

Evolution-guided optimization of biosynthetic pathways

Srivatsan Raman^{a,b,1,2}, Jameson K. Rogers^{a,c,1}, Noah D. Taylor^{a,b,1}, and George M. Church^{a,b}

^aWyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115; ^bDepartment of Genetics, Harvard Medical School, Boston, MA 02115; and ^cSchool of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02143

Edited by David Baker, University of Washington, Seattle, WA, and approved October 30, 2014 (received for review May 21, 2014)

Engineering biosynthetic pathways for chemical production requires extensive optimization of the host cellular metabolic machinery. Because it is challenging to specify a priori an optimal design, metabolic engineers often need to construct and evaluate a large number of variants of the pathway. We report a general strategy that combines targeted genome-wide mutagenesis to generate pathway variants with evolution to enrich for rare high producers. We convert the intracellular presence of the target chemical into a fitness advantage for the cell by using a sensor domain responsive to the chemical to control a reporter gene necessary for survival under selective conditions. Because artificial selection tends to amplify unproductive cheaters, we devised a negative selection scheme to eliminate cheaters while preserving library diversity. This scheme allows us to perform multiple rounds of evolution (addressing $\sim 10^9$ cells per round) with minimal carryover of cheaters after each round. Based on candidate genes identified by flux balance analysis, we used targeted genome-wide mutagenesis to vary the expression of pathway genes involved in the production of naringenin and glucaric acid. Through up to four rounds of evolution, we increased production of naringenin and glucaric acid by 36- and 22-fold, respectively. Naringenin production (61 mg/L) from glucose was more than double the previous highest titer reported. Whole-genome sequencing of evolved strains revealed additional untargeted mutations that likely benefit production, suggesting new routes for optimization.

evolution | metabolic engineering | synthetic biology | sensors | biosynthetic pathways

Microbial production of chemicals presents an alternative to ubiquitous chemical synthesis methods. Biosynthetic production is attractive because it can use a broad assortment of organic feedstocks, proceed under benign physiological conditions, and avoid environmentally deleterious byproducts. Biosynthetic alternatives are being pursued for a wide range of chemicals, from bulk commodity building blocks to specialty chemicals.

Natural cells are seldom optimized to produce a desired molecule. To achieve economically viable production, extensive modifications to host cell metabolism are often required to improve metabolite titer, production rate, and yield. The optimizations of biosynthetic pathways for 1,3-propanediol (1), flavonoids (2, 3), L-tyrosine (4), and 1,4-butanediol (5) illustrate this complexity. Fortunately, computational models of cellular metabolism, such as flux balance analysis (FBA), aid in predicting metabolic changes likely to improve the production of a target molecule. Powerful methods including oligonucleotide-directed genome engineering (6) (multiplex automated genome engineering, MAGE) and Cas9-mediated editing can specifically mutate genomic targets predicted by FBA. The combinatorial space of these genomic mutations quickly outstrips the throughput of current analytical methods for evaluating chemical production in individual clones ($<10^3$ samples per machine per day).

Biosensors that report on the concentration of a chemical within each individual cell can alleviate this screening bottleneck (10). Such sensor reporters transduce the binding of a target small molecule by a sensory protein or RNA into a gene expression

readout (7). The resulting expression of a fluorescent reporter gene or antibiotic resistance gene allows facile identification of mutant cells with increased production of the target chemical.

Sensor reporters have been used to screen for increased microbial production of several chemicals, including the isoprenoid precursor mevalonate (8), L-lysine (9, 10), 1-butanol (11), and triacetic acid lactone (12). These studies evaluated a set of variants that altered the expression or coding sequences of one or two key enzyme genes encoded on a plasmid (8, 10–12). Similarly, a lysine-responsive sensor reporter was used to uncover new endogenous enzyme mutants in *Corynebacterium glutamicum* implicated in higher L-lysine production (9).

We sought to expand the scope of sensor-directed metabolic engineering to the directed evolution of whole endogenous pathways. Using FBA as a guide, we simultaneously targeted up to 18 *Escherichia coli* genomic loci to induce mutations in regulatory or coding sequence of genes implicated in biosynthesis of a target molecule. We established a robust selection, using a sensor protein responsive to the target chemical to regulate the expression of an antibiotic resistance gene. Nearly a billion pathway variants could be evaluated simultaneously, enriching for the best producers when selection pressure was applied.

A major challenge faced by this selection approach (and a difficulty for most genetic selections) is the incidence of cheater cells that survive without producing the target molecule. These cheaters evolve to survive selection by mutating the sensor or selection machinery, rather than through higher target molecule synthesis. Lacking a metabolic burden, these “evolutionary escapees”

Significance

Microbes can be made to produce industrially valuable chemicals in high quantities by engineering their central metabolic pathways. This process may require evaluating billions of cells, each containing a unique pathway design, to identify the rare cells with high production phenotypes. We mutated targeted locations across the genome to modify several genes identified as key players. We used sensory proteins responsive to a number of target chemicals to couple the concentration of the target chemical in each cell to individual cell fitness. This coupling of chemical production to fitness allows us to harness evolution to progressively enrich superior pathway designs. Through iterations of genetic diversification and selection, we increased the production of naringenin and glucaric acid 36- and 22-fold, respectively.

Author contributions: S.R., J.K.R., N.D.T., and G.M.C. designed research; S.R., J.K.R., and N.D.T. performed research; S.R., J.K.R., N.D.T., and G.M.C. analyzed data; and S.R., J.K.R., N.D.T., and G.M.C. wrote the paper.

Conflict of interest statement: The authors have a pending patent application.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database, www.ncbi.nlm.nih.gov/bioproject/267705.

¹S.R., J.K.R., and N.D.T. contributed equally to this work.

²To whom correspondence should be addressed. Email: sraman@genetics.med.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1409523111/-DCSupplemental.

outcompete the top producers during a selection. Multiple selection cycles compound escape, obscuring productive cells and making further pathway evolution infeasible. We therefore devised a selection scheme that, by toggling between negative and positive selection, allows us to remove escapees from the population when they arise. This strategy maintained high selection fidelity, permitting multiple rounds of evolution to progressively enrich for higher-producing cells.

For sensor reporter metabolic engineering to be generalizable, sensor domains specific to many different target molecules must be available. Fortunately, natural sensors exist for a wide array of industrially relevant chemicals, including aliphatic hydrocarbons, short-chain alcohols, sugars, amino acids, polymer building blocks, and vitamins. Many more sensor domains are likely to be present among the thousands of additional bacterial regulators known from sequence (13–15) that remain to be characterized. We adapted 10 regulators to our selection system, creating synthetic dependence on their cognate inducer molecules, and demonstrated the utility of two of these for genome-wide metabolic engineering.

Results

Sensor selectors are a specific example of the sensor reporter paradigm that use a gene whose product confers a fitness advantage (e.g., antibiotic resistance) as the reporter. Our sensor selector architecture encodes a chemical-responsive sensor domain together with its cognate promoter, which controls a selectable reporter (Fig. 1A). We show that this general implementation is suitable for transcriptional regulators (both activators and repressors) and riboswitches that collectively respond to a wide variety of chemicals (Fig. 2A and *SI Appendix, Table S1*).

Each sensor selector exhibits unique behavior, dependent on sensor affinity for the chemical, sensor type, and induction response; for example, the escape rate and operational range can vary over orders of magnitude for different sensors (Fig. 2A). For each sensor, the operational range is defined as the chemical concentration range over which cells continue to experience a marginal fitness advantage with increasing concentration. The lower bound of the range reflects the lowest concentration of exogenously supplied inducer that provides a selective advantage.

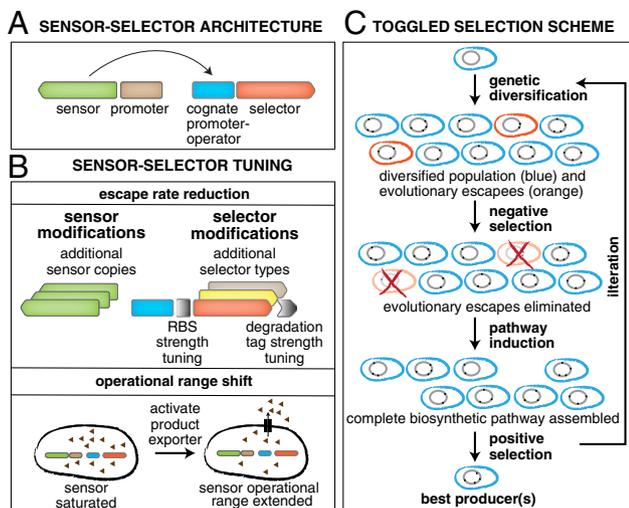


Fig. 1. Sensor selector design and pathway optimization through toggled selection. (A) Sensor selector genetic architecture. (B) Methods for tuning sensor selectors to reduce escape rate and shift operational range. Escape rate is reduced by (i) adding a degradation tag, (ii) mutating the RBS of the selector, (iii) including multiple orthogonal selectors, or (iv) including an additional copy of the sensor. Activating an exporter shifts the sensor selector operational range. (C) Toggled selection protocol for biosynthetic pathway optimization through multiple rounds of evolution. Negative selection eliminates cheaters; subsequent positive selection identifies higher-producing clones from a diverse library.

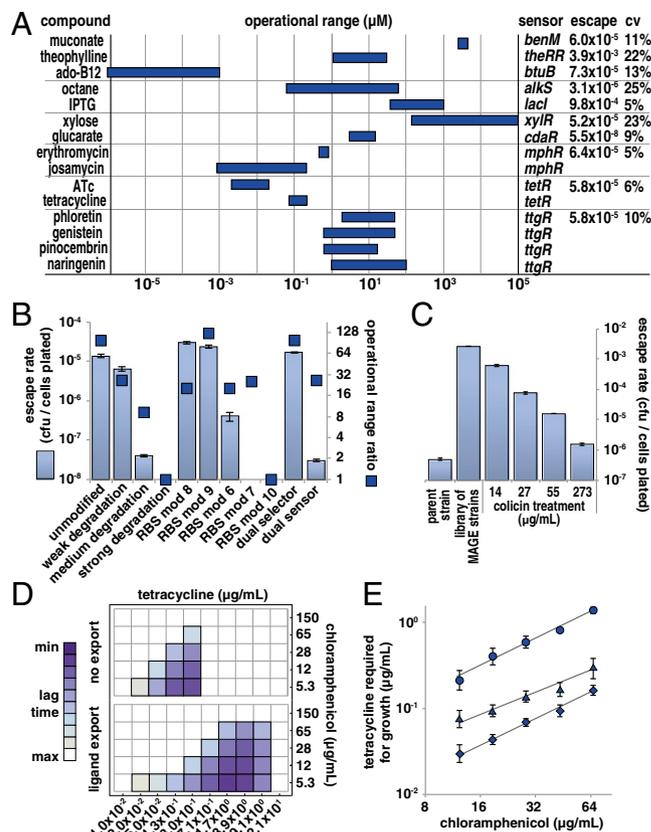


Fig. 2. Characterization of sensor selector modifications. (A) Escape rate and operational range of 10 sensors with cognate inducer chemicals and ToIC as a selector. Horizontal bars depict the operational range. The lower bound of the range reflects the lowest concentration of exogenously supplied inducer that provides a selective advantage. The upper bound of the range indicates that higher inducer concentration does not increase fitness advantage. (B) Effect of genetic modifications on the TtgR-TolC sensor-selector escape rate and operational range. Escape rate (light blue bars, left axis) is the proportion of cells that evade selection (cfu per cells plated). Escape rate not shown if below the limit of detection (10^{-10} cfu per cells plated). Escape rate operational range ratio (blue boxes, right axis) is the ratio of the high concentration of the operational range to the low concentration of the operational range. (C) MAGE mutagenesis increases the escape rate (cfu per cells plated) in the CdaR-TolC strain. Treatment with colicin E1 removes escapees in a dose-dependent manner. (D) Tetracycline exporter (*tetA*) expression shifts the operational range of the TetR-CAT (chloramphenicol acetyltransferase) sensor selector. Growth lag times reported for orthogonal concentration gradients of tetracycline vs. chloramphenicol in the absence of *tetA* (Top) compared with *tetA* expression (Bottom). (E) The shift in TetR-CAT operational range is tunable by titration of *tetA* expression. The minimum tetracycline concentration required for growth (y axis) at a given selection pressure (x axis) for three *tetA* expression levels: none (diamonds), intermediate (triangles), high (circles). Error bars represent SEM of three biological replicates.

The upper bound of the range indicates that higher inducer concentration provides no additional fitness advantage. This range informs the utility of a sensor for optimizing a pathway. We measured the operational range of 10 sensor selectors; the MphR, TtgR, and TetR operational ranges were measured for multiple inducers (Fig. 2A and *SI Appendix, Table S1*).

Under selection pressure, most cells in a sensor selector strain population survive only when the target chemical is detected. However, a small fraction of the cells survive absent the chemical. “Evolutionary” escape results from mutations that permanently reduce selection sensitivity, and, additionally, natural sensors may not have evolved to completely repress the basal expression level of the genes they regulate. In our selections, the resultant constitutive or leaky selector expression generates false

positives, making it difficult to identify rare winners. Promoter engineering to optimize the placement of operator sites can yield very tight repression (16), but this approach requires specific development for each sensor. Instead, our standardized construction allows us to reduce the effect of leaky selector expression through common *cis*-regulatory modifications that are sensor independent. These modifications include appending a degradation tag to the selector to accelerate its proteolysis and mutating the ribosome binding site (RBS) of the selector gene to attenuate translation (Fig. 1B).

We implemented several modifications in the TtgR–TolC sensor selector strain for comparison. Appending *ssrA* degradation tag variants to TolC reduced escape, in correlation to the strength of the degradation tag (17), by as much as six orders of magnitude (Fig. 2B). However, reduced escape also reduced the operational range. We adjusted the spacing between the RBS and translation start site of TolC to achieve fine-grained translation control (18). Five of 10 spacing mutations reduced escape rate while maintaining a measurable operational range (Fig. 2B and *SI Appendix*, Fig. S6). For a dual selector strain, in which TtgR regulates both *tolC* and a kanamycin resistance gene, observed escape rates support the hypothesis of escape through leaky reporter expression: With both SDS and kanamycin present, the escape rate was much lower ($5.2 \pm 0.21 \times 10^{-8}$ cell/cell) than with either SDS alone ($1.7 \pm 0.092 \times 10^{-5}$ cell/cell) or kanamycin alone ($4.4 \pm 0.44 \times 10^{-4}$ cell/cell). Finally, we observed substantial escape rate reduction using two copies of the *ttgR* sensor gene and a single TolC selector (Fig. 2B). Because TtgR acts as a transcriptional repressor, evolutionary escape requires inactivating mutations to both gene copies, and higher sensor expression may reduce escape through tighter basal repression of the selector.

Sensors are useful for pathway optimization only when the intracellular concentration of the target chemical is within the operational range of the sensor. We hypothesized that expressing an exporter of the target chemical should decrease the intracellular concentration, shifting the operational range (Fig. 1B). We studied this export effect by expressing a tetracycline exporter, TetA, in cells that place the tetracycline-responsive sensor, TetR, in control of chloramphenicol acyltransferase (CAT) expression. When this strain expressed TetA, the entire operational range for tetracycline, including both the lower detection threshold and upper saturation point, shifted about 10-fold higher (Fig. 2D). This effect was tunable by controlling TetA expression from the arabinose-inducible pBAD promoter (Fig. 2E). The CAT selector was used here due to improved titration of drug sensitivity.

Pathway Evolution by Toggled Selection. To maximize the likelihood of identifying rare cells with a higher-production phenotype, we developed a toggled selection scheme (Fig. 1C) that preserves library complexity while eliminating evolutionary escapees. Evolutionary escapees are cells that acquire mutations to survive selection without producing the target chemical. This escape prevents the identification of rare winners in a selection, and confounds multiple rounds of evolution as these escapees out-compete the productive cells. Through toggled selection, we can selectively kill the escapees at each round, and carry over the productive cells for further improvements in subsequent rounds. Central to toggled selection is our choice of TolC (19) as a selector; its use was motivated by its utility for both positive selection (using sodium dodecyl-sulfate, SDS) and negative selection (using colicin E1). MAGE is highly mutagenic, increasing the escape rate from below 10^{-7} to above 10^{-3} after five cycles in the CdaR–TolC sensor selector strain. This increase could be reversed by incubation with colicin E1 (Fig. 2C), because evolutionary escapees evade SDS toxicity through mutations that constitutively express *tolC*, making them highly susceptible to colicin E1. Crucially, we ensure that productive cells are not also killed during negative selection by maintaining a pathway gene under tight transcriptional control, which prevents prematurely triggering the sensor (Fig. 1C). Following negative

selection, we induce the regulated enzyme to allow cells to produce the target chemical, and the sensor expresses *tolC* in proportion to chemical production. By toggling to positive TolC selection with SDS, we enrich for higher producers, and these can be characterized for their production phenotypes or subjected to further pathway evolution (Fig. 1C).

Naringenin Pathway. We implemented the toggled selection scheme to evolve *E. coli* toward higher production of two chemicals: naringenin and glucaric acid. Naringenin, a pharmacologically useful plant flavonoid molecule, was chosen because previous efforts serve to benchmark our optimization (2, 3, 20). *E. coli* requires four heterologous enzymes to synthesize naringenin from glucose: tyrosine ammonia lyase (TAL), 4-coumaroyl ligase (4CL), chalcone synthase (CHS), and chalcone isomerase (CHI) (3) (Fig. 3A). Because this pathway consumes tyrosine and malonyl-CoA, our strain engineering strategy targeted endogenous *E. coli* gene regulatory and coding loci to increase the availability of these precursors (*SI Appendix*, Table S4). As the focus of this work was genomic mutagenesis, the heterologous genes were left untargeted.

We performed FBA toward increased malonyl-CoA, because its availability limits naringenin production (*SI Appendix*, Table S6) (2, 20). FBA identified three key pathways: glycolysis, fatty acid biosynthesis, and the tricarboxylic acid (TCA) cycle (Fig. 3A and *SI Appendix*, Table S3). Greater flux through glycolysis by up-regulation of *gapA*, *pgk*, and *pdh* should increase pools of acetyl-CoA, which is converted to malonyl-CoA by acetyl-CoA carboxylase enzymes *accABCD*. Because acetyl-CoA is oxidized in the TCA cycle, we targeted for down-regulation TCA enzymes *mdh*, *fumBC*, and *acnAB*. To throttle acetyl- and malonyl-CoA consumption in fatty acid biosynthesis, we targeted *fabBDFH* for down-regulation. Availability of tyrosine, the other precursor for naringenin production, is limited by activity of two enzymes in aromatic biosynthesis, *aroG* (21) and *tyrA* (4) that are inhibited by 3-deoxy-D-arabinoheptulosonate 7-phosphate (DAHP) and chorismate, respectively. We targeted *aroG* and *tyrA* for coding sequence changes shown to alleviate product inhibition. These predictions (Fig. 3A) corroborate interventions experimentally shown to increase production of malonyl-CoA (20), tyrosine (22), and naringenin (2, 3).

Previous efforts to engineer the naringenin pathway have largely relied on plasmid-based overexpression or complete knockouts (20); for tightly regulated or essential central metabolism genes, such drastic modifications can have deleterious growth defects. For finer control of gene expression states, which can more closely balance biosynthetic and survival objectives, we used MAGE (6). Oligonucleotides for MAGE mutagenesis were targeted to Shine–Dalgarno sequences to finely increase or decrease translation efficiency, to alternative start codons (CTG, GTG, or TTG) to yield larger translational attenuation, or to premature stop codons or coding frameshifts for complete inactivation (*SI Appendix*, Table S4). Seven genes were identified by FBA for overexpression to increase flux through glycolysis and to convert acetyl-CoA to malonyl-CoA. MAGE oligonucleotides containing T7 promoter.

Four rounds of evolution by toggled selection were performed on the strain containing two copies of the *ttgR* gene controlling TolC, due to its favorable combination of escape rate and operational range (Fig. 2B). We verified that TtgR responds only to naringenin and cannot be induced by pathway intermediate coumaric acid (*SI Appendix*, Fig. S3). After four rounds, each consisting of about 15 cycles of targeted mutagenesis followed by toggled selection, the best strain identified produced 36 times more naringenin than the parent strain (Fig. 3C). We screened ~20 colonies to identify the highest producer at each round. With a supernatant concentration of 39 mg/L, the production titer of this strain surpasses the highest published production of naringenin (29 mg/L) directly from glucose (3) (Fig. 3C). We further enhanced the production titer to 61 mg/L by overexpressing *E. coli* acetyl-CoA carboxylase genes (*accABCD*), which have been

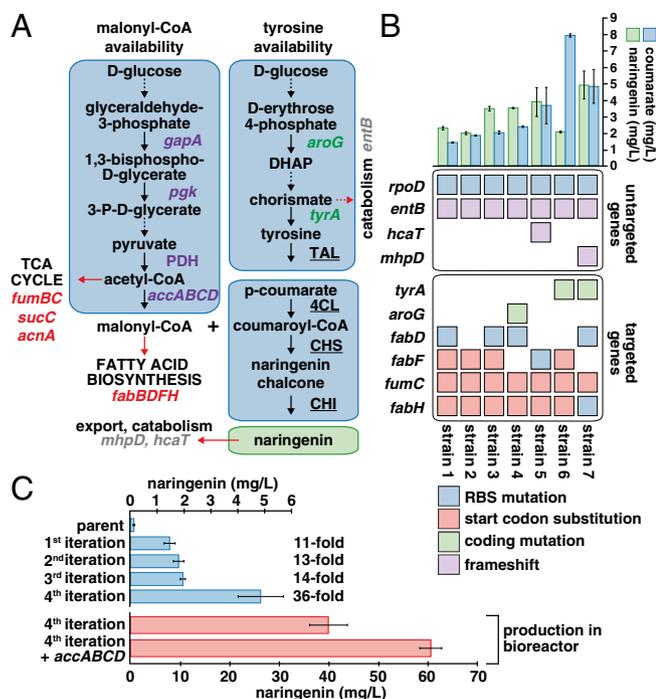


Fig. 3. Optimization of the naringenin biosynthetic pathway. (A) Endogenous *E. coli* genes targeted by MAGE to increase malonyl-CoA and tyrosine availability for naringenin production; targeted genes are colored: purple, up-regulation; red, down-regulation; green, coding changes; gray, untargeted knocked out genes. (B) Genotype and production phenotype of the top seven producers (in no particular order) from the fourth round of toggled selection. Colored boxes denote the type of genetic modification. Shown are mutations found at targeted genes (Bottom) and those at untargeted genes (Center). Naringenin (green bars) and coumaric acid (blue bars) concentrations for single production measurements are shown above the corresponding genotype (Top). Error bars represent SEM of three biological replicates. (C) Average naringenin production titers for parent and highest producer after each round of evolution (blue bars). Production titer from fed batch bioreactor fermentation of the highest producer and highest producer with *accABCD* overexpressed (red bars).

strain to increase endogenous malonyl-CoA levels (Fig. 3C and *SI Appendix*, Fig. S1). Through genetic changes alone, we were able to nearly recapitulate the high-naringenin titer (84 mg/L) previously achieved by addition of cerulenin, an inhibitor of fatty acid biosynthesis, which is prohibitively expensive for industrial-scale production (3).

We sequenced the genomes of the starting strain and seven high-producing strains isolated after round four. All seven strains incorporated RBS or start codon changes at several targeted loci (Fig. 3B). We found a number of mutations associated with malonyl-CoA production (Fig. 3B and *SI Appendix*, Table S7). In the TCA cycle, fumarase was down-regulated by a *fumC* start codon mutation in all seven strains (likely due to its selection in an early round). Several fatty acid genes were also down-regulated. Fatty acid biosynthesis genes whose products initiate synthesis from acetyl-CoA (*fabH*) or malonyl-CoA (*fabD*) were down-regulated by start codon or RBS mutations in seven and four strains, respectively. The fatty acid elongation gene *fabF* had start codon attenuation (GTG to TTG) in four strains and a purine to pyrimidine mutation in the RBS predicted to lower translation rate (23) in a fifth strain (Fig. 3B and *SI Appendix*, Table S7). None of the seven strains had a down-regulation target knocked out, and none of the strains had mutations affecting *fabB*, an essential gene, reflecting a balance between production and growth objectives. Computational prediction of translation rate shows that selected clones enrich for RBS and start codon mutations that

attenuate translation of genes, consistent with FBA predictions (*SI Appendix*, Fig. S4).

Three strains exhibited targeted mutations in tyrosine biosynthetic genes shown to alleviate product inhibition. All three produced substantially more coumaric acid, including two strains with the *tyrA* mutation A354V, which produced at least an order of magnitude more coumaric acid (Fig. 3B). This large coumaric acid buildup suggests that malonyl-CoA may be limiting for naringenin production in these strains. In support of this idea, overexpression of the enzymes *accABCD*, which convert acetyl-CoA to malonyl-CoA, increased naringenin production almost 1.5-fold in the evolved strain (Fig. 3C).

Although the MAGE process concentrates diversity generation on targeted loci and increases the probability of sampling specific mutations hypothesized to confer beneficial phenotypes, it also has unintended mutagenic effects. Whole-genome sequencing revealed many nontargeted mutations in the producer strains (Fig. 3B and *SI Appendix*, Table S8), including several mutations likely involved in higher naringenin production. Frameshifts inactivated *mhpD*, which catabolizes aromatic compounds similar to coumaric acid (24), and *hcaT*, a putative transporter of phenylpropionates like coumaric acid (25). Similarly, a frameshift in *entB*, which diverts chorismate from aromatic biosynthesis, may increase tyrosine production (26). We speculate that knocking out all three enzymes facilitates production of naringenin by increasing the concentration of the precursor, p-coumaric acid. Attributing function to noncoding regulatory mutations is more tenuous. However, we observed a mutation in the Shine–Dalgarno sequence of *rpoD*, mutation of which increases tyrosine production (22).

Glucaric Acid Pathway. To validate directed evolution by sensor selectors as a generalizable method, we optimized the production of glucaric acid in *E. coli*. Glucaric acid was chosen for two reasons. First, unlike naringenin production, previous work to modulate endogenous pathways was absent. Second, glucaric acid was identified as a key renewable chemical for the replacement of petroleum-based polymer production. Glucaric acid can be synthesized in *E. coli* by expression of three exogenous enzymes: myo-inositol-1-phosphate synthase (Ino1), myo-inositol oxygenase (MIOX), and uronate dehydrogenase (Udh) (27) (Fig. 4A).

To ensure that the heterologous enzymes were functional and provided a growth advantage under selective conditions, we measured growth lag times in the CdaR-TolC sensor selector strain after exogenously providing pathway intermediates (glucose, myo-inositol, and glucuronic acid). Furthermore, we verified CdaR is specifically activated by glucaric acid, and does not respond to pathway intermediates myo-inositol and glucuronic acid (*SI Appendix*, Fig. S7).

Increasing concentrations of glucaric acid result in lower lag times for cells grown in the presence of SDS. Under selective conditions, decreasing growth lag times reflect the decreasing number of enzymatic reactions required to produce glucaric acid for CdaR-TolC activation (Fig. 4B). Higher concentrations of myo-inositol and glucuronic acid resulted in shorter lag times under selective conditions, but increasing glucose or glucaric acid concentrations in the media did not result in a growth advantage. In the case of glucaric acid, this is expected, as both 1 mM and 10 mM are above the operational range. With glucose, one possible explanation is that an increase in glucose in the media results in additional flux through glycolysis and central metabolism rather than increased flux through the glucaric acid pathway, which likely operates slower than glycolysis. The lag time observed at the high glucuronic acid concentration is comparable to the lag time observed with glucaric acid, supporting the previous finding that the Udh enzyme acts on a fast time scale compared with the selection (27, 28). A long lag time even at a high concentration of myo-inositol indicates that the MIOX enzyme is less efficient, as reported in previous work (29).

Efforts to increase glucaric acid production in *E. coli* have focused on colocalization of pathway enzymes (30) and improving MIOX solubility (29). To date, modifying endogenous *E. coli* pathways has not been explored. We hypothesized that glycolysis and the pentose phosphate pathway were competing with Ino1 for glucose-6-phosphate (g6p), the branch-point for glucaric acid production. We used MAGE to introduce degeneracy in the RBS of genes involved in catabolism of g6p (*SI Appendix, Table S5*). We similarly targeted the RBS sequences of *mdh* and *suhB*, the endogenous phosphatase responsible for dephosphorylating myo-inositol-1-phosphate (31) (Fig. 4A). Degeneracy in the RBS sequences allowed the selection to sample both up- and down-regulation of the genes. We hypothesized that tuning the rate of glycolysis would allow the glucaric acid pathway to compete for glucose more effectively while still facilitating robust cell growth. The product of the *pgi* gene shuttles g6p into glycolysis and its disruption has been shown to increase the intracellular pool of g6p (32), the substrate of Ino1. The growth defect of a *pgi* mutant can be rescued by overexpression of *sthA* (31) and thus *pgi* and *sthA* were chosen for simultaneous expression modulation. The other major pathway for g6p catabolism is the pentose phosphate pathway and is initiated by the product of *zwf*, which was also targeted for expression modification. To prevent flux diversion of the intermediate molecule glucuronic acid into the Entner–Doudoroff pathway, we targeted uronate isomerase (*uxaC*) for a knockout. To avoid catabolism of glucaric acid, we also targeted glycerate kinase (*garK*) for a knockout.

We performed five cycles of MAGE on seven genomic targets (*SI Appendix, Table S5*) to achieve a predicted prevalence of $\sim 1 \times 10^{-6}$ for strains incorporating mutations at all seven loci. The statistically most common strain contained a single mutation and was predicted to account for 40% of the cell population.

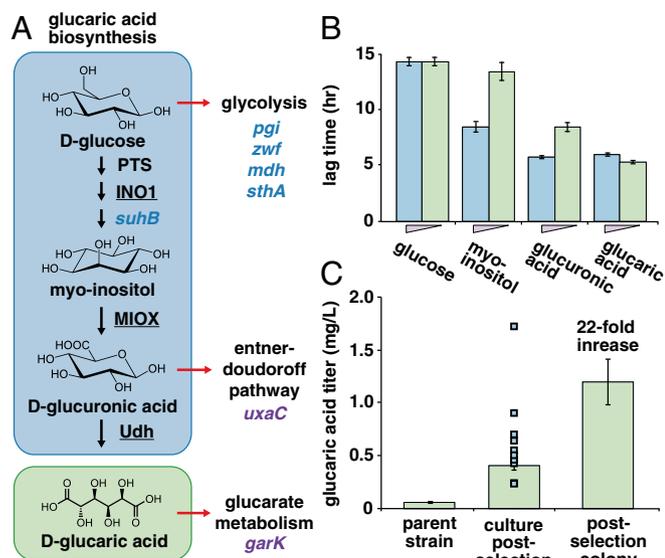


Fig. 4. Optimization of the glucaric acid biosynthetic pathway. (A) Glucaric acid biosynthetic pathway showing key intermediate metabolites and enzymes. Heterologous gene names are underlined. Endogenous *E. coli* genes targeted by MAGE for expression modification: blue, RBS modification; purple, knockout. (B) Lag time in growth reflects time required for the pathway enzymes to produce activating levels of glucaric acid in the sensor selector strain CdaR–ToIC. Pathway intermediates are supplied exogenously (blue, 10 mM; green, 1 mM). Error bars represent SEM from three biological replicates. (C) Glucaric acid titers produced by the parent strain, the postselection mixed population, and the highest-producing clone (bars). Squares indicate titers produced by clones isolated from the postselection population. Error bars represent SEM from three biological replicates.

After MAGE followed by toggled selection, the enriched non-clonal culture produced sevenfold more glucaric acid than the parent. The best clone isolated from this population produced 22-fold more than the parent (Fig. 4C and *SI Appendix, Table S7*). This highest-producing strain contained a targeted nonsense mutation in *garK*, a gene not previously shown to enhance glucaric acid production. None of the other targeted genes were mutated, but an untargeted nonsense mutation in the L-glyceraldehyde 3-phosphate reductase gene (*yghZ*) was found. As an aldo-keto reductase, *yghZ* has fairly broad substrate specificity (33) and could be diverting carbon flux away from glucaric acid by reducing glucuronate to gluconate.

Glucaric acid titers were improved 22-fold over the parent strain; however, absolute production of glucaric acid remained substantially lower (1.2 mg/L, Fig. 4C) than previously reported titers (27). Moon et al. carried out glucaric acid production in an *E. coli* B strain (BL21), whereas we optimized the pathway in the MAGE-competent *E. coli* K strain. To investigate the possible role of strain background (B vs. K strains) in glucaric acid production, we measured glucaric acid titer in our parent K strain and BL21. We found that glucaric acid titer was 300 times higher in BL21 with the same glucaric acid enzymes and culture conditions (*SI Appendix, Fig. S5*).

There are substantial differences between B and K strains of *E. coli* that are difficult to bridge through naïve mutagenesis. Notably, B strains have altered carbohydrate metabolism compared with K strains, as well as an enhanced capacity for recombinant protein production. Previous work to produce glucaric acid in *E. coli* has revealed MIOX to be a highly unstable enzyme (27), and the primary limit on production may lie in protein folding and stability, rather than host cell glucose metabolism. Our evolved K strain grew just slightly worse than the parent strain, ruling out gross metabolic deficiency as the cause of low production (*SI Appendix, Fig. S8*). In light of these considerations, subsequent rounds of diversification and selection were not pursued in the K-strain background. Currently, work is underway to enable MAGE in BL21 for optimization of production pathways better suited for *E. coli* B strains. These results highlight that directed evolution is not a replacement for the careful choice of a host strain, but should complement thoughtful strain selection.

Discussion

Rapid advances in DNA sequencing and DNA synthesis technologies (34, 35) have not been accompanied by similar advances to enable the high-throughput evaluation of phenotypes. Our implementation of small-molecule sensors coupled to selection advances a versatile platform that can transform biosynthetic phenotypes into fitness differences. These differences empower us to use evolution followed by sequencing to reveal clues to potential metabolic pathway inefficiencies and to identify targets for subsequent rounds of evolution. The multiplex mutations facilitated by MAGE enable us to target all candidate genes predicted by FBA without prior assumption of the relative importance of each target. Because selection amplifies faster-dividing cells, we indirectly enrich for variants that suitably balance biomass and biosynthetic objectives. We show that toggled selection refreshes the pool of productive cells by removing evolutionary escapees. Toggled selection enables multiple rounds of evolution to progressively enrich for higher-producing variants. Combining beneficial mutations from independently evolved strains could lead to even higher metabolite production due to epistatic synergies. The incidence of evolutionary escapees and off-target mutations is likely to be significantly reduced by transiently repressing mismatch repair (36). Although this may decrease untargeted beneficial mutations (*mhpD*, *entB*, and *hcaT* in naringenin biosynthesis) in a single round of evolution, mutations that provide significant selective advantage will ultimately be enriched over multiple rounds.

Besides pathway optimization, we can use sensor selectors to screen libraries of synthetic or metagenomic sequences for novel

biosynthetic operons, new enzyme functions, and transporters. The vast reservoir of natural chemicals found in microbial species remains largely inaccessible because the enzymatic pathways for their synthesis are not known. With sensor selectors, large libraries encoding natural or synthetic operons can be interrogated to identify the putative pathway for a target chemical.

Natural sensor domains exist for many classes of molecules that are of economic interest; however, some metabolite targets have no known sensor to detect them. We expect this challenge to be addressed by advances in protein design and by efforts to characterize new transcription factors encoded in metagenomes. Clever use of existing sensors will also allow the optimization of multiple pathways that use common intermediates. For biosynthetic pathways diverging only in late “decoration” steps, we can leverage class-specific sensors to optimize the production of many related molecules by simply exchanging terminal enzymes. For example, our best naringenin production strain likely has an elevated intracellular concentration of malonyl-CoA, which could be used immediately for the improved production of fatty acid-derived targets or polyketides.

Evolution is a powerful tool for resolving the complexity of biology. Using evolution to guide rational design should ultimately lead to a better understanding of the genotypic basis of biological function.

Methods

Sensor Selector Strain Construction. All sensor selector strains were constructed from *E. coli* MG1655 derivative EcNR2 (Δ bioAB::Red- λ prophage-*bla* Δ mutS::Crm) to facilitate recombineering and MAGE (6). Sensor selector constructs were genomically integrated using a standard genetic architecture (Fig. 1A).

Glucaric Acid Pathway Construction and Optimization. A plasmid (pT7GAEXP) enabling glucaric acid biosynthesis in *E. coli* was constructed, encoding: the *Mus musculus* myo-inositol-oxygenase (MIOX) gene; the *Saccharomyces cerevisiae* inositol-1-phosphate synthase (INO1) gene; and the *Agrobacterium tumefaciens* uronate dehydrogenase (Udh) gene. MAGE (6) mutagenesis was used to target seven genes (SI Appendix, Table S5) for expression changes in strain CdaR-TolC. One cycle of toggled negative and positive selection was used to enrich for mutations benefiting glucaric acid production, as assayed by clonal production and mass spectrometry.

Naringenin Pathway Construction and Optimization. Four heterologous genes enabling naringenin production were cloned into two plasmids for expression in a TtgR-TolC sensor-selector strain: tyrosine-ammonia lyase (TAL), 4-coumarate:CoA ligase (4CL), chalcone synthase (CHS), and chalcone isomerase (CHI) (3). MAGE (6) mutagenesis targeted 20 endogenous genes (SI Appendix, Table S3) for expression and coding changes in this strain. Four iterations of mutation and toggled negative and positive selection enriched for mutations benefiting naringenin production, as assayed by clonal production and mass spectrometry.

See SI Appendix for full methods.

- Nakamura CE, Whited GM (2003) Metabolic engineering for the microbial production of 1,3-propanediol. *Curr Opin Biotechnol* 14(5):454–459.
- Leonard E, Lim K-H, Saw P-N, Koffas MAG (2007) Engineering central metabolic pathways for high-level flavonoid production in *Escherichia coli*. *Appl Environ Microbiol* 73(12):3877–3886.
- Santos CNS, Koffas M, Stephanopoulos G (2011) Optimization of a heterologous pathway for the production of flavonoids from glucose. *Metab Eng* 13(4):392–400.
- Lütke-Eversloh T, Stephanopoulos G (2005) Feedback inhibition of chorismate mutase/prephenate dehydrogenase (TyrA) of *Escherichia coli*: Generation and characterization of tyrosine-insensitive mutants. *Appl Environ Microbiol* 71(11):7224–7228.
- Yim H, et al. (2011) Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat Chem Biol* 7(7):445–452.
- Wang HH, et al. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460(7257):894–898.
- van Sint Fiet S, van Beilen JB, Witholt B (2006) Selection of biocatalysts for chemical synthesis. *Proc Natl Acad Sci USA* 103(6):1693–1698.
- Tang S-Y, Cirino PC (2011) Design and application of a mevalonate-responsive regulatory protein. *Angew Chem Int Ed Engl* 50(5):1084–1086.
- Binder S, et al. (2012) A high-throughput approach to identify genomic variants of bacterial metabolite producers at the single-cell level. *Genome Biol* 13(5):R40.
- Yang J, et al. (2013) Synthetic RNA devices to expedite the evolution of metabolite-producing microbes. *Nat Commun* 4:1413.
- Dietrich JA, Shis DL, Alikhani A, Keasling JD (2013) Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis. *ACS Synth Biol* 2(1):47–58.
- Tang S-Y, et al. (2013) Screening for enhanced triacetic acid lactone production by recombinant *Escherichia coli* expressing a designed triacetic acid lactone reporter. *J Am Chem Soc* 135(27):10099–10103.
- Gallegos MT, Schleich R, Bairoch A, Hofmann K, Ramos JL (1997) AraC/XylS family of transcriptional regulators. *Microbiol Mol Biol Rev* 61(4):393–410.
- Ramos JL, et al. (2005) The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev* 69(2):326–356.
- Tropel D, van der Meer JR (2004) Bacterial transcriptional regulators for degradation pathways of aromatic compounds. *Microbiol Mol Biol Rev* 68(3):474–500.
- Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/11-12 regulatory elements. *Nucleic Acids Res* 25(6):1203–1210.
- Andersen JB, et al. (1998) New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl Environ Microbiol* 64(6):2240–2246.
- Chen H, Bjercknes M, Kumar R, Jay E (1994) Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. *Nucleic Acids Res* 22(23):4953–4957.
- DeVito JA (2008) Recombineering with tolC as a selectable/counter-selectable marker: Remodeling the rRNA operons of *Escherichia coli*. *Nucleic Acids Res* 36(1):e4.
- Xu P, Ranganathan S, Fowler ZL, Maranas CD, Koffas MAG (2011) Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. *Metab Eng* 13(5):578–587.
- (2005) US Patent 7,482,140.
- Santos CNS, Xiao W, Stephanopoulos G (2012) Rational, combinatorial, and genomic approaches for engineering L-tyrosine production in *Escherichia coli*. *Proc Natl Acad Sci USA* 109(34):13538–13543.
- Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* 27(10):946–950.
- Burlingame R, Chapman PJ (1983) Catabolism of phenylpropionic acid and its 3-hydroxy derivative by *Escherichia coli*. *J Bacteriol* 155(1):113–121.
- Diaz E, Ferrández A, García JL (1998) Characterization of the hca cluster encoding the dioxygenolytic pathway for initial catabolism of 3-phenylpropionic acid in *Escherichia coli* K-12. *J Bacteriol* 180(11):2915–2923.
- Gehring AM, Bradley KA, Walsh CT (1997) Enterobactin biosynthesis in *Escherichia coli*: Isochorismate lyase (EntB) is a bifunctional enzyme that is phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2,3-dihydroxybenzoate. *Biochemistry* 36(28):8495–8503.
- Moon TS, Yoon S-H, Lanza AM, Roy-Mayhew JD, Prather KLJ (2009) Production of glucaric acid from a synthetic pathway in recombinant *Escherichia coli*. *Appl Environ Microbiol* 75(3):589–595.
- Yoon S-H, Moon TS, Iranpour P, Lanza AM, Prather KJ (2009) Cloning and characterization of uronate dehydrogenases from two pseudomonads and *Agrobacterium tumefaciens* strain C58. *J Bacteriol* 191(5):1565–1573.
- Shiue E, Prather KLJ (2014) Improving D-glucaric acid production from myo-inositol in *E. coli* by increasing MIOX stability and myo-inositol transport. *Metab Eng* 22:22–31.
- Moon TS, Dueber JE, Shiue E, Prather KLJ (2010) Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli*. *Metab Eng* 12(3):298–305.
- Canonaco F, et al. (2001) Metabolic flux response to phosphoglucose isomerase knock-out in *Escherichia coli* and impact of overexpression of the soluble transhydrogenase UdhA. *FEMS Microbiol Lett* 204(2):247–252.
- Fraenkel DG (1968) The accumulation of glucose 6-phosphate from glucose and its effect in an *Escherichia coli* mutant lacking phosphoglucose isomerase and glucose 6-phosphate dehydrogenase. *J Biol Chem* 243(24):6451–6457.
- Grant AW, Steel G, Waugh H, Ellis EM (2003) A novel aldo-keto reductase from *Escherichia coli* can increase resistance to methylglyoxal toxicity. *FEMS Microbiol Lett* 218(1):93–99.
- Carr PA, Church GM (2009) Genome engineering. *Nat Biotechnol* 27(12):1151–1162.
- Kosuri S, et al. (2013) Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proc Natl Acad Sci USA* 110(34):14024–14029.
- Nyerges Á, et al. (2014) Conditional DNA repair mutants enable highly precise genome engineering. *Nucleic Acids Res* 42(8):e62.