Inducible directed evolution of complex phenotypes in bacteria

Ibrahim S. Al'Abri, Daniel J. Haller, Zidan Li and Nathan Crook [ID]*

Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC, USA

ABSTRACT

Directed evolution is a powerful method for engineering biology in the absence of detailed sequence-function relationships. To enable directed evolution of complex phenotypes encoded by multigene pathways, we require large library sizes for DNA sequences >5–10 kb in length, elimination of genomic hitchhiker mutations, and decoupling of diversification and screening steps. To meet these challenges, we developed Inducible Directed Evolution (IDE), which uses a temperate bacteriophage to package large plasmids and transfer them to naive cells after intracellular mutagenesis. To demonstrate IDE, we evolved a 5-gene pathway from *Bacillus licheniformis* that accelerates tagatose catabolism in *Escherichia coli*, resulting in clones with 65% shorter lag times during growth on tagatose after only two rounds of evolution. Next, we evolved a 15.4 kb, 10-gene pathway from *Bifidobacterium breve* UC2003 that aids *E. coli*'s utilization of melezitose. After three rounds of IDE, we isolated evolved pathways that both reduced lag time by more than 2-fold and enabled 150% higher final optical density. Taken together, this work enhances the capacity and utility of a whole pathway directed evolution approach in *E. coli*.

INTRODUCTION

Many important phenotypes emerge from the interactions between multiple genes [1,2]. These 'complex' phenotypes have traditionally encompassed small molecule biosynthesis [3], tolerance to inhibitors [4] and growth in new habitats [5]. However, advances in synthetic biology and metabolic engineering have revealed that even supposedly 'simple' phenotypes, such as production of a recombinant protein, become increasingly complex as higher performance is desired. This is because auxiliary cellular functions, such as chaperone proteins, cell wall synthesis and secretion machinery can become limiting [6–8]. Clearly, engineering these 'systems-level' phenotypes requires systems-level techniques.

In bacteria, complex phenotypes can be improved via adaptive laboratory evolution (ALE) [5,9,10] or iterative genome-wide sequence diversification or expression perturbation screens (for example, multiplex automated genome engineering (MAGE) [11], asRNA [12,13] and CRISPRi/a [14–16]). One downside of adaptive evolution is the accumulation of genomic hitchhiker mutations, which are changes to off-target genomic DNA sequences. Hitchhiker mutations can be particularly troublesome for biosensor-coupled screens and that make learning from these experiments time-consuming. Furthermore, while the use of hypermutator strains can accelerate ALE [17], hypermutation cannot be turned off in these strains, confounding downstream characterization. For asRNA and CRISPRi/a, the researcher is limited to sampling changes to expression space, rather than the much larger space of protein bioactivity.

For these reasons, directed evolution is useful because it enhances and directs mutations to defined DNA sequences and samples a much wider evolutionary sequence space. However, due to the limited length of DNA that can be evolved using most methods, it has been difficult to apply directed evolution to complex phenotypes. For example, traditional error-prone PCR-based libraries are effectively limited to sequences <10 kb in length due to reductions in polymerase processivity, cloning efficiency, and transformation rate above this size. Although recent methods for directed evolution in bacteria have eliminated many hands-on steps via the use of filamentous phage (phage-assisted continuous and non-continuous evolution (PACE, PANCE) [18–20], and phagemid-assisted continuous evolution (PACEmid) [21]), they are usually limited to small regions of DNA (<5 kb) and often couple the mutagenesis and screening steps. This is because the phage used in these techniques (M13) has a strict packaging limit (5 kb) and is engineered to replicate as soon as a certain threshold of biological activity has been reached [22]. Additionally, guided directed evolution approaches such as EvolvR [23], eMutaT7 [24], T7-DIVA [25] offer rapid molecular diversification for desired sequences. However, the window of diversification with these methods is limited (<3 kb [24]) and off-target mutations are bound to occur as the number of cycles increases.

---

*To whom correspondence should be addressed. Tel: +1 919 513 2429; Fax: +1 919 515 3465; Email: nccrook@ncsu.edu

Here we add Inducible Directed Evolution (IDE), a new approach that combines the perks of recently developed directed evolution methods to evolve complex phenotypes, to the directed evolution toolkit. IDE harnesses the large genomes of temperate phages (40–100 kb) to evolve large DNA segments (26), avoids the accumulation of off-target genomic mutations, and decouples mutagenesis and screening steps.

MATERIALS AND METHODS

Strains and media

All strains (Supplementary Table S1) and oligos (Supplementary Table S2) used in this study are in the supplementary file. *Escherichia coli* Top10, NEB 5α and NEB 10β were used for plasmid construction. *E. coli* C600 (CGSC 5394 C600) was used for optimization of infection and packaging rates. *E. coli* Nissle 1917 and *E. coli* MG1655 were used to test IDE's applicability to different *E. coli* strains. All *E. coli* strains were grown in lysogeny broth (LB) (5 g/L yeast extract, 10 g/L tryptone, 10 g/l NaCl) at 37°C supplemented with ampicillin (Amp) (100 μg/mL), kanamycin (Kan) (50 μg/mL) or chloramphenicol (Cm) (34 μg/mL). Strains containing mutagenesis plasmids were grown in LB containing 1% (w/v) D-glucose and appropriate antibiotics. *Bacillus licheniformis* (ATCC® 14580™) was grown in Difco™ Nutrient Broth. All infection and packaging experiments were performed in Phage Lysate Media (PLM, LB with 100 mM MgCl$_2$ and 5 mM CaCl$_2$) or enhanced PLM (ePLM, LB 140 mM MgCl$_2$ and 7 mM CaCl$_2$). P1kcΔcoi::*kanR* and P1kc:10kb::*kanR* phages were gifts from Dr. Chase Beisel.

Preparation and transformation of electrocompetent cells

Overnight cultures of the desired strains were inoculated (1:100 dilution) into 50 mL LB media containing appropriate antibiotics and grown to OD$_{600}$ 0.8 (37°C, 250 rpm). Cells were chilled on ice for 15 minutes before being pelleted by centrifugation at 3000 × g for 5 min. Next, cells were washed twice via resuspension in 25 mL of 10% glycerol and pelleting via centrifugation at 3000 × g for 5 min. Washed cells were then resuspended in 1 mL 10% glycerol and pelleted at 4000 × g for 3 min in eppendorf microcentrifuge tubes. The resulting pellets were resuspended in 0.5 mL of 10% glycerol and divided into 50 μl aliquots that were either used for transformation immediately or stored at −80°C. Frozen cells were thawed on ice for 10 min before being used for transformation. Fresh cells yielded higher transformation efficiency.

Cloning

Plasmids and phagemids are listed in Supplementary Table S1. Primers used in this study were obtained from Eurofins Genomics and are listed in Supplementary Table S2. NEBuilder® HiFi DNA Assembly Master Mix was used for plasmid and phagemid construction. SGI-DNA Gibson Assembly® (GA) HiFi 1-Step Kit assembly was used for construction of large phagemids. NEB Q5® Site-Directed Mutagenesis (SDM) Kit was used to introduce point mutations according to manufacturer instructions. Addgene #40782 was used as the backbone for all phagemid cloning, and primers 1 and 2 were used to amplify this backbone for Gibson cloning.

Construction of large phagemids

A 24.0 kb (PM-24) P1 phagemid (Addgene #40784) was used as a backbone for constructing a 42 kb phagemid (PM-42.0). To construct PM-42.0, 3 PCR reactions were used to amplify 5–7 kb fragments from the *Saccharomyces cerevisiae* genome in addition to the phagemid backbone. NEB Q5® High-Fidelity 2X Master Mix was used to amplify the parts and SGI-DNA Gibson Assembly® (GA) HiFi 1-Step Kit was used to assemble the parts. The assembled product was transformed into *E. coli* DH10B (ISA585). ZymoPURE™ II Plasmid Midiprep Kit (Catalog No. D4200 & D4201) was used to extract PM-42.0 from ISA585.

Phage production

Overnight cultures of strains containing P1 and the phagemid were subinoculated into ePLM (1:100 dilution) with appropriate antibiotics. At OD$_{600}$ 0.8–1.0 cell cultures were induced with 20% L-arabinose (1/100 culture volume) and put back to the shaking incubator (37°C, 250 rpm). After 2 hours, the cultures were removed from the incubator and transferred to 15 mL centrifuge tubes containing chloroform (1/40 culture volume). The tubes were left on ice for 5 min with gentle mixing or pipetting every minute. The tubes were then centrifuged at 3000 × g for 10 min at 4°C. The produced phage (present in the supernatant) was then transferred to sterile tubes for storage. Phage lysate is stable at 4°C for 1 year and indefinitely at −80°C.

Phage infection

An overnight culture (37°C, 250 rpm) grown in LB with appropriate antibiotics was subinoculated into ePLM (1:100 dilution). At OD$_{600}$ 1.0, the cells were spun down at 3000 × g for 5 min. The supernatant was discarded and the pellet was resuspended in $\frac{1}{3}$ volume of fresh ePLM. The cells were then added to 1 mL phage lysate in a 14 mL falcon culture tube. The infection mixture was incubated (37°C, 250 rpm) for 20 min and then moved to a 37°C standing incubator for 20 min. The infection activity was quenched with 1 mL of Super Optimal Broth with glucose (SOC) containing 200 mM sodium citrate. The mixture was then incubated for 40 min (37°C, 250 rpm) before being transferred to 50 ml of LB media with the appropriate antibiotics or plated on LB agar plates containing the appropriate antibiotics.

Cell growth phase experiments

Three *E. coli* C600 colonies grown overnight in LB containing appropriate antibiotics were subinoculated (at 1:100, 1:200 and 1:300 dilution) into three 50 mL aliquots of PLM and grown to OD$_{600}$ 0.5, 1.0 and 1.5 in a shaking incubator (37°C, 250 rpm). Cell cultures were harvested (3000 × g

for 5 min) and resuspended with PLM to obtain an identical cell count per mL (3×10^8 CFU/mL). The cultures were then infected with 1 mL phage lysate produced from ISA199 (*E. coli* C600 containing P1kc and 12 kb phagemid). Infected cultures were plated on LB with chloramphenicol to count CFUs.

Large phagemid infection rate experiments

PM-9 (9.0 kb), PM-12 (12.8 kb), PM-24 (24.0 kb) and PM-42.0 (42.0 kb) phagemids were transformed into *E. coli* C600 containing only P1kc:10kb or P1kc:10kb and aTc-MP. Phage lysate was produced from each strain as described above and then used to infect wild type *E. coli* C600 or *E. coli* C600 containing P1kc::10kb. Infected strains were plated on LB with chloramphenicol plates to count CFU/mL.

Infecting different *E. coli* strains

The 12 kb phagemid (PM-12) was transformed into *E. coli* C600, MG1655, and Nissle containing P1kc:10kb::*kanR*. Three biological replicates from the transformed strains were grown overnight in LB with chloramphenicol and used for phage production. Phage lysate from each strain was used to infect wild type *E. coli* C600, MG1655 and Nissle.

Flip recombinase to delete kanR from P1kcΔcoi::kanR and P1kc:10kb::kanR

Flip recombinase protocol adapted from barricklab.org (27). In brief, *E. coli* C600 containing P1kcΔcoi::*kanR* (ISA137) and P1kc:10kb::*kanR* (ISA138) were first transformed with pCP20 and plated in LB with ampicillin agar plates and grown overnight at 30°C. Cultures were inoculated into 1.0 mL of LB in eppendorf microcentrifuge tubes and grown overnight at 43°C to induce FLP recombinase expression and plasmid loss. Colonies were then screened for loss of *kanR* from the phage genome via plating in LB with kanamycin, LB with ampicillin and LB only plates.

Stop codon reversion

ampR was inserted into a phagemid backbone (ISA012, Plasmid Map 1 (28)) via Gibson assembly. Mutations were introduced via NEB Q5 SDM to introduce either one or two stop codons in *cmR*. This phagemid, with a dysfunctional *cmR,* was transformed into *E. coli* C600 containing aTc-MP and P1 phage (ISA308 and ISA311). Three biological replicates were grown overnight in LB containing 1% (w/v) D-glucose, kanamycin, and ampicillin. The overnight cultures were plated in LB with chloramphenicol agar plates to check for escapers and background aTc-MP activity; no escapers were detected. The cultures were spun down at $4000 \times g$ for 3 min in eppendorf microcentrifuge tubes and washed with $1\times$ PBS twice to remove the residual glucose from the media. Washed cells were then transferred to LB media containing kanamycin and ampicillin (1:1000) and with or without 200 ng/mL aTc to induce aTc-MP. The cultures were plated after 8 or 16 hours of induction on LB with chloramphenicol plates to count the number of resistant colonies. Random colonies were picked to verify stop codon reversion via colony PCR.

Rifampicin resistance mutation rate assay

Mutation rate was measured using the rifampicin resistance assay as previously described (29,30). MP6 and aTc-MP were transformed into *E. coli* C600 containing P1kc. Transformations were plated on LB agar containing 1% (w/v) D-glucose and appropriate antibiotics. Three colonies from each plate were grown in LB liquid media containing 1% (w/v) D-glucose and appropriate antibiotics. Overnight cell cultures were harvested (3000 × g for 5 min), resuspended in LB and then subinoculated (1:100 dilution) in LB with appropriate antibiotics and either with or without 1% (w/v) D-glucose. The cultures were grown for 1 hours (37°C, 250 rpm). Cultures without D-glucose were induced with 10 mM L-arabinose (MP6) or with 200 ng/mL anhydrotetracycline. Cultures were then grown for 20 hours (37°C, 250 rpm). Serial dilutions (10^0–10^6) of each sample were plated on LB only and LB containing 100 μg/mL rifampicin agar plates. Plates were incubated overnight at 37°C. CFUs were counted the next day. Substitutions per base pair of the *E. coli* genome per generation (mutation rates: μ_{bp}) were calculated using the equation previously described in (29,30). $\mu_{bp} = f/[R \times \ln(N/N_0)]$, where *f* is the frequency of rifampin-resistant mutants (CFUs counted on rifampicin plates divided by CFUs counted on glucose plates for each sample), *R* is the number of unique sites yielding rifampin resistance (77 previously identified sites), *N* is the final population size (5×10^8) and N_0 is the population size at which resistance is first observed ($\sim 1.5 \times 10^7$ CFU/mL based on prior work)

Improving pSC101-sfGFP fluorescence via IDE (see Supplementary methods 1)

The pSC101 origin and *sfGFP* were inserted into ISA012 (Plasmid Map 1 (28)) via Gibson assembly. The resulting phagemid (pSC101-sfGFP) was then transformed into *E. coli* C600 containing aTc-MP and P1kc:10kb (ISA426). The resulting strain was mutated 2 times before selection for increased colony fluorescence. Colonies that exhibited higher GFP fluorescence were visually identified and analyzed via flow cytometry to quantify GFP production. Mutations that were found in pSC101 were introduced to the unevolved phagemid via NEB Q5 SDM, transformed into *E. coli* C600, and analyzed via flow cytometry to confirm their causal effects.

Tagatose pathway evolution (see Supplementary methods 2)

The tagatose pathway from *Bacillus licheniformis* (ATCC® 14580™) was inserted into ISA012 (Plasmid Map 1 (28)) via Gibson assembly. The resulting phagemid (ISA179, Plasmid Map 2) was then transformed into *E. coli* C600 containing aTc-MP and P1kc:10kb. The resulting strain was evolved through two rounds of diversification and selection. The resulting phage lysate was used to infect wild-type *E. coli* C600, and the resulting cells were plated on tagatose minimal media agar plates. Large colonies were picked and grown overnight in LB media containing chloramphenicol. The cultures were then washed with tagatose minimal media and grown for 40 hours in a plate reader (see Materials and Methods: growth curves). The pathways in the

colonies with faster growth rates were sequenced via Sanger sequencing. Resulting mutations were then reintroduced to unevolved phagemid via Q5 SDM and analyzed again for growth in tagatose minimal media.

Melezitose pathway evolution (see Supplementary methods 3)

The melezitose pathway from *Bifidobacterium Breve* UC2003 was inserted into the phagemid backbone from the last cycle of the tagatose pathway evolution experiment via 3-part Gibson assembly. The backbone was amplified in one PCR reaction and the melezitose pathway was amplified in PCR reactions yielding 7.5 kb and 8.0 kb fragments. The resulting phagemid (ISA861, Plasmid Map 3) was then transformed into *E. coli* C600 containing aTc-MP and P1kc:10kb. The resulting strain was evolved through three rounds of diversification and selection. The produced phage lysate was used to infect wild-type *E. coli* C600, and the infected cells were plated on melezitose minimal media agar plates. Large colonies were picked and grown overnight in LB media containing chloramphenicol. The cultures were then washed with melezitose minimal media and grown for 24 hours in a plate reader (see methods: growth curves). The pathways in the colonies with faster growth rates were sequenced with plasmidsaurus sequencing services (www.plasmidsaurus.com). Evolved clones were re-transformed to wildtype *E. coli* C600 to confirm enhanced growth curves.

Tagatose and melezitose minimal media

The selection media for the tagatose minimal media was adapted from Van der Heiden, *et al.* (2013) (31) with few modifications for *E. coli* C600. The $20\times$ Salt solution was composed of KH_2PO4 (54.4 g), K_2HPO_4 (208.8 g) and NH_4Cl (12 g) for 1 L solution. $500\times$ mineral solution was composed of $MgCl_2 \cdot 6H_2O$ (1 g), $CaCl_2 \cdot 2H_2O$ (0.25 g), $FeCl_2 \cdot 4H_2O$ (25 mg), $ZnSO_4 \cdot 7H_2O$ (25 mg), $CoCl_2 \cdot H_2O$ (12.5 mg), $CuSO_4 \cdot 5H_2O$ (0.5 mg) and $MnSO_4 \cdot H_2O$ (0.14 g) in 100 mL solution. $100\times$ vitamin mix (50 mL) was composed of 5 mg of thiamine-HCl, nicotinic acid, folic acid, D-L-pantothenic acid, D-biotin, leucine, lysine, homoserine and riboflavin and 10 mg of pyridoxal–HCl. 0.1% (w/v) I-casamino acids and 1% (w/v) tagatose or melezitose were added to the minimal media.

Growth curves

Three biological replicates of each isolate were grown overnight in 96-deep-well plates (VWR International, cat #10755-248) in LB media. Saturated cultures were spun down at $2500 \times g$ for 10 min, the supernatant was discarded, and the pellet was resuspended in tagatose minimal media. These suspensions were spun again at $2500 \times g$ for 10 min, the supernatant was discarded, and the pellet was resuspended in tagatose minimal media one last time. This culture was then transferred to 96-well-plates (Costar, Corning™ 3788) containing tagatose minimal media (1:200 dilution) and grown for 40 hours in a plate reader (BioTek Synergy™ H1, Shake Mode: Double Orbital, Orbital Frequency: continuous shake 365 cpm, Interval: 10 min).

Next generation sequencing (NGS) of evolved tagatose phagemids

Phage lysates produced from cultures containing unevolved tagatose phagemid (WT) and phagemids after 2 cycles of IDE (Cy1 and Cy2) were used for DNA extraction with QIAamp DNA Kit (Qiagen Cat. No./ID: 56304). DNA from each lysate was then prepped using NEBnext (NEBNext® DNA Library Prep Master Mix Set for Illumina®) for sequencing with an Illumina iSeq 100. Alignment and detection of mutations were performed using bowtie2 (32) to align reads (with parameters –very-sensitive-local and -local), samtools (33) to generate mpileup files, and VarScan (34) to call mutations (with parameter –min-var-freq 0.001).

RESULTS AND DISCUSSION

Overview of IDE

The IDE workflow is both simple and flexible (Figure 1). Pathways of interest are assembled in a plasmid containing phage packaging recognition sequences (e.g. a phagemid (PM)) and transformed to a bacterium containing a helper phage. The master regulator for this phage is placed under inducible control. Next, mutagenesis is induced to create random mutations. Then, the phage lytic cycle is induced to initiate phagemid packaging and cell lysis. The resulting phage particles can then be applied to an unmutated strain to start a screening step or another mutagenesis step. We demonstrated IDE using the P1 phage. P1 is a temperate phage that primarily infects *E. coli*, has a 93 kb circular dsDNA genome, and exists as an extrachromosomal plasmid during lysogeny (35). In P1, lysis is repressed by the product of the *c1* gene. *coi* encodes a repressor of *c1*, and therefore overexpressing *coi* promotes P1 packaging and lysis (35). In prior work, placement of *coi* under the control of an arabinose-inducible promoter on a P1 phagemid (PM) enabled lysis in response to the addition of arabinose (28). In the uninduced state, *c1* is expressed and maintains P1 lysogeny. Inducible mutagenesis was achieved using a previously-described plasmid (MP6) (30).

Removing coi from P1 reduces variation in library size

We first focused on P1 phagemid packaging and infection levels, as these metrics define the number of library members that can be passaged between evolutionary rounds and affect the explorable sequence space. We hypothesized that removing the native, non-arabinose-inducible copy of *coi* from the P1 genome would result in an even more controlled lytic cycle, and therefore used a modified version of the P1 phage (P1kcΔcoi (ISA221)). We found that this modified phage enabled similar average levels of transduction as wild-type ($P > 0.05$, *t*-test) but with a reduced variance ($P < 0.04$, *F*-test) (Figure 2A).

Increasing media salt content increases library size

Prior reports of lysate production using P1 were performed by diluting stationary, P1-containing cultures 100-fold into phage lysate medium (PLM) for 1 hour (37°C, 250 rpm) followed by induction of lysis via addition of 13 mM arabinose
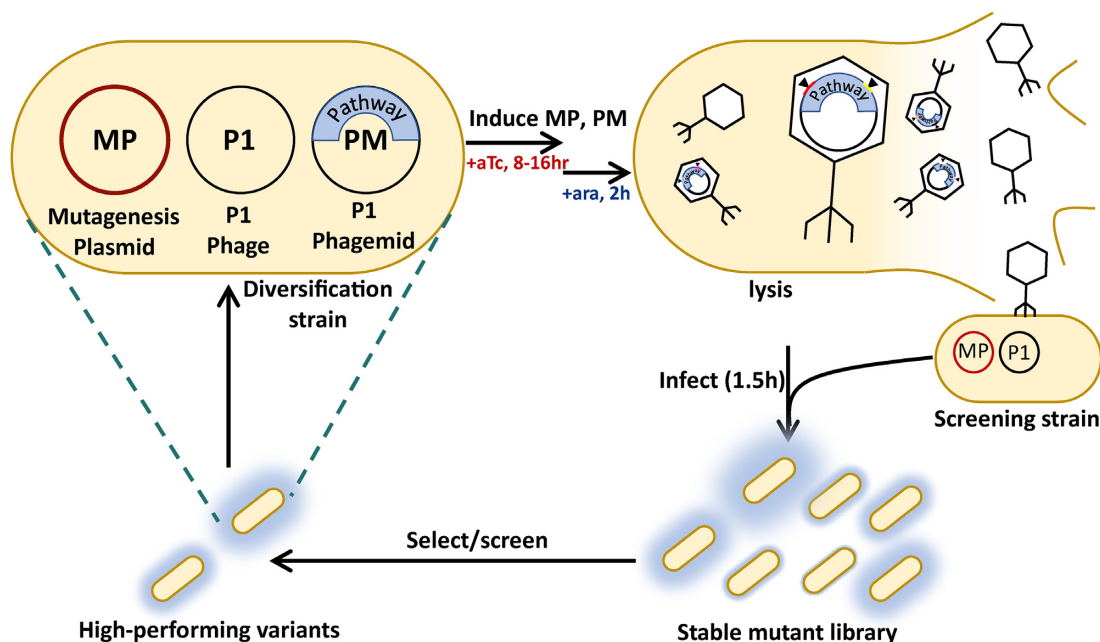
Figure 1. Conceptual overview of Inducible Directed Evolution (IDE). The pathway of interest is cloned into a P1 phagemid (PM) and transformed into a diversification strain containing the mutagenesis plasmid (MP) and P1 phage. MP is induced with anhydrotetracycline (+aTc) to create mutations, and then P1 lysis is induced with arabinose (+ara) to produce phage particles containing mutated P1 and PM. Phage lysate is then used to infect a screening strain to select/screen for the desired phenotype. After selection/screening, successful strains can enter another cycle of IDE.
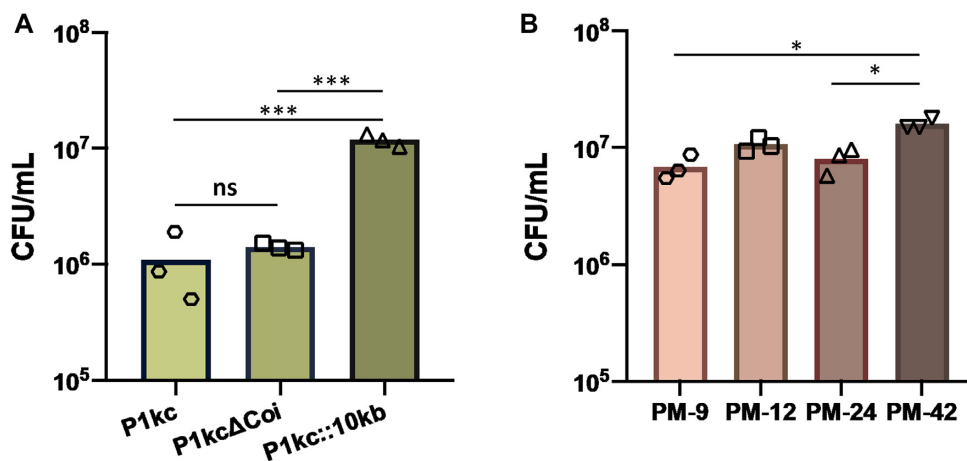
Figure 2. Optimizing P1 phagemid infection rate. (**A**) Engineered P1kc (P1kcΔcoi and P1kc::10kb) increases packaging/infection rates of phagemid compared to wild-type P1kc by >10-fold (ns $P > 0.05$, ***$P < 0.002$, Student's t test, Bonferroni correction). (**B**) The effect of insert size on phagemid transfer is negligible. PM-9 (9.7 kb), PM-12 (12.8 kb), PM-24 (24.0 kb) and PM-42 (42.0 kb) phagemids were packaged from the same strain (*E. coli* C600) containing P1kc::10kb and the same amount of phage lysate was used to infect wild-type *E. coli* C600 (*$P < 0.05$, Student's t test, Bonferroni correction).

for 1–4 hours (37°C, 250 rpm) (28,36). The produced phage lysate was used to infect a stationary-phase culture of *E. coli* KL739 at 37°C for 30 min without shaking (28,36). This lysate production and infection approach produced ∼10^5 CFU/1 mL lysate (28).

To improve transduction, we began by inducing phage production in a much larger *E. coli* C600 culture (OD$_{600}$ 1, 3×10^8 cells/mL) containing P1kcΔcoi and a 12 kb Phagemid (PM-12). The resulting lysate (1 mL) was applied to 3×10^8 *E. coli* C600 cells containing P1kcΔcoi and plated on media selective for PM. 6.1×10^5 PM-containing

cells were obtained, a 6.1-fold increase over previous methods. To further increase this value, we turned to the composition of the media in which cell growth and infection was performed. Ca^{2+} (CaCl$_2$) is required for P1 adsorption to lipopolysaccharide (LPS), and adding Mg^{2+} (MgCl$_2$) helps Gram-negative bacteria stabilize negatively charged lipopolysaccharides on the membrane (37–39). Therefore, 100 mM MgCl$_2$ and 5 mM CaCl$_2$ are commonly added to LB media (forming phage lysate medium (PLM)) in studies of P1 phage (28,36). We hypothesized that increasing the concentration of MgCl$_2$ and CaCl$_2$ would enhance P1 in-

fection rates. To test this hypothesis, we increased the concentrations of both salts by 20%, 40% and 60%. We found that only when the concentration of both salts in PLM is increased by 40%, the number of PM-containing cells increased a further 2.2-fold to 1.3×10^6. We called the new medium ePLM (Supplementary Figure S1).

We next varied the optical density to which the recipient cells were grown, while keeping the total number of recipient cells constant. We found that our initial strategy of growing cells to an optical density of 1.0 yielded the highest infection rates (Supplementary Figure S2).

Increasing P1 genome size promotes PM packaging

Next, we hypothesized that introducing an insert to the P1kc genome would decrease its packaging efficiency, like other phages (40), allowing more phagemids to be packaged. We found that inserting 10 kb of non-coding yeast DNA into P1kc (generating P1kc:10kb::*kanR* and P1kc::10kb) increased transferable library size by an additional 9.2-fold, to 1.2×10^7 (Figure 2A). As expected, this also increased the proportion of phagemids containing PM, relative to those containing P1 (Figure 2A). Similarly, we found that PM packaging was copy-number dependent, with lower PM copy numbers resulting in a reduced number of transduced cells (Supplementary Figure S3). Taken together, over the course of these optimization experiments we were able to increase P1 transduction rates by more than 123 fold over previous methods (28,41).

Although we expected that increasing the amount of lysate applied to naive cells would increase the number of transduced cells, we found that the ratio we had been using (lysate from 5×10^8 cells applied to 3×10^8 cells, defined as a ratio of 1:1) was past its saturation level, with reduced amounts of lysate providing similar values (Supplementary Figure S4). Therefore, we varied the number of naive cells, holding the lysate volume constant. As expected, we observed a linear relationship ($R^2 = 0.88$) between the number of naive cells and the number of infected cells (Supplementary Figure S5), indicating that IDE library sizes can be easily increased by scaling up lysate and cell amounts.

P1 enables packaging and transfer of pathways up to 36 kb in length

Using these improved conditions, we investigated the effect of PM size on the number of transduced cells. While we observed small fluctuations in library size with increasing cargo length, up to the largest phagemid we have tested (PM-42, 42 kb phagemid, 36 kb insert), the average library size passaged in each cycle remained relatively constant $\sim 10^7$ (Figure 2B). This indicates that IDE is capable of efficiently evolving large multi-gene pathways, hypothetically up to 87 kb as this would generate a phagemid the size of the P1 genome. In comparison, when these same phagemids were transformed via electroporation, the transformation efficiency decreased dramatically as the size of the phagemid increased, up to the size of the P1 genome (93 kb) (Supplementary Figure S6). In addition to the inefficiency of electroporating large plasmids, isolating and transforming large phagemids (for example, via a kit) is costly and difficult to

scale up compared to IDE (which only requires the addition of inducer), making IDE a desirable approach for directed evolution of large phagemids.

IDE enables tunable mutation rates

Having established efficient transfer of phagemids between cell populations, we next asked whether we could achieve mutations at rates sufficient for directed evolution. For this purpose, we modified a previously-reported plasmid enabling arabinose-inducible mutagenesis (MP6) (30). MP6 enhances cellular mutation rates by expressing a mutagenic operon (*danQ926, dam, seqA, emrR, ugi* and *cda1*) that collectively lowers the potency of DNA repair processes such as proofreading, translesion synthesis, mismatch repair and base excision and selection (30). To ensure compatibility with the arabinose-inducible PM, we switched this plasmid to an anhydrotetracycline (aTc)-inducible promoter and a kanamycin selection marker, forming aTc-MP. To test the mutation rate conferred by aTc-MP, we first used a previously described rifampicin resistance assay. In the presence of the inducer (anhydrotetracycline), the mutation rate of aTc-MP was similar to the mutation rate of the original MP6 (3.4×10^{-7} substitutions per base pair (bp) (29,30)). Additionally, omitting the inducer revealed that aTc-MP has a tighter off state, indicating that aTc-MP is suitable for inclusion in cells during selection or screening steps (Figure 3A). We then measured the rate of a specified mutation by inserting a chloramphenicol resistance gene (*cmR*) with one premature stop codon into PM (ISA308 and ISA311). We expected that inducing aTc-MP would randomly mutate *cmR* and yield variants with the stop codon reverted to a functional codon. We observed time-dependent increases in the number of *cmR*-resistant cells, supporting the notion that IDE enables tunable mutagenesis of defined DNA cargo (e.g. by inducing aTc-MP for differing lengths of time) (Figure 3B). Other mutagenesis inducers, such as UV light and chemical mutagens, may also be applicable in IDE.

Use of two different strains for mutagenesis and selection in IDE

Due to P1's ability to mediate generalized transduction between different *E. coli* strains, we hypothesized that the strain that is used for screening does not have to be the same as the strain used for library generation (Figure 4A). This would be beneficial if the ideal screening strain has a limited phage production capacity. To test this hypothesis, we chose three *E. coli* strains that are useful for different applications. *E. coli* C600 is a lab strain (42,43), *E. coli* MG1655 is often used in industrial fermentations (44–46), and *E. coli* Nissle 1917 is a probiotic strain (5,47,48). We found that phage lysate produced from 10^8 *E. coli* C600 cells could passage $>10^7$ variants to *E. coli* MG1655 and *E. coli* Nissle 1917 (Figure 4B). On the other hand, the same number of *E. coli* MG1655 and *E. coli* Nissle cells could only passage 4.9×10^6 and 6.9×10^4 variants back to C600 (Figure 4B), respectively. These results indicate that *E. coli* C600 is well-suited for production and packaging of large libraries, enabling these libraries to be screened in a more appropriate strain, for example incorporating biosensors or production-coupled growth circuits.
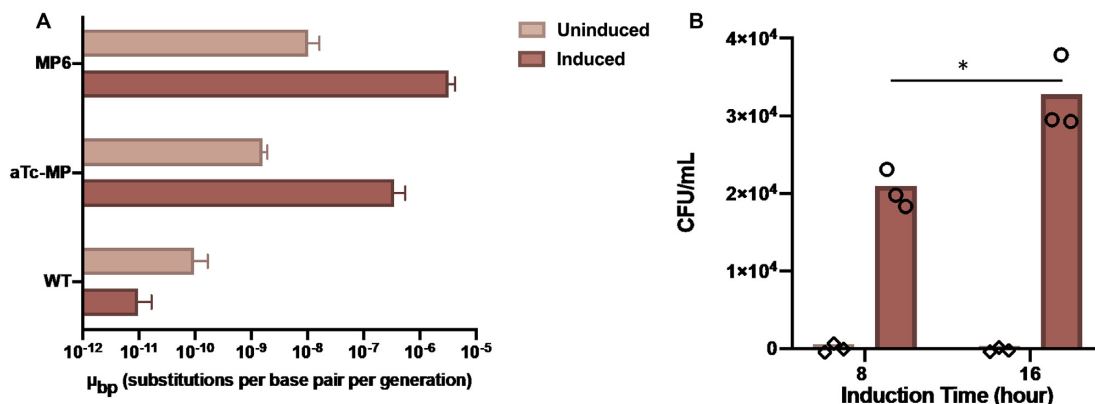
Figure 3. IDE allows high, tunable mutation rates. (**A**) Mutation rate of original MP6 induced with 20 mM arabinose, compared to aTc-MP induced with 200 ng/ml anhydrotetracycline and the base mutation rate of *E. coli* C600. The aTc-MP mutation rate is 6-fold lower than that of MP6 during induction, and 10-fold lower without induction. (**B**) Single stop codon reversion in *CmR*. A single stop codon was introduced to *CmR* (W16*) and reverted via IDE. Colonies from both induced and uninduced aTc-MP were counted after 8 and 16 hours to test mutation rate tunability with IDE (*$P < 0.02$, Student's *t* test). Three replicate cultures were used for both rifampicin assay and *CmR* stop codon reversion.
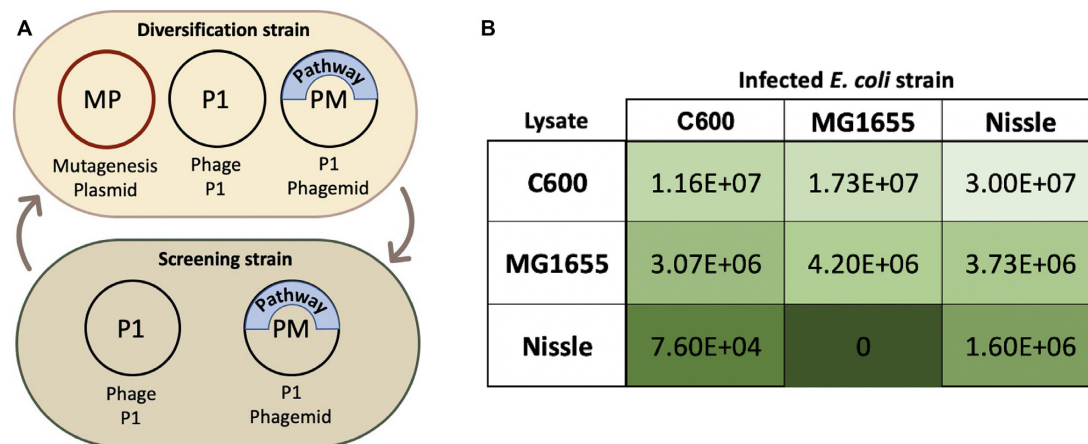
Figure 4. P1 phage's broad infectivity extends IDE screening possibilities to different *E. coli* strains. (**A**) Overview of using different *E. coli* strains in an IDE cycle for diversification and screening steps. (**B**) Heat map summarizing infection/packaging rates (CFU/mL) of phage lysate produced from different *E. coli* strains (C600 is a lab strain, MG1655 is an industrial strain, and Nissle 1917 is a probiotic strain) and used to infect the same 3 strains.

Phage co-infection rates are insufficient to impact IDE

After infection, some phages prevent subsequent infections by other phage through a process called superinfection exclusion. While superinfection exclusion is conferred by the *sim* gene in P1 (35,49), no such gene is present on PM. Therefore, we hypothesized that superinfection might occur during infection of fresh cells with mutant pathways, either between P1 and PM, or between two PM variants, thereby allowing recipient *E. coli* cells to express more than one pathway variant at once. To measure the rate of superinfection in IDE, we used two phagemids with different resistance genes and origins of replications (*cmR-pSC101* and *ampR-p15A*) and transformed them to *E. coli* C600 containing P1kc:10kb::*kanR*. Phage lysate from both strains was mixed (1:1 ratio) and used to infect WT *E. coli* C600. We found that the single variant infection rate was 330 times higher than the co-infection rate, supporting the notion that the vast majority of recipient cells carry a unique pathway variant (Figure 5).

Increasing GFP expression using IDE

To demonstrate IDE's capability to evolve a simple phenotype, we assembled *sfGFP* on a phagemid containing the *pSC101* origin. We chose the *pSC101* origin because of its stringent *Rep101*-dependent replication mechanism and its low copy number (50). In this setting, GFP fluorescence is controlled by at least 4 different genetic elements (the GFP coding sequence and its promoter (*PglpT*), as well as *Rep101* and its promoter). We wished to know which of these elements (or combination thereof) would lead to increased cellular fluorescence when mutated. We found that after two sequential rounds of mutagenesis and passage to fresh cells (Supplementary Figure S7a), we were able to verify via flow cytometry seven highly fluorescent isolates out of 20 colonies visually selected from ∼600 colonies on LB agar plates (1.2% 'hit' rate). Sequencing *Rep101* and *sfGFP* in these seven isolates yielded mutations exclusively in *Rep101*. To separate these mutations from unknown mutations potentially present in other parts of PM, we cloned
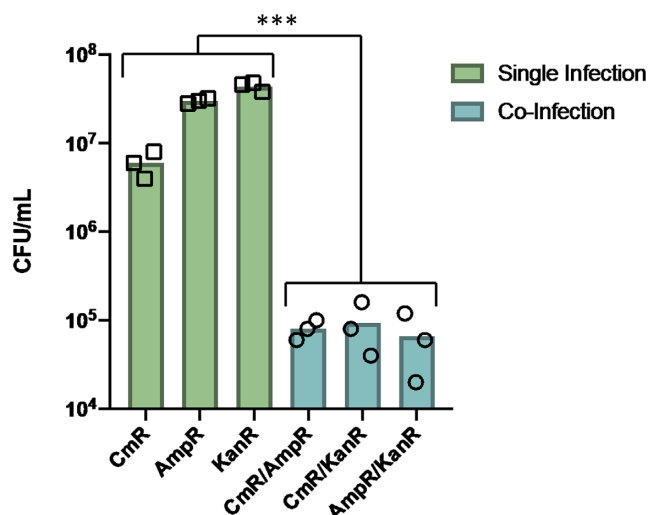
Figure 5. IDE allows for the selection of single variants. Co-infection of lysate produced from strains containing P1kc:10kb::kanR (kanR) with PM-AmpR (AmpR) or PM-CmR (CmR). Phage lysate produced from these two strains was used to infect WT *E. coli* C600 (1:1 ratio, 0.5 mL from each lysate and 1 mL resuspended cells in ePLM) and plated on LB plates containing Cm, Amp, Kan, Cm/Amp, Cm/Kan or Amp/Kan to select for cells infected with either P1kc:10kb::*kanR*, PM-Amp, PM-CmR or combinations of two constructs. The rates single infections are >330-fold higher than co-infection (***$P < 0.0001$, Student's *t* test). Phage lysate was produced from three biological replicates; each dot represents one biological replicate.

these *Rep101* variants into an unmutated PM-*sfGFP* vector and measured fluorescence via flow cytometry. All variants yielded significantly higher fluorescence than wild-type (Supplementary Figure S7b). Most of the *Rep101* mutations present in these clones (R46W, M78I, E93G, E93K, K102E and E115K) were previously found to increase the copy number of the *pSC101* origin (51), while one highly beneficial variant (I94N) was novel. It is therefore likely that the increase in GFP production in these isolates is due to an increased phagemid copy number. This result is reasonable because *sfGFP* has already been optimized for high stability and fluorescence in prior studies (52), and so increasing the copy number of the plasmid may be an easier path to achieve high GFP expression.

Improving a multigene carbon utilization pathway using IDE

We applied IDE to improve a heterologous tagatose consumption pathway in *E. coli* (Figure 6A). Tagatose is a hexose monosaccharide with a low glycemic index, making it attractive as an artificial sweetener (53). The cost of tagatose is decreasing due to advances in enzymatic synthesis (54,55), and therefore tagatose may find use as an ingredient in prebiotic functional foods. *E. coli* C600 consumes tagatose poorly, taking over 12.5 hours to reach an optical density (OD) of 0.25 in media containing tagatose as a sole carbon source from an initial OD_{600} of 0.05. We chose a five-gene tagatose consumption pathway (5.2 kb) from *Bacillus licheniformis* for insertion in PM (31). This pathway consists of *orf48* (encoding a predicted transcriptional regulator in the *murR/rpiR* family), *fruA2* and *orf51* (encoding a predicted phosphotransferase system that trans-

ports D-tagatose into the cell and converts it to tagatose 1-phosphate), *fruK2* (encoding a predicted kinase that converts tagatose 1-phosphate to tagatose 1,6-bisphosphate), and *gatY* (encoding a predicted aldolase that converts tagatose 1,6-bisphosphate to dihydroxyacetone phosphate and D-glyceraldehyde 3-phosphate). This pathway encodes different functions that collectively elicit the phenotype of interest and is therefore a good test case for improving complex phenotypes with IDE. Insertion of this pathway into PM yielded a C600 strain with a 29% reduction in lag time compared to wild type *E. coli* C600, taking 8.8 hours to achieve an OD_{600} of 0.25 in tagatose media (Figure 6C).

We expected that evolving the *B. licheniformis* tagatose consumption pathway would lead *E. coli* to consume tagatose more efficiently. After two IDE cycles comprising mutagenesis, growth-based selection, and transfer to fresh cells (Figure 6A), we plated ∼300 cells on tagatose minimal media agar plates. We picked 28 colonies based on increased size and then selected for variants that grew faster in liquid tagatose media than strains containing the unmutated pathway. We also performed a parallel selection comprising the same steps, except that mutagenesis was not performed. Phagemids present in cells passing both selections were transferred to fresh *E. coli* C600, and the growth of strains forming large colonies on tagatose minimal media agar plates was assayed in microtiter plates (Supplementary Figure S8). Because evolved strains seemed to exploit different strategies (for example, increased growth rate, reduced lag time, and/or final cell density) to pass selection, we computed the area under the growth curve (AUC) for each strain. Strains from the mutagenic selection exhibited significantly higher AUC in tagatose minimal media than strains from the non-mutagenic selection ($P < 0.0001$, Student's *t* test) (Figure 6B). Eight strains from the mutagenic selection exhibiting the best combinations of growth rate and final optical density were cloned into a wild type phagemid backbone (forming ISA179, Plasmid Map 2) to verify increased growth (Supplementary Figure S9a). Of these, four pathways conferred increased growth, relative to the wild-type sequence (Figure 6C and Supplementary Figure S9b).

All four tagatose pathway variants exhibited some combination of higher optical density (strain E3 exhibited a 2.6-fold higher cell density at 500 min than a strain containing the unmutated pathway) and reduced lag time (strain E3 exhibited a 64% reduction in time to reach an optical density of 0.25 than a strain containing the unmutated pathway). Mutations were identified in different genes, as shown in Supplementary Table S3. Isolates E1 and E2 have mutations across two sets of three different genes (*orf48, fruA2,* and *gatY* for E1, *fruK2, fruA2* and *gatY* for E2), while isolates E3 and E4 share one mutation in the ribosome binding site (RBS) of *fruK2*. This is the only mutation present in E4, while E3 contains two others (a silent mutation in *fruK2* and a coding mutation in *orf48*) that together further increase growth. Using the RBS calculator (56) to estimate the effect of the RBS mutation, we found that the predicted translation rate of the variant increased by 2-fold compared to the wild type sequence. This indicates that increasing *fruK2* expression could help increase the utilization of tagatose. Introducing the E3 mutations one by one revealed a potential selective path, with
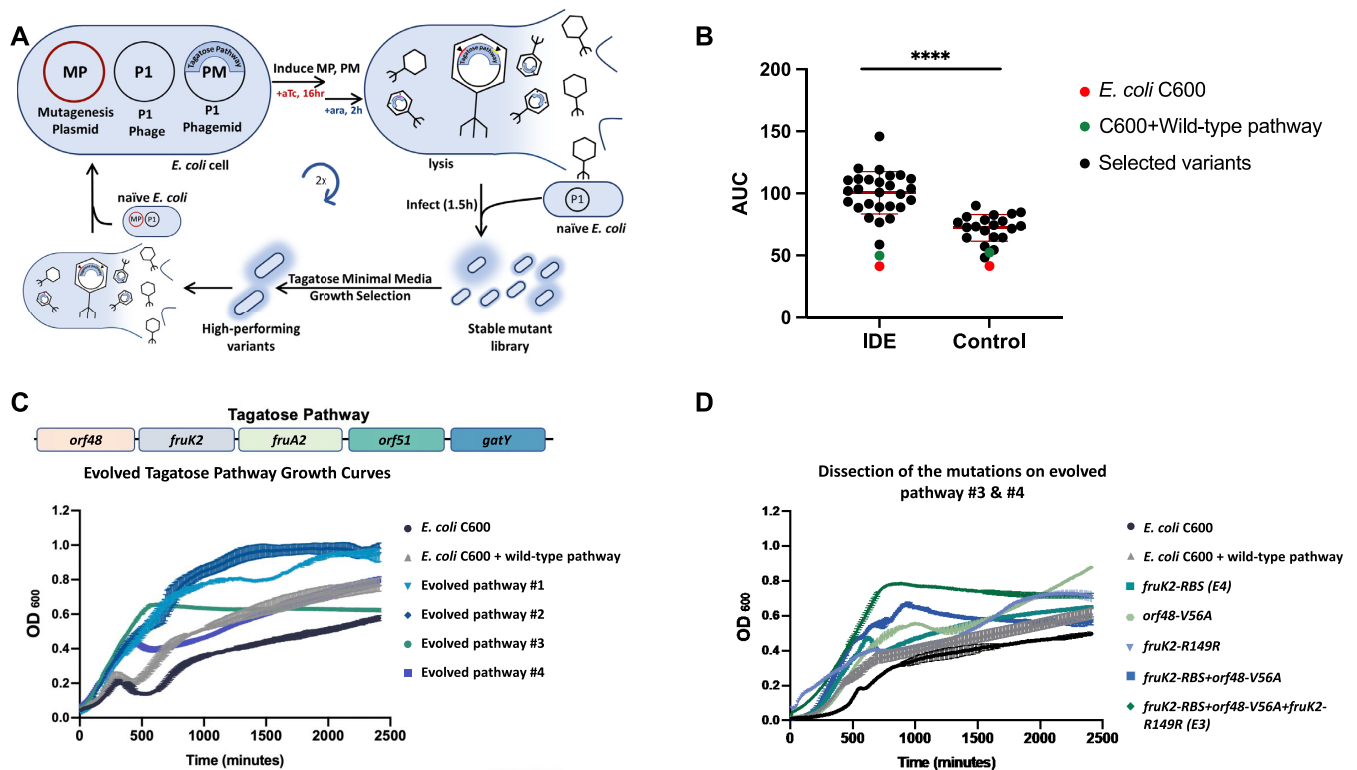
Figure 6. Directed evolution of a 5.2 kb tagatose pathway. (**A**) Overview of evolving a tagatose pathway via IDE. (**B**) Comparison of growth characteristics of variants isolated from IDE and control (mutagenesis uninduced) tagatose selections. Variants from IDE and control selections were grown in tagatose minimal media and optical density was measured over time in a microplate reader. AUC (Area Under the Curve) is calculated by summing OD_{600} values obtained over the course of the experiment. **** $P < 0.0001$, Student's t test. (**C**). Isolated tagatose pathway variants show improved growth on tagatose minimal media after the evolved pathway was re-cloned into the wild-type backbone of the phagemid. The growth curves of the four evolved isolates were compared to the growth curves of *E. coli* C600 and *E. coli* C600 + wild-type pathway. (**D**) Single and double mutations from evolved pathway #3 and #4 were introduced into wild-type pathway and grown in tagatose media. three biological replicates were used for all growth curves.

orf48-V56A+*fruK2*-RBS exhibiting improved growth over *fruK2*-RBS (E4), and *fruK2*-R149R+*orf48*-V56A+*fruK2*-RBS exhibiting improved growth over *orf48*-V56A+*fruK2*-RBS (Figure 6D). The accumulation of fitness-enhancing mutations across multiple genes agrees with prior studies pointing to the utility of a pathway-wide approach to directed evolution (57,58).

To probe the full complement of mutations occurring more deeply during evolution, we deep sequenced the lysate produced from tagatose control experiment (uninduced culture) and the lysate produced after cycle 1 (Cy1) and cycle 2 (Cy2) of IDE. Using VarScan (34), we identified the mutations shown in Supplementary Table S4. Mutations were identified in the tagatose pathway and phagemid packaging components. No mutations were identified in the origin of replication (*p15A*) or the resistance gene (*cmR*). Some of the mutations that were found in *coi*, Pac and AraC, while currently of unknown effect, may help phagemids to replicate more efficiently, as was observed during the development of PACE (20). These vector mutations could be incorporated into PM in future work to increase IDE throughput.

IDE of a 15.4 kb melezitose utilization pathway

We next asked whether IDE could operate on a pathway over 10 kb in length, as these long pathways are especially challenging for other directed evolution methods. To an-

swer this question, we turned to a 10-gene pathway (15.4 kb) taken from *Bifidobacterium breve* UC2003 that enables *E. coli* to consume the trisaccharide melezitose. Melezitose is a natural sugar that is also a prebiotic candidate (59,60). Wildtype *E. coli* C600 grows slowly in melezitose minimal media, taking more than 6 hours to reach OD_{600} 0.2 and 24 hours to reach OD 0.43. However, when the melezitose pathway from *B. breve* is introduced, *E. coli* C600 can grow to OD_{600} 0.3 in 6 hours and OD_{600} 0.97 in 24 hours (Figure 7). This pathway consists of two transcriptional regulators (*MelR1* and *MelR2*), two sugar binding proteins (*Bbr_1862* and *MelA*), two sugar permeases (*MelB* and *MelC*), two alpha-galactosidases (*MelD* and *Agl1*), an alpha-glucosidase (*MelE*), and a hypothetical protein (*Bbr_1861*) (Figure 7) (61,62). We hypothesized that further improvements to this pathway were possible, particularly to reduce the exceptionally long (~560 min) lag phase exhibited by strains expressing this pathway. We further reasoned that a pathway-wide evolutionary approach would be appropriate as it would be difficult to determine which genes were rate limiting (e.g. transport, sugar hydrolysis, or gene regulation) without time-consuming and costly experiments.

To evolve the melezitose pathway, we inserted it into the P1 phagemid and undertook three rounds of IDE, with each round consisting of a mutagenesis step, phage-mediated transfer to fresh cells, and three serial subcultures in melezi-
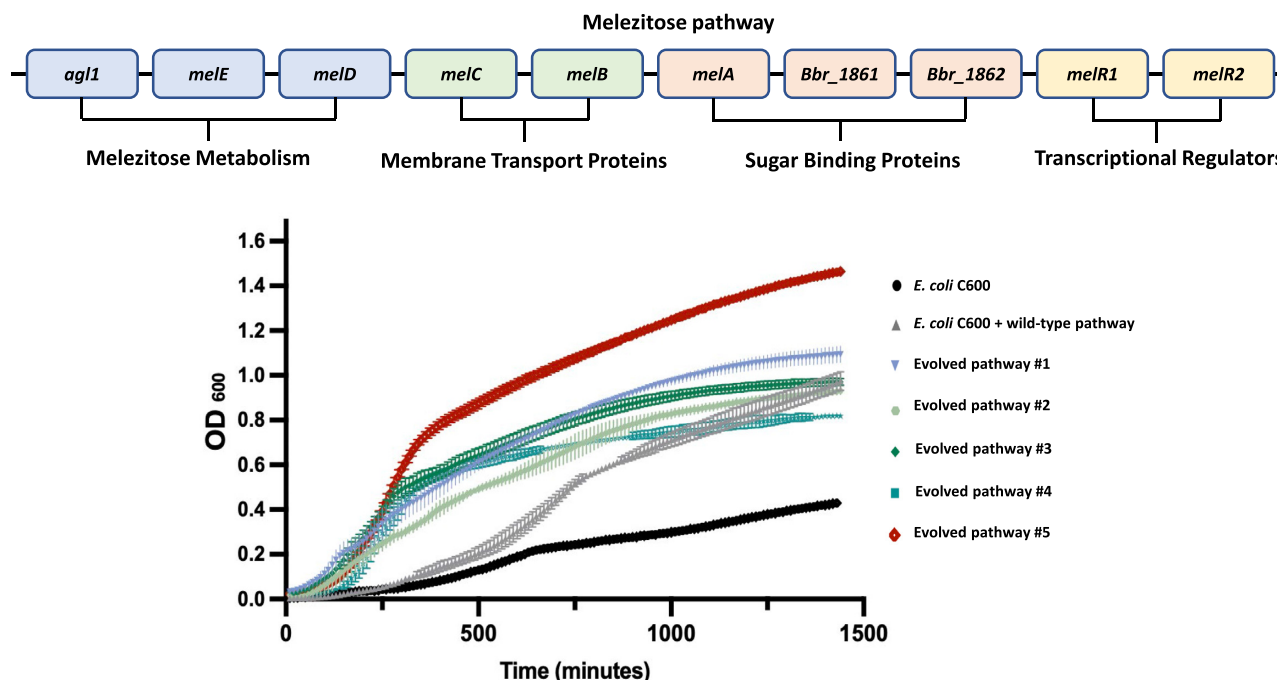
Figure 7. Directed evolution of a 15.4 kb melezitose consumption pathway. **Top:** Layout of a melezitose utilization operon from *B. breve* UCC2003. **Bottom:** Isolated melezitose pathway variants show improved growth on melezitose minimal media after the evolved phagemids were transformed to wildtype *E. coli* C600. The growth curves of the 5 evolved isolates were compared to the growth curves of *E. coli* C600 and *E. coli* C600 + wild-type pathway.

tose minimal liquid media. From the last culture, we spread cells on melezitose minimal media agar plates to identify the fastest-growing variants. Out of 60 isolates we picked, we chose to focus on the 5 that showed the greatest degree of improvement over the strain harboring the wildtype pathway (Supplementary Figure S10).

To determine whether the growth enhancements we observed were due to genomic background mutations, the phagemids from these five strains were transferred to wild-type *E. coli* C600. As we observed for the tagatose pathway, all 5 re-transformed strains exhibited some combination of higher optical density (strain E5 exhibited a 7.4-fold higher cell density at 280 min than a strain containing the unmutated pathway) and reduced lag time (strain E5 exhibited a 63% reduction in time to reach an optical density of 0.25 than a strain containing the unmutated pathway) (Figure 7). This result was reasonable, because although the wild-type reached a similar optical density as many of the evolved variants after 24 h, performance at early times determines eventual dominance in a mixed culture (63). Collectively, this result indicated that phagemid mutations alone led to the improved growth characteristics of the strains we isolated.

We next wished to identify the mutations present in the five highest-performing phagemids and determine how these phagemids were related. Whole-plasmid sequencing revealed that out of the 10 pathway genes, only 6 were mutated across these five phagemids: *Agl1, MelE, MelD, MelC, Bbr_1862* and *MelR1* (Supplementary Table S5). Further, while isolates 1, 2, and 3 each carry a different set of mutations, all three contained coding changes in *MelD* and *MelR1*. Isolates 2 and 3 share the same coding change in *MelE* (P172T) and isolates 1 and 3 contain different cod-

ing changes in *Bbr_1862*. On the other hand, isolate 5 appears to have evolved from isolate 4, as isolate 5 contains additional mutations (in *Agl1, MelE,* and *MelD*) beyond those present in isolate 4 (in *MelE, MelD, MelC, Bbr_1862* and *MelR1*) (Supplementary Table S5). Isolate 5 also has the highest number of pathway nucleotide mutations (10) out of all the isolates and is the only isolate with a mutation in *Agl1*. It would be interesting to measure the contribution of each mutation to improved growth, potentially using barcoded combinatorial libraries (64), to determine whether mutations to multiple genes can work together, as we observed for the evolved tagatose pathways. Taken together, this experiment indicated that IDE can be applied to long pathways that are challenging to evolve using other approaches, and that whole-pathway evolution can identify genes that act as key bottlenecks of pathway performance.

CONCLUSION

As microbial engineering moves toward applications demanding ever-higher performance (for example, green production of fuels (65) and chemicals (66), sensing (67) and biosynthesis on host-associated sites (68)), the ability to engineer complex phenotypes is becoming increasingly important. Currently, optimizing the performance of multi-gene pathways is a challenging task. IDE offers the ability to perform directed evolution on long (at least up to 36 kb) sequences of DNA with tunable error rates (up to 3.4×10^{-7} substitutions per bp per generation) and library sizes that scale trivially with culture volume. Importantly, biosynthesis pathways for many natural products are quite long and contain many genes. Therefore, we suspect that IDE may accelerate the engineering of such pathways. We expect that

the use of different mutagenesis methods (for example, ultraviolet light, chemical mutagens, and CRISPR-based systems(23–25)) in place of a mutagenesis plasmid can add further mutational flexibility and potentially reduce mutations in the vector backbone. Importantly, the use of temperate phages (such as P1) to passage variants to fresh hosts greatly reduces the impact of off-target mutations and decouples mutagenesis and screening steps, providing a large degree of flexibility when designing selections. In particular, we expect this approach to be highly amenable to automation, enabling rapid and highly parallel evolution campaigns similar to the eVOLVR (10) and PACE (20) systems. The data presented here indicate that IDE is a valuable addition to the existing continuous evolution toolkit, and demonstrates that temperate phages are promising vehicles for directed evolution of complex phenotypes.

DATA AVAILABILITY

Sequencing reads have been deposited in SRA under Bio-Project ID PRJNA801315.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Harvey,A.L., Edrada-Ebel,R. and Quinn,R.J. (2015) The re-emergence of natural products for drug discovery in the genomics era. *Nat. Rev. Drug Discov.*, **14**, 111–129.
2. Crook,N. and Alper,H.S. (2012) Classical strain improvement. In: Patnaik,R. (ed). *Engineering Complex Phenotypes in Industrial Strains*.
3. Lee,S.Y., Kim,H.U., Chae,T.U., Cho,J.S., Kim,J.W., Shin,J.H., Kim,D.I., Ko,Y.-S., Jang,W.D. and Jang,Y.-S. (2019) A comprehensive metabolic map for production of bio-based chemicals. *Nat. Catal.*, **2**, 18–33.
4. Pham,H.L., Wong,A., Chua,N., Teo,W.S., Yew,W.S. and Chang,M.W. (2017) Engineering a riboswitch-based genetic platform for the self-directed evolution of acid-tolerant phenotypes. *Nat. Commun.*, **8**, 411.
5. Crook,N., Ferreiro,A., Gasparrini,A.J., Pesesky,M.W., Gibson,M.K., Wang,B., Sun,X., Condiotte,Z., Dobrowolski,S., Peterson,D. *et al.* (2019) Adaptive strategies of the candidate probiotic *E. coli* nissle in the mammalian gut. *Cell Host Microbe*, **25**, 499–512.
6. Farkas,Z., Kalapis,D., Bódi,Z., Szamecz,B., Daraba,A., Almási,K., Kovács,K., Boross,G., Pál,F., Horváth,P. *et al.* (2018) Hsp70-associated chaperones have a critical role in buffering protein production costs. *Elife*, **7**, e29845.
7. Tsoi,R., Wu,F., Zhang,C., Bewick,S., Karig,D. and You,L. (2018) Metabolic division of labor in microbial systems. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 2526–2531.
8. Rosano,G.L. and Ceccarelli,E.A. (2014) Recombinant protein expression in Escherichia coli: advances and challenges. *Front. Microbiol.*, **5**, 172.
9. Gleizer,S., Ben-Nissan,R., Bar-On,Y.M., Antonovsky,N., Noor,E., Zohar,Y., Jona,G., Krieger,E., Shamshoum,M., Bar-Even,A. *et al.* (2019) Conversion of *Escherichia coli* to generate all biomass carbon from CO_2. *Cell*, **179**, 1255–1263.
10. Wong,B.G., Mancuso,C.P., Kiriakov,S., Bashor,C.J. and Khalil,A.S. (2018) Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER. *Nat. Biotechnol.*, **36**, 614–623.
11. Wang,H.H., Isaacs,F.J., Carr,P.A., Sun,Z.Z., Xu,G., Forest,C.R. and Church,G.M. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, **460**, 894–898.
12. Georg,J. and Hess,W.R. (2011) cis-antisense RNA, another level of gene regulation in bacteria. *Microbiol. Mol. Biol. Rev.*, **75**, 286–300.
13. Meng,J., Kanzaki,G., Meas,D., Lam,C.K., Crummer,H., Tain,J. and Xu,H.H. (2012) A genome-wide inducible phenotypic screen identifies antisense RNA constructs silencing Escherichia coli essential genes. *FEMS Microbiol. Lett.*, **329**, 45–53.
14. Bikard,D., Hatoum-Aslan,A., Mucida,D. and Marraffini,L.A. (2012) CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe*, **12**, 177–186.
15. Reis,A.C., Halper,S.M., Vezeau,G.E., Cetnar,D.P., Hossain,A., Clauer,P.R. and Salis,H.M. (2019) Simultaneous repression of multiple bacterial genes using nonrepetitive extra-long sgRNA arrays. *Nat. Biotechnol.*, **37**, 1294–1301.
16. Dong,C., Fontana,J., Patel,A., Carothers,J.M. and Zalatan,J.G. (2018) Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. *Nat. Commun.*, **9**, 2489.
17. Muteeb,G. and Sen,R. (2010) Random mutagenesis using a mutator strain. *Methods Mol. Biol.*, **634**, 411–419.
18. Roth,T.B., Woolston,B.M., Stephanopoulos,G. and Liu,D.R. (2019) Phage-assisted evolution of *Bacillus methanolicus* methanol dehydrogenase 2. *ACS Synth. Biol.*, **8**, 796–806.
19. Suzuki,T., Miller,C., Guo,L.-T., Ho,J.M.L., Bryson,D.I., Wang,Y.-S., Liu,D.R. and Söll,D. (2017) Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nat. Chem. Biol.*, **13**, 1261–1266.
20. Esvelt,K.M., Carlson,J.C. and Liu,D.R. (2011) A system for the continuous directed evolution of biomolecules. *Nature*, **472**, 499–503.
21. Brödel,A.K., Rodrigues,R., Jaramillo,A. and Isalan,M. (2020) Accelerated evolution of a minimal 63-amino acid dual transcription factor. *Sci. Adv.*, **6**, eaba2728.
22. Marintcheva,B. (2018) Viral tools for in vitro manipulations of nucleic acids. In: *Harnessing the Power of Viruses*. Elsevier, pp. 27–67.
23. Halperin,S.O., Tou,C.J., Wong,E.B., Modavi,C., Schaffer,D.V. and Dueber,J.E. (2018) CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature*, **560**, 248–252.
24. Park,H. and Kim,S. (2021) Gene-specific mutagenesis enables rapid continuous evolution of enzymes in vivo. *Nucleic Acids Res.*, **49**, e32.
25. Álvarez,B., Mencía,M., de Lorenzo,V. and Fernández,L.Á. (2020) In vivo diversification of target genomic sites using processive base deaminase fusions blocked by dCas9. *Nat. Commun.*, **11**, 6436.
26. Coren,J.S., Pierce,J.C. and Sternberg,N. (1995) Headful packaging revisited: the packaging of more than one DNA molecule into a bacteriophage P1 head. *J. Mol. Biol.*, **249**, 176–184.

27. Baba,T., Ara,T., Hasegawa,M., Takai,Y., Okumura,Y., Baba,M., Datsenko,K.A., Tomita,M., Wanner,B.L. and Mori,H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.*, **2**, 2006.0008.

28. Kittleson,J.T., DeLoache,W., Cheng,H.-Y. and Anderson,J.C. (2012) Scalable plasmid transfer using engineered P1-based phagemids. *ACS Synth. Biol.*, **1**, 583–589.

29. Ho,J.M.L., Miller,C.A., Parks,S.E., Mattia,J.R. and Bennett,M.R. (2020) A suppressor tRNA-mediated feedforward loop eliminates leaky gene expression in bacteria. *Nucleic Acids Res.*, **49**, e25.

30. Badran,A.H. and Liu,D.R. (2015) Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nat. Commun.*, **6**, 8425.

31. Van der Heiden,E., Delmarcelle,M., Lebrun,S., Freichels,R., Brans,A., Vastenavond,C.M., Galleni,M. and Joris,B. (2013) A pathway closely related to the (D)-tagatose pathway of gram-negative enterobacteria identified in the gram-positive bacterium Bacillus licheniformis. *Appl. Environ. Microbiol.*, **79**, 3511–3515.

32. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.

33. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

34. Koboldt,D.C., Zhang,Q., Larson,D.E., Shen,D., McLellan,M.D., Lin,L., Miller,C.A., Mardis,E.R., Ding,L. and Wilson,R.K. (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.*, **22**, 568–576.

35. Łobocka,M.B., Rose,D.J., Plunkett,G., Rusin,M., Samojedny,A., Lehnherr,H., Yarmolinsky,M.B. and Blattner,F.R. (2004) Genome of bacteriophage P1. *J. Bacteriol.*, **186**, 7032–7068.

36. Westwater,C., Schofield,D.A., Schmidt,M.G., Norris,J.S. and Dolan,J.W. (2002) Development of a P1 phagemid system for the delivery of DNA into Gram-negative bacteria. *Microbiology*, **148**, 943–950.

37. Coughlin,R.T., Tonsager,S. and McGroarty,E.J. (1983) Quantitation of metal cations bound to membranes and extracted lipopolysaccharide of *Escherichia coli*. *Biochemistry*, **22**, 2002–2007.

38. Cvirkaite-Krupovic,V., Krupovic,M., Daugelavicius,R. and Bamford,D.H. (2010) Calcium ion-dependent entry of the membrane-containing bacteriophage PM2 into its Pseudoalteromonas host. *Virology*, **405**, 120–128.

39. Tomás,J.M. and Kay,W.W. (1984) Effect of bacteriophage P1 lysogeny on lipopolysaccharide composition and the lambda receptor of Escherichia coli. *J. Bacteriol.*, **159**, 1047–1052.

40. Tridgett,M., Ababi,M., Osgerby,A., Ramirez Garcia,R. and Jaramillo,A. (2021) Engineering bacteria to produce pure phage-like particles for gene delivery. *ACS Synth. Biol.*, **10**, 107–114.

41. Kasman,L.M., Kasman,A., Westwater,C., Dolan,J., Schmidt,M.G. and Norris,J.S. (2002) Overcoming the phage replication threshold: a mathematical model with implications for phage therapy. *J. Virol.*, **76**, 5557–5564.

42. Allué-Guardia,A., Nyong,E.C., Koenig,S.S.K., Vargas,S.M., Bono,J.L. and Eppinger,M. (2019) Closed genome sequence of *Escherichia coli* K-12 group strain C600. *Microbiol. Resour. Announc.*, **8**, e01052-18.

43. Lennox,E.S. (1955) Transduction of linked genetic characters of the host by bacteriophage P1. *Virology*, **1**, 190–206.

44. Rugbjerg,P., Dyerberg,A.S.B., Quainoo,S., Munck,C. and Sommer,M.O.A. (2020) Short and long-read ultra-deep sequencing profiles emerging heterogeneity across five platform Escherichia coli strains. *Metab. Eng.*, **65**, 197–206.

45. Wang,J., Huang,C., Guo,K., Ma,L., Meng,X., Wang,N. and Huo,Y.-X. (2020) Converting Escherichia coli MG1655 into a chemical overproducer through inactivating defense system against exogenous DNA. *Synth. Syst. Biotechnol.*, **5**, 333–342.

46. Rosales-Colunga,L.M. and Martínez-Antonio,A. (2014) Engineering *Escherichia coli* K12 MG1655 to use starch. *Microb. Cell Fact.*, **13**, 74.

47. Crook,N., Ferreiro,A., Condiotte,Z. and Dantas,G. (2020) Transcript barcoding illuminates the expression level of synthetic constructs in *E. coli* nissle residing in the mammalian gut. *ACS Synth. Biol.*, **9**, 1010–1021.

48. Hwang,I.Y., Koh,E., Wong,A., March,J.C., Bentley,W.E., Lee,Y.S. and Chang,M.W. (2017) Engineered probiotic *Escherichia coli* can eliminate and prevent *Pseudomonas aeruginosa* gut infection in animal models. *Nat. Commun.*, **8**, 15028.

49. Kliem,M. and Dreiseikelmann,B. (1989) The superimmunity gene sim of bacteriophage P1 causes superinfection exclusion. *Virology*, **171**, 350–355.

50. Furuno,S., Watanabe-Murakami,Y., Takebe-Suzuki,N. and Yamaguchi,K. (2000) Negative control of plasmid pSC101 replication by increased concentrations of both initiator protein and iterons. *J. Gen. Appl. Microbiol.*, **46**, 29–37.

51. Thompson,M.G., Sedaghatian,N., Barajas,J.F., Wehrs,M., Bailey,C.B., Kaplan,N., Hillson,N.J., Mukhopadhyay,A. and Keasling,J.D. (2018) Isolation and characterization of novel mutations in the pSC101 origin that increase copy number. *Sci. Rep.*, **8**, 1590.

52. Pédelacq,J.-D., Cabantous,S., Tran,T., Terwilliger,T.C. and Waldo,G.S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.*, **24**, 79–88.

53. Oh,D.-K. (2007) Tagatose: properties, applications, and biotechnological processes. *Appl. Microbiol. Biotechnol.*, **76**, 1–8.

54. Bober,J.R. and Nair,N.U. (2019) Galactose to tagatose isomerization at moderate temperatures with high conversion and productivity. *Nat. Commun.*, **10**, 4548.

55. Xu,Z., Qing,Y., Li,S., Feng,X., Xu,H. and Ouyang,P. (2011) A novel l-arabinose isomerase from *Lactobacillus fermentum* CGMCC2921 for d-tagatose production: gene cloning, purification and characterization. *J. Mol. Catal. B Enzym.*, **70**, 1–7.

56. Salis,H.M., Mirsky,E.A. and Voigt,C.A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.*, **27**, 946–950.

57. Crook,N., Abatemarco,J., Sun,J., Wagner,J.M., Schmitz,A. and Alper,H.S. (2016) In vivo continuous evolution of genes and pathways in yeast. *Nat. Commun.*, **7**, 13051.

58. Johnston,C.W., Badran,A.H. and Collins,J.J. (2020) Continuous bioactivity-dependent evolution of an antibiotic biosynthetic pathway. *Nat. Commun.*, **11**, 4202.

59. Swears,R.M. and Manley-Harris,M. (2021) Composition and potential as a prebiotic functional food of a Giant Willow Aphid (*Tuberolachnus salignus*) honeydew honey produced in New Zealand. *Food Chem.*, **345**, 128662.

60. Behera,P. and Balaji,S. (2021) The forgotten sugar: a review on multifarious applications of melezitose. *Carbohydr. Res.*, **500**, 108248.

61. O'Connell,K.J., O'Connell Motherway,M.O., Liedtke,A., Fitzgerald,G.F., Ross,R.P., Stanton,C., Zomer,A. and van Sinderen,D. (2014) Transcription of two adjacent carbohydrate utilization gene clusters in Bifidobacterium breve UCC2003 is controlled by LacI- and repressor open reading frame kinase (ROK)-type regulators. *Appl. Environ. Microbiol.*, **80**, 3604–3614.

62. O'Connell,K.J., O'Connell Motherway,M., O'Callaghan,J., Fitzgerald,G.F., Ross,R.P., Ventura,M., Stanton,C. and van Sinderen,D. (2013) Metabolism of four α-glycosidic linkage-containing oligosaccharides by *Bifidobacterium breve* UCC2003. *Appl. Environ. Microbiol.*, **79**, 6280–6292.

63. Ram,Y., Dellus-Gur,E., Bibi,M., Karkare,K., Obolski,U., Feldman,M.W., Cooper,T.F., Berman,J. and Hadany,L. (2019) Predicting microbial growth in a mixed culture from growth curve data. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 14698–14707.

64. Zurek,P.J., Knyphausen,P., Neufeld,K., Pushpanath,A. and Hollfelder,F. (2020) UMI-linked consensus sequencing enables phylogenetic analysis of directed evolution. *Nat. Commun.*, **11**, 6023.

65. Gronenberg,L.S., Marcheschi,R.J. and Liao,J.C. (2013) Next generation biofuel engineering in prokaryotes. *Curr. Opin. Chem. Biol.*, **17**, 462–471.

66. Beller,H.R., Lee,T.S. and Katz,L. (2015) Natural products as biofuels and bio-based chemicals: fatty acids and isoprenoids. *Nat. Prod. Rep.*, **32**, 1508–1526.

67. Daeffler,K.N.-M., Galley,J.D., Sheth,R.U., Ortiz-Velez,L.C., Bibb,C.O., Shroyer,N.F., Britton,R.A. and Tabor,J.J. (2017) Engineering bacterial thiosulfate and tetrathionate sensors for detecting gut inflammation. *Mol. Syst. Biol.*, **13**, 923.

68. Fischbach,M.A. and Segre,J.A. (2016) Signaling in host-associated microbial communities. *Cell*, **164**, 1288–1300.