AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology

Identification of Genes Conferring Tolerance to Lignocellulose-Derived Inhibitors by Functional Selections in Soil Metagenomes

Kevin J. Forsberg,[a]* Sanket Patel,[a,b] Evan Witt,[a] Bin Wang,[a,b] Tyler D. Ellison,[a] Gautam Dantas[a,b,c,d]

Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, Missouri, USA[a]; Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri, USA[b]; Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, Missouri, USA[c]; Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, Missouri, USA[d]

The production of fuels or chemicals from lignocellulose currently requires thermochemical pretreatment to release fermentable sugars. These harsh conditions also generate numerous small-molecule inhibitors of microbial growth and fermentation, limiting production. We applied small-insert functional metagenomic selections to discover genes that confer microbial tolerance to these inhibitors, identifying both individual genes and general biological processes associated with tolerance to multiple inhibitory compounds. Having screened over 248 Gb of DNA cloned from 16 diverse soil metagenomes, we describe gain-of-function tolerance against acid, alcohol, and aldehyde inhibitors derived from hemicellulose and lignin, demonstrating that uncultured soil microbial communities hold tremendous genetic potential to address the toxicity of pretreated lignocellulose. We recovered genes previously known to confer tolerance to lignocellulosic inhibitors as well as novel genes that confer tolerance via unknown functions. For instance, we implicated galactose metabolism in overcoming the toxicity of lignin monomers and identified a decarboxylase that confers tolerance to ferulic acid; this enzyme has been shown to catalyze the production of 4-vinyl guaiacol, a valuable precursor to vanillin production. These metagenomic tolerance genes can enable the flexible design of hardy microbial catalysts, customized to withstand inhibitors abundant in specific bioprocessing applications.

Many lignocellulosic feedstocks (e.g., switchgrass) are preferred to maize, sugarcane, and other traditional food crops for the production of fuels and chemicals because they are able to grow on marginal land, often require little attention or energy input, and do not compete directly with the food supply (1–6). However, lignocellulose requires harsh thermochemical pretreatment methods to liberate fermentable monosaccharides (7–9), producing an additional compendium of compounds inhibitory to microbial growth that ultimately reduce production efficiencies (10–12). These small-molecule inhibitors derived from lignocellulose pretreatment (here called lignocellulosic inhibitors) are typically aldehydes, organic acids, furans, or phenolics and can originate from the cellulosic, hemicellulosic, and lignified fractions of the feedstock (11–15).

Engineering hardier microbial production hosts with elevated tolerance to these inhibitors offers potential to ameliorate the toxic effects of these compounds without incurring the high process costs associated with detoxifying the lignocellulosic hydrolysate (16, 17). Unfortunately, the modes of toxicity of many of these toxins are poorly described, and genes conferring tolerance to many of these compounds have not been identified (18, 19). An expanded catalog of tolerance-conferring genotypes may shed light on mechanisms of toxicity and enable synthetic biology approaches for the design of diverse production hosts with broad-spectrum tolerance.

Soil microorganisms, including white-rot fungi (20) and many bacteria (21), are likely a valuable reservoir of genetic elements that confer tolerance to lignocellulosic inhibitors and next-generation biofuels (22–25). Indeed, our previous work using large-insert (40- to 50-kb) functional metagenomic libraries identified three such genes conferring improved tolerance to two biomass inhibitors (19). Characterization of these large-insert libraries required numerous rounds of sequencing to assemble tolerance-conferring DNA fragments, followed by transposon mutagenesis and targeted subcloning to identify the three causal tolerance genes on the two fragments interrogated. These time- and cost-intensive methods are not feasible strategies for discovering genes conferring tolerance to many inhibitors, and accordingly, we applied a higher-throughput approach.

We functionally interrogated 16 agricultural and grassland soils for their repertoire of lignocellulosic inhibitor tolerance-conferring genotypes (26–28). Since our earlier work with large-insert libraries demonstrated that tolerance phenotypes can be conferred by individual open reading frames (ORFs) (19), we elected to construct small-insert (1- to 5-kb) metagenomic libraries, which allow screening with greater clone diversity and sampling depth compared to screening with large-insert libraries (29, 30). To facilitate the interrogation of large, diverse clone libraries with direct application to bioprocessing goals, we performed functional selections in *Escherichia coli*, as it exhibits high transformation efficiencies and is a leading producer of advanced biofuels and commodity chemicals (31–34).

Received 31 August 2015 **Accepted** 1 November 2015

Accepted manuscript posted online 6 November 2015

Citation Forsberg KJ, Patel S, Witt E, Wang B, Ellison TD, Dantas G. 2016. Identification of genes conferring tolerance to lignocellulose-derived inhibitors by functional selections in soil metagenomes. Appl Environ Microbiol 82:528–537. doi:10.1128/AEM.02838-15.

Editor: R. M. Kelly

Address correspondence to Gautam Dantas, dantas@wustl.edu.

* Present address: Kevin J. Forsberg, Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.02838-15.

MATERIALS AND METHODS

Construction of soil metagenomic libraries. Construction of soil metagenomic libraries was performed exactly as described in previous work (28). Briefly, metagenomic DNA was extracted from soils (see Table S1 in the supplemental material) using the PowerMax soil DNA isolation kit (MoBio Laboratories), and 10 g was used for the construction of a given library. Small-insert metagenomic libraries were created by shearing this DNA into 1- to 5-kb fragments before ligation into the pZE21 expression vector (35) and electroporation into *E. coli* MegaX cells (Invitrogen). These cells do not contain the TetR repressor, and therefore, expression from the PLtetO-1 promoter of pZE21 is constitutive. Titers of libraries were determined by plating out 0.1 μl and 0.01 μl of recovered cells onto Luria-Bertani (LB) agar (5 g yeast extract, 5 g NaCl, 10 g tryptone, 12 g agar in 1 liter of water) plates containing 50 μg/ml kanamycin (Kan). For each library, insert size distribution was estimated by gel electrophoresis of PCR products obtained by amplifying the insert from 12 randomly picked clones using primers flanking the HincII site of the multiple cloning site of the pZE21MCS1 vector (which contains a selectable marker for kanamycin resistance). The average insert size across all libraries was determined to be 2,000 bp, and library size estimates were calculated by multiplying the average PCR-based insert size by the number of CFU from titer determination after transformation recovery. The rest of the recovered cells were inoculated into 50 ml of LB medium containing 50 μg/ml kanamycin and grown overnight. The overnight culture was frozen with 15% glycerol and stored at −80°C for subsequent selection.

Determining MICs of lignocellulosic inhibitors. MICs for solid and liquid toxicity assays were determined separately, each in LB medium using *E. coli* MegaX cells (DH10B-derived) containing an empty pZE21MCS vector. Initial MIC predictions were informed by previous work (13–15, 19) and tested using LB medium formulated with 50 μg/ml kanamycin and a range of concentrations for each inhibitor spanning the predicted MIC (typically, four inhibitor concentrations were chosen for agar MICs and seven for liquid MICs). For agar MIC determination, overnight cultures were plated using glass sterile beads and growth was monitored at 37°C for 2 days. For liquid MICs, a 2.5% inoculum of mid-log culture was added to 200 μl of broth medium and growth was monitored in 96-well plates using the Synergy H1 microplate reader (BioTek Instruments) for up to 5 days at 37°C. If an empirically determined MIC fell between two tested concentrations, a subsequent round of MIC determination was performed to narrow the step size between putative MICs. If the MIC fell outside the initial tested range, the initial assay was repeated to set boundaries for a MIC estimate and a third assay was performed to narrow the MIC further. Because the range of inhibitor concentrations chosen for the first round of MIC testing varied across compounds, so, too, did the resolution for the final rounds of MIC determination.

Functional selections for tolerance to lignocellulosic inhibitors. For each soil metagenomic library, selections for tolerance to each of the 20 inhibitors tested were performed on LB agar containing 50 μg/ml kanamycin and supplemented with concentrations of lignocellulosic inhibitor outlined in Table S2 in the supplemental material. For each metagenomic library, the number of cells plated on each type of selective medium represented 10 times the number of unique CFU in the library, as determined from titers during library creation. After plating (using sterile glass beads), selections for compound tolerance were incubated at 37°C for up to 5 days to allow the growth of tolerant *E. coli* transformants. If microbial growth was observed, cells were collected in two ways to interrogate tolerance-conferring genotypes: (i) by picking and validating individual transformants, as described below, or (ii) by collecting all microbial growth. For the latter method, after individual transformants were picked, all remaining transformants from a single plate (soil by inhibitor selection) were collected by adding 750 μl of 15% LB-glycerol to the plate and scraping with an L-shaped cell scraper (Fisher Scientific catalog no. 03-392-151) to gently remove colonies from the agar. The liquid "plate scrape culture" was then collected, and this process was performed a second time to ensure that all colonies were removed from the plate. The bacterial cells were then stored at −80°C before PCR amplification of metagenomic fragments and Illumina library creation.

Picking individual transformants and verifying tolerance phenotypes. After microbial growth was observed, but before total growth was collected, between four and 10 individual colonies were picked from a positive selection, amplified in LB broth with 50 μg/ml kanamycin (LB-Kan broth), and stored in 96-well plates with 15% glycerol at −80°C. These picked clones were subsequently subjected to liquid growth assays in Mueller-Hinton (MH) broth (2 g beef infusion solids, 1.5 g starch, 17.5 g casein hydrolysate, pH 7.4, in a final volume of 1 liter) supplemented with 50 μg/ml kanamycin and containing MICs of the compound on which the clones were originally selected, per Table S2 in the supplemental material. Before liquid growth assays, clones were passaged twice through LB-Kan broth by transfer with a 96-pin stamp transfer tool (each passage was allowed to grow overnight) to eliminate residual glycerol from freezer stocks. To test for tolerance in liquid medium, cultures were stamp transferred into 96-well plates containing 250 μl of medium with inhibitory compound and growth was profiled by absorbance measurements at 600 nm (optical density at 600 nm [OD_{600}]) taken every 20 min using the Synergy H1 microplate reader (BioTek Instruments) for up to 4 days at 37°C. Clones positive for growth were rearrayed, stored at −80°C, and subjected to a second round of liquid growth assays, performed exactly as described above. Only clones showing growth in both liquid growth assays were considered tolerant and maintained for downstream analyses.

Amplification of metagenomic DNA fragments from selected transformants. Amplification of metagenomic fragments from picked transformants was performed by PCR, using 1 μl of cells from overnight culture. A sample PCR mixture consisted of 1 μl of template, 2.5 μl of ThermoPol reaction buffer (New England BioLabs), 0.5 μl of 10 mM deoxynucleotide triphosphates (dNTPs; New England BioLabs), 0.5 μl of *Taq* polymerase (New England BioLabs; 5 U/μl), 3 μl of a custom primer mix (see Table S3 in the supplemental material), and 17.5 μl of nuclease-free H_2O to bring the final reaction volume to 25 μl. PCR mixtures were then amplified using the following thermocycler conditions: 94°C for 10 min and 35 cycles of 94°C for 5 min, 55°C for 45 s, 72°C for 5.5 min, and 72°C for 10 min. The amplified metagenomic inserts were then purified using the QIAquick 96-well PCR purification kit (Qiagen), per the manufacturer's recommendations. After purification, total eluate for fragments from colonies on the same selection was pooled for Illumina library preparation and sequencing, resulting in 50 pools, each representing a unique selection.

Amplification of metagenomic DNA fragments from total bacterial growth. Freezer stocks of plate scrape cultures from positive selections were thawed, and 200 to 250 μl of cells was pelleted by centrifugation at 13,000 rpm for 2 min and gently washed with 1 ml of nuclease-free H_2O. Cells were then pelleted again and resuspended in 20 to 25 μl nuclease-free water (1/10 original volume), frozen at −20°C, and then thawed to promote cell lysis. Subsequently, 2.5 μl of thawed resuspensions was used as the template for amplification of resistance-conferring DNA fragments by PCR with *Taq* DNA polymerase (New England BioLabs) exactly as described above, except that 25 cycles of PCR, rather than 35, were performed.

Illumina sample preparation and sequencing. Pooled amplicons from picked colonies and amplified metagenomic inserts from total bacterial growth were processed similarly as previously described (28). For each input, 500 ng of PCR product was sheared to 150- to 200-bp fragments using nine 10-minute cycles of 30 s on and 30 s off on the Bioruptor XL sonicator (high-power setting). The ends of sheared DNA fragments were blunted with T4 DNA ligase and polymerase (New England BioLabs), and barcoded adapters were ligated to blunt fragments by T-A cloning with T4 DNA ligase (New England BioLabs). Next, 10 μl of adapter-ligated samples was combined into pools of 12 samples and concentrated by elution through a Qiagen MinElute PCR purification column. DNA sized 300 to 400 bp was purified by gel extraction, and adapter-

ligated samples were enriched by PCR using 12.5 μl 2× Phusion HF master mix, 1 μl of 10 μM Illumina PCR primer mix, and 2 μl of purified DNA as the template in a 25-μl PCR mixture. PCR proceeded as follows: 30 s at 98°C followed by 17 cycles of 98°C for 10 s, 65°C for 30 s, and 72°C for 30 s before a final 5-min extension at 72°C. Finally, a 9 pM sample was used for Illumina HiSeq 101-bp sequencing with the HiSeq 2500 platform at GTAC (Genome Technology Access Center, Washington University in St. Louis, MO, USA).

Assembly and annotation of metagenomic contigs. Illumina paired-end sequence reads were binned by barcode (perfect match required), such that independent selections were assembled and annotated in parallel. Metagenomic DNA fragments from each selection were assembled using PARFuMS (parallel annotation and reassembly of functional metagenomic selections), a tool developed for the high-throughput assembly and annotation of functional metagenomic selections (36). Assembly with PARFuMS consists of three iterations of variable job size with the short-read assembler Velvet (37), two iterations of assembly with PHRAP (38), and custom scripts to clean sequence reads, remove chimeric assemblies, and link contigs by coverage and common annotation, as described previously (36). In total, 116 sequencing libraries were assembled with PARFuMS, encompassing 84 unique selections (for 32 selections, picked colonies and total bacterial growth were sequenced and assembled separately). Fifty-two selections yielded metagenomic fragments greater than 500 bp, totaling 1,932 metagenomic contigs assembled. To annotate these contigs, we predicted ORFs using the gene-finding algorithm MetaGeneMark (39) and predicted gene function by searching the amino acid sequences against the TIGRFAMs (40) and Pfam (41) profile HMM databases with HMMER3 (42). MetaGeneMark was run using default gene-finding parameters while hmmscan (HMMER3) was run with the option "-cut_ga," requiring that genes meet profile-specific gathering thresholds (rather than a global, more permissive, default log odds cutoff) before receiving annotation. The hit with the best E value was used to designate an annotation, and ORFs over 350 bp were used in downstream analyses.

Percent identity comparisons of recovered ORFs against NCBI database. All recovered ORFs greater than 350 bp were compared against the NCBI protein Non-Redundant (NR) database (retrieved 15 September 2014) using BLASTX to determine the amino acid identity between recovered ORFs and their nearest neighbor in an NCBI protein. For each ORF, the NCBI entry that generated the best local alignment was used to seed global alignments with estwise (http://dendrome.ucdavis.edu/resources/tooldocs/wise2/doc_wise2.html). The following options were used in global alignment: "-init global" and "-alg 333." From this alignment, global percent identities were calculated as the number of matched amino acids divided by the full length of the shorter of the two sequences compared.

Subcloning putative tolerance-conferring genes. Putative tolerance-conferring genes were amplified from picked colonies twice verified for tolerance by liquid growth assay. Plasmids (pZE21 plus a soil metagenomic fragment) were purified by minipreparation using a Qiagen plasmid purification kit and used as the template for PCR of individual tolerance genes. Open reading frames were amplified using PFX polymerase (Life Technologies), with the following components in a 50-μl reaction mixture: 1.5 μl purified plasmid as the template, PFX buffer (10 μl), 10 mM dNTPs (New England BioLabs; 1.5 μl), 50 mM $MgSO_4$ (1 μl), PFX enhancer solution (5 μl), polymerase (0.4 μl), 10 μM forward primer (0.75 μl), and 10 μM reverse primer (0.75 μl). Thermocycler conditions were as follows: 95°C for 5 min followed by 35 cycles of 95°C for 45 s, 55°C for 45 s, and 72°C for 3.5 min, finishing with a 5-min extension at 72°C. Reaction-specific primers can be found in Table S4 in the supplemental material. Amplicons were then purified using RapidTip2 PCR purification tips (Diffinity Genomics) and quantified using the BR Qubit fluorometer assay kit. All PCR products were cloned 9 bp downstream of the ribosome binding site within the pZE21 expression vector (35) and transformed into *E. coli* MegaX cells (Invitrogen). Amplicons were then phosphorylated using T4 polynucleotide kinase (PNK) (Epicentre; catalog no. P0505H) per the manufacturer's recommendations and ligated into pZE21 using the Fast-Link DNA ligation kit (Epicentre; catalog no. LK11025) with suggested protocols. Subcloned constructs were transformed into chemically competent *E. coli* MegaX by heat shock at 42°C, and clones were verified for proper size, sequence, and orientation by diagnostic PCR.

Quantifying tolerance genes by liquid growth assay. The putative tolerance-conferring capability of each transformant was tested against the inhibitor on which the genotype was identified, in LB medium supplemented with kanamycin at 50 μg/ml. Growth assays were performed in 96-well plates, and inhibitor concentrations were arrayed via serial dilution in LB-Kan broth. Each construct was tested against a range of inhibitor concentrations generated via these dilutions; Fig. 3 depicts growth curves at the highest inhibitor concentration for which tolerant growth was reproducibly observed. All growth curves were performed in triplicate and compared directly against an empty-vector control strain grown in identical medium on the same 96-well plate. Medium was inoculated with 5 μl of cells grown to mid-log phase at a final volume of 200 μl, and plates were sealed with a Breathe-Easy membrane (Sigma-Aldrich; catalog no. Z380059) permeable to oxygen, carbon dioxide, and water vapor. Growth was profiled at 37°C by absorbance measurements at 600 nm using the Synergy-H1 microplate reader (BioTek Instruments, Inc.), under constant shaking at "medium" intensity; readings were taken every 15 min for a minimum of 48 h.

For defined-medium growth assays, M9 medium was prepared using M9 salts (Sigma-Aldrich; catalog no. M6030) supplemented with 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 10 mg/ml glucose, 50 μg/ml leucine, 50 μg/ml thiamine, and 50 μg/ml kanamycin. For toxicity assays, stationary-phase cultures (grown in M9 plus glucose) for each putative tolerance strain (and control) were subcultured and grown to mid-log (OD_{600}, 0.4 to 0.7) using M9-plus-glucose medium. Then, 5 μl was inoculated into 200 μl of M9-plus-glucose medium which also contained the indicated concentration of toxin (see Fig. S3 in the supplemental material). Inhibitor concentrations were arrayed via serial dilution, and growth assays were performed in triplicate on 96-well plates; Fig. S3 in the supplemental material depicts growth curves at the highest inhibitor concentration for which tolerant growth was observed. Plates were handled by the BioStack automated microplate stacker, with absorbance readings taken as described above at 37°C every 45 min for 3 days (cultures were shaken at medium intensity for 2 min before measurement).

Nucleotide sequence accession numbers. All assembled sequences have been deposited in GenBank and are available via the BioProject identifier PRJNA294310 (GenBank accession numbers KT941423 to KT943354).

RESULTS

We constructed 16 small-insert metagenomic libraries using soils from two sites in the Long Term Ecological Research network: grassland soils from Cedar Creek (CC) and agricultural soils from Kellogg Biological Station (KBS) (26–28). Metagenomic libraries ranged in size from 1.4 to 16.9 Gb (see Table S1 in the supplemental material) and were selected for tolerance to one of 20 compounds (15 common lignocellulosic inhibitors and five current and next-generation biofuels) on solid medium at the MIC of each compound against the *E. coli* host (see Table S2). Growth was monitored for 5 days, with tolerance identified to seven of 20 compounds (four lignocellulosic inhibitors and three fuels). Tolerance appeared to be largely compound specific (as opposed to library or soil specific), indicating that for most compounds, edaphic properties did not influence the ability to capture a tolerance phenotype (Fig. 1). For a subset of the 13 compounds where tolerant clones were not identified, organisms within the soil are known to metabolize the chemical (23, 43), and in at least one

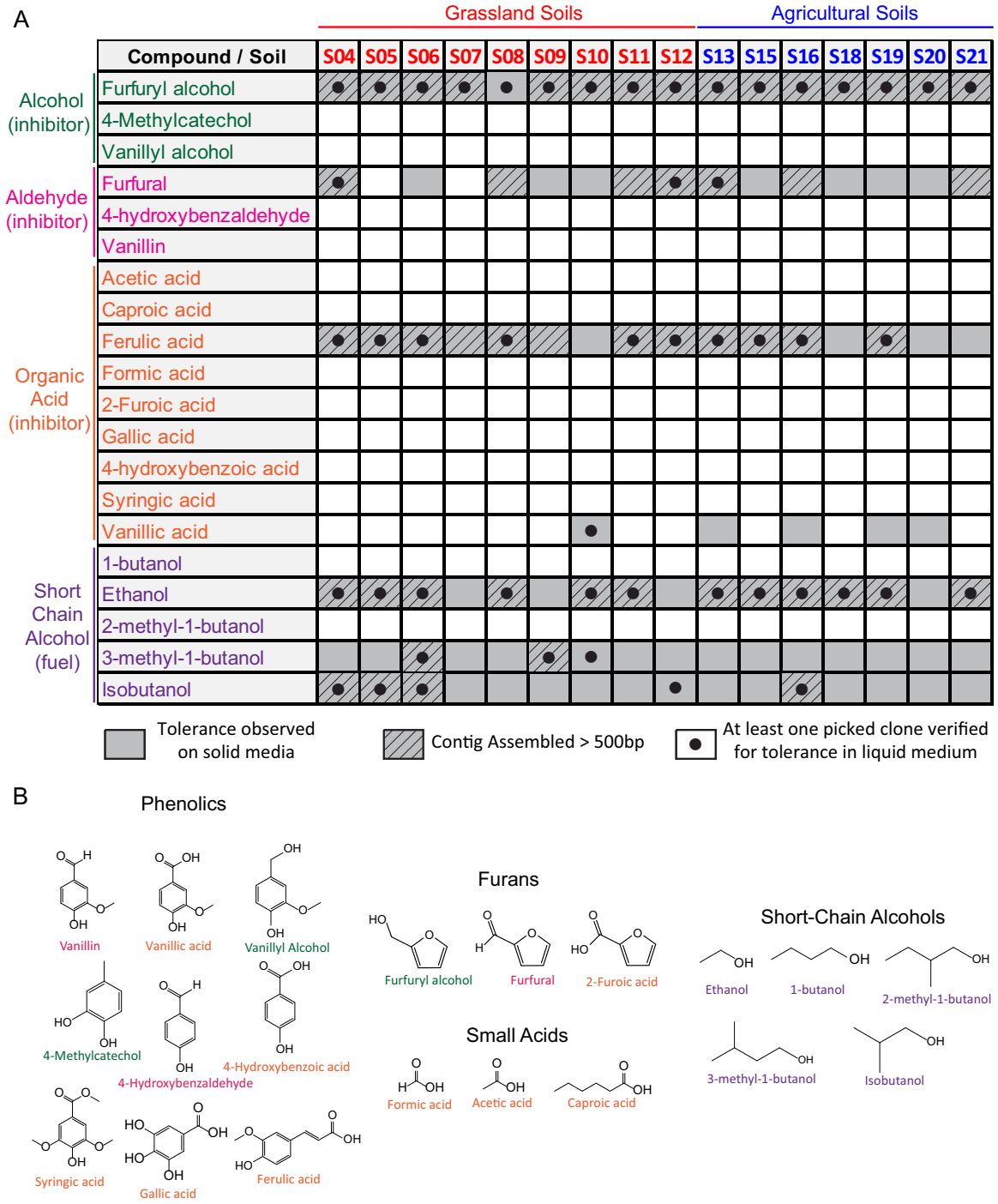


FIG 1 Functional selections of 16 soil metagenomes for tolerance to 20 inhibitory compounds. (A) Gray cells designate selections where a tolerance phenotype was observed, hatched cells designate selections where a metagenomic contig greater than 500 bp was assembled, and dotted cells indicate selections from which individual clones were twice validated for tolerance in liquid medium. (B) Structures of the inhibitors and fuels used in functional selections.

instance, tolerance to 2-furoic acid has been transferred to *E. coli* using soil-derived large-insert metagenomic libraries (19). Thus, the absence of tolerant growth in our experiments does not signify the absence of these properties within the soil microbial community or the inability to transfer the trait to *E. coli*. Instead, it is likely that tolerance to these compounds either requires the action of multiple, distal genes or is encoded by genes at low to intermediate abundance, such that the functional gene was not incorporated into the original metagenomic libraries. Alternative strategies (e.g., *in situ* enrichment of soil communities [44, 45]) are likely required to expand both the number of compounds to which tolerance is observed and the diversity of tolerance-conferring genes recovered.

Some selections yielded unambiguous tolerance phenotypes,

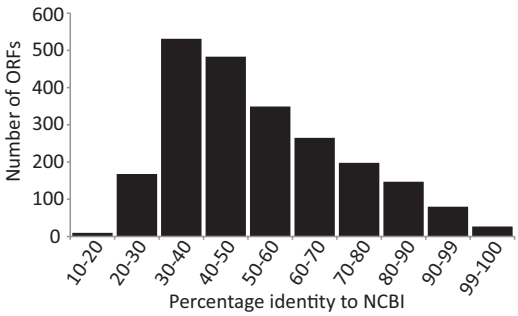


FIG 2 Amino acid identities between all ORFs ($n = 2,283$) and their top hits in the NCBI protein database.

defined as punctate colonies on experimental plates (those with metagenomic libraries) and no observable growth on matched, empty-vector controls. In other cases, small satellite colonies surrounded larger, likely bona fide tolerant clones. This phenotype may result from a larger colony that detoxifies the compound and allows for the survival of nearby nontolerant cells (e.g., see Fig. S1 in the supplemental material). Accordingly, further validation was often required to distinguish truly tolerant clones from background. As a means of validation, 483 colonies from the 99 selections that showed a positive phenotype (Fig. 1) were randomly selected and verified for tolerance by passage through liquid medium containing MICs of the appropriate compound (see Table S2). In total, 132 clones from 50 selections reproducibly exhibited a tolerance phenotype across two independent liquid growth assays and were selected for downstream analyses (Fig. 1).

For 66 of the 99 positive selections (~67%), all bacterial growth from solid selection plates was collected and putative tolerance-conferring metagenomic fragments were amplified by PCR and prepared in Illumina sequencing libraries. These 66 libraries were barcoded by selection and combined with another 50 libraries constructed from tolerance-conferring DNA fragments amplified from the 132 verified tolerant clones, also barcoded by selection. The 116 sequencing libraries included (i) 34 selections where total growth, but not verified clones, was interrogated; (ii) 18 selections where only verified clones were sequenced; and (iii) 32 selections where data of both types were generated. All 116 libraries were sequenced and assembled using PARFuMS (36), with 52 of 84 selections yielding at least one assembled contig longer than 500 bp (Fig. 1). In total, 1,932 metagenomic fragments were assembled (N_{50} of >1.35 kb) containing 2,283 ORFs which generally appeared unlike sequences deposited in public repositories, consistent with past observations from soil metagenomes (28, 46–51). The average amino acid identity of all 2,283 ORFs to their closest homolog in the NCBI database was 52.7% ± 19.0% (mean ± standard deviation) (Fig. 2).

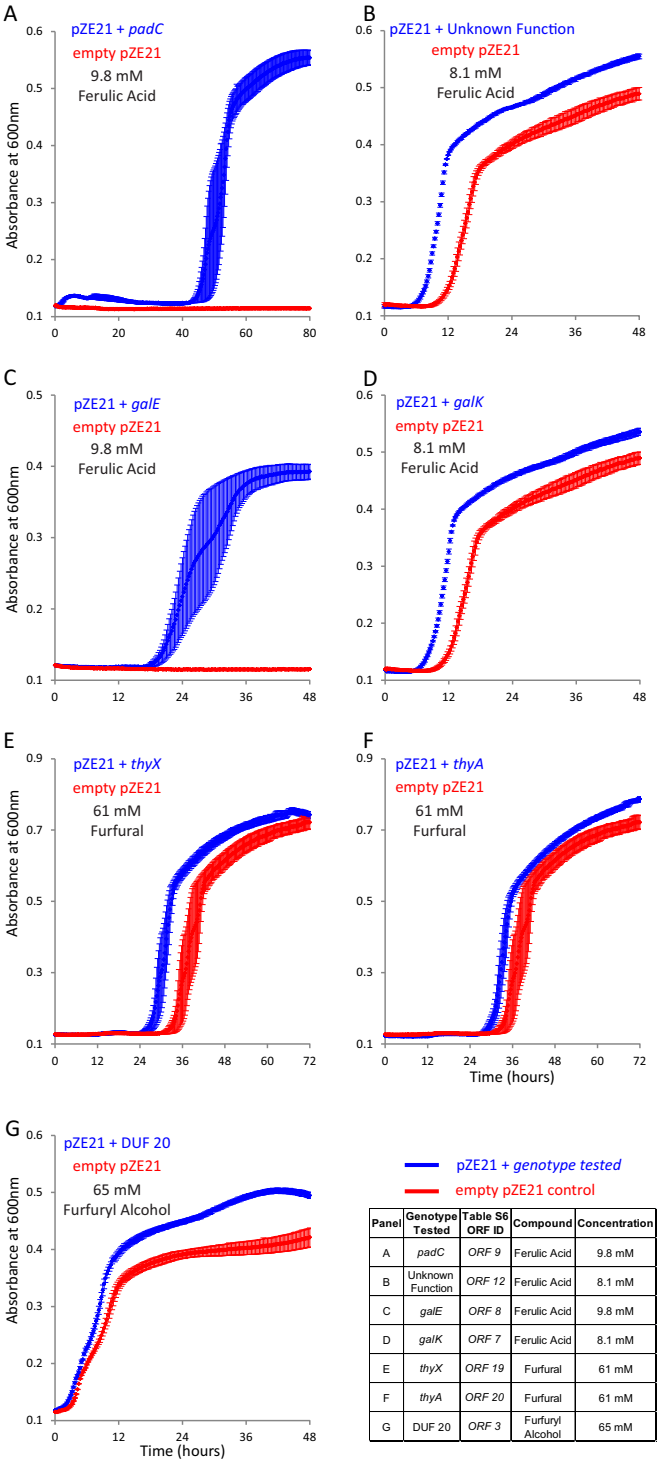
The most comprehensive sequence data were available for selections using furfuryl alcohol, ferulic acid, and furfural, and we focused downstream work on sequences identified from selections using these compounds. These chemicals encompass alcohol, acid, and aldehyde inhibitors and include structures derived from both the hemicellulosic and lignified fractions of lignocellulose.

Glyoxylase resistance proteins were the most common function identified in furfural selections (see Table S5 in the supplemental material), which are archetypal aldehyde-detoxifying enzymes (52). The heterologous expression of a glyoxalase in *E. coli* from the soil bacterium *Pseudomonas putida* has been demonstrated to improve tolerance and fuel yield in the presence of the aldehyde inhibitor methylglyoxal (53), but our data are the first to implicate a putative lactoylglutathione lyase in furfural tolerance. From selections on furfuryl alcohol, we recovered ORFs with vague or uncharacterized predicted function (e.g., HD domain and domain of unknown function [DUF] 20 [see Table S5]). Given the sequence divergence of the cloned soil metagenome from known functions (Fig. 2), vague or unassignable annotations are not surprising and may be particularly common with selections using furfuryl alcohol given the compound's poorly characterized modes of toxicity (14).

The most frequently recovered genes from ferulic acid selections were homologous to *galE*, a UDP-glucose 4-epimerase which catalyzes the interconversion between UDP-glucose and UDP-galactose (54) (see Table S5 in the supplemental material). Interestingly, many substituted phenolic compounds inhibit UDP-glucose 4-epimerases (54, 55). Functional deficiencies in these enzymes result in cell wall defects (in the absence of galactose) (56) or cell death (in the presence of galactose, a component of plant hemicellulose [6]) (57). The overexpression of *galE* homologs may rescue the native UDP-glucose 4-epimerase of *E. coli* in the presence of ferulic acid, a guaiacyl phenolic compound. Given the structural diversity of phenolic compounds capable of inhibiting GalE (54, 55), its rescue may represent a generalized strategy for engineering tolerance toward products of lignin hydrolysate. Indeed, the overexpression of *galE* homologs confers tolerance against the syringyl phenolic monomer syringaldehyde (19) and the substituted aromatic menadione (58) when expressed in *E. coli*.

To identify individual ORFs that confer a tolerance phenotype and to quantify an ORF's impact on tolerance when removed from its native genetic context (as would occur in engineered production strains), we subcloned ORFs from their original metagenomic fragment and assayed for tolerance to the compound on which the gene was originally selected. We prioritized functions predicted to encompass broad mechanistic diversity, amplifying ORFs from colonies previously verified for tolerance (rather than from total bacterial growth on functional selections), as this allowed us to select transformants that displayed strong tolerance phenotypes. Across furfuryl alcohol, ferulic acid, and furfural selections, we cloned 23 predicted ORFs from 14 assembled contigs, seven of which were confirmed to provide tolerance when removed from their native context, spanning all three compounds (Fig. 3; see also Table S6 in the supplemental material). No tolerance-conferring gene impaired growth when expressed in the absence of inhibitor (see Fig. S2), suggesting that these genotypes do not generally reduce fitness and are therefore suitable for bioprocessing applications.

Two genes with unknown function conferred tolerance to ferulic acid and furfuryl alcohol, with shortened lag times and increased growth rates when expressed heterologously in *E. coli* compared to empty-vector controls (Fig. 3B and G). This highlights the potential of uncharacterized gene products from the soil metagenome to impact bioprocessing applications. We also identified two thymidylate synthases, homologs of ThyA and ThyX, which confer furfural tolerance (Fig. 3E and F). These enzymes catalyze the production of deoxythymidine monophosphate (dTMP) from dUMP, and their detection is consistent with pre-



Panel	Genotype Tested	Table S6 ORF ID	Compound	Concentration
A	*padC*	*ORF 9*	Ferulic Acid	9.8 mM
B	Unknown Function	*ORF 12*	Ferulic Acid	8.1 mM
C	*galE*	*ORF 8*	Ferulic Acid	9.8 mM
D	*galK*	*ORF 7*	Ferulic Acid	8.1 mM
E	*thyX*	*ORF 19*	Furfural	61 mM
F	*thyA*	*ORF 20*	Furfural	61 mM
G	DUF 20	*ORF 3*	Furfuryl Alcohol	65 mM

FIG 3 Open reading frames (ORFs) confer tolerance to lignocellulosic inhibitors when removed from their native context in the soil metagenome. The *y* axes of each panel give the absorbance at 600 nm, values are the averages from three trials, and error bars depict standard errors. Growth of *E. coli* expressing each genotype is in blue, and an empty-vector control (pZE21) is in red.

vious work demonstrating that ThyA overexpression confers furfural tolerance on ethanologenic *E. coli* (59). We show that homologs of both ThyA and ThyX provide furfural tolerance, even though these enzymes are evolutionarily unrelated and function via different reductive mechanisms (60).

A phenolic acid decarboxylase (PadC) with perfect amino acid identity to homologs from the genus *Pantoea* (NCBI GenInfo number 497939809) was cloned and confirmed to confer tolerance to ferulic acid, a lignin monomer (Fig. 3A). Ferulic acid decarboxylases are of high industrial interest, as they catalyze the conversion of ferulic acid to 4-vinyl guaiacol, an important precursor for vanillin production with approximately 30 times the value of ferulic acid (61). Indeed, multiple groups have expressed PadC homologs in *E. coli* and have shown that these enzymes can produce 4-vinyl guaiacol from ferulic acid (61, 62). In contrast to previous work, which suggested that expressing a phenolic acid decarboxylase impaired the growth of *E. coli* and other Gram-negative bacteria (62), we demonstrate improved ferulic acid tolerance when expressing the decarboxylase (Fig. 3A). Toxicity differences between our study and previous work may result from differences in PadC expression level or the time course used in growth assays (PadC confers delayed ferulic acid tolerance at some inhibitor concentrations [see Fig. S3A in the supplemental material]). Nonetheless, PadC holds the potential to couple biomass detoxification with the production of the high-value chemical 4-vinyl guaiacol, offering a means to offset costs associated with biomass conversion.

The observed enrichment for galactose metabolism in ferulic acid selections, particularly *galE* homologs (see Table S5 in the supplemental material), suggested that the overexpression of *galE* may provide ferulic acid tolerance. Consistent with this hypothesis, a *galE* homolog conferred strong tolerance to ferulic acid when removed from its native context; a recombinant *E. coli* strain expressing this gene grew robustly at concentrations of ferulic acid entirely inhibitory to a vector-only control (Fig. 3C). Because substituted phenolics inhibit UDP-glucose 4-epimerases (54, 55), we speculate that *galE* overexpression may provide tolerance by rescuing a compromised native UDP-glucose 4-epimerase. In addition to *galE* homologs, ferulic acid selections also yielded numerous homologs of galactokinase (*galK*), another nucleotide-galactose active enzyme (see Table S5), which conferred tolerance to ferulic acid when overexpressed (Fig. 3D). The *galK* homologs identified by functional metagenomic selections all contain mutations at residues predicted to eliminate galactokinase activity (63–65) (see Table S7), suggesting that tolerance is conferred by functionally impaired GalK variants.

Sugars are the predominant carbon source in most bioprocessing applications, whereas they constitute a minor source of carbon in LB medium compared to small peptides. To determine whether the observed ferulic acid tolerance genes improved *E. coli* growth under conditions more representative of real-world bioreactors, we tested each tolerance gene against a panel of lignin-derived phenolic monomers in M9 minimal medium supplemented with glucose as the primary carbon source. These minimal medium experiments typically reproduced the tolerance phenotypes observed using LB medium, though the growth curves occasionally differed between experiments (compare Fig. 3 to Fig. S3 in the supplemental material). Although it is tempting to attribute these growth differences to medium composition, other growth conditions also differed between experiments (namely, periods of orbital shaking, as described in Materials and Methods), and we therefore compare growth differences within, but not across, experiments.

With the exception of *galK*, each gene conferred tolerance to multiple phenolic monomers in minimal medium (Table 1; see

TABLE 1 Summary of tolerance genes active against phenolic inhibitors[a]

Compound	Tolerance activity of ORF/genotype			
	ORF9/ *padC*	ORF12/unknown function	ORF8/ *galE*	ORF7/ *galK*
Ferulic acid (LB medium)	+	+	+	+
Ferulic acid (M9 medium)	+	+	+	−
Vanillin	+	+	+	ND
Vanillic acid	+	+	+	−
Vanillyl alcohol	+	+	+/−	−
4-Methylcatechol	−	+	−	−
Syringic acid	+	−	+/−	−
4-Hydroxybenzoic acid	+	+	+	−
4-Hydroxybenzaldehyde	+	−	−	−

[a] All data were generated using M9 defined medium, except as indicated. +, tolerance conferred; −, no tolerance; +/−, ambiguous tolerance; ND, not determined.

also Fig. S3 in the supplemental material), despite being recovered from only ferulic acid in the original selections using LB medium. Interestingly, expression of *galE* conferred tolerance in both LB medium, which contains galactose (66), and M9-plus-glucose medium, which lacks galactose. Expressing our *galK* variant, however, provided tolerance only in the presence of galactose (Fig. 3; see also Fig. S3). This is consistent with previous observations that suggest that an impaired GalE enzyme can exhibit galactose-independent toxicity (56) but indicates that our GalK variants likely ameliorate only galactose-dependent forms of toxicity.

DISCUSSION

Analogous to the predicted mechanism for *galE*-mediated ferulic acid tolerance, simple complementation of an impaired GalK enzyme could underlie the tolerance conferred by *galK* overexpression. Although intuitive, this hypothesis is unlikely to explain our observed ferulic acid tolerance, as it is inconsistent with previous observations: *galK* mutants rescue *galE* mutants (56, 67), likely by preventing the accumulation of GalK's reaction product, galactose-1-phosphate (67). Therefore, increased GalK enzyme is expected to exacerbate the deleterious effects of an impaired GalE, and the overexpression of a functional *galK* gene is unlikely to compensate for a phenolic-inhibited GalE.

Instead, we hypothesize that the tolerance-conferring GalK variants may exhibit a dominant negative phenotype. For instance, a GalK variant may be catalytically inactive and yet able to inhibit *E. coli*'s native GalK by competition for substrate, relieving galactose-1-phosphate accumulation and ameliorating the galactose-dependent toxicity associated with an impaired GalE. Supporting this hypothesis, each of the five full-length *galK* homologs assembled from ferulic acid selections contains mutations in core motifs that are conserved across diverse bacteria, archaea, yeast, and mammals (see Table S7 and Text S1 in the supplemental material). In four of five cases (including in the enzyme expressed in Fig. 3D), mutations at these residues have been shown to eliminate galactokinase activity (63–65), with the fifth enzyme predicted to contain an alternative start codon that would truncate a key functional motif (68) (see Table S7), suggesting that loss-of-function variants of *galK* provide ferulic acid tolerance. Phenolic acids are also key toxins in bioproduction efforts that utilize the yeast *Saccharomyces cerevisiae* (11), for which GALE deficiency similarly triggers galactose toxicity which is rescued upon GALK deletion (69). Though phenolic acids inhibit GalE in many organisms (e.g., trypanosomes and mammals [55]), whether they exert toxicity via GALE impairment in *S. cerevisiae* is unknown and warrants future work.

We also identified two thymidylate synthase variants, ThyA and ThyX, which can confer furfural tolerance on *E. coli* when expressed heterologously. Despite furfural toxicity resulting (in part) from NADPH starvation (59, 70), both ThyA and ThyX are known to cause NADPH oxidation as a consequence of dTMP synthesis, though oxidation occurs at different catalytic steps with each enzyme. ThyA can use methylenetetrahydrofolate (MTHF) as both a carbon donor and a reductant to synthesize dTMP from dUMP, producing dihydrofolate (DHF). DHF must be reduced to tetrahydrofolate (THF) before it can be recycled back into the original methyl donor, MTHF (60, 71, 72); this reduction requires the oxidation of NADPH. In contrast, ThyX produces THF as a direct product of dTMP synthesis but oxidizes NADPH during this reaction, obviating the need for DHF reduction but not NADPH oxidation (60, 71, 72). Because both thymidylate synthases result in oxidized NADPH and yet both confer furfural tolerance when overexpressed, these enzymes likely ameliorate furfural toxicity independently of the compound's effect on NADPH levels (59, 70, 73, 74).

Expression of ThyX, but not ThyA, reduces dependence on DHF-reducing enzymes (60, 71, 72), and yet the two thymidylate synthases confer similar levels of furfural tolerance (Fig. 3E and F). This suggests that the furfural tolerance conferred by these enzymes is also independent of their differential folate-reducing capacities and implies that furfural does not inhibit folate reduction reactions, as has been theorized elsewhere (59). Furfural is also known to cause DNA damage (75, 76), perhaps via the induction of reactive oxygen species (77). Accordingly, the increased furfural tolerance that results from ThyA and ThyX overexpression may result from an increased supply of pyrimidine deoxyribonucleotides to aid in DNA repair (59). Since both enzymes are expected to increase dTMP production, this explanation for furfural-induced toxicity most satisfactorily describes the similar tolerance profiles conferred by each thymidylate synthase.

We describe gain-of-function tolerance to acid, alcohol, and aldehyde inhibitors derived from hemicellulose and lignin, demonstrating that functional selections using soil metagenomes have tremendous potential to address the toxicity associated with thermochemical lignocellulose depolymerization. Because bioprocessing applications are necessarily diverse and are expected to utilize variable lignocellulosic feedstocks (8, 9, 78), our capacity to understand and engineer tolerance into microbial catalysts must become comparatively flexible. By performing selections against individual chemical inhibitors, rather than actual pretreated hydrolysate, we identified tolerance genes that yield insight into the mechanisms of individual inhibitor toxicity (e.g., the putative inhibition of GalE by phenolic compounds) and associated microbial tolerance (e.g., increased pyrimidine supply to provide furan tolerance). This strategy allows for flexible bioengineering efforts, where custom production hosts may be designed with inhibitor tolerance tailored particularly to their bioprocessing applications.

The toxicity associated with treated lignocellulosic hydrolysate likely results from the combination of many inhibitors at low concentrations (9). In contrast, the tolerance genes that we describe were recovered from selections on individual compounds. These genes nonetheless hold the potential to confer broad-spectrum

tolerance. We provide several examples of genes identified on ferulic acid selections that confer tolerance to many additional phenolic compounds (see Fig. S3 in the supplemental material). Future work should test combinations of genes for broad-spectrum tolerance to synthetic inhibitor mixtures and real-world hydrolysates, enabling the empirical construction of custom tolerance operons for diverse inhibitor profiles.

Despite the diversity of our functional metagenomic libraries, selections did not identify genes for tolerance to several lignocellulosic inhibitors known to be tolerated by soil microorganisms (19, 23, 43). This outcome is likely a result of the incredible diversity of soil metagenomes, which confounds efforts to identify low-abundance genotypes. Methods to enrich these important yet low-abundance functions should improve the capture efficiency of functional selections and enable the discovery of new gene functions (44, 45). Even without applying enrichment strategies, we discover numerous tolerance genes active against hemicellulose- and lignin-derived inhibitors, including genes of entirely unknown function (e.g., Fig. 3B and G) as well as enzymes with well-defined applications (61, 62) (e.g., PadC [Fig. 3A]). Our discoveries highlight the potential for the soil metagenome to enable process improvements in industrial biotechnology. The tremendous genetic diversity in soil (28, 46–51) (Fig. 2), however, remains mostly unexplored, and this potential is largely waiting to be realized.

ACKNOWLEDGMENTS

We thank members of the Dantas lab for their help with protocol optimization and for thoughtful discussions on the results and analyses presented here, the Genome Technology Access Center at Washington University School of Medicine for Illumina sequencing services, and N. Fierer and C. Lauber for providing soil samples.

Research reported in this publication was supported in part by awards to G.D. through the Children's Discovery Institute (MD-II-2011-117), the International Center for Advanced Renewable Energy and Sustainability at Washington University, the National Academies Keck Futures Initiatives (Synthetic Biology, SB2), and the NIH Director's New Innovator Award (DP2-DK-098089). K.J.F. is a National Science Foundation graduate research fellow (award number DGE-1143954) and received support from training grants through the National Human Genome Research Institute (NHGRI; https://www.genome.gov/) and the NIGMS through award numbers T32HG000045 and T32GM007067. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

FUNDING INFORMATION

Children's Discovery Institute provided funding to Gautam Dantas under grant number MD-II-2011-117. International Center for Advanced Renewable Energy and Sustainability at Washington University provided funding to Gautam Dantas under grant number 2011-Dantas. National Academies - Keck Futures Initiative provided funding to Gautam Dantas. National Science Foundation (NSF) provided funding to Kevin J. Forsberg under grant number DGE-1143954. HHS | NIH | National Human Genome Research Institute (NHGRI) provided funding to Kevin J. Forsberg under grant number T32 HG000045. HHS | NIH | National Institute of General Medical Sciences (NIGMS) provided funding to Kevin J. Forsberg under grant number T32 GM007067. HHS | NIH | NIH Office of the Director (OD) provided funding to Gautam Dantas under grant number DP2 DK098089.

REFERENCES

1. **Simmons BA, Loque D, Blanch HW.** 2008. Next-generation biomass feedstocks for biofuel production. Genome Biol **9:**242. http://dx.doi.org/10.1186/gb-2008-9-12-242.
2. **Hill J, Nelson E, Tilman D, Polasky S, Tiffany D.** 2006. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. Proc Natl Acad Sci U S A **103:**11206–11210. http://dx.doi.org/10.1073/pnas.0604600103.
3. **Rubin EM.** 2008. Genomics of cellulosic biofuels. Nature **454:**841–845. http://dx.doi.org/10.1038/nature07190.
4. **Ragauskas AJ, Beckham GT, Biddy MJ, Chandra R, Chen F, Davis MF, Davison BH, Dixon RA, Gilna P, Keller M, Langan P, Naskar AK, Saddler JN, Tschaplinski TJ, Tuskan GA, Wyman CE.** 2014. Lignin valorization: improving lignin processing in the biorefinery. Science **344:**1246843. http://dx.doi.org/10.1126/science.1246843.
5. **Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ, Jr, Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T.** 2006. The path forward for biofuels and biomaterials. Science **311:**484–489. http://dx.doi.org/10.1126/science.1114736.
6. **Chundawat SP, Beckham GT, Himmel ME, Dale BE.** 2011. Deconstruction of lignocellulosic biomass to fuels and chemicals. Annu Rev Chem Biomol Eng **2:**121–145. http://dx.doi.org/10.1146/annurev-chembioeng-061010-114205.
7. **Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD.** 2007. Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science **315:**804–807. http://dx.doi.org/10.1126/science.1137016.
8. **Du B, Sharma LN, Becker C, Chen SF, Mowery RA, van Walsum GP, Chambliss CK.** 2010. Effect of varying feedstock-pretreatment chemistry combinations on the formation and accumulation of potentially inhibitory degradation products in biomass hydrolysates. Biotechnol Bioeng **107:**430–440. http://dx.doi.org/10.1002/bit.22829.
9. **van der Pol EC, Bakker RR, Baets P, Eggink G.** 2014. By-products resulting from lignocellulose pretreatment and their inhibitory effect on fermentations for (bio)chemicals and fuels. Appl Microbiol Biotechnol **98:**9579–9593. http://dx.doi.org/10.1007/s00253-014-6158-9.
10. **Weber C, Farwick A, Benisch F, Brat D, Dietz H, Subtil T, Boles E.** 2010. Trends and challenges in the microbial production of lignocellulosic bioalcohol fuels. Appl Microbiol Biotechnol **87:**1303–1315. http://dx.doi.org/10.1007/s00253-010-2707-z.
11. **Klinke HB, Thomsen AB, Ahring BK.** 2004. Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. Appl Microbiol Biotechnol **66:**10–26. http://dx.doi.org/10.1007/s00253-004-1642-2.
12. **Almeida JRM, Modig T, Petersson A, Hahn-Hagerdal B, Liden G, Gorwa-Grauslund.** 2007. Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by *Saccharomyces cerevisiae*. J Chem Technol Biotechnol **82:**340–349. http://dx.doi.org/10.1002/jctb.1676.
13. **Zaldivar J, Ingram LO.** 1999. Effect of organic acids on the growth and fermentation of ethanologenic *Escherichia coli* LY01. Biotechnol Bioeng **66:**203–210.
14. **Zaldivar J, Martinez A, Ingram LO.** 2000. Effect of alcohol compounds found in hemicellulose hydrolysate on the growth and fermentation of ethanologenic *Escherichia coli*. Biotechnol Bioeng **68:**524–530.
15. **Zaldivar J, Martinez A, Ingram LO.** 1999. Effect of selected aldehydes on the growth and fermentation of ethanologenic *Escherichia coli*. Biotechnol Bioeng **65:**24–33. http://dx.doi.org/10.1002/(SICI)1097-0290(19991005)65:1<24::AID-BIT4>3.0.CO;2-2.
16. **Almeida JR, Bertilsson M, Gorwa-Grauslund MF, Gorsich S, Liden G.** 2009. Metabolic effects of furaldehydes and impacts on biotechnological processes. Appl Microbiol Biotechnol **82:**625–638. http://dx.doi.org/10.1007/s00253-009-1875-1.
17. **Wilson DB.** 2009. Cellulases and biofuels. Curr Opin Biotechnol **20:**295–299. http://dx.doi.org/10.1016/j.copbio.2009.05.007.
18. **Jarboe LR, Zhang X, Wang X, Moore JC, Shanmugam KT, Ingram LO.** 2010. Metabolic engineering for production of biorenewable fuels and chemicals: contributions of synthetic biology. J Biomed Biotechnol **2010:**761042. http://dx.doi.org/10.1155/2010/761042.
19. **Sommer MO, Church GM, Dantas G.** 2010. A functional metagenomic approach for expanding the synthetic biology toolbox for biomass conversion. Mol Syst Biol **6:**360. http://dx.doi.org/10.1038/msb.2010.16.
20. **Suzuki H, MacDonald J, Syed K, Salamov A, Hori C, Aerts A, Henrissat B, Wiebenga A, VanKuyk PA, Barry K, Lindquist E, LaButti K, Lapidus A, Lucas S, Coutinho P, Gong Y, Samejima M, Mahadevan R, Abou-Zaid M, de Vries RP, Igarashi K, Yadav JS, Grigoriev IV, Master ER.** 2012. Comparative genomics of the white-rot fungi, *Phanerochaete car-*

nosa and *P. chrysosporium*, to elucidate the genetic basis of the distinct wood types they colonize. BMC Genomics **13:**444. http://dx.doi.org/10.1186/1471-2164-13-444.
21. **Bugg TD, Ahmad M, Hardiman EM, Singh R.** 2011. The emerging role for bacteria in lignin degradation and bio-product formation. Curr Opin Biotechnol **22:**394–400. http://dx.doi.org/10.1016/j.copbio.2010.10.009.
22. **Koopman F, Wierckx N, de Winde JH, Ruijssenaars HJ.** 2010. Identification and characterization of the furfural and 5-(hydroxymethyl)furfural degradation pathways of *Cupriavidus basilensis* HMF14. Proc Natl Acad Sci U S A **107:**4919–4924. http://dx.doi.org/10.1073/pnas.0913039107.
23. **Overhage J, Kresse AU, Priefert H, Sommer H, Krammer G, Rabenhorst J, Steinbuchel A.** 1999. Molecular characterization of the genes *pcaG* and *pcaH*, encoding protocatechuate 3,4-dioxygenase, which are essential for vanillin catabolism in *Pseudomonas* sp. strain HR199. Appl Environ Microbiol **65:**951–960.
24. **Rojas A, Duque E, Mosqueda G, Golden G, Hurtado A, Ramos JL, Segura A.** 2001. Three efflux pumps are required to provide efficient tolerance to toluene in *Pseudomonas putida* DOT-T1E. J Bacteriol **183:**3967–3973. http://dx.doi.org/10.1128/JB.183.13.3967-3973.2001.
25. **Dunlop MJ, Dossani ZY, Szmidt HL, Chu HC, Lee TS, Keasling JD, Hadi MZ, Mukhopadhyay A.** 2011. Engineering microbial biofuel tolerance and export using efflux pumps. Mol Syst Biol **7:**487. http://dx.doi.org/10.1038/msb.2011.21.
26. **Ramirez KS, Lauber CL, Knight R, Bradford MA, Fierer N.** 2010. Consistent effects of nitrogen fertilization on soil bacterial communities in contrasting systems. Ecology **91:**3463–3470. http://dx.doi.org/10.1890/10-0426.1.
27. **Fierer N, Lauber CL, Ramirez KS, Zaneveld J, Bradford MA, Knight R.** 2012. Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. ISME J **6:**1007–1017. http://dx.doi.org/10.1038/ismej.2011.159.
28. **Forsberg KJ, Patel S, Gibson MK, Lauber CL, Knight R, Fierer N, Dantas G.** 2014. Bacterial phylogeny structures soil resistomes across habitats. Nature **509:**612–616. http://dx.doi.org/10.1038/nature13377.
29. **Daniel R.** 2005. The metagenomics of soil. Nat Rev Microbiol **3:**470–478. http://dx.doi.org/10.1038/nrmicro1160.
30. **Riesenfeld CS, Schloss PD, Handelsman J.** 2004. Metagenomics: genomic analysis of microbial communities. Annu Rev Genet **38:**525–552. http://dx.doi.org/10.1146/annurev.genet.38.072902.091216.
31. **Chen Y, Nielsen J.** 2013. Advances in metabolic pathway and strain engineering paving the way for sustainable production of chemical building blocks. Curr Opin Biotechnol **24:**965–972. http://dx.doi.org/10.1016/j.copbio.2013.03.008.
32. **Bokinsky G, Peralta-Yahya PP, George A, Holmes BM, Steen EJ, Dietrich J, Lee TS, Tullman-Ercek D, Voigt CA, Simmons BA, Keasling JD.** 2011. Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*. Proc Natl Acad Sci U S A **108:**19949–19954. http://dx.doi.org/10.1073/pnas.1106958108.
33. **Wargacki AJ, Leonard E, Win MN, Regitsky DD, Santos CN, Kim PB, Cooper SR, Raisner RM, Herman A, Sivitz AB, Lakshmanaswamy A, Kashiyama Y, Baker D, Yoshikuni Y.** 2012. An engineered microbial platform for direct biofuel production from brown macroalgae. Science **335:**308–313. http://dx.doi.org/10.1126/science.1214547.
34. **Ling H, Teo W, Chen B, Leong SS, Chang MW.** 2014. Microbial tolerance engineering toward biochemical production: from lignocellulose to products. Curr Opin Biotechnol **29:**99–106. http://dx.doi.org/10.1016/j.copbio.2014.03.005.
35. **Lutz R, Bujard H.** 1997. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Res **25:**1203–1210. http://dx.doi.org/10.1093/nar/25.6.1203.
36. **Forsberg KJ, Reyes A, Wang B, Selleck EM, Sommer MO, Dantas G.** 2012. The shared antibiotic resistome of soil bacteria and human pathogens. Science **337:**1107–1111. http://dx.doi.org/10.1126/science.1220761.
37. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.
38. **de la Bastide M, McCombie WR.** 2007. Assembling genomic DNA sequences with PHRAP, 2008/04/23. Curr Protoc Bioinformatics Chapter 11:Unit11.4. http://dx.doi.org/10.1002/0471250953.bi1104s17.
39. **Zhu W, Lomsadze A, Borodovsky M.** 2010. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res **38:**e132. http://dx.doi.org/10.1093/nar/gkq275.
40. **Haft DH, Loftus BJ, Richardson DL, Yang F, Eisen JA, Paulsen IT, White O.** 2001. TIGRFAMs: a protein family resource for the functional identification of proteins. Nucleic Acids Res **29:**41–43. http://dx.doi.org/10.1093/nar/29.1.41.
41. **Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL.** 2000. The Pfam protein families database. Nucleic Acids Res **28:**263–266. http://dx.doi.org/10.1093/nar/28.1.263.
42. **Finn RD, Clements J, Eddy SR.** 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res **39:**W29–W37. http://dx.doi.org/10.1093/nar/gkr367.
43. **Priefert H, Rabenhorst J, Steinbuchel A.** 1997. Molecular characterization of genes of *Pseudomonas* sp. strain HR199 involved in bioconversion of vanillin to protocatechuate. J Bacteriol **179:**2595–2607.
44. **Narihiro T, Suzuki A, Yoshimune K, Hori T, Hoshino T, Yumoto I, Yokota A, Kimura N, Kamagata Y.** 2014. The combination of functional metagenomics and an oil-fed enrichment strategy revealed the phylogenetic diversity of lipolytic bacteria overlooked by the cultivation-based method. Microbes Environ **29:**154–161. http://dx.doi.org/10.1264/jsme2.ME14002.
45. **Jones MD, Rodgers-Vieira EA, Hu J, Aitken MD.** 2014. Association of growth substrates and bacterial genera with benzo[a]pyrene mineralization in contaminated soil. Environ Eng Sci **31:**689–697. http://dx.doi.org/10.1089/ees.2014.0275.
46. **Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, Owens S, Gilbert JA, Wall DH, Caporaso JG.** 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proc Natl Acad Sci U S A **109:**21390–21395. http://dx.doi.org/10.1073/pnas.1215210110.
47. **Fierer N, Ladau J, Clemente JC, Leff JW, Owens SM, Pollard KS, Knight R, Gilbert JA, McCulley RL.** 2013. Reconstructing the microbial diversity and function of pre-agricultural tallgrass prairie soils in the United States. Science **342:**621–624. http://dx.doi.org/10.1126/science.1243768.
48. **Pehrsson EC, Forsberg KJ, Gibson MK, Ahmadi S, Dantas G.** 2013. Novel resistance functions uncovered using functional metagenomic investigations of resistance reservoirs. Front Microbiol **4:**145. http://dx.doi.org/10.3389/fmicb.2013.00145.
49. **Gibson MK, Forsberg KJ, Dantas G.** 2015. Improved annotations of antibiotic resistance functions reveals microbial resistomes cluster by ecology. ISME J **9:**207–216. http://dx.doi.org/10.1038/ismej.2014.106.
50. **Delmont TO, Robe P, Cecillon S, Clark IM, Constancias F, Simonet P, Hirsch PR, Vogel TM.** 2011. Accessing the soil metagenome for studies of microbial diversity. Appl Environ Microbiol **77:**1315–1324. http://dx.doi.org/10.1128/AEM.01526-10.
51. **Delmont TO, Prestat E, Keegan KP, Faubladier M, Robe P, Clark IM, Pelletier E, Hirsch PR, Meyer F, Gilbert JA, Le Paslier D, Simonet P, Vogel TM.** 2012. Structure, fluctuation and magnitude of a natural grassland soil metagenome. ISME J **6:**1677–1687. http://dx.doi.org/10.1038/ismej.2011.197.
52. **MacLean MJ, Ness LS, Ferguson GP, Booth IR.** 1998. The role of glyoxalase I in the detoxification of methylglyoxal and in the activation of the KefB K^+ efflux system in *Escherichia coli*. Mol Microbiol **27:**563–571. http://dx.doi.org/10.1046/j.1365-2958.1998.00701.x.
53. **Zhu MM, Skraly FA, Cameron DC.** 2001. Accumulation of methylglyoxal in anaerobically grown *Escherichia coli* and its detoxification by expression of the *Pseudomonas putida* glyoxalase I gene. Metab Eng **3:**218–225. http://dx.doi.org/10.1006/mben.2001.0186.
54. **Thoden JB, Frey PA, Holden HM.** 1996. High-resolution X-ray structure of UDP-galactose 4-epimerase complexed with UDP-phenol. Protein Sci **5:**2149–2161. http://dx.doi.org/10.1002/pro.5560051102.
55. **Urbaniak MD, Tabudravu JN, Msaki A, Matera KM, Brenk R, Jaspars M, Ferguson MA.** 2006. Identification of novel inhibitors of UDP-Glc 4′-epimerase, a validated drug target for African sleeping sickness. Bioorg Med Chem Lett **16:**5744–5747. http://dx.doi.org/10.1016/j.bmcl.2006.08.091.
56. **Nikaido H.** 1961. Galactose-sensitive mutants of Salmonella. I. Metabolism of galactose. Biochim Biophys Acta **48:**460–469. http://dx.doi.org/10.1016/0006-3002(61)90044-0.
57. **Yarmolinsky MB, Wiesmeyer H, Kalckar HM, Jordan E.** 1959. Hereditary defects in galactose metabolism in *Escherichia coli* mutants. II. Galactose-induced sensitivity. Proc Natl Acad Sci U S A **45:**1786–1791. http://dx.doi.org/10.1073/pnas.45.12.1786.
58. **Mori T, Suenaga H, Miyazaki K.** 2008. A metagenomic approach to the identification of UDP-glucose 4-epimerase as a menadione resistance pro-

tein. Biosci Biotechnol Biochem **72:**1611–1614. http://dx.doi.org/10.1271/bbb.70815.

59. **Zheng H, Wang X, Yomano LP, Shanmugam KT, Ingram LO.** 2012. Increase in furfural tolerance in ethanologenic *Escherichia coli* LY180 by plasmid-based expression of *thyA*. Appl Environ Microbiol **78:**4346–4352. http://dx.doi.org/10.1128/AEM.00356-12.
60. **Myllykallio H, Lipowski G, Leduc D, Filee J, Forterre P, Liebl U.** 2002. An alternative flavin-dependent mechanism for thymidylate synthesis. Science **297:**105–107. http://dx.doi.org/10.1126/science.1072113.
61. **Gu W, Li X, Huang J, Duan Y, Meng Z, Zhang KQ, Yang J.** 2011. Cloning, sequencing, and overexpression in *Escherichia coli* of the *Enterobacter* sp. Px6-4 gene for ferulic acid decarboxylase. Appl Microbiol Biotechnol **89:**1797–1805. http://dx.doi.org/10.1007/s00253-010-2978-4.
62. **Licandro-Seraut H, Roussel C, Perpetuini G, Gervais P, Cavin JF.** 2013. Sensitivity to vinyl phenol derivatives produced by phenolic acid decarboxylase activity in *Escherichia coli* and several food-borne Gram-negative species. Appl Microbiol Biotechnol **97:**7853–7864. http://dx.doi.org/10.1007/s00253-013-5072-x.
63. **Timson DJ, Reece RJ.** 2003. Functional analysis of disease-causing mutations in human galactokinase. Eur J Biochem **270:**1767–1774. http://dx.doi.org/10.1046/j.1432-1033.2003.03538.x.
64. **Platt A, Ross HC, Hankin S, Reece RJ.** 2000. The insertion of two amino acids into a transcriptional inducer converts it into a galactokinase. Proc Natl Acad Sci U S A **97:**3154–3159. http://dx.doi.org/10.1073/pnas.97.7.3154.
65. **Hartley A, Glynn SE, Barynin V, Baker PJ, Sedelnikova SE, Verhees C, de Geus D, van der Oost J, Timson DJ, Reece RJ, Rice DW.** 2004. Substrate specificity and mechanism from the structure of *Pyrococcus furiosus* galactokinase. J Mol Biol **337:**387–398. http://dx.doi.org/10.1016/j.jmb.2004.01.043.
66. **Xu J, Banerjee A, Pan SH, Li ZJ.** 2012. Galactose can be an inducer for production of therapeutic proteins by auto-induction using *E. coli* BL21 strains. Protein Expr Purif **83:**30–36. http://dx.doi.org/10.1016/j.pep.2012.02.014.
67. **Mumma JO, Chhay JS, Ross KL, Eaton JS, Newell-Litwa KA, Fridovich-Keil JL.** 2008. Distinct roles of galactose-1P in galactose-mediated growth arrest of yeast deficient in galactose-1P uridylyltransferase (GALT) and UDP-galactose 4′-epimerase (GALE). Mol Genet Metab **93:**160–171. http://dx.doi.org/10.1016/j.ymgme.2007.09.012.
68. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119. http://dx.doi.org/10.1186/1471-2105-11-119.
69. **Ross KL, Davis CN, Fridovich-Keil JL.** 2004. Differential roles of the Leloir pathway enzymes and metabolites in defining galactose sensitivity in yeast. Mol Genet Metab **83:**103–116. http://dx.doi.org/10.1016/j.ymgme.2004.07.005.
70. **Miller EN, Jarboe LR, Yomano LP, York SW, Shanmugam KT, Ingram LO.** 2009. Silencing of NADPH-dependent oxidoreductase genes (*yqhD* and *dkgA*) in furfural-resistant ethanologenic *Escherichia coli*. Appl Environ Microbiol **75:**4315–4323. http://dx.doi.org/10.1128/AEM.00567-09.
71. **Graziani S, Bernauer J, Skouloubris S, Graille M, Zhou CZ, Marchand C, Decottignies P, van Tilbeurgh H, Myllykallio H, Liebl U.** 2006. Catalytic mechanism and structure of viral flavin-dependent thymidylate synthase ThyX. J Biol Chem **281:**24048–24057. http://dx.doi.org/10.1074/jbc.M600745200.
72. **Griffin J, Roshick C, Iliffe-Lee E, McClarty G.** 2005. Catalytic mechanism of *Chlamydia trachomatis* flavin-dependent thymidylate synthase. J Biol Chem **280:**5456–5467. http://dx.doi.org/10.1074/jbc.M412415200.
73. **Glebes TY, Sandoval NR, Reeder PJ, Schilling KD, Zhang M, Gill RT.** 2014. Genome-wide mapping of furfural tolerance genes in *Escherichia coli*. PLoS One **9:**e87540. http://dx.doi.org/10.1371/journal.pone.0087540.
74. **Wang X, Yomano LP, Lee JY, York SW, Zheng H, Mullinnix MT, Shanmugam KT, Ingram LO.** 2013. Engineering furfural tolerance in *Escherichia coli* improves the fermentation of lignocellulosic sugars into renewable chemicals. Proc Natl Acad Sci U S A **110:**4021–4026. http://dx.doi.org/10.1073/pnas.1217958110.
75. **Hadi SM, Shahabuddin, Rehman A.** 1989. Specificity of the interaction of furfural with DNA. Mutat Res **225:**101–106. http://dx.doi.org/10.1016/0165-7992(89)90125-5.
76. **Khan QA, Hadi SM.** 1993. Effect of furfural on plasmid DNA. Biochem Mol Biol Int **29:**1153–1160.
77. **Allen SA, Clark W, McCaffery JM, Cai Z, Lanctot A, Slininger PJ, Liu ZL, Gorsich SW.** 2010. Furfural induces reactive oxygen species accumulation and cellular damage in *Saccharomyces cerevisiae*. Biotechnol Biofuels **3:**2. http://dx.doi.org/10.1186/1754-6834-3-2.
78. **Tilman D, Hill J, Lehman C.** 2006. Carbon-negative biofuels from low-input high-diversity grassland biomass. Science **314:**1598–1600. http://dx.doi.org/10.1126/science.1133306.