to construct pFA-*mEGFP-HPHNT1* plasmid. This was done by inserting *mEGFP* gene (insert 1) and *ENO1* terminator fragment (insert 2) into pFA-*hphNT1* (vector) using Gibson assembly. All three fragments were PCR amplified (PCR15, PCR16 and PCR17, respectively) and treated with 20 units of DpnI for 1 h at 37 °C before being purified on magnetic beads.

Yeast strain validation by complementation (Supplementary Figure 5C)

In order to validate whether *PDR1* expressed from the plasmid complements its genomic counterpart, we selected a condition where *PDR1* expression is beneficial (in presence of antifungal) and assessed the growth of *ScPDR1*-DOX strain transformed either with an empty plasmid or with one containing *PDR1* WT sequence. After adjusting cell density of precultures to an OD_{600} of 1, three serial 1/5 dilutions were prepared in 200 µL of water (i.e., 40 µL of cells in 160 µL of water). Then, 5 µL of each dilution were spotted on YPD + NAT agar plates supplemented with either 0.1% DMSO (control), 1 µg/mL ITR, 2 µg/mL ITR or 4 µg/mL ITR, in the presence or not of DOX (10 µg/mL). Plates were incubated at 30 °C for 48 h before imaging.

Library in yeast

The DMS plasmid library was introduced into *ScPDR1*-DOX yeast cells via lithium acetate transformation, performed fragment by fragment using a standard protocol [53]. For each fragment, approximately 100,000 colonies were recovered to ensure sufficient library coverage and barcode diversity. To collect transformants, 5 mL of YPD medium was added directly to each plate, and colonies were scraped using a sterile glass spreader. All transformants corresponding to a given fragment were pooled together. Aliquots were prepared, each corresponding to an optical density (OD_{600}) of 80 per fragment, and stored at −80 °C in 25% glycerol.

Pooled competition assay

A pooled competition assay was performed using the *PDR1* F13 variant library, in the presence or absence of an antifungal. All cultures were grown in YPD medium supplemented with NAT (100 µg/mL), to maintain plasmid selection, and DOX (10 µg/mL), to repress the endogenous *PDR1* promoter. As the used antifungal compound was dissolved in DMSO, the control condition also included 0.1% DMSO. Posaconazole (POSA) was used at a final concentration of 0.40 µg/mL, corresponding to the concentration that inhibits approximately 50% of WT growth, as previously reported [5].

The *PDR1* F13 variant library was first pre-cultured overnight in YPD + NAT, starting with an estimated inoculum of ~10,000 cells per variant (Time Point 0, TP0). Saturated cultures were diluted to an initial OD_{600} of 0.05 into fresh YPD + NAT + DOX + POSA, and grown until an OD_{600} of 0.80 was reached (TP1). A second dilution and growth cycle under the same conditions was performed (OD_{600} = 0.05 to 0.80), corresponding to a total of approximately eight mitotic generations (TP2). Identical conditions were used for the control (DMSO) condition.

At TP0, TP1 and TP2, cells were harvested by collecting 5 OD_{600} units per sample in 24 deep-well plate, followed by centrifugation at 916 × g for 5 min. Pellets were resuspended in 1 mL of YPD, transferred to 1.5 mL microcentrifuge tubes, and centrifuged again at 600 × g for 2 min. Supernatants were removed by aspiration, and resulting pellets were stored at −80 °C until plasmid extraction.

Plasmid recovery

Plasmid recovery from yeast was performed using the Zymoprep Yeast Plasmid Miniprep II kit, with the following modified protocol based on [5,24] to improve yield. Frozen cell pellets were resuspended in 200 µL of Solution 1, followed by the addition of 30 U of Zymolyase 20T. Cells

were gently vortexed and incubated at 37 °C for 2 h. Following incubation, tubes were briefly vortexed and frozen at −80 °C for at least 20 min (up to overnight). After thawing at 37 °C for 3 min, 200 µL of Solution 2 was added and mixed, followed by the addition of 400 µL of Solution 3. Tubes were centrifuged at maximum speed for 3 min, and the supernatant (~800 µL) was transferred to a Zymo-Spin I column. Columns were washed with 550 µL of ethanol-containing wash buffer and centrifuged for 1.5 min at maximum speed. After removing the flow-through, columns were spun for an additional min to dry. Plasmids were eluted by adding 10 µL of Tris buffer (10 mM, pH 8.0) and centrifuging for 1 min at maximum speed. Eluted DNA was stored at −20 °C until further processing.

High-throughput sequencing

To validate whether DNA-barcode inference mutations corresponded accurately to the actual mutations present in the F13 sequence of *PDR1*, two types of library were generated: one containing the F13 region of *PDR1* and one consisting of the barcodes.

Sequencing from yeast minipreps was performed using a protocol similar to that described in the *Plasmid library quality control by sequencing* section. Starting from 3 µL of yeast miniprep, an initial PCR was carried out to amplify only *PDR1* from the plasmid, using Phusion High-Fidelity DNA Polymerase and M13 primers flanking the *PDR1* insert on pRS31N (PCR18). This step avoids sequencing genomic *PDR1* and ensures that the same sequencing material could be used for both F13 region and barcode-based analyses, allowing a direct and unbiased comparison between the two datasets.

In a standard barcode-based DMS experiment, DNA-barcode could be directly amplified using the PBS_i5 and PBS_i7 flaking sites, which would simplify library preparation and improve sequencing data quality.

To generate the F13 sequencing libraries, two rounds of PCR were then performed. Starting with 2 µL of M13-purified amplicons, a first PCR made with Phusion High-Fidelity DNA Polymerase (PCR19) was used to amplify the F13 region while introducing PBS_i5 and PBS_i7 sequences at the 5′ and 3′ ends, respectively. In addition, a 0 to 3 nucleotides spacer (N) was added to the primers to increase sequencing diversity. A second PCR (PCR20) was then used to add the appropriate Illumina i5 and i7 sample indexes, as described in the *Plasmid library quality control by sequencing* section.

To generate the DNA barcode sequencing libraries, 2 µL of the M13-purified amplicons were used as a template for a single PCR reaction to add the appropriate Illumina i5 and i7 indexes (PCR21), as described in the *Plasmid library quality control by sequencing* section.

Both the F13 and DNA barcode sequencing libraries were sent for high-throughput sequencing in paired-end 150 bp at the Centre de recherche du CHU de Québec – Université Laval sequencing platform (CHUL, Canada) using an Illumina NovaSeq 6000 S4 flow cell (sample description in Supplementary Table 10). All samples were sequenced to achieve a minimum of 500 reads per expected variant (i.e., ~0.4 million reads per sample for the F13 libraries and ~4 million reads per sample for the DNA barcode libraries). Only samples from timepoints TP0 and TP2 were sequenced. TP1 was excluded, as the limited frequency changes observed after just four generations hindered robust genotype-phenotype associations.

Data analysis

Selection coefficient and variant characterization

Selection coefficients were obtained with *gyōza*, a Snakemake-based workflow for the analysis of deep mutational scanning (DMS) data [54]. The pipeline includes adapter trimming with *Cutadapt*, read merging using *PANDAseq*, and clustering of identical reads via *VSEARCH*.

Singleton reads (i.e., sequences observed only once) were excluded to reduce noise from sequencing errors. A custom script then aggregated read statistics across all processing steps.

The F13 samples were processed using the 'codon' mode of gyōza v1.1.8 [54], with constant 20 bp sequences on either side of the mutated locus specified for trimming. For the barcoded library, the 'barcode' mode of gyōza v1.1.8 was used instead, by providing the dataframe of barcode-variant associations. Briefly, (i) all singletons (read count of 1) were discarded, (ii) $\log_2$ fold changes of allele frequencies were calculated between TP0 and TP2, (iv) selection coefficients were calculated by normalizing with the number of mitotic generations (to account for different growth rates across cultures and make sure functional impact scores reflect the per-generation selective advantage), and (v) selection coefficients were further normalized by subtracting the median $\log_2$ fold-change of silent variants (excluding the WT nucleotide sequence).

Variants were assigned confidence scores (as defined by gyōza) based on their initial read count at TP0 across replicates. A minimum threshold of five reads per nucleotide sequence was used to define high-confidence variants, which were retained for downstream analysis. Selection coefficients were then aggregated by amino acid position, averaging across synonymous codons (Supplementary Tables 11 and 12). The resulting tables were used to classify mutations from resistant to sensitive (0.6 to -0.6) based on their deviation from the WT. Finally, heatmaps were generated using the selection coefficient for each variant to enable clear visualization and comparison between F13 region and DNA barcode sequencing (Figure 2B and Supplementary Figure 9).

To assess the quality of the barcoded library, we compared the final selection coefficients per mutation obtained from sequencing of either mutated F13 region or DNA-barcode inference. To evaluate how barcode diversity influences the reliability of these coefficients, we progressively

increased the number of barcodes required per mutation (from 1 to 10) and, at each level, performed 100 random samplings. To ensure consistent comparisons across variants with different barcode counts, we implemented a standardized subsampling procedure during correlation analysis. If a variant had fewer barcodes than the target number (*n*), some barcodes were randomly resampled multiple times to reach *n*. Conversely, if a variant had more than *n* barcodes, a random subset of *n* barcodes was selected. This approach allowed us to normalize barcode representation across variants and to accurately assess correlation metrics based on 100 independent subsamplings. For each sampling, we calculated the correlation between selection coefficients inferred from DNA barcodes and those obtained from the F13 region (Figure 2C). An analysis of the correlation stratified by the number of barcodes per variant confirms the robustness of our strategy, even for variants associated with a lower barcode count (Supplementary Figure 9).

Prediction and visualization of Pdr1 tertiary structure (Figure 2D)

As no experimentally determined structure was available for Pdr1 in the Protein Data Bank (PDB, RCSB), we used the AlphaFold 3 server [55] to predict its tertiary structure (pTM score = 0.72; seed = 264428788). The amino acid sequence of *S. cerevisiae* Pdr1 was obtained from UniProt (accession: P12383). The predicted structure was visualized using ChimeraX [56].

Data and materials availability

All data are available in the main text or the supplementary materials. Sequencing data are available in the NCBI Sequence Read Archive (SRA) under BioProject PRJNA1269095. All code and processed data to reproduce the analyses, results and figures are available at https://github.com/Landrylab/Jann_et_al_2025.

Acknowledgements

We thank Dan Evans-Yamamoto, Philippe Després and other Landrylab members for discussions and feedback on the design of the cloning strategy.

Funding

This work was supported by a Genome Québec and Genome Canada grant (6569), a Canadian Institutes of Health Research (CIHR) Foundation grant (387697), a CIHR project grant (202409PJT), and a Natural Sciences and Engineering Research Council of Canada (NSERC) CREATE grant (EvoFunPath) to CRL. JJ and RD are supported by postdoctoral fellowships from NSERC. JJ benefits also from FRQS postdoctoral fellowship (343898). CRL holds the Canada Research Chair in Cellular Systems and Synthetic Biology.

Authors' contributions

AKD, JJ, IGA, CRL designed the experimental strategy; JJ & IGA performed experiments; JJ, IGA, AP, AF and RD performed the data analysis; JJ wrote the manuscript with input from IGA and CRL; All authors read and commented on the manuscript; CRL and IGA supervised and administered the project; CRL obtained funding.

References

1. Fowler DM, Fields S. Deep mutational scanning: a new style of protein science. Nat Methods. 2014;11: 801–807.

2. Wei H, Li X. Deep mutational scanning: A versatile tool in systematically mapping genotypes to phenotypes. Front Genet. 2023;14: 1087267.

3. Escobedo A, Voigt G, Faure AJ, Lehner B. Genetics, energetics, and allostery in proteins with randomized cores and surfaces. Science. 2025;389: eadq3948.

4. Tabet DR, Coté AG, Lancaster MC, Weile J, Rayhan A, Fotiadou I, et al. The functional landscape of coding variation in the familial hypercholesterolemia gene LDLR. Science. 2025; eady7186.

5. Bédard C, Gagnon-Arsenault I, Boisvert J, Plante S, Dubé AK, Pageau A, et al. Most azole resistance mutations in the Candida albicans drug target confer cross-resistance without intrinsic fitness cost. Nat Microbiol. 2024;9: 3025–3040.

6. Kitzman JO, Starita LM, Lo RS, Fields S, Shendure J. Massively parallel single-amino-acid mutagenesis. Nat Methods. 2015;12: 203–6, 4 p following 206.

7. Wrenbeck EE, Klesmith JR, Stapleton JA, Adeniran A, Tyo KEJ, Whitehead TA. Plasmid-based one-pot saturation mutagenesis. Nat Methods. 2016;13: 928–930.

8. Stiffler MA, Hekstra DR, Ranganathan R. Evolvability as a function of purifying selection in TEM-1 β-lactamase. Cell. 2015;160: 882–892.

9. Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, et al. High-resolution mapping of protein sequence-function relationships. Nat Methods. 2010;7: 741–746.

10. Matreyek KA, Starita LM, Stephany JJ, Martin B, Chiasson MA, Gray VE, et al. Multiplex assessment of protein variant abundance by massively parallel sequencing. Nat Genet. 2018;50: 874–882.

11. Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M, et al. Accurate classification of BRCA1 variants with saturation genome editing. Nature. 2018;562: 217–222.

12. Sadhu MJ, Bloom JS, Day L, Siegel JJ, Kosuri S, Kruglyak L. Highly parallel genome variant engineering with CRISPR-Cas9. Nat Genet. 2018;50: 510–514.

13. Howard CJ, Abell NS, Osuna BA, Jones EM, Chan LY, Chan H, et al. High resolution deep mutational scanning of the melanocortin-4 receptor enables target characterization for drug discovery. 2025. doi:10.7554/elife.104725.2

14. Dadonaite B, Brown J, McMahon TE, Farrell AG, Figgins MD, Asarnow D, et al. Spike deep mutational scanning helps predict success of SARS-CoV-2 clades. Nature. 2024;631: 617–626.

15. Shepherdson JL, Granas DM, Li J, Shariff Z, Plassmeyer SP, Holehouse AS, et al.

Mutational scanning of CRX classifies clinical variants and reveals biochemical properties of the transcriptional effector domain. Genome Res. 2024;34: 1540–1552.

16. Després PC, Cisneros AF, Alexander EMM, Sonigara R, Gagné-Thivierge C, Dubé AK, et al. Asymmetrical dose responses shape the evolutionary trade-off between antifungal resistance and nutrient use. Nat Ecol Evol. 2022;6: 1501–1515.

17. Romanowicz KJ, Resnick C, Hinton SR, Plesa C. Exploring antibiotic resistance in diverse homologs of the dihydrofolate reductase protein family through broad mutational scanning. Sci Adv. 2025;11: eadw9178.

18. Kinsler G, Schmidlin K, Newell D, Eder R, Apodaca S, Lam G, et al. Extreme sensitivity of fitness to environmental conditions: Lessons from #1BigBatch. J Mol Evol. 2023;91: 293–310.

19. Maes S, Deploey N, Peelman F, Eyckerman S. Deep mutational scanning of proteins in mammalian cells. Cell Rep Methods. 2023;3: 100641.

20. Fowler DM, Adams DJ, Gloyn AL, Hahn WC, Marks DS, Muffley LA, et al. An Atlas of Variant Effects to understand the genome at nucleotide resolution. Genome Biol. 2023;24: 147.

21. Wong ASL, Choi GCG, Cui CH, Pregernig G, Milani P, Adam M, et al. Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. Proc Natl Acad Sci U S A. 2016;113: 2544–2549.

22. Maciá Valero A, Prins RC, de Vroet T, Billerbeck S. Combining oligo pools and Golden Gate cloning to create protein variant libraries or guide RNA libraries for CRISPR applications. Methods Mol Biol. 2025;2850: 265–295.

23. Kuiper BP, Prins RC, Billerbeck S. Oligo pools as an affordable source of synthetic DNA for cost-effective library construction in protein- and metabolic pathway engineering. Chembiochem. 2022;23: e202100507.

24. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. Cell. 2020;182: 1295–1310.e20.

25. Weile J, Ferra G, Boyle G, Pendyala S, Amorosi C, Yeh C-L, et al. Pacybara: accurate long-read sequencing for barcoded mutagenized allelic libraries. Bioinformatics. 2024;40. doi:10.1093/bioinformatics/btae182

26. Jones EM, Lubock NB, Venkatakrishnan AJ, Wang J, Tseng AM, Paggi JM, et al. Structural and functional characterization of G protein-coupled receptors with deep mutational scanning. Elife. 2020;9. doi:10.7554/eLife.54895

27. Howard CJ, Abell NS, Osuna BA, Jones EM, Chan LY, Chan H, et al. High-resolution deep mutational scanning of the melanocortin-4 receptor enables target characterization for drug discovery. Elife. 2025;13. doi:10.7554/eLife.104725

28. Mighell TL, Toledano I, Lehner B. SUNi mutagenesis: Scalable and uniform nicking for efficient generation of variant libraries. PLoS One. 2023;18: e0288158.

29. Macdonald CB, Nedrud D, Grimes PR, Trinidad D, Fraser JS, Coyote-Maestas W. DIMPLE: deep insertion, deletion, and missense mutation libraries for exploring protein variation in evolution, disease, and biology. Genome Biol. 2023;24: 36.

30. Aubé S, Dubé AK, Landry CR. Adaptation is less accessible through mutations in promoters than in coding sequences when large effect sizes are needed. bioRxiv. 2025. doi:10.1101/2025.09.09.675213

31. Mateyko N, de Boer CG. Culture wars: Empirically determining the best approach for Plasmid library amplification. ACS Synth Biol. 2024;13: 2328–2334.

32. Fisher MC, Alastruey-Izquierdo A, Berman J, Bicanic T, Bignell EM, Bowyer P, et al. Tackling the emerging threat of antifungal resistance to human health. Nat Rev Microbiol. 2022;20: 557–571.

33. Berman J, Krysan DJ. Drug resistance and tolerance in fungi. Nat Rev Microbiol. 2020;18: 319–331.

34. Langner S, Staber PB, Neumeister P. Posaconazole in the management of refractory invasive fungal infections. Ther Clin Risk Manag. 2008;4: 747–758.

35. Boutin C-A, Luong M-L. Update on therapeutic approaches for invasive fungal infections in adults. Ther Adv Infect Dis. 2024;11: 20499361231224980.

36. Maquera-Afaray J, Luna-Vilchez M, Salazar-Mesones B, Portillo-Alvarez D, Uribe-Ramirez L, Taipe-Sedano G, et al. Antifungal prophylaxis with posaconazole in immunocompromised children younger than 13 years. J Pediatr Pharmacol Ther. 2022;27: 57–62.

37. Pardiwala A, Datarkar AN, Manekar V, Daware S. Posaconazole as single or adjunctive antifungal in mucormycosis of the maxilla: systematic review and meta-analysis. J Oral Med Oral Surg. 2023;29: 41.

38. Bédard C, Pageau A, Fijarczyk A, Mendoza-Salido D, Alcañiz AJ, Després PC, et al. FungAMR: A comprehensive portrait of antimicrobial resistance mutations in fungi. bioRxiv. 2024. doi:10.1101/2024.10.07.617009

39. Gale AN, Pavesic MW, Nickels TJ, Xu Z, Cormack BP, Cunningham KW. Redefining pleiotropic drug resistance in a pathogenic yeast: Pdr1 functions as a sensor of cellular stresses in Candida glabrata. Mitchell AP, editor. mSphere. 2023;8. doi:10.1128/msphere.00254-23

40. Ferrari S, Ischer F, Calabrese D, Posteraro B, Sanguinetti M, Fadda G, et al. Gain of function mutations in CgPDR1 of Candida glabrata not only mediate antifungal resistance but also enhance virulence. PLoS Pathog. 2009;5: e1000268.

41. Mizoguchi H, Watanabe M, Nishimura A. Characterization of a PDR1 mutant allele from a clotrimazole-resistant sake yeast mutant with improved fermentative activity. J Biosci Bioeng. 1999;88: 20–25.

42. Carvajal E, van den Hazel HB, Cybularz-Kolaczkowska A, Balzi E, Goffeau A. Molecular and phenotypic characterization of yeast PDR1 mutants that show hyperactive transcription of various ABC multidrug transporter genes. Mol Gen Genet. 1997;256: 406–415.

43. Bautista C, Gagnon-Arsenault I, Utrobina M, Fijarczyk A, Bendixsen DP, Stelkens R, et al. Hybrid adaptation is hampered by Haldane's sieve. Nat Commun. 2024;15: 10319.

44. Conway TP, Vu BG, Beattie SR, Krysan DJ, Moye-Rowley WS. Similarities and distinctions in the activation of the Candida glabrata Pdr1 regulatory pathway by azole and non-azole drugs. bioRxivorg. 2024. doi:10.1101/2024.09.19.613905

45. Cisneros AF, Gagnon-Arsenault I, Dubé AK, Després PC, Kumar P, Lafontaine K, et al. Epistasis between promoter activity and coding mutations shapes gene evolvability. Sci Adv. 2023;9: eadd9109.

46. Dibyachintan S, Dubé AK, Bradley D, Lemieux P, Dionne U, Landry CR. Cryptic genetic variation shapes the fate of gene duplicates in a protein interaction network. Nat Commun. 2025;16: 1530.

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

48. Weile J, Sun S, Cote AG, Knapp J, Verby M, Mellor JC, et al. A framework for exhaustively mapping functional missense variants. Mol Syst Biol. 2017;13: 957.

49. Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, et al. Functional discovery via a compendium of expression profiles. Cell. 2000;102: 109–126.

50. New England Biolabs. High Efficiency Transformation Protocol for 96-tube format (C2987U). [cited 9 Nov 2025]. Available: https://www.neb.com/en-ca/protocols/2015/11/24/high-efficiency-transformation-protocol-for-96-tube-format-c2987u

51. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA 3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009;6: 343–345.

52. Mnaimneh S, Davierwala AP, Haynes J, Moffat J, Peng W-T, Zhang W, et al. Exploration of essential gene functions via titratable promoter alleles. Cell. 2004;118: 31–44.

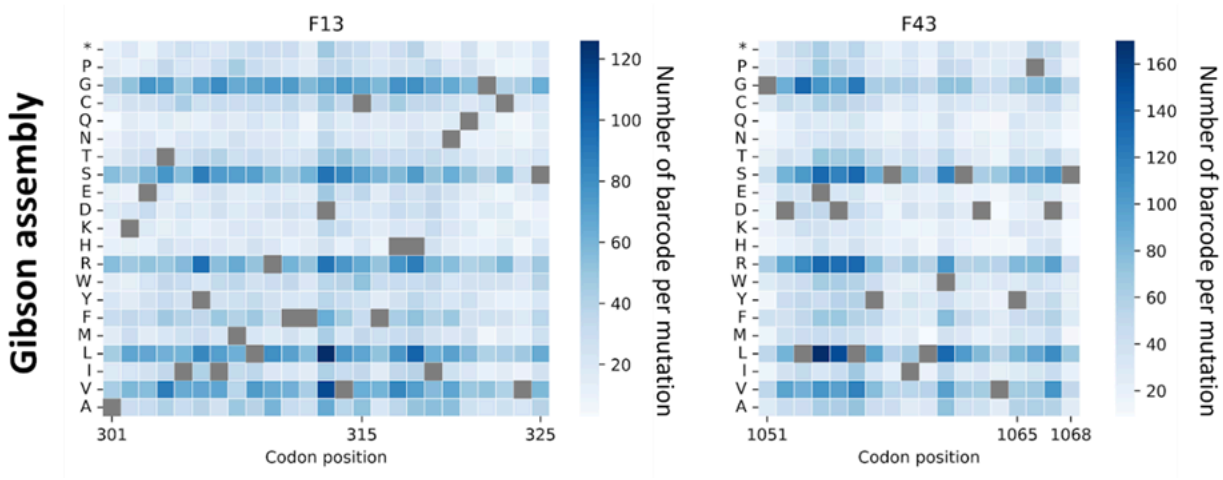
53. Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc. 2007;2: 31–34.

54. Durand R, Pageau A, Landry CR. gyōza: a Snakemake workflow for modular analysis of deep-mutational scanning data. bioRxiv. 2025. doi:10.1101/2025.02.19.639168

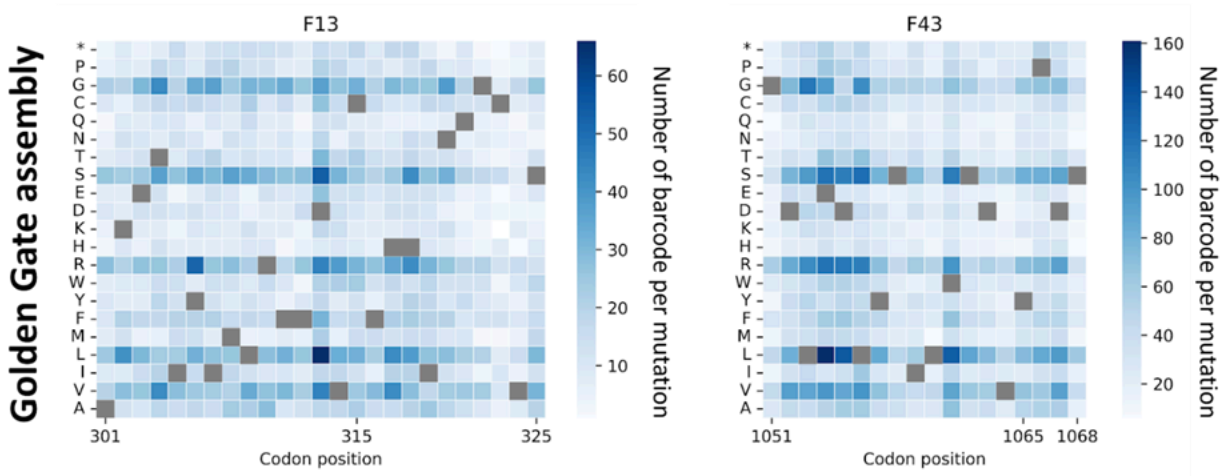
55. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630: 493–500.

56. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, et al. UCSF ChimeraX: Tools for structure building and analysis. Protein Sci. 2023;32: e4792.

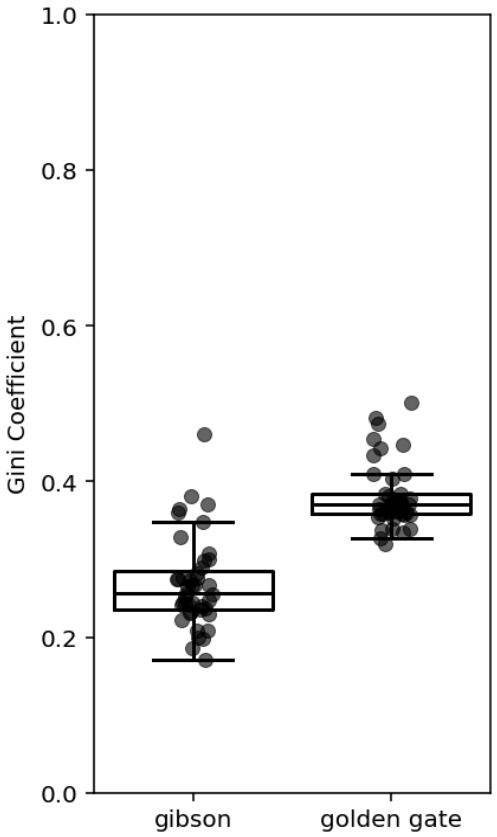
Supporting Information:



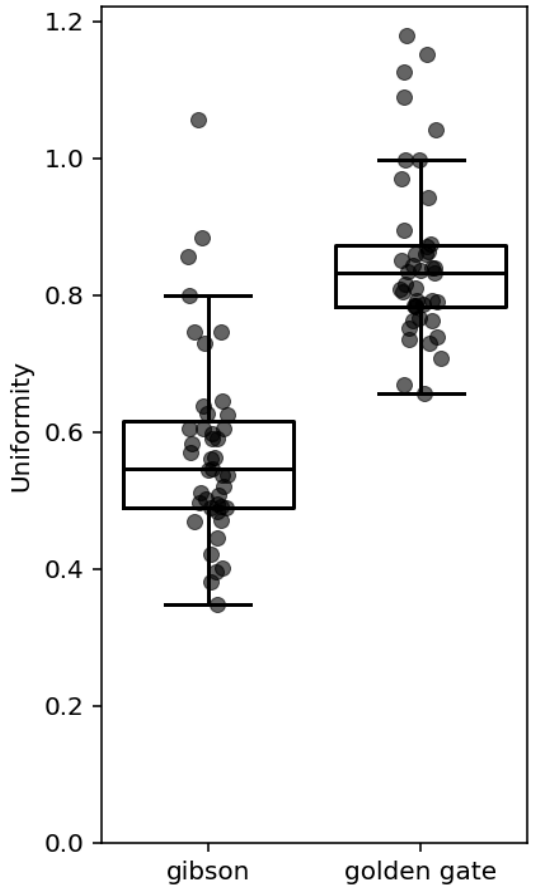
Supplementary Figure 1: Barcode diversity after Gibson assembly. Heatmaps show the total number of unique barcodes for each possible amino acid substitution at each codon position in *PDR1* fragments F13 and F43. For each fragment, a total of 25,000 transformants were recovered and analyzed. Some amino acids, such as serine, leucine, and arginine, are more accessible with the NNK library (six codons each instead of four) and thus show higher diversity. Grey squares indicate wild-type (WT) amino acids.



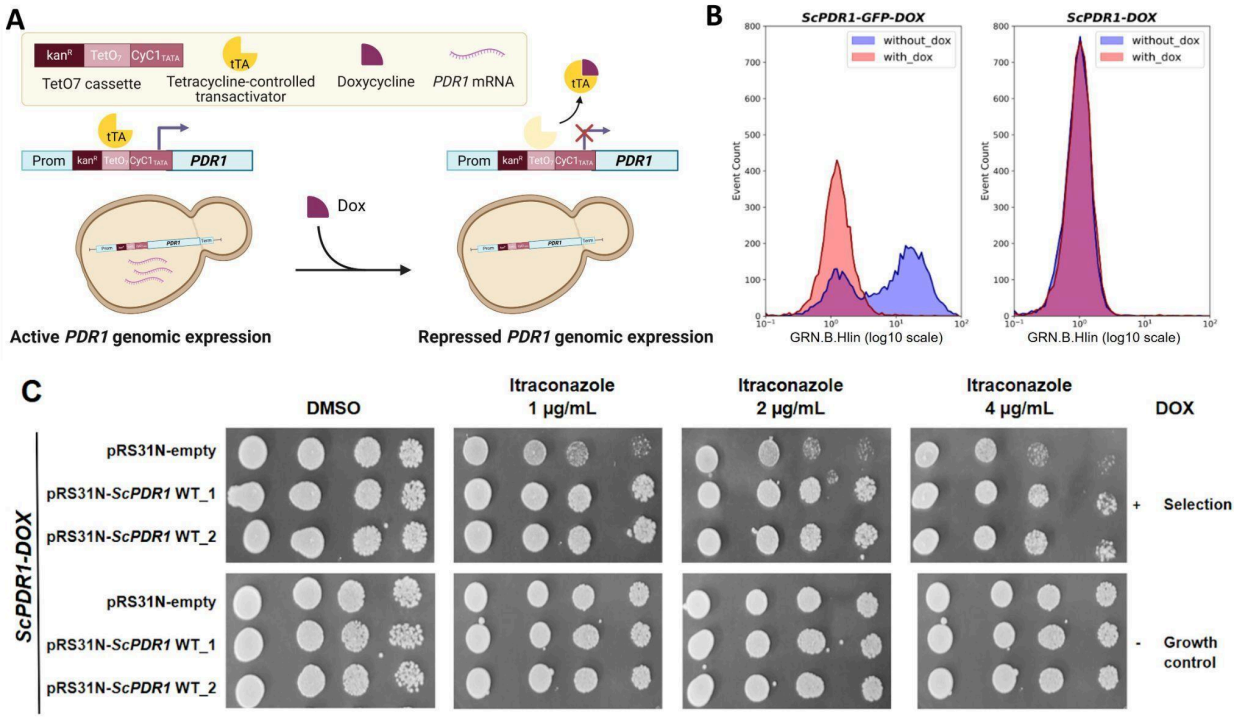
Supplementary Figure 2: Barcode diversity after Golden Gate assembly. Heatmaps show the total number of unique barcodes for each possible amino acid substitution at each codon position in *PDR1* fragments F13 and F43. For each fragment, a total of 100,000 transformants were recovered and analyzed. Some amino acids, such as serine, leucine, and arginine, are more accessible with the NNK library (six codons each instead of four) and thus show higher diversity. Grey squares indicate wild-type (WT) amino acids.



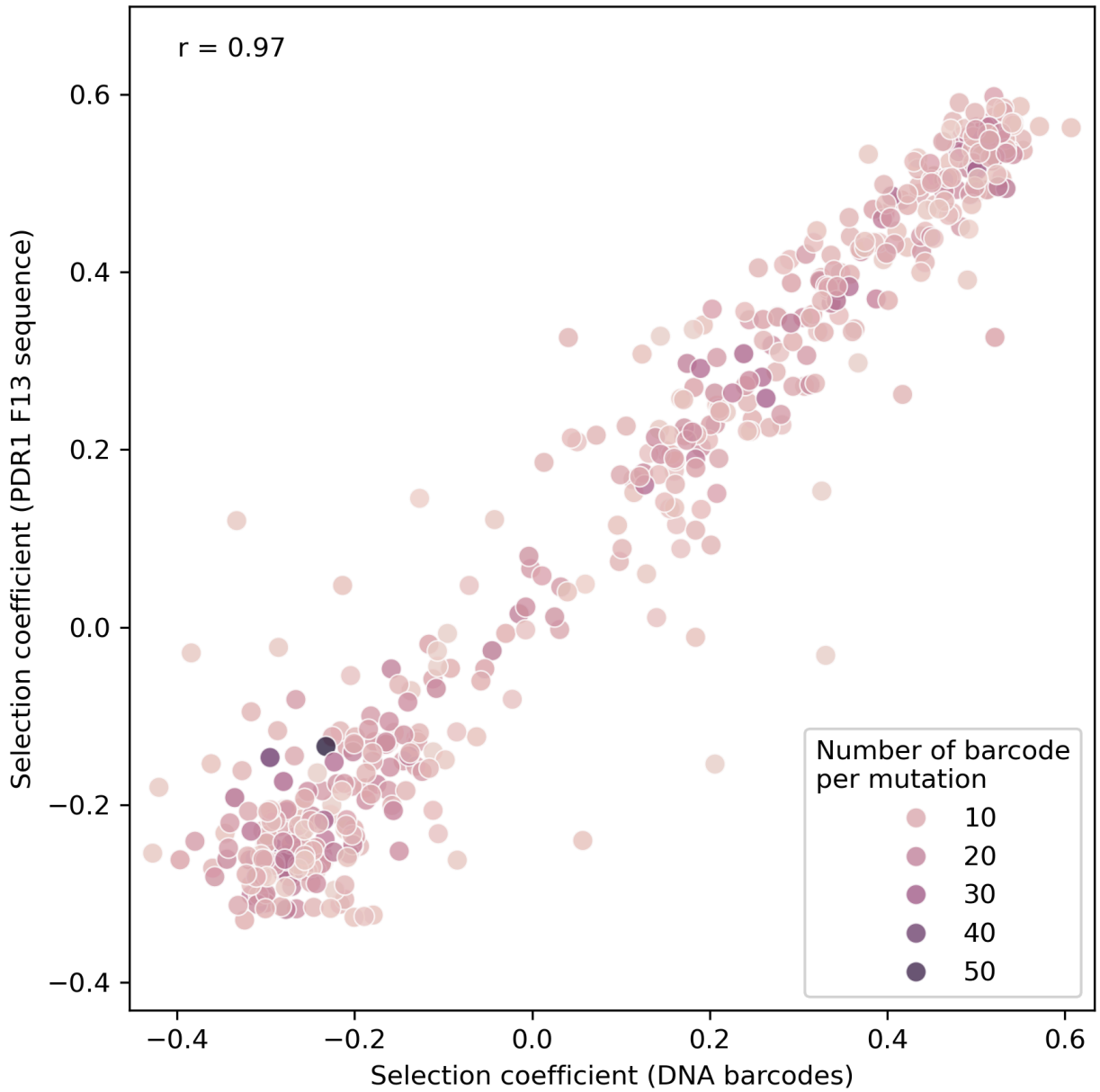
Supplementary Figure 3: Assessment of library uniformity after each cloning step (Gibson and Golden Gate assemblies) using the Gini coefficient [31]**.** Boxplots represent the distribution of Gini coefficients (ranging from 0 to 1) calculated for each fragment, where lower values indicate more uniform coverage and higher values indicate increased inequality in representation.



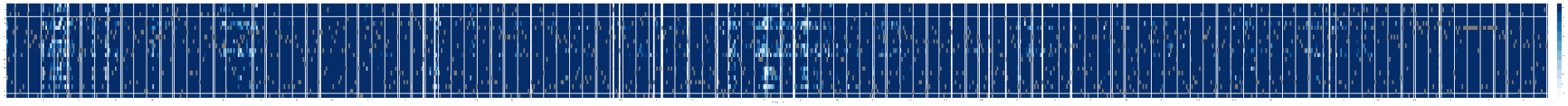
Supplementary Figure 4: Assessment of library uniformity after each cloning step (Gibson and Golden Gate assemblies) using the uniformity score [28]**.** Boxplots show the distribution of uniformity scores for each fragment, defined as the log difference between the 90th and 10th percentiles of mutant read counts. Lower scores indicate more uniform representation, while higher scores indicate greater inequality.



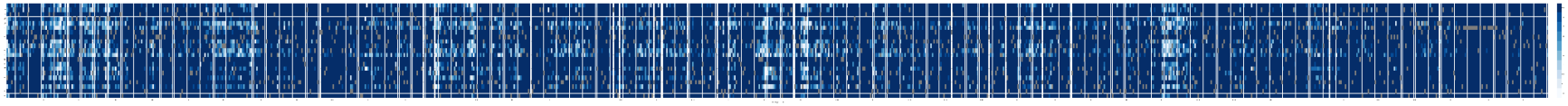
Supplementary Figure 5: Construction and validation of a *S. cerevisiae* strain to measure the impact of *PDR1* mutations on drug resistance. A) Inducible control of genomic *PDR1* expression in *S. cerevisiae* using a doxycycline-repressible promoter (TetO7). Created in BioRender. Barff, T. (2025) https://BioRender.com/zo8wlgo. **B)** Validation of genomic *PDR1* repression in the *ScPDR1*-DOX strain. *PDR1* expression level was assessed by flow cytometry using a GFP fusion as a reporter. Fluorescence intensity is shown for the strain expressing GFP (left panel: *ScPDR1-GFP-DOX*) and the negative control strain without GFP (right panel: *ScPDR1-DOX*) in media with (red) or without (blue) DOX. GFP intensity was analyzed on the main population of morphologically normal cells (FSC-H<15000 and FSC-A<15000). 5,000 events were recorded per replicate. Three independent biological replicates were analyzed per condition. **C)** Spot dilution complementation assay of genomic *PDR1* expression in *ScPDR1-DOX* by plasmid-expression of *ScPDR1* (pRS31N-*ScPDR1* WT) or not (pRS31N-empty (control)). Growth conditions: DMSO (control) or Itraconazole (antifungal).



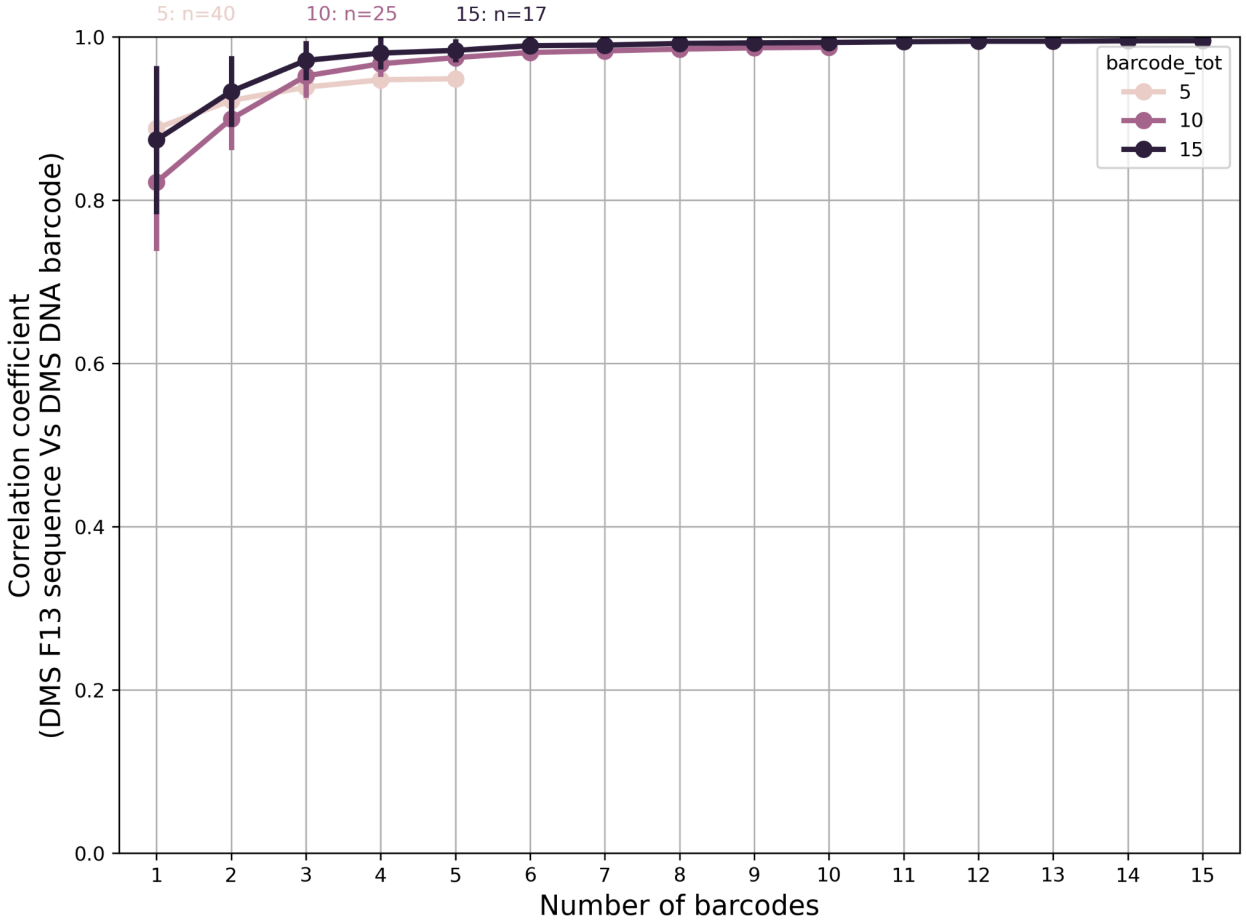
Supplementary Figure 6: Correlation between the selection coefficients measured with direct sequencing of the mutated *PDR1* F13 coding region and DNA-barcode inference, using all usable barcodes. The strong correlation (Pearson r = 0.97) indicates that barcode-based estimates accurately recapitulate the directly measured fitness effects. The number of barcodes per mutation does not affect the correlation.



Supplementary Figure 7: Library quality control after Gibson assembly of the full-length *PDR1* sequence (43 fragments). Heatmap displays the barcode diversity associated with each possible amino acid substitution at each codon position, with values clipped at 10 barcodes per mutation for clarity.



Supplementary Figure 8: Library quality control after Golden Gate assembly of the full-length *PDR1* sequence (43 fragments). Heatmap displays the barcode diversity associated with each possible amino acid substitution at each codon position, with values clipped at 10 barcodes per mutation for clarity.



Supplementary Figure 9: Correlation between the selection coefficients measured with direct sequencing of the mutated F13 coding region and DNA-barcode inference, as a function of the number of barcodes per variant considered. Curves correspond to variants with exactly 5 (n = 40), 10 (n=25), or 15 (n=17) barcodes. Means and confidence intervals were calculated from 100 random barcodes subsamplings per variant. The high correlation observed even for variants with only five barcodes indicates that the overall correlation is not driven only by highly diversified variants.