

Coexisting/Coexpressing Genomic Libraries (CoGeL) identify interactions among distantly located genetic loci for developing complex microbial phenotypes

Sergios A. Nicolaou, Stefan M. Gaida and Eleftherios T. Papoutsakis*

Department of Chemical & Biomolecular Engineering, Molecular Biotechnology Laboratory, Delaware Biotechnology Institute, University of Delaware, Newark, DE 19711, USA

Received August 17, 2011; Revised September 14, 2011; Accepted September 15, 2011

ABSTRACT

In engineering novel microbial strains for biotechnological applications, beyond *a priori* identifiable pathways to be engineered, it is becoming increasingly important to develop complex, ill-defined cellular phenotypes. One approach is to screen genomic or metagenomic libraries to identify genes imparting desirable phenotypes, such as tolerance to stressors or novel catabolic programs. Such libraries are limited by their inability to identify interactions among distant genetic loci. To solve this problem, we constructed plasmid- and fosmid-based *Escherichia coli* Coexisting/Coexpressing Genomic Libraries (CoGeLs). As a proof of principle, four sets of two genes of the L-lysine biosynthesis pathway distantly located on the *E. coli* chromosome were knocked out. Upon transformation of these auxotrophs with CoGeLs, cells growing without supplementation were found to harbor library inserts containing the knocked-out genes demonstrating the interaction between the two libraries. CoGeLs were also screened to identify genetic loci that work synergistically to create the considerably more complex acid-tolerance phenotype. CoGeL screening identified combination of genes known to enhance acid tolerance (*gadBC* operon and *adiC*), but also identified the novel combination of *arcZ* and *recA* that greatly enhanced acid tolerance by 9000-fold. *arcZ* is a small RNA that we show increases pH tolerance alone and together with *recA*.

INTRODUCTION

Complex phenotypes arise in cells from interactions among genes, pathways, signaling events and cellular

programs under changing environmental conditions. Many important cellular traits are the result of such complex interactions, and they typically involve many genes, which in most cases are not precisely known. Identifying such complex gene interactions would enable the construction of superior strains and cellular programs for biotechnological applications. Tolerance to toxic chemicals is an important example of a complex phenotype. It is also an important bioprocessing trait that has been long sought after but with variable success (1). For example, solvent tolerance is the result of several simultaneous mechanisms of action including molecular pumps, changes in membrane properties, changes in cell-wall composition and altered energy metabolism (1). Each mechanism may engage many genes, and may act alone or in conjunction with other mechanisms or programs. Frequently, the genes involved in such mechanisms are dispersed on the chromosome thus hindering identification with conventional genomic-library screening (1,2). Furthermore, a phenotypic improvement can be achieved by different genomic coordinates that are not always compatible, which makes searching the total genome challenging, and can be due to pre- or post-translational effects (3,4).

Identification and/or development of desirable microbial phenotypes has been pursued by various strategies (1), including the use of genomic (5,6) and metagenomic libraries (7,8), whole-genome shuffling (9,10), biosynthetic pathway engineering (4) and recombineering approaches, such as multiplex automated genome engineering (MAGE) and trackable multiplex recombineering (TRMR) (11,12).

Genomic libraries allow for the screening of an entire genome (or a collection of genomes) by digesting genomic DNA (gDNA), cloning it into vectors and transforming these into cells that can be screened for a desired phenotype (5,13–15). Library-insert carrying cells are exposed to selective pressure with the assumption that some gene(s) represented in the library will allow enrichment under

*To whom correspondence should be addressed. Tel: +1 302 831 8376; Fax: +1 302 831 4841; Email: epaps@udel.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

pressure. The population is eventually taken over by cells carrying vectors with these library insert(s), and the trait-conferring gene(s) or genetic loci can be identified by sequencing or DNA-microarray analysis (13–15). Similarly, screening can be carried out using metagenomic libraries (8,16), but in these cases only single genetic loci made up of one or a few adjacent genes can be identified (17).

Standard, i.e. single genomic or metagenomic libraries cannot capture interactions among distantly located loci on a chromosome (and/or multiple chromosomes for metagenomic libraries) necessary to create or improve a complex phenotype. This derives from the fact that each library cell contains one type of library vector and thus one DNA fragment. For plasmid-borne libraries, the insert size can be up to 6–8 kb of DNA or in the case of fosmid libraries ~35 kb of contiguous DNA. Employment of large-insert libraries, like fosmids or bacterial artificial chromosomes (BACs) [with insert sizes of exceeding 35 kb (18)] does not effectively overcome this limitation because of the poor expression of heterologous genes in the host organism (19,20).

This inability to have multiple DNA fragments (library inserts) in a single cell (clone) combined with the DNA-fragment size limitation constrains the combinatorial genomic space that can be sampled and hinders the identification of beneficial interactions among distantly-located genetic loci. To overcome these limitations, we propose and demonstrate the use of the Coexisting/Coexpressing Genomic Libraries (CoGeLs) (Figure 1). CoGeLs enable two (and more, but practically only a few) genomic (and/or metagenomic) libraries to coexist in one cell thus allowing to screen for necessary or cooperative gene interactions in the development or improvement of a screenable phenotype. Genomic fragments normally distantly located in a genome or multiple genomes can be expressed together in a single cell and screened for beneficial interactions. For small-insert plasmid-based libraries, the number of binary (let alone trinary) combinations of DNA-fragment inserts to be screened is very large because even single-insert libraries require a large number of individual clones to achieve a desired genome-coverage probability (discussed below). This limitation can be overcome by utilizing a fosmid library (~35 kb insert size) in combination with a coexisting plasmid library, and/or by using enriched libraries. Thus, the number of individual CoGeL clones necessary for a desirable genome-coverage probability is reduced by one order of magnitude.

The stable maintenance and interactions among two CoGeLs was demonstrated first in a proof-of-concept study in a well-defined genetic background. We constructed double knockouts (dKOs; resulting in auxotrophic strains) in the L-lysine biosynthesis pathway using genes which are distantly located on the *E. coli* chromosome. Introduction of *E. coli* CoGeLs reinstated the knocked out genes and recovered the wild-type (WT) phenotype. Distantly located genes were carefully selected for knock out to ensure that a dKO cannot be repaired with a single library insert. To demonstrate that the CoGeL technology can be applied to generate a useful, but largely genetically

undefined, complex phenotype, we searched for and identified genetic loci that impart low-pH tolerance in *E. coli*. Among several acid-resistance systems (AR) described in the literature (21), four have been well characterized. AR1 is known as the oxidative or glucose-repressed system (22) and requires the alternative sigma factor RpoS and CRP (cyclic AMP receptor protein), AR2 is a glutamate dependent (23), AR3 is arginine dependent (24) and AR4 is lysine dependent (25). Here, we were able to identify some loci linked to these established AR mechanisms, but also some potentially novel determinants.

MATERIALS AND METHODS

Media, strains and plasmids

Strains, plasmids and media compositions are listed in the Supplementary Methods section. DNA isolation and manipulation are described in the Supplementary Methods section.

Plasmid library construction

Sheared gDNA was blunt-ended, dephosphorylated and adenylated (see Supplementary Methods section). The final gDNA was used in a TOPO reaction with the pCR[®]8/GW/TOPO[®] TA Cloning[®] kit (Invitrogen) to construct an entry vector library for the Gateway[®] System (Invitrogen). Plasmids were transformed in TOP10 cells (Invitrogen) to obtain $\sim 1.9 \times 10^5$ clones. This entry-clone library was designated as **1MgL**. Sequencing of 20 clones yielded an average insert size of ~ 3.5 kb.

The 1MgL library was shuttled to the destination vector pDEST[™]14 (Invitrogen) using the LR recombination reaction, as per the manufacturer's instructions. The recombinated plasmids were transformed in TOP10 cells, to obtain a total of $\sim 3.4 \times 10^4$ clones and designated as **2MgL** library. Similarly, the 1MgL library was also shuffled in the pACYCdest plasmid (see Supplementary Methods section) to create the **4MgL** library, yielding 2.25×10^4 clones.

2MgL has the colE1 (pBR322) origin of replication (ori) and the ampicillin resistance gene. 4MgL has the p15A ori and contains the tetracycline resistance gene. Thus, the two libraries can coexist in a single host cell.

Fosmid library construction

The CopyControl[™] Fosmid Library Production Kit (Epicentre) was utilized as per the manufacturer's instructions. A total of 50 μ g *E. coli* K-12 substr. MG1655 (MG1655) gDNA was sheared by passage through a 50 μ l microsyringe 30 times. The **FosMg** library generated consists of 1800 clones with 35-kb inserts.

CoGeL expression in wild-type *E. coli* str. K-12 substr. MG1655

Because the 2MgL and 4MgL plasmid libraries were constructed in a restriction and methylation negative cloning

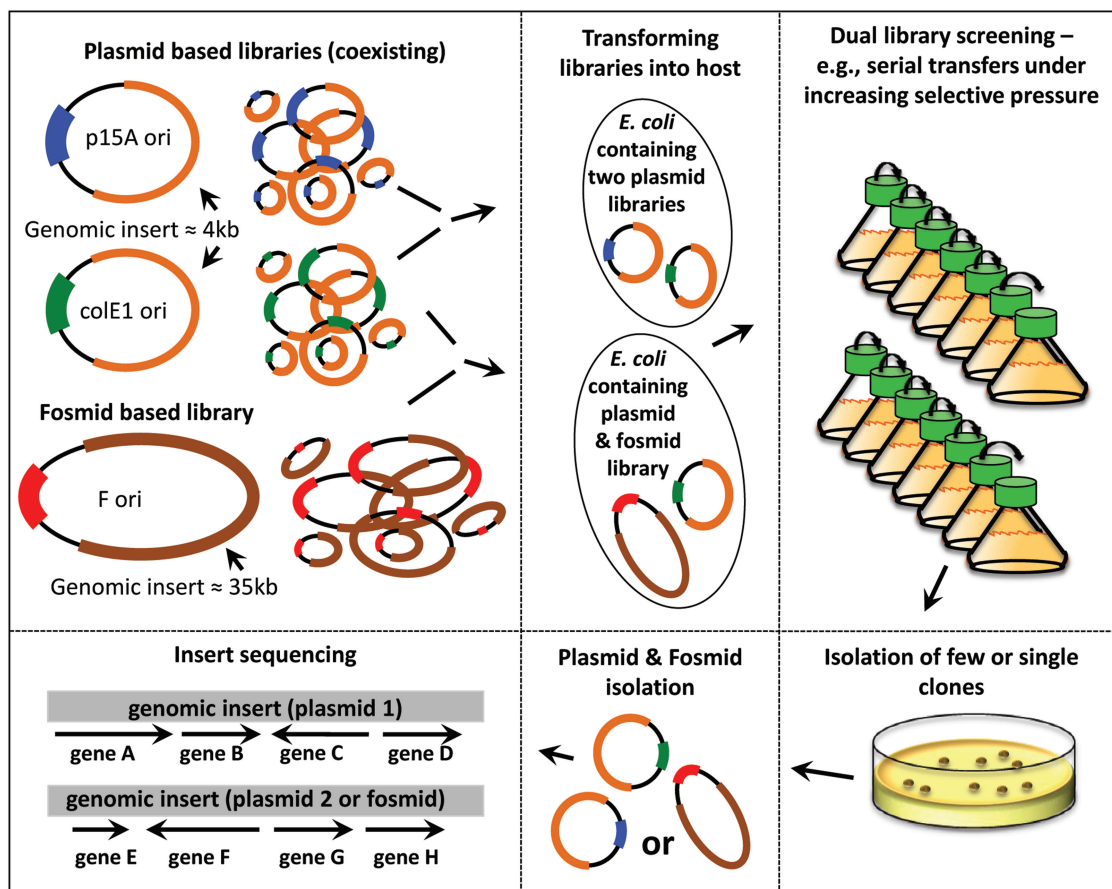


Figure 1. The CoGeL technology. Multiple libraries with compatible origins of replication (blue: p15A ori, green: colE1 ori, red: F ori) and different insert sizes (plasmid ≈ 3.5 kb, fosmid ≈ 35 kb) are constructed and transformed in a desired host (dual plasmid: blue and green, plasmid-fosmid combination: red and green). Cells containing two CoGeLs are screened for a specific phenotype, here shown as serial transfers under selective pressure, e.g. increasing concentrations of a toxic chemical (stressant) when selecting for a resistance phenotype. After enrichment, single clones are isolated by plating and non-chromosomal DNA is extracted. Genes on the selected clones can be identified by sequencing library inserts.

strain (TOP10), they would be digested in the WT strain MG1655. So the libraries were introduced into a restriction negative but methylation positive cloning strain, namely NEB-5alpha (New England Biolabs), to reinstate the correct methylation pattern. The size of the methylated library 2MgL was 1.2×10^7 , and 8×10^6 clones for the methylated 4MgL library in the NEB-5alpha cells. Methylated plasmid DNA was then used to transform MG1655. First, the methylated 4MgL library was transformed yielding a total size of about 1×10^6 clones. This library was made electrocompetent again and transformed with methylated 2MgL. A total of 2.6×10^8 clones were obtained for the (dual) library containing both 2MgL and 4MgL.

Disrupting the lysine pathway to generate double gene-knockout *E. coli* strains

To disrupt the metabolic pathway of the lysine biosynthesis, double gene knockouts (dKOs) were constructed using the inactivation method via linear PCR products (26) (Supplementary Tables S1 and S2). The dKOs were

confirmed by PCR and for auxotrophy, by growth in non-supplemented minimal M9 media.

Acid-resistance assay

Cells were grown in 10 ml LB media containing 10 μ g/ml tetracycline and 100 μ g/ml ampicillin to an OD_{600} between 1 and 1.3, and then transferred in a 1:10 ratio to acidified minimal M9 media (adjusted with HCl to pH 2) supplemented with 1.5 mM glutamate. Cells were exposed for 1 h in pre-warmed M9 media at 37°C under shaking (220 rpm). Serial dilutions were plated on LB plates containing antibiotics before and after the passage through the acidified media. Survival was determined as (% survival) = $[(CFU/ml)_{\text{post-stress}} / (CFU/ml)_{\text{initial}}] * 100$. As control, the survival of the strain containing the control plasmids pC1 and pC2 (see below) was determined. Increased acid resistance was defined as higher survival than this control strain.

Selection of acid tolerant phenotypes

MG1655 containing the two libraries 2MgL + 4MgL was outgrown in 100 ml LB media (inoculated with 2 ml 15%

glycerol frozen stock) containing antibiotics to exponential growth phase (OD₆₀₀ of 1). Two milliliters of culture was then diluted 1:10 in pH 2 adjusted pre-warmed M9 media (18 ml) supplemented with 1.5 mM glutamate. After exposure for 1 h, serial dilutions of the stressed library were plated on LB agar plates containing antibiotics. From these plates, individual clones were picked and verified for acid tolerance with the AR assay. Plasmid DNA was extracted from clones showing an increased survival and used to transform MG1655 cells never exposed to acids before. The retransformed plasmid combinations of the selected clones were again checked for acid tolerance via the AR assay and clones with an increased acid survival were identified as clones with a true acid tolerant phenotypes originating from the selected plasmid combination.

Construction of control strains for characterization of acid resistance

Control plasmids were constructed by an LR recombination reaction of the library destination plasmid (pDESTTM14 and pACYCdest) with a control entry plasmid pENTRTM_{gus} provided by the manufacture (Invitrogen). The entry plasmid carries the promoterless *gus* gene from *Arabidopsis thaliana* which codes for a β-glucuronidase. Since neither destination plasmid contains a promoter recognized by MG1655, transcription from this insert is impossible. The pDESTTM14 based control plasmid was named pC1, and the pACYCdest-based control plasmid pC2. MG1655 was transformed with control plasmids pC1 + pC2 and used as the control strain for assessing acid resistance. The survival of this strain was used as a baseline and only clones with higher survival were selected. After identification of true acid

tolerant clones (see above), another two control strains were constructed. First, the plasmids of the selected clone were separated by plating transformants on either ampicillin or tetracycline containing plates. Clones from these plates, which were sensitive towards the other antibiotic, were transformed with a compatible control plasmid. Thus, strains containing one plasmid of the selected plasmid combination and the compatible control plasmid were generated and used as additional control strains in characterization of the combined impact of selected plasmid combinations (Table S3).

Statistical analysis

Minitab16 (27) was used to perform statistical analysis. The difference in acid survival of the individual strains was tested for significance by applying one-sided two-sample *t*-test. The prerequisite of normality was tested by applying an Anderson–Darling test to the individual data sets.

RESULTS

Construction of genomic libraries that can coexist in a single host cell

A starting genomic library was constructed in the pCR[®]8/GW/TOPO[®] vector using 3–6 kb sheared *E. coli* str. K-12 substr. MG1655 gDNA. The library contains approximately 190 000 clones and was designated 1MgL (Table 1). This entry library was recombined into two destination vectors, the commercially available pDESTTM14 (Invitrogen) and the constructed plasmid pACYCdest, to create two destination libraries, termed 2MgL and 4MgL, respectively, comprising of 34 000 and 22 500 clones. Both

Table 1. CoGeL libraries and strains used to coexpress them

Library	Vector type/Name/Reference or Source of plasmid	Origin of replication	Average library insert size (kb)	Library size (number of clones)	Antibiotic marker	<i>E. coli</i> hosts ^a
1MgL	Plasmid/ pCR [®] 8/GW/TOPO [®] / Invitrogen ^b	colE1 (pBR322) ori	3.5	190 000	Spectinomycin	A
2MgL	Plasmid/pDEST TM 14/ Invitrogen ^b	colE1 (pBR322) ori	3.5	34 000	Ampicillin	A, C, D, dKO1, dKO2, dKO3, dKO4
4MgL	Plasmid/pACYCdest/This study	p15A	3.5	22 500	Tetracycline	A, C, D, dKO1, dKO2
FosMg	Fosmid/ Copy Control TM pCC1Fos TM Vector/Epicentre ^c	F-factor single-copy origin of replication. Inducible oriV (high copy)	35	1800	Chloramphenicol	B, dKO3, dKO4

^aA, TOP10: F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ80*lacZ*ΔM15 Δ*lacX74* *recA1* *araD139* Δ(*ara-leu*)7697 *galU galK rpsL* (Str^R) *endA1 nupG* (*Invitrogen*); B, Epi300TM-T1^R Phage T1-resistant *E. coli*: F- *mcrA*.Δ(*mrr-hsdRMS-mcrBC*).(Str^R).φ80*lacZ*ΔM15 Δ*lacX74* *recA1 endA1 araD139* Δ(*ara-leu*)7697 *galU galK rpsL nupG trfA tonA dhfr* (Epicentre); C, *Escherichia coli* str. K-12 substr. MG1655 (WT); D, NEB 5 alpha competent *E. coli*: *shuA2Δ(argF-lacZ)U169 phoA glnV44* Φ80 Δ(*lacZ*)M15 *gyrA96 recA1 relA1 endA1 thi-1 hsdR17* (New England Biolabs, Ipswich, MA, USA); dKO1, *E. coli* K-12 MG1655 Δ*dapA*Δ*lysA* (constructed in this study); dKO2, *E. coli* K-12 MG1655 Δ*dapD*Δ*dapE* (constructed in this study); dKO3, Epi300TM-T1^R Δ*dapB*Δ*lysA* (constructed in this study); dKO4, Epi300TM-T1^R Δ*dapD*Δ*lysA* (constructed in this study).

^bInvitrogen, Life Technologies, Carlsbad, CA, USA.

^cEpicentre, Madison, WI, USA.

libraries are estimated to cover 99% of the *E. coli* genome with at least 3.5 times coverage. The 2MgL and 4MgL libraries have different origins of replication and antibiotic resistance markers and are thus compatible and can coexist in a single *E. coli* host (Table 1). A fosmid library of *E. coli* str. K-12 MG1655 gDNA was generated using the CopyControl™ fosmid library production kit (Epicentre). The fosmid library (FosMg) contains 35 kb inserts and is made up of 1800 clones. This allows for 3-fold coverage of the genome. The FosMg and 2MgL libraries are compatible and can also coexist in a single host cell.

Complementation of dKOs with CoGeLs recovers the disrupted lysine biosynthesis pathway and identifies the knocked out genes

To demonstrate the stable coexistence of and interactions between two CoGeLs, dKOs of the lysine biosynthesis pathway (Figure 2) were constructed and complemented with CoGeLs. WT *E. coli* K-12 MG1655 was used to

create two double knockouts, MG1655 $\Delta dapA \Delta lysA$ and MG1655 $\Delta dapD \Delta dapE$ (see Figure 2, Table 1 and 'Materials and Methods' section). These pairs of genes were selected because they are distantly located on the chromosome, and thus cannot both be included on a single library insert, and because they lead to lysine and meso-diaminopimelate (meso-DAP) auxotrophy. The two dKO strains were confirmed by PCR and for auxotrophy by cultivation in defined M9 minimal media with and without supplementation with meso-DAP (100 $\mu\text{g}/\text{ml}$) and L-lysine (10 $\mu\text{g}/\text{ml}$). Inability of the dKO strains to grow in non-supplemented media verified their auxotrophy and the need for two-gene complementation. Each of the two verified dKO strains was sequentially electrotransformed with two plasmid CoGeLs, 2MgL and 4MgL (Table 1). Transformants were grown in LB media containing meso-DAP, lysine and the antibiotics ampicillin and tetracycline (for library maintenance). Transformation of MG1655 $\Delta dapA \Delta lysA$ with the two CoGeLs yielded 8×10^7 clones. Similarly, 6×10^6 clones

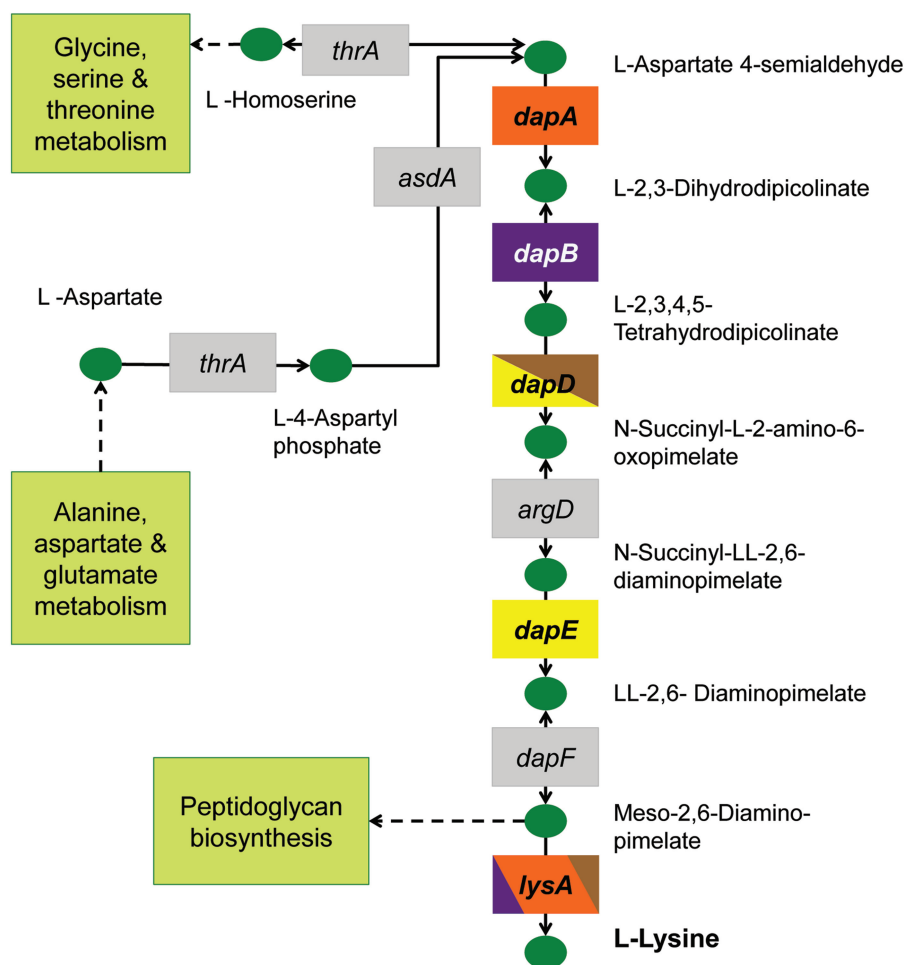


Figure 2. Construction of multiple dKOs. The lysine biosynthesis pathway in *E. coli* [based on KEGG (40)] was disrupted by dKOs to generate four auxotrophs. *Escherichia coli* K-12 MG1655 was used to generate the knockouts M1655 $\Delta dapA \Delta lysA$ (orange), designated **dKO1**, and MG1655 $\Delta dapD \Delta dapE$ (yellow), designated **dKO2**, as described in 'Materials and Methods' section. Similarly, *E. coli* strain Epi300 T1^R $\Delta dapB \Delta lysA$ (purple) and Epi300 T1^R $\Delta dapD \Delta lysA$ (brown), designated **dKO3** and **dKO4**, respectively. dKOs were confirmed by PCR and for lysine auxotrophy. See 'Materials and Methods' section for details on the generation and confirmation of all gene knockouts.

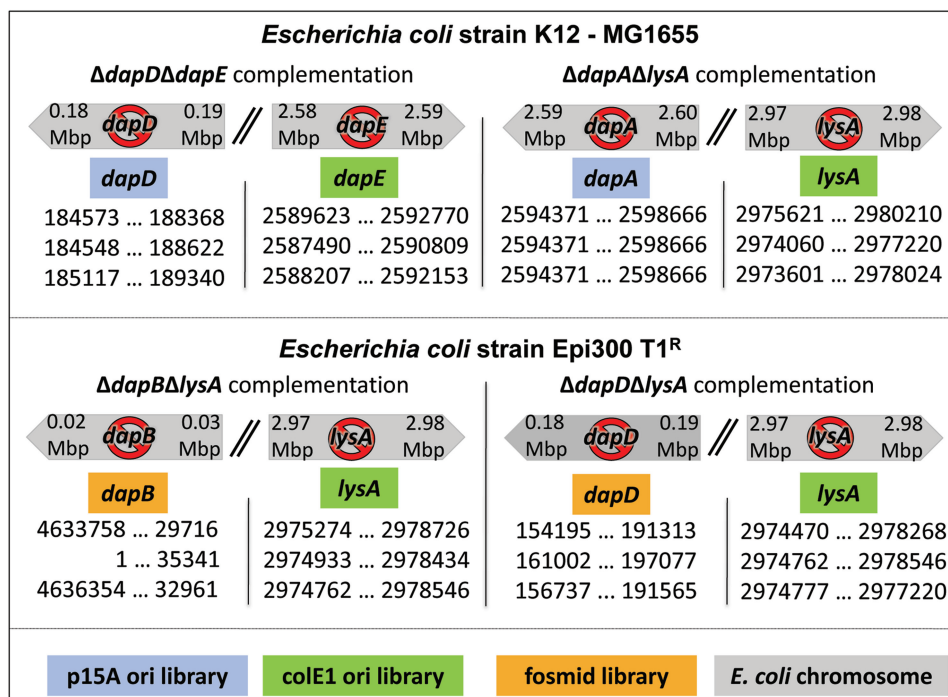


Figure 3. Complementation of dKOs with plasmid and fosmid CoGeLs. The approximate chromosomal position of the knocked out genes in the *E. coli* chromosome are shown in grey for each knockout combination and the genomic regions isolated from three clones of the complementation CoGeLs are listed below each knockout. In the upper part of the figure the complementation of the dKOs in strain *E. coli* K-12 MG1655, Δ dapD Δ dapE and Δ dapA Δ lysA, are shown, which were complemented with two plasmid libraries, namely 4MgL (blue, contains a p15A origin of replication) and 2MgL (green, contains a ColE1 origin of replication). In the lower part, the complementation of the dKOs in *E. coli* Epi300T1^R, Δ dapD Δ lysA and Δ dapB Δ lysA, are shown, which were complemented with a fosmid library (orange, F-plasmid origin of replication) and a plasmid library (2MgL, green, ColE1 origin of replication). The genomic coordinates from three independent clones were obtained by isolating the library vectors, sequencing the inserts, and using BLAST to obtain the specific chromosomal location of the DNA contained in the inserts.

were obtained for the CoGeL-transformed MG1655 Δ dapD Δ dapE. In biological transformation replicates, the number of transformants was 1×10^8 for both dKO strains.

The two different MG1655 dKO strains complemented with the two CoGeLs were washed twice with M9 minimal media to remove added supplements. Washed cells were concentrated 100-fold and were plated on M9 plates without supplementation. We hypothesized that without supplementations only clones carrying library plasmids that contain the knocked out genes could produce viable clones. Colonies were collected after 48 h and grown in liquid M9 media to ensure recovery of the phenotype. Plasmids from the strains exhibiting a recovered phenotype were isolated, and the inserts were sequenced for each strain and plasmid (Figure 3). All sequenced inserts show that one library insert contained one knocked out gene and the other the second knocked out gene. These data show that two libraries ‘interacted’ to provide the necessary genes to complete the disrupted pathway and restore the phenotype. No single insert alone was sufficient to complement the dKO phenotype.

Larger library inserts contained in fosmid libraries (~35 kb) span longer positions of the chromosome and allow for a larger combinatorial genomic space when coupled with a plasmid library, as compared to two plasmid libraries that are smaller. Thus, we next

demonstrated the successful employment of a fosmid library in combination with a plasmid library. The phage resistant *E. coli* Epi300 T1^R cells were used to construct two dKO strains: Epi300 T1^R Δ dapD Δ lysA and Epi300 T1^R Δ dapB Δ lysA (Figure 2 and Table 1). After strain verification as above, each of the two dKO strains was infected with a packaged fosmid library containing *E. coli* K-12 MG1655 DNA, was then made electrocompetent and transformed with the 2MgL plasmid library. The first replicate yielded 8×10^6 and 3×10^6 clones for the Epi300 T1^R Δ dapA Δ lysA and Epi300 T1^R Δ dapB Δ lysA strains, respectively. The second biological replicate produced more than 1×10^7 transformants for each strain. Transformants for the Epi300 T1^R dKO strains were grown in LB medium with ampicillin and chloramphenicol, as well as the required supplementation of lysine and meso-DAP to an OD₆₀₀ of ~1 (early exponential phase). They were then washed with M9 medium and plated on plates of modified M9 medium as follows. The parent strain Epi300 T1^R requires leucine (100 µg/ml) and thiamine (10 µg/ml) to grow on minimal media, and thus these two supplements were added to the M9 medium to prepare plates. After 48 h of growth, colonies from the modified M9 plates were grown in liquid M9 medium supplemented with antibiotics, leucine and thiamine. Cultures were transferred twice in fresh media to ensure

that the strains could grow without lysine or meso-DAP supplementation. The selected strains were then transferred to LB media in order to achieve higher cell densities, and the library vectors – plasmids and fosmids – contained in these cells were isolated (see ‘Materials and Methods’ section). Inserts from the isolated fosmids and plasmids were sequenced to identify the genes on the isolated library clones (Figure 3). Fosmids and plasmids with different genetic inserts were isolated. The fosmid inserts always contained the *dapB* or *dapD* genes, while the plasmid inserts always contained the *lysA*. These data suggest that the cells preferentially selected a high copy plasmid for the *lysA* complementation, and the single copy fosmid for the other gene.

Identifying genetic determinants for acid (low-pH) tolerance

We then aimed to examine the use of CoGeLs in identifying genetic determinants imparting a complex, less defined phenotype. We examined several tolerance phenotypes (to low pH, ethanol and oxidative stress), and here we report data on the acid-tolerant phenotype. As discussed, several mechanisms of *E. coli* tolerance to low pH have been reported. Among those, the most extensively studied mechanism is the glutamate-dependent system, which can protect cells against a pH as low as 2 (22). The mechanism involves two parts, a glutamate-decarboxylase (coded as two isoforms by *gadA* and *gadB*) and a glutamate:γ-aminobutyric acid (GABA) antiporter (coded by *gadC*). Extracellular glutamate is transported into the cells and converted to GABA, whereby one proton is consumed by the reaction. GABA is then transported out of the cell in exchange for a glutamate, which results in a net loss of one proton from the cell, thus counteracting the decrease of intracellular pH (28). This system is acid inducible (pH 5) and fully active in stationary phase (22). Based on this work, we adapted an assay for screening CoGeL libraries and assessing tolerance to pH (see ‘Materials and Methods’ section). The idea behind the assay is that it should make possible to identify via screening of CoGeLs genes/genetic loci both related to this glutamate-dependent system, as well as genes not related to this system, and especially mechanisms that work during active, exponential cell growth. Screening of a CoGeL library [methylated (see ‘Materials and Methods’ section) 2MgL/4MgL; Table 1] in acidified

(pH = 2) M9 media supplemented with 1.5 mM glutamate resulted in the identification of several clones with increased pH tolerance. After exposing the CoGeL library for 1 h in pH 2 acidified media, aliquots were plated and surviving clones were selected from these plates. These clones were first re-checked for acid tolerance by exposing exponential-phase cells to pH 2 acidified media. As described in ‘Materials and Methods’ section, plasmid DNA was extracted from these clones, the plasmids were separated, and then used to transform *E. coli* K12 MG1655 cells never exposed to low pH before. This was done as a precautionary step because in the acidic screening environment, the mutation rate is increased (29), potentially leading to acid tolerance originating from unknown chromosomal mutations. The retransformed plasmid combination in fresh WT genetic background was again checked for acid tolerance by exposing exponential-phase cells to pH-2 acidified media. Sequencing of library inserts identified several pairs of genetic determinants, and those contained whole annotated genes are listed in Table 2.

Among the identified pairs of library inserts, several contained the *adiC* gene and one pair contained the *adiC* gene on one plasmid and the *gadBC* operon on the other plasmid. Both genetic determinants are directly associated with acid resistance systems in *E. coli*, either the arginine (*adiC*) (24) or the glutamate-dependent (*gadBC*) (23) AR systems. Another isolated pair of clones contained genes previously not directly associated with acid resistance, namely the *arcZ* gene (coding for a small RNA) on one plasmid, and *recA*, on the other coexist plasmid. Since the two genes have not been previously associated with resistance to low pH, we carried out experiments to characterize the combined impact of these two genes, each on a separate plasmid (*arcZ* on pI1 and *recA* on pI2) co-transformed in fresh WT cells. We also generated three control strains: one carrying the two control (no insert) plasmids (Figure 4; one for each of the two plasmid libraries (Table 1): pC1 for pI1, and pC2 for pI2), the second strain carrying pI1+pC2, and the third strain carrying pI2+pC1. The four strains were then tested for acid tolerance by exposing them for 1 h to pH 2 (see ‘Materials and Methods’ section) and the data are shown in Figure 4. Individually, overexpression of either *arcZ* or *recA* increased cell survival at low pH (compare the single-control strains pI1+pC2 and pI2+pC1 against the pC1+pC2 strain), but *recA* alone had a small beneficial

Table 2. Sequencing results of clones selected after screening the methylated CoGeL libraries 2MgL/4MgL in acidified media (pH 2)

Isolated clone	colE1 ori based plasmid library			p15A ori based plasmid library		
	Insert		Complete genes present on the insert	Insert		Complete genes present on the insert
	Start	End		Start	End	
# 1	1 406 172	1 408 811	<i>fnrS</i>	2 823 110	2 820 296	<i>recA, ygaD</i>
# 2	4 336 034	4 331 323	<i>eptA, adiC, adiY</i>	4 141 040	4 135 764	<i>gldA, fsaB, ptsA</i>
# 3	4 336 034	4 331 323	<i>eptA, adiC, adiY</i>	968 670	966 649	<i>lpxK</i>
# 4	3 348 467	3 352 408	<i>arcZ, arcB, yhcC</i>	2 818 605	2 821 867	<i>recX, recA</i>
# 5	4 336 034	4 331 323	<i>eptA, adiC, adiY</i>	1 565 705	1 570 771	<i>gadB, gadC</i>

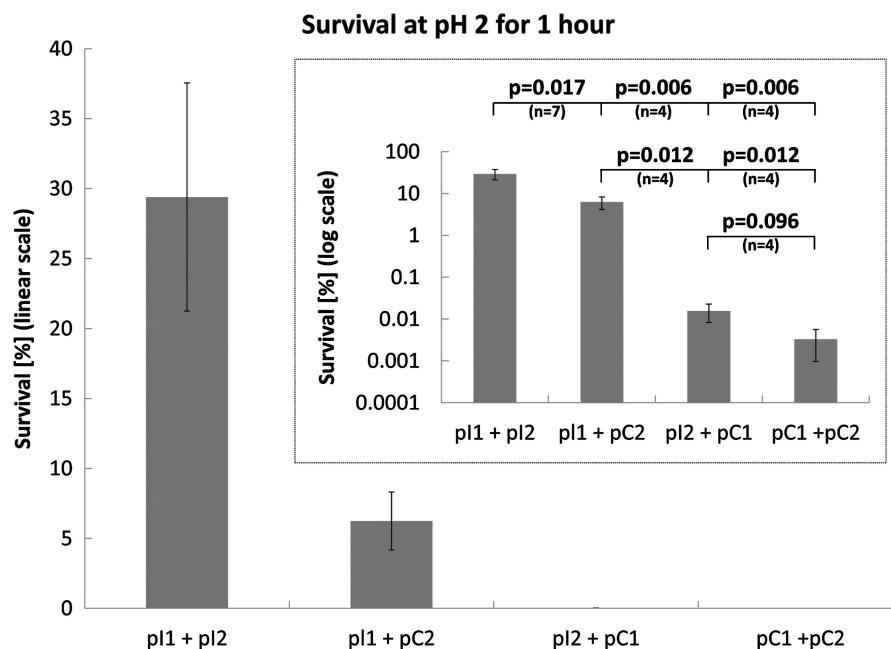


Figure 4. Superior acid tolerance of one identified CoGeL clone after strain reconstitution into unstressed WT cells. Survival of the isolated clone carrying the sRNA *arcZ* on the pDEST based plasmid (pI1) and *recA* in the pACYC based plasmid (pI2). The reconstituted strain is shown as (pI1 + pI2). Two single-plasmid control strains, each with one control plasmid containing the promoterless *gus* gene, are termed (pI1 + pC2) and (pI2 + pC1). The dual plasmid-control strain is termed (pC1 + pC2). Average percent is shown in linear scale to emphasize the higher survival of the reconstituted clone (pI1 + pI2) compared to the single control strain carrying the *arcZ* containing plasmid (pI1 + pC2). Insert: log scale plot showing the survival rates of the other strains that are not visible in the linear scale plot. Error bars indicate the standard error of the biological replicates ($n = 7$ or 4). Statistical significance for differences in survival between strains was examined with a one-sided two-sample *t*-test and the determined *P*-values are shown.

effect. On the other hand, *arcZ* by itself dramatically increased cell survival compared to the double-plasmid control. The combined overexpression of *arcZ* or *recA* (strain pI1+pI2) had a profound supra-additive effect: ~5-fold better survival than *arcZ* alone (strain pI1+pC2), 1900-fold better than *recA* alone (pI2+pC1) and 9000-fold better than the double plasmid control (Figure 4).

DISCUSSION

We demonstrated that plasmid- and/or fosmid-borne genomic libraries can be designed to coexist in cells aiming to facilitate the identification of pairs of genetic loci that, together, can develop or improve a complex phenotype. We first demonstrated the concept in the well-defined genetic background of four dKOs of the lysine biosynthesis pathway (Figure 2). All four dKO *E. coli* strains constructed were successfully complemented through either dual-plasmid CoGeLs, or via fosmid-plasmid CoGeLs to reinstate the non-auxotrophic phenotype. Each CoGeL vector provided one of the chromosomally distant located knocked out genes. This proves the concept and supports the hypothesis that plasmid-plasmid as well as fosmid-plasmid CoGeLs can be stably maintained in the cells and can interact to establish a desired phenotype.

Sequencing data revealed that the libraries contained multiple inserts spanning a similar region of the chromosome (Figure 3) as evidenced by the overlapping DNA

regions of the selected inserts. Thus, our CoGeLs provide good coverage of the genomic space. High genome coverage ensures that all genes are represented in the libraries and could provide the desirable beneficial interactions. The number of colonies (N) required to ensure a coverage probability (P) is based on the fraction (f) of the insert size relative to the entire genome, as follows: $N = \ln(1 - P)/\ln(1 - f)$ (30). For a plasmid library with 3.5 kb inserts, a total of approximately 6100 clones are required for 99% coverage. Our fosmid library requires approximately 610 clones. For a combination of two plasmids, an estimated 3.7×10^7 clones are required for a full coverage on each plasmid, assuming a perfectly separated library that has no duplicate clones. One of our initial transformants with both libraries, namely the MG1655 $\Delta dapD \Delta dapE$ strain, produced fewer than this number of clones. However, the others, and our replicates, yielded higher and satisfactory number of transformants. Using a fosmid library in tandem with the plasmid library decreases the total number of necessary clones to 3.7×10^6 . This lower benchmark is much easier to obtain in transformations, thus facilitating high coverage with a smaller number of clones. In our experiments we exceeded this figure in all but the complementation of the Epi300 T1^R $\Delta dapB \Delta lysA$ strain. In our biological replicates, the desired minimum number of clones was obtained for all strains.

To demonstrate the utility of CoGeLs for identifying and developing a more complex phenotype, we screened for acid tolerance thus identifying several combinations of

genetic loci (Table 2). While some of the inserts included genes previously associated with established AR mechanisms, others did not. We studied the combination of the *arcZ* and *recA* genes since *arcZ* has not been previously associated with acid tolerance. The WT strain harboring this gene combination exhibited dramatic increase in survival under stress during active growth (Figure 4). sRNA *arcZ* is a translational regulator of the alternative sigma factor *rpoS*, apparently functionally similar to the well-characterized sRNA *dsrA* (31), which is involved in acid resistance. *rpoS* is required for activation of AR1 and is indirectly also involved in regulation of AR3 by activating another regulator, *gadX* (32). A possible explanation of the effect of *arcZ* is based on its ability to activate the translation of *rpoS*, which triggers AR1 activation, thus leading to a higher acid resistance of actively growing cells. Regarding *recA*, it is known that low pH severely damages chromosomal DNA and that cells counteract such detrimental effects via DNA repair systems (29). RecA is involved in the SOS response of *E. coli* induced by DNA damage (33). Knockout of *recA* was reported to increase the *E. coli* acid sensitivity (29). Here, while plasmid expression of *recA* alone had only a small impact on enhancing acid tolerance, the combination of pI1 and pI2 increased the acid tolerance in a supra-additive way. Known AR systems are associated with stationary-phase cells (22) and some depend directly (AR1) or indirectly (AR2) on *rpoS* (34). Here, we report an assay which focuses on screening actively growing cells. Using this assay, we were able to identify genetic determinants conferring acid tolerance in the acid-sensitive growth phase of *E. coli*. Acid tolerance during active growth is a desirable bioprocessing trait. Resistance to toxic acid stresses, such as those due to production of organic acids or chemicals released during hydrolysis of lignocelluloses, is paramount in efforts to make biorefining and biofuels economically competitive (1,35,36).

We have explored and will report the use of CoGeLs for developing other tolerant phenotypes (solvent and oxidative-stress tolerance). CoGeLs can be used also to screen biased or enriched libraries for identifying interacting genetic loci after partially enriching single libraries. For example, full libraries can be first screened alone or in CoGeLs to generate enriched libraries, which can be screened again with full-genomic partner libraries for a second round of enrichment. Enriched libraries with a few hundred clones each can be screened in pairs or in ternary combinations [three coexisting libraries using three compatible origins of replication (37)]. Furthermore, CoGeLs are well suited for screening metagenomic libraries, which are extensively used to assay for enzymatic function (38). Use of the CoGeL technology would enable the discovery of enzymatic functions that are based on enzyme complexes with genetic determinants that are not proximally coded. The CoGeL technology offers other advantages as well because of the modularity of the technology (Gateway®) used to construct the libraries. Shuffling of an entry library into different destination plasmids requires only the construction of one initial library, thus enabling the construction of

multiple libraries from small amounts or precious metagenomic DNA (39). The library can be recombined into a range of destination vectors, and library coverage can be maintained by performing sufficient recombination reactions. In addition, destination plasmid(s) for constructing CoGeLs can be endowed with strong host promoters in order to enhance expression of library inserts in the desirable host. The use of strong host promoters on CoGeL plasmid libraries can facilitate the expression of metagenomic DNA that would not be otherwise likely expressed in the CoGeL host. These approaches can be used to identify novel catabolic and biosynthetic capabilities and programs utilizing CoGeLs based on both homologous and allogeneic genomic as well as metagenomic libraries.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Data and Methods, Supplementary Tables S1–S3, and Supplementary References [41,42].

ACKNOWLEDGEMENTS

The authors thank Changhao Bi for assistance in generating the fosmid library. Both S.A.N. and S.M.G. designed experiments, performed experiments, analyzed data and wrote the article. E.T.P. conceived the original idea, detailed strategies, designed and supervised experiments, analyzed data and wrote the article.

FUNDING

National Science Foundation (USA) (Grant CBET-0756451). Funding for open access charge: National Science Foundation (USA) (Grant CBET-1033926).

Conflict of interest statement. None declared.

REFERENCES

- Nicolaou,S.A., Gaida,S.M. and Papoutsakis,E.T. (2010) A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: From biofuels and chemicals, to biocatalysis and bioremediation. *Metab. Engineer.*, **12**, 307–331.
- Patnaik,R. (2008) Engineering complex phenotypes in industrial strains. *Biotechnol. Prog.*, **24**, 38–47.
- Stephanopoulos,G. (2002) Metabolic engineering by genome shuffling. *Nat. Biotechnol.*, **20**, 666–668.
- Santos,C.N. and Stephanopoulos,G. (2008) Combinatorial engineering of microbes for optimizing cellular phenotype. *Curr. Opin. Chem. Biol.*, **12**, 168–176.
- Borden,J.R., Jones,S.W., Indurthi,D., Chen,Y. and Papoutsakis,E.T. (2010) A genomic-library based discovery of a novel, possibly synthetic, acid-tolerance mechanism in *Clostridium acetobutylicum* involving non-coding RNAs and ribosomal RNA processing. *Metab. Eng.*, **12**, 268–281.
- Gill,R.T., Wildt,S., Yang,Y.T., Ziesman,S. and Stephanopoulos,G. (2002) Genome-wide screening for trait conferring genes using DNA microarrays. *Proc. Natl Acad. Sci. USA*, **99**, 7033–7038.
- Jin,D., Lu,W., Ping,S., Zhang,W., Chen,J., Dun,B., Ma,R., Zhao,Z., Sha,J., Li,L. *et al.* (2007) Identification of a new gene

- encoding EPSPs with high glyphosate resistance from the metagenomic library. *Curr. Microbiol.*, **55**, 350–355.
8. Chauhan, N.S., Ranjan, R., Purohit, H.J., Kalia, V.C. and Sharma, R. (2009) Identification of genes conferring arsenic resistance to *Escherichia coli* from an effluent treatment plant sludge metagenomic library. *FEMS Microbiol. Ecol.*, **67**, 130–139.
 9. Patnaik, R., Louie, S., Gavrilovic, V., Perry, K., Stemmer, W.P., Ryan, C.M. and del Cardayre, S. (2002) Genome shuffling of *Lactobacillus* for improved acid tolerance. *Nat. Biotechnol.*, **20**, 707–712.
 10. Zhang, Y.X., Perry, K., Vinci, V.A., Powell, K., Stemmer, W.P.C. and del Cardayre, S.B. (2002) Genome shuffling leads to rapid phenotypic improvement in bacteria. *Nature*, **415**, 644–646.
 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. Warner, J.R., Reeder, P.J., Karimpour-Fard, A., Woodruff, L.B.A. and Gill, R.T. (2010) Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat. Biotechnol.*, **28**, 856–862.
 13. Gill, R.T., Wildt, S., Yang, Y.T., Ziesman, S. and Stephanopoulos, G. (2002) Genome-wide screening for trait conferring genes using DNA microarrays. *Proc. Natl Acad. Sci. USA*, **99**, 7033–7038.
 14. Borden, J.R. and Papoutsakis, E.T. (2007) Dynamics of genomic-library enrichment and identification of solvent tolerance genes for *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.*, **73**, 3061–3068.
 15. Lynch, M.D., Warnecke, T. and Gill, R.T. (2007) SCALES: multiscale analysis of library enrichment. *Nat. Methods*, **4**, 87–93.
 16. Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.*, **68**, 669–685.
 17. Park, S.J., Kang, C.H., Chae, J.C. and Rhee, S.K. (2008) Metagenome microarray for screening of fosmid clones containing specific genes. *FEMS Microbiol. Lett.*, **284**, 28–34.
 18. Simon, C. and Daniel, R. (2011) Metagenomic analyses: past and future trends. *Appl. Environ. Microbiol.*, **77**, 1153–1161.
 19. de Lorenzo, V. (2005) Problems with metagenomic screening. *Nat. Biotechnol.*, **23**, 1045; author reply 1045–1046.
 20. Gabor, E., Liebeton, K., Niehaus, F., Eck, J. and Lorenz, P. (2007) Updating the metagenomics toolbox. *Biotechnol. J.*, **2**, 201–206.
 21. Lin, J.S., Smith, M.P., Chapin, K.C., Baik, H.S., Bennett, G.N. and Foster, J.W. (1996) Mechanisms of acid resistance in enterohemorrhagic *Escherichia coli*. *Appl. Environ. Microbiol.*, **62**, 3094–3100.
 22. Castanie-Cornet, M.P., Penfound, T.A., Smith, D., Elliott, J.F. and Foster, J.W. (1999) Control of acid resistance in *Escherichia coli*. *J. Bacteriol.*, **181**, 3525–3535.
 23. Castanie-Cornet, M.P. and Foster, J.W. (2001) *Escherichia coli* acid resistance: cAMP receptor protein and a 20 bp cis-acting sequence control pH and stationary phase expression of the *gadA* and *gadBC* glutamate decarboxylase genes. *Microbiology*, **147**, 709–715.
 24. Iyer, R., Williams, C. and Miller, C. (2003) Arginine-aggmatine antiporter in extreme acid resistance in *Escherichia coli*. *J. Bacteriol.*, **185**, 6556–6561.
 25. Moreau, P.L. (2007) The lysine decarboxylase CadA protects *Escherichia coli* starved of phosphate against fermentation acids. *J. Bacteriol.*, **189**, 2249–2261.
 26. Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA*, **97**, 6640–6645.
 27. Minitab 16 Statistical Software. (2010) [Computer software]. Minitab, Inc, State College, PA.
 28. Richard, H. and Foster, J.W. (2004) *Escherichia coli* glutamate- and arginine-dependent acid resistance systems increase internal pH and reverse transmembrane potential. *J. Bacteriol.*, **186**, 6032–6041.
 29. Jeong, K.C., Hung, K.F., Baumler, D.J., Byrd, J.J. and Kaspar, C.W. (2008) Acid stress damage of DNA is prevented by Dps binding in *Escherichia coli* O157: H7. *BMC Microbiol.*, **8**, 181.
 30. Clarke, L. and Carbon, J. (1976) Colony bank containing synthetic Col EI hybrid plasmids representative of entire *Escherichia coli* genome. *Cell*, **9**, 91–99.
 31. Mandin, P. and Gottesman, S. (2010) Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. *EMBO J.*, **29**, 3094–3107.
 32. Lease, R.A., Smith, D., McDonough, K. and Belfort, M. (2004) The small noncoding DsrA RNA is an acid resistance regulator in *Escherichia coli*. *J. Bacteriol.*, **186**, 6179–6185.
 33. Little, J.W. and Mount, D.W. (1982) The *Sos* regulatory system of *Escherichia coli*. *Cell*, **29**, 11–22.
 34. Foster, J.W. (2004) *Escherichia coli* acid resistance: Tales of an amateur acidophile. *Nat. Rev. Microbiol.*, **2**, 898–907.
 35. Warnecke, T. and Gill, R.T. (2005) Organic acid toxicity, tolerance, and production in *Escherichia coli* biorefining applications. *Microb. Cell Fact.*, **4**, 25.
 36. Sandoval, N.R., Mills, T.Y., Zhang, M. and Gill, R.T. (2011) Elucidating acetate tolerance in *E. coli* using a genome-wide approach. *Metab. Engineer.*, **13**, 214–224.
 37. de Marco, A., Deuerling, E., Mogk, A., Tomoyasu, T. and Bukau, B. (2007) Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. *BMC Biotechnol.*, **7**, 32.
 38. Taupp, M., Mewis, K. and Hallam, S.J. (2011) The art and design of functional metagenomic screens. *Curr Opin. Biotechnol.*, **22**, 465–472.
 39. Gabor, E.M., de Vries, E.J. and Janssen, D.B. (2003) Efficient recovery of environmental DNA for expression cloning by indirect extraction methods. *FEMS Microbiol. Ecol.*, **44**, 153–163.
 40. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M. and Hirakawa, M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, **34**, D354–D357.
 41. Sambrook, J. and Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*. 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
 42. Chang, A.C.Y. and Cohen, S.N. (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from p15A cryptic miniplasmid. *J. Bacteriol.*, **134**, 1141–1156.