ABSTRACT: Feedback control allows cells to dynamically sense and respond to environmental changes. However, synthetic controller designs can be challenging because of implementation issues, such as determining optimal expression levels for circuit components within a feedback loop. Here, we addressed this by coupling rational design with selection to engineer a synthetic feedback circuit to optimize tolerance of *Escherichia coli* to the biofuel pinene. *E. coli* can be engineered to produce pinene, but it is toxic to cells. Efflux pumps, such as the AcrAB-TolC pump, can improve tolerance, but pump expression impacts growth. To address this, we used feedback to dynamically regulate pump expression in response to stress. We developed a library with thousands of synthetic circuit variants and subjected it to three types of pinene treatment (none, constant, and varying pinene). We were able to select for strains that were biofuel tolerant without a significant growth cost in the absence of biofuel. Using next-generation sequencing, we found common characteristics in the designs and identified controllers that dramatically improved biofuel tolerance.

KEYWORDS: feedback control, biofuel tolerance, next generation sequencing, efflux pump
The ideal feedback control system would only turn on pumps in the presence of pinene. Therefore, the AcrAB-TolC efflux pump improves pinene tolerance; however, there is a trade-off between the biofuel tolerance benefits and costs associated withacrAB expression. We confirmed thatacrAB increases pinene tolerance (Figure 1A). We next induced pump expression by using IPTG to control a LacI-regulated promoter. We observed a decrease in growth as we increased induction (Figure 1B). These results confirm that the AcrAB-TolC efflux pump improves pinene tolerance; however, there is a trade-off between the biofuel tolerance benefits and costs associated withacrAB expression.

To accomplish this, we designed a library of synthetic promoters that vary the number and locations of a binding site for a repressor that responds to biofuel stress. In addition, we introduced additional diversity in the −10 and −35 regions of the promoter to alter the unregulated expression level. The synthetic promoters were used to control expression of theacrAB pump. We subjected the library to three different treatment conditions including no pinene, constant pinene, and varying pinene. The pinene treatments selected for library variants with biofuel tolerance without a significant cost to pump expression. By conducting deep sequencing on our library results, we were able to identify common features in the feedback loop designs. This approach can be broadly generalized by coupling it with sensor-selector systems.

**RESULTS AND DISCUSSION**

We began by confirming that the AcrAB-TolC efflux pump improves pinene tolerance. Working inE. coli BW25113 ΔacrBwe tested the pinene tolerance of cells with and without a plasmid containingacrAB. Because all three pump components are necessary for function,ΔacrB is sufficient to eliminate pump activity. We confirmed thatacrAB increases pinene tolerance (Figure 1A). We next induced pump expression by using IPTG to control a LacI-regulated promoter. We observed a decrease in growth as we increased induction (Figure 1B). These results confirm that the AcrAB-TolC efflux pump improves pinene tolerance; however, there is a trade-off between the biofuel tolerance benefits and costs associated withacrAB expression.

To design a regulatory system to control efflux pump expression to balance these trade-offs. Although in principle it may be possible to select a pump expression level...
that balances the benefits and costs, the exact expression level required to optimize growth is dependent on the level of biofuel present and will change over time and across production conditions. Because these factors are uncertain and dynamic, we sought instead to use feedback to automatically optimize pump expression. With this approach, we used a sensor to measure pinene stress within the cell and turn on pumps in response (Figure 1C). We used MexR, a transcriptional repressor derived from *Pseudomonas aeruginosa*, to control *acrAB* expression (Figure 1D).

To confirm that a MexR-regulated promoter can respond to pinene stress, we used a synthetic promoter containing MexR binding sites to drive expression of green fluorescent protein (gfp). We observed an increase in fluorescence in response to pinene (Figure 1E). This initial design served as a proof of principle that MexR could regulate gene expression in response to pinene. However, different designs of the synthetic promoter could adjust the basal expression level, fold-change in induction, or the pinene concentration that turns on gene expression from the promoter.

We next built a library of synthetic promoters. We varied the number and location of MexR binding sites and also altered the transcription levels through modifications to the −10 and −35 sites within the promoter. This design was initially inspired by related studies using a modular promoter structure, but which lacked the library diversity introduced here. As our promoter template, we started with the phage T7 A1 promoter. This is a strong promoter from T7 that uses *E. coli* RNA polymerase for transcription. We considered three potential MexR binding sites, one upstream of the −35 site, one in between the −35 and −10 sites, and one downstream of the −10 site (Figure 1F). In *P. aeruginosa*, the transcription factor MexR regulates the *mexAB-oprM* multidrug efflux pump operon and binds to a promoter within the *mexA-mexR* intergenic region. We isolated the binding sequence and used this in our synthetic promoter. In our design, each promoter variant can have a MexR binding site or no site in each of these three spots, for a total of eight possible promoters. We developed a naming system to describe the promoter classes where "M" indicates presence and "O" indicates absence of a MexR site. For instance, pOOM is a promoter that has a single MexR binding site downstream of the −10 site.

We sought to further increase the diversity of our library by allowing for degeneracies in the −35 and −10 sites. We started with the native sequence for the T7 promoter and incorporated degenerate sites at locations within the promoter where natural sequence diversity is observed. In total, our design has ~5000 possible promoters. We confirmed that the differences in −35

Figure 2. Selection experiments under three types of pinene treatment. (A) Schematic of the selection experiment. The library of plasmids with synthetic promoters controlling *acrAB* was cotransformed with pBbS8a-mexR into *E. coli ΔacrB*. Then cultures were subjected to one of three pinene treatments. Every 12 h, the culture was diluted into fresh selective medium with pinene as required. (B) Treatments included no pinene, constant 0.05% pinene, and varying pinene. (C) Growth with and without pinene for colonies that survived the selection. Shaded region shows standard error of *n* = 3 biological replicates.
−10 sites generate diversity in expression levels by cloning library variants of the pOOO promoter upstream of gfp (Figure 1G, Figure S1).

We next added feedback by using the synthetic promoters to control expression of the acrAB pump. We ran selection experiments over the course of 60 h, with dilution into fresh media every 12 h, with pinene as necessary (Figure 2A). We grew cultures in the absence of pinene, with a constant level of 0.05% (v/v) pinene, and with a treatment that varied between 0.05% and 0% pinene every 12 h (Figure 2B). Our goal was to isolate library members that were optimized for growth without pinene, with constant pinene, and those that balance these two effects by trading off pump expression benefits and costs.

As expected, our selections returned results consistent with the environments they were isolated in. Samples isolated from the 0% pinene treatment experiment exhibited normal growth in 0% pinene and no growth in 0.05% pinene (Figure 2C). In contrast, those strains isolated in the presence of constant or varying pinene maintained growth in pinene. In all cases, growth under 0% pinene was similar, suggesting that the selection was able to isolate library members where pumps do not place a major burden on their growth.

We were interested in which classes of promoter sequences led to pinene tolerance. For an in depth look at the results of our selections, we used next generation sequencing to identify the promoter regions before and after selection. In all cases, sequencing returned 8−15 million reads, providing excellent coverage of our synthetic promoter library.

First, we looked at the composition of our initial library before pinene treatment. We sorted these sequencing results into the eight promoter classes (pOOO, pMOO, etc.) and looked at the distributions of library membership. As expected, we observed a roughly even distribution across the promoter classes (Figure 3A). Further details on the sequences of the overrepresented promoters are provided in Table S2.

Figure 3. Library composition before and after pinene treatment. (A) Initial library composition before pinene treatment, grouped by promoter class. (B) Relative abundance is the ratio of sequence frequency after selection to the original library before selection. Data are shown for three biological replicates. (C) Parity plot showing the number of occurrences of each sequence before and after selection. These data are for the pOOO promoter class with no pinene treatment. Three biological replicates are shown. (D) Stacked bar chart showing repeated sequences within each of the eight promoter classes at the end of the selection experiment. Magenta shows most abundant, cyan second most abundant, and so on. Three biological replicates are shown. Further details on the sequences of the overrepresented promoters are provided in Table S2.
results did not align well with any of the promoter designs. We grouped these sequences into a ninth class called “Other”. We note that even for promoters that grouped with known classes, we observed variation beyond that included in the initial designs. Overall, we estimate that our starting library contains ~15,000 unique sequences that appear at least 10 times each in the sequencing results.

We compared the initial library composition to the results after selection (Figure 3B). We calculated the number of members in each of the promoter classes for each of three biological replicates. We normalized these values to those from the initial library distribution (Figure 3A) and looked for changes relative to this initial distribution. For 0% pinene treatment we observed no major changes in the promoter class distribution relative to the initial library.

We wondered whether these results indicated that there was little pump expression cost observed across any of the promoter classes, or, alternatively, if there was selective pressure within the subclasses, but not a particular bias toward the MexR binding site configuration. To answer this, we looked at which promoter sequences were present after 0% pinene treatment relative to those that were present in the initial library prior to treatment. Interestingly, we observed dramatic selection within these promoter classes, with many of the sequences disappearing during treatment. As an example, within the pOOO promoter class we plotted the number of appearances of all promoter sequences relative to the number of appearances in the initial library (Figure 3C). We found examples of promoter sequences that were observed only once in the initial library that were present thousands or tens of thousands of times after selection. So, although we observed no particular preference for the presence or absence of MexR binding sites, library membership was winnowed during selection. The sequences that have dropped out likely represent cases where pump expression is high enough to be toxic, as there would be strong selective pressure against these sequences under 0% pinene treatment.

In contrast to 0% pinene, 0.05% constant pinene treatment resulted in large shifts in the distributions of promoter classes (Figure 3B). In particular, we noted that results were biased away from constitutive promoters (pOOO) and toward promoters with a MexR binding site in the region just downstream of the −10 site (pOOM, pOMM, pMOM, pMMM). The results from the varying pinene treatment were similar to the 0.05% constant pinene treatment, though we observed more variation between the replicates. This result suggests that there may be multiple ways to solve the trade-off between pump benefits and toxicity.

We were interested in understanding whether results had converged to a small number of sequences, and if these were consistent across replicates and treatments. To address this, we searched for repeated sequences within each of the promoter classes. The initial library prior to selection had relatively few repeats, with no sequences dominating the library (Figure S2). We observed similar results under no pinene treatment (Figure 3D). For these conditions, we retained a broad distribution of promoter sequences and results were consistent across the replicates. In contrast, under pinene treatment we observed convergence to a small subset of sequences. In the case of 0.05% constant pinene treatment, we found that the promoter classes that were overrepresented were composed of a small number of promoter sequences. For instance, the top 10 most abundant sequences from pMMM in one replicate made up 56% of all samples sequenced from that promoter class. Across replicates we observed some overlap in promoter sequence results, though there were also cases where distinct promoters emerged (Table S2). These results indicated that it is possible to select for feedback control strategies that balance biofuel and pump toxicity, and that multiple sequences may exist within a library that solve the problem.

Feedback control is a powerful strategy that allows cells to respond to changing environments and uncertainty. Here, we designed a synthetic feedback controller to sense the presence of pinene and turn on efflux pumps in response. Although the benefits of feedback are widely appreciated, it can be challenging to design biological controllers. Therefore, we used a library and selection-based approach here to identify synthetic feedback systems that are optimized for performance in the presence, absence, and under fluctuating pinene stress conditions. In each case, we were able to select for library variants that optimized for the pinene treatment they were subjected to. Next generation sequencing revealed sequenced-based trends in the outcomes of the selections. In future applications of this work, it may be possible to use library-based approaches to identify features of synthetic control system designs that emerge that could be used in rational design. For instance, in this study we found that MexR binding sites immediately upstream of the acrAB gene tend to be selected when cells are grown in the presence of pinene. In addition, we did not see a strong selective pressure for specific promoter classes in the absence of pinene; however, we did observe selection within the classes. In cases with pinene treatment we often observed convergence to a small number of sequences; these sequences varied between biological replicates, suggesting that there are likely several library variants capable of optimizing performance. Additional experiments measuring the response of these synthetic promoters to pinene could help to illustrate whether the pinene response properties are similar between the selected promoters. In particular, it would be interesting to understand where feedback strategies, as opposed to static strategies, are favored.

In this study, we added pinene exogenously; however, the ultimate application of these controllers is in production strains. As such, it will be important to match the pinene treatment conditions applied to those that are likely to be encountered in the production setting to ensure that the optimal controllers for those conditions are selected.

Tolerance engineering allows us to directly perform selection experiments because library members that perform poorly become underrepresented in the culture. However, this approach can be applied even in cases where there is not a direct selection target. Alternative strategies can use screening-based approaches or be coupled with sensor-selectors to screen for optimal designs. Overall, coupling synthetic feedback circuit design with library-based selection offers a powerful strategy for optimizing tolerance for synthetic biology and metabolic engineering applications.

**METHODS**

**Strains and Plasmids.** For all experiments, we used E. coli BW25113 ΔacrB from the Keio collection, where we removed the kanamycin resistance marker following the protocol with pCP20 from Datsenko et al. We used the T7 phage P1 as the base sequence for the synthetic promoter design. P1 is a strong “early” promoter for T7 that uses E. coli RNA polymerase. This
sequence is the promoter we refer to as pOOO, which has no MexR binding sites. We derived the MexR binding site from a footprinting study on the mexAB-ompM promoter.33 We focused on binding site I from this study, and developed a truncated version of the site that preserves the repeated sequence GTTGA and its inverse while maintaining the optimal 17 bp spacing between the −35 and −10 sites. The MexR binding site sequence we used is AGTTGACCTTATCAACC, where the inverted repeats are underlined. Following the approach from Zhang et al.,10 we placed MexR binding sites at three regions within the promoter. Further sequence details are provided in Table S1.

We introduced variability into the promoter sequence at the −35 and −10 sites. To do this, we started with the original T7 P_Ai sequence, which is −35 = TTTTACT and −10 = GACTAT. Working with a previously published list of known E. coli promoters and the frequency of degeneracy within the −35 and −10 sites,37 we first determined which base pairs within this region were more likely to be variable. We defined an arbitrary threshold of 75% and determined that we would only introduce variability at sites where the most frequent base pair in that position appeared in less than 75% of cases. As an example, if a given position is a T in 80% of cases, we fixed this position as a T. Applying this upper threshold to T7 P_Ai results in sequences of the form −35 = TTNNN and −10 = GANNNT. Because there are certain base pairs that appear very rarely in E. coli promoters we also set a lower threshold of 15%, where we excluded any base that appeared less frequently. The final promoter degeneracies are −35 = TTTKHMW and −10 = GADDDHT (K = G or T; H = A, C, or T; M = A or C; W = A or T; D = A, G, or T).

We constructed the library for pOOO and used it to control expression of superfolder green fluorescent protein (pOOO-gfp). We used the pBbAsk vector from38 and amplified the region containing the kanamycin resistance marker and p15A origin of replication, but not the lacI gene. Using Gibson assembly,39 we cloned in the superfolder gfp gene (AddGene #50550) in place of rfp to make pBbAsk-gfp. We next built the pOOO promoter library by starting with single-stranded oligonucleotides for pOOO (Table S1). We PCR amplified this sequence, which contained degenerate bases, using overhangs for the vector. The promoter insert was cloned into the vector with gfp to create pOOO-gfp. After cloning, we obtained several thousand colonies per plate (compared to tens of colonies on the control plate without the insert). Using a sterilized razor blade, we scraped all colonies together and resuspended them in a Lucio Broth (LB) solution with 20% glycerol. We made 50 μL aliquots of this pooled mixture and stored them at −80 °C for subsequent testing.

The complete library was constructed in a similar fashion. Here, we used pBbAsk-acrAB40 as the vector. Using single-stranded oligos for each of the eight promoters (pOOO, pOOM, etc.) as shown in Table S1, we amplified each of the eight promoters using PCR with identical overhangs for Gibson assembly to generate inserts. We combined these with the acrAB vector to create eight libraries. We then isolated plasmids and transformed them into E. coli ΔacrB with pBB8a-mexR (described below). After plating each of the eight libraries, we scraped the plates with a sterilized razor blade and pooled all eight libraries together in equal proportion. We made glycerol stocks of the pooled library by resuspending them in a 20% glycerol LB solution. Aliquots of this pooled mixture were stored at −80 °C for subsequent testing.

We synthesized a codon optimized version of mexR as a gBlock (Integrated DNA Technologies). Starting with the sequence for mexR from P. aeruginosa, we codon optimized the gene for E. coli by selecting the most abundant codon. We cloned the codon-optimized mexR gene into the vector pBB8a41 to make pBB8a-mexR.

Measuring Pinene Sensor. We used E. coli cotransformed with pMMM-gfp and pBB8a-mexR, where we used the original T7 P_Ai sequence for the −35 and −10 sites. This strain was grown from a single colony in selective LB medium with 100 μg/mL carbenicillin and 30 μg/mL kanamycin. Overnight cultures were diluted 1:100 in fresh LB with antibiotics and grown at 37 °C with 200 rpm orbital shaking for 2 h. The optical density (600 nm) of the cultures was then adjusted to 0.2, and cultures were transferred to a 24-well plate with 1 mL of culture per well. Pinene was added directly to the wells at final concentrations of 0%, 0.05%, 0.075%, 0.1%, 0.15%, and 0.2%. The plate was sealed with a gas permeable membrane (Thermo Scientific AB-0580) and incubated at 37 °C with 200 rpm shaking for 8 h. Next, 1.5 μL of each sample was diluted into 150 μL phosphate buffered saline (PBS) and samples were read on a Guava easyCyte HT Sampling Flow Cytometer with GFP excitation and emission values of 485 and 510 nm.

Selection Experiment. For the selection, we grew cultures in 24-well plates covered with a gas-permeable membrane (Thermo Scientific AB-0580). At the start, each well contained 1 mL of LB with 100 μg/mL carbenicillin and 30 μg/mL kanamycin, 10 μL of the pooled library glycerol stock, and pinene when required. Cultures were grown at 37 °C with 200 rpm shaking for 12 h, then diluted 1:100 into fresh LB medium with selective antibiotics and pinene when required. This process was repeated every 12 h over the course of 60 h, for a total of five treatment periods. We ran three biological replicates for each of three pinene treatment conditions, for a total of nine samples.

For the three treatment conditions we grew cells with (i) no pinene, (ii) constant 0.05% (v/v) pinene, and (iii) varying pinene. In the case of varying pinene, we added 0.05% pinene at the first, third, and fifth 12-h periods. The second and fourth 12-h periods had no pinene added. At the conclusion of the 60-h selection we made 20% glycerol stocks from the final cultures and froze them at −80 °C for subsequent analysis.

Toxicity Measurements. Glycerol stocks of the selection experiment results were plated on LB agar plates containing 100 μg/mL carbenicillin and 30 μg/mL kanamycin. Three colonies were chosen at random from each of three replicates of three treatments (total of 27 colonies selected). These colonies were grown overnight in LB with antibiotics. After overnight growth, each culture was refreshed at a ratio of 1:100 in a 24-well plate containing LB with antibiotics. After a 2 h growth period, pinene was added at 0.05% (v/v) to half of the wells, while the remaining wells received no treatment. The optical density at 600 nm of all wells was measured at 10 min intervals for 14 h thereafter using a BioTek Synergy H1m plate reader.

Next Generation Sequencing. Plasmid DNA was extracted from saved glycerol stocks from the selection experiment using the Qiagen Miniprep Kit (Qiagen). After quantification, the miniprep products were adjusted to a concentration of 1 ng/μL. Products were then PCR amplified for 13 cycles using the primers 5’-CGGCGTAGAGGA-TCGAG-3’ (forward) and 5’-CTCCTTCTTTAAAGATCCT-TTTG-3’ (reverse), which target the synthetic promoter region. Library preparation and sequencing-by-synthesis were
carried out by the University of Vermont Cancer Center Advanced Genome Technologies Core.

Samples were labeled and barcoded with the NEXTflex Rapid DNA Sequencing Kit (Bioo) and sequenced with the Illumina HiSeq 1500 using single-end sequencing with 150 base pair reads at the University of Vermont Cancer Center Massively Parallel Sequencing facility.

**Bioinformatics.** We used bowtie2 to categorize the promoter sequences returned from next generation sequencing reads as members of the classes (pOOO, pMOO, etc.). To accomplish this, we built a custom index using bowtie2 and the eight promoter sequences (Table S1). We then ran bowtie2 and generated SAM output files, which classified the reads as one of the eight promoter classes, or if the sequence failed to be classified we called it Other. Analysis of the SAM output files to calculate the frequency of sequence occurrence was conducted using custom Perl and Matlab scripts.

**REFERENCES**


