Genomics, Proteomics, and Bioinformatics:

Global Gene Expression Profiling in *Escherichia coli* K12: THE EFFECTS OF OXYGEN AVAILABILITY AND FNR

Kirsty Salmon, She-pin Hung, Kathy Mekjian, Pierre Baldi, G. Wesley Hatfield and Robert P. Gunsalus

J. Biol. Chem. 2003, 278:29837-29855.
doi: 10.1074/jbc.M213060200 originally published online May 15, 2003

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 278, No. 32, Issue of August 8, pp. 29837–29855, 2003

Printed in U.S.A.

Global Gene Expression Profiling in *Escherichia coli* K12

THE EFFECTS OF OXYGEN AVAILABILITY AND FNR*[S]

Received for publication, December 20, 2002, and in revised form, May 15, 2003
Published, JBC Papers in Press, May 15, 2003, DOI 10.1074/jbc.M213060200

Kirsty Salmon‡§, She-pin Hung§¶‖, Kathy Mekjian‡, Pierre Baldi‖‡‡, G. Wesley Hatfield¶‖§§, and Robert P. Gunsalus‡¶¶**

*From the ‡Department of Microbiology, Immunology, and Molecular Genetics, and the Molecular Biology Institute, UCLA, Los Angeles, California 90095-1489, the ¶Department of Microbiology and Molecular Genetics, the **School of Information and Computer Science, the ‡‡Department of Biological Chemistry, and the ‖Institute for Genomics and Bioinformatics, University of California, Irvine, California 92697*

The work presented here is a first step toward a long term goal of systems biology, the complete elucidation of the gene regulatory networks of a living organism. To this end, we have employed DNA microarray technology to identify genes involved in the regulatory networks that facilitate the transition of *Escherichia coli* cells from an aerobic to an anaerobic growth state. We also report the identification of a subset of these genes that are regulated by a global regulatory protein for anaerobic metabolism, FNR. Analysis of these data demonstrated that the expression of over one-third of the genes expressed during growth under aerobic conditions are altered when *E. coli* cells transition to an anaerobic growth state, and that the expression of 712 (49%) of these genes are either directly or indirectly modulated by FNR. The results presented here also suggest interactions between the FNR and the leucine-responsive regulatory protein (Lrp) regulatory networks. Because computational methods to analyze and interpret high dimensional DNA microarray data are still at an early stage, and because basic issues of data analysis are still being sorted out, much of the emphasis of this work is directed toward the development of methods to identify differentially expressed genes with a high level of confidence. In particular, we describe an approach for identifying gene expression patterns (clusters) obtained from multiple perturbation experiments based on a subset of genes that exhibit high probability for differential expression values.

The enteric bacterium *Escherichia coli,* like many commensal and pathogenic microorganisms, thrives in the gastrointestinal tract of humans and other warm-blooded animals. In this environment, oxygen required for respiration and energy generation is in limited supply. Thus, the cell must derive energy from anaerobic respiration with alternative electron acceptors such as nitrate and fumarate or by fermentation of simple sugars. Metabolic transitions between aerobic and anaerobic growth states occur when *E. coli* cells enter an animal host and colonize the gastrointestinal tract, and when individual cells reposition themselves in new microenvironments inside the host. Each of these transitions is accompanied by fluctuations in oxygen tension. The cell responds to these fluctuations by modulating its central metabolic pathways for carbon and energy flow (1). Depending on the availability of oxygen, the cell can transition to the utilization of a variety of small carbon compounds as electron donors and/or acceptors for respiration (2). In addition, *E. coli* cells respond to these fluctuations in oxygen availability by altering the expression of a number of membrane-associated nutrient uptake or excretion systems, as well as a number of metabolic pathways such as those required for heme and quinone synthesis (1).

E. coli controls many of these systems in response to oxygen by altering gene expression levels. For example, expression of genes involved in oxygen utilization are switched off as oxygen is fully depleted from the environment. In a reciprocal fashion, expression of genes encoding alternative anaerobic electron transport pathways or genes needed for fermentation are switched on. Many of these metabolic transitions are controlled at the transcriptional level by the activities of a global regulatory protein, FNR,[1] and a two-component regulatory system ArcAB (3, 4). FNR is a CAP (catabolic activator protein) homologue that contains an oxygen labile iron-sulfur center as a sensor element for anaerobiosis (5, 6). Mutations in the *fnr* gene are known to affect the synthesis of nitrite, nitrate, and fumarate reductases (7), as well as fermentation pathway genes. Over 70 genes in 31 operons are currently recognized as members of the FNR gene regulatory network. The ArcAB (aerobic respiratory control) two-component regulatory system is composed of a classical OmpR-like receiver regulator, ArcA, and a membrane-associated sensor transmitter, ArcB (8). Examples of ArcA-regulated genes include genes for the Krebs cycle (*sdhCDAB, icd, fumA, mdh, gltA, acnA*, and *acnB*), for pyruvate metabolism and superoxide dismutase (*pfl* and *sodA*), and genes for the cytochrome *o* oxidase (*cyoABCDE*) and cytochrome *d* oxidase (*cydAB*) (1).

The purpose of this genome-based study is to identify additional genes differentially expressed in response to oxygen availability and to define further the network of genes controlled by the global regulatory protein, FNR. To identify the

* This work was supported in part by National Institutes of Health Grants GM49694, AI21678 (to R. G.), GM55073, and GM68903 (to G. W. H.) and by the University of California, Irvine, Institute of Genomics and Bioinformatics. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at *http://www.jbc.org*) contains the raw and processed data for the experimental results and may be downloaded in tabular format as Excel files.

§ Both authors contributed equally to this work.

§§ To whom correspondence (computation) may be addressed. Tel.: 949-824-5344; Fax: 949-824-8595; E-mail: gwhatfie@uci.edu.

¶¶ To whom correspondence (biology) may be addressed. Tel.: 310-206-8201; Fax: 310-206-5231; E-mail: robg@microbio.ucla.edu.

[1] The abbreviations used are: FNR, fumarate nitrate reduction regulatory protein; PPDE, posterior probability of differential expression; ORF, open reading frame; PCA, principal component analysis; Lrp, leucine-responsive regulatory protein.

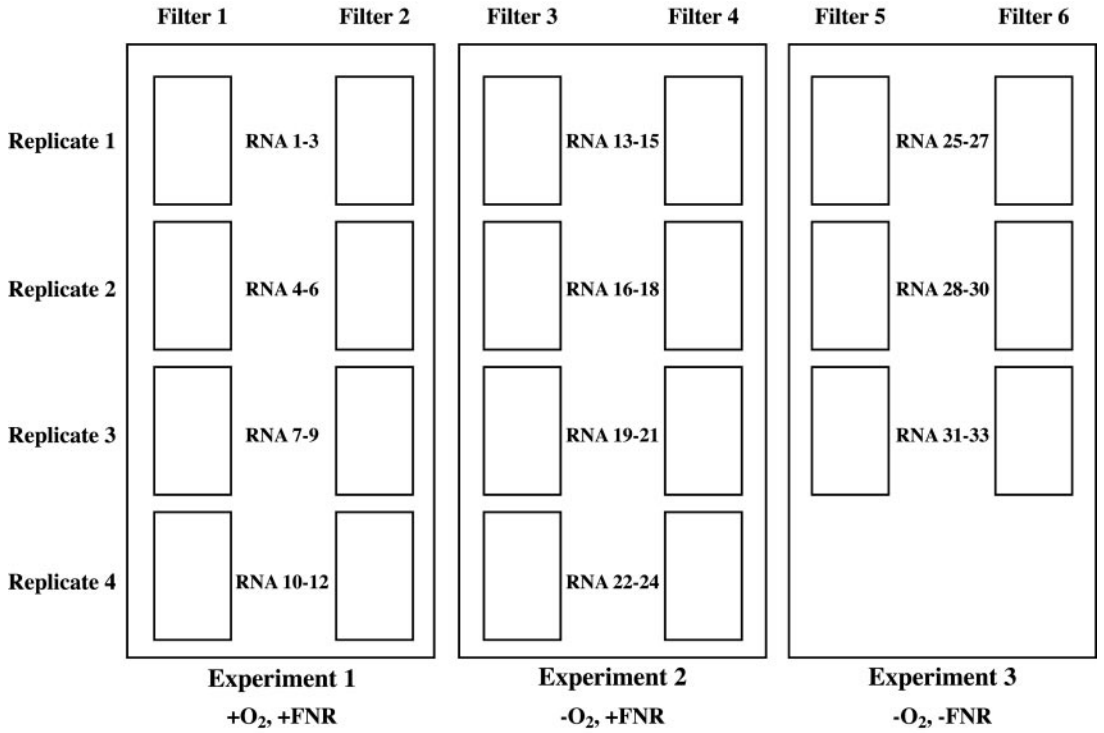

FIG. 1. **Experimental design.** See "Materials and Methods" for details.

global changes and adjustments of gene expression patterns that facilitate a change from aerobic to anaerobic growth conditions, we used DNA microarrays to analyze *E. coli* gene expression profiles of cells cultured at steady state growth rates under aerobic or anaerobic growth conditions ($+O_2$ or $-O_2$). To identify the genes controlled by FNR, we analyzed gene expression profiles of cells cultured under anaerobic growth conditions in the presence or absence of FNR ($-O_2$, +FNR or $-O_2$, −FNR) in otherwise isogenic strains.

MATERIALS AND METHODS

Chemicals and Reagents—Avian myeloblastosis virus (AMV)-reverse transcriptase and Sephadex G-25 Quickspin Columns were obtained from Roche Applied Science. Phenol and the DNA-*free* Kit were purchased from Ambion Inc. Ribonuclease Inhibitor III was purchased from Panvera/Takara. Ultrapure deoxynucleoside triphosphates were purchased from Amersham Biosciences. Random hexamer oligonucleotides and T4 polynucleotide kinase were obtained from New England Biolabs, and [α-^{33}P]dCTP (2–3000 Ci/mmol) was obtained from PerkinElmer Life Sciences. DNA filter arrays (Panorama *E. coli* Gene Arrays) were obtained from Sigma-Genosys Biotechnologies. SYBR Gold was purchased from Molecular Probes. All other chemicals were obtained from Sigma. All reagents and baked glassware used in RNA manipulations were treated with diethylpyrocarbonate.

Bacterial Strains and Growth Conditions—E. coli strains MC4100 (*F*$^-$ *araD139* Δ*(argF-lac)U169 rpsL150 relA1 flb-5301 deoC1 ptsF25 rbsR*) (9) and PC2 (MC4100 Δ*fnr-2*) (10) were used in this study. Aerobic cultures were grown in 125-ml Erlenmeyer flasks with constant aeration. Anaerobic cultures were grown in 15-ml anaerobic tubes fitted with butyl rubber stoppers (10). The medium was made anaerobic by flushing with O_2-free N_2 gas for 20 min and then dispensed anaerobically into N_2-flushed tubes. Cultures of the indicated strain were inoculated from overnight cultures grown under identical conditions (10).

Total RNA Isolation, cDNA Synthesis, and Target Labeling Conditions—Total RNA was isolated from 10-ml cultures; cDNA was synthesized and labeled with [α-^{33}P]dCTP, and filters were hybridized exactly as described by Hung *et al.* (11).

Stripping and reusing filters four times as described here results in a less than 3% increase in variance (12).[2]

Data Acquisition—A commercial software package obtained from Research Imaging Inc. (DNA ArrayVision) was used to grid the 16-bit image file, obtained from the PhosphorImager, to record the pixel density of each of the 18,432 addresses on each filter and to perform the background subtractions. 8,580 of the addresses on each filter are spotted with duplicate copies of each of the 4,290 *E. coli* ORFs. The remaining 9,852 empty addresses were used for background measurements. Because the backgrounds were quite constant, a global average background measurement was subtracted from each experimental measurement, although local background calculations are possible. Greater than 4 logs of linearity for the PhosphorImager derived data were observed.

Experimental Design—The experimental design for the experiments reported here is diagrammed in Fig. 1. In Experiment 1, filters 1 and 2 were hybridized with ^{33}P-labeled, random hexamer-generated cDNA fragments complementary to each of three RNA preparations (RNA 1–3) obtained from the cells of three individual cultures of the FNR$^+$ strain, MC4100, grown under aerobic conditions. These three ^{33}P-labeled cDNA target preparations were pooled prior to hybridization. Equal aliquots were hybridized to the duplicate *E. coli* Sigma-Genosys Panorama™ nylon filter arrays (Experiment 1, replicate 1, filters 1 and 2). Following PhosphorImager analysis, these filters were stripped and again hybridized with pooled ^{33}P-labeled cDNA target fragments complementary to each of another three independently prepared RNA preparations (RNA 4–6) from the same strain (MC4100; Experiment 1, replicate 2). This procedure was repeated two more times with filters 1 and 2 using two more independently prepared pools of cDNA targets (Experiment 1, replicates 3 and 4; RNA 7–9 and RNA 10–12). In Experiment 2, filters 3 and 4 were hybridized with ^{33}P-labeled, random hexamer-generated cDNA fragments complementary to each of three RNA preparations (RNA 13–15) obtained from the cells of three individual cultures of the FNR$^+$ strain MC4100 grown under anaerobic conditions. As for Experiment 1, these three ^{33}P-labeled cDNA target preparations were pooled prior to hybridization to the full-length ORF probes on the filters (Experiment 2, replicate 1, filters 3 and 4). Following PhosphorImager analysis, these filters were stripped and again hybridized with pooled, ^{33}P-labeled cDNA target fragments complementary to each of another three independently prepared RNA preparations (RNA 16–18) from the same strain (MC4100; Experiment 2, replicate 2). This procedure was repeated two more times with filters 3 and 4 using two more independently prepared pools of cDNA targets (Experiment 2, replicates 3 and 4; RNA 19–21 and RNA 22–24). In Experiment 3, filters 5 and 6 were hybridized with ^{33}P-labeled, random hexamer-generated, cDNA fragments complementary to each of three RNA

[2] S. P. Hung, G. W. Hatfield, S. Sundaresh, and P. Baldi, unpublished results.

preparations (RNA 25–27) obtained from the cells of three individual cultures of the FNR$^-$ strain PC2 grown under anaerobic conditions. These three ^{33}P-labeled cDNA target preparations were pooled prior to hybridization to the full-length ORF probes on the filters (Experiment 3, replicate 1, filters 5 and 6). Following PhosphorImager analysis, these filters were stripped and again hybridized with pooled, ^{33}P-labeled cDNA target fragments complementary to each of another three independently prepared RNA preparations (RNA 28–30) from the same strain (PC2; Experiment 3, replicate 2). This procedure was repeated one more time with filters 5 and 6 with another independently prepared pool of cDNA targets (Experiment 3, replicates 3; RNA 31–33). The data for the fourth replicate of this experiment was lost.

This experimental design produces duplicate filter data for four replicates performed with cDNA targets complementary to four independent sets of pooled RNA preparations for each experiment. Thus, because each filter contains duplicate spots for each ORF and duplicate filters were used for each experiment, a total of 16 measurements were obtained, 4 measurements for each ORF from each of 4 replicates for wild-type experiments, and 4 for each of the 3 replicates for the FNR$^-$ experiments (12 measurements).

Data Analysis—For each target signal, a background-subtracted estimate of expression level was obtained and scaled to total counts on the membrane by dividing each individual gene expression value by the total of all target signals on the membrane. Thus, each normalized gene level is expressed as a fraction of the total mRNA hybridized to each DNA array. For any given measurement, a value greater than zero (indicating an expression level) or a zero (indicating an expression level lower than background) is obtained. Only those genes exhibiting an expression level greater than zero in all replicates were used for statistical analysis. These gene expression level measurements were analyzed by a regularized *t* test based on a Bayesian statistical framework (11–15). For the analysis of the data reported here, we ranked the mean gene expression levels of the replicate experiments in ascending order, used a sliding window of 101 genes, and we assigned the average standard deviation of the 50 genes ranked below and above each gene as the Bayesian standard deviation for that gene. The *p* values for each gene measurement based on a regularized *t* test with a confidence value of 10 are reported in the Supplemental Material. A comprehensive discussion of the use of a regularized *t* test and the modifications applicable to the analysis of DNA microarray data of the type presented here are described in detail elsewhere (12).

Gene measurements containing zero expression values in one or more replicates were set aside. Among this set of genes, those with zero expression values for all replicates in one experiment, and all values greater than zero for all measurements of another experiment were identified. Because these gene measurements could not be analyzed with a *t* test, the significance of these results was evaluated by ranking these genes in ascending order according to their coefficients of variance of the four greater than zero measurements of each experiment(s).

To interpret the results of a high dimensional DNA array experiment, it is necessary to determine the global false-positive and -negative levels inherent in the data set being analyzed. We have implemented a mixture model-based method described by Allison *et al.* (16) for the computation of the global false-positive and -negative levels inherent in a DNA microarray experiment (11, 12). The basic idea is to consider the *p* values as a new data set and to build a probabilistic model for these new data. When control data sets are compared with one another (*i.e.* no differential gene expression), it is easy to see that the *p* values ought to have a uniform distribution between zero and one. In contrast, when data sets from different genotypes or treatment conditions are compared with one another, a non-uniform distribution will be observed in which *p* values will tend to cluster more closely to zero than one (Fig. 2), *i.e.* there will be a subset of differentially expressed genes with "significant" *p* values. The computational method of Allison (16) is used to model this mixture of uniform and non-uniform distributions to determine the probability, PPDE(*p*) ranging from 0 to 1, that any gene at any given *p* value is differentially expressed, *i.e.* it is a member of the uniform (not differentially expressed) or the non-uniform (differentially expressed) distribution. With this method, we can estimate the rates of false positives and false negatives as well as true positives and true negatives at any given *p* value threshold, PPDE(<*p*). In other words, we can obtain a posterior probability of differential expression PPDE(*p*) value for each gene measurement and a PPDE(<*p*) value at any given *p* value threshold based on the experiment-wide global false-positive level and the *p* value exhibited by that gene (11, 12). It should also be emphasized that this information allows us to infer the genome-wide number of genes that are differentially expressed, *i.e.* the fraction of genes in the non-uniform distribution

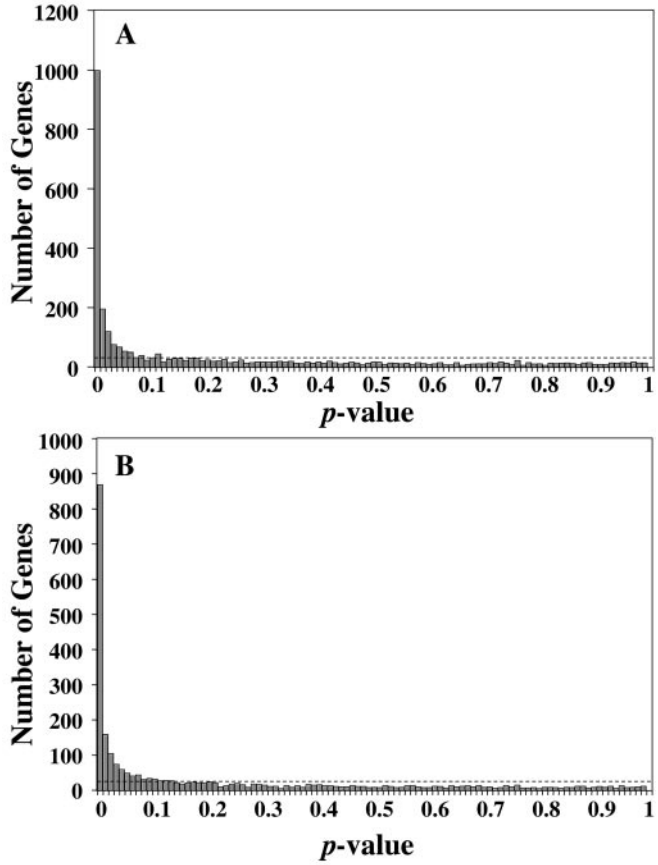

FIG. 2. ***p* value distributions.** The *p* values, based on a regularized *t* test distribution, of the 2,820 genes (*A*, $+O_2$, +FNR *versus* $-O_2$, +FNR) or the 2,402 genes (*B*, $-O_2$, +FNR *versus* $-O_2$, −FNR) expressed at a value above background in all replicate experiments grouped into 100 bins and plotted against the number of genes in each bin. The *dashed line* in each plot indicates the uniform distribution of *p* values under conditions of no differential expression.

(differentially expressed) and the fraction of genes in the uniform distribution (not differentially expressed). The PPDE(<*p*) and PPDE(*p*) values plotted against *p* values for the gene measurements of the $+O_2$ *versus* $-O_2$ and $-O_2$, +FNR *versus* $-O_2$, −FNR experiments are shown in Fig. 3. In Fig. 3*A* we see that at a *p* value less than 1, which includes all gene measurements, the PPDE(<*p*) is 0.63. This means that 63% of the 2,820 genes expressed above background in all of four replicate experiments are inferred to be differentially expressed between growth in the presence and absence of oxygen. In most instances, PPDE(<*p*) values are reported in the text and tables of this article. However, both PPDE(*p*) and PPDE(<*p*) values are given for each gene in the Supplemental Material.

The statistical methods described above are implemented in the Cyber-T software package available for on-line use at the website of the Institute for Genomics and Bioinformatics at the University of California, Irvine (www.igb.uci.edu). The clustering methods used to determine the regulatory patterns reported below are those implemented in the GeneSpring™ (Silicon Genetics, Redwood City, CA) software package.

RESULTS AND DISCUSSION

Differential Gene Expression in the Presence or Absence of Oxygen—In the following discussions we often simply refer to the fold change for differentially expressed genes. However, it is important to emphasize that reporting fold changes is incomplete and can be misleading (12). For this reason, the mean expression levels, standard deviations, *p* values, and PPDE(<*p*) values for all differentially expressed genes are included in the Supplemental Material. However, in the tables of this article we report only *p* values, PPDE (<*p*) values, and fold changes.

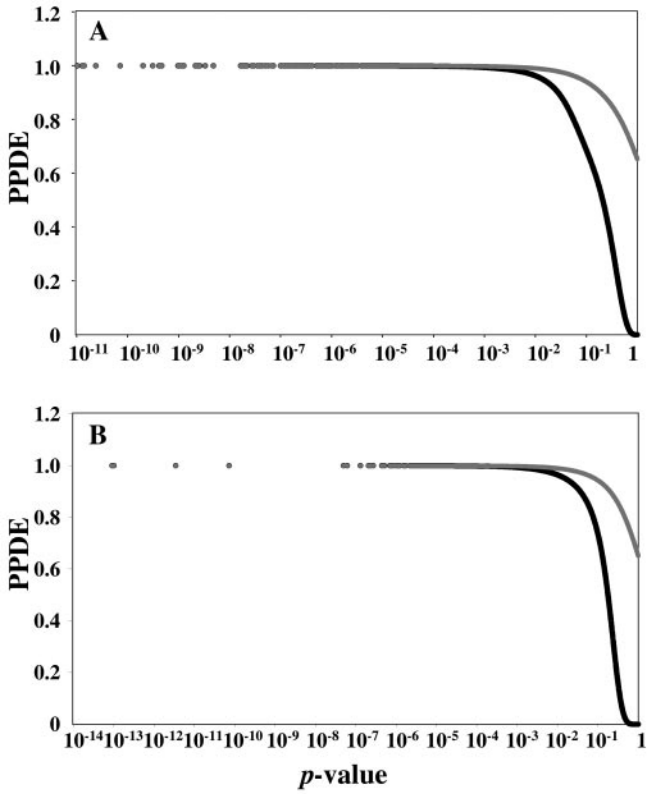

FIG. 3. **PPDE(*p*) or PPDE(<*p*) *versus* *p* values.** PPDE(*p*), *black dots*, is the posterior probability for differential expression for any gene of a given *p* value. PPDE(<*p*), *gray dots*, is the posterior probability for differential gene expression of the group of genes below any given *p* value. *A*, $+O_2$, +FNR *versus* $-O_2$, +FNR. *B*, $-O_2$, +FNR *versus* $-O_2$, −FNR.

A comparison of the gene expression levels between cells grown in the presence and absence of oxygen revealed 2,820 genes that exhibited expression levels above background for all replicates of Experiments 1 and 2 ($+O_2$, +FNR *versus* $-O_2$, +FNR; Fig. 1). If two data sets for which no differences are expected (*e.g.* $+O_2$ *versus* $+O_2$) were compared, then the *p* values would be equally distributed between 0 and 1.0 (*dashed lines* in Fig. 2). On the other hand, if differences among measurement levels of some genes are present (*e.g.* $+O_2$ *versus* $-O_2$; Fig. 2*A*), then the *p* values for those genes will be low and cluster toward 0. In Fig. 2*A,* the *p* values for all 2,820 gene measurements are distributed into 100 bins ranging from 0 to 1.0 and plotted against the number of genes in each bin. It is evident from an examination of the *p* value distribution in Fig. 2*A* that about one-half of the genes expressed during aerobic growth are modulated during the transition to anaerobic growth. Whereas the demarcation between differentially and non-differentially expressed genes is arbitrary, the data in Fig. 2*A* suggest that a lower threshold of $p = 0.05$, which corresponds to a PPDE(<*p*) value of 0.96 described under "Materials and Methods," is reasonable. Thus, of the 1,445 differentially expressed genes that exceed this threshold, 58 are expected to be false positives. Furthermore, it must be kept in mind that the remaining genes classified as not differentially expressed contain false negatives. The complete computational method to determine the fraction of differentially expressed genes and the fraction of falsely identified differentially expressed genes at any given PPDE(<*p*) value has been described by Hung *et al.* (11).

The *p* values and PPDE(<*p*) values, as well as additional statistical data, for all genes are contained in the Supplemental Material.

Differential Gene Expression in the Absence of Oxygen in the Presence and Absence of the FNR Global Regulatory Protein—A comparison of the gene expression levels between cells grown in the absence of oxygen and in the presence or absence of FNR revealed 2,402 genes that exhibited expression levels above background for all replicates of Experiments 2 and 3 ($-O_2$, +FNR *versus* $-O_2$, −FNR; Fig. 1). Again, about one-half of the gene expression levels are modulated by this treatment condition. An examination of the distribution of *p* values, shown in Fig. 2*B*, suggest that the expression levels of 1,256 genes with *p* values less than 0.05 are modulated, either directly or indirectly, by FNR during growth under anaerobic conditions (Fig. 2*B*). Again the PPDE(<*p*) value for this group of genes is 0.96; thus, 50 false positives are expected among this set of differentially expressed genes. The individual *p* values and PPDE values, as well as additional statistical data, for all genes are contained in the Supplemental Material.

Identification of Differential Gene Expression Patterns Resulting from Two-variable Perturbation Experiments—A basic paradigm for understanding regulatory networks at a system level involves performance of perturbation experiments. When only one parameter is perturbed, gene regulation patterns can be of only two types, they can go up or they can go down. However, when two or more parameters are perturbed, data mining methods designed to identify more complex gene expression patterns are needed. This is the case for the set of experiments described here where we examined the effects of perturbing two variables, one genetic variable and one environmental variable.

To identify the global changes and adjustments of gene expression patterns that facilitate a transition from aerobic to anaerobic growth conditions, and to determine the effects of genotype on these gene expression patterns, we analyzed *E. coli* gene expression profiles obtained from cells cultured under aerobic or anaerobic growth conditions ($+O_2$ or $-O_2$) and under anaerobic growth conditions in the presence or absence of the global regulatory protein for anaerobic metabolism, FNR ($-O_2$, +FNR or $-O_2$, −FNR). Because FNR is presumed to be inactive under aerobic ($+O_2$) conditions, we did not perform experiments comparing *fnr* genotypes under aerobic conditions. Only two general regulatory patterns can be observed when only two experimental conditions are compared, for example growth in the presence or absence of oxygen. However, when two conditions are compared, at least eight general regulatory patterns are expected. The data in Fig. 4 diagram the eight basic regulatory patterns that could be observed among three experiments conducted in the presence and absence of oxygen in an fnr^+ strain and in the absence of oxygen in an fnr^- strain. For simplicity, only three expression levels for each of these three experimental conditions are assumed: low, medium, and high.

An intuitive method to identify genes with these regulatory patterns could be to simply use any of several popular clustering methods on the entire data set. However, in experiments like the ones presented here where a limited number of replications or sample measurements are performed, resulting in many genes being measured with low confidence levels that result in false positives as well as false negatives, such a clustering approach could be misleading placing many genes in wrong clusters. To circumvent this problem, the approach described here is based on selecting those genes differentially expressed with high confidence levels for the initial clustering. Once the genes of these regulatory patterns are established, it is possible to "fish" for other genes with similar regulatory patterns with lower confidence levels that can be included at the discretion of the investigator.

To identify genes differentially expressed at a high confi-

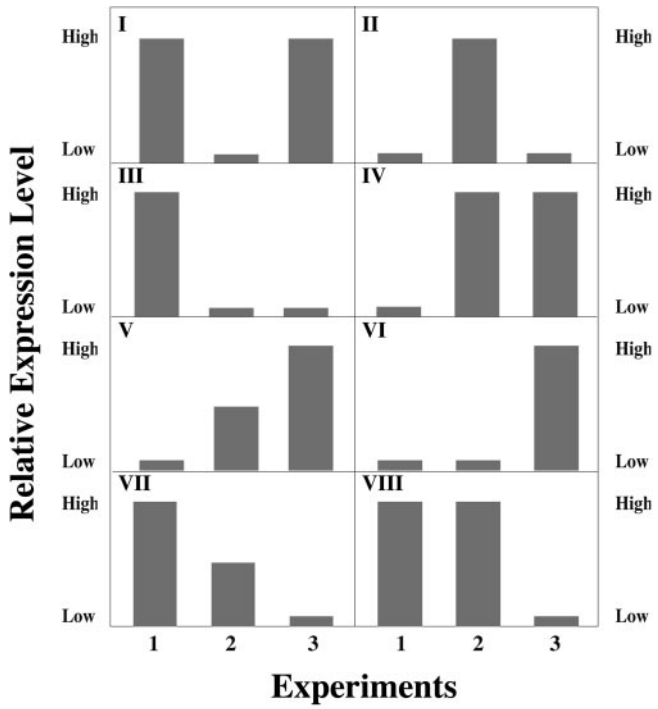

FIG. 4. **Gene expression regulatory patterns expected from the comparison of DNA array experiments with one control and two treatment conditions.** Control condition (*Experiment 1*), gene expression levels during growth under aerobic conditions in an *fnr*$^+$ *E. coli* strain; treatment condition 1 (*Experiment 2*), gene expression levels during growth under anaerobic conditions in an *fnr*$^+$ *E. coli* strain; treatment condition 2 (*Experiment 3*), gene expression levels during growth under anaerobic conditions in an *fnr*-deficient *E. coli* strain. Each regulatory pattern is designated by *roman numerals I–VIII*.

dence level that correspond to each of the patterns (I–VIII) diagrammed in Fig. 4, the genes differentially expressed due to the treatment condition of Experiments 1 and 2 were sorted in ascending order according to their p values based on the regularized t test as described under "Materials and Methods." Next, the genes differentially expressed due to the treatment condition of Experiments 2 and 3 were sorted in ascending order according to their p values. 100 genes with the lowest p values present in both lists were selected. These genes exhibited either an increased or decreased expression level between both treatment conditions (*i.e.* between Experiment 1 and 2 and Experiment 2 and 3; see Fig. 4).

To identify those genes differentially expressed with a high level of confidence under the treatment conditions of Experiments 1 and 2 but expressed at the same or similar levels under the treatment conditions of Experiments 2 and 3 (Patterns III and IV; see Fig. 4), the 500 genes of Experiments 1 and 2 with the lowest p values were compared with the 500 genes with the highest p values (genes unchanged with values closest to 1) from Experiments 2 and 3. This comparison identified 57 genes that were present in both lists. Likewise, to identify those genes differentially expressed under the treatment conditions of Experiments 2 and 3 but expressed at the same or similar levels under the treatment conditions of Experiments 1 and 2 (Patterns VI and VIII; see Fig. 4), the 500 genes of Experiments 2 and 3 with the lowest p values were compared with the 500 genes with the highest p values from Experiments 1 and 2. This comparison identified 48 genes that were present in both lists. These gene lists were combined into a single list of 205 genes differentially expressed under at least one treatment condition. All of the differentially expressed genes of this list exhibit p values less than 0.0013 and a global confidence based on the experiment-wide false-positive level of 99% (PPDE($<p$) = 0.99).

Hierarchical Clustering and Principal Component Analysis—In the previous section we described the identification of 205 genes differentially expressed with a high level of confidence and their supervised clustering into the eight regulatory groups of Fig. 4. These clusters can be used as a "gold standard" for the selection of clustering algorithms and parameters. Such a gold standard is important because all supervised clustering programs will produce clusters, but the members of each cluster will be a function of the clustering parameters chosen (12).

If we use the default settings of the popular GeneSpring software for hierarchical or k-means clustering of our selected gene set, we obtain many improperly grouped clusters. Therefore, we found it necessary to employ a trial-and-error approach to produce clusters of genes exhibiting the regulatory patterns of Fig. 4 (12). We empirically determined the parameters that result in the hierarchical clustering of our gold standard gene set. In other words, we used our gold standard gene set for supervised clustering.

As an independent test to corroborate the accuracy of our supervised hierarchical clustering method, we used an unsupervised principal component analysis (PCA) method to cluster and visualize the same set of 205 genes (17). The PCA clustering results shown in Fig. 5 show that this unsupervised method produces the same results as the supervised hierarchical clustering method. An advantage of the PCA method is that it visually highlights correlations among individual gene clusters and identifies relationships of outliers that are not as apparent with hierarchical trees.

Depending on the statistical threshold selected, some biologically significant genes will appear as false negatives, *i.e.* they will not meet the standards of an arbitrary statistical threshold. We have found, however, that it is possible to rescue some of these genes by selecting other genes with higher p values that have expression patterns similar to the genes of our gold standard. For example, five of the repressed genes of regulatory pattern I (Fig. 4) are previously known FNR regulated genes, and four of these have a documented FNR-binding site in their promoter region (Fig. 6). When we use this set of FNR-repressed genes and select additional genes with a similar regulatory pattern to four genes in the center of each PCA cluster (correlation coefficient = 0.97), we pick up an additional 137 oxygen- and FNR-regulated genes, 11 of which are documented in the literature (8) (*aceE, aceF, cydB, cyoB, icdA, mdh, sdhA, sdhB, sucA, sucB,* and *sucC*). These genes exhibit p values under one treatment condition or the other ranging from 1×10^{-11} to 0.01 and PPDE(p) values from 0.99 to 0.50. With this method the PPDE and p value limits of genes deemed worthy of further investigation can be set by the experimenter.

Interpretation of Clustering Results—For the experimental example described here, we can assess the efficacy of the clustering methods described above by correlating the known regulatory pattern of certain well studied genes with the patterns identified with the clustering algorithms. However, in other less well defined systems, such a standard is not available. Nevertheless, if the methods described here, based on a statistically reliable subset of genes, produce expected results, then it is reasonable to assume that this is a valid method that can be used as a general approach. For example, the genes of patterns I, V, and VI are negatively regulated by FNR; and those of pattern II, VII, and VIII are positively regulated by FNR (Figs. 4 and 5). Whereas some of these genes are expected to be affected only indirectly by the presence or absence of FNR, others whose expression is directly regulated by FNR should possess a DNA-binding site(s) upstream of their transcriptional start sites.

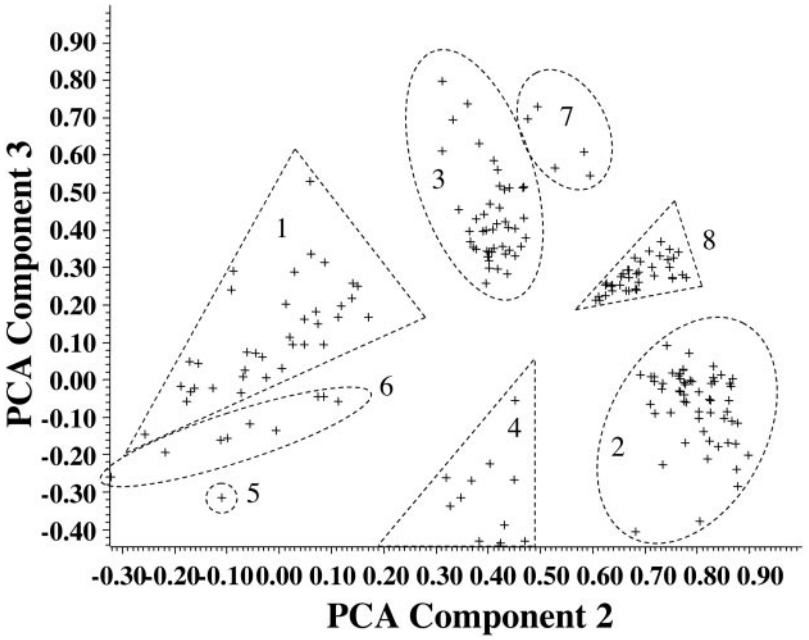

FIG. 5. **PCA clustering of differentially expressed gene regulatory patterns.** A two-dimensional projection onto a plane spanned by the second and third principal components. Each cluster is enclosed by a *dotted line.* The clusters are numbered according to the regulatory patterns of Fig. 4.

FNR protein is a homodimer that contains symmetrical helix-turn-helix structures that interact with a highly conserved, dyad consensus DNA sequence (5). This consensus sequence, obtained from mutational analyses and chromosome footprinting experiments performed with nearly a dozen FNR-binding sites, is TTGAT-N4-ATCAA. When FNR acts as an activator of gene expression, it most often binds to a site(s) upstream of and including a site centered at an average distance of 41.5 bp before the transcriptional start site of the affected gene or operon. When it acts as a repressor of gene expression, it binds to other sites often located near the transcriptional start site of the affected gene or operon (5).

Of the 46 genes down-regulated in the presence of FNR (*patterns I, V,* and *VI*, Figs. 4 and 5), 20 contain a documented or predicted FNR-binding site at or near the transcriptional start site with less than a 2-bp mismatch to the 10-bp FNR consensus sequence. Of the 63 genes up-regulated in the presence of FNR (*patterns II, VII*, and *VIII*, Figs. 4 and 5), 30 contain an upstream documented or predicted FNR-binding site. Because of the high statistical significance of the genes of these groups that do not exhibit upstream FNR-binding sites, we can be confident that these genes are members of the FNR regulatory network whose expression levels are indirectly affected by FNR. Without appropriate statistical analyses, it would not have been possible to distinguish these genes from false positives. Furthermore, because the expression levels of the 57 genes of patterns III and IV are not affected by the presence or absence of FNR, they are not expected to possess binding sites for this regulatory protein. This is the case. None of these genes possess a documented FNR-binding site or any site with less than a 2-bp mismatch to the 10-bp FNR consensus sequence. Thus, the statistical and clustering methods described here produce results consistent with biological expectations.

Functional Classes of Genes Affected by Oxygen Availability and FNR—Here we limit our discussion to the 205 genes of regulatory patterns I–VIII (Fig. 4). They represent many genes known to be oxygen-controlled and a larger set for which no previous information is available. These genes are listed in Tables I–VIII. Schematics of the genes in putative or known operons with putative or documented FNR-binding sites are shown in Figs. 6–11. Many of these genes are involved in aerobic or anaerobic electron transport processes: carbon flow through the Krebs cycle and/or fermentation; small molecule biosynthesis; macromolecular synthesis; and a variety of nutrient uptake or nutrient excretion reactions (Fig. 12). Regardless of their metabolic role, these genes are discussed below in the context of their expression patterns (Fig. 4).

Expression Pattern I: Decreased Expression during Anaerobiosis and Increased Expression in an FNR Strain—Among the 205 genes displayed in the clustering procedures described above, 36 are repressed in anaerobic conditions due to regulation by FNR (Table I). Of these 36 genes, 5 have been reported to be regulated directly by FNR (8), and 31 are newly discovered genes that are either directly or indirectly regulated by this global regulatory protein. The genes previously known to be regulated by FNR will be discussed first followed by a discussion of the newly discovered FNR-"repressed" genes.

The *cydA* gene (part of the *cydAB* operon) encodes the high affinity terminal oxidase of the oxygen respiratory chain, cytochrome *d* oxidase. The data obtained here show that *cydA* is repressed ~2-fold during anaerobic growth and is expressed at a 4.7-fold higher level in the FNR-deficient strain. In agreement with these findings, previous studies using a *cydA*::*lacZ* fusion and *in vitro* studies confirm that transcription of the *cydAB* operon is repressed when oxygen becomes limiting (18). Other studies have shown that ArcA functions to activate *cydAB* transcription as oxygen becomes limiting (19, 20), whereas FNR is required for repression when the oxygen tension is decreased further (18, 21, 22). This FNR-mediated repression of *cydA* is consistent with observations that FNR binds *in vitro* at two different sites, centered at positions −53.5 and −1.5, respectively, relative to the P1 transcriptional start (Fig. 6) (22). Three other promoters, P2, P3, and P4, are coordinately regulated with P1 in response to oxygen, ArcA, and FNR; however, a fifth promoter, P5, is not (Fig. 6) (23).

The *cyoA* and *cydA* genes encode cytochrome oxidases. The *cyoA* gene is the first member of the *cyoABCDE* operon that encodes all of the subunits of the cytochrome *o* ubiquinol oxidase. The *cyoA* gene is expressed 10-fold higher when cells are grown aerobically and 13-fold higher when grown anaerobically in the FNR-deficient strain (Table I). Previous studies by our laboratory using a *cyoA*::*lacZ* fusion in the same isogenic FNR^+ and FNR^- strains used in this study showed the same regulatory pattern (20). However, a search of the *cyo* promoter region does not reveal a site similar to the FNR consensus

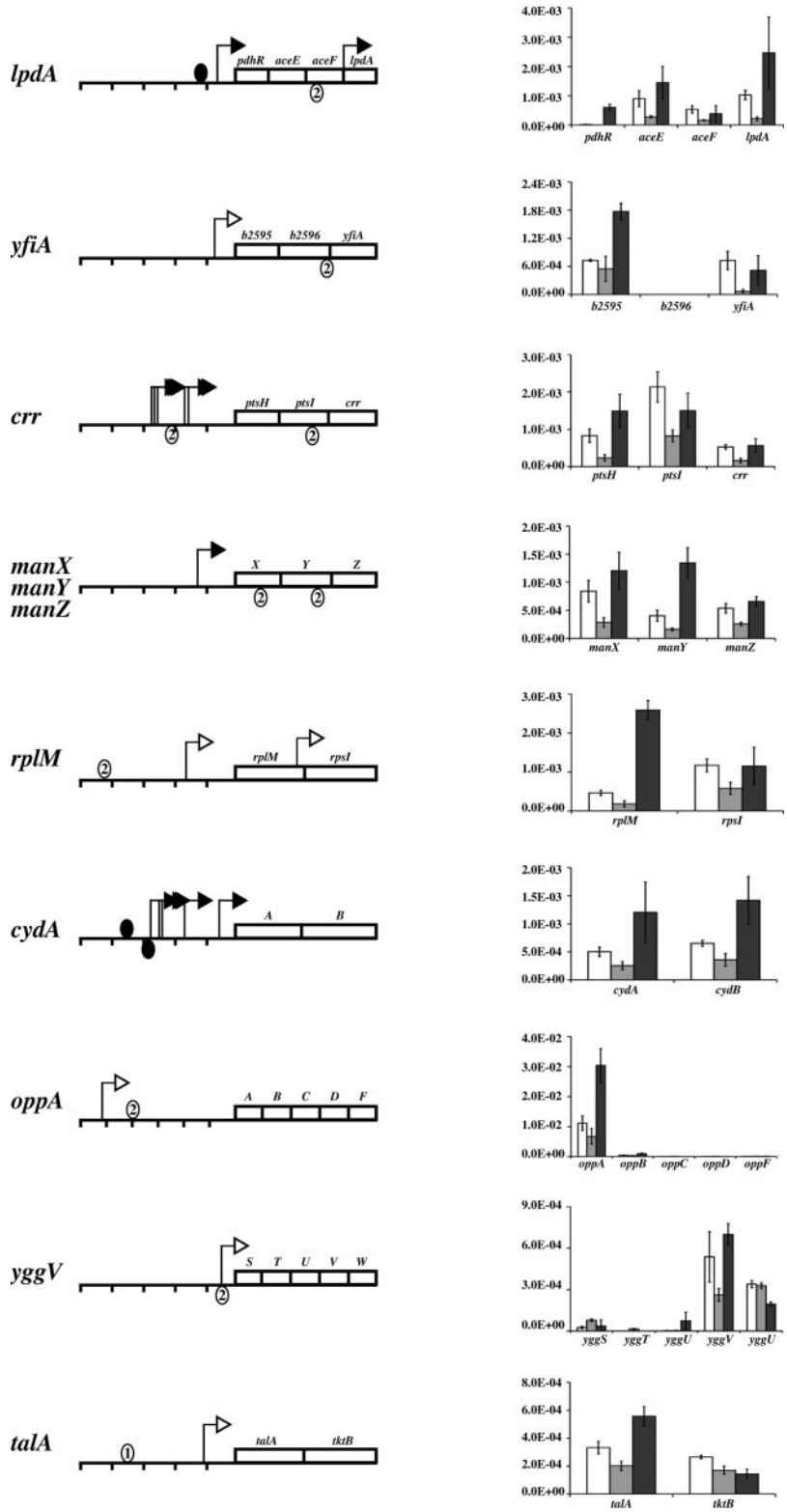

FIG. 6. **Genes of regulatory pattern I.** ORFs of genes (operons) are represented as *bars.* The 500 bp upstream of each ORF is represented by a *straight line*; *tick bars* are spaced at 100-bp intervals. *Open arrows* identify the position of predicted transcriptional start sites. *Black arrows* identify the position of documented transcriptional start sites. *Open ovals* identify the position of predicted FNR-binding sites. *Black ovals* identify the position of documented FNR-binding sites. The *number inside the open ovals* indicates the bp mismatch to the consensus FNR-binding site. The level of expression of each gene for Experiments 1–3 (Figs. 1 and 4) are indicated in the graphs by *white, light gray*, and *dark gray bars*, respectively. The operon organizations, the positions of the transcriptional start sites, and the documented FNR-binding sites were obtained either from GenBank™ (www.ncbi.nlm.nih.gov/Genbank) or RegulonDB (www.cifn.unam.mx/Computational_Biology/regulondb). The putative binding sites were identified using the search algorithm in the GeneSpring™ software package.

sequence. This suggests that FNR-mediated repression of this gene is indirect.

The gene for the non-proton translocating NADH dehydrogenase (NdhII), *ndh*, shows a 2.5-fold higher expression during aerobic *versus* anaerobic conditions, and a 5-fold increase under anaerobic conditions in the FNR-deficient strain (Table I). Previous studies (24–27) have demonstrated that this repression of *ndh* gene expression is due to the binding of FNR to two sites centered at bp positions −50.5 and −94.5 (Table I).

In addition to the above members of the FNR regulatory network, another gene, *nuoE*, has been described as negatively regulated by FNR (28). The *nuoA-N* operon encodes NADH dehydrogenase I (NdhI), a membrane-associated, multisubunit, proton-translocating enzyme similar to complex I of eukaryotic mitochondria (29). Expression of the *nuoE* gene (Fig. 6) is 3.8-fold lower under anaerobic conditions and elevated 4.5-fold in the FNR mutant (Table IX). Previous studies using a *nuo-lacZ* fusion established that *nuo* expression is also sub-

TABLE I

Regulatory pattern I: genes that exhibit decreased levels during anaerobic growth and increased levels in an fnr-deficient strain

Gene	*p* value ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	PPDE ($<p$) ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	*p* value ($-O_2$, +FNR *vs.* $-O_2$, −FNR)	PPDE ($<p$) ($-O_2$, +FNR *vs.* $-O_2$, −FNR)	Fold ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	Fold ($-O_2$, +FNR *vs.* $-O_2$, −FNR)
lpdA	1.29E-11	1.0000000	9.15E-05	0.9996050	−4.75	11.43
yfiA	7.16E-11	1.0000000	6.46E-05	0.9996903	−10.41	7.38
yceD	2.00E-10	1.0000000	2.14E-05	0.9998566	−3.90	4.10
trpB	3.12E-10	1.0000000	4.94E-04	0.9987177	−3.85	9.24
cyoA	9.75E-10	0.9999999	6.96E-07	0.9999869	−9.98	13.05
gpmA	1.27E-09	0.9999999	1.73E-04	0.9993858	−7.18	10.41
crr	2.14E-09	0.9999998	6.26E-05	0.9996970	−3.23	3.47
nuoE	2.34E-09	0.9999998	1.17E-04	0.9995302	−3.79	4.49
manX	2.74E-08	0.9999990	1.95E-04	0.9993304	−2.96	4.24
yadF	4.26E-08	0.9999986	6.78E-05	0.9996795	−2.28	2.60
tufA	1.59E-07	0.9999963	9.97E-14	1.0000000	−1.66	3.49
prlA	2.35E-07	0.9999952	2.05E-07	0.9999944	−2.12	3.63
rplM	3.40E-07	0.9999937	5.72E-06	0.9999430	−2.57	14.32
gatY	3.49E-07	0.9999936	1.92E-04	0.9993372	−2.74	8.00
trmD	5.65E-07	0.9999909	1.64E-04	0.9994076	−2.18	2.50
ndh	7.80E-07	0.9999886	8.73E-06	0.9999234	−2.50	5.06
manY	8.58E-07	0.9999878	2.49E-04	0.9992066	−2.51	8.35
manZ	9.60E-07	0.9999867	1.07E-04	0.9995607	−2.06	2.53
ompA	1.48E-06	0.9999819	6.57E-06	0.9999372	−2.27	3.41
rplT	4.41E-06	0.9999606	3.68E-05	0.9997910	−2.78	8.41
rpsJ	6.28E-06	0.9999492	6.23E-04	0.9984930	−2.10	4.27
speD	6.40E-06	0.9999485	2.35E-05	0.9998469	−2.30	4.62
speE	7.18E-06	0.9999442	3.52E-06	0.9999593	−2.10	4.79
cydA	1.14E-05	0.9999221	3.15E-04	0.9990636	−1.98	4.73
rpsT	1.90E-05	0.9998883	8.79E-05	0.9996159	−2.26	3.19
thlD	2.80E-05	0.9998525	3.94E-04	0.9989047	−1.66	2.19
rplS	3.02E-05	0.9998445	2.70E-07	0.9999932	−2.23	6.24
ptsG	3.30E-05	0.9998342	3.41E-04	0.9990114	−1.82	3.17
atpB	5.43E-05	0.9997637	1.74E-04	0.9993825	−1.66	3.82
rpsA	7.17E-05	0.9997120	1.19E-04	0.9995268	−1.91	2.67
yajG	7.73E-05	0.9996961	3.03E-05	0.9998173	−1.95	3.45
oppA	8.78E-05	0.9996672	9.06E-14	1.0000000	−1.66	4.53
yggV	9.69E-05	0.9996430	4.79E-05	0.9997486	−2.05	2.68
rpmB	1.25E-04	0.9995714	2.74E-04	0.9991518	−3.54	10.49
yajC	1.60E-04	0.9994903	4.62E-04	0.9987776	−2.01	3.02
talA	2.21E-04	0.9993581	7.62E-05	0.9996525	−1.64	2.76

ject to ArcA-mediated anaerobic repression and NarL nitrate-mediated anaerobic activation (28). In this same report, FNR and IHF were reported to have roles as weak repressors under anaerobic conditions, and it was suggested that the effect of FNR may be indirect as a consequence of its role in regulating ArcA expression (30). This is consistent with the fact that no FNR-binding site is observed in the promoter region of the *nuo* operon.

The *lpd* (*lpdA*) gene, encoding lipoamide dehydrogenase, a member of the *phdR-aceEF-lpdA* operon (Fig. 6), shows a 4.8-fold decrease in expression when grown under anaerobic conditions and a 11.4-fold increase in the FNR deletion strain. This operon encodes a pyruvate-responsive repressor (PhdR), pyruvate dehydrogenase (AceE), and lipoate acetyltransferase (AceF), components of the pyruvate dehydrogenase (PDH) complex, as well as Lpd, a component of both the PDH and 2-oxoglutarate (ODH) complexes (31). Transcription of *lpd* is coordinated from both the *pdhR* promoter as well as an internal *lpd* promoter (Fig. 6). Although neither the *pdhR* nor *lpd* promoters have been reported to be strongly controlled by FNR using *lacZ* fusions (31, 32), the results reported here are supported by the observation that the *pdh* promoter has a previously reported FNR site at bp position −49.5 (31), and a previously unidentified putative FNR site (2-mismatch) located upstream of the internal *lpd* promoter (Fig. 6).

Although the genes discussed above are previously known members of the FNR regulatory network important for anaerobic metabolism, no prior evidence for either oxygen- or FNR-mediated regulation of the remaining genes of this cluster has been reported. However, it is interesting that these newly discovered genes that are regulated by FNR under anaerobic conditions fall into to the same functional classes as the genes regulated by the leucine-responsive regulatory protein (Lrp) under aerobic conditions (11). These functional classes include genes for small molecule biosynthesis and transport and macromolecule biosynthesis. In fact, of the 31 remaining genes of this expression group regulated by FNR under anaerobic growth conditions, 10 are regulated by Lrp under aerobic conditions. A primary role for FNR is to coordinate carbon and energy metabolism during growth under anaerobic conditions. A primary role for Lrp is to coordinate metabolic activities with the nutritional and environmental growth state of the cell. Because many of the same genes are regulated by these two global regulatory proteins, these results suggest functional interactions between these two global regulatory networks.

The genes of this cluster that belong to the small molecule metabolism and transport groups include the three genes of the *manXYZ* operon (mannose phosphotransferase system): *crr* (phosphocarrier protein for glucose transport); and *ptsG* (glucose phosphotransferase enzyme II); *gpmA* (phosphoglyceromutase); *gatY* (D-tagatose-1,6-bisphosphate aldolase); *talA* (transaldolase A); *trpB* (tryptophan synthase); *speD* and *speE* (biosynthesis of spermidine); *thiD* (thiamin biosynthesis); and *prlA* (*secY*), a protein translocator of the *secYEG* operon. The *crr* gene is a member of the *ptsH-ptsI-crr* operon that contains an upstream putative FNR-binding site (2 mismatch) (Fig. 6). The *ompA* gene encodes an outer membrane protein.

The remaining genes of this expression group that belong to the macromolecule synthesis class are as follows: *rplM, rplT, rpsJ, rpsT, rplS, rpsA, rpmB* (ribosomal protein); *tufA* (elongation factor Tu); *atpB (uncB)* (ATP synthase); *trmD*, (tRNA methyltransferase); and *oppA* (oligopeptide permease). The

TABLE II
Regulatory pattern II: genes that exhibit increased levels during anaerobic growth and further decreased levels in an fnr-deficient strain

Gene	*p* value ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	PPDE(≤*p*) ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	*p* value ($-O_2$, +FNR *vs.* $-O_2$, −FNR)	PPDE(≤*p*) ($-O_2$, +FNR *vs.* $-O_2$, −FNR)	Fold ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	Fold ($-O_2$, +FNR *vs.* $-O_2$, −FNR)
fdhF	1.40E-11	1.0000000	1.84E-04	0.9993577	20.39	−2.28
iadA	4.48E-10	0.9999999	1.62E-06	0.9999764	23.04	−4.55
yhiV	1.18E-09	0.9999999	4.00E-05	0.9997782	3.91	−2.51
ydbA_2	9.98E-08	0.9999974	2.41E-05	0.9998444	2.33	−2.31
caiT	8.85E-07	0.9999875	2.28E-07	0.9999940	3.06	−6.62
yjiH	9.41E-07	0.9999869	2.96E-06	0.9999640	3.95	−7.44
b2866	5.29E-06	0.9999551	2.08E-05	0.9998596	2.10	−2.94
pyrD	5.36E-06	0.9999547	6.25E-06	0.9999394	4.95	−13.74
mobB	5.57E-06	0.9999534	1.77E-04	0.9993726	3.26	−3.07
yhjW	1.57E-05	0.9999023	3.52E-04	0.9989885	2.15	−2.23
b2878	2.10E-05	0.9998801	3.35E-05	0.9998038	3.79	−5.93
yjiZ	2.61E-05	0.9998598	4.22E-05	0.9997709	2.66	−4.25
yjeH	2.61E-05	0.9998595	8.83E-05	0.9996144	3.78	−4.90
recC	2.98E-05	0.9998457	6.38E-05	0.9996931	1.78	−2.29
tdh	3.20E-05	0.9998378	1.06E-05	0.9999120	1.92	−3.01
araB	3.80E-05	0.9998167	1.21E-04	0.9995131	2.55	−3.59
yhjN	3.80E-05	0.9998166	2.19E-04	0.9992745	3.44	−4.18
narY	3.92E-05	0.9998124	5.30E-04	0.9986469	2.39	−2.64
glnE	5.30E-05	0.9997677	7.84E-05	0.9996455	1.80	−2.34
yeiA	5.98E-05	0.9997469	1.70E-04	0.9994013	4.22	−4.15
ybeS	6.67E-05	0.9997265	2.70E-04	0.9991591	8.48	−10.12
nanT	7.22E-05	0.9997104	1.13E-05	0.9999085	1.87	−3.04
acrF	7.45E-05	0.9997039	1.53E-06	0.9999773	2.25	−6.83
uraA	7.91E-05	0.9996910	2.11E-04	0.9992913	1.97	−2.48
yjdH	8.23E-05	0.9996823	1.51E-04	0.9994398	2.30	−3.13
yicM	8.54E-05	0.9996737	1.38E-05	0.9998947	1.88	−2.95
xylE	8.75E-05	0.9996680	3.07E-04	0.9990818	3.08	−3.85
glnD	8.77E-05	0.9996675	1.09E-05	0.9999105	2.55	−6.32
yjcS	9.51E-05	0.9996478	3.05E-04	0.9990846	1.72	−2.09
yggF	9.79E-05	0.9996405	3.54E-05	0.9997973	2.38	−4.29
pstS	1.17E-04	0.9995927	4.63E-05	0.9997545	2.40	−4.35
yidK	1.29E-04	0.9995616	8.97E-06	0.9999219	2.13	−4.86
yjiE	1.32E-04	0.9995565	8.13E-07	0.9999854	1.53	−2.77
tra5_2	1.38E-04	0.9995403	1.64E-04	0.9994073	1.90	−2.51
metL	1.45E-04	0.9995253	3.16E-06	0.9999623	2.07	−5.65
yhcL	1.51E-04	0.9995109	4.25E-05	0.9997694	3.18	−7.80
yidE	1.72E-04	0.9994636	2.69E-05	0.9998320	1.82	−3.16
mhpF	1.72E-04	0.9994624	2.77E-05	0.9998287	2.78	−5.92
tynA	1.73E-04	0.9994605	3.06E-05	0.9998163	1.82	−2.78
glgA	1.74E-04	0.9994591	3.51E-04	0.9989903	1.85	−2.33
b3051	1.83E-04	0.9994389	1.14E-05	0.9999069	2.07	−4.82
yhfS	2.06E-04	0.9993890	1.06E-04	0.9995625	2.92	−5.50
sbp	2.22E-04	0.9993566	6.05E-04	0.9985233	5.80	−8.42
sbm	2.39E-04	0.9993217	3.51E-06	0.9999594	2.20	−7.63
mrcA	2.50E-04	0.9993008	5.92E-05	0.9997084	2.00	−3.42
yheF	2.72E-04	0.9992563	7.99E-05	0.9996408	2.01	−3.13
yadQ	2.80E-04	0.9992420	9.92E-05	0.9995821	1.99	−3.13
speC	2.86E-04	0.9992303	2.49E-06	0.9999681	1.68	−3.54
yhjJ	2.90E-04	0.9992217	3.92E-04	0.9989098	2.08	−2.92
fliP	2.97E-04	0.9992091	6.45E-04	0.9984574	2.69	−3.58
yabM	3.01E-04	0.9992020	1.25E-05	0.9999021	2.30	−6.97
yhhT	3.06E-04	0.9991925	1.86E-05	0.9998701	2.14	−4.56
dinG	3.20E-04	0.9991662	4.66E-05	0.9997532	1.83	−3.14
yjeM	3.30E-04	0.9991485	1.49E-04	0.9994424	2.20	−3.66
yhhJ	3.34E-04	0.9991405	4.74E-04	0.9987544	2.26	−3.04
proW	3.36E-04	0.9991367	3.73E-06	0.9999577	1.92	−4.97
rarD	3.61E-04	0.9990914	7.62E-05	0.9996524	1.99	−3.35
sbcC	3.89E-04	0.9990428	5.66E-04	0.9985907	2.17	−3.06

rplM gene is predicted to be part of a two-gene operon and has a putative FNR-binding site (2 mismatch) upstream of the putative promoter for this operon (Fig. 6).

The functions of the remaining six genes in this list, *yfiA, yceD, yadF, yajG, yggV,* and *yajC,* remain to be characterized. Two of these genes, *yfiA* and *yggV,* are members of predicted operons, which have putative FNR sites (2 mismatch) in their upstream regions (Fig. 6).

Expression Pattern II: Increased Expression during Anaerobiosis and Decreased Expression in an FNR Strain—The transcription of the 58 genes of expression pattern II (Table II) are both induced in the absence of oxygen and positively regulated by FNR. Among the 23 previously reported FNR-activated genes, none are present in this expression group; however, 10 genes in Table II possess a putative FNR-binding site (Fig. 7). In agreement with the suggestion that there is a significant interaction between the Lrp and FNR regulatory networks, 23 of the 58 genes of this group are regulated by Lrp under aerobic growth conditions (11). Furthermore, the 58 genes of this group again can be grouped into the functional categories of small molecule metabolism and transport and macromolecule biosynthesis characteristic of Lrp-regulated genes.

Among the genes of the small molecule metabolism group, *fdhF, tdh, pyrD, metL,* and *mhpF* are dehydrogenases. The dehydrogenases can be divided into the following two groups: those involved in the supply of reducing potential for energy

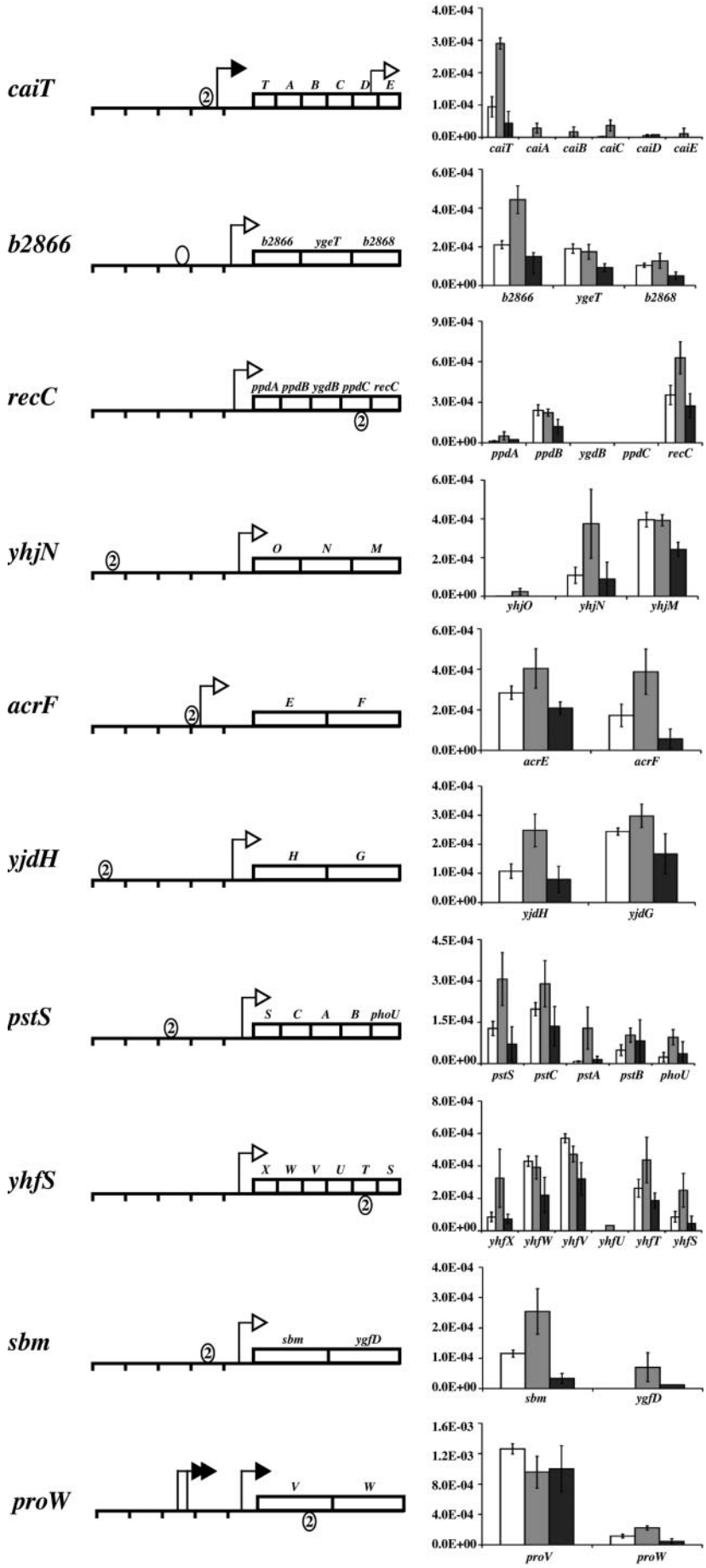

FIG. 7. **Genes of regulatory pattern II.** See legend for Fig. 6.

generation (*fdhF, tdh,* and *mhpF*, and those involved in anabolic processes required for the supply of metabolic intermediates (*pyrD* and *metL*). The *fdhF* gene encodes formate dehydrogenase-H (part of the formate dehydrogenase-H complex). Previous studies using an *fdhF*::*lacZ* gene fusion confirm that its transcription is activated anaerobically in the presence of FNR (33). The *tdh* gene encodes threonine dehydrogenase, involved in the conversion of threonine to glycine. This gene is a known member of the Lrp regulon (34). The *mhpF* gene is reported to encode acetaldehyde dehydrogenase, part of the gene cluster encoding the 3-(3-hydroxyphenyl)propionate catabolic pathway (35). These pathways might provide electrons for anaerobic respiration. The *mobB* gene product is involved in molybdenum cofactor biosynthesis. The *narY* gene encodes a cryptic nitrate reductase subunit, and *tynA* encodes a tyramine oxidase.

Among the dehydrogenases involved in anabolic processes, *pyrD* encodes the pyrimidine biosynthetic enzyme dihydro-orotate dehydrogenase, and *metL* encodes the aspartokinase II-homoserine dehydrogenase, a member of the *metBL* operon. Other genes of this group include two transferases: *glnE* encoding glutamine synthetase/adenyltransferase and *glnD,* encoding a uridylyl transferase; *speC*, ornithine decarboxylase; and *sbp,* a periplasmic sulfate-binding protein.

Nine genes of this expression pattern belong to the small molecule transport functional class: *caiT, nanT, uraA, xylE, proW, pstS, yadQ, yabM,* and *yhcL.* The *caiT* gene encodes a putative carnitine/betaine transport protein involved in osmoprotection, reported to be transcribed during anaerobic growth (36). Recent results from our laboratory employing a *caiT*::*lacZ* fusion confirm the DNA microarray data reported here.[3] The *proW* gene encodes the high affinity transport protein for glycine, betaine, and proline, whereas *pstS* encodes a high affinity phosphate-specific transporter. The *nanT* gene encodes a sialic acid transporter; the *uraA* gene encodes an inner membrane protein involved in the transport of uracil in the cell, and *xylE* encodes a xylose-proton symporter. Finally, the three additional genes encode a homologue of a mammalian chloride channel protein (*yadQ*), a sugar efflux transporter that can transport both glucose and lactose (*yabM*), and *yhcL* (*dcuD*), which shares significant homology to *dcuC*, a C4-dicarboxylate carrier that functions during anaerobic growth. The function of the YhcL (DcuD) protein is currently unclear (37).

The genes of this expression pattern belonging to the macromolecular synthesis class include genes for DNA repair. These genes are as follows: *recC*, a subunit of the RecBCD enzyme complex; *dinG,* encoding a LexA-regulated DNA repair enzyme; and *sbcC*, a cosuppressor of *recBC* mutations. The *acrF* gene encodes a lipoprotein, and the *glgA* gene is required for glycogen synthesis. The *fliP* gene is a gene required for flagella synthesis, and *iadA* encodes an isoaspartyl dipeptidase.

Rounding out the genes of this expression group are the following: *araB* and *araA*, encoding the ribulokinase and L-arabinose isomerase, of the *araBAD* operon; *mrcA*, encoding penicillin-binding protein 1A; *yheF (gspD*), a member of the general secretory pathway; and *rarD,* a gene involved in chloramphenicol resistance. The 26 remaining members of this expression pattern are currently uncharacterized.

Expression Pattern III: Decreased Expression during Anaerobiosis, No Change in an FNR Strain—Genes with reduced anaerobic expression but unaffected by FNR were clustered into expression pattern III (Table III). With one exception, the 45 genes of this cluster have not been studied previously for their expression under anaerobic growth conditions; however, 14 contain putative FNR-binding sites (Fig. 8). Again, the genes of this cluster are members of the same functional classes of expression patterns I and II. Nine genes (*ftsN, rfbX, wbbJ, fabG, fdoH, goaG, katE, pheA,* and *ilvH*) are involved in small molecule metabolism, and five (*lamB, malM, yabK(thiP), dppA,* and *acpP*) are involved in small molecule transport. Thirteen genes (*rplB, rplO, rplC, rpsE, rplD, rplQ, rplI, rpsH, rplE, nfi, hflC, hupB,* and *tig*) are involved in macromolecule synthesis or degradation. In this expression pattern, only five genes are known to be regulated by Lrp (*lamB, malM, yabI, pheA,* and *hupB*); however, 17 genes of this cluster are of unclassified function.

The single gene of this cluster previously known to be regulated by oxygen and not affected by FNR is formate dehydrogenase-O, the product of the *fdoH* gene. This gene is expressed 2.3-fold higher during growth under aerobic than under anaerobic conditions (Table III). Previous transcriptional fusion studies have reported a 3-fold increase in the expression level of this gene during aerobic growth, and this decreased expression during anaerobiosis is independent of FNR, as well as ArcA (38).

Expression Pattern IV: Increased Expression during Anaerobiosis, No Change in an FNR Strain—The 12 genes of this cluster (Table IV) show elevated expression under anaerobic growth conditions but are not affected by the deletion of the FNR allele; none of these genes contain documented or putative FNR-binding sites near their promoter regions. Of these genes, *poxB* and *ubiC* have been reported to be affected by oxygen availability. The *poxB* gene product, pyruvate oxidase, is growth phase-dependent exhibiting maximal expression during early stationary phase. It has also been reported that *poxB* expression is repressed during anaerobiosis in a manner that is not mediated by ArcAB or FNR (39). The *ubiC* gene product, chorismate lyase involved in the first two steps of ubiquinone biosynthesis, is induced during growth under anaerobic conditions (40). No FNR-binding site is evident in the promoter regulatory region of either of these genes. The *appBC* operon encodes subunits of a second putative cytochrome oxidase similar to CydAB.

Previous studies (41) have confirmed that the expression of the *gadA* and *gadB* genes, two highly homologous glutamate decarboxylases, are elevated under anaerobic conditions 23- and 24-fold, respectively. Recent *lacZ* transcriptional fusion experiments in our laboratory confirm these results and further demonstrate that these genes are regulated by ArcA but, as reported here, not by FNR.[3] The remaining members of this cluster include *cyaA*, the adenylate cyclase gene, and *mrr* involved in the restriction of methylated adenine residues. In addition, four genes of unknown function cluster to this group: *hdeA, hedB*, *ybeD,* and *ygjD.*

Expression Pattern V: Increased Expression during Anaerobiosis, Increased Expression in an FNR Strain—This cluster contains only a single gene of unknown function, *ybjX* (Table V).

Expression Pattern VI: No Change during Anaerobiosis, Increased Expression in an FNR Strain—This cluster contains nine genes (Table VI). Two, *purN* and *gapC-1*, contain putative FNR-binding sites (Fig. 9), and two are of unknown function (*ydcF* and *yhbJ*). Seven genes of this cluster are involved in small molecule metabolism and transport: *gapA* and *gapC-1,* structural genes for glyceraldehyde-3-phosphate dehydrogenase A and C, essential for glycolysis, and *purN,* the 5′-phosphoribosylglycinamide transformylase I structural gene. The *nrdB* gene of the *nrdAB* operon is the structural gene for ribonucleoside di-phosphate reductase subunit B2, an aerobic nucleotide

[3] R. P. Gunsalus, unpublished results.

TABLE III
Regulatory pattern III: genes that exhibit decreased levels during anaerobic growth that are unaffected in an fnr-deficient strain

Gene	*p* value ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	PPDE(≤*p*) ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	*p* value ($-O_2$, +FNR *vs.* $-O_2$, −FNR)	PPDE(≤*p*) ($-O_2$, +FNR *vs.* $-O_2$, −FNR)	Fold ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	Fold ($-O_2$, +FNR *vs.* $-O_2$, −FNR)
ylcB	4.14E-10	0.9999999	8.83E-01	0.6790325	−8.17	−1.05
b1200	9.88E-10	0.9999999	5.56E-01	0.7704788	−4.81	−1.23
ycgC	2.10E-08	0.9999991	6.63E-01	0.7381401	−2.68	−1.11
ftsN	2.94E-07	0.9999943	6.09E-01	0.7542667	−2.48	−1.11
rplB	3.15E-07	0.9999940	7.77E-01	0.7062095	−2.25	−1.06
rplO	6.12E-07	0.9999904	4.45E-01	0.8075183	−2.15	−1.15
dppA	1.83E-06	0.9999790	4.80E-01	0.7953656	−1.84	1.13
rplC	2.61E-06	0.9999729	6.23E-01	0.7499961	−2.15	−1.10
ymfN	6.49E-06	0.9999480	9.59E-01	0.6609094	−1.98	−1.01
goaG	7.84E-06	0.9999405	8.74E-01	0.6813837	−1.84	1.04
hflC	9.54E-06	0.9999316	8.14E-01	0.6966782	−2.10	1.06
ymfM	9.86E-06	0.9999300	8.40E-01	0.6899644	−1.74	−1.04
lamB	1.06E-05	0.9999264	5.66E-01	0.7674463	−4.54	1.18
ilvH	1.15E-05	0.9999220	6.28E-01	0.7483795	−1.62	1.08
rpsE	1.53E-05	0.9999044	4.78E-01	0.7962193	−2.02	1.19
ygiC	2.44E-05	0.9998664	9.32E-01	0.6673165	−1.52	1.01
yfhG	2.71E-05	0.9998560	4.64E-01	0.8009367	−1.55	1.13
ykgG	5.68E-05	0.9997560	5.28E-01	0.7799862	−3.72	1.46
ykgI	6.29E-05	0.9997376	5.62E-01	0.7688640	−1.63	1.13
katE	6.37E-05	0.9997351	5.98E-01	0.7575827	−1.60	1.10
rplD	6.91E-05	0.9997194	4.13E-01	0.8186765	−1.82	1.25
tig	1.15E-04	0.9995974	8.87E-01	0.6780300	−1.60	1.03
yliG	1.33E-04	0.9995540	9.30E-01	0.6678157	−1.54	−1.02
rplQ	1.40E-04	0.9995370	4.13E-01	0.8188175	−2.05	−1.31
ybfH	1.44E-04	0.9995263	5.23E-01	0.7814155	−1.58	1.13
wbbJ	1.45E-04	0.9995253	4.02E-01	0.8224808	−2.04	−1.30
ytfL	1.46E-04	0.9995223	7.70E-01	0.7081760	−1.46	−1.05
ychG	1.53E-04	0.9995067	5.71E-01	0.7658627	−1.52	−1.11
fabG	1.60E-04	0.9994904	9.07E-01	0.6731365	−1.80	−1.03
malM	1.90E-04	0.9994243	9.13E-01	0.6718650	−3.20	−1.03
rplI	2.21E-04	0.9993587	6.49E-01	0.7421794	−2.02	1.11
rpsH	2.22E-04	0.9993558	7.20E-01	0.7217522	−1.80	1.07
ycdT	2.67E-04	0.9992674	5.05E-01	0.7871838	−1.56	−1.16
yabI	2.80E-04	0.9992415	6.57E-01	0.7397388	−1.44	1.07
pheA	3.09E-04	0.9991868	4.17E-01	0.8173707	−1.94	1.26
b1808	3.15E-04	0.9991757	4.65E-01	0.8006586	−1.48	1.13
nfi	3.46E-04	0.9991185	6.32E-01	0.7473926	−1.62	1.12
yojH	3.48E-04	0.9991162	4.05E-01	0.8216075	−1.37	1.15
fdoH	3.98E-04	0.9990271	4.25E-01	0.8145795	−2.29	1.34
acpP	4.73E-04	0.9989008	4.46E-01	0.8071028	−1.71	1.21
yfiB	5.20E-04	0.9988240	7.16E-01	0.7231582	−1.69	−1.09
rplE	5.65E-04	0.9987536	4.93E-01	0.7913104	−1.74	−1.19
hupB	5.74E-04	0.9987394	5.76E-01	0.7644347	−1.85	1.15
rfbX	6.71E-04	0.9985912	4.32E-01	0.8120677	−1.74	1.23
yabK	7.69E-04	0.9984493	9.20E-01	0.6700117	−1.37	1.02

reductase. Previous studies (42) demonstrate that *nrd* expression is decreased during fermentative or NO_3 respiring conditions. The *artI* gene of the *artPIQMJ* operon is involved in arginine transport. The *potF* gene is a member of the *potFGHI* operon involved in the transport of putrescine. Both *artI* and *potF* are regulated by Lrp under aerobic growth conditions (11). The *his* gene of the *hisTJQMP* operon encodes a histidine-binding protein that is part of the periplasmic permeases for the high affinity uptake of histidine.

Expression Pattern VII: Decreased Expression during Anaerobiosis, Decreased Expression in an FNR Strain—Two of the five genes in this cluster are unknown (*ybdE* and *ylcD*), and two contain documented FNR-binding sites (8) (Table VII and Fig. 10). One gene encodes ribosomal protein L29, *rpmC*, and the remaining two genes, *frdA* and *nirB*, encode oxidoreductases involved in fumarate and nitrate reduction during anaerobiosis. Previous studies (43) from our laboratory using a *frdA*::*lacZ* fusion have demonstrated a 10-fold increase during anaerobic *versus* aerobic cell growth, and have demonstrated that FNR is responsible for this anaerobic activation. However, the data in Table VII show a 2-fold reduction in the expression of this gene during anaerobiosis. This discrepancy is likely due to the fact that the *frdA* and *sdhA* genes are paralogs that show a high degree of sequence identity. Because these genes are regulated in a reciprocal manner, the array results probably reflect hybridization of both of these targets to the *frdA* probe. It should be noted that systematic experimental errors of these types elude statistical screening. In this specific case, the only way to remove this source of error is to redesign the array probes to avoid cross-hybridization with multiple targets.

The *nirB*-encoding nitrite reductase is a soluble siroheme-containing enzyme that uses NADH as an electron donor to reduce nitrite in the cytoplasm. Although the array data presented in Table VII indicate that *nirB* expression is decreased during anaerobic growth, a *lacZ* transcriptional fusion showed that expression of the *nirBDC* operon is activated by FNR during anaerobic cell growth (44, 45). Therefore, the *nirB* gene, like the *frdA* gene, should have clustered to Expression Pattern I. Again, this discrepancy might be due to the high similarity of a region of the *nirB* gene to other nitrate and sulfate reductase subunits encoded by the *nirD, cysI,* and *cysJ* genes.

Expression Pattern VIII: No Change during Anaerobiosis, Decreased Expression in an FNR Strain—This cluster contains 33 genes of which 15 are of unknown function. Of the 18 genes of known function (Table VIII), 4 are known to be regulated by

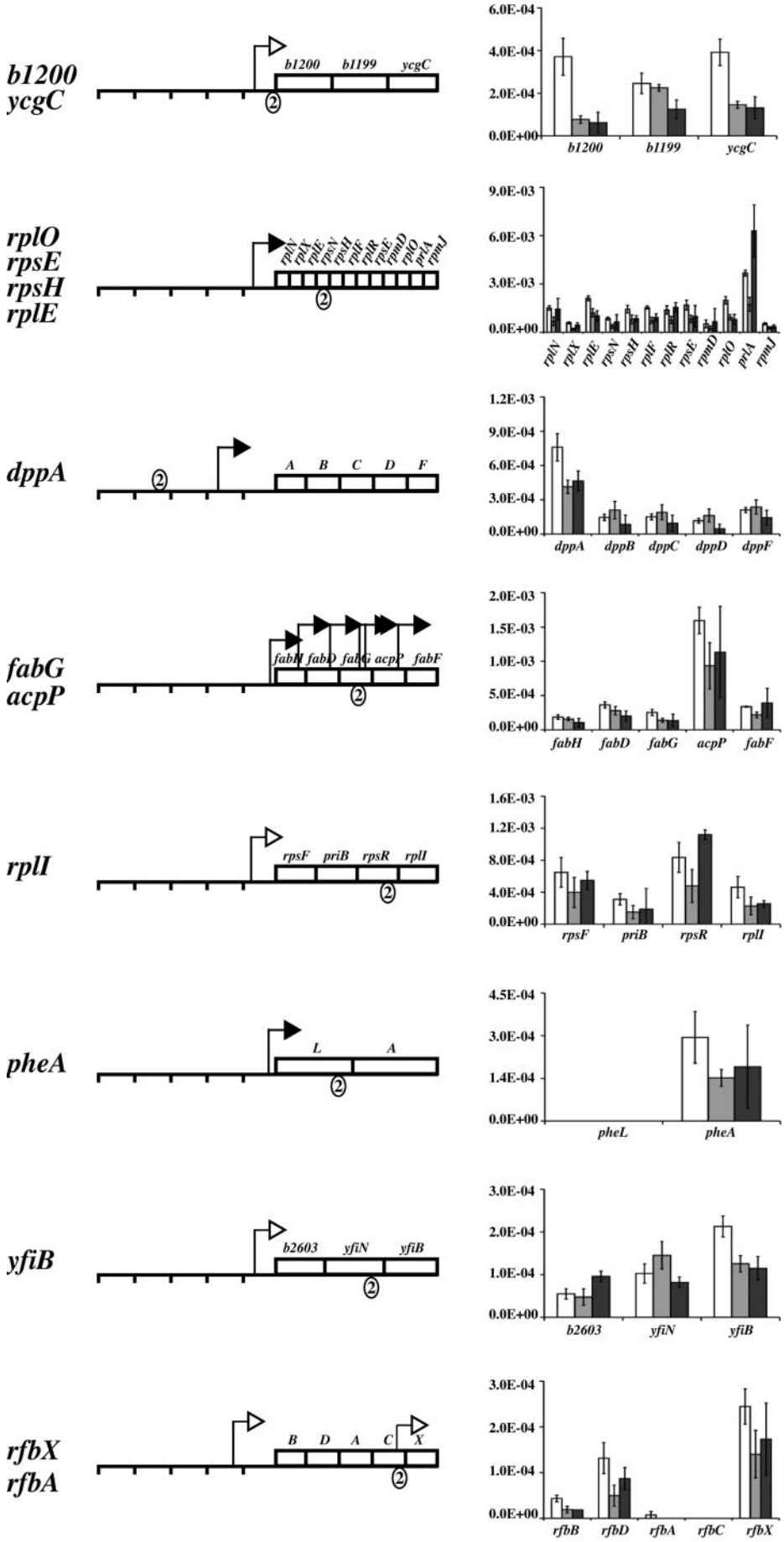

FIG. 8. **Genes of regulatory pattern III.** See legend for Fig. 6.

TABLE IV

Regulatory pattern IV: genes that exhibit increased levels during anaerobic growth that are unaffected in an fnr-deficient strain

Gene	*p* value ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	PPDE(<*p*) ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	*p* value ($-O_2$, +FNR *vs.* $-O_2$, $-$FNR)	PPDE(<*p*) ($-O_2$, +FNR *vs.* $-O_2$, $-$FNR)	Fold ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	Fold ($-O_2$, +FNR *vs.* $-O_2$, $-$FNR)
appB	3.31E-09	0.9999998	7.05E-01	0.7260438	9.26	−1.10
hdeA	4.82E-09	0.9999997	6.47E-01	0.7429059	40.69	−1.10
appC	4.83E-09	0.9999997	6.89E-01	0.7306487	3.50	−1.08
gadB	1.87E-08	0.9999992	5.19E-01	0.7825475	23.98	−1.23
hdeB	6.73E-08	0.9999980	6.48E-01	0.7425627	28.76	−1.11
gadA	5.14E-07	0.9999915	9.85E-01	0.6548496	22.98	−1.01
poxB	1.93E-05	0.9998871	8.53E-01	0.6866352	2.73	1.04
cyaA	4.00E-05	0.9998098	8.79E-01	0.6802070	1.60	−1.03
ybeD	6.14E-05	0.9997421	6.30E-01	0.7479069	4.36	1.14
mrr	2.30E-04	0.9993396	6.80E-01	0.7332206	3.33	1.11
ygjD	4.07E-04	0.9990107	5.91E-01	0.7597149	5.13	1.18
ubiC	7.43E-04	0.9984854	4.28E-01	0.8135841	3.08	1.24

TABLE V

Regulatory pattern V: genes that exhibit increased levels during anaerobic growth and further increased levels in an fnr-deficient strain

Gene	*p*-value ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	PPDE(<*p*) ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	*p*-value ($-O_2$, +FNR *vs.* $-O_2$, $-$FNR)	PPDE(<*p*) ($-O_2$, +FNR *vs.* $-O_2$, $-$FNR)	Fold ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	Fold ($-O_2$, +FNR *vs.* $-O_2$, $-$FNR)
ybjX	3.47E-05	0.9998281	1.73E-04	0.9993848	4.45	3.73

TABLE VI

Regulatory pattern VI: genes that exhibit similar levels during aerobic and anaerobic growth but increased levels in an fnr-deficient strain

Gene	*p* value ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	PPDE(<*p*) ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	*p* value ($-O_2$, +FNR *vs.* $-O_2$, $-$FNR)	PPDE(<*p*) ($-O_2$, +FNR *vs.* $-O_2$, $-$FNR)	Fold ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	Fold ($-O_2$, +FNR *vs.* $-O_2$, $-$FNR)
gapA	6.54E-01	0.7420081	4.38E-07	0.9999905	−1.07	3.76
hisJ	8.98E-01	0.6771080	1.03E-04	0.9995715	−1.08	25.08
ydcF	9.76E-01	0.6587620	1.50E-04	0.9994422	1.01	4.32
purN	5.10E-01	0.7858719	3.71E-04	0.9989497	1.14	3.07
yhbJ	7.67E-01	0.7107023	6.71E-04	0.9984125	1.04	2.27
nrdB	6.54E-01	0.7419271	8.05E-04	0.9981985	−1.05	2.47
gapC_1	5.46E-01	0.7746369	8.88E-04	0.9980705	−1.19	8.09
artI	9.90E-01	0.6553850	1.33E-03	0.9974374	1.00	4.54
potF	6.51E-01	0.7428616	1.28E-03	0.9975029	−1.08	2.56

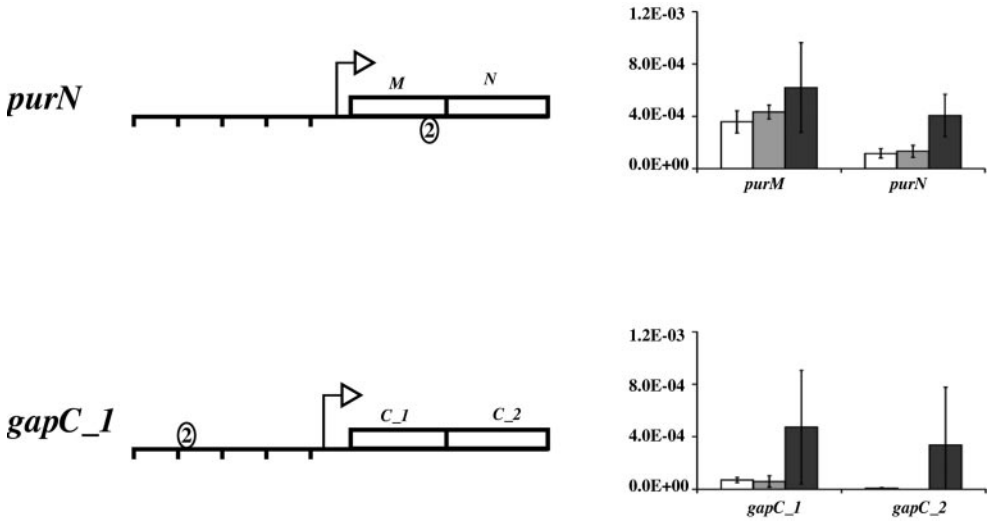

FIG. 9. **Genes of regulatory pattern VI.** See legend for Fig. 6.

oxygen and/or FNR under anaerobic growth conditions, and 8 contain putative FNR-binding sites (Fig. 11).

Three of the genes of this cluster (*gcvT*, *lysU*, and *dadX*) are known to be regulated by Lrp under aerobic growth conditions. Again, the genes of this cluster are mostly members of the small molecule metabolism and transport or molecular biosynthesis functional classes.

The four genes reported to be regulated by oxygen and/or FNR are *fumB, lysU, nikA,* and *narX*. The *fumB* gene encodes the anaerobic fumarase involved in the reductive pathway from oxaloacetate to succinate during anaerobic growth. Woods and Guest (46) demonstrated that the *fumB* gene is elevated 4-fold during anaerobic fermentative growth, and Tseng (47) showed that both ArcA and FNR are responsible for this anaerobic activation. Although our microarray data indicate that *fumB* is not regulated with respect to oxygen, this is probably a result of

TABLE VII

Regulatory pattern VII: genes that exhibit decreased levels during anaerobic growth and further decreased levels in an fnr-deficient strain

Gene	*p* value ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	PPDE(<*p*) ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	*p* value ($-O_2$, +FNR *vs.* $-O_2$, $-$FNR)	PPDE(<*p*) ($-O_2$, +FNR *vs.* $-O_2$, $-$FNR)	Fold ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	Fold ($-O_2$, +FNR *vs.* $-O_2$, $-$FNR)
frdA	1.41E-07	0.9999966	2.20E-06	0.9999708	−2.04	−5.03
ybdE	2.33E-04	0.9993343	4.65E-05	0.9997536	−1.77	−3.29
nirB	6.62E-07	0.9999898	3.40E-04	0.9990127	−2.04	−2.09
ylcD	2.00E-08	0.9999992	2.24E-04	0.9992522	−3.04	−2.99
rpmC	2.45E-09	0.9999998	5.10E-05	0.9997380	−3.20	−11.32

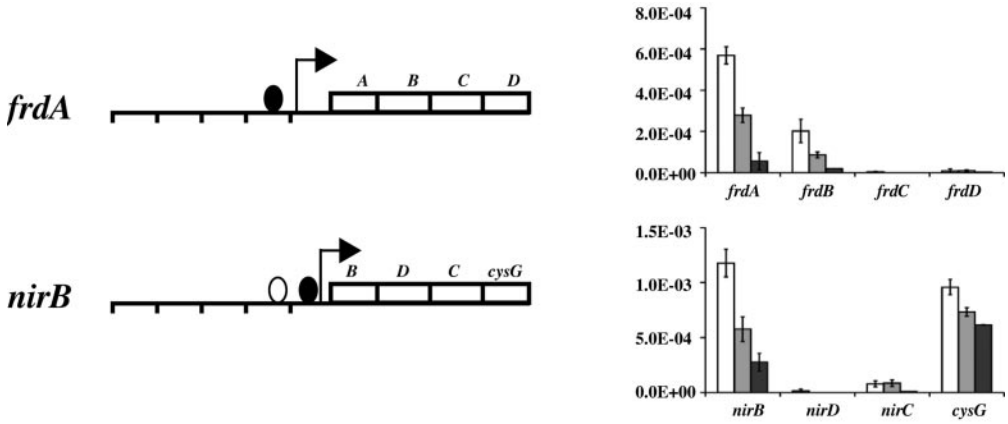

FIG. 10. **Genes of regulatory pattern VII.** See legend for Fig. 6.

TABLE VIII

Regulatory pattern VIII: genes that exhibit similar levels during aerobic and anaerobic growth but decreased levels in an fnr-deficient strain

Gene	*p* value ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	PPDE(<*p*) ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	*p* value ($-O_2$, +FNR *vs.* $-O_2$, $-$FNR)	PPDE(<*p*) ($-O_2$, +FNR *vs.* $-O_2$, $-$FNR)	Fold ($+O_2$, +FNR *vs.* $-O_2$, +FNR)	Fold ($-O_2$, +FNR *vs.* $-O_2$, $-$FNR)
nikA	7.62E-01	0.7119911	5.89E-06	0.9999420	−1.04	−5.43
nadC	9.76E-01	0.6586042	6.93E-05	0.9996743	−1.01	−6.53
ycbO	7.90E-01	0.7045299	9.46E-05	0.9995954	1.04	−4.34
narX	5.50E-01	0.7733437	1.23E-04	0.9995151	1.07	−2.42
yjiL	5.29E-01	0.7797786	1.30E-04	0.9994941	1.10	−4.13
emrK	8.36E-01	0.6925384	1.32E-04	0.9994898	−1.03	−4.05
ycfT	5.72E-01	0.7666014	1.45E-04	0.9994572	1.10	−4.15
yaaU	9.99E-01	0.6533494	1.53E-04	0.9994334	−1.00	−1.95
yi41	7.46E-01	0.7163046	1.67E-04	0.9993988	−1.03	−1.89
ilvY	8.29E-01	0.6944700	1.70E-04	0.9993902	−1.02	−2.18
lysU	9.48E-01	0.6652543	1.78E-04	0.9993727	−1.01	−2.69
yqiG	9.91E-01	0.6553178	1.86E-04	0.9993519	1.00	−2.35
ujcP	7.28E-01	0.7212742	2.53E-04	0.9991966	1.04	−2.55
ybhK	9.92E-01	0.6549812	2.84E-04	0.9991456	1.00	−3.26
parC	6.15E-01	0.7534451	2.92E-04	0.9991158	1.11	−5.24
yiaY	5.43E-01	0.7755093	3.32E-04	0.9990277	1.08	−2.43
yfbS	7.28E-01	0.7211014	3.36E-04	0.9990207	−1.03	−2.26
sdaB	7.39E-01	0.7181570	3.86E-04	0.9989210	−1.04	−2.40
fhuC	7.66E-01	0.7108355	4.81E-04	0.9987423	1.03	−1.94
yjeT	5.58E-01	0.7707909	5.08E-04	0.9986922	1.13	−5.55
b1012	7.62E-01	0.7117933	5.31E-04	0.9986516	−1.04	−2.43
b2438	9.69E-01	0.6602961	5.37E-04	0.9986372	1.01	−5.31
gcvT	7.82E-01	0.7067209	5.38E-04	0.9986435	1.05	−3.20
fumB	6.59E-01	0.7406371	5.71E-04	0.9985813	1.04	−1.82
yceL	8.64E-01	0.6854618	5.80E-04	0.9985671	1.02	−2.10
dfp	6.09E-01	0.7552905	6.01E-04	0.9985299	1.06	−2.18
umuC	5.60E-01	0.7701812	6.49E-04	0.9984495	1.07	−2.25
aroK	7.26E-01	0.7217079	7.46E-04	0.9982947	−1.05	−2.91
dadX	6.28E-01	0.7496791	7.61E-04	0.9982677	−1.10	−4.94
aroP	6.23E-01	0.7509828	8.32E-04	0.9981554	1.04	−1.67
aat	9.88E-01	0.6559196	8.64E-04	0.9981017	1.00	−4.58
yaaJ	5.41E-01	0.7761592	9.12E-04	0.9980332	1.05	−1.73
yigN	8.54E-01	0.6879241	9.67E-04	0.9979519	−1.02	−1.87

the high sequence identity (80%) between *fumB* and the aerobically expressed fumarase, *fumA*. The *lysU* gene encodes one of the two lysyl-tRNA synthetases (the other being *lysS*), reported previously to be regulated by Lrp under aerobic conditions (48) and induced under anaerobic conditions but not by FNR (49). However, the *lysU* ORF shares 79% sequence identity with *lysS*, which is apparently expressed constitutively (50).

The *nikA* gene encodes a periplasmic binding protein that acts as the primary substrate receptor for nickel uptake. Wu

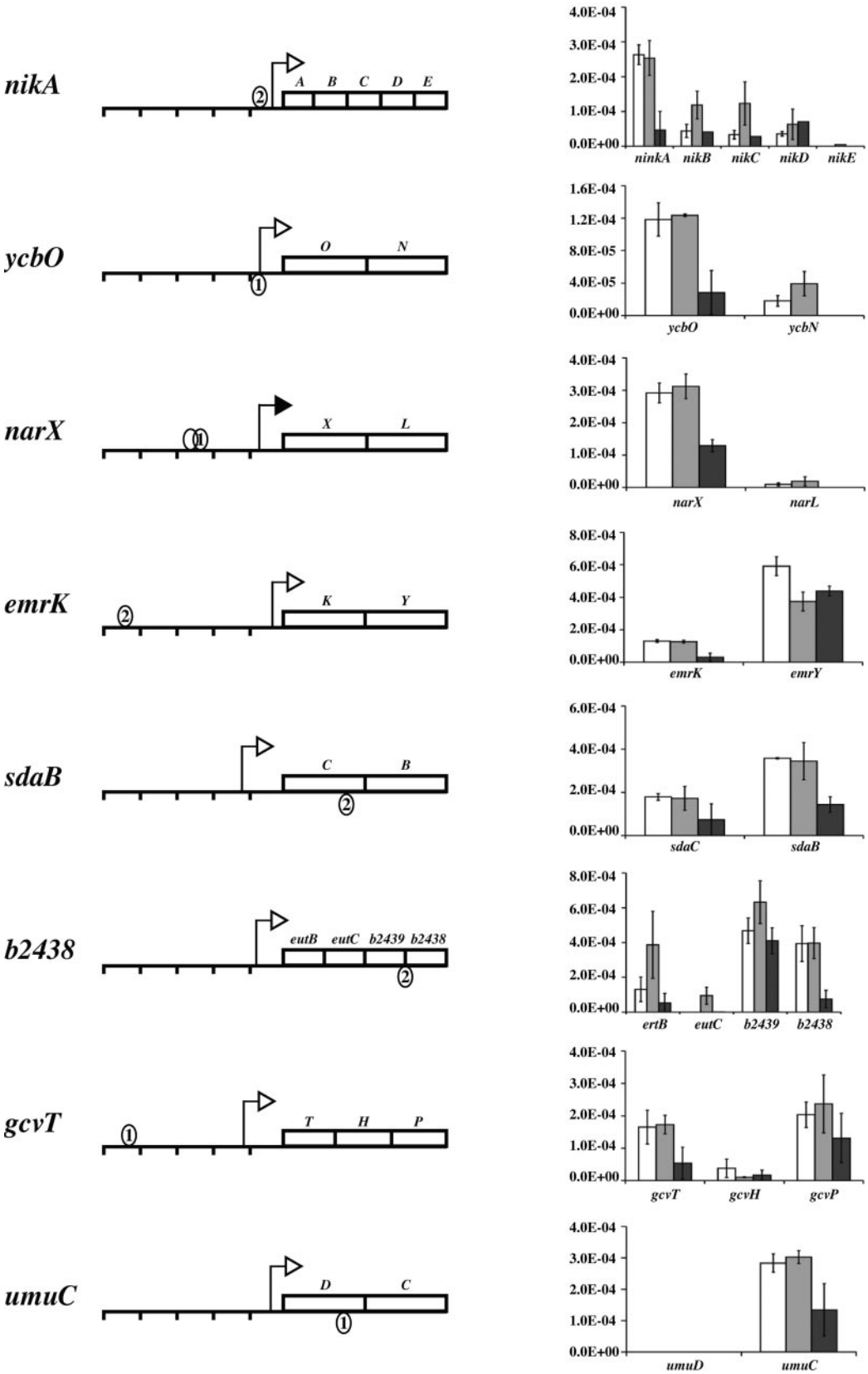

FIG. 11. **Genes of regulatory pattern VIII.** See legend for Fig. 6.

and Mandrand-Berthelot (51) have demonstrated that the transcription of the *nikABCDE* operon is anaerobically induced and that this induction is FNR-dependent. The final gene of this group, *narX*, encodes one of the two nitrate response membrane-bound sensor kinases that control the expression of many anaerobic electron transport and fermentation-related genes in response to nitrate. Darwin and Stewart (52) used a *narX-lacZ* operon fusion to demonstrate no change in expression with respect to oxygen but a slight repression by FNR in anaerobiosis.

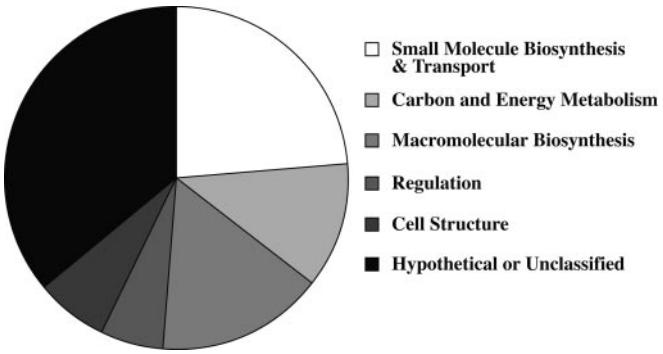

FIG. 12. **Distribution of functions for genes affected by oxygen availability and FNR.** The distribution of the 205 genes with PPDE(<*p*) values greater than 0.99 and *p* values less than 0.0013 are as follows: small molecules biosynthesis and transport 48; carbon and energy metabolism 24; macromolecular biosynthesis 32; regulation 12; cell structure 14; hypothetical and unclassified 75.

Six genes of this regulatory pattern are involved in small molecule metabolism, two of which are known to be regulated by Lrp (53). These are as follows: *nadC*, encoding quinolinate phosphoribosyltransferase; *sdaB*, encoding an Lrp-regulated L-serine deaminase; *dfp*, encoding a flavoprotein affecting DNA synthesis and pantothenate metabolism; *aroK*, encoding shikamate kinase; *dadX*, encoding an Lrp-regulated alanine racemase; and, *aat*, encoding aminoacyl-tRNA-protein transferase.

Three genes are involved in small molecule transport. These are as follows: *emrK*, encoding a putative multidrug resistance pump; *fhuC*, involved in ferric hydroxamate-dependent iron uptake; and *aroP*, encoding a general aromatic amino acid transport protein.

Three genes are involved in macromolecular synthesis: *eutK* (*b2438*), encoding a putative carboxysome structural protein; *umuC*, involved in UV induction of mutations; and *parC*, involved in cell partitioning.

The final members of this cluster are *ilvY* and *gcvT*. The *ilvY* gene is the positive regulator for *ilvC* gene expression, a gene required for branched chain amino acid biosynthesis. These pathways are known to be regulated by Lrp under aerobic growth conditions (11, 54). The *gcvTHP* operon encodes proteins that cleave glycine to produce one-carbon units and ammonia. This gene is also regulated by Lrp under aerobic growth conditions (53).

The functional class distribution of the 205 genes of regulatory patterns I to VIII is shown in Fig. 12. Roughly 35% are hypothetical or unclassified, and another 23% are involved in small molecule metabolism. Most of the previously documented oxygen-controlled genes fall into the categories of carbon/energy metabolism (15%).

Genes Not Expressed in at Least One Experiment—Only those genes exhibiting an expression level greater than zero in all experiments were used for statistical analysis as described above. To identify differentially expressed genes that are not expressed under one condition but turned on under another treatment condition (or vice versa), gene measurements containing zero expression values were set aside and are listed in Table IX. This set contains 25 genes with zero expression values for all measurements in at least one experiment, and expression levels greater than 1×10^{-5} of total mRNA for all measurements in at least one of the remaining experiments. The significance of these results was evaluated by the coefficients of variance of the four greater than zero measurements. These genes are grouped according to their regulatory patterns, described above, in ascending order of their coefficients of variance (Table IX). Four of these genes contain putative FNR-binding sites (Fig. 13).

SUMMARY AND CONCLUSIONS

In this report we describe and employ statistical methods appropriate for high dimensional DNA microarray experiments for the identification of differentially expressed genes with a high level of confidence. These methods allow us to infer differential expression for more than one-third of the 4,290 genes of *E. coli* during growth in the presence or absence of oxygen (Fig. 2*A*). We further observe that about one-half of these changes in expression levels are mediated, either directly or indirectly, by FNR (Fig. 2*B*). These results show that the network of genes required for the transition of cells from aerobic to anaerobic growth conditions is as much as 10 times larger than previously suspected. This discovery emphasizes the power of global gene expression profiling experiments over previous one-gene or one-operon approaches.

Previous approaches have revealed about 70 genes that are regulated by FNR or ArcAB under anaerobic growth conditions. The results presented here identify 166 among 205 genes with *p* values less than 0.0013 and PPDE(<*p*) values above 0.99 whose expression is affected by FNR. However, if we include all genes expressed at a level above background and examine the PPDE *versus p* value plots, we see that our confidence for any gene expressed at a *p* value less than 1 (PPDE(<*p*)) is 0.63. This means that we can be 63% confident that any gene in our data set is differentially expressed, *i.e.* 63% of the 2,820 or about 1,700 of the genes analyzed in these experiments are differentially expressed between cells grown in the presence or absence of oxygen (Fig. 3*A*). In the same manner, the data in Fig. 3*B* show that about 60% of these 1,700 genes or 1,000 genes are either directly or indirectly regulated by FNR. Thus, these results greatly expand our knowledge of genes that compose these regulatory networks.

Regulatory patterns I (anaerobically repressed, decreased expression in the presence of FNR) and II (anaerobically activated, increased expression in the presence of FNR) where FNR acts as a repressor (Table I, Fig. 4, and Fig. 6) or activator (Table II, Fig. 4, and Fig. 7) are most easily reconciled with previous reports. Of the 94 genes of these patterns about one-quarter (23) contain known or putative FNR-binding site motifs. Among this subset four contain previously characterized FNR-binding sites. These results suggest that we might expect the total number of genes directly activated or repressed by FNR to be in the range of 200 genes.

Regulatory patterns III (anaerobically repressed, not affected by FNR) and IV (anaerobically activated, not affected by FNR) are most easily explained as genes affected by the ArcAB system or as yet unidentified global regulators such as Lrp, IHF, FIS, or H-NS. Furthermore, because only 2 of the 57 genes of these groups are known to be regulated by ArcA, our results suggest that, like FNR, the number of ArcA-regulated genes may also be much larger than expected previously.

It is more difficult to understand the physiological roles that many of the genes of regulatory patterns V–VIII might play in anaerobic metabolism. However, it is striking that the genes of these patterns fall into the same functional classes as the functional classes of genes regulated by Lrp under aerobic conditions (11). It is tempting to speculate that these results reveal common functions between FNR and Lrp under aerobic or anaerobic growth conditions. This is perhaps not unreasonable because the roles of both of these global regulatory proteins are to coordinate gene expression levels with nutritional and environmental growth conditions. To illustrate this overlap between genes regulated by FNR and Lrp, we compared the 500 genes with the highest PPDE(*p*) values (>0.991) and the lowest *p* values (<0.0014) obtained from the array experiments reported here comparing *fnr* isogenic strains under anaerobic

TABLE IX
Genes not expressed in at least one experiment

Gene	Mean ($+O_2$ +FNR)	Mean ($-O_2$ +FNR)	Mean ($-O_2$ −FNR)	Coefficient of variance ($+O_2$ +FNR)	Coefficient of variance ($-O_2$ +FNR)	Coefficient of variance ($-O_2$ −Fnr)
Regulatory pattern I						
cspA	1.18E-03	0.00E + 00	7.96E-05	0.45	—[a]	0.76
b1586	2.17E-05	0.00E + 00	1.54E-05	0.70	—	0.88
hha	7.11E-06	0.00E + 00	2.28E-05	1.00	—	1.12
Regulatory pattern III						
rfbA	7.30E-06	0.00E + 00	0.00E + 00	1.05	—	—
nirD	1.52E-05	0.00E + 00	0.00E + 00	1.05	—	—
b3914	9.29E-05	0.00E + 00	0.00E + 00	1.04	—	—
Regulatory pattern IV						
hyaC	0.00E + 00	1.15E-04	9.89E-05	—	0.16	1.40
hyaF	0.00E + 00	1.57E-04	1.83E-04	—	0.28	1.42
hyaA	0.00E + 00	3.49E-04	6.63E-04	—	0.41	1.26
yhiE	0.00E + 00	4.69E-04	4.06E-04	—	0.57	1.17
Regulatory pattern V						
dps	0.00E + 00	4.90E-05	1.18E-03	—	0.52	1.39
ygiN	0.00E + 00	1.79E-05	1.25E-04	—	0.91	0.54
baeR	0.00E + 00	6.59E-06	1.01E-04	—	0.74	0.30
Regulatory pattern VI						
osmB	0.00E + 00	0.00E + 00	3.29E-04	—	—	0.17
trxC	0.00E + 00	0.00E + 00	2.77E-05	—	—	0.63
marR	0.00E + 00	0.00E + 00	4.67E-05	—	—	1.06
b2833	0.00E + 00	0.00E + 00	8.48E-05	—	—	1.54
Regulatory pattern VII						
galF	1.36E-04	7.95E-05	0.00E + 00	0.08	0.31	—
sohA	4.34E-05	1.09E-05	0.00E + 00	0.37	0.70	—
Regulatory pattern VIII						
ansB	3.05E-04	2.49E-04	0.00E + 00	0.44	0.14	—
metR	3.36E-05	5.16E-05	0.00E + 00	0.50	0.40	—
thrA	9.13E-06	1.09E-05	0.00E + 00	0.58	0.54	—
b1629	1.39E-05	1.65E-05	0.00E + 00	0.83	0.72	—
b1345	9.67E-06	1.30E-05	0.00E + 00	1.38	0.86	—
b0817	1.19E-05	2.12E-05	0.00E + 00	0.83	1.10	—

[a] —, insufficient information for calculation.

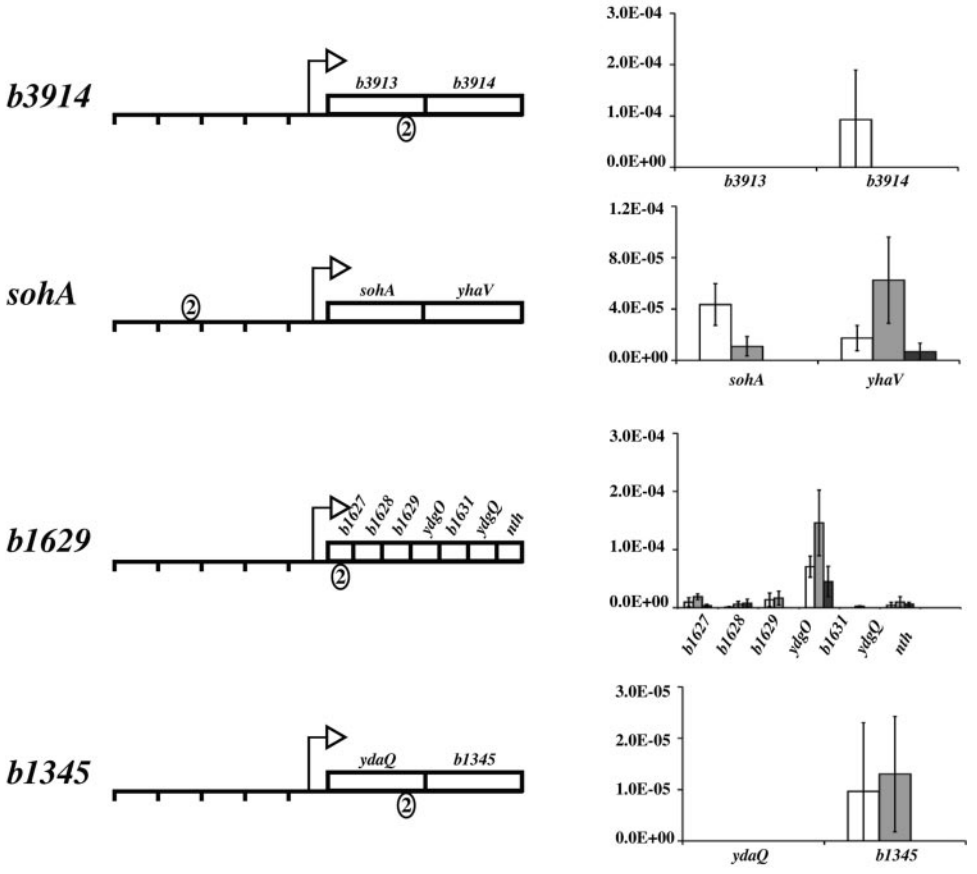

FIG. 13. **Genes not expressed in at least one experiment.** See legend for Fig. 6.

growth conditions with the 500 genes with the highest PPDE(p) values (>0.80) and the lowest p values (<0.027) obtained from Lrp array experiments comparing *lrp* isogenic strains under aerobic growth conditions (16). Among these two gene sets, nearly 25% (120 genes) are present in both lists. It is of further interest that *lrp* expression is increased 5-fold (p = 0.0002; PPDE(p) = 0.99) in the FNR-deficient strain during growth under anaerobic conditions. This suggests a functional role for Lrp during anaerobic growth in an FNR-deficient strain. However, because there has been no previous motivation to examine the roles of Lrp on anaerobic gene expression, the test of this possibility must await further experimentation. As additional gene expression profiling experiments are performed, a better understanding of interactions among these and other large regulatory networks, and the systems biology of *E. coli*, will surely emerge.

Acknowledgments—We are grateful to Cambridge University Press (Cambridge, UK) for permitting reproduction of materials contained in the book "*DNA Microarrays and Gene Expression: From Experiments to Data Analysis and Modeling*" (12).

REFERENCES

1. Gunsalus, R. P., and Park, S. J. (1994) *Res. Microbiol.* **145,** 437–450
2. Gennis, R., and Stewart, V. (1996) in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (Neidhartt, F. C., ed) Vol. 1, pp. 217–261, American Society for Microbiology, Washington, D. C.
3. Guest, J. R., Attwood, M. M., Machado, R. S., Matqi, K. Y., Shaw, J. E., and Turner, S. L. (1997) *Microbiology* **143,** 457–466
4. Lynch, A. S., and Lin, E. C. C. (1996) in *Regulation of Gene Expression in Escherichia coli* (Lin, E. C. C., and Lynch, A. S., ed) pp. 362–381, Chapman & Hall, New York
5. Guest, J. R., Green, J., Irvine, A. S., and Spiro, S. (1996) in *Regulation of Gene Expression in Escherichia coli* (Lin, E. C. C., and Lynch, A. S., eds) pp. 317–342, Chapman & Hall, New York
6. Kiley, P. J., and Beinert, H. (1998) *FEMS Microbiol. Lett.* **22,** 341–352
7. Lambden, P. R., and Guest, J. R. (1976) *J. Gen. Microbiol.* **97,** 145–160
8. Lynch, A. S., and Lin, E. C. C. (1996) in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (Neidhardt, F. C., ed) Vol. 1, 2nd Ed., pp. 1526–1538, American Society of Microbiology, Washington, D. C.
9. Casadaban, M. J. (1976) *J. Mol. Biol.* **104,** 541–555
10. Cotter, P. A., and Gunsalus, R. P. (1989) *J. Bacteriol.* **171,** 3817–3823
11. Hung, S. P., Baldi, P., and Hatfield, G. W. (2002) *J. Biol. Chem.* **277,** 40309–40323
12. Baldi, P., and Hatfield, G. W. (2002) *DNA Microarrays and Gene Expression: From Experiments to Data Analysis and Modeling*, Cambridge University Press, Cambridge, UK
13. Hatfield, G. W., Hung, S. P., and Baldi, P. (2003) *Mol. Microbiol.* **47,** 871–877
14. Long, A. D., Mangalam, H. J., Chan, B. Y., Tolleri, L., Hatfield, G. W., and Baldi, P. (2001) *J. Biol. Chem.* **276,** 19937–19944
15. Baldi, P., and Long, A. D. (2001) *Bioinformatics* **17,** 509–519
16. Allison, D. B., Gadbury, G. L., Heo, M., Fernandez, J. R., Lee, C. K., Prolla, T. A., and Weindruch, R. (2002) *Comput. Statist. Data Anal.* **39,** 1–20
17. Baldi, P., and Hatfield, G. W. (2002) *DNA Microarrays and Gene Expression: From Experiments to Data Analysis and Modeling*, pp. 73–96, Cambridge University Press, Cambridge, UK
18. Tseng, C. P., Albrecht, J., and Gunsalus, R. P. (1996) *J. Bacteriol.* **178,** 1094–1098
19. Iuchi, S., Chepuri, V., Fu, H. A., Gennis, R. B., and Lin, E. C. (1990) *J. Bacteriol.* **172,** 6020–6025
20. Cotter, P. A., and Gunsalus, R. P. (1992) *FEMS Microbiol. Lett.* **70,** 31–36
21. Cotter, P. A., Chepuri, V., Gennis, R. B., and Gunsalus, R. P. (1990) *J. Bacteriol.* **172,** 6333–6338
22. Cotter, P. A., Melville, S. B., Albrecht, J. A., and Gunsalus, R. P. (1997) *Mol. Microbiol.* **25,** 605–615
23. Govantes, F., Albrecht, J. A., and Gunsalus, R. P. (2000) *Mol. Microbiol.* **37,** 1456–1469
24. Spiro, S., Roberts, R. E., and Guest, J. R. (1989) *Mol. Microbiol.* **3,** 601–608
25. Sharrocks, A. D., Green, J., and Guest, J. R. (1991) *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **245,** 219–226
26. Green, J., and Guest, J. R. (1994) *Mol. Microbiol.* **12,** 433–444
27. Meng, W., Green, J., and Guest, J. R. (1997) *Microbiology* **143,** 1521–1532
28. Bongaerts, J., Zoske, S., Weidner, U., and Unden, G. (1995) *Mol. Microbiol.* **16,** 521–534
29. Weidner, U., Geier, S., Ptock, A., Friedrich, T., Leif, H., and Weiss, H. (1993) *J. Mol. Biol.* **233,** 109–122
30. Compan, I., and Touati, D. (1994) *Mol. Microbiol.* **11,** 955–964
31. Quail, M. A., Haydon, D. J., and Guest, J. R. (1994) *Mol. Microbiol.* **12,** 95–104
32. Cunningham, L., Georgellis, D., Green, J., and Guest, J. R. (1998) *FEMS Microbiol. Lett.* **169,** 403–408
33. Birkmann, A., Sawers, R. G., and Bock, A. (1987) *Mol. Gen. Genet.* **210,** 535–542
34. Ravnikar, P. D., and Somerville, R. L. (1987) *J. Bacteriol.* **169,** 2611–2617
35. Ferrandez, A., Garcia, J. L., and Diaz, E. (1997) *J. Bacteriol.* **179,** 2573–2581
36. Eichler, K., Bourgis, F., Buchet, A., Kleber, H. P., and Mandrand-Berthelot, M. A. (1994) *Mol. Microbiol.* **13,** 775–786
37. Janausch, I. G., and Unden, G. (1999) *Arch. Microbiol.* **172,** 219–226
38. Abaibou, H., Pommier, J., Benoit, S., Giordano, G., and Mandrand-Berthelot, M. A. (1995) *J. Bacteriol.* **177,** 7141–7149
39. Chang, Y. Y., Wang, A. Y., and Cronan, J. E., Jr. (1994) *Mol. Microbiol.* **11,** 1019–1028
40. Soballe, B., and Poole, R. K. (1997) *FEBS Lett.* **414,** 373–376
41. Blankenhorn, D., Phillips, J., and Slonczewski, J. L. (1999) *J. Bacteriol.* **181,** 2209–2216
42. Casado, C., Llagostera, M., and Barbe, J. (1991) *FEMS Microbiol. Lett.* **67,** 153–157
43. Jones, H. M., and Gunsalus, R. P. (1987) *J. Bacteriol.* **169,** 3340–3349
44. Bell, A. I., Cole, J. A., and Busby, S. J. (1990) *Mol. Microbiol.* **4,** 1753–1763
45. Jayaraman, P. S., Cole, J., and Busby, S. (1989) *Nucleic Acids Res.* **17,** 135–145
46. Woods, S. A., and Guest, J. R. (1988) *Biochim. Biophys. Acta* **954,** 14–26
47. Tseng, C. P. (1997) *FEMS Microbiol. Lett.* **157,** 67–72
48. Lin, R., Ernsting, B., Hirshfield, I. N., Matthews, R. G., Neidhardt, F. C., Clark, R. L., and Newman, E. B. (1992) *J. Bacteriol.* **174,** 2779–2784
49. Leveque, F., Gazeau, M., Fromant, M., Blanquet, S., and Plateau, P. (1991) *J. Bacteriol.* **173,** 7903–7910
50. Hirshfield, I. N., Tenreiro, R., Vanbogelen, R. A., and Neidhardt, F. C. (1984) *J. Bacteriol.* **158,** 615–620
51. Wu, L. F., and Mandrand-Berthelot, M. A. (1986) *Biochimie* (*Paris*) **68,** 167–179
52. Darwin, A. J., and Stewart, V. (1995) *J. Bacteriol.* **177,** 3865–3869
53. Newman, E. B., Lin, R. T., and D'Ari, R. (1996) in *Escherichia coli and Salmonella. Cellular and Molecular Biology* (Neidhardt, F. C., Curtis, R. I., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., eds) Vol. 1, 2nd Ed., pp. 1513–1525, American Society for Microbiology, Washington, D. C.
54. Rhee, K. Y., Parekh, B. S., and Hatfield, G. W. (1996) *J. Biol. Chem.* **271,** 26499–26507