Metabolic Engineering 14 (2012) 579–590

Contents lists available at SciVerse ScienceDirect

Metabolic Engineering

journal homepage: www.elsevier.com/locate/ymben

Visualizing evolution in real time to determine the molecular mechanisms of *n*-butanol tolerance in *Escherichia coli* ☆

Luis H. Reyes, Maria P. Almario, James Winkler, Margarita M. Orozco, Katy C. Kao *

Department of Chemical Engineering, Texas A&M University, College Station, TX, USA

ARTICLE INFO

Article history:
Received 2 December 2011
Received in revised form
30 April 2012
Accepted 17 May 2012
Available online 29 May 2012

Keywords:
n-butanol
Evolutionary engineering
Solvent tolerance
Complex phenotype
Genomics

ABSTRACT

Toxicity of products or feedstock components poses a challenge in the biocatalyst-based production of fuels and chemicals. The genetic determinants that are involved in increased resistance to an inhibitor form the adaptive landscape for the phenotype; so in order to engineer more robust biocatalysts, a better understanding of the adaptive landscape is required. Here, we used an adaptive laboratory evolution method called visualizing evolution in real time (VERT) to help map out part of the adaptive landscape of *Escherichia coli* tolerance to the biofuel *n*-butanol. VERT enables identification of adaptive events (population expansions triggered by adaptive mutants) via visualization of the relative proportions of different fluorescently-labeled cells. Knowledge of the occurrence of adaptive events allows for a more systematic isolation of adaptive mutants while simultaneously reducing the number of missed adaptive mutants (and the underlying adaptive mechanisms) that result from clonal interference during the course of in vitro evolution. Based on the evolutionary dynamics observed, clonal interference was found to play a significant role in shaping the population structure of *E. coli* during exposure to *n*-butanol, and VERT helped to facilitate the isolation of adaptive mutants from the population. We further combined adaptive laboratory evolution with genome shuffling to significantly enhance the desired *n*-butanol tolerance phenotype. Subsequent transcriptome analysis of the isolated adaptive mutants revealed different mechanisms of *n*-butanol resistance in different lineages. In one fluorescently-marked subpopulation, members of the Fur regulon were upregulated; which was not observed in the other subpopulation. In addition, genome sequencing of several adaptive mutants revealed the genetic basis for some of the observed transcriptome profiles. We further elucidated the potential role of the iron-related gene in *n*-butanol tolerance via overexpression and deletion studies and hypothesized that the upregulation of the iron-related genes indirectly led to modifications in the outer membrane, which contributed to enhanced *n*-butanol tolerance.

Published by Elsevier Inc.

1. Introduction

Biochemical and biofuel production using microbial systems have been the renewed focus of innumerable research in the last decade. Among the many accomplishments in this area, metabolic engineering efforts have significantly improved the rate and yield of production (Atsumi et al., 2008; Basso et al., 2011; Connor and Liao, 2009; Shen and Liao, 2008; Sillers et al., 2009; Smith and Liao, 2011; Trinh et al., 2011); however, the maximum final titer achievable is a function of the robustness of the microbial system in the production environment (Fischer et al., 2008). The genetic determinants underlying tolerance form the adaptive landscape of the phenotype of interest, are generally complex, and involve interactions between multiple genes (Nicolaou et al., 2010). Further understanding of the adaptive landscape for these complex phenotypes of interest is necessary for the rational engineering of strains. However, these complex phenotypes are largely not well understood, and therefore pose a challenge for the rational design of more robust microbial producers.

Advances in genomic tools have led to the development of reverse engineering tools for the determination of the underlying genetic determinants (points on the adaptive landscape) and molecular mechanisms associated with complex phenotypes; such as transcriptome analysis (Oh et al., 2002; Rutherford et al., 2010), genomic enrichment (Borden and Papoutsakis, 2007; Reyes et al., 2011), SCALEs (Lynch et al., 2007), TRMR (Warner et al., 2010), and coupling high throughput genomic technology with the use of adaptive laboratory evolution (Conrad et al., 2011; Deng and Fong, 2011; Minty et al., 2011). In adaptive laboratory evolution, adaptive mutants (strains with increased fitness) arise spontaneously and expand in a population under a

☆ Author contributions: LHR and KCK conceived and designed the experiments; LHR, MPA, and MMO conducted the experiments; LHR, JW, and KCK analyzed the data; LHR, JW, and KCK wrote the manuscript.

* Corresponding author at: 3122 TAMU, College Station, TX 77843–3122, USA; Fax: +1 979 845 6446.
E-mail address: kao.katy@mail.che.tamu.edu (K.C. Kao).

1096-7176/$ - see front matter Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.ymben.2012.05.002

specific selective pressure (Agrawal et al., 2011; Cooper and Lenski, 2010; Fong et al., 2005; Lee and Palsson, 2010; Minty et al., 2011; Paquin and Adams, 1983a; Paquin and Adams, 1983b). The adaptive mutants harbor genetic mutations that render them fitter than the background population, allowing them to expand in the population. Identification of the underlying genetic mutations and their relative fitness effects (mapping the adaptive landscape) allow researchers to use adaptive evolution as a tool to identify important parameters for the rational engineering for the specific complex phenotype. In addition, the evolutionary trajectories can be elucidated using adaptive evolution to shed additional insight into the fundamental evolutionary processes involved in the acquisition of complex phenotypes. Typically, mutants are isolated from evolving populations for further analysis after an arbitrarily chosen elapsed time and/or at the end of the evolution experiment. However, depending on the population size, the mutation rate, the rate of adaptive mutations, and the relative fitness associated with beneficial mutations, the population may consist of multiple beneficial mutants that co-exist and compete in a phenomenon known as clonal interference (Desai and Fisher, 2007; Fogle et al., 2008; Gerrish and Lenski, 1998; Kim and Orr, 2005; Notley-McRobb and Ferenci, 2000). Clonal interference results in a heterogeneous population and creates complex evolution dynamics during in vitro evolution, where beneficial mutations that arise in the population may be outcompeted and lost through the course of asexual evolution. Thus, depending on when samples are isolated, some beneficial mutations may not be identified, thus imposing a limit on our ability to map the adaptive landscape and in the determination of the trajectories involved during the evolution for the desired phenotype.

To help alleviate the limitation of traditional evolutionary engineering approaches, here, we demonstrate the use of a recently developed adaptive evolution method, called visualizing evolution in real-time (VERT) (Huang et al., 2011; Kao and Sherlock, 2008), combined with genome shuffling, for the engineering and characterization of complex phenotypes in microbial systems. Using different fluorescently marked (but otherwise isogenic) strains, VERT can be used to readily determine when adaptive events occur (the expansion of adaptive mutants in the population) and to facilitate the isolation of those mutants from the population. Additionally, during adaptive evolution, genetic recombination between bacteria does not occur, and therefore beneficial mutations arising in different lineages cannot be combined synergistically to create a superior genotype. To overcome this limitation, we complemented our adaptive laboratory evolution experiment with genome shuffling to allow for recombination between mutant linages. Using this approach, we were able recombine beneficial mutations arising in different individuals in a manner analogous to sexual evolution. Based on prior work by us and others, we hypothesize that butanol tolerance in *E. coli* is a complex phenotype, where the use of adaptive evolution will result in complex population dynamics with multiple beneficial lineages competing as a result of clonal interference, and the use of VERT combined with genome shuffling will help to enhance *n*-butanol tolerance and aid in the identification of additional molecular mechanisms involved.

2. Materials and methods

2.1. Bacterial strains and plasmid construction

The *E. coli* K-12 strain, BW25113 (Δ(*araD-araB*)*567*, Δ*lacZ4787*(::*rrnB*-3), *lambda*-, *rph*-1, Δ(*rhaD-rhaB*)568, *hsdR514*), was used in this study. The conditional replication, integration, and modular (CRIM) system (Haldimann and Wanner, 2001) was used to integrate the plasmids into the genome. Overnight cultures were growth in Luria–Bertani (LB) medium or on solid LB agar plates supplemented with the appropriate antibiotics and incubated at 37 °C. For the construction of the fluorescent-marked strains, the pBAD promoter on the plasmid pTB108ext (Kao et al., 2005) was replaced by the growth linked promoter *rrnB* P1 (Bartlett and Gourse, 1994; Rao et al., 1994), by PCR amplification from the *E. coli* genome using the forward primer 5′-GGC CAA GCT TCT ATA CAA ATA ATA ACT GCA GCC AAG-3′ and the reverse primer 5′-GGC CAA GCT TGA GCT CGG CCG TTG CTT CGC AAC-3′. The engineered promoter has a decreased activity by approximately 50 fold in comparison with the unaltered promoter (−150 to +1 region), by using only the −50 to +1 region (Rao et al., 1994). The DNA sequences corresponding to the *rrnB* P1 promoter, the GFP protein and *rrnB* terminator were cloned into the CRIM plasmid vector pAH144 in the multi-cloning-site. The CRIM plasmid was integrated into the genome using the HK022 phage attachment site, following the procedure described in Haldimann and Wanner (2001). The gene encoding the yellow fluorescent protein (YFP) was amplified from the plasmid pKKGS5, used to construct GSY31137 (Kao and Sherlock, 2008), using the forward and reverse primers 5′-CCG GTC TAG ATC AAA GAT GAG TAA AGG AGA AGA ACT T-3′ and 5′-CCG GCT GCA GGC GGC CGC CTA TTT GTA TAG-3′, respectively, and cloned into the plasmid pAH144. The plasmid was integrated as described above.

2.2. Adaptive evolution experiment

The adaptive evolution experiments were carried out in two chemostat bioreactors for the *n*-butanol challenged experiments in M9 minimum medium supplemented with 5 g/l of D-glucose under aerobic conditions. The system was maintained at steady state with a constant volume of 30 ml, flow rate of 7 ml/h and temperature of 37 °C. The concentration of *n*-butanol in the feed was increased in a step-wise fashion with an initial concentration of 0.5% (v/v) and raised to 1.3% (v/v) (approximate steady-state concentrations inside the bioreactor) during the course of the evolution experiment. The concentration of *n*-butanol was increased arbitrarily every 20 to 30 generations. At generation 88 the chemostats were restarted from glycerol stocks due to mechanical problems in the feeding peristaltic pump. During the re-start, the concentration of *n*-butanol was decreased while the chemostats reached steady state. At regular intervals (approximately every eight generations), samples were taken from the populations and a portion was stored at −80 °C in 17% glycerol for future analysis. The relative proportions of each of the two fluorescent populations were monitored using the fluorescent activated cell sorter (FACS) (BD FACScan™).

2.3. Genome shuffling

Protoplasts of *E. coli* were generated using a modified protocol of Dai et al. (2005). *E. coli* cells were grown at 37 °C in LB medium until mid-exponential phase, harvested by centrifugation, washed three times using ice-cold 0.01 M Tris/HCl (pH 8.0) and resuspended in 0.01 M Tris/HCl pH 8.0 supplied with 0.5 M sucrose. Removal of outer membrane was initiated by using EDTA to a final concentration of 0.01 M. Cells were washed using SMM buffer (0.5 M sucrose, 20 mM sodium maleate monohydrate, 20 mM $MgCl_2$, pH 7.1) and resuspended in SMM buffer with 2 mg/ml lysozyme to remove the cell membrane, generating the protoplasts. The resulting protoplasts of *E. coli* were treated with DNase I (Promega) to digest DNA to prevent transformations due to the release of DNA from lysed cells, incubating at room temperature for 10 min. The protoplasts were harvested by

centrifugation and resuspended in PEG buffer (SMM+40% v/v PEG 6000, 10 mM $CaCl_2$, 5% v/v DMSO), and incubated for 6 min to allow the fusion of the protoplasts. The protoplasts were harvested again and resuspended in SMM buffer. Cells were plated in LB plates, supplied with 0.5 M sucrose, at various dilutions and incubated at 37 °C for 3 day. All cells were recovered from the plates, and used to initiate a new set of chemostats.

2.4. Isolation of mutants

To isolate the adaptive mutants from each of the observed adaptive events, eight clones were isolated from the expanding colored-subpopulation and pairwise competition experiments were carried out in chemostats fed with M9 minimum medium supplemented with 5 g/l D-glucose and 0.8% (v/v) *n*-butanol at 37 °C to measure the relative fitness coefficient of each clone against the previous adaptive mutant (the clones isolated from the first adaptive event were competed against a wild-type expressing a different fluorescent protein). Supplemental Fig. S1 shows the flow of the procedure. The clone with the largest relative fitness coefficient was selected as the adaptive clone from the expanding subpopulation.

For each pairwise competition experiment, equal numbers of the two strains (expressing different fluorescent proteins) to be competed were used to seed each chemostat. The relative proportions of each colored subpopulation were tracked and measured approximately every four generations using FACS. The fitness coefficient was calculated as shown in Eq. (1), where P_i is the relative proportion of strain i in the population and t_j is the generation at which the sample was analyzed.

$$s = \frac{ln((P_{\text{Adaptive mutant @ } t_2}/P_{\text{Adaptive mutant @ } t_1}) \times (P_{\text{Reference @ } t_1}/P_{\text{Reference @ } t_2}))}{t_2 - t_1} \quad (1)$$

2.5. n-butanol-shock experiment

An *n*-butanol-shock experiment was performed in 2% *n*-butanol (v/v). Cells isolated from each adaptive event were grown overnight at 37 °C in M9 minimal medium supplemented with 5 g/l D-glucose, then diluted at a ratio of 1:100 into fresh M9 minimal medium, and grown at 37 °C to mid-exponential phase (OD_{600} 0.6–1.0). The OD_{600} of each culture were normalized to 0.5 using fresh M9 minimal medium and *n*-butanol was added to a final concentration of 2% (v/v). Samples were incubated for 1 h at 37 °C and 220 rpm. After incubation, the samples were serial diluted in fresh M9 minimal medium and plated on LB plates. After 18 h, pictures of the plates were taken to assess relative survival rates between the different strains.

2.6. Cell harvest for RNA extraction

Cells were grown in chemostats with a constant volume of 30 ml, flow rate of 7 ml/h, fed with M9 minimal medium supplemented with 5 g/l D-glucose and 0.8% (v/v) *n*-butanol at 37 °C. After the system reached steady state (approximately 36 h), the cells were harvested by filtration using NALGENE analytical test filter funnels (Nalgene) and immediately resuspended in 10 ml of RNA*later* (Sigma) and stored at −80 °C for later processing.

2.7. Total RNA extraction

The total RNA was extracted using the RNeasy Mini Kit (QIAGEN). For each of the isolated mutants, 1.5 ml of sample stored at −80 °C in RNA*later* were centrifuged and the pellet was recovered. Total RNA was extracted by following manufacturer's instructions. The extracted RNA was treated with DNase I and purified using acid phenol/chloroform followed by ethanol precipitation. The RNA was quantified using the Qubit fluorometer using the RNA quantification kit (Invitrogen).

2.8. Labeled cDNA generation, microarray hybridization and data analysis

The SuperScript indirect cDNA labeling system (Invitrogen) was used to generate cDNA incorporating amino-allyl dUTP. cDNA was recovered using ice cold ethanol precipitation. Cy3- and Cy5-dUTP dyes (GE Healthcare) were used to label cDNA samples by following manufacture's protocol. The labeled cDNA was hybridized to the *E. coli* Gene Expression Microarray (Agilent Technologies). The arrays were scanned using the GenePix 4100A Microarray Scanner and images analysis performed using GenePix Pro 6.0 Software (Molecular Devices).

The Microarray Data Analysis System (MIDAS) software was used to normalize the data using LOWESS based normalization algorithm (Quackenbush, 2002; Yang et al., 2001). Differentially expressed (DE) genes were identified using the rank product method with a critical *p*-value $P < 0.01$ (Breitling et al., 2004). The MeV (TM4) (Saeed et al., 2003) microarray analysis software was used for clustering and other expression profile analysis.

2.9. Calculation of growth kinetic parameters

The parameters "Percentage of Inhibition" and "Relative increase in maximum specific growth rate" were calculated as shown in Eqs. (2) and (3), respectively. These parameters were determined by measuring the maximum specific growth rate ($\mu_{\max,i}$) of each strain (strain i) in M9 minimal medium (supplied with 0.5% (w/v) D-glucose) supplemented with 0% and 0.5% (v/v) *n*-butanol. The growth kinetics for each strain was measured using a TECAN Infinite M200 Microplate reader (TECAN). Four biological replicas were obtained per sample. Student's *t*-test was used to determine if there was a significant improvement in the *n*-butanol tolerance in overexpression or deletion strains.

$$\text{Inhibition} = 1 - \left(\frac{\mu_{\text{strain @ 0.5\% } n\text{-butanol}}}{\mu_{\text{strain @ 0\% } n\text{-butanol}}}\right) \quad (2)$$

$$\text{Relative increase in } \mu_{\text{Clone}} = \frac{((\mu_{\text{Clone @ 0.5\% n-Butanol}}/\mu_{\text{Clone @ 0\% } n\text{-butanol}})) - ((\mu_{\text{wild-type @ 0.5\% } n\text{-butanol}}/\mu_{\text{wild-type @ 0\% } n\text{-butanol}}))}{((\mu_{\text{wild-type @ 0.5\% } n\text{-butanol}}/\mu_{\text{wild-type @ 0\% } n\text{-butanol}}))} \quad (3)$$

2.10. Overexpression and deletion studies

The ASKA Clone (-) collection (Kitagawa et al., 2005), which contains a set of plasmids overexpressing all predicted *E. coli* K-12 ORFs under the control of an IPTG-inducible promoter, was used as the source of all plasmid constructs for overexpression studies. *E. coli* K-12 strain AG1 (*recA1*, *endA1*, *gyrA96*, *thi-1*, *hsdR17* (r_k^- m_k^+), *supE44*, *relA1*) is the host for the ASKA collection. Plasmids of the clones from the ASKA Clone (−) collection

overexpressing the genes of interest were isolated and transformed into *E. coli* strain BW25113. All the overexpression analyses in this study were also performed in AG1 to reduce strain-specific effects. For deletion studies, the Keio collection (Baba et al., 2006), which contains single gene deletions for all non-essential genes in BW25113, was used.

2.11. Whole genome resequencing

Genomic DNA (5 μg) was isolated from single colonies for each of the isolated mutants to be sequenced using Solexa technology (Illumina Inc., San Diego, CA, USA), with multiplexing of six strains per lane. DNA library generation and sequencing was performed by the Borlaug Genomics and Bioinformatics core facility at Texas A&M University. The software CLC Genomics Workbench 4 (CLC bio, Germany) was used to assemble the different sequences and to identify single-nucleotide polymorphisms (SNPs) and possible genome rearrangements. Verification of possible SNPs and translocations identified from the genome resequencing were performed using traditional Sanger sequencing (MCLAB, South San Francisco, CA, USA).

2.12. Allele tracking

Genomic DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen). The CFX384 Real-Time PCR Detection System (Bio-Rad) was used to carry out the qRT-PCR experiments. GoTaq® qPCR Master Mix (Promega) was used for all the quantitative real-time PCR (qRT-PCR) experiments using 20 ng of total genomic DNA and 0.5 μM of primers. The primers used for the tracking of the different alleles are: (A) the *rhoS82F* allele, For-Wild-Type: CCG GTC CTG ATG ACA TCT ACG TTT C, For-Mutation: CCG GTC CTG ATG ACA TCT ACG TTT T, Rev: TTA CCA GAC GCT GCA TCT CGG, (B) the *feoA201*::*IS5* allele, For-Wild-Type: CCA CGG CGC GCC ATT AC, For-Mutation: CCA CGG CGC GCC ATT AT, Rev: TCA TCG CGT TCG GTG GGA, (C) the *FBSmut* allele, For: ATG CAA TAC ACT CCA GAT ACT GCG TGG, Rev-Wild-Type: CTG CGA TGA GAT GGT GGT CAG AGA ATA, Rev-Mutation: CAA GGG GTT GAT GAA AGA CGA TAA CCA AC, (D) the *nusAE212A* allele, For-Wild-Type: TCT TCG CCG ATT TCT GGC ACT T, For-Mutation: TCT TCG CCG ATT TCT GGC ACT G, Rev: TCC TGC GCG AAG ATA TGC TGC, and (E) the *relA142*::*IS2* allele, For-Wild-Type: GCA ATA CGC TCC GCC AGT TTG AT, For-Mutation: GGA GAT TCA GGG GGC CAG TCT A, Rev: AGC TGA AAG CGA CGC ACA CTG. The thermocycler protocol used was: 95 °C for 3 min followed by 39 cycles of 95 °C for 10 s followed by 60 °C for 30 s. The allele frequency calculations were done as described by Kao and Sherlock (2008).

3. Results

3.1. Confirmation of neutrality and definition of adaptive events

The primary utility of the VERT method is to identify adaptive events (population expansions) by tracking the relative proportions of different fluorescent subpopulations during the evolutionary time-course. As a beneficial mutant arises and expands in the population, the corresponding labeled subpopulation increases in proportion. By tracking the different subpopulations using FACS, it is possible to determine when mutants with increased relative fitness arise in the evolving population; allowing us to isolate adaptive mutants from the populations in a more rational manner. Thus, it is important that any fitness biases in the fluorescent proteins used are taken into consideration in the analysis in order to avoid isolation of false adaptive mutants. The basal variability of relative subpopulation proportions from FACS measurements of the different fluorescently-marked strains were analyzed using neutrality tests, which typically follow the short time propagation ($<\sim$20 generations typically) of the labeled subpopulations to detect any fitness differences arising from different fluorescent protein expression, as beneficial mutations are unlikely to occur within such a short time scale (Ferea et al., 1999). To assess the neutrality between the two fluorescently-marked strains used for this work (GFP and YFP-labeled), we used the early time points from control experiments, where we seeded parallel continuous culture experiments with approximately equal numbers of the two fluorescently-marked strains in M9 minimum medium supplemented with glucose. The data for the first 24 generations after the populations reached steady-state (>6 volume replacements) were used to assess the variation in fitness between the two fluorescently-marked strains (see Fig. S2); a slight increase in the yellow subpopulation was observed in both populations, suggesting a potentially slight fitness benefit in the yellow subpopulation either due to fluorophore differences or due to jackpot mutations present in the overnight inoculum. The neutrality data was then used to measure the expected levels of measurement noise.

The analysis for estimating measurement noise focuses on the average change in population proportion per generation (μ_R) and its standard deviation (σ_R) within the neutrality dataset. Since no (or few) adaptive events occur during the first few generations of an evolution experiment, it is very likely that any change observed arises from random variations during measurement or the sampling procedure. Evolved populations can be also analyzed in a similar fashion using regions in which the subpopulations are in a meta-stable state (i.e., no net expansion or contraction). These meta-stable states may occur if there are no significant differences between the relative fitness of the adaptive mutants in the different colored subpopulations. If the fluorescent proteins are in fact neutral, the average slope should be close to zero ($\mu_R \approx 0$) to reflect equivalent fitness of the two colored subpopulations. These parameters together intuitively describe how much a given population proportion can be expected to fluctuate between subpopulations with equal fitness. Based on data generated during this study, $\mu_R=0.001$ (95% CI: [0.0031, −0.0009]) and $\sigma_R=0.004$ ($N=16$) based on the assumption that deviations from the mean are normally distributed. These statistical inferences derived from the neutrality data may also be applied to determine if an arbitrary population expansion meets a statistically significant threshold (Winkler and Kao, 2012) and will be used to identify adaptive events based on the evolutionary dynamics described below.

3.2. The evolutionary dynamics of E. coli during evolution for enhanced n-butanol tolerance

Approximately equal numbers of GFP and YFP marked strains were used to seed two *n*-butanol challenged populations (P1 and P2) in continuous cultures in chemostats with a starting concentration of 0.5% (v/v) *n*-butanol. The concentration of *n*-butanol in the chemostats was increased to 1.3% (v/v) in a step-wise manner over the course of approximately 144 generations (see Fig. 1). The relative proportions of the two fluorescently-marked subpopulations throughout the course of the evolution experiment were monitored using FACS (the evolutionary dynamics during the in vitro evolution are shown in Fig. 1). An expansion in a colored subpopulation is indicative of the occurrence and expansion of an adaptive mutant in that subpopulation. Thus by tracking the expansions and contractions of the two colored-subpopulations, we were able to identify adaptive events occurring throughout the course of the in vitro evolution. The expansions and contractions in both colored-subpopulations are clear indications that

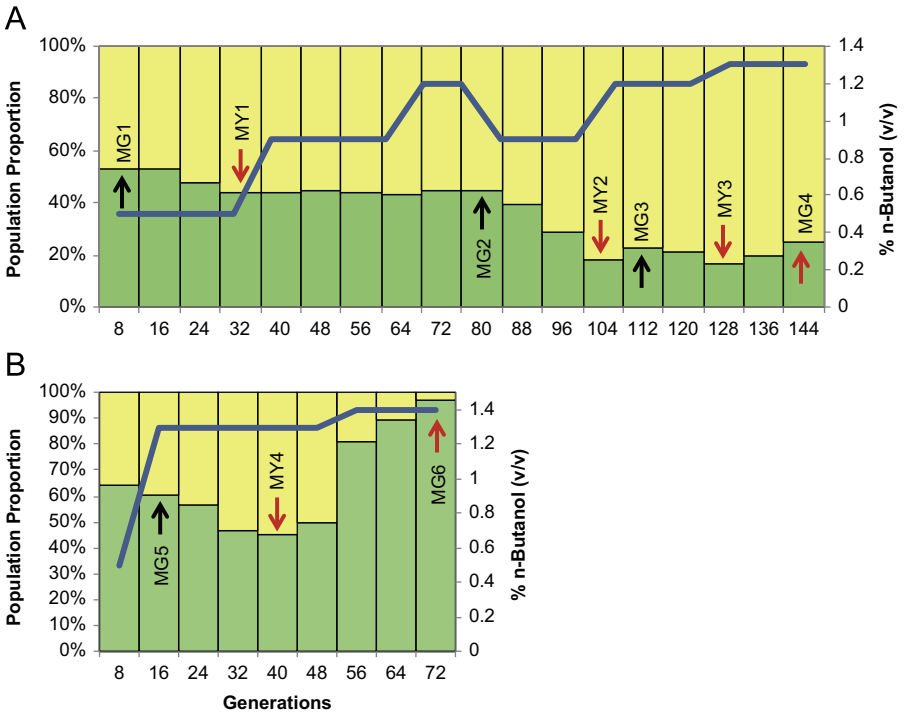

Fig. 1. The evolutionary dynamics of the in vitro evolution for (A) *n*-butanol challenged population P2, and (B) the second round of evolution after genome shuffling in the presence of *n*-butanol. The yellow bars indicate the proportion of yellow subpopulation and the green bars indicate the proportion of the green subpopulation. The blue line indicates the approximate steady-state concentration of *n*-butanol in the chemostat. A visual inspection of the evolution dynamics allows us to determine different adaptive events, defined as an expanding subpopulation when such expansion reaches a maximum, marked with arrows. The red arrows indicate the adaptive events determined using the computational algorithm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

clonal interference plays a role in shaping the population structure in our populations.

Upon completion of the adaptive evolution experiment, population P2 was arbitrarily chosen for further analysis. We identified seven adaptive events, where the expanding subpopulations reached their maximum proportion at generations 7, 32, 104, 112, 128, and 144, in this population via visual inspection. Between generation 32 and 80, a meta-stable region was observed, which led us to hypothesize that an adaptive event occurred in the green subpopulation that was able to impede the continued expansion of the yellow subpopulation at generation 32. Thus, generation 80 was also identified as the end of an adaptive event (in this case, the end of the meta-stable region). Using the computational algorithm, 4 adaptive events, reaching maximum proportion at generations 32, 104, 128, and 144, were identified (Fig. 1A). We isolated adaptive mutants from all the potential adaptive events identified visually (all the computationally identified adaptive events were also identified visually).

To isolate the adaptive mutants responsible for the observed expansions in population P2, eight clones were randomly picked for each of the observed adaptive event, from the expanding colored subpopulation at the generation when the relative proportions are at its' maximum (or at the end of the meta-stable region) (marked by arrows in Fig. 1A). Using pair-wise competition experiments in chemostats with 0.8% (v/v) *n*-butanol, the clone with the highest relative fitness coefficient against the previous adaptive mutant was selected as the adaptive mutant from that expanding subpopulation. These adaptive mutants are named according to their color and sequence in the evolution (labeled and shown in Fig. 1A and Table S1). The relative fitness coefficients for each isolated adaptive mutant was also measured against a differentially-marked parental strain (Table S1 and Fig. 2). All but 2 (MG3 and MG4) of the isolated adaptive mutants showed positive or neutral fitness coefficients

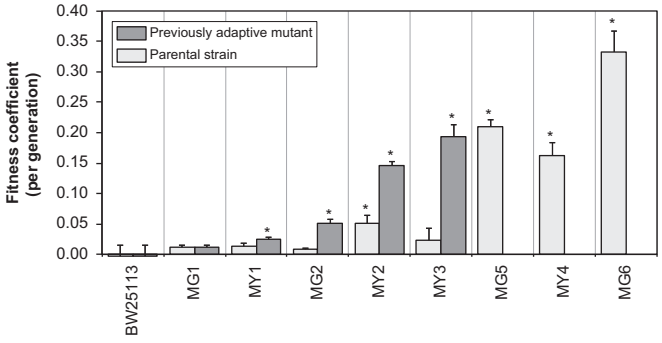

Fig. 2. The relative fitness coefficients of isolated adaptive mutants from population P2 in the presence of 0.8% *n*-butanol. Light gray bars: the fitness coefficients compared to the previously isolated mutant as reference. Dark gray bars: the fitness coefficient compared to a differentially-labeled parental strain. Measurements that are statistically significantly different from the wild-type (student t-test with a *p*-value cut-off of 0.05) are marked by asterisks.

in the presence of 0.8% (v/v) *n*-butanol; MG3 and MG4 were not analyzed further.

3.3. Recombination significantly enhances the desired phenotype

At generation 144, the *n*-butanol-challenged populations began to washout at 1.3% (v/v) *n*-butanol, marking the end of the first round of adaptive evolution experiments. At this point, we subjected samples from the two *n*-butanol-challenged populations to genome shuffling via protoplast fusion (Dai et al., 2005; Winkler et al., 2010) in an attempt to further enhance tolerance by recombining the beneficial and potentially synergistic mutations between different lineages. The resulting protoplast fusion library was used to seed a new chemostat in the presence of *n*-butanol for a second round of evolution (Fig. 1B). An

enhancement in *n*-butanol tolerance was detected immediately, as the population was stably maintained at 1.3% (v/v) *n*-butanol (whereas they were washing out prior to genome shuffling). Adaptive mutants were identified and isolated from this second round of evolution, and were named MG5 (generation 16 post genome shuffling), MY4 (generation 40), and MG6 (generation 72).

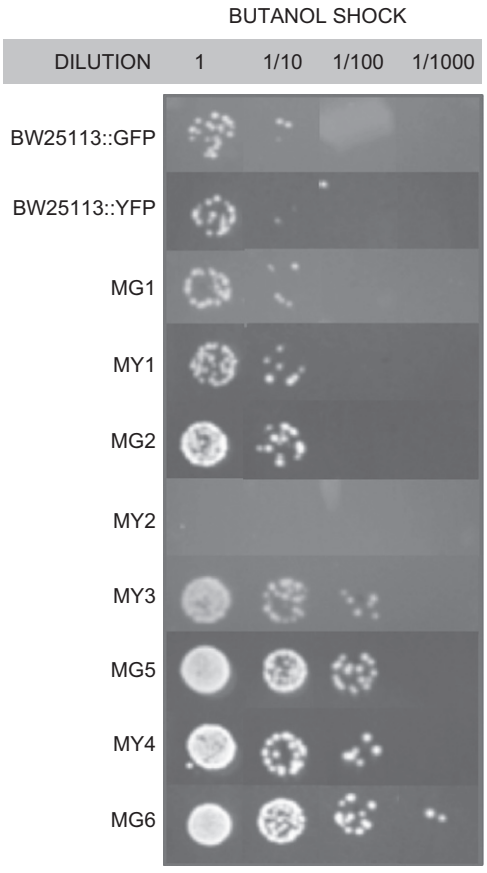

Fig. 3. Relative cell viability of fluorescently-marked wild-type and isolated adaptive mutants after short-term exposure to 2% (v/v) *n*-butanol.

3.4. Enhanced survival upon n-butanol shock

The survival rate after short-term exposure to a higher concentration of *n*-butanol was conducted to determine the tolerance level of the adaptive mutants. The cells were exposed to 2% (v/v) *n*-butanol for 1 h and their viabilities were assessed (see Fig. 3). Most adaptive mutants (MG2, MY3, MG5, MY4 and MG6) showed at least a 10-fold increase in survival rate upon *n*-butanol exposure compared to the parental strains. The adaptive mutants isolated after the genome shuffling (MG5, MY4, and MG6) showed ~10–100-fold increase in survival rate compared to the parental strains. This corresponded to the higher relative fitness coefficient in these strains compared to the parental strains in the presence of *n*-butanol. Similarly, mutant MG1 that showed little increase in its *n*-butanol tolerance also showed a lack of relative fitness coefficient gain relative to the wild-type. MY2 is an exception: it showed a decreased viability after 2% (v/v) *n*-butanol exposure, but had a higher fitness coefficient relative to the previous adaptive mutant (Fig. 2), suggesting that this particular mutant either only has a higher resistance to lower concentrations of *n*-butanol (i.e. the concentration in the bioreactor that it was isolated from) or it may have mutation(s) that confer a fitness advantage for more efficient nutrient assimilation.

3.5. Transcriptome analysis

DNA microarrays were used to assess the transcriptional changes in the isolated adaptive mutants compared to the parental strains in response to *n*-butanol stress. The transcriptome data showed some similarities and also some major differences between the adaptive mutants isolated from the two lineages (GFP vs. YFP). Some of the key transcriptional perturbations and subsequent experimental validations are described below.

3.5.1. Changes in regulation of fatty acids and membrane composition

We found significant expression changes in genes involved in cell wall and membrane biosynthesis in every isolated mutant analyzed. Table 1 summarizes the differentially regulated genes related to fatty acids and membrane composition in each isolated mutant. Genes related to the biotin biosynthetic pathway were

Table 1
Selected genes involved in fatty acid or membrane composition that were upregulated (bold) or downregulated (non-bolded) in the isolated adaptive mutants.

Strain	Function						
	A[a]	B[b]	C[c]	D[d]	E[e]	F[f]	G[g]
MG1							***yhhK***
MY1		***rffM***		*plsB*, *mdoB*, *plsX*	***mraW***, ***bamB***	***bioA***	
MG2		***rffH***, ***glmM***, ***basR***, *lpxB*, *lpxD*, *lpxH*	*mrdA*, *prc*, *amiA*, *mtgA*, *mltA*		***lplA***, ***mraW***, ***borD***, ***yqeF***	***bioA***, ***bioB***, ***bioC***, ***bioF***	*panD*, *coaA*
MY2		*rffD*, *htrL*,	*ddlA*, *murD*, *oppA*, *oppC*, *oppD*	*ugpC*, *ugpE*			
MY3		***lpcA***, ***lpxK***	***bacA***,	*psd*, *plsX*	***ybfP***, ***lplA***, *yqeF*	*bioA*	
MG5	***cfa***	***lpxL***, *kdsA*, *glmS*, *wbbJ*, *rfaG*, *rfaP*	*prc*, *mrcA*	*ybgC*	***lplA***, ***ycfL***, *csgG*, *bamB*, *smpA*	***bioA***	
MY4	***scpA***, ***lipB***, ***mhpE***	***glf***, ***rfbD***, ***rffA***, ***rffM***, ***lpxK***, *gmm*	***murE***, ***murF***, ***amiB***, *murB*, *hipA*, *oppD*		***lplA***, *ycfL*		
MG6	***cfa***	***lptA***, ***basR***, *gmm*			***lplA***		

[a] Fatty acid biosynthesis.
[b] Lipopolysaccharide biosynthesis.
[c] Peptidoglycan synthesis.
[d] Phospholipids biosynthesis.
[e] Lipoprotein.
[f] Biotin biosynthesis.
[g] Pantothenate synthesis.

upregulated in several adaptive mutants in both lineages (MY1, MG2 and MG5). Biotin is an essential cofactor for acetyl-CoA carboxylase, a key enzyme in the biosynthesis of fatty acids (Gavin and Umbreit, 1965). The biotin-related genes were also found to be enriched in a recent study using an *E. coli* genomic library to identify genes involved in *n*-butanol resistance (Reyes et al., 2011). Overexpression of the gene *bioA*, which encodes the 7,8-diaminopelargonic acid synthase enzyme in the biotin biosynthesis pathway in the wild-type strain, showed significant improvements in its' *n*-butanol tolerance ($30 \pm 4\%$) (Fig. 4).

The lipopolysaccharide (LPS) biosynthesis genes were also differentially regulated in all the isolated adaptive mutants (Table 1). The bacterial LPS is the main component of the outer membrane, composed of a hydrophobic domain (lipid A), a phosphorylated oligosaccharide, and an O-antigen; however, *E. coli* K-12 normally does not make O-antigen due to a mutation in the *rfb* gene cluster (Yao and Valvano, 1994). Increases in the expression of genes involved in lipid A production were observed in adaptive mutants from both lineages (MY3, MG5, MY4 and MG6), suggesting a possible change in the hydrophobicity of the cell wall in these strains. Changes in the expression of many enterobacterial antigen related genes were also observed (Table 1) and may also contribute to changes in cell wall hydrophobicity. Increase in the LPS content may lead to less hydrophobic cell surfaces, hindering the penetration of hydrophobic compounds into the cytosol, and increasing its tolerance to organic solvents (Aono and Kobayashi, 1997).

After genome shuffling, we saw an upregulation of genes involved in fatty acids biosynthesis, including the genes *cfa* (MG5 and MG6) and *scpA*, *lipB* and *mhpE* (in MY4). The cyclopropane fatty acyl phospholipid synthase gene (*cfa*) modifies the double bond of unsaturated fatty acids into their cyclopropane derivatives (Taylor and Cronan, 1979). This conversion, known as homeoviscous adaptation, occurs in different bacteria in response to environmental changes. This change in the membrane of some of the adaptive mutants was confirmed through Fatty Acid Methyl Ester (FAME) analysis and the results are shown in Table S2. This was further experimentally verified via overexpression of *cfa* in the wild-type strain, which led to significant improvement in *n*-butanol tolerance (Fig. 4).

Long-chain alcohols can affect membrane structure via insertion into the lipid bilayer in any orientation, increasing membrane fluidity. This phenomenon has been demonstrated by electron-spin resonance (Grisham and Barnett, 1973) and differential spin calorimetry (Hui and Barton, 1973). Modifications in the membrane that counteract these disruptions may be obtained by increasing the saturated/unsaturated fatty acids ratio or by increasing the length of the fatty acid (Sikkema et al., 1995). The results from FAME analysis revealed a decrease in the proportion of unsaturated and short-chain fatty acids in MY3 and MY4, whereas an increase of more than five times the amount of *cis*-10,11-Methylene-nonadecanoic acid was observed in MG6.

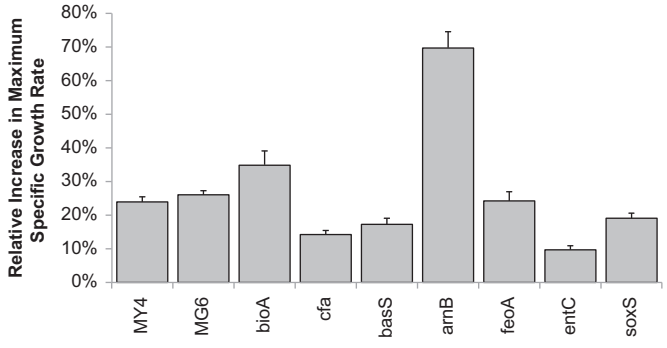

Fig. 4. The relative increase in maximum specific growth rate via overexpression of individual genes compared to the wild-type in the presence of 0.5% (v/v) *n*-butanol. From left to right: MY4, MG6, BW25113/pCA24N::*bioA*, BW25113/pCA24N::*cfa*, BW25113/pCA24N::*basS*, BW25113/pCA24N::*arnB*, BW25113/pCA24N::*feoA*, BW25113/pCA24N::*entC* and BW25113/pCA24N::*soxS*.

3.5.2. Iron-ion transport and metabolism

In several isolated adaptive mutants, we observed an unexpected transcriptional perturbation in the iron metabolic process (Table 2). Iron uptake is carried out by siderophores, organic ligands that sequester the insoluble Fe^{+3} into the cell (Neilands, 1995). We found increased expression in genes related to the biosynthesis and transport of enterobactin and genes involved in different high affinity iron transport mechanisms, such as *feoA*, *feoB*, *ftnB*, *fiu* and *fieF* in mutants isolated from the green lineage (MG2, MG5 and MG6). Interestingly, the mutants from the yellow lineage exhibited the opposite response in iron ion related genes, suggesting the two lineages had evolved different mechanisms for enhanced tolerance.

Fur is the main transcription factor involved in the regulation of iron-related genes. The increase in the expression of such a large number of these genes suggest a decrease in the activity of Fur (Bagg and Neilands, 1987) in these mutants. Using Network Component Analysis (NCA) (Liao et al., 2003; Tran et al., 2005), Fur activity was identified to be significantly repressed in several green mutants (MG2, MG5, MG6) and significantly increased in several yellow mutants (MY2, MY3, MY4) (see Fig. 5). Fur was the only transcription factor from the NCA compliant set of transcription factors that

Table 2
Selected genes involved in iron transport and metabolism that were upregulated (bold) and downregulated (non-bolded) in the isolated adaptive mutants.

Strain	Function	Genes
MG1	Iron ion transport	***fecC***, ***fiu***, ***ftnB ftnB***
MG2	Iron ion transport	***ftnB***, ***fiu***, ***fepA***, ***fhuE cirA***, ***basR***, ***basS*** *yaaA*, *fhuA*,
	Siderophore biosynthesis	***entA***, ***entB***, ***entC***, ***entD***, ***entE***
MY2	Iron ion transport	*fhuD*, *entF*
	Siderophore biosynthesis	*entE*, *entF*
MY3	Iron ion transport	*fecC*, *exbB*
MG5	Iron ion transport	***fhuF***, ***feoA***, ***feoB***, ***feoC***, *yaaA*
	Siderophore biosynthesis	***entB***, ***entC***, ***entD***, ***entE***
MY4	Iron ion transport	*fecR*, *fecI*
	Siderophore biosynthesis	*entA*, *entC*, *entS*
MG6	Iron ion transport	***basR***, ***feoA***, ***feoB***, ***feoC***, ***fecA***, ***fecR***, ***fecI***
	Siderophore biosynthesis	***entC***, ***entE***

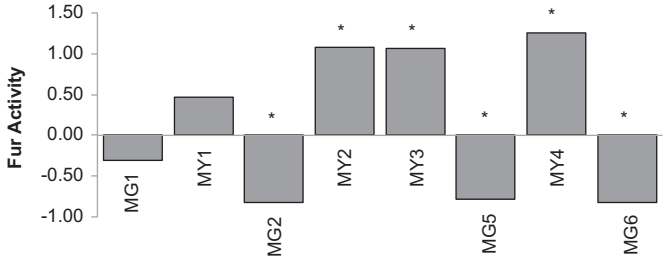

Fig. 5. Fur activity in the isolated mutants determined using NCA. The asterisks indicate statistical significance with a *p*-value threshold of 0.05 using 1000 permutations.

showed such a significant difference in activity between the two different lineages.

3.5.3. Verification of iron-related genes via overexpression studies

If decreased Fur activity was responsible for increased *n*-butanol tolerance in the mutants in the green lineage, we would expect a decrease in tolerance if Fur was overexpressed in the mutant. To test this hypothesis, we overexpressed *fur* in MG6 and compared it to wild-type overexpressing *fur*. In the wild-type strain, overexpression of *fur* does not significantly increase the inhibition by *n*-butanol (Fig. 6). However, overexpression of *fur* in MG6 inhibited growth significantly in the presence of *n*-butanol, supporting the model that the activity of Fur is involved in *n*-butanol tolerance in this mutant.

The contribution of iron-related genes in *n*-butanol tolerance was validated by overexpressing *entC* and *feoA* in the wild-type strain. In the presence of 0.5% (v/v) *n*-butanol, the overexpression of *entC* and *feoA* genes improved growth compared to wild-type under the same conditions by $10 \pm 1\%$ and $24 \pm 3\%$, respectively, demonstrating that the genes involved in enterobactin biosynthesis and transport are involved in *n*-butanol tolerance (see Fig. 4). The genes *entC* and *feoA* were also enriched in a previous study of *n*-butanol tolerance in *E. coli* (Reyes et al., 2011). The enhanced tolerance conferred by overexpression of the iron-related genes was not specific to BW25113, as an increase in tolerance was also achieved when overexpressed in strain AG1.

The effects of the overexpression of iron-related genes on cell viability under *n*-butanol shock were also assessed (Fig. 7). The overexpression of *entC* and *feoA* increased the viability of the strain to *n*-butanol exposure by more than 10-fold. Thus, the iron-related genes not only increase tolerance towards *n*-butanol in terms of fitness advantage, but also in terms of cell viability.

3.5.4. Downregulation of Fur activity may have resulted in changes in lipid A composition

To the best of our knowledge, iron transport and metabolism has not been shown to play a role in *n*-butanol tolerance. Thus, to investigate the role of iron in *n*-butanol tolerance, the transcriptome data was analyzed for potential membrane modifications that result in changes in the cell wall or membrane, which is the most common mechanism of solvent tolerance. Both the MG2 and MG6 strains exhibited a general upregulation of iron-regulated genes including the *basR* gene, which encodes the bacterial adaptive response transcriptional regulatory protein. BasR belongs to the BasS/BasR two-component system, which responds to high concentration of Fe^{+3}. Activation of BasR via phosphorylation by BasS leads to

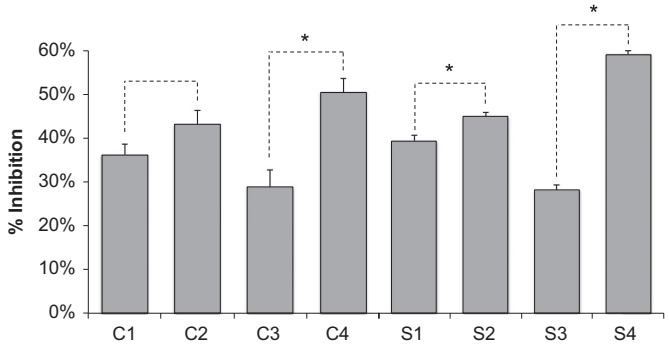

Fig. 6. Growth inhibition in the presence of 0.5% (v/v) *n*-butanol. C1. BW25113/pCA24N, C2. BW25113/pCA24N::*fur*, C3. MG6/pCA24N, C4. MG6/pCA24N::*fur*. S1. BW25113/pCA24N, S2. BW25113Δ*basR*/pCA24N, S3. BW25113/pCA24N::*entC*, and S4. BW25113Δ*basR*/pCA24N::*entC*. Asterisks indicate measurements that are statistically significantly different between the two strains (determined using a student *t*-test with a *p*-value cut-off of 0.05).

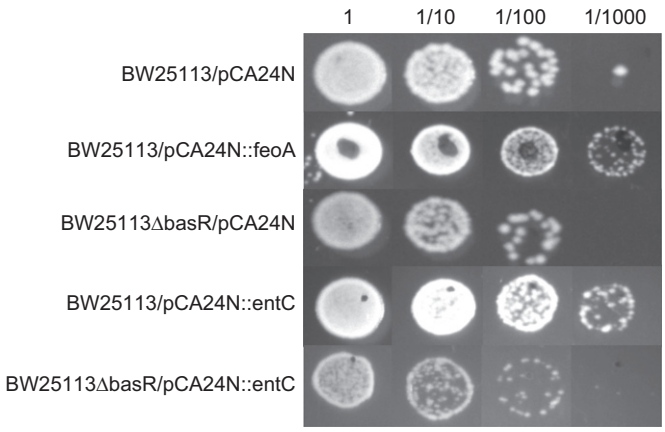

Fig. 7. Cell viability after short-term exposure to 2% (v/v) *n*-butanol to determine the effects of *Δe th* on the *n*-butanol tolerance of strains overexpressing iron-related genes.

increased expression of genes involved in modification of lipopolysaccharide to prevent excessive Fe^{+3} binding (Hagiwara et al., 2004). We found that overexpression of *basS* increased *n*-butanol tolerance (Fig. 4). Members of the BasSR regulon include the *arn* operon (*arnBCADTEF*) (Hagiwara et al., 2004), which is involved in the resistance of gram-negative bacteria to cationic antimicrobial peptides (CAMPs) via the addition of positively charged 4-amino-4-deoxy-l-arabinose to lipid A of the LPS (Gatzeva-Topalova et al., 2005). Even though the *arn* operon was not significantly upregulated in several other strains that showed overexpressions of the iron-related genes (MG2, MG5 and MG6), we decided to test whether increased expression of the *arn* operon would increase *n*-butanol tolerance. Overexpression of *arnB* increased *n*-butanol tolerance by approximately 70% in comparison to the wild-type (Fig. 4). This suggests that the potential LPS change resulting from the overexpression of the *arn* operon increases *n*-butanol resistance.

Based on the upregulation of the iron-related genes and *basR* and the increased *n*-butanol tolerance when we overexpressed *entC*, *feoA*, *basS* and *arnB* in the wild-type strain, we hypothesize that the decreased activity of Fur led to an increase in activation of BasSR, which led to changes in the LPS, ultimately increasing *n*-butanol tolerance. To demonstrate that the increased *n*-butanol tolerance conferred by overexpression of iron-related genes may be due to the downstream effects of increased BasSR activity, we overexpressed *entC* in the Δ*basR* strain. This combination resulted in a significant increase in susceptibility to *n*-butanol compared to wild-type, while overexpression of *entC* in the wild-type strain reduced the susceptibility to *n*-butanol (Fig. 6 and Fig. 7), suggesting that the effects on *n*-butanol tolerance conferred by the iron-related genes are through the BasSR system. If our hypothesis were correct, then we would expect to see an increase in resistance of mutants MG2, MG5 and MG6 (GFP-labeled mutants exhibiting increased expression of iron-related genes) to the CAMP polymyxin B. Indeed, the data (Fig. 8) agreed with the expected results, demonstrating that the mutants were more resistant to the antibiotic compared to the wild-type. On the other hand, the mutants that did not show an upregulation in the iron-related genes did not exhibit any increase in resistance to polymyxin B.

3.6. Genome sequencing

The genomes of all the isolated adaptive mutants were re-sequenced via ultra high throughput sequencing using 32 bp single-end sequencing for an average coverage of 30X. All confirmed mutations are shown in Table 3. Note that due to the short

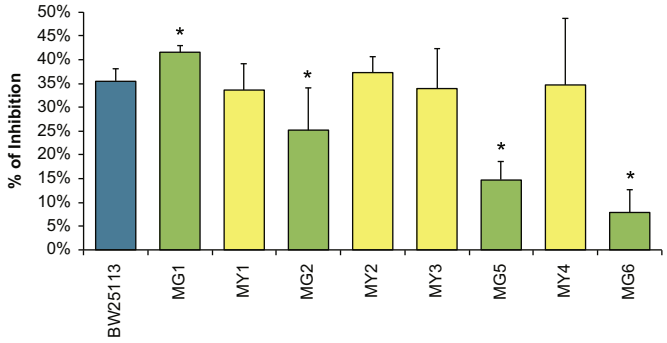

Fig. 8. Growth inhibition of BW25113, MG1, MY1, MG2, MY2, MY3, MG5, MY4 and MG6 in the presence of 3 μg/ml of the cationic antibiotic peptide Polymyxin B. Asterisks indicate adaptive mutants that are affected significantly compared to the wild-type (determined using a student *t*-test with a *p*-value cut-off of 0.05).

Table 3
List of validated mutations detected using genome resequencing.

Strain	Mutation
MG1	No mutations identified
MY1	No mutations identified
MG2	No mutations identified
MY2	*relA142*::*IS2*
MY3	No mutations identified
MG5	*rhoS82F*
	feoA201::*IS5*
MY4	*nusAE212A*
MG6	*rhoS82F*
	feoA201::*IS5*
	FBSmut, a SNP (C→T) Fur binding site *fes-ybdZ-entF-fepE* (at 141 bp upstream of fes) and *fepA-entD* (157 bp upstream of *fepA*)

single-end sequencing, we were confident in our ability to identify SNPs, but not chromosomal rearrangements.

In the green mutants, we only identified mutations in the last two mutants. The genome sequence of the mutants MG5 and MG6 revealed a single-nucleotide polymorphism (a C→T transition (S82F)) in *rho*. A recent study using adaptive evolution in *E. coli* demonstrated that a point mutation in *rho* plays an important role in ethanol resistance (Goodarzi et al., 2010). However, the defective termination factor *rho* imparts pleiotropic phenotypes due to the physiological role of this protein (Gulletta et al., 1983), making its effect unpredictable; further characterization is needed to elucidate the exact role this mutation plays on *n*-butanol tolerance.

Even with the single-end reads we managed to identify and validate two chromosomal rearrangements. One is a translocation of the transposable insertion sequence IS5 within the 3′ end of the *feoA* gene (*feoA201*::*IS5*) shared by both MG5 and MG6. Studies have demonstrated that the insertion of IS5 increases transcription of downstream genes in the same operon (Sawers, 2005; Schnetz and Rak, 1992; Schoner and Kahn, 1981). Since the insertion of the IS5 is located at the end of the *feoA*, the genes *feoB* and *feoC* in the *feoABC* operon can potentially increase their transcript abundances by being transcribed from promoters located inside the transposable element, and bypass the repression by Fur. Indeed, in MG5 and MG6, we observed an upregulation of *feoB* and *feoC*.

The only genotypic difference between MG5 and MG6 is the SNP (C→T) located in the Fur binding site (*FBSmut*) of the operons *fes-ybdZ-entF-fepE* (at 141 bp from *fes*) and *fepA-entD* (at 157 bp from *fepA*) found in MG6 but not in MG5. The SNP was likely the cause of the higher transcriptional activation of the *fepA-entD* operon in MG6 compared to MG5. The data indicates that this mutation is beneficial since the fitness advantage of MG6 is greater than MG5 (as shown in Figs. 2 and 3). Furthermore, the data strongly suggest that the MG6 is direct progeny of MG5.

The other chromosomal rearrangement we identified and validated was a translocation of the 1331 bp transposable insertion sequence IS2 within the 3′ end of the gene *relA* (*relA142*::*IS2*) in the yellow mutant MY2. RelA is the key enzyme involved in the activation of the stringent response. Studies have shown that mutant alleles of *relA* confers temperature-sensitive phenotypes (Yang and Ishiguro, 2003). IS2 has been associated as a controller element, depending on the integration orientation (Saedler et al., 1974). Inclusion of stringent response regulon in NCA confirmed that the stringent response was significantly perturbed in MY2 (data not shown). Further analysis is needed to elucidate the role of *relA* in the tolerance to this solvent.

In the mutant MY4 (yellow lineage) a point mutation was detected in *nusA* (*nusAE212A*). The allele *nusAE212K* has been identified to confer a slower growth phenotype at 30 °C (Craven and Friedman, 1991). NusA interacts with many proteins involved with transcription and termination, including *rpoB*, *rpoC*, *rho*, and core RNAP (Platt, 1986). The presence of two independent mutations affecting transcriptional termination implies that the mutated *rho* and *nusA* potentially increases the *n*-butanol tolerance phenotype. In addition, since mutations in both of these genes have been shown to confer temperature-sensitive phenotypes, there may be a correlation between temperature-sensitivity and *n*-butanol tolerance in *E. coli*.

No SNPs were identified in any of the other mutants. However, as stated earlier, there may be potential genome rearrangements present in these mutants that we were not able to detect.

3.7. Dynamics of allele frequencies during evolution

We subjected the genomic DNA extracted from population samples at different generations to qRT-PCR in order to measure the allele frequencies throughout the course of the evolution. Calibration experiments for each set of allelic primers were conducted using known ratios of genomic DNA with wild-type and mutant alleles; all sets of primers performed well (Supplementary Fig. S3). The allelic data for the population samples are shown in Table 4. The pattern shown for the mutation *rhoS82F*, for samples isolated after the genome shuffling, correlates with the evolution dynamics traced using VERT for the green subpopulation (Fig. 9). This result allows us to conclude that *rhoS82F* was fixed in the green subpopulation shortly after the initiation of the second round of evolution. Based on the fact that *rhoS82F* and *feoA201*::*IS5* were both found in MG5, we expected to see similar frequencies of both alleles in the population samples; however as illustrated in Fig. 9, the frequency of *feoA201*::*IS5* allele was constantly lower than *rhoS82F*. To determine if such difference was due to an actual reduced proportion of *feoA201*::*IS5* or inefficiency of the qRT-PCR process (although the allelic primers performed well in the calibration experiments), the presence of *feoA201*::*IS5* in 16 randomly isolated colonies chosen from generation 72 was assessed. We found that ~93.8% of the whole population contains such mutation (15 out of 16 isolated colonies), which is very close to the frequency of the *rhoS82F* (~92%) at generation 72. Both of these mutations were nearly fixed in the green subpopulation (~97%). However, the reason why a 50% reduction in the proportion of *feoA201*::*IS5* using the qRT-PCR was seen is still unclear. In the case of the mutation *FBSmut*, it appeared in between generations 40 and 48 in the green subpopulation and rapidly became the majority of the population. This result correlates with the higher fitness advantage observed in the mutant MG6 in comparison with MG5 (Fig. 2).

Table 4
The mutant allele frequencies at various time points during evolution.

	rhoS82F (%)	*feoA201*::*IS5* (%)	*FBSmut* (%)	*relA142*::*IS2* (%)	*nusAE212A* (%)
Before genome shuffling (Population P2)					
Gen 8	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.01	0.00 ± 0.00
Gen 40	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.02 ± 0.03	0.00 ± 0.00
Gen 80	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Gen 120	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	4.59 ± 1.48	0.00 ± 0.00
Gen 144	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	18.66 ± 7.08	0.00 ± 0.00
Afer genome shuffling (Population P2)					
Gen 8	64.94 ± 10.05	22.86 ± 3.37	9.74 ± 4.97	0.00 ± 0.00	13.18 ± 1.63
Gen 24	47.89 ± 7.97	15.42 ± 2.88	7.66 ± 1.34	0.00 ± 0.00	17.07 ± 2.15
Gen 32	28.69 ± 5.38	5.66 ± 3.88	8.54 ± 2.22	0.00 ± 0.00	10.06 ± 8.36
Gen 40	23.43 ± 4.60	4.38 ± 2.84	9.73 ± 4.12	0.00 ± 0.00	13.71 ± 5.63
Gen 48	38.84 ± 4.05	7.39 ± 2.84	7.95 ± 0.19	0.00 ± 0.00	3.76 ± 3.70
Gen 56	52.95 ± 0.12	15.47 ± 5.04	58.04 ± 13.37	0.00 ± 0.00	2.02 ± 0.41
Gen 72	93.28 ± 1.67	49.16 ± 2.55	86.79 ± 0.64	0.00 ± 0.00	1.93 ± 0.42
Before genome shuffling (Population P1)					
Gen 144	8.88 ± 3.20	21.99 ± 12.41	0.00 ± 0.00	0.00 ± 0.00	0.05 ± 0.05

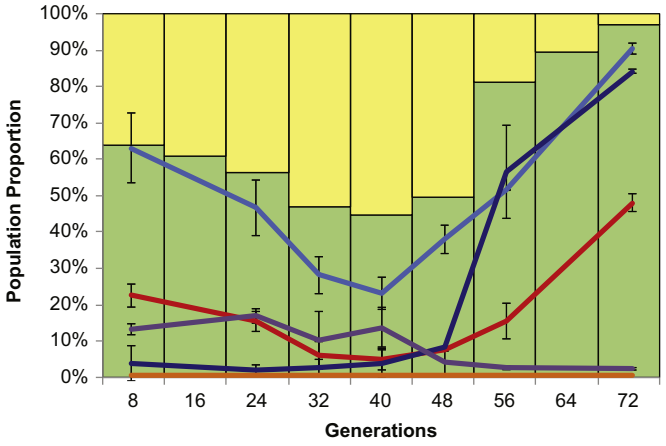

Fig. 9. Mutant allele frequencies in the evolving population samples isolated after genome shuffling. Yellow bars: proportion of the yellow subpopulation. Green bars: proportion of the green subpopulation. Blue line: approximate steady-state concentration of *n*-butanol in the chemostat. Light blue line: frequency of the *rhoS82F* allele. Red line: frequency of the *feoA201*::*IS5* allele. Blue line: frequency of the *FBSmut* allele. Purple line: frequency of the *nusAE212A* allele. Orange line: frequency of the *relA142*::*IS2* allele. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

To determine whether *feoA201*::*IS5* and *rhoS82F* were present in population P2 prior to the genome shuffling, we analyzed the genomic DNA from population samples at various time points prior to genome shuffling. Neither of these mutations was present in P2, indicating that they either came from population P1 or were generated during the genome shuffling process. A quick analysis of the population sample isolated from generation 144 in population P1 showed that the proportion of *rhoS82F* was 8.88 ± 3.20% and *feoA201*::IS5 was 21.99 ± 12.41% prior to genome shuffling.

The mutations identified in the yellow mutants were also tracked throughout the course of the evolution. The data showed that *relA142*::*IS2* (identified in MY2) appeared between generations 80 and 120 during the first round of evolution and was lost from the population after the genome shuffling process, and *nusAE212A* (identified in MY4) was not present in populations P1 or P2 prior to genome shuffling (Table 4). The *nusAE212A* mutation was either in too low of a frequency in the population to be detected using qRT-PCR or may have arisen during the genome shuffling process; it remained in the population during the second round of evolution at a fairly constant proportion until generation 40 and was rapidly lost due to the expansion of MG6 (Fig. 9).

4. Discussion

4.1. Exploring evolution dynamics using VERT

In this study, we used a two-colored VERT system to track the evolutionary dynamics of *E. coli* under increasing concentrations of *n*-butanol. Unlike traditional evolutionary engineering methods, VERT allows adaptive events to be easily identified via visualization of the relative proportions of different fluorescently-labeled cells. It can help to facilitate a more rational scheme for the isolation of adaptive mutants for further characterization and can be broadly applicable to the adaptive evolution of microbial systems. In addition, while it was not employed in this work, a formal schedule for the ramp-up of selective pressure can also be established by the use of VERT.

Since adaptive events are identified via significant changes in the relative proportion of the different colored subpopulation, we developed a computational approach to identify adaptive events and compared it with visual inspection. Adaptive events visually identified at generations 8, 80, and 112 in Fig. 1 were not identified computationally and may therefore represent transient fluctuations. MG1 (isolated from generation 8) showed no significant increase in relative fitness compared to the wild-type (and no genomic changes were identified from whole genome resequencing; see below), and is therefore likely a false positive. All 8 isolates from generation 112 showed negative fitness compared with both the wild-type and the previous adaptive mutant; this may be due to differences in the conditions during the actual evolution and that in the pairwise competition, but does not preclude the possibility of it being a falsely identified adaptive event. An adaptive event apparently occurred in the green subpopulation after generation 32 that gave rise to a mutant that was able to impede the expansion of the yellow subpopulation. This episode led to a meta-stable state, visible from generation 32 to 80, where there are negligible subpopulation proportion changes. Indeed, generation 80 was an actual adaptive event, as MG2 that was isolated from the expanding subpopulation at that generation showed an increase in relative fitness compared to the previous adaptive mutant (MY1). This result was not identified by the computational algorithm due to the lack of significant change in relative proportions of the two colored-subpopulations, which

suggests that MY1 and MG2 may be equally fit under evolution conditions. Based on these results, our computational algorithm has the potential to be more sensitive and accurate compared to visual inspection, but may not be able to identify adaptive events where the fitness differential between populations is small; further development of the computational algorithm is underway to incorporate meta-stable regions into the analysis.

4.2. Role of Fur in n-butanol tolerance

Upon isolation of the adaptive mutants from the evolved populations, transcriptome studies and subsequent NCA analysis revealed differential activation of Fur in the yellow and the green lineages. Fur activity was down in several of the green mutants, and is a potentially newly discovered adaptive mechanism of *E. coli* for *n*-butanol tolerance. We determined that the decreased activation of Fur might have led to an increase in the activity of the BasS/BasR system, which ultimately led to changes in LPS composition. The hypothesis was further supported by the increased resistance to the CAMP, polymyxin B, in MG2, MG5 and MG6, which showed increased gene expression of many iron-related genes.

Some of our mutants exhibited an overexpression of *cfa*, which was further validated to lead to an increase in *n*-butanol tolerance. However, this result seems to contradict previous results in *E. coli* mutants isolated from populations evolved in isobutanol (Minty et al., 2011), where a downregulation in *cfa* was observed. This contradiction may be due to differences in membrane disruptions due to structural differences between the two isomers or the differences in other adaptive mechanisms present in the mutants.

It was clear from our data that certain potential adaptive mechanisms arose independently. For example, we found the upregulation of the biotin biosynthesis genes in both the green and the yellow lineages (MY1 (yellow) and MG2 (green)), suggesting that mutations that lead to an upregulation in biotin biosynthesis are frequently acquired. Unfortunately, we did not identify any SNPs in MY1 and MG2; longer paired-end reads will be needed to identify any potential genome rearrangements in these mutants. As mentioned previously, the upregulation of the iron-related genes were only observed in the green lineage; in fact, several genes related to iron transport were downregulated in two of the yellow adaptive mutants (MY2 and MY3). After the genome shuffling, adaptive mutants isolated from the green subpopulation exhibited increases in the iron-related genes. However, the adaptive mutant isolated from the yellow subpopulation (MY4) showed a decrease in iron-related gene expression, suggesting that genetic interactions are potentially involved in the adaptive mechanism associated with Fur.

Even though several genes we verified to be involved in *n*-butanol tolerance were already upregulated in some of the earlier isolated green mutants, such as *basS*, *entC* and *bioA*, the tolerance levels of these earlier green mutants were not significantly higher than the wild-type strain at 0.8% (v/v) *n*-butanol. It was not until after the genome shuffling did we observe a significant increase in *n*-butanol tolerance. These results further confirm the complex nature of *n*-butanol tolerance in *E. coli*; involvement of multiple genes and potential genetic interactions being involved in conferring the desired phenotype.

5. Conclusions

In this proof-of-principle study, our results showed that while subsequent adaptive mutants have fitness advantages over the previous adaptive mutant, they might not always be better than the wild-type. This type of phenomenon was observed previously with yeast evolved under glucose-limitation (Paquin and Adams, 1983a), and is likely due to the heterogeneities in the evolving population and potential interactions between clones. Thus, this result further highlights the benefit of tracking the adaptive events in a population to isolate each adaptive mutant, as the last one isolated from the population may not necessarily be the best one. By isolating mutants from each observed adaptive event, we are able to identify tolerance mechanisms that may otherwise be missed due to the effects of clonal interference. In addition, by combining VERT with genome shuffling, we were able to significant enhance the rate of adaptation.

Data availability

All raw data is MIAME compliant and have been deposited in the GEO database with accession number GSE30005.

Acknowledgments

The authors wish to thank Albert Lee for helpful comments and discussions and partial financial support from US NSF grants CBET-1032487 (KCK), MCB-1054276 (KCK), and Graduate Research Fellowship program (JW) and the Texas Engineering Experimental Station (TEES).

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.ymben.2012.05.002.

References

Agrawal, M., Mao, Z., Chen, R.R., 2011. Adaptation yields a highly efficient xylose-fermenting Zymomonas mobilis strain. Biotechnol. Bioeng. 108, 777–785.
Aono, R., Kobayashi, H., 1997. Cell surface properties of organic solvent-tolerant mutants of *Escherichia coli* K-12. Appl. Environ. Microbiol. 63, 3637–3642.
Atsumi, S., Cann, A.F., Connor, M.R., Shen, C.R., Smith, K.M., Brynildsen, M.P., Chou, K.J.Y., Hanai, T., Liao, J.C., 2008. Metabolic engineering of *Escherichia coli* for 1-butanol production. Metab. Eng. 10, 305–311.
Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., Mori, H., 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Sys. Biol., 2.
Bagg, A., Neilands, J.B., 1987. Ferric uptake regulation protein acts as a repressor, employing iron (II) as a cofactor to bind the operator of an iron transport operon in *Escherichia coli*. Biochemistry 26, 5471–5477.
Bartlett, M.S., Gourse, R.L., 1994. Growth rate-dependent control of the rrnB P1 core promoter in *Escherichia coli*. J. Bacteriol. 176, 5560–5564.
Basso, T.O., de Kok, S., Dario, M., do Espirito-Santo, J.C., Muller, G., Schlolg, P.S., Silva, C.P., Tonso, A., Daran, J.M., Gombert, A.K., van Maris, A.J., Pronk, J.T., Stambuk, B.U., 2011. Engineering topology and kinetics of sucrose metabolism in Saccharomyces cerevisiae for improved ethanol yield. Metab. Eng. 13, 694–703.
Borden, J.R., Papoutsakis, E.T., 2007. Dynamics of genomic-library enrichment and identification of solvent tolerance genes for clostridium acetobutylicum. Appl. Environ. Microbiol. 73, 3061–3068.
Breitling, R., Armengaud, P., Amtmann, A., Herzyk, P., 2004. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett. 573, 83–92.
Connor, M.R., Liao, J.C., 2009. Microbial production of advanced transportation fuels in non-natural hosts. Curr. Opin. Biotechnol. 20, 307–315.
Conrad, T.M., Lewis, N.E., Palsson, B.O., 2011. Microbial laboratory evolution in the era of genome-scale science. Mol. Sys. Biol. 7, 509.
Cooper, T., Lenski, R., 2010. Experimental evolution with E. coli in diverse resource environments. I. Fluctuating environments promote divergence of replicate populations. BMC Evol. Biol. 10, 11.
Craven, M.G., Friedman, D.I., 1991. Analysis of the *Escherichia coli* nusA10(Cs) allele: relating nucleotide changes to phenotypes. J. Bacteriol. 173, 1485–1491.

Dai, M., Ziesman, S., Thomas, R., Ryan, T.G., Copley, S.D., 2005. Visualization of protoplast fusion and quantitation of recombination in fused protoplasts of auxotrophic strains of *Escherichia coli*. Metab. Eng. 7, 45–52.
Deng, Y., Fong, S.S., 2011. Laboratory evolution and multi-platform genome re-sequencing of the cellulolytic actinobacterium thermobifida fusca. J. Biol. Chem. 286, 39958–39966.
Desai, M.M., Fisher, D.S., 2007. Beneficial mutation-selection balance and the effect of linkage on positive selection. Genetics.
Ferea, T.L., Botstein, D., Brown, P.O., Rosenzweig, R.F., 1999. Systematic changes in gene expression patterns following adaptive evolution in yeast. PNAS 96, 9721–9726.
Fischer, C.R., Klein-Marcuschamer, D., Stephanopoulos, G., 2008. Selection and optimization of microbial hosts for biofuels production. Metab. Eng. 10, 295–304.
Fogle, C.A., Nagle, J.L., Desai, M.M., 2008. Clonal interference, multiple mutations, and adaptation in large asexual populations. Genetics.
Fong, S.S., Joyce, A.R., Palsson, B.O., 2005. Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. Genome Res. 15, 1365–1372.
Gatzeva-Topalova, P.Z., May, A.P., Sousa, M.C., 2005. Crystal structure and mechanism of the *Escherichia coli* ArnA (PmrI) transformylase domain. An enzyme for lipid A modification with 4-Amino-4-deoxy-l-arabinose and polymyxin resistance. Biochemistry 44, 5328–5338.
Gavin, J.J., Umbreit, W.W., 1965. Effect of biotin on fatty acid distribution in *Escherichia coli*. J. Bacteriol. 89, 437 &.
Gerrish, P.J., Lenski, R.E., 1998. The fate of competing beneficial mutations in an asexual population. Genetica 102-3, 127–144.
Goodarzi, H., Bennett, B.D., Amini, S., Reaves, M.L., Hottes, A.K., Rabinowitz, J.D., Tavazoie, S., 2010. Regulatory and metabolic rewiring during laboratory evolution of ethanol tolerance in E. coli. Mol. Syst. Biol., 6.
Grisham, C.M., Barnett, R.E., 1973. The effects of long-chain alcohols on membrane lipids and the (Na++K+)-ATPase. Biochim. Biophys. Acta, Biomembr. 311, 417–422.
Gulletta, E., Das, A., Adhya, S., 1983. The pleiotropic ts15 mutation of E. coli is an IS1 insertion in the rho structural gene. Genetics 105, 265–280.
Hagiwara, D., Yamashino, T., Mizuno, T., 2004. A genome-wide view of the Escherichia coli BasS-BasR two-component system implicated in iron-responses. Biosci. Biotechnol., Biochem. 68, 1758–1767.
Haldimann, A., Wanner, B.L., 2001. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. J. Bacteriol. 183, 6384–6393.
Huang, M., McClellan, M., Berman, J., Kao, K.C., 2011. Evolutionary dynamics of candida albicans during in vitro evolution. Eukaryotic Cell 10, 1413–1421.
Hui, F.K., Barton, P.G., 1973. Mesomorphic behaviour of some phospholip1ds with aliphatic alcohols and other non-ionic substances. Biochim. Biophys. Acta, Lipids Lipid Metab. 296, 510–517.
Kao, K.C., Sherlock, G., 2008. Molecular characterization of clonal interference during adaptive evolution in asexual populations of *Saccharomyces cerevisiae*. Nat. Genet. 40, 1499–1504.
Kao, K.C., Tran, L.M., Liao, J.C., 2005. A global regulatory role of gluconeogenic genes in *Escherichia coli* revealed by transcriptome network analysis. J. Biol. Chem. 280, 36079–36087.
Kim, Y., Orr, H.A., 2005. Adaptation in sexuals vs. asexuals: clonal interference and the Fisher–Muller model. Genetics 171, 1377–1386.
Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., Mori, H., 2005. Complete set of ORF clones of *Escherichia coli* ASKA library (A complete Set of E. coli K-12 ORF archive): unique resources for biological research. DNA Res. 12, 291–299.
Lee, D.-H., Palsson, B.O., 2010. Adaptive evolution of *Escherichia coli* K-12 MG1655 during growth on a nonnative carbon source, L-1,2-Propanediol. Appl. Environ. Microbiol. 76, 4158–4168.
Liao, J.C., Boscolo, R., Yang, Y.L., Tran, L.M., Sabatti, C., Roychowdhury, V.P., 2003. Network component analysis: reconstruction of regulatory signals in biological systems. PNAS 100, 15522–15527.
Lynch, M.D., Warnecke, T., Gill, R.T., 2007. SCALEs: multiscale analysis of library enrichment. Nat. Methods 4, 87–93.
Minty, J., Lesnefsky, A., Lin, F., Chen, Y., Zaroff, T., Veloso, A., Xie, B., McConnell, C., Ward, R., Schwartz, D., Rouillard, J.-M., Gao, Y., Gulari, E., Lin, X., 2011. Evolution combined with genomic study elucidates genetic bases of isobutanol tolerance in *Escherichia coli*. Microb. Cell Factories 10, 18.
Neilands, J.B., 1995. Siderophores: structure and function of microbial iron transport compounds. J. Biol. Chem. 270, 26723–26726.
Nicolaou, S.A., Gaida, S.M., 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. Eng. 12, 307–331.
Notley-McRobb, L., Ferenci, T., 2000. Experimental analysis of molecular events during mutational periodic selections in bacterial evolution. Genetics 156, 1493–1501.
Oh, M.K., Rohlin, L., Kao, K.C., Liao, J.C., 2002. Global expression profiling of acetate-grown *Escherichia coli*. J. Biol. Chem. 277, 13175–13183.
Paquin, C., Adams, J., 1983a. Frequency of fixation of adaptive mutations is higher in evolving diploid than haploid yeast populations. Nature 302, 495–500.
Paquin, C.E., Adams, J., 1983b. Relative fitness can decrease in evolving asexual populations of S. cerevisiae. Nature 306, 368–371.
Platt, T., 1986. Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. 55, 339–372.
Quackenbush, J., 2002. Microarray data normalization and transformation. Nat. Genet.
Rao, L., Ross, W., Appleman, J.A., Gaal, T., Leirmo, S., Schlax, P.J., Record, M.T., Gourse, R.L., 1994. Factor independent activation of rrnB P1: an "Extended" promoter with an upstream element that dramatically increases promoter strength. J. Mol. Biol. 235, 1421–1435.
Reyes, L.H., Almario, M.P., Kao, K.C., 2011. Genomic library screens for genes involved in *n*-butanol tolerance in *Escherichia coli*. PLoS One 6, e17678.
Rutherford, B.J., Dahl, R.H., Price, R.E., Szmidt, H.L., Benke, P.I., Mukhopadhyay, A., Keasling, J.D., 2010. Functional genomic study of exogenous *n*-butanol stress in *Escherichia coli*. Appl. Environ. Microbiol. 76, 1935–1945.
Saedler, H., Reif, H.J., Hu, S., Davidson, N., 1974. IS2, a genetic element for turn-off and turn-on of gene activity in *E. coli*. Mol. Gen. Genet. 132, 265–289.
Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V., Quackenbush, J., 2003. TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34, 374 +.
Sawers, R.G., 2005. Transcript analysis of *Escherichia coli* K-12 insertion element IS5. FEMS Microbiol. Lett. 244, 397–401.
Schnetz, K., Rak, B., 1992. IS5: a mobile enhancer of transcription in *Escherichia coli*. PNAS 89, 1244–1248.
Schoner, B., Kahn, M., 1981. The nucleotide sequence of IS5 from *Escherichia coli*. Gene 14, 165–174.
Shen, C.R., Liao, J.C., 2008. Metabolic engineering of *Escherichia coli* for 1-butanol and 1-propanol production via the keto-acid pathways. Metab. Eng. 10, 312–320.
Sikkema, J., Debont, J.A.M., Poolman, B., 1995. Mechanisms of membrane toxicity of hydrocarbons. Microbiol. Rev 59, 201–222.
Sillers, R., Al-Hinai, M.A., Papoutsakis, E.T., 2009. Aldehyde-alcohol dehydrogenase and/or thiolase overexpression coupled with CoA transferase downregulation lead to higher alcohol titers and selectivity in Clostridium acetobutylicum fermentations. Biotechnol. Bioeng. 102, 38–49.
Smith, K.M., Liao, J.C., 2011. An evolutionary strategy for isobutanol production strain development in *Escherichia coli*. Metab. Eng. 13, 674–681.
Taylor, F.R., Cronan, J.E., 1979. Cyclopropane fatty acid synthase of *Escherichia coli*. Stabilization, purification, and interaction with phospholipid vesicles. Biochemistry. 18, 3292–3300.
Tran, L.M., Brynildsen, M.P., Kao, K.C., Suen, J.K., Liao, J.C., 2005. gNCA: a framework for determining transcription factor activity based on transcriptome: identifiability and numerical implementation. Metab. Eng. 7, 128–141.
Trinh, C.T., Li, J., Blanch, H.W., Clark, D.S., 2011. Redesigning *Escherichia coli* metabolism for anaerobic production of isobutanol. Appl. Environ. Microbiol. 77, 4894–4904.
Warner, J.R., Reeder, P.J., Karimpour-Fard, A., Woodruff, L.B., Gill, R.T., 2010. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. Nat. Biotechnol. 28, 856–862.
Winkler, J., Kao, K.C., 2012. Computational identification of adaptive mutants using the VERT system. J. Biol. Eng. 6, 3.
Winkler, J., Rehmann, M., Kao, K., 2010. Novel *Escherichia coli* hybrids with enhanced butanol tolerance. Biotechnol. Lett. 32, 915–920.
Yang, X., Ishiguro, E.E., 2003. Temperature-sensitive growth and decreased thermotolerance associated with relA mutations in *Escherichia coli*. J. Bacteriol. 185, 5765–5771.
Yang, Y.H., Dudoit, S., Luu, P., Speed, T.P., 2001. Normalization for cDNA microarray data. In: Bittner, M.L., et al., Eds.), microarrays: optical technologies and informatics. vol. 2. SPIE Int. Soc. Opt. Eng., Bellingham, pp. 141–152.
Yao, Z., Valvano, M.A., 1994. Genetic analysis of the O-specific lipopolysaccharide biosynthesis region (rfb) of *Escherichia coli* K-12 W3110: identification of genes that confer group 6 specificity to Shigella flexneri serotypes Y and 4a. J. Bacteriol. 176, 4133–4143.