

# SCIENTIFIC REPORTS



OPEN

## Systems-level understanding of ethanol-induced stresses and adaptation in *E. coli*

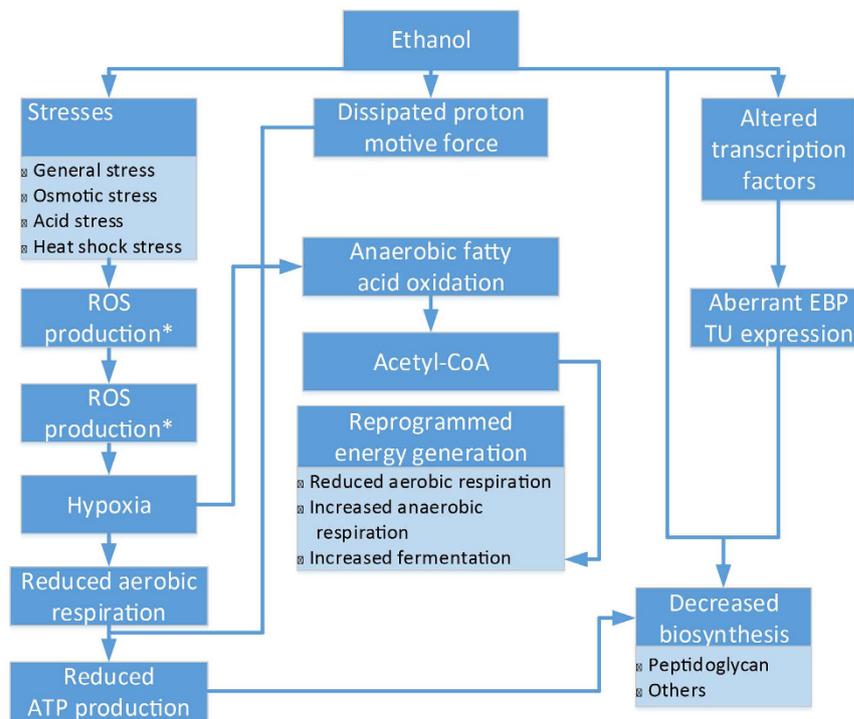
Received: 27 October 2016  
Accepted: 02 February 2017  
Published: 16 March 2017

Huansheng Cao<sup>1,2,\*</sup>, Du Wei<sup>1,3,\*</sup>, Yuedong Yang<sup>4</sup>, Yu Shang<sup>1,3</sup>, Gaoyang Li<sup>1,3</sup>, Yaoqi Zhou<sup>4</sup>, Qin Ma<sup>5,6</sup> & Ying Xu<sup>1,2,3</sup>

Understanding ethanol-induced stresses and responses in biofuel-producing bacteria at systems level has significant implications in engineering more efficient biofuel producers. We present a computational study of transcriptomic and genomic data of both ethanol-stressed and ethanol-adapted *E. coli* cells with computationally predicated ethanol-binding proteins and experimentally identified ethanol tolerance genes. Our analysis suggests: (1) ethanol damages cell wall and membrane integrity, causing increased stresses, particularly reactive oxygen species, which damages DNA and reduces the O<sub>2</sub> level; (2) decreased cross-membrane proton gradient from membrane damage, coupled with hypoxia, leads to reduced ATP production by aerobic respiration, driving cells to rely more on fatty acid oxidation, anaerobic respiration and fermentation for ATP production; (3) the reduced ATP generation results in substantially decreased synthesis of macromolecules; (4) ethanol can directly bind 213 proteins including transcription factors, altering their functions; (5) all these changes together induce multiple stress responses, reduced biosynthesis, cell viability and growth; and (6) ethanol-adapted *E. coli* cells restore the majority of these reduced activities through selection of specific genomic mutations and alteration of stress responses, ultimately restoring normal ATP production, macromolecule biosynthesis, and growth. These new insights into the energy and mass balance will inform design of more ethanol-tolerant strains.

Ethanol generation by bacteria represents a promising approach to biofuel production at an industrial scale. Substantial research efforts have been invested into strain optimization for the cost-effective biofuel production by microbes. One of the key challenges in realizing this lies in the toxicity of the produced ethanol to the host cells, which makes the bacteria sick with reduced ethanol production and viability. Previous studies have identified cell membrane damage<sup>1–3</sup>, reduced cross-membrane proton potential<sup>4,5</sup>, and reduced peptidoglycan cross-linking in cell wall<sup>2,6</sup> as the main reasons for the reduced productivity and viability of the affected cells. To understand the detailed mechanisms, a number of ethanol-response regulons such as ArcA and FNR along with numerous ethanol-tolerance genes have been identified using omics techniques and extensively studied<sup>7–12</sup>. However, the overall understanding about the impact of ethanol-induced stress, short-term response and long-term adaptation remains to be somewhat fragmented, which is yet to offer very informative guidance to engineering highly ethanol-tolerant biofuel producers. One important reason is that there is a lack of a systems-level understanding about the impact of ethanol stress, specifically how different impacts are related to each other and which may be the root causes of the observed ethanol-induced cellular level alterations. For example, we are yet to fully understand which cellular subsystems are directly affected by ethanol, which are the results of stress responses, and how the reduced functionalities of these subsystems are compensated by other subsystems. In addition, information is still lacking regarding which macromolecules can directly bind with ethanol, hence having their functions disrupted at the cellular level.

<sup>1</sup>Computational Systems Biology Laboratory, Department of Biochemistry and Molecular Biology, and Institute of Bioinformatics, the University of Georgia, Athens, GA 30602, USA. <sup>2</sup>BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. <sup>3</sup>College of Computer Science and Technology, Jilin University, Changchun, 130012, China. <sup>4</sup>Institute for Glycomics and School of Information and Communication Technology, Griffith University, Parklands Dr., Southport, QLD 4222, Australia. <sup>5</sup>Department of Agronomy, Horticulture and Plant Science, South Dakota State University, Brookings, SD 57007, USA. <sup>6</sup>BioSNTR, Brookings, SD, 57007, USA. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to Y.X. (email: xyn@uga.edu)



**Figure 1. Impact of ethanol on the NEA cells.** ROS: reactive oxygen species; EBP: ethanol-binding proteins; and TU: transcription unit. Boxes with asterisks are inferred based on information collected through literature search.

We present a computational study of published microarray gene-expression data in the public domain of *E. coli* cells with ethanol treatment *versus* ethanol-free controls in the non-ethanol-adapted (NEA) samples, aiming to derive how ethanol affects the normal functions of the cells. We have also examined gene-expression and genomic data of *E. coli* cells that are ethanol-adapted (EA) after 2,496 hours of evolution, to elucidate how the cells have adapted to the ethanol-induced stress and regained their viability. Our overall findings include: (1) ethanol disrupts and damages cell wall and membranes, resulting in increased oxidative and hypoxic stresses among others; (2) decreased cross-membrane proton potential due to membrane damage, along with hypoxia, leads to reduced ATP production by aerobic respiration, which drives the cells to turn up other means for ATP production, namely anaerobic respiration and fermentation; (3) the reduced ATP production results in reduction in macromolecular biosynthesis and cell proliferation, and triggers increased catabolism of carbohydrates and fatty acid oxidation for energy production; (4) ethanol can directly bind with hundreds of proteins, including transcription factors, which alters their normal functions; (5) these impacts are functionally linked with each other, resulting in reduced viability and growth in the affected cells; and (6) ethanol-adapted *E. coli* cells have restored some key activities through selection of specific mutations and changes in stress response, leading to restored ATP production and macromolecule biosynthesis, reduced stresses, repaired membranes, and ultimately regained cellular viability and growth. Based on these findings, we have constructed a model for ethanol stress and response as well as adaptation by *E. coli* cells.

The main contribution of this study includes: (i) a systems-level model for ethanol-induced stresses in *E. coli* cells as well as for adaptation to the stresses, which integrate new findings and previous observations; (ii) cell membrane damage and the stress-induced ROS (reactive oxygen species) production may be at the root of the observed changes as outlined above; and (iii) a number of new observations regarding how ethanol may contribute to the reduced cellular viability and growth, as well as how the adapted cells overcome the ethanol-induced adversity.

## Results

Computational analyses of transcriptomic and genomic data in the public domain (Supplementary Table S1) combined with ethanol tolerance genes (Supplementary Table S2) and stress resistance pathways (Supplementary Table S3) were performed to infer how ethanol-induced stresses impact *E. coli* cells at a systems level and how the affected cells adapt to the stresses.

**Impacts of ethanol on NEA samples and cellular responses.** We present here the identified impacts of ethanol and corresponding responses in *E. coli* cells, which are summarized in Fig. 1.

*Multiple stresses are observed along with increased ROS level.* We first examined the types of stress responses to ethanol treatment in the NEA samples. Known responses to acid and osmotic stress, envelope stress, and heat-shock stress were all observed in our results, with varying degrees of up-regulation of the relevant marker

genes. For examples, *asr*, *gadE*, *gadA*, and *gadB* are up-regulated for acid stress; *osmC*, *osmB*, and *ompR* for osmotic stress; *ostA*, *ostB*, *treB*, and *treC* for trehalose production in resisting the osmotic stress; *rpoE*, *degP*, *asnB*, and *opgG* for envelope stress; and *groS*, *groL*, *grpE*, and *metA* for heat-shock stress, as shown in Supplementary Fig. S1.

Furthermore, we have observed increased response to ROS, which has not received much attention, although it has been found to be induced by ethanol in yeast<sup>13</sup>. Specifically, two key regulons, OxyR and SoxRS, responding to hydrogen peroxide H<sub>2</sub>O<sub>2</sub> and superoxide (O<sub>2</sub><sup>-</sup>), respectively, are up-regulated (Fig. 2A and B), including key genes: *sodA*, *fumC*, and *katE* in at least five out of the seven samples under consideration. Additionally, the expressions of these two regulons correlate with the ethanol concentration, with correlation coefficients 0.53 and 0.49, respectively.

Fenton reaction (Fe<sup>2+</sup> + H<sub>2</sub>O<sub>2</sub> -> Fe<sup>3+</sup> + •HO + OH<sup>-</sup>) has also been observed in *E. coli* when the cells produce substantial amounts of H<sub>2</sub>O<sub>2</sub>, which can react with Fe<sup>2+</sup> in the iron-sulfur (Fe-S) clusters in proteins<sup>14–16</sup>, producing Fe<sup>3+</sup>, hydroxyl radical (•HO) and hydroxide (OH<sup>-</sup>). Sulfurs in the same cluster can reduce Fe<sup>3+</sup> back to Fe<sup>2+</sup><sup>17</sup>, giving rise to another round of Fenton reaction and more damaged iron-sulfur clusters. We found that most genes in Fe-S cluster assembly process are up-regulated (Supplementary Fig. S2C), suggesting that the Fe-S clusters are damaged and hence need to be replaced. We also detect a correlation between the Fe-S assembly genes and ethanol concentration, with a correlation coefficient of 0.5. All these provide evidence of Fenton reaction. It is noteworthy that •HO is a most reactive molecule, which can cause substantial damages to nearby molecular structures.

As ROS oxidizes and hence damages macromolecules such as DNA, we also observed an up-regulated SOS response, an error-prone DNA repair system (Fig. 2D). About half of the genes of the SOS response are up-regulated in all the samples, including key genes *recA*, *recN*, *sulA*, and *ssb* in at least six of the samples (Fig. 2D).

Since ROS production consumes O<sub>2</sub> and thus reduces the O<sub>2</sub> level, hypoxia might occur. The AcrAB two-component system is a switch that dually regulates genes in micro-aerobic conditions<sup>18</sup>. As expected, about 1/3 genes in the ArcAB regulon are up-regulated for anaerobic metabolism, such as *ackA* and *pflB* for fermentation (Supplementary Fig. S3). In addition, a correlation is observed between the SoxRS or OxyR regulon and the ArcAB regulon, with correlation coefficients 0.33 and 0.25, respectively, hence suggesting that the observed hypoxia is related to the increased ROS production.

**Impact of reduced cross-membrane proton gradient and response.** It has been established that ethanol can fluidize cell membrane, leading to decreased cross-membrane proton gradient and thus reduction in ATP production by respiration<sup>4</sup>. Our analysis confirms this in the NEA samples with multiple supporting evidence. First, the major energy sensor genes *aer* and *tsr* are up-regulated in six of seven samples (Fig. 3A). In addition, two other chemotaxis sensor genes *tap* and *tar* are also up-regulated, responding to the signaling molecules imported by four transporters (Fig. 3B). Together, these sensors work synergistically in activating flagellum assembly and cell motility (Supplementary Fig. S4), and genes involved in the two processes are found to be up-regulated for 20 out of the 26 genes (77%) in all the NEA samples (Supplementary Fig. S5). A correlation was also observed between these responses and ethanol concentration, with a correlation coefficient 0.42.

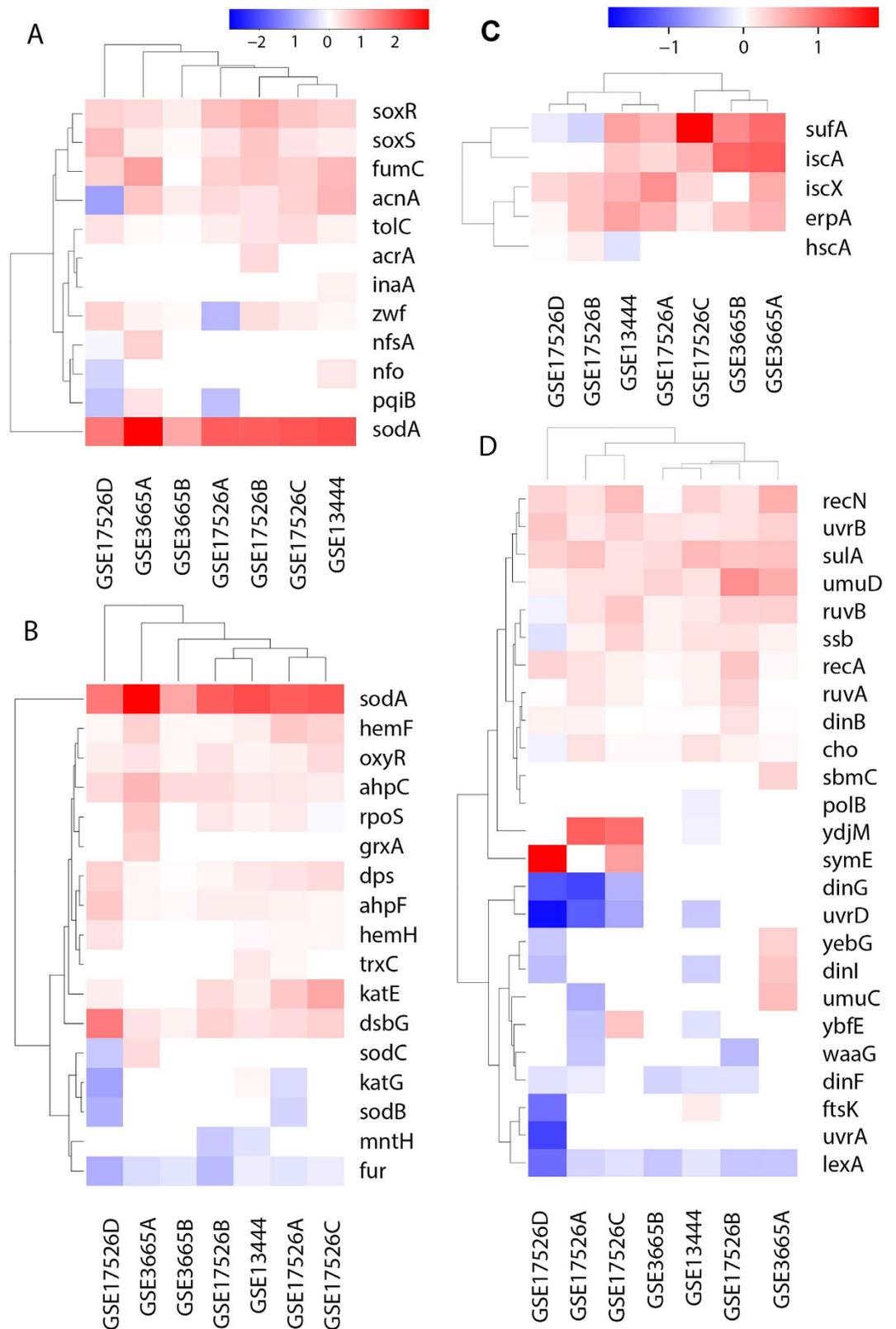
Additional evidence for decreased ATP production includes: (1) DNA replication shows negative correlation with increased ethanol concentration (correlation coefficient: -0.45); (2) major biosynthetic processes, such as cofactor/prosthetic group biosynthesis, phospholipid and amino acid syntheses, all decrease with the increase in ethanol concentration (Supplementary Table S4); and (3) these decreased biosynthetic processes are found to be positively correlated with the decrease in DNA replication (Supplementary Table S5).

We then examined cellular responses to the reduced ATP production. We observed that some genes in the electron transport chain are down-regulated with varying degrees of proportions across different samples, such as NADH dehydrogenase (*nuoA-N*), terminal oxidoreductases *cyoABCDE* and *cydAB* (Fig. 3C). Interestingly, two anaerobic oxidoreductases (*frdABCD* and *torA/torC*) are up-regulated (Fig. 3D). Some key fermentation enzymes such as pyruvate-formate lyase (*pflB*), phosphate acetyltransferase (*pta*), and acetate kinase (*ackA*) are up-regulated, hence indicating ATP generation through conversion of pyruvate to acetate (Fig. 4C). These changes suggest that with the reduced ATP production by aerobic respiration, the affected cells rely on anaerobic respiratory and fermentation enzymes for ATP production.

Another major response to the reduced ATP production is anaerobic  $\beta$ -oxidation of fatty acid in four of the seven samples, which also have up-regulated fermentation activities. The key step of this pathway is to break up long-chain fatty acids to acetyl-CoA that is catalyzed by three anaerobic enzymes (*fadI*, *fadJ*, and *fadK*), which are all up-regulated in four samples, while one (*fadD*) of the three aerobic counterparts (*fadA*, *fadB*, and *fadD*) is down-regulated. In addition, the level of  $\beta$ -oxidation is found to correlate with the increase in ethanol concentration with a correlation coefficient of 0.49.

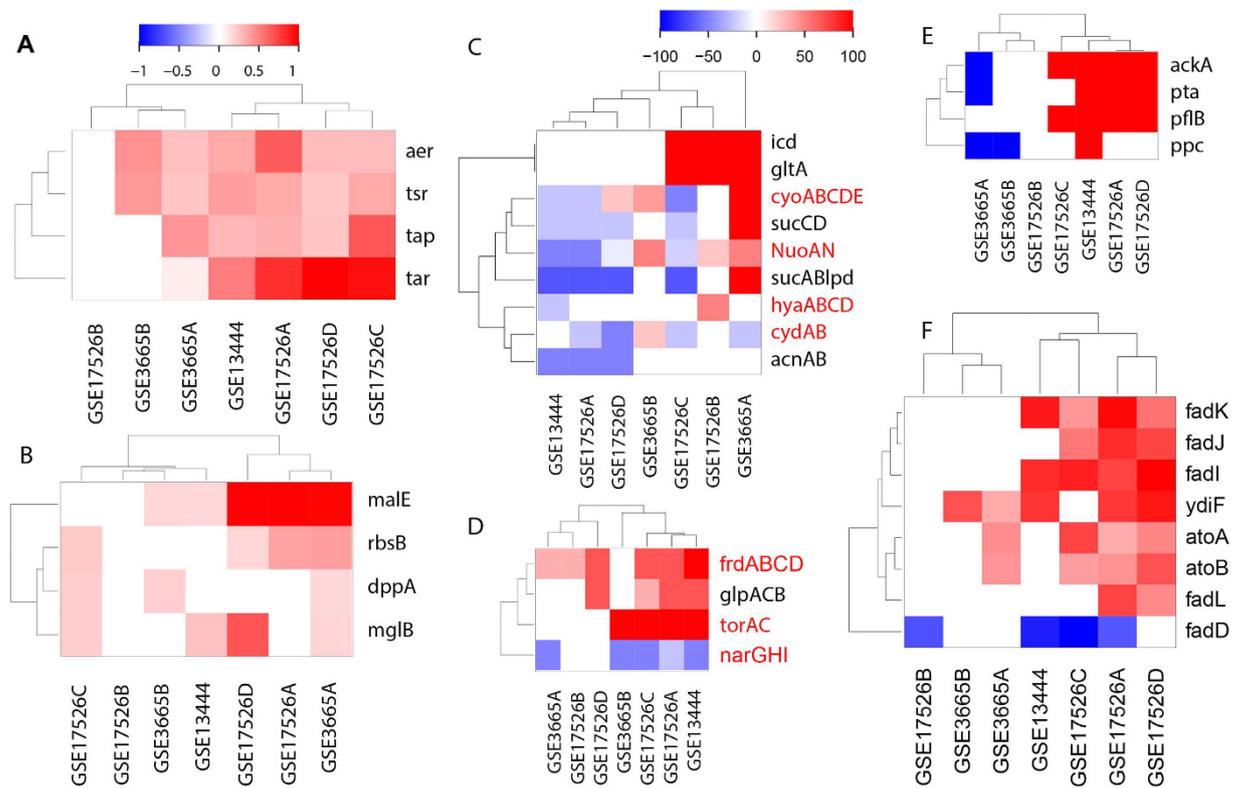
**Ethanol affects protein functions.** We have also investigated which proteins may directly bind with ethanol and thus show altered functions by computationally identifying ethanol-binding proteins (EBPs) and examining the possible consequence of ethanol binding. A total of 213 EBPs are predicted, as detailed in Supplementary Table S6, which are involved in 31 biological processes (Supplementary Table S7). Among the most highly enriched processes are oxidation reduction, carbohydrate/polysaccharide biosynthesis, lipid biosynthesis, rRNA/RNA modification, and metabolism of vitamins and cofactors, all of which are consistent with our earlier observation that the affected cells have substantially reduced biosynthesis of macromolecules and increased ROS production. Among all the NEA samples, 78 out of the 213 (37%) EBP genes are up-regulated while 54 (25%) are down-regulated.

We also examined whether ethanol binding may alter protein functions, by comparing the expression profiles of the transcription units (TUs) regulated by EBP transcription factor genes (TFs) with the experimentally validated regulatory relationships as documented in the RegulonDB database<sup>19</sup>. Here, only those TUs with EBP



**Figure 2. Differential gene expressions in seven pairs of NEA samples.** (A) SoxRS regulon; (B) OxyR regulon; (C) iron-sulfur cluster assembly genes; and (D) SOS response. In each panel, the x-axis represents the samples used, and the y-axis denotes the proportion of differentially expressed genes out of all. The color definitions for Panels A and B, and for Panels C and D are the same.

TFs as sole regulators were checked, and 10–100% such TUs were found to be aberrantly expressed (Fig. 4). The total 48 genes in these TUs are involved in multiple biosynthetic processes, such as amine biosynthetic process



**Figure 3. Pathways in response to reduced ATP production in the seven NEA samples.** (A) energy and chemotaxis sensor genes; (B): genes encoding transporters for chemotaxis signal molecules; (C–E): the proportion of differentially expressed genes encoding enzymes out of all encoding the aerobic respiration pathway (C), anaerobic respiration (D), and fermentation (E) and (F) differential expression of genes in fatty acid oxidation pathway. For each panel, the *x*-axis represents samples and the *y*-axis is for genes. Panels A, B and F have the same color definition while panels (C,D and E) have the same definition, which represents the proportion of genes in each pathway with red and blue representing up- and down-regulation, respectively.

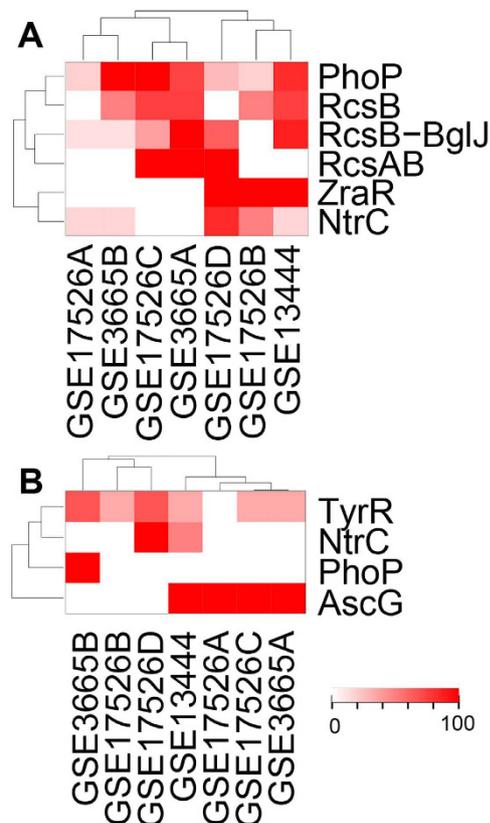
(Supplementary Table S8). These aberrant TU expressions suggest that functions of these EBPs are likely to be altered by ethanol binding.

**Reduced peptidoglycan and other macromolecule biosynthesis.** Besides interfering with normal cellular processes presented above, ethanol also reduces peptidoglycan biosynthesis among other biosynthetic processes (Supplementary Table S9). One key crosslinking gene *murG* is an EBP (Supplementary Table S6) and is down-regulated in most samples (Supplementary Fig. S6B). Its product, *N*-acetylglucosaminyl transferase, catalyzes the final step of intracellular peptidoglycan crosslinking (Figures S7–S9). Furthermore, all three stages of peptidoglycan biosynthesis (initiation, crosslinking, and maturation) are down-regulated, with 75% (3/4), 92% (12/13), and 76% (16/21) genes down-regulated at each stage, respectively, in six of the seven NEA samples (Supplementary Fig. S6). The down-regulation of the entire pathway correlates negatively with ethanol concentration with a correlation coefficient  $-0.54$ . Besides, the expression level of the peptidoglycan biosynthesis has high correlation with other reduced biosynthetic processes as presented in Supplementary Table S9.

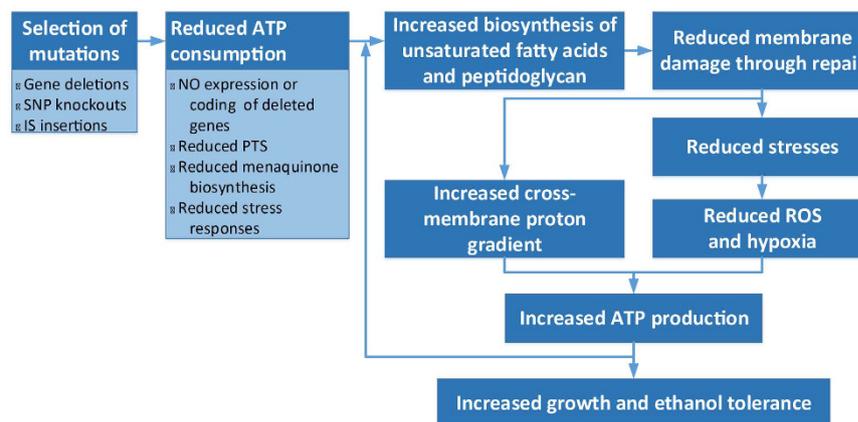
We summarize the GO biological processes which are highly enriched by differentially expressed genes in Supplementary Table S10.

**Adaptation in the evolved strains.** Here we address the question: how may the affected cells have adapted to the stresses induced by ethanol? To answer this question, we collected ethanol tolerance genes that are experimentally identified (Supplementary Table S2) from datasets in the public domain, along with transcriptomic and genomic data of six strains of *E. coli* that have adapted to the ethanol toxicity in parallel<sup>20,21</sup>. For each evolved strain, the transcriptomic data were collected at six different time points: h0, h384, h744, h1,224, h1,824 and h2,496. We also examined genomic mutations in these evolved strains. Based on our analyses, we constructed a model for how the EA strains adapt to ethanol at a systems level (Fig. 5).

**Genomic mutations in the evolved strains.** We have examined the mutations accumulated in the ethanol-adapted strains. Of the six strains that evolved from a common parent strain, strain A has a mutant DNA mismatch repair gene *mutS* and hence is deficient in mismatch repair (commonly known as a mutator strain). It has 131 mutations: 125 SNP and 6 indels<sup>20</sup>. A total of 27 mutated genes are found in the other five EA strains: 4 in strain B; 8 in C; and 5 in each of D, E and F. The details of these mutations and their roles in adaptation are provided in



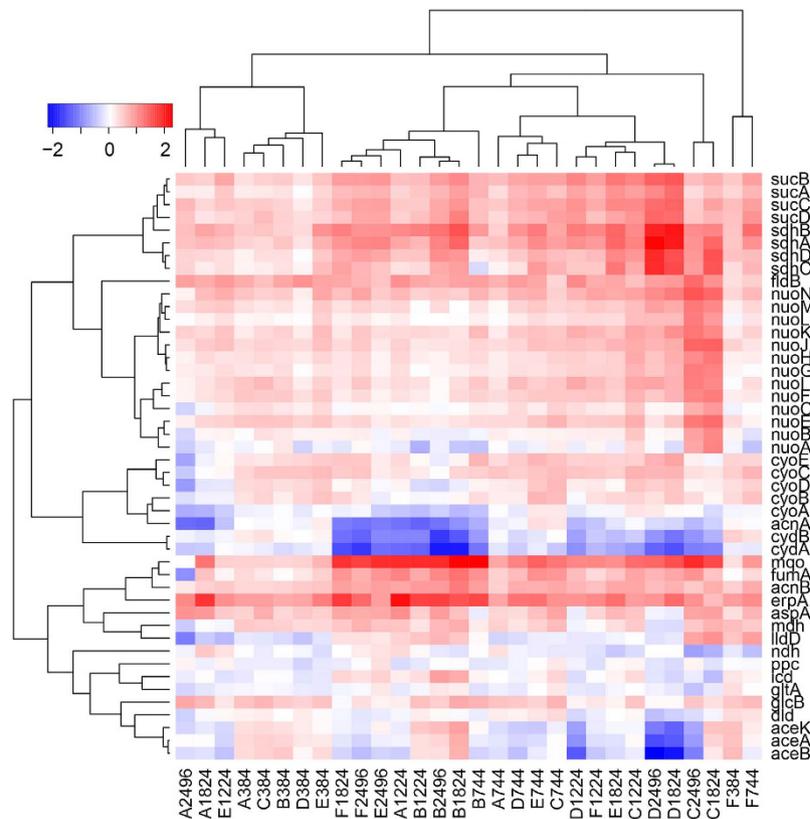
**Figure 4.** Proportions of differentially regulated TUs by EBP TFs in the NEA samples. The x-axis represents the seven NEA samples and the y-axis is for the EBP TFs. (A) activating EBP TFs; and (B) repressive EBP TFs.



**Figure 5.** A model for how the evolved NEA strains have adapted to ethanol-induced stresses. Asterisks indicate that our predicted changes are also experimentally reported in literature. Major processes are highlighted in bold. Different colors represent different processes.

Supplementary Table S11. 11 of these genes are involved in gene regulation (Supplementary Fig. S10B), including four global regulator genes: *hns*, *relA*, *rpoA* and *rpoC*. Besides, strain C has a large duplication of ~220 Kb<sup>20</sup>, which contains 180 genes (Supplementary Table S12). 96 (53%) of these genes are expressed and up-regulated between h1,224 and h1,824 (Wilcoxon test,  $p < 0.01$ ) and continues to h2,496 (Fig. S11A and B); these genes show higher expressions than those in other strains (Supplementary Fig. S11C). Functionally, these up-regulated genes are involved in biosynthetic processes, such as biosynthesis of carbohydrates, polysaccharides and lipopolysaccharides (Supplementary Table S13).

We have also examined 1,623 ethanol tolerance genes that have been experimentally identified either by overexpression or knockout (Supplementary Table S14). 60% (977/1623) of these genes are expressed in all six EA samples, and 49% (480/977) are differentially expressed: 277 up-regulated and 203 down-regulated



**Figure 6. Differentially expressed genes involved in aerobic respiration in EA samples.** The *x*-axis is for the six strains (A–F) at five time points (384, 744, 1,224, 1,824, and 2,496) and the *y*-axis is for genes involved in aerobic respiration.

(Supplementary Fig. S12). Of the 480 genes, 91% showed monotonic changes over time in terms of their differential expressions and match their overexpression or knockout patterns (Supplementary Table S14), i.e., these genes show step-wise up- or down-regulation over the course of evolution as overexpression or knockout (Supplementary Fig. S12). Functionally, most of these genes are involved in transcription, translation, and biosynthesis of vitamins, cofactors, carbohydrates and polysaccharides, and cell wall macromolecules including peptidoglycan (Supplementary Table S15).

**Reduced non-essential ATP consumption and restored ATP production.** We then examined whether the ROS production is still high in the evolved strains by analyzing the level of response to ROS. Compared to the parent strain, response to ROS shows a step-wise decline in mean expression levels of all response genes throughout the evolution (Fig. S13A and D). Most (6/9) genes are down-regulated persistently while only two (*soda* and *ahpF*) are up-regulated. This is consistent with the fact that the majority of the genes in SoxRS and OxyR regulons are consistently down-regulated (Supplementary Fig. S13B, C, E, and F). These down-regulations suggest that ROS level is significantly reduced and the O<sub>2</sub> level is increased in the EA strains.

To verify that O<sub>2</sub> level is indeed increased, we examined the expression level of the *nrdD* gene that is essential to anaerobic respiration and is highly repressed by O<sub>2</sub><sup>22</sup>. Indeed, this gene is dramatically repressed (Supplementary Fig. S14) compared to the parent strain, and its activator *nrdG*, which is also essential for anaerobic respiration<sup>22</sup>, is not expressed. In addition, genes for three key anaerobic enzymes (*narGHI*, *narZXV*, and *frdABCD*) are not expressed, and the only two expressed genes *narX* and *frdA* are down-regulated progressively during evolution (Supplementary Fig. S15). Furthermore, most fermentation genes are also down-regulated (Supplementary Fig. S16).

Additionally, most genes involved in aerobic respiration are up-regulated (Fig. 6), such as NADH dehydrogenase (*nuoA–N*), succinate dehydrogenase (*sdhABCD*), fumarate dehydrogenase (*fumA*), terminal oxidoreductases (*cydAB* and *cyoABCDE*), and lactate dehydrogenase (*ldhA* and *lldD*). Furthermore, we observed (1) no genes encoding energy sensors are expressed in the evolved strains; (2) no genes are expressed for fatty-acid oxidation, except the repressor of this pathway *fadR* and a fatty acid porin *fadL*, which are up-regulated in most EA strains (Supplementary Fig. S17); and (3) 65% (28/43) of biosynthetic processes show positive correlation with DNA replication in the EA samples (Supplementary Table S16), as opposed to the negative correlation between DNA replication and biosynthesis in the NEA samples. All these strongly support that ATP production by aerobic respiration is greatly improved.

We also observed that reduction in non-essential ATP consumption contributes to ATP sufficiency during the adaptive evolution. For example, the *nagE* gene in strain C has a CCG insertion in its coding sequence, causing an insertion of a bulky proline residue right in the middle of a transmembrane helix between residues 282 and 302 in the protein sequence (Supplementary Table S11). The insertion is likely to disrupt the transmembrane channel of the *N*-acetylglucosamine PTS permease, causing a loss of function. This suggests that the entire ATP-consuming PTS system may no longer be needed. This postulation is supported by the fact that only 45% (20/44) of the PTS genes are expressed, of which 55% (11/20) are down-regulated (Supplementary Fig. S18A). Furthermore, seven PTS genes show step-wise decrease in expression level throughout the entire 2,496-hour evolution (Supplementary Fig. S18B). Additionally, the PTS genes show negative correlations with the levels of biosynthetic processes (Supplementary Table S17).

Another energy-saving mutation is in *menC* in strain C, which has an in-frame IS186 (1343 bp) insertion (Supplementary Table S11). As *menC* is involved in the biosynthesis of menaquinone, the quinone for anaerobic respiration, which is not needed as most anaerobic genes are not expressed in the EA samples. Indeed, all five genes in this biosynthesis pathway are down-regulated in all six evolved strains (Supplementary Fig. S19A), and two (*menC* and *menF*) show step-wise decrease in expression level across all six strains (Supplementary Fig. S19B). Overall, a total of 11 genes are deleted or mutated, leading to reduced ATP consumption in all six evolved strains (Supplementary Table S11).

**Strengthened cell envelope through evolution.** Ethanol disrupts peptidoglycan crosslinking in cell wall and even disrupts its physical support for cell structure integrity<sup>6</sup>. We hypothesized that both components of the cell envelope, peptidoglycan and unsaturated fatty acids, of the evolved strains are reinforced through their increased biosynthesis. Indeed, most genes involved in peptidoglycan biosynthesis are consistently up-regulated in all six strains: 75% (3/4) of the initiation-stage genes, 62% (8/13) of the crosslinking stage including the committed-step gene *murA*, and 63% of the maturation-stage genes (Supplementary Fig. S20). Among the up-regulated genes, eight (*glmU*, *murA*, *murB*, *murC*, *murI*, *mrcA*, *dacA*, and *pbpG*,) have been experimentally identified as ethanol tolerance genes (Supplementary Tables S2 and S14). Furthermore, the effectiveness of this cell-wall reinforcement has been experimentally confirmed by over-expressing a *Lactobacillus plantarum murA* gene in *E. coli*<sup>23</sup>, and increased peptidoglycan synthesis provides physical protection against ethanol lytic effects<sup>6</sup>.

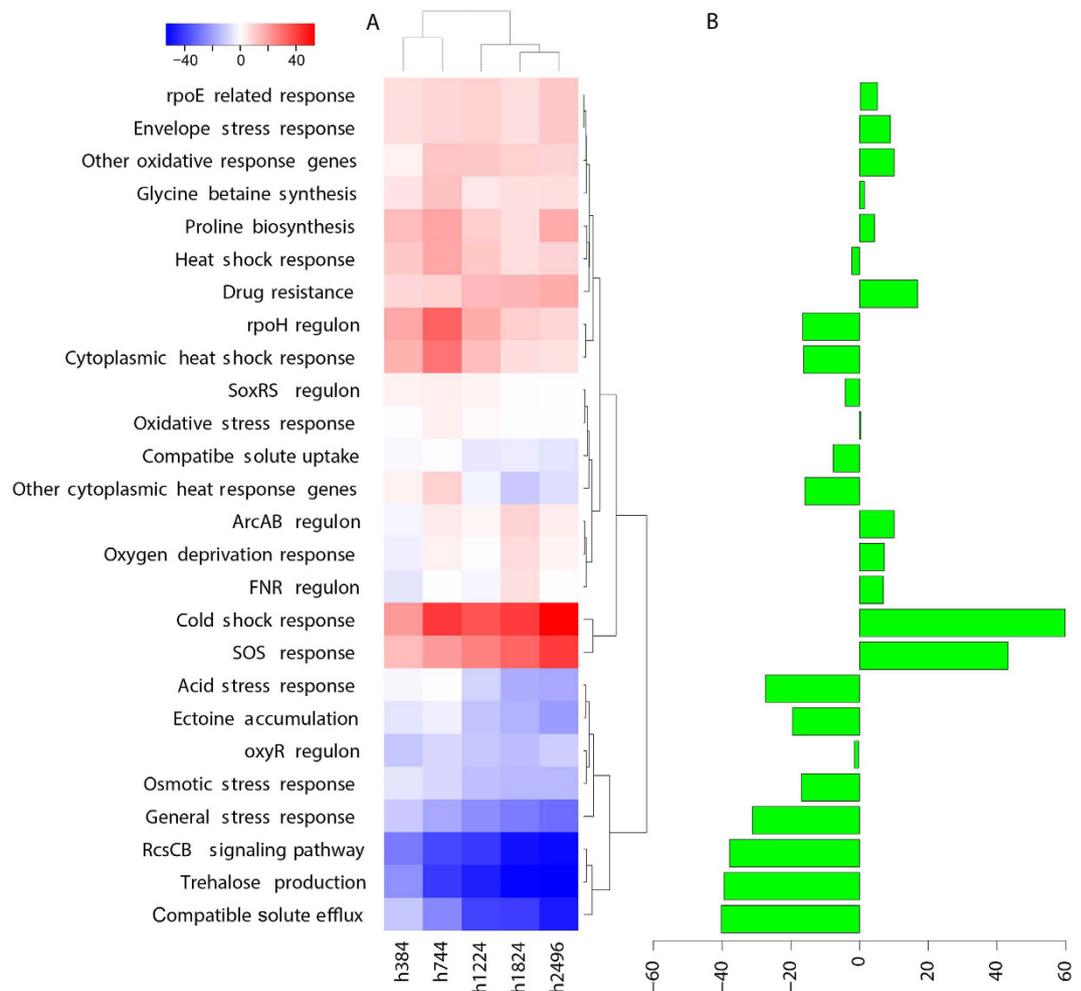
The proportion of unsaturated fatty acids, such as C18:1, in phospholipids is found to increase in response to ethanol<sup>1,2,6,24</sup>. The increased gene expression involved in fatty acid biosynthesis is given in Supplementary Results (Figs S27–S30). Such increased synthesis of unsaturated fatty acids will ameliorate the movement constraint on fatty acids imposed by ethanol inserted into membranes and aid restoring membrane proton potential<sup>1,3</sup>.

**Improved intracellular environment during evolution.** Reduced stresses as an indicator for improved cellular milieu: Given the increased ATP generation and restored membrane potential, we expected that cells will have less stressful intracellular environment. A first indicator is the reduction in aberrant TU expression by EBP TFs. Specifically, no aberrantly expressed TUs of the repressive EBP TFs are found in the evolved strains; and only 42% of the TUs of two activating EBP TFs, NtrC and PhoP, are aberrantly expressed (Supplementary Fig. S21). Therefore, aberrant expression of TUs due to ethanol binding is significantly reduced compared to those in the NEA samples (Wilcoxon test,  $p < 0.001$ ). Since no mutations are directly involved, this reduction could be attributed to the strengthened cell wall and membranes, which may limit the invasion of ethanol molecules into the cells<sup>25</sup>. In addition, 12 stress-response pathways in the evolved strains see steady decrease compared to the parental strain, including the response levels of acid stress, general stress, osmotic stress, trehalose production, and the RcsCB signaling pathway (Fig. 7).

Mutations in *cspC*, *proQ* and *hns* may also contribute to the reduced stress responses. *cspC* encodes a  $\rho$ -independent anti-terminator, which has an IS5 insertion in strains E and F. Its knockout is shown to increase growth rate<sup>20</sup>, consistent with the reported growth gain due to function loss of this gene<sup>26</sup>. Mechanistically, mutated *cspC* decreases the stability of the *rpoS* mRNA, which in turn decreases the expression level of stress response gene *proP* (osmolyte:H<sup>+</sup> symporter)<sup>27</sup>, giving rise to reduced levels of both the general and osmotic stress responses (Figs 7 and S22). *proQ* is an RNA chaperone that post-transcriptionally controls the expression level of *proP*<sup>28</sup>. A mutation (Leu91Gln) occurs in its critical protein region, contributing to the reduced expression of *proP*. *hns* is a global dual-regulator, and an IS5 insertion in the promoter regions in five of the six evolved strains (Supplementary Table S11) leads to its up-regulation in these strains (Supplementary Fig. S23A). Consequently, some *hns* regulon genes are differentially expressed (Supplementary Fig. S23B). Among the down-regulated genes, *gadW*, *nhaA* and *nhaR* are involved in acid-stress response, and *osmC* involved in osmotic stress response (Supplementary Fig. S23D), resulting in reduced responses to these two stress types.

**Up-regulated stress response as effective counter resistance.** Besides the above reduced stress responses, ten stress responses are increased to varying degrees. SOS response and cold shock response are among the most up-regulated (Fig. 7). These two processes have high positive correlations with DNA replication with correlation coefficients of 0.75 and 0.72, respectively. In contrast to the decreasing stress pathways as ‘passive’ indicators of improved cellular milieu, these two up-regulated responses may act as ‘active’ counter resistance to ethanol-induced stresses.

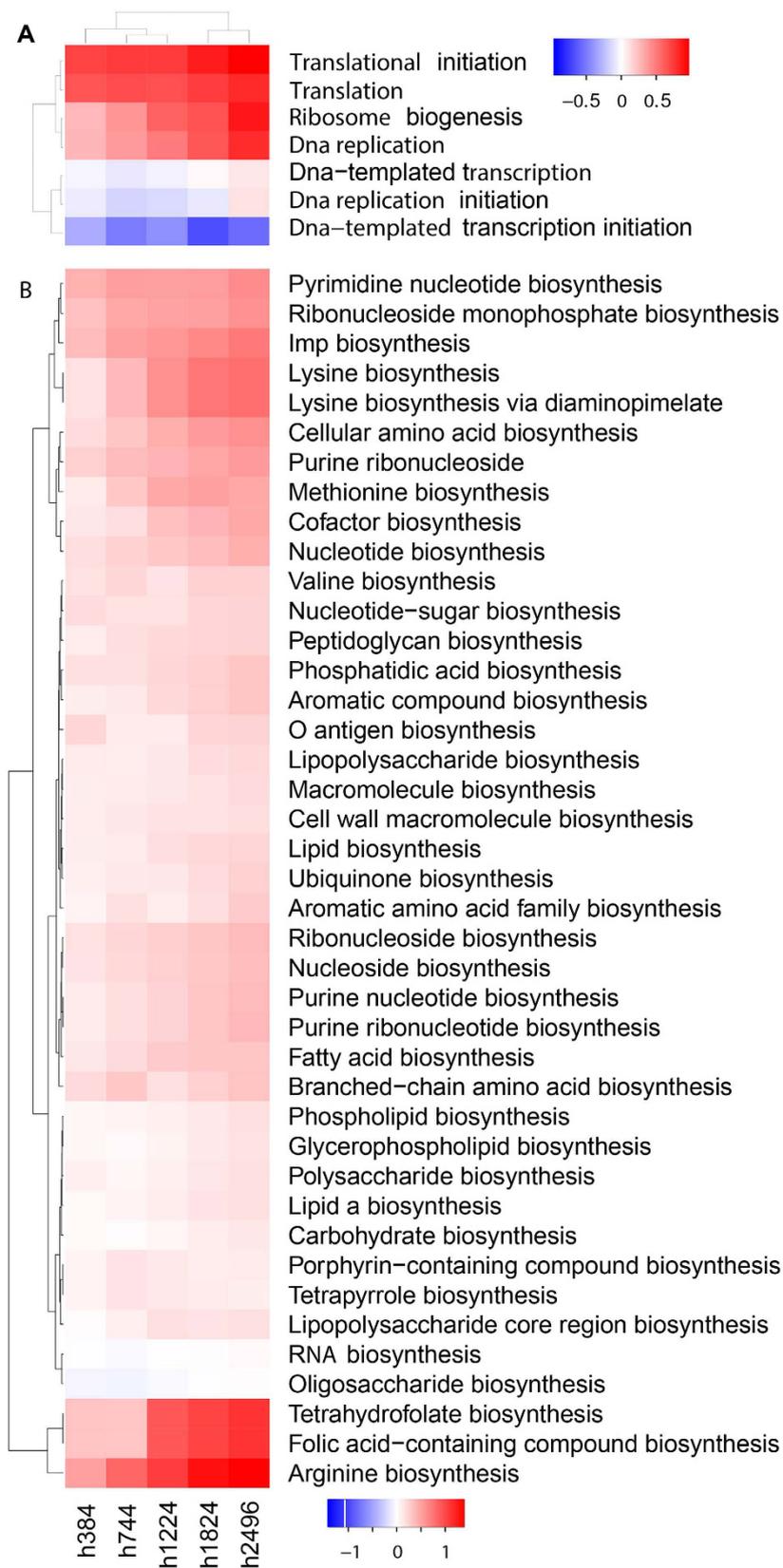
Ethanol induces lipid modification and stalling of gene expression<sup>1,24,29</sup>, analogous to the main effects of cold shock<sup>24,30</sup>. For the SOS response, 78% (21/27) genes are expressed (Supplementary Fig. S24A) and nine are consistently up-regulated during evolution (Supplementary Fig. S24B), including five key DNA repair genes *recN*, *uvrD*, *cho*, *dinG*, and *dinB*. The specific function of each differentially expressed gene is provided in Supplementary Table S18. Since the ROS level is low in the EA samples and therefore low DNA damage, SOS response may be active to induce error-prone mismatches, thus increasing the mutation rate for adaptation



**Figure 7. 26 stress responses in six EA samples.** (A) percentage of changes in mean expression level of each of the six strains to that of the parent strain in each response; (B) the difference (%) between time 0 and h2,496 in EA samples. The *x* axis represents the samples, and the *y* axis is for the proportion of differentially expressed genes.

to ethanol<sup>31,32</sup>. In addition, most cold-shock protein (*csp*) genes are up-regulated with genes involved in DNA repair, transcription, and translation (Supplementary Fig. S25A). Among them, six genes: *cspA*, *deaS*, *hns*, *hscA*, *hscB* and *rbfA* show step-wise up-regulation (Supplementary Fig. S25B), with *hscA* being a known ethanol tolerance gene (Supplementary Table S14). The functions of the cold-shock response genes are provided in Supplementary Table S19. Two mutated genes are involved in this response: *hns* and *relA*. *hns*, a global dual-regulator, is a key organizing protein involved in chromosome structure (supercoiling) and shows step-wise up-regulation due to an IS5 insertion in five of the six EA strains (Supplementary Fig. S23A). Another mutated gene *relA* has three missense SNPs, each in one EA strain (C, E and F) and two of them in critical domains (Supplementary Table S11); this functional loss in *relA* is consistent with the observation that *relA* is down-regulated in all six EA strains. Furthermore, mutated *relA* has been shown to confer the biggest fitness gain among all identified mutations in strain F<sup>20</sup>. *RelA* is a synthase for the global regulatory molecules (p)ppGpp, which activates the expression of cold-shock proteins<sup>33</sup>. Therefore, mutated *relA* appears to synthesize fewer (p)ppGpp, causing lower expressions of the cold-shock proteins. We verified this with the expression level of ribosomal RNA genes, which is reduced by (p)ppGpp<sup>34</sup>. Therefore, the up-regulation of ribosomal biogenesis genes should be a result of mutated or down-regulated *relA*; indeed, this is what we observed in six EA strains (Supplementary Fig. S26). Therefore, both these reduced ‘passive’ and increased ‘active’ stress response both contributed to growth restoration in the EA strains.

**Coordinated restoration of biological processes in the EA strains.** To fully understand the adaptation process to ethanol-induced stresses, we compared differential expressions of genes involved in stress responses, transcription, translation, DNA replication, and biosynthetic processes during evolution, to detect time-dependent patterns, particularly monotonic patterns, at a systems level. First, most processes gradually recover or improve with multiple, not one, steps (time points) (Figs 7 and 8). Second, these improvements appear to proceed in a coordinated fashion with certain processes taking place prior to others. For example, translation and biosynthesis



**Figure 8.** Time-dependent changes in transcription, translation, DNA replication and biosynthetic processes in six EA strain during evolution. (A) transcription, translation, DNA replication; and (B) biosynthetic processes.

of amino acids and nucleotides recover first, followed by biosynthesis of macromolecules including peptidoglycan and fatty acids; and the last to recover is transcription and DNA replication (Fig. 8). Most stress responses show a steady decline or increase (e.g., glycine betaine biosynthesis and heat shock response). Third, there is a systems-level optimization of biological processes: essential pathways are enhanced (e.g., aerobic respiration) and non-essential pathways (e.g., PTS system) are repressed at certain points during the adaptive evolution. Last, an optimized system provides stronger tolerance against ethanol and improved growth in the presence of ethanol.

## Discussion

A key limiting factor in ethanol production is ethanol's toxicity to the host cells. Despite extensive studies on this subject since 1970s, e.g., refs 2, 6, 35–37, new targets affected by ethanol continue to be identified, e.g., ref. 29. Therefore, a comprehensive systems-level understanding of ethanol toxicity is needed. By comparatively analyzing transcriptomic and genomic data of NEA vs. EA samples and known ethanol tolerance genes, we have proposed a model for how ethanol stresses the host cells and how stressed cells develop tolerance through evolution, both at a systems level. In our model, ethanol induces stresses at multiple levels of cellular activities, spanning a wide range of cellular processes. They include cell envelope integrity, multiple stress types, increased ROS production and reduced O<sub>2</sub> level, compensated ATP production via three different mechanisms, macromolecule biosynthesis, and protein-ethanol binding and altered protein functions.

Our key findings are that ethanol disrupts the following processes: (i) weakening cell wall and cell membranes, leading to reduced cross-membrane proton gradient and hence reduced ATP production; (ii) induction of multiple stress types; (iii) increased ROS production due to stresses, leading to reduced intracellular O<sub>2</sub> level and hence decreased aerobic respiration for ATP production; (iv) the reduced ATP production leading to a number of cellular-level changes such as reduced biosynthesis of macromolecules and therefore reduced cell proliferation, and increased fatty acid oxidation for energy production; and (v) altered functions of over 200 proteins including dozens of transcription factors, leading to extensive impact on many biological processes; and (vi) that these changes are functionally linked with each other, all stemming from the reduced cross-membrane proton gradient and increased intracellular ROS. We anticipate that this new understanding would guide more effective designing of better ethanol-tolerant strains.

We also proposed a model for how the evolved *E. coli* strains adapted to the ethanol-induced stresses and regained viability through selection of specific mutations, which enabled changes in certain basic behaviors of the cells, particularly in reducing ATP consumption for non-essential processes. The regained ATP sufficiency and biosynthesis of macromolecules put the cells on an evolutionary trajectory towards restoration. Our key findings are: (1) reduced non-essential energy consumption enabled by mutations allows the stressed cells to use ATP towards more essential activities for their survival, such as increased biosynthesis towards strengthened cell membrane, which in turn enables the cells to gradually regain their cross-membrane proton gradients and hence ATP production through respiration, leading to reduced overall stress level and hence ROS; (2) the reduced ROS production increases the intracellular O<sub>2</sub> level for aerobic respiration, offering the most essential foundation for recovering the normal cellular functions; (3) the increased biosynthesis of macromolecules coupled with up-regulation of ethanol tolerance genes, allow for increased cell division; and (4) all these activities together improve the overall intracellular condition.

One defining characteristic of our model is that all processes affected by ethanol are interconnected through energy and stresses, in both the stressed and evolved cells. Thus, our model can serve as a unifying framework for integrating all ethanol targets identified so far. Previously identified impacts such as cell membrane proton gradient<sup>4</sup> and ROS production<sup>13,38</sup> can be viewed as the foundational changes; alterations in respiration (e.g., ArcAB regulon)<sup>9</sup> and anaerobic fatty acid oxidation are the next-level responses; and the reduced translation and biosynthesis of macromolecules are the last and most comprehensive impact. The whole process is reverted in the evolved cells but starts with mutation-enabled reduction in non-essential ATP consumption, which gradually leads to improved biosynthesis, reduced stresses and improved ATP generation in an iterative manner, and finally increased DNA replication and growth. We have found two experimental supports for our model. For example, fortifying cell envelope by overexpressing genes, e.g., *murAG* for peptidoglycan biosynthesis and for unsaturated fatty biosynthesis<sup>3,23,24</sup> have been successful in making ethanol tolerance bacteria. Ethanol-affected proteins (e.g., *rho*, *metJ*, and *rpsQ*) in transcription and translation are reinforced through mutations, so as to be more robust in the presence of ethanol<sup>29</sup>.

ROS production is an important finding in our study, which can be generated due to stresses as well as from damaged electron transport chain<sup>39</sup>. Another major finding is that ethanol can directly affect protein functions, a question long suspected but not confirmed at the physiological concentrations in bacteria<sup>6,40,41</sup>. A most 'direct' evidence in this regard comes from *rps*, which acquired beneficial mutations and reduced translational misreading in ethanol stressed cells<sup>29</sup>. We showed that ethanol alters protein functions by demonstrating that TFs can have their functions altered when bound with ethanol, hence impacting the functions at a pathway level. This may be the very reason why overexpressing protein chaperones GroESL increases ethanol tolerance in *E. coli*<sup>42</sup>.

Besides mutations and adaptive gene expression, microRNAs and non-coding RNAs have also been found to confer tolerance to stressors<sup>43–46</sup>. As most of microarray probes used in this study (Table 1) are not designed for these small RNAs, we cannot confirm their contribution. Future work is needed to study this issue.

Our qualitative model may need mechanistic validation with experimental studies, which could lead to a further improved quantitative model and better understanding about the fundamentals of ethanol stress, responses and adaptation. The merit of the study lies in a systems-level integration of a collection of observations about individual changes. Together, the model sheds new light on the complex issues about ethanol stresses and adaptation.

Accession No.	Sample name	Ethanol (v/v)	Replicate	<i>E. coli</i> K12 strain	Microarray platform	Medium
GSE3665	GSE3665A <sup>b</sup>	0 vs. 2.5%	6 (3 + 3)	DH5 $\alpha$	GPL3154	M9 with 5 g/L D-glucose and 1 mM thiamine
	GSE3665B <sup>b</sup>	0 vs. 5%	12 (8 + 4)	DH5 $\alpha$	GPL3154	M9 with 5 g/L D-glucose and 1 mM thiamine
GSE13444	GSE13444	3%	12 (8 + 4)	BW25113	GPL5113	MOPS with 0.2% glucose
GSE17526	GSE17526A <sup>b</sup>	15% (15 min)	6 (3 + 3)	BW25113	GPL3154	M9 + 4 g/L glucose
	GSE17526B <sup>b</sup>	15% (15 min)	6 (3 + 3)	BW25113	GPL3154	M9 + 4 g/L glucose
	GSE17526C <sup>b</sup>	15% (15 min)	6 (3 + 3)	BW25113	GPL3154	M9 + 4 g/L glucose
	GSE17526D <sup>b</sup>	15% (15 min)	6 (3 + 3)	BW25113	GPL3154	M9 + 4 g/L glucose
GSE59050 <sup>a</sup>	Parent/control	5%	6	W3310	GPL13336	M9 + 4 g/L glucose
	h384	5%	6	W3310	GPL13336	M9 + 4 g/L glucose
	h744	5%	6	W3310	GPL13336	M9 + 4 g/L glucose
	h1,224	5%	6	W3310	GPL13336	M9 + 4 g/L glucose
	h1,824	5%	6	W3310	GPL13336	M9 + 4 g/L glucose
	h2,496	5%	6	W3310	GPL13336	M9 + 4 g/L glucose

**Table 1. Four microarray datasets containing seven control-treatment comparisons in response to ethanol in two *E. coli* strains that are not evolved, and one strain that is evolved for 2,496 h with six replicates in the presence of ethanol.** <sup>a</sup>This data set contains parent (non-evolved) strain and evolved strains for 2,496 h in the presence of ethanol. <sup>b</sup>These samples are used in correlation analyses with different biological processes.

## Data and Methods

***E. coli* transcriptomic data.** Four sets of microarray gene-expression data, collected on three *E. coli* strains with ethanol treatment and control, are downloaded from the NCBI Gene Expression Omnibus<sup>47</sup> and used in this study. The three strains are W3110, BW25113, and DH5 $\alpha$ , which differ in coding regions of a few genes. For the collected data, these strains are grown in minimal (M9 or MOPS) media supplemented with glucose, and most are treated with 2.5%~5% (v/v) ethanol and one with a 15% ethanol shock for 15 min. Three of the four datasets are produced on Affymetrix microarrays (Santa Clara, CA, USA) and one on custom arrays, all covering at least 4,305 genes of *E. coli* K12 MG1655. Three datasets are for non-ethanol-adapted cells and one for ethanol-adapted cells. Table 1 summarizes the detailed information of these datasets. In the NEA samples, each treatment has three or eight replicates and the corresponding control has three or four replicates. In the EA samples, data are collected on six strains evolved in parallel from a common parent (GSE59050) at five time points: 384, 744, 1,224, 1,824, and 2,496 hours with ethanol treatment.

**Differential gene-expression analysis and functional enrichment.** The three NEA datasets have a total of seven treatment-*versus*-control pairs (Table 1 and S1). The EA dataset has five evolved-*versus*-the-parental-strain pairs. Differential gene-expression analyses were conducted between each treatment or evolved sample and the matching control sample. For the NEA pairs, gene-expression data were analyzed using the R packages GEOquery<sup>48</sup> and affy<sup>49</sup>. A differentially expressed gene (DEG) between each pair of samples is defined as having a fold change in expression  $>1.2$  or  $<1/1.2$  for up-regulation or down-regulation in Wilcoxon signed-rank test ( $p < 0.05$ ), respectively. For the EA datasets, differential expression analyses were performed between the parental strain and each of the six evolved strains, each collected at five time points during the adaptive evolution in ethanol<sup>20</sup> (Supplementary Table S1). For the EA samples, we used more stringent criteria for determining DEGs since there was only one control, i.e., the parental strain, *versus* six evolved strains: the six evolved strains must all have greater (or less) expression levels than the parent strain, and the mean difference is greater than 20% compared to the expression level in the parent strain.

A pathway enrichment analysis was conducted separately over the up- and down-regulated genes using DAVID<sup>50</sup> against the Gene Ontology database<sup>51</sup>.

**Identification of ethanol-binding proteins.** We predicted ethanol-binding proteins (EBPs) using a structure-based computational approach. The reference structure is taken from alcohol dehydrogenase (AdhE), the only known ethanol-binding enzyme in *E. coli*, which catalyzes the inter-conversion between ethanol and acetyl-CoA<sup>9,52</sup>. Specifically, we searched the protein structural data in the Protein Data Bank<sup>53</sup> using keyword “alcohol dehydrogenase” in three enzyme classes: 1.1.1.1, 1.1.1.2 and 1.1.2.8, which gives rise to 130, 27 and 3 structures, respectively. Since all structures in the same class were highly similar, only one representative was chosen from each class: namely 2jhf, 4cvb and 1zk4 for ethanol binding analysis. We then downloaded the structural models for 3,784 *E. coli* proteins predicted by QUARK<sup>54</sup>, a state-of-the-art prediction program. These models were structurally aligned to 2jhf, 4cvv, and 1zk4 using SP-align<sup>55</sup> for identification of possible binding sites with ethanol. It is noteworthy that structural alignment for binding pocket identification by tools like SP-align has been shown to be powerful enough for identifying proteins with the same molecular function<sup>55–58</sup>. 213 *E. coli* proteins are found to have similar binding pockets with the three alcohol dehydrogenase proteins with SP-score  $>0.5$ , a widely used threshold for identifying proteins with the same structural fold. Functional information of these proteins, including proteins that interact with, are retrieved from the UniProt database<sup>59</sup>.

**Altered gene regulation by ethanol-bound transcription factors.** Among the 213 predicted EBPs, 23 are transcription factors (TFs). So we analyzed the expression profiles of the transcription units (TUs) regulated by each of these TFs as defined in regulonDB<sup>60</sup>, to determine whether their regulatory functions are altered by ethanol binding. We used TUs that are regulated by EBP TFs only. To quantify aberrant expressions for each such TU, we first calculate the differential expression of the TF and genes in each TU between control and ethanol treated samples. We then divide the TUs into two categories: activated or repressed based on the direction of the regulation. For an activated TU, it is considered as having aberrant expression if (1) the TF is not differentially expressed but a TU is down-regulated, or (2) the TF is up-regulated but the TU is not. For a repressed TU, it is considered as aberrantly expressed if (1) the TF is not differentially expressed but a TU is up-regulated, or (2) the TF is up-regulated and the TU is also up-regulated. The level of aberrantly expressed TUs is defined as the percentage of the all the TUs of TF's regulon.

**Ethanol tolerance genes and stress response pathways.** Numerous genes have been found to increase ethanol tolerance in *E. coli* when overexpressed or knocked out<sup>7,42,61</sup>. We found these genes (Supplementary Table S2) through a systematic literature search using keywords 'ethanol tolerance' and '*E. coli*' in Web of Science and PubMed. We also collected 26 stress-response genes (Supplementary Table S3) as identified by Storz and Hengge<sup>62</sup> and curated by Goodarzi *et al.*<sup>9</sup>, respectively.

## References

- Ingram, L. O. Adaptation of membrane lipids to alcohols. *J. Bacteriol.* **125**, 670–678 (1976).
- Dombek, K. M. & Ingram, L. O. Effects of ethanol on the *Escherichia coli* plasma membrane. *J. Bacteriol.* **157**, 233–239 (1984).
- Huffer, S., Clark, M. E., Ning, J. C., Blanch, H. W. & Clark, D. S. Role of alcohols in growth, lipid composition, and membrane fluidity of yeasts, bacteria, and archaea. *Appl. Environ. Microbiol.* **77**, 6400–6408, doi: 10.1128/aem.00694-11 (2011).
- Cartwright, C. P. *et al.* Ethanol dissipates the proton-motive force across the plasma membrane of *Saccharomyces cerevisiae*. *J. Gen. Microbiol.* **132**, 369–377, doi: 10.1099/00221287-132-2-369 (1986).
- Damore, T., Panchal, C. J., Russell, I. & Stewart, G. G. A study of ethanol tolerance in yeast. *Crit. Rev. Biotechnol.* **9**, 287–304, doi: 10.3109/07388558909036740 (1989).
- Ingram, L. O. & Vreeland, N. S. Differential effects of ethanol and hexanol on the *Escherichia coli* cell envelope. *J. Bacteriol.* **144**, 481–488 (1980).
- Woodruff, L. B. A. *et al.* Genome-scale identification and characterization of ethanol tolerance genes in *Escherichia coli*. *Metab. Eng.* **15**, 124–133, doi: http://dx.doi.org/10.1016/j.ymben.2012.10.007 (2013).
- Gonzalez, R. *et al.* 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, doi: 10.1021/bp025658q (2003).
- Goodarzi, H. *et al.* Regulatory and metabolic rewiring during laboratory evolution of ethanol tolerance in *E. coli*. *Mol. Syst. Biol.* **6**, 1–12, doi: 10.1038/msb.2010.33 (2010).
- Woodruff, L. B. A., Boyle, N. R. & Gill, R. T. Engineering improved ethanol production in *Escherichia coli* with a genome-wide approach. *Metab. Eng.* **17**, 1–11, doi: http://dx.doi.org/10.1016/j.ymben.2013.01.006 (2013).
- Brynildsen, M. P. & Liao, J. C. An integrated network approach identifies the isobutanol response network of *Escherichia coli*. *Mol. Syst. Biol.* **5**, 1–13, doi: 10.1038/msb.2009.34 (2009).
- Horinouchi, T. *et al.* Transcriptome analysis of parallel-evolved *Escherichia coli* strains under ethanol stress. *BMC Genomics* **11**, 579 (2010).
- Pérez-Gallardo, R. V. *et al.* Reactive oxygen species production induced by ethanol in *Saccharomyces cerevisiae* increases because of a dysfunctional mitochondrial iron–sulfur cluster assembly system. *FEMS Yeast Res.* **13**, 804–819, doi: 10.1111/1567-1364.12090 (2013).
- Imlay, J., Chin, S. & Linn, S. Toxic DNA damage by hydrogen peroxide through the Fenton reaction *in vivo* and *in vitro*. *Science* **240**, 640–642, doi: 10.1126/science.2834821 (1988).
- Henle, E. S. *et al.* Sequence-specific DNA cleavage by Fe<sup>2+</sup> + -mediated Fenton reactions has possible biological implications. *J. Biol. Chem.* **274**, 962–971, doi: 10.1074/jbc.274.2.962 (1999).
- Luo, Y., Henle, E. S. & Linn, S. Oxidative damage to DNA constituents by iron-mediated Fenton reactions: The deoxycytidine family. *J. Biol. Chem.* **271**, 21167–21176, doi: 10.1074/jbc.271.35.21167 (1996).
- Flint, D. H. & Allen, R. M. Iron-sulfur proteins with nonredox functions. *Chem. Rev.* **96**, 2315–2334, doi: 10.1021/cr950041r (1996).
- Shalel Levanon, S., San, K.-Y. & Bennett, G. N. Effect of oxygen on the *Escherichia coli* ArcA and FNR regulation systems and metabolic responses. *Biotechnol. Bioeng.* **89**, 556–564, doi: 10.1002/bit.20381 (2005).
- Salgado, H. *et al.* RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res.* **41**, D203–D213, doi: 10.1093/nar/gks1201 (2013).
- Horinouchi, T. *et al.* Phenotypic convergence in bacterial adaptive evolution to ethanol stress. *BMC Evol. Biol.* **15**, 1–14, doi: 10.1186/s12862-015-0454-6 (2015).
- Horinouchi, T. *et al.* Transcriptome analysis of parallel-evolved *Escherichia coli* strains under ethanol stress. *BMC Genomics* **11**, 1–11, doi: 10.1186/1471-2164-11-579 (2010).
- Garriga, X. *et al.* *nrdD* and *nrdG* genes are essential for strict anaerobic growth of *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **229**, 189–192, doi: http://dx.doi.org/10.1006/bbrc.1996.1778 (1996).
- Yuan, Y. *et al.* Overexpression of the *Lactobacillus plantarum* peptidoglycan biosynthesis *murA2* gene increases the tolerance of *Escherichia coli* to alcohols and enhances ethanol production. *Appl. Microbiol. Biotechnol.* **98**, 8399–8411, doi: 10.1007/s00253-014-6004-0 (2014).
- Berger, B., Carty, C. E. & Ingram, L. O. Alcohol-induced changes in the phospholipid molecular species of *Escherichia coli*. *J. Bacteriol.* **142**, 1040–1044 (1980).
- Ingram, L. O. Ethanol tolerance in bacteria. *Crit. Rev. Biotechnol.* **9**, 305–319, doi: 10.3109/07388558909036741 (1989).
- Rath, D. & Jawali, N. Loss of expression of *cspC*, a cold shock family gene, confers a gain of fitness in *Escherichia coli* K-12 strains. *J. Bacteriol.* **188**, 6780–6785, doi: 10.1128/jb.00471-06 (2006).
- Phadtare, S. & Inouye, M. Role of CspC and CspE in regulation of expression of RpoS and UspA, the stress response proteins in *Escherichia coli*. *J. Bacteriol.* **183**, 1205–1214, doi: 10.1128/jb.183.4.1205-1214.2001 (2001).
- Chaulk, S. G. *et al.* ProQ Is an RNA chaperone that controls ProP levels in *Escherichia coli*. *Biochemistry* **50**, 3095–3106, doi: 10.1021/bi101683a (2011).
- Haft, R. J. F. *et al.* Correcting direct effects of ethanol on translation and transcription machinery confers ethanol tolerance in bacteria. *Proc. Natl. Acad. Sci. USA.* **111**, E2576–E2585, doi: 10.1073/pnas.1401853111 (2014).
- Phadtare, S. Recent developments in bacterial cold-shock response. *Curr. Issues Mol. Biol.* **6**, 125–136 (2004).

31. Boshoff, H. I. M., Reed, M. B., Barry Iii, C. E. & Mizrahi, V. DnaE2 polymerase contributes to *in vivo* survival and the emergence of drug resistance in *Mycobacterium tuberculosis*. *Cell* **113**, 183–193, doi: [http://dx.doi.org/10.1016/S0092-8674\(03\)00270-8](http://dx.doi.org/10.1016/S0092-8674(03)00270-8) (2003).
32. Cirz, R. T. *et al.* Inhibition of mutation and combating the evolution of antibiotic resistance. *PLoS Biol.* **3**, e176, doi: [10.1371/journal.pbio.0030176](https://doi.org/10.1371/journal.pbio.0030176) (2005).
33. Jones, P. G., Cashel, M., Glaser, G. & Neidhardt, F. C. Function of a relaxed-like state following temperature downshifts in *Escherichia coli*. *J. Bacteriol.* **174**, 3903–3914 (1992).
34. Chatterji, D., Fujita, N. & Ishihama, A. The mediator for stringent control, ppGpp, binds to the  $\beta$ -subunit of *Escherichia coli* RNA polymerase. *Genes Cells* **3**, 279–287, doi: [10.1046/j.1365-2443.1998.00190.x](https://doi.org/10.1046/j.1365-2443.1998.00190.x) (1998).
35. Horne, J. R. & Erdmann, V. A. Effects of ethanol, methanol and different antibiotics on the ATPase and GTPase activities associated with *B. Stearothermophilus* 5 S RNS protein complex. *FEBS Lett.* **42**, 42–45, doi: [10.1016/0014-5793\(74\)80274-7](https://doi.org/10.1016/0014-5793(74)80274-7) (1974).
36. Nagodawithana, T. W., Whitt, J. T. & Cutaia, A. J. Study of the feedback effect of ethanol on selected enzymes of the glycolytic pathway. *J. Am. Soc. Brew. Chem.* **35**, 179–183, doi: [10.1094/ASBCJ-35-0179](https://doi.org/10.1094/ASBCJ-35-0179) (1977).
37. Ingram, L. O., Dickens, B. F. & Buttke, T. M. In *Biological Effects of Alcohol* Vol. 126 *Advances in Experimental Medicine and Biology* (ed. Henri, Begleiter) Ch. 24, 299–337 (Springer US, 1980).
38. Frick, K., Schulte, M. & Friedrich, T. Reactive oxygen species production by *Escherichia coli* respiratory Complex I. *Biochemistry* **54**, 2799–2801, doi: [10.1021/acs.biochem.5b00160](https://doi.org/10.1021/acs.biochem.5b00160) (2015).
39. González-Flecha, B. & Demple, B. Metabolic sources of hydrogen peroxide in aerobically growing *Escherichia coli*. *J. Biol. Chem.* **270**, 13681–13687, doi: [10.1074/jbc.270.23.13681](https://doi.org/10.1074/jbc.270.23.13681) (1995).
40. Millar, D., Griffiths-Smith, K., Algar, E. & Scopes, R. Activity and stability of glycolytic enzymes in the presence of ethanol. *Biotechnol. Lett.* **4**, 601–606, doi: [10.1007/BF00127792](https://doi.org/10.1007/BF00127792) (1982).
41. Osman, Y. A. & Ingram, L. O. Mechanism of ethanol inhibition of fermentation in *Zymomonas mobilis* CP4. *J. Bacteriol.* **164**, 173–180 (1985).
42. Zingaro, K. A. & Terry Papoutsakis, E. GroESL overexpression imparts *Escherichia coli* tolerance to i-, n-, and 2-butanol, 1,2,4-butanetriol and ethanol with complex and unpredictable patterns. *Metab. Eng.* **15**, 196–205, doi: <http://dx.doi.org/10.1016/j.ymben.2012.07.009> (2013).
43. Borden, J. R., Jones, S. W., Indurthi, D., Chen, Y. & Terry Papoutsakis, E. A genomic-library based discovery of a novel, possibly synthetic, acid-tolerance mechanism in *Clostridium acetobutylicum* involving non-coding RNAs and ribosomal RNA processing. *Metab. Eng.* **12**, 268–281, doi: <http://dx.doi.org/10.1016/j.ymben.2009.12.004> (2010).
44. Gómez-Lozano, M., Marvig, R. L., Tulstrup, M. V. & Molin, S. Expression of antisense small RNAs in response to stress in *Pseudomonas aeruginosa*. *BMC Genomics* **15**, 783, doi: [10.1186/1471-2164-15-783](https://doi.org/10.1186/1471-2164-15-783) (2014).
45. Zeng, X., Zhang, X. & Zou, Q. Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks. *Briefings in Bioinformatics* **17**, 193–203, doi: [10.1093/bib/bbv033](https://doi.org/10.1093/bib/bbv033) (2015).
46. Tang, W., Liao, Z. & Zou, Q. Which statistical significance test best detects oncomiRNAs in cancer tissues? An exploratory analysis. *Oncotarget* **7** (2016).
47. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–D995, doi: [10.1093/nar/gks1193](https://doi.org/10.1093/nar/gks1193) (2013).
48. Davis, S. & Meltzer, P. S. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846–1847, doi: [10.1093/bioinformatics/btm254](https://doi.org/10.1093/bioinformatics/btm254) (2007).
49. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315, doi: [10.1093/bioinformatics/btg405](https://doi.org/10.1093/bioinformatics/btg405) (2004).
50. Huang, H., Lu, X., Liu, Y., Haaland, P. & Marron, J. S. R/DWD: distance-weighted discrimination for classification, visualization and batch adjustment. *Bioinformatics* **28**, 1182–1183, doi: [10.1093/bioinformatics/bts096](https://doi.org/10.1093/bioinformatics/bts096) (2012).
51. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat Genet* **25**, 25–29, doi: [10.1038/75556](https://doi.org/10.1038/75556) (2000).
52. Membrillo-Hernández, J. *et al.* 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, doi: [10.1074/jbc.M005464200](https://doi.org/10.1074/jbc.M005464200) (2000).
53. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242, doi: [10.1093/nar/28.1.235](https://doi.org/10.1093/nar/28.1.235) (2000).
54. Xu, D. & Zhang, Y. *Ab initio* protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins: Struct., Funct., Bioinf.* **80**, 1715–1735, doi: [10.1002/prot.24065](https://doi.org/10.1002/prot.24065) (2012).
55. Yang, Y., Zhan, J., Zhao, H. & Zhou, Y. A new size-independent score for pairwise protein structure alignment and its application to structure classification and nucleic-acid binding prediction. *Proteins: Struct., Funct., Bioinf.* **80**, 2080–2088, doi: [10.1002/prot.24100](https://doi.org/10.1002/prot.24100) (2012).
56. Zhao, H., Yang, Y., von Itzstein, M. & Zhou, Y. Carbohydrate-binding protein identification by coupling structural similarity searching with binding affinity prediction. *J. Comput. Chem.* **35**, 2177–2183, doi: [10.1002/jcc.23730](https://doi.org/10.1002/jcc.23730) (2014).
57. Brylinski, M. & Skolnick, J. FINDSITE<sup>EHM</sup>: a threading-based approach to ligand homology modeling. *PLoS Comput Biol* **5**, e1000405, doi: [10.1371/journal.pcbi.1000405](https://doi.org/10.1371/journal.pcbi.1000405) (2009).
58. Roy, A. & Zhang, Y. Recognizing protein-ligand binding sites by global structural alignment and local geometry refinement. *Structure* **20**, 987–997, doi: <http://dx.doi.org/10.1016/j.str.2012.03.009> (2012).
59. Consortium, T. U. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–D212, doi: [10.1093/nar/gku989](https://doi.org/10.1093/nar/gku989) (2015).
60. Huerta, A. M., Salgado, H., Thieffry, D. & Collado-Vides, J. RegulonDB: A database on transcriptional regulation in *Escherichia coli*. *Nucleic Acids Res.* **26**, 55–59, doi: [10.1093/nar/26.1.55](https://doi.org/10.1093/nar/26.1.55) (1998).
61. Chong, H. *et al.* Improving ethanol tolerance of *Escherichia coli* by rewiring its global regulator cAMP receptor protein (CRP). *PLoS ONE* **8**, e57628, doi: [10.1371/journal.pone.0057628](https://doi.org/10.1371/journal.pone.0057628) (2013).
62. Storz, G. & Hengge, R. (American Society of Microbiology, Washington, 2011).

## Acknowledgements

This work was supported by the U.S. Department of Energy BioEnergy Science Center [DE-PS02-06ER64304]; National Health and Medical Research Council of Australia (1059775 and 1083450); and State of South Dakota Research Innovation Center, the Agriculture Experiment Station of South Dakota State University.

## Author Contributions

H.C., W.D. and Y.X. conceived and designed project; H.C. and W.D. collected analyzed transcriptional and curated ethanol tolerance gene data; Y.S., G.L., and Q.M. analyzed genome resequencing data; Y.Y. and Y.Z. contributed ethanol binding proteins. All authors wrote the manuscripts.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing Interests:** The authors declare no competing financial interests.

**How to cite this article:** Cao, H. *et al.* Systems-level understanding of ethanol-induced stresses and adaptation in *E. coli*. *Sci. Rep.* 7, 44150; doi: 10.1038/srep44150 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017