Molecular Systems Biology 6; Article number 378; doi:10.1038/msb.2010.33
Citation: *Molecular Systems Biology* 6:378
© 2010 EMBO and Macmillan Publishers Limited All rights reserved 1744-4292/10
www.molecularsystemsbiology.com

molecular systems biology

Regulatory and metabolic rewiring during laboratory evolution of ethanol tolerance in *E. coli*

Hani Goodarzi[1], Bryson D Bennett[2], Sasan Amini[1], Marshall L Reaves[2], Alison K Hottes[1], Joshua D Rabinowitz[2] and Saeed Tavazoie[1,*]

[1] Department of Molecular Biology, Princeton University, Princeton, NJ, USA and [2] Department of Chemistry and Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA
* Corresponding author. Department of Molecular Biology, Princeton University, Carl Icahn Bldg. Rm. 245, Washing St., Princeton, NJ 08544, USA. Tel.: +1 609 258 0331; Fax: +1 609 258 5738; E-mail: tavazoie@genomics.princeton.edu

Received 4.2.10; accepted 18.4.10

Understanding the genetic basis of adaptation is a central problem in biology. However, revealing the underlying molecular mechanisms has been challenging as changes in fitness may result from perturbations to many pathways, any of which may contribute relatively little. We have developed a combined experimental/computational framework to address this problem and used it to understand the genetic basis of ethanol tolerance in *Escherichia coli*. We used fitness profiling to measure the consequences of single-locus perturbations in the context of ethanol exposure. A module-level computational analysis was then used to reveal the organization of the contributing loci into cellular processes and regulatory pathways (e.g. osmoregulation and cell-wall biogenesis) whose modifications significantly affect ethanol tolerance. Strikingly, we discovered that a dominant component of adaptation involves metabolic rewiring that boosts intracellular ethanol degradation and assimilation. Through phenotypic and metabolomic analysis of laboratory-evolved ethanol-tolerant strains, we investigated naturally accessible pathways of ethanol tolerance. Remarkably, these laboratory-evolved strains, by and large, follow the same adaptive paths as inferred from our coarse-grained search of the fitness landscape.
Molecular Systems Biology **6**: 378; published online 8 June 2010; doi:10.1038/msb.2010.33
Subject Categories: functional genomics; metabolic and regulatory networks
Keywords: adaptation; ethanol tolerance; evolution; fitness profiling

This is an open-access article distributed under the terms of the Creative Commons Attribution Noncommercial Share Alike 3.0 Unported License, which allows readers to alter, transform, or build upon the article and then distribute the resulting work under the same or similar license to this one. The work must be attributed back to the original author and commercial use is not permitted without specific permission.

Introduction

Microbial organisms are constantly adapting to environmental changes. When perturbations are limited to those commonly encountered in the native habitat, physiological processes allow rapid adaptation through both homeostatic and predictive behaviors (Tagkopoulos *et al*, 2008). However, environmental perturbations beyond the structure of the native habitat, set the stage for the emergence of fitter mutants through mutation and natural selection (Yokoyama, 2002). Revealing the genetic basis of adaptation to extreme environments is a formidable challenge due to the potential involvement of many cellular components and pathways. Evolution of ethanol tolerance—the capacity to grow at high concentrations of ethanol—represents an ideal model system for studying such adaptation, which at the same time has significant implications for commercialization of bioethanol as an environmentally sustainable source of energy (Zaldivar *et al*, 2001). As a byproduct of fermentation, ethanol is thought to cause toxicity through effects on membrane integrity (Dombek and Ingram, 1984; Ingram, 1986), the activity of membrane bound and soluble enzymes (Ingram, 1976; Nagodawithana *et al*, 1977; Millar *et al*, 1982), and proton flux balance across the membrane (Cartwright *et al*, 1986; D'Amore *et al*, 1990). No single-genetic modification can substantially increase the level of ethanol tolerance, suggesting the involvement of multiple pathways (D'Amore and Stewart, 1987).

The first whole-genome attempt to discover the genetic modifiers of ethanol tolerance involved comparing the gene-expression levels in a laboratory-evolved ethanol-tolerant strain to its parental ethanologenic strain (Gonzalez *et al*, 2003). However, due to the number of generations needed to reach the maximal tolerance level obtainable in laboratory timescales, the observed changes in the expression levels may be phenotypically neutral (e.g. *marAB* upregulation in the tolerant strain; Gonzalez *et al*, 2003). In addition, many alterations are most likely targeting other aspects of adaptation to the media and not ethanol *per se*. Furthermore, given that a

single-tolerant strain may not adopt all possible strategies for higher tolerance, this approach may fail to provide a comprehensive genetic portrait of ethanol tolerance.

In this study, we have used a whole-genome experimental and computational framework to dissect the adaptive mechanisms of ethanol tolerance in *Escherichia coli*. To systematically identify the loci that positively or negatively contribute to ethanol tolerance, we measured the fitness consequences of gene-level perturbations through whole-genome fitness profiling (Girgis *et al*, 2007). We used high-coverage mutant libraries to profile the effects of single-genetic perturbations (i.e. suppression or overexpression) on the growth rate of wild-type *E. coli* in the presence of ethanol (Figure 1A). Our study revealed many potential target loci with large or small effects on relative growth; however, genes rarely work in isolation and the contribution of the identified loci are not necessarily independent. Thus, to increase our analytical sensitivity, we used a modular computational framework to systematically identify the cellular components and pathways whose modifications are beneficial or detrimental to higher levels of ethanol tolerance.

We found ethanol tolerance to be affected by a diverse range of genetic modules, including stress response pathways (e.g. osmotolerance and acid stress response), metabolic processes (e.g. aerobic respiration), and structural components (e.g. cell wall and fimbriae). We subsequently tested whether the identified modules act independently or interact as part of an 'ethanol-tolerance' pathway. As a result, we discovered intracellular ethanol degradation as a potential adaptive mechanism for ethanol tolerance in *E. coli*. The functional relevance of the discovered pathways was then assessed through metabolic concentration measurements (Fiehn, 2001; Nicholson *et al*, 2002; Wikoff *et al*, 2007) and stable-isotope labeling (Sauer, 2006; Yuan *et al*, 2006) of laboratory-evolved ethanol-tolerant strains (Figure 1B). Using liquid chromatography coupled tandem-mass spectrometry (LC-MS/MS), we discovered that naturally evolved ethanol tolerance benefits from the contribution of the pathways we have identified. In what follows, we detail the framework used to identify the associated pathways along with their potential functions in bringing about a higher level of ethanol tolerance in *E. coli*. Our results suggest that the combination of whole-genome

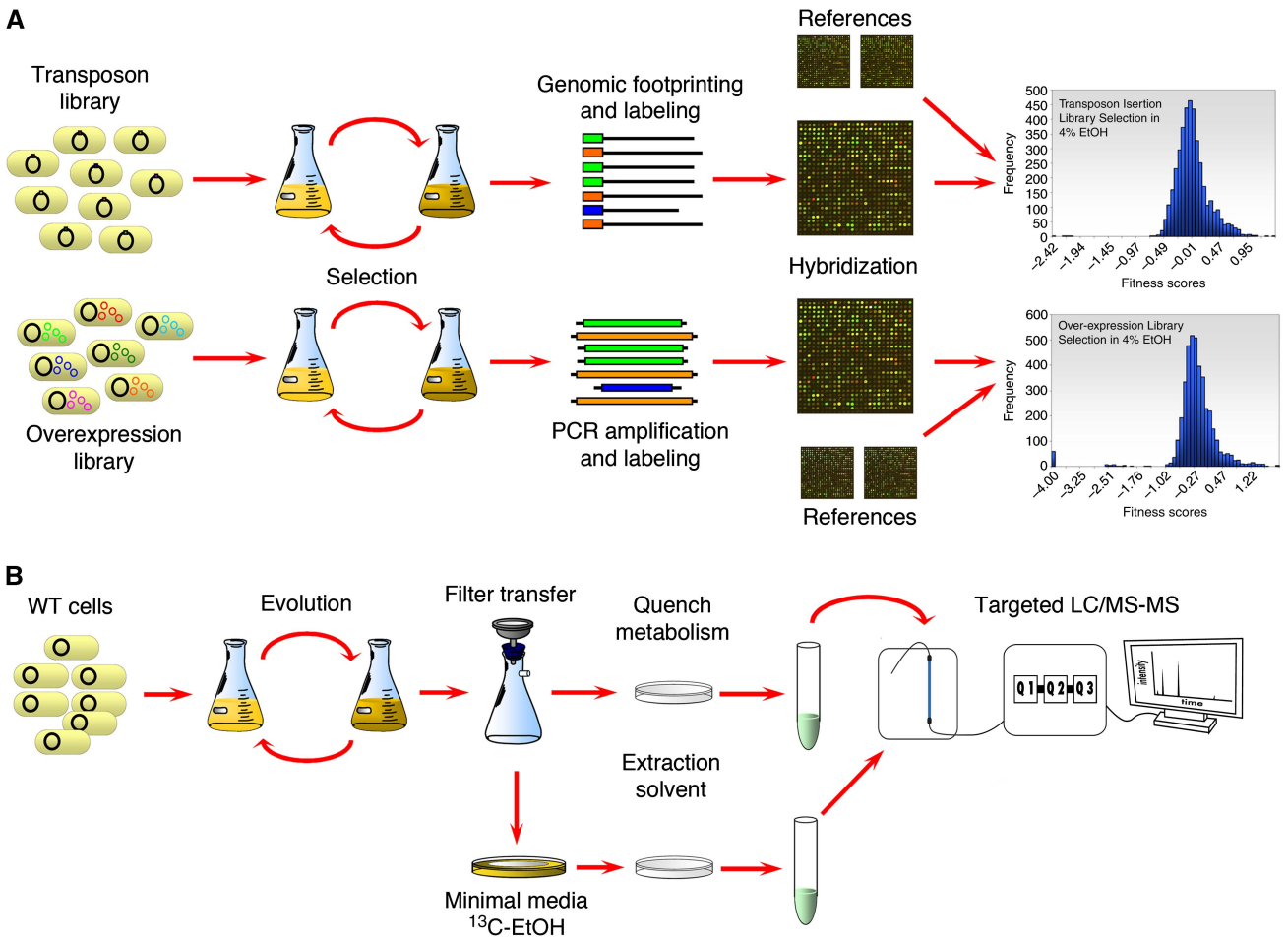

Figure 1 A schematic representation of the overall strategy. (**A**) Starting from the transposon insertion or overexpression libraries, we enriched for relative increase in ethanol tolerance through several rounds of selection in the presence of ethanol. The changes in the frequency of each mutant in the selected versus unselected samples were then translated into a whole-genome fitness profile. (**B**) The wild-type strain (MG1655) was grown in minimal media plus glucose and exogenous ethanol to select for increased ethanol tolerance. The resulting tolerant strains were subjected to metabolomic analyses for measuring changes in the steady-state levels of intracellular metabolites. In a parallel experiment, stable-isotope labeling (^{13}C-ethanol) was used to test ethanol assimilation as a possible mechanism for tolerance.

 © 2010 EMBO and Macmillan Publishers Limited

fitness profiling and metabolite concentration and flux measurements is a powerful framework for studying adaptive evolution to extreme environments.

Results

A coarse-grained fitness landscape of ethanol tolerance

Starting from a comprehensive transposon mutant library (Girgis *et al*, 2007), we used rich media (LB) plus ethanol (4 and 5.5% v/v) to select for mutants with higher levels of ethanol tolerance. For wild-type *E. coli* strain MG1655, ethanol concentrations higher than 6% v/v in rich media resulted in complete growth inhibition. Thus, our selections included both 4% v/v (mild) and 5.5% v/v (harsh) ethanol concentrations to capture different toxicity levels (see Supplementary Figure S1). The frequency of insertions in each locus (both in the selected samples and unselected controls) was then determined through a microarray-based genetic footprinting approach (Girgis *et al*, 2007). In genetic footprinting, we selectively amplify the sequence adjacent to the transposon insertion site, which subsequently serves as a tag for its identification (Badarinarayana *et al*, 2001). A microarray-based quantification of these tags is then used to measure transposon insertion frequencies as a function of the hybridization signal at each locus across the population (Girgis *et al*, 2007). After several rounds of selection ($\sim$5–10 generations), a fitness score is assigned to each locus based on its associated hybridization signal in the selected versus unselected samples (see Materials and methods for details). As transposon mutagenesis typically results in gene inactivation, genes that when disrupted decrease fitness in ethanol, have negative fitness scores. In other words, the loci with negative scores are beneficial to higher tolerance, whereas the ones with positive scores have an adverse impact on growth in ethanol.

To capture the genes that may be essential or affect general growth as well as ethanol tolerance, we also used a pBR322-based overexpression library in which the bacteria carry 1–3 kb random fragments of the *E. coli* genome cloned into a pBR322 vector (Amini *et al*, 2009). This overexpression library was similarly selected in the presence of ethanol (4 or 5.5% v/v) and the changes in the frequency of the overexpressed loci were subsequently determined through cloning site amplification and microarray hybridization (see Materials and methods). Similar to the transposon library, the hybridization signals were translated into fitness scores by comparing the selected and unselected samples. In this case, however, the beneficial loci have positive scores resulting in a positive correlation between the fitness scores and ethanol tolerance.

Detecting pathways and cellular components involved

On determining the fitness scores associated with each locus in the two libraries (transposon insertion and overexpression) under both conditions (4 and 5.5% v/v ethanol), we sought to identify the genes that significantly affect the ethanol tolerance capacity of *E. coli*. However, due to the limited effect of single-gene perturbations on ethanol tolerance, few genes passed our gene-level statistical threshold. Although the fitness effects can be accentuated through increasing the number of generations, the occurrence of beneficial spontaneous mutations during selection can adversely affect the quality of fitness profiles. Thus, to boost the sensitivity of our approach without increasing the number of generations, we used a module-level analysis of these whole-genome fitness profiles. To this end, we combined the data from GO annotations (Ashburner *et al*, 2000), transcription factor regulons (Salgado *et al*, 2006), and known stress response pathways (Storz and Hengge-Aronis, 2000) to compile predefined gene sets representing the prominent modules in the *E. coli* genome. Starting from these gene sets, we subsequently used a mutual-information-based approach (termed iPAGE; Goodarzi *et al*, 2009a) to discover the genetic modules that are significantly informative of our fitness profiles. In this approach, we sorted and quantized the fitness scores in each sample into equally populated bins (10 bins in this case) where each gene is assigned to a single bin. Then, for every module, we calculated the mutual information (Cover and Thomas, 2006) between the quantized fitness profile and the module-membership profile across all the genes (see Materials and methods; Supplementary Figure S2). On the basis of their mutual information values, the significantly informative modules were identified and their enrichment/depletion patterns were visualized through a heat map (see Materials and methods). The most prominent modules emerging from this analysis can be seen in Figure 2. Our results imply that the genes active in propionate catabolism (PrpR regulon), glycine cleavage complex (GcvA regulon), and glycine-betaine synthesis (BetI regulon) boost ethanol tolerance capacity (low scores in transposon libraries and high scores in overexpression samples), whereas fimbriae and acid stress response genes have a significant negative contribution (high scores in transposon libraries and low scores in overexpression samples). Similarly, heat-shock stress response and cell-wall biogenesis pathways are significantly beneficial for ethanol tolerance; as are the genes involved in aerobic respiration (i.e. FNR/ArcA regulons). Here, we have focused on the modules that, in addition to their enrichments among the genes with significant fitness consequences, are also discovered in multiple samples across both transposon and overexpression libraries. In what follows, we discuss these pathways and their potential functions in ethanol tolerance.

Contribution from stress response pathways

Ethanol alters the physical characteristics of the aqueous environment, thus perturbing protein folding both in the cytoplasm and periplasm (Ingram and Buttke, 1984). Ethanol is one of the most powerful elicitors of the heat-shock stress response (Neidhardt and VanBogelen, 1987; Thomas and Baneyx, 1997) and a known activator of the envelope stress response (Storz and Hengge-Aronis, 2000). Indeed, our framework has successfully captured the crucial function of these two pathways in attenuating the adverse effects of ethanol on protein folding (Figure 2). Ethanol tolerance is also affected by the concentration of osmoprotectants inside the cell (D'Amore and Stewart, 1987). Compatible solutes involved in osmoregulation, for example glycine and glycine-betaine in *E. coli* LY01 (an ethanologenic derivative of *E. coli* B) and trehalose and glycerol in yeast strains,

© 2010 EMBO and Macmillan Publishers Limited

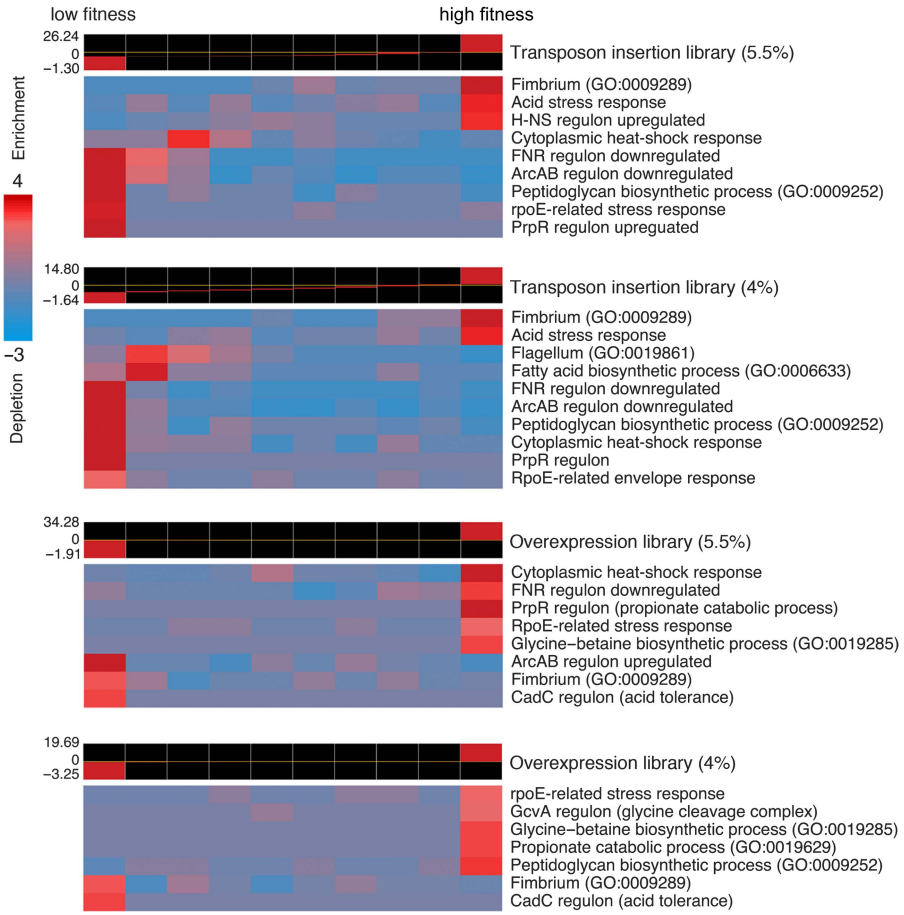

Figure 2 Modular analysis of fitness profiles. The genes in each sample are quantized into 10 equally populated bins based on their fitness scores. These bins are then ordered from lowest fitness (left) to highest fitness (right). For each sample, the red bar in the top panel shows the range of fitness scores contained in each bin. A set of significantly informative (empirical *P*-value <0.005) modules are listed along with a heat map representing their enrichment/depletion patterns across all the bins. In these heat maps, red and blue entries represent the log(*P*-values) calculated for every fitness bin and module using the hypergeometric distribution, highlighting the enrichment/depletion pattern of each module across the fitness range (see Materials and methods). Note that in transposon insertion libraries, the modules that favor growth in ethanol are enriched in lower fitness scores, whereas those with adverse effects are enriched in higher bins. In case of the overexpression libraries, the opposite trend occurs.

are known enhancers of ethanol tolerance (Mansure *et al*, 1994; Gonzalez *et al*, 2003). Consistently, we discovered that glycine and glycine-betaine synthetic genes (GcvA and BetI regulons) substantially contribute to ethanol tolerance in MG1655, presumably through higher production of osmolytes. Given its size and polarity, ethanol readily permeates through the membrane, rendering this compound as an unlikely elicitor of the osmotic shock response. However, it has been shown that cells are more sensitive to osmotic stress in the presence of ethanol (Gonzalez *et al*, 2003). This effect may be attributed to higher membrane fluidity induced by ethanol, which accentuates the effects of osmotic pressures through membrane leakage. Also, the fact that heat shock, ethanol, and osmotic stress similarly activate the envelope stress response and the phage shock protein pathways indicates membrane fluidity as the common target of these stresses (Rowley *et al*, 2006). The function of compatible solutes in neutralizing the effects of ethanol has been extensively studied (for review see Hallsworth, 1998). For example, trehalose can inhibit the leakage induced by ethanol in both intact yeast cells and liposomes (Mansure *et al*, 1994).

Remarkably, we also discovered that the acid stress response pathway (Foster, 2004) antagonizes ethanol tolerance (Figure 2). We observed that the overexpression of the genes in this pathway increases ethanol sensitivity (Figure 3A). To further validate this effect, we made a partial deletion of the acid fitness island (Δ*afi*: b3506–b3511), which includes four of the genes presented in Figure 3A (Mates *et al*, 2007), and found that in comparison with wild type, the resulting strain shows a significantly increased survival rate in 7% (v/v) ethanol (*P*-value <0.001; Figure 3B).

The function of structural components

We also found a number of structural components with significant positive or negative effects on ethanol tolerance. We were not surprised to find that the cell-wall biogenesis pathway is crucial for ethanol tolerance given its function in supporting membrane integrity. A number of peptidoglycan biosynthesis genes show beneficial contributions to ethanol tolerance (Figure 4A). We also observed that *slt*, which encodes a murein-degrading 'soluble lytic transglycosylase'

© 2010 EMBO and Macmillan Publishers Limited

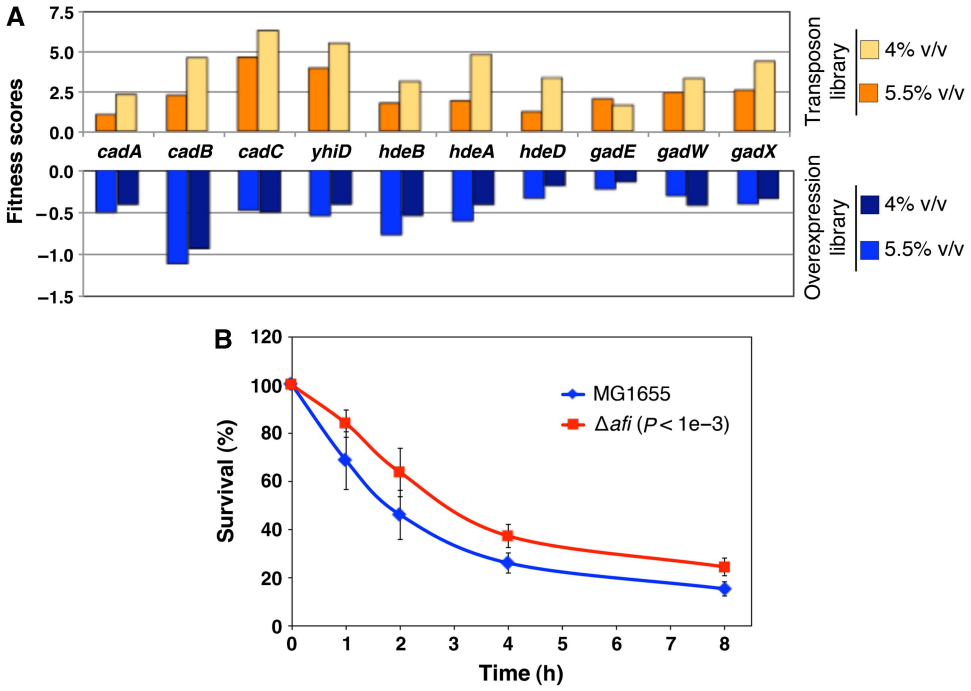

Figure 3 Acid stress response and ethanol tolerance. (**A**) The fitness scores of the genes active in acid tolerance are shown here for all the samples. Disrupting the activity of these genes is beneficial for ethanol tolerance, whereas their overexpression is deleterious. (**B**) We used kill curves (i.e. the number of CFUs at each timepoint) to compare the ethanol tolerance of the wild-type strain and Δ*afi* strain (containing a partial deletion in the acid fitness island) in 7% v/v ethanol (left). Δ*afi*, an acid-sensitive strain, showed a significant increase in ethanol tolerance (*P*-value < 1e−3). The experiments were performed in triplicate, and the error bars mark the minimum and maximum for each point.

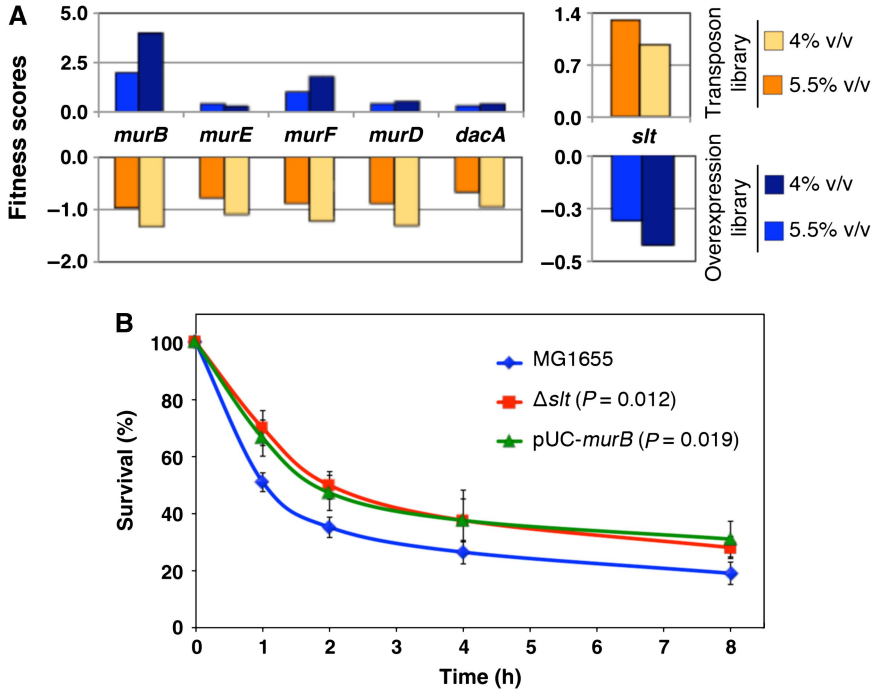

Figure 4 Cell-wall biogenesis and ethanol tolerance. (**A**) The fitness scores assigned to the peptidoglycan biosynthesis genes (*mur* genes) and a murein-hydrolyzing enzyme (*slt*) are shown across all four samples. (**B**) As shown in these kill curves, deleting *slt* (*P*-value < 0.012) or overexpressing *murB* (*P*-value = 0.019) significantly increases the level of ethanol tolerance compared with the MG1655 strain. The experiments were performed in triplicate with error bars showing the minimum and maximum values for each point.

© 2010 EMBO and Macmillan Publishers Limited

(Engel *et al*, 1991) negatively affects ethanol tolerance (Figure 4A). To validate these observations, we showed that *E. coli* strains harboring a plasmid overexpressing *murB*, the enzyme that catalyzes the production of UDP-GlcNAc-enolpyruvate, or those lacking the *slt* gene show higher survival rates at 7% (v/v) ethanol compared with the wild-type strain (Figure 4B). We also discovered that null mutations in fimbrial and fimbrial-like genes significantly increase ethanol tolerance (see Supplementary Figure S3). The lower expression of the non-essential periplasmic proteins, including fimbriae, may help the cell cope with its envelope stress. The structural strain imposed on the membrane by these components may also result in membrane leakage or breakage.

Changes in the lipid composition of the membrane in response to ethanol stress has been extensively studied (Ingram, 1977). In *E. coli* CSH2, lipids with unsaturated fatty acids increase in frequency, as a result of saturated fatty acid synthesis inhibition (Buttke and Ingram, 1980). As shown in Figure 2, transposon insertion events in the fatty acid biosynthetic genes result in loss of fitness in 4% v/v ethanol. This observation further highlights the function of membrane composition in mounting a response against ethanol. However, the absence of this pathway in the 5.5% sample also signifies the slow nature of this adaptive process.

Ethanol tolerance through ethanol degradation and assimilation

In addition to osmoregulatory transcription factors, we also identified other regulatory proteins with significant contributions to ethanol tolerance. The key regulators we identified include FNR/ArcA, PrpR (Figure 2), and CafA (Figure 5A). FNR and ArcA, controllers of the aerobic to anaerobic switch (Green and Paget, 2004), largely regulate the central carbon metabolism enzymes. On the other hand, *cafA* codes for ribonuclease G, which is involved in rRNA processing (Umitsuki *et al*, 2001). We asked whether the contributions from these loci are additive by combining deletions in *fnr*, *arcA*, and *cafA*. These deletions, which individually increase ethanol tolerance, result in a large cumulative effect (Figure 5B). Prior studies had shown a decrease in the *fnr* transcript level in the ethanol-tolerant strain LY01; however, this was hypothesized to contribute through increased osmoprotection (Gonzalez *et al*, 2003). Our observations in MG1655, on the other hand, suggest that the activity of central metabolic enzymes, as part of the FNR/ArcA regulon, is the key contributor to ethanol tolerance (Figure 5A).

Although FNR and ArcA are the transcriptional regulators of the respiratory proteins, CafA is a post-transcriptional regulator that mainly functions in rRNA processing. A CafA

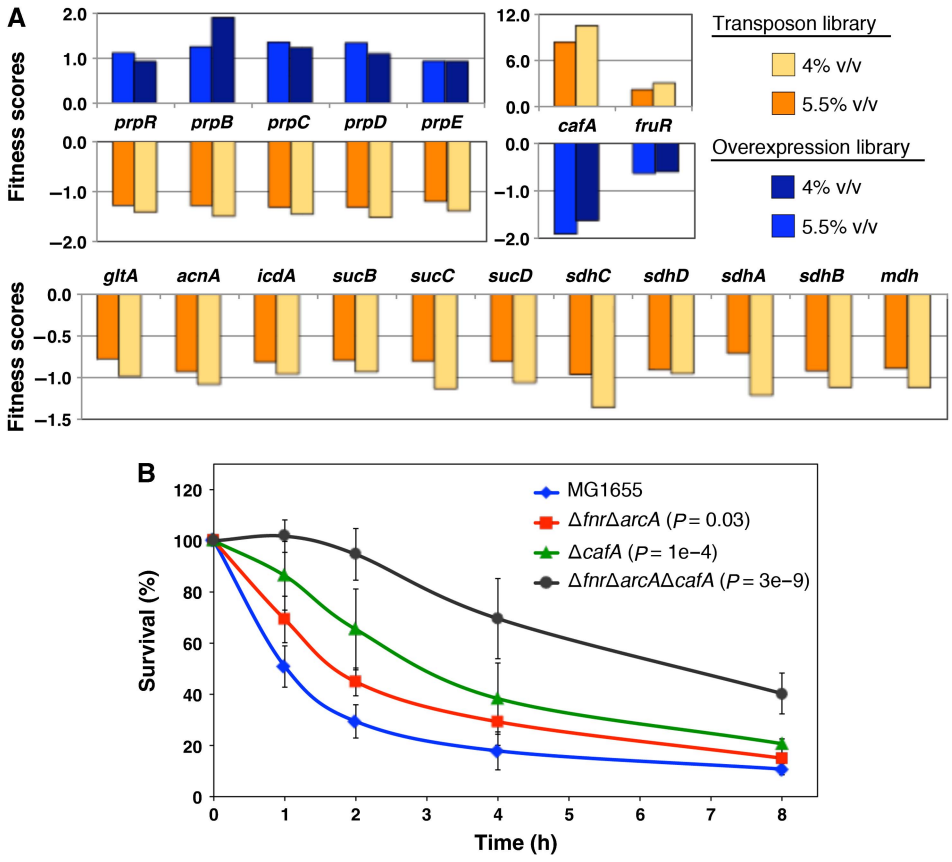

Figure 5 Ethanol degradation as a mechanism for tolerance. (**A**) The fitness profile of the genes active in the TCA cycle along with propionate catabolism genes (*prp* genes) and the negative regulators of alcohol dehydrogenase (*cafA* and *fruR*). (**B**) *fnr/arcA* and *cafA* deletions, both separately and in combination, significantly boost the ethanol tolerance capacity of *E. coli*. These experiments were performed in triplicate. Error bars mark the minimum and maximum values.

© 2010 EMBO and Macmillan Publishers Limited

knockout strain (Δ*cafA*) does not elicit a growth defect under normal conditions, largely due to the activity of ribonuclease E whose function in part overlaps with that of CafA (Ow *et al*, 2003). Consequently, we focused on the genes that are regulated by CafA and not ribonuclease E as key potential players in ethanol tolerance. CafA specifically downregulates *adhE* through *in vivo* mRNA degradation and in the absence of this ribonuclease, the mRNA half-life of *adhE* increases by 2.5-fold. (Umitsuki *et al*, 2001). *adhE* codes for the fermentative alcohol dehydrogenase, which converts acetyl-coenzyme A to ethanol under anaerobic conditions. Interestingly, in addition to *cafA*, overexpression of the transcription factor FruR—which negatively regulates *adhE*—also shows a deleterious effect on ethanol tolerance, whereas its disruption is to some extent beneficial (Figure 5A).

In total, our observations suggest that the ethanol-tolerant mutant Δ*fnr*Δ*arcA*Δ*cafA* (Figure 5B) has a higher level of AdhE and a more active aerobic respiration apparatus. This led us to hypothesize that high levels of ethanol tolerance may be reached through breakdown of ethanol to acetyl-coenzyme A (acetyl-CoA) by the reversible enzyme AdhE and its subsequent assimilation into the TCA cycle. This hypothesis is strengthened by the advantageous effects of overexpressing propionate catabolic genes (*prp* operon in Figure 5A), which replenish the carbon backbone of the TCA cycle through succinate biosynthesis (Palacios and Escalante-Semerena, 2000). In addition, we also observed that exogenous addition of succinate to the media slightly enhances ethanol tolerance (see Supplementary Figure S4). Under normal conditions, wild-type *E. coli* is not capable of significant ethanol degradation. In particular, AdhE is largely inactive under aerobic conditions due to the expression of its negative regulators (Membrillo-Hernandez *et al*, 2000). However, through a combination of mutations, an *E. coli* strain capable of growing on ethanol as a sole source of carbon and energy was successfully evolved in the laboratory (Membrillo-Hernandez *et al*, 2000). Thus, enhancing ethanol degradation to decrease intracellular ethanol concentration seems to be a viable mechanism for ethanol tolerance. In fact, in organisms capable of ethanol detoxification, active alcohol dehydrogenases have been associated earlier with ethanol tolerance (e.g. *Kluyveromyces lactis*; Heipieper *et al*, 2000).

Laboratory-evolved ethanol-tolerant strains use naturally occurring perturbations to the discovered pathways

Our systematic genetic approach helped us acquire a broad understanding of the pathways associated with ethanol tolerance. We next sought to investigate whether laboratory-evolved ethanol-tolerant strains use perturbations in the same pathways identified here. To this end, we used laboratory experimental evolution in media containing exogenously added ethanol to select for mutations that confer higher levels of ethanol tolerance.

We first tested the anti-correlation observed between ethanol tolerance and acid resistance. We grew wild-type *E. coli* for 80 generations in rich media plus high concentrations of ethanol, which resulted in strains capable of growing

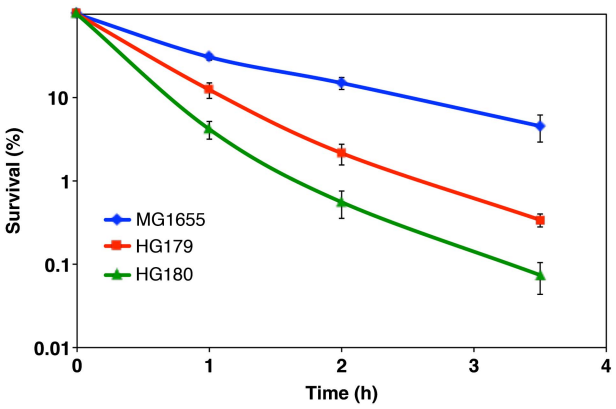

Figure 6 Acid tolerance in laboratory-evolved ethanol-tolerant strains. Two laboratory-evolved ethanol-tolerant strains, HG179 and HG180, showed a highly significant decline in their survival in acidic environments (pH=3.0) compared with the wild-type strain.

in 7.0% (v/v) ethanol (strains HG179 and HG180). We then assayed the activity of the acid response pathway in these ethanol-tolerant backgrounds through measuring their survival in LB with a low pH (pH=3.0). As shown in Figure 6, these strains were at least an order of magnitude more sensitive to low pH than the wild-type strain MG1655.

For testing the metabolic aspects of ethanol tolerance, however, LB is not the medium of choice, as the compounds already present in the medium interfere with metabolite measurements. To measure the metabolic alterations in ethanol-tolerant backgrounds, we evolved wild-type *E. coli* (strain MG1655) in minimal media plus glucose in the presence of increasing concentrations of ethanol. We focused our analysis on two timepoints along the evolutionary trajectory, one early (HG227: ∼30 generation in 3% ethanol) and one late (HG228: ∼160 generations in 5% ethanol). We then used LC-MS/MS to measure the metabolite pool sizes in HG227 and HG228 and compare them to the wild-type strain. For HG227, the intracellular pool sizes of glycolytic compounds are largely unaffected. On the other hand, in HG228, a strain with higher ethanol tolerance, our results support a highly active TCA cycle: we observed an increase in the concentration of many of the TCA cycle components including citrate, succinate, and fumarate (see Supplementary Figure S5).

In addition to metabolite measurements, we also performed gene-expression profiling to compare transcript abundances in HG228 and WT. The expression values serve as additional information for analyzing the metabolite pool sizes. Remarkably, and consistent with our prior discoveries, HG228 shows a significant downregulation in acid stress response genes accompanied by upregulations in peptidoglycan biosynthesis, glycine cleavage system, ArcA regulon, and heat-shock stress response pathways (see Supplementary Figure S6). As mentioned in the earlier sections, the cell-wall biogenesis pathway is beneficial for ethanol tolerance. Both ethanol-tolerant strains HG227 and HG228 show significant reductions in the steady-state concentrations of UDP-glucose and UDP-*N*-acetyl-glucosamine, whereas HG227 also shows a significant reduction in UDP-glucuronate, indicating an increase in their consumption by peptidoglycan and colanic acid biosynthesis

© 2010 EMBO and Macmillan Publishers Limited

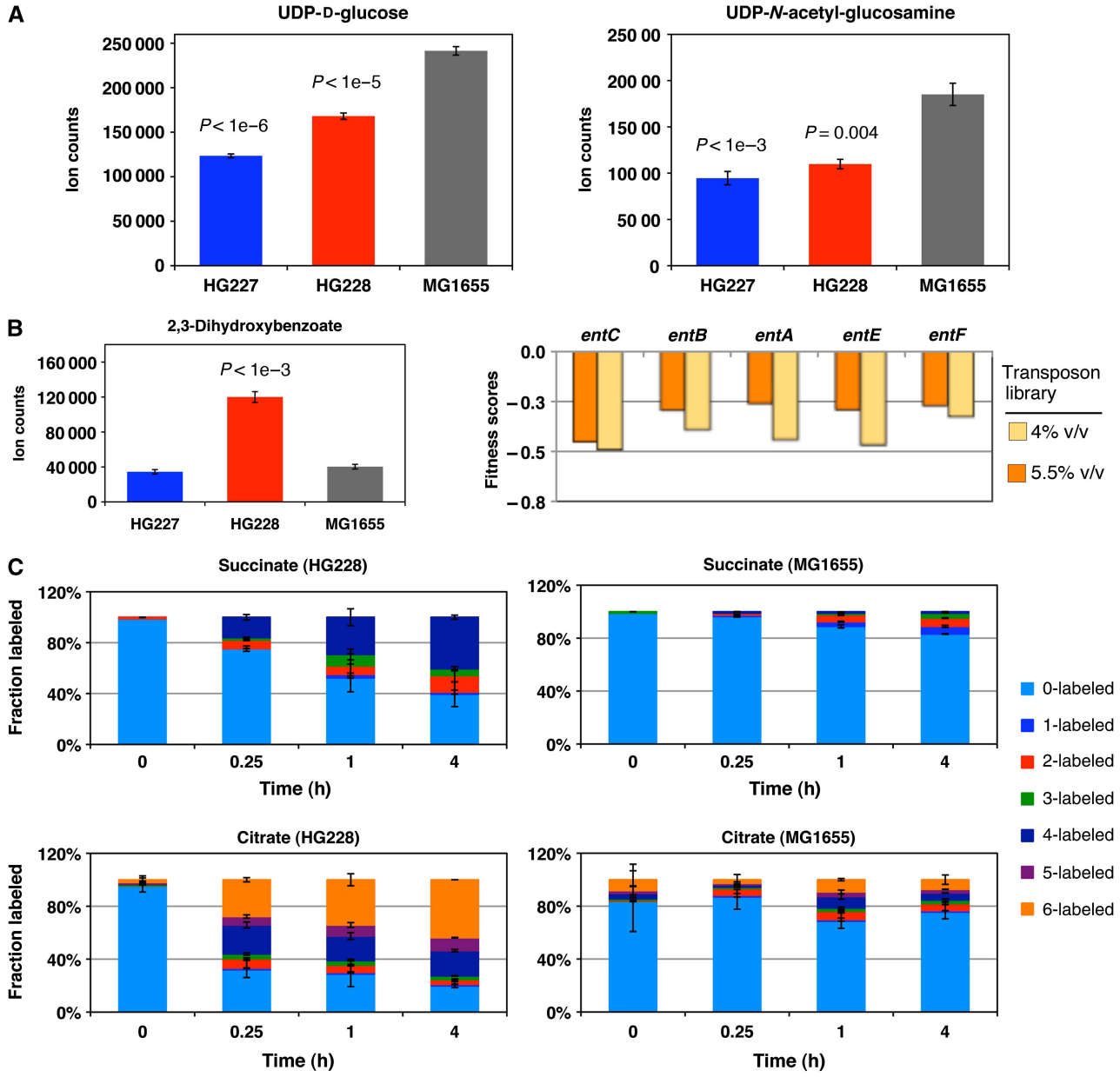

Figure 7 Metabolite concentration measurements and stable-isotope labeling of ethanol-tolerant strains. (**A**) The steady-state level of UDP-glucose and UDP-*N*-acetyl-glucoseamine in ethanol-tolerant strains (HG227 and HG228) compared with wild-type MG1655. The samples were assayed in quadruplicate, and the error bars mark the s.d. (**B**) A significantly higher concentration of 2,3-dihydroxybenzoate was observed in HG228 compared with wild type. Also, the genes in the *ent* pathway have negative fitness scores in the fitness profiling of transposon insertion libraries. (**C**) Shown are the fraction of citrate/isocitrate or succinate metabolites that are labeled with carbon from the ^{13}C-ethanol that was added at time 0 in both HG228 and wild-type strain MG1655. The samples were measured in triplicate with the error bars marking the s.d.

(Figure 7A). The higher expression of genes functioning in peptidoglycan biosynthesis pathway in HG228 further underlines an increase in cell-wall biogenesis.

Moreover, HG228 showed a significant increase in 2,3-dihydroxybenzoate concentration, the only component of the enterobactin biosynthesis pathway we were able to measure (Figure 7B). Enterobactin, a high-affinity siderophore and a component of iron acquisition pathways in *E. coli*, is essential for the activity of many enzymes including the respiratory complexes (Sprencel *et al*, 2000). Our fitness profiling results also show beneficial contributions from the enterobactin biosynthesis pathway (namely the *ent* genes), consistent with iron acquisition being required to support biosynthesis of enzymes for oxidative metabolism (Figure 7B). Accordingly, in HG228, *entA*, *entB*, and *entH* show a significant 24% increase in their expression accompanied by a slight but significant increase (~10%) in the transcript level of *entC* and *entE*.

In HG228, an increase in TCA cycle metabolites (e.g. citrate and succinate) and a high free CoA to acetyl-CoA ratio suggests the capacity for metabolism of ethanol to acetyl-CoA.

 © 2010 EMBO and Macmillan Publishers Limited

Moreover, almost all the TCA cycle genes are upregulated significantly as part of the ArcA regulon (Supplementary Figure S6). To test for ethanol degradation in this strain, we measured its ability to assimilate ethanol in comparison with the wild type. On addition of ^{13}C-ethanol, we used LC-MS/MS to detect the fraction of labeled key metabolites within central metabolism at 0, 0.25, 1, and 4 h timepoints. Our goal was to compare the rate at which ^{13}C gets incorporated into different metabolite pools in HG228 and wild type, thus eliciting an ethanol degradation pathway in HG228. Figure 7C shows the label composition of citrate/isocitrate and succinate across these timepoints. In the case of citrate/isocitrate, whereas wild-type strain MG1655 showed $<10\%$ of the molecules as labeled, HG228 showed $\sim 40\%$ as six-labeled (fully labeled citrate) after 4 h. As shown in Figure 7C, a significant increase in the fraction of labeled metabolites was also observed for succinate where $>40\%$ of the pool is fully labeled in HG228, whereas $<20\%$ is detected as labeled in wild type (Figure 7C). Other metabolites in (and near) central carbon metabolism pathways were also significantly labeled in HG228 compared with the wild type (for additional metabolites see Supplementary Figure S7). We also confirmed that knocking out *adhE* in this background results in a significant decrease in ethanol tolerance (see Supplementary Figure S8). These results indicate that the ethanol-tolerant strain HG228 has adaptively augmented its ethanol degradation capacity as a mechanism for tolerance.

Discussion

The relationship between genotype and phenotype is at the core of classical and modern genetics. However, complex phenotypes, involving many cellular processes, present significant challenges due to the limited sensitivity with which genotype–phenotype relationships can be mapped. Here, we have combined a comprehensive exploration of adaptive potentials together with a robust modular data analysis approach to reveal the genetic basis of a complex phenotype. Through the use of both transposon insertion and overexpression libraries, we surveyed the adaptive potential of all genetic loci with respect to ethanol tolerance. The fitness consequences of transposon insertion events were more pronounced compared with that of overexpression. This is not surprising given the nature of these perturbations. In many cases, like that of core enzymes in the TCA cycle, the overexpression of a single gene has little effect on the output of the pathway as a whole. Nevertheless, we observed three conditions in which overexpression perturbations result in a pronounced fitness effect: (1) overexpression of key regulatory components (e.g. *cadB* in acid stress response), (2) the upregulation of a key enzyme in the pathway (e.g. *slt* in cell-wall biogenesis), or (3) the simultaneous overexpression of multiple genes in the same pathway (e.g. *bet* regulon). Our ability to observe the latter is the consequence of the size of the cloned fragments in the overexpression library and the cistronic structure of the bacterial genomes in which all the genes in a small pathway exist together as part of a single operon.

On measuring the fitness consequences of both transposon insertions and overexpressions, we used a modular analysis of the fitness profiles to identify the relevant underlying pathways. For example, transposon insertions in the envelope stress response genes cause a slight decrease in fitness that is not significant enough for these genes individually to pass our gene-level statistical threshold. Whereas, in a modular analysis, the significance of this pathway can be detected as a collective effect of all these genes (Goodarzi *et al*, 2009a).

Our study revealed many pathways and processes that collectively contribute to ethanol tolerance in *E. coli*. We found modifications to endogenous pathways (e.g. upregulation of osmoprotectants and suppression of acid stress response pathway) or metabolic reprogramming to boost ethanol degradation capacity as potential mechanisms for adaptive ethanol tolerance. Our results argue for the dominance of regulatory network perturbations in adaptation to extreme environments. The fitness contribution of genes regulated by a range of transcription factors such as *betI*, *gcvA*, *arcA*, *fnr*, and *hns* signifies the adaptive potential of regulatory perturbations. This is to be contrasted with a model in which subtle amino-acid modifications in effector proteins are the dominant contributors to adaptation. Discovering adaptive mutations in different environments would ultimately test this hypothesis; nevertheless, we have previously catalogued adaptive mutations in two evolved strains: the ethanol-tolerant strain (HG179) and a strain (ASN*) capable of growing in minimal media plus asparagine eight times faster than the wild-type strain (Goodarzi *et al*, 2009b). In HG179, we found the major contributor to ethanol tolerance to be a point mutation in *rho*, the gene coding for the Rho transcription terminator (Goodarzi *et al*, 2009b). It has been shown earlier that Rho is a global regulator of gene expression and PrpC/D (propionate catabolic process) and CadA (acid stress response pathway) are among the proteins most affected by Rho inhibition (Cardinale *et al*, 2008). These proteins and their corresponding pathways were also identified as key players in ethanol tolerance in this study. Similarly, in ASN*, we discovered three adaptive mutations (an IS2 insertion, a single-nucleotide insertion, and a mismatch) that were all upstream of their respective ORFs, modifying their expression levels rather than their amino-acid sequence (Goodarzi *et al*, 2009b). Similar studies in other environments may further highlight the importance of regulatory perturbations in adaptation.

Using metabolomic approaches as a measure for downstream effects of the adaptation process, we have shown that some of the pathways identified through our global genetic approach are also modified in laboratory-evolved strains for enhanced ethanol tolerance, most notably biosynthesis of peptidoglycans, colanic acid, and enterobactin. Interestingly, neither of the evolved strains (HG227 and HG228) shows significant changes in glycine or glycine betaine levels (Supplementary Figure S9). The fact that the evolved strains did not show changes in all beneficial pathways is not surprising, as a single strain is unlikely to explore the entire fitness landscape on a relatively short evolutionary timescale, emphasizing the importance of approaching the analysis of evolution of complex traits through more systematic methods rather than simple strain selection under the desired condition.

© 2010 EMBO and Macmillan Publishers Limited

Through stable-isotope labeling in the ethanol-tolerant strain, HG228, we observed a substantial boost in ethanol assimilation as compared with the wild-type strain (also see Supplementary Figure S10). As mentioned earlier, ethanol consumption has been associated with ethanol tolerance in bacteria with active ethanol degradation pathways (Heipieper *et al*, 2000). However, in the case of our laboratory-evolved *E. coli* strain, ethanol degradation capacity emerges as part of the adaptation process, through regulatory and metabolic rewiring. Moreover, the anti-correlation between ethanol tolerance and ethanol production has been noted earlier in yeast: typically ethanol-tolerant strains are poor ethanol producers and *vice versa* (del Castillo Agudo, 1985). If ethanol degradation is a mechanism for tolerance, selecting for this phenotype results in an adaptive metabolic rewiring, which maximizes ethanol degradation (i.e. enhancing the reactions that deplete acetyl-CoA) rather than ethanol production.

In this study, we have introduced a framework based on coarse-grained sampling of the fitness landscape followed by a modular analysis for identification of pathways that contribute to emergence of complex adaptation. Given that we are directly assaying fitness, the identified pathways are either directly responsible for the observed effects (e.g. osmoregulation in ethanol tolerance), or function as part of an emerging pathway (e.g. *adhE* activity in ethanol degradation, in contrast with its normal function as an ethanol-producing enzyme). In parallel, we have used metabolomic approaches to probe the status of the identified pathways in the evolved strains. Validating the function of these pathways in the laboratory-evolved strains highlights the biological relevance of our approach and its ability to reveal the actual genetic mechanisms used during the evolutionary process.

Materials and methods

Strains and media

The strains, phages, and plasmids used in this study are listed in Supplementary Table S1. All the experiments were performed in LB (1% Trypton, 0.5% yeast extract, and 0.05% NaCl), or M9 minimal media plus glucose (4% w/v), supplemented with ethanol or antibiotics as required: ethanol, 4, 5.5, or 7% (v/v); ampicillin, 100 μg/ml; and kanamycin, 50 μg/ml, unless otherwise mentioned.

Quantitative analysis of mutant libraries

Library generation and microarray-based footprinting were carried out as described earlier (Girgis *et al*, 2007). To determine the fitness contribution of each gene in the transposon library, we compared its normalized hybridization signal in the selected samples to a set of five unselected samples in to capture the effect of selection on the frequency of each mutant (Girgis *et al*, 2007). For this, we calculated a *z*-score for gene *i* using $z_i=(x_i-\mu_i)/\sigma_i$, where x is the normalized hybridization signal in the selected sample, μ is the mean, and σ is the s.d. of normalized hybridization signals from the unselected samples. For each sample, the *z*-scores were then variance normalized across all the genes in a given sample to calculate the fitness scores. The selections were performed in biological replicates and the resulting fitness scores were averaged and reported.

The generation and manipulation of the overexpression library was performed as described earlier (Amini *et al*, 2009). The selections were performed in duplicates, and the fitness scores were calculated similarly to those of the transposon library. The fitness profiles are available in the Supplementary information.

DNA manipulation

Chromosomal deletions were either obtained from the Keio collection (Baba *et al*, 2006) and transferred by generalized transduction with P1*vir* phage into the MG1655 background (Silhavy *et al*, 1984) or created using the previously described methods (Datsenko and Wanner, 2000).

Modular analysis of fitness profiles

We used iPAGE (Goodarzi *et al*, 2009a), a mutual information-based approach, to discover the genetic modules that show non-random patterns across the fitness profiles (see Supplementary information and Supplementary Figure S2). The iPAGE outputs are included in the Supplementary information. iPAGE is also available for download at http://tavazoielab.princeton.edu/iPAGE/ and can be used online at https://iget.princeton.edu/.

Metabolite concentration measurements and stable-isotope labeling of the ethanol-tolerant strains

Culture filtering, metabolome quenching, and extraction procedures used features from previously described protocols (Rabinowitz and Kimball, 2007; Bennett *et al*, 2008). Briefly, overnight cultures in LB media were diluted in minimal media with 1.5% ethanol (for ^{13}C-ethanol labeling experiment) or 0.2% glucose and 3.5% ethanol (for relative metabolite concentrations). Cells were then placed on nylon filters by vacuum filtration, and the filters were placed cell-side up on plates of identical composition embedded in triply washed agarose. For metabolite concentrations, metabolism was quenched and metabolites were simultaneously extracted by moving the filter to a solution of 40:40:20 methanol, acetonitrile, and water with 0.1 M formic acid at −20°C. For ^{13}C-ethanol labeling, cell-loaded filters were first placed on plates with unlabeled ethanol, then either extracted (time 0) or moved to plates with ^{13}C-ethanol for 0.25, 1, or 4 h.

Relative pool sizes were quantified using two different LC methods coupled by electrospray ionization to TSQ Quantum triple quadrupole mass spectrometers (Thermo Scientific) operating in MRM mode. Incorporation of labeled ethanol was monitored using LC coupled to a high-resolution, high-mass accuracy Exactive mass spectrometer (Thermo Scientific). For compete details on cell culture conditions, analytical methods, and complete data set, see Supplementary information.

Assessing the effects of mutations on ethanol tolerance

We used kill curves in LB plus 7% (v/v) ethanol to compare the tolerance level of different mutants to the wild-type strain MG1655. We counted the colony forming units (CFUs) at 1, 2, 4, and 8 h on addition of ethanol in triplicate. To assess the significance level at which the mutants differed from the wild-type strain, we used an ANCOVA test with an exponential model to calculate the associated *P*-values.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We thank members of the Tavazoie laboratory for critical reading of the manuscript. S.T. was supported by grants from the NIGMS (P50 GM071508) and the NIH Director's Pioneer Award (1DP10D003787-01).

© 2010 EMBO and Macmillan Publishers Limited

Conflict of interest

The authors declare that they have no conflict of interest.

References

Amini S, Goodarzi H, Tavazoie S (2009) Genetic dissection of an exogenously induced biofilm in laboratory and clinical isolates of E. coli. *PLoS Pathog* **5:** e1000432

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25:** 25–29

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2:** 2006.008

Badarinarayana V, Estep PW, Shendure J, Edwards J, Tavazoie S, Lam F, Church GM (2001) Selection analyses of insertional mutants using subgenic-resolution arrays. *Nat Biotech* **19:** 1060–1065

Bennett BD, Yuan J, Kimball EH, Rabinowitz JD (2008) Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. *Nat Protoc* **3:** 1299–1311

Buttke TM, Ingram LO (1980) Ethanol-induced changes in lipid composition of Escherichia coli: inhibition of saturated fatty acid synthesis *in vitro*. *Arch Biochem Biophys* **203:** 565–571

Cardinale CJ, Washburn RS, Tadigotla VR, Brown LM, Gottesman ME, Nudler E (2008) Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E-coli. *Science* **320:** 935–938

Cartwright CP, Juroszek JR, Beaven MJ, Ruby FMS, De Morais SMF, Rose AH (1986) Ethanol dissipates the proton-motive force across the plasma membrane of Saccharomyces cerevisiae. *J Gen Microbiol* **132:** 369–377

Cover T, Thomas J (2006) *Elements of Information Theory*, 2nd edn, Hoboken, NJ: Wiley-Interscience

D'Amore T, Panchal CJ, Russell I, Stewart GG (1990) A study of ethanol tolerance in yeast. *Crit Rev Biotechnol* **9:** 287–304

D'Amore T, Stewart GG (1987) Ethanol tolerance of yeast. *Enzyme Microb Technol* **9:** 322–330

Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci USA* **97:** 6640–6645

del Castillo Agudo L (1985) Genetic aspects of ethanol tolerance and production by Saccharomyces cerevisiae. *Current Microbiol* **12:** 41–44

Dombek KM, Ingram LO (1984) Effects of ethanol on the Escherichia coli plasma membrane. *J Bacteriol* **157:** 233–239

Engel H, Kazemier B, Keck W (1991) Murein-metabolizing enzymes from Escherichia coli: sequence analysis and controlled overexpression of the slt gene, which encodes the soluble lytic transglycosylase. *J Bacteriol* **173:** 6773–6782

Fiehn O (2001) Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp Funct Genomics* **2:** 155–168

Foster JW (2004) Escherichia coli acid resistance: tales of an amateur acidophile. *Nat Rev Microbiol* **2:** 898–907

Girgis HS, Liu Y, Ryu WS, Tavazoie S (2007) A comprehensive genetic characterization of bacterial motility. *PLoS Genet* **3:** 1644–1660

Gonzalez R, Tao H, Purvis JE, York SW, Shanmugam KT, Ingram LO (2003) Gene array-based identification of changes that contribute to ethanol tolerance in ethanologenic Escherichia coli: comparison of KO11 (parent) to LY01 (resistant mutant). *Biotechnol Prog* **19:** 612–623

Goodarzi H, Elemento O, Tavazoie S (2009a) Revealing global regulatory perturbations across human cancers. *Mol Cell* **36:** 900–911

Goodarzi H, Hottes AK, Tavazoie S (2009b) Global discovery of adaptive mutations. *Nat Methods* **6:** 581–583

Green J, Paget MS (2004) Bacterial redox sensors. *Nat Rev Microbiol* **2:** 954–966

Hallsworth JE (1998) Ethanol-induced water stress in yeast. *J Ferment Bioeng* **85:** 125–137

Heipieper HJ, Isken S, Saliola M (2000) Ethanol tolerance and membrane fatty acid adaptation in adh multiple and null mutants of Kluyveromyces lactis. *Res Microbiol* **151:** 777–784

Ingram LO (1976) Adaptation of membrane lipids to alcohols. *J Bacteriol* **125:** 670–678

Ingram LO (1977) Preferential inhibition of phosphatidyl ethanolamine synthesis in E. coli by alcohols. *Can J Microbiol* **23:** 779–789

Ingram LO (1986) Microbial tolerance to alcohols: role of the cell membrane. *Trends Biotechnol* **4:** 40–44

Ingram LO, Buttke TM (1984) Effects of alcohols on micro-organisms. *Adv Microb Physiol* **25:** 253–300

Mansure JJ, Panek AD, Crowe LM, Crowe JH (1994) Trehalose inhibits ethanol effects on intact yeast cells and liposomes. *Biochim Biophys Acta* **1191:** 309–316

Mates AK, Sayed AK, Foster JW (2007) Products of the Escherichia coli acid fitness island attenuate metabolite stress at extremely low pH and mediate a cell density-dependent acid resistance. *J Bacteriol* **189:** 2759–2768

Membrillo-Hernandez J, Echave P, Cabiscol E, Tamarit J, Ros J, Lin EC (2000) Evolution of the adhE gene product of Escherichia coli from a functional reductase to a dehydrogenase. Genetic and biochemical studies of the mutant proteins. *J Biol Chem* **275:** 33869–33875

Millar DG, Griffiths-Smith K, Algar E, Scopes RK (1982) Activity and stability of glycolytic enzymes in the presence of ethanol. *Biotechnol Lett* **4:** 601–606

Nagodawithana TW, Whitt JT, Cutaia AJ (1977) Study of the feedback effect of ethanol on selected enzymes of the glycolytic pathway. *J Am Soc Brew Chem* **35:** 179–183

Neidhardt FC, VanBogelen RA (1987) *The Heat Shock Response*. Washington, DC: American Society for Microbiology Press

Nicholson JK, Connelly J, Lindon JC, Holmes E (2002) Metabonomics: a platform for studying drug toxicity and gene function. *Nat Rev Drug Discov* **1:** 153–161

Ow MC, Perwez T, Kushner SR (2003) RNase G of Escherichia coli exhibits only limited functional overlap with its essential homologue, RNase E. *Mol Microbiol* **49:** 607–622

Palacios S, Escalante-Semerena JC (2000) prpR, ntrA, and ihf functions are required for expression of the prpBCDE operon, encoding enzymes that catabolize propionate in Salmonella enterica serovar typhimurium LT2. *J Bacteriol* **182:** 905–910

Rabinowitz JD, Kimball E (2007) Acidic acetonitrile for cellular metabolome extraction from Escherichia coli. *Anal Chem* **79:** 6167–6173

Rowley G, Spector M, Kormanec J, Roberts M (2006) Pushing the envelope: extracytoplasmic stress responses in bacterial pathogens. *Nat Rev Microbiol* **4:** 383–394

Salgado H, Gama-Castro S, Peralta-Gil M, Diaz-Peredo E, Sanchez-Solano F, Santos-Zavaleta A, Martinez-Flores I, Jimenez-Jacinto V, Bonavides-Martinez C, Segura-Salazar J, Martinez-Antonio A, Collado-Vides J (2006) RegulonDB (version 5.0): Escherichia coli K-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Res* **34:** D394–D397

Sauer U (2006) Metabolic networks in motion: 13C-based flux analysis. *Mol Syst Biol* **2:** 62

Silhavy TJ, Berman ML, Enquist LW (1984) *Experiments with Gene Fusions*. Plainview, NY: Cold Spring Harbor Press

Sprencel C, Cao Z, Qi Z, Scott DC, Montague MA, Ivanoff N, Xu J, Raymond KM, Newton SM, Klebba PE (2000) Binding of ferric enterobactin by the Escherichia coli periplasmic protein FepB. *J Bacteriol* **182:** 5359–5364

© 2010 EMBO and Macmillan Publishers Limited

Storz G, Hengge-Aronis R (2000) *Bacterial Stress Responses*. ASM Press, American Society for Microbiology, Washington, DC

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102:** 15545–15550

Tagkopoulos I, Liu YC, Tavazoie S (2008) Predictive behavior within microbial genetic networks. *Science* **320:** 1313–1317

Thomas JG, Baneyx F (1997) Divergent effects of chaperone overexpression and ethanol supplementation on inclusion body formation in recombinant Escherichia coli. *Protein Expr Purif* **11:** 289–296

Umitsuki G, Wachi M, Takada A, Hikichi T, Nagai K (2001) Involvement of RNase G in *in vivo* mRNA metabolism in Escherichia coli. *Genes Cells* **6:** 403–410

Wikoff WR, Gangoiti JA, Barshop BA, Siuzdak G (2007) Metabolomics identifies perturbations in human disorders of propionate metabolism. *Clin Chem* **53:** 2169–2176

Yokoyama S (2002) Evaluating adaptive evolution. *Nat Genet* **30:** 350–351

Yuan J, Fowler WU, Kimball E, Lu W, Rabinowitz JD (2006) Kinetic flux profiling of nitrogen assimilation in Escherichia coli. *Nat Chem Biol* **2:** 529–530

Zaldivar J, Nielsen J, Olsson L (2001) Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration. *Appl Microbiol Biotechnol* **56:** 17–34

SOME RIGHTS RESERVED

Molecular Systems Biology is an open-access journal published by *European Molecular Biology Organization* and *Nature Publishing Group*. This article is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License.

© 2010 EMBO and Macmillan Publishers Limited