S8-5

Comparative Genome-Scale Expression Analysis of Growth Phase-dependent Genes in Wild Type and *rpo*S Mutant of *Escherichia coli*

Tae Jeong Oh[1], Il Lae Jung[2], Sook Kyung Woo[1], Myung Soon Kim[1], Sun Woo Lee[1], Keun Ha Kim[1], In Gyu Kim[2], and Sung Whan An[1]

[1]R&D Center, GenomicTree, Inc. Daejeon, Korea; [2]Environmental Radiation Research Group, Korea Atomic Energy Research Institute, Daejeon, Korea

Abstract

Numerous genes of *Escherichia coli* have been shown to growth phase-dependent expression throughout growth. The global patterns of growth phase-dependent gene expression of *E. coli* throughout growth using oligonucleotide microarrays containing a nearly complete set of 4,289 annotated open reading frames. To determine the change of gene expression throughout growth, we compared RNAs taken from timecourses with common reference RNA, which is combined with equal amount of RNA pooled from each time point. The hierarchical clustering of the conditions in accordance with timecourse expression revealed that growth phases were clustered into four classes, consistent with known physiological growth status. We analyzed the differences of expression levels at genome level in both exponential and stationary growth phase cultures. Statistical analysis showed that 213 genes are shown to growth phase-dependent expression. We also analyzed the expression of 256 known operons and 208 regulatory genes. To assess the global impact of RpoS, we identified 193 genes coregulated with *rpoS* and their expression levels were examined in the isogenic *rpoS* mutant. The results revealed that 99 of 193 were novel RpoS-dependent stationary phase-induced genes and the majority of those are functionally unknown. Our data provide that global changes and adjustments of gene expression are coordinately regulated by growth transition in *E. coli*.

Introduction

For past five decades, extensive studies for *Escherichia coli* K-12 have been established and provided useful information on the genes, operons and proteins required for diverse physiologies (1). Most researchers have focused on the physiology of exponential growth, nutrient shifts and entry into stationary phase (2). Recently, comprehensive gene expression profiling using high-density microarray with *E. coli* have been reported (3-6). The *E. coli* microarray have allowed the monitoring of individual gene's expression profiles. The current availability of DNA microarray containing a complete set of 4,290 ORFs of *E. coli* permits an unprecedented ability not only to look at a single aspect of physiology but also to see

how a particular gene, regulon, or modulon interact with every other aspect of physiology.

During *E. coli* cells growing in the medium, numerous genes were transcriptionally regulated depending on growth phases. In exponentially growing cells, it is necessary to increase the transcription/translation rate, the flow of carbon and electron and the rate of building block biosynthesis to support fast growth (4, 7-9). When *E. coli* cells enter the stationary phase due to nutrient depletion, extensive reprogramming of physiology and gene expression is accomplished to cell survival under extreme conditions (10, 11). The RpoS, alternative sigma factor, can partially replace the vegetative sigma factor 70 under many stress conditions and is required to induce genes needed in the stationary phase (12-14). Even though a number of RpoS-dependent genes have been identified, much remains to be elucidated about RpoS-mediated response (14). Nevertheless, there have been less systematic approaches to analyze the expression profile at the genome level in responses to growth transition throughout growth.

In the present work, We determined gene expression changes in *E. coli* throughout growth on LB medium in the wild type using oligonucleotide microarrays containing a nearly complete set of 4,289 annotated open reading frames in the *E. coli* K-12 genome. We also analyzed growth phase-dependent genes by statistical analysis. To assess the global impact of master regulator RpoS on the gene expression, we identified coregulated genes with *rpoS* and their expression levels were compared with isogenic *rpoS* mutant. Our data will provide new insights into growth phase-dependent gene expression, global regulation of biosynthetic genes and stress responses that appear to be involved in growth on LB medium.

Materials and Methods

Bacterial strains and growth conditions

Bacterial strains used in this study are *E. coli* K-12 MG1655 [F- λ- *rph*-] (CGSC) and isogenic *rpoS* mutant. JIL645 (MG1655 *rpoS13*::Tn*10* (15)) was constructed by P1 mediated transduction and routinely checked for their increased sensitivity to hydrogen peroxide.

Isolation of total RNA and preparation of common reference pool

Total RNA was isolated as described previously (5). To cross compare multiple data sets from timecourse experiments, we prepared a single common reference RNA pool combined with RNAs taken during the timecourses for optimal broad gene coverage. Common reference RNA was prepared by combining of the equal amount (20 μg) of total RNA isolated from wild type MG1655 every hour (0 to 12 hour).

Microarray fabrication

The library consisted of 65 nucleotides probes represented 4,289 coding sequences were purchased from Sigma-Genosys (Woodlands, Texas, USA). All oligonucleotides were resuspended in printing buffer (Telechem International, Inc., USA) at final concentration of 50 pmole/μl.

Preparation of cDNA probe and microarray hybridization

The synthesis of target cDNA probes and hybridization were performed according to previously

described (Tani et al. 2002). The arrays were hybridized at 42°C for 12 - 16 h in the humidified hybridization chamber (Array Chamber X, GenomicTree Inc., Korea). The hybridized miroarrays were washed and quantification was performed using Axon 4000B scanner (Axon Instruments, CA).

Data acquisition and analysis

The hybridization images were quantitated by GenePix Pro 4.0 (Axon Instruments, CA). The average fluorescence intensity for each spot was calculated and local background was subtracted. Microarray data analysis was performed using GeneSpring 6.2 (Silicon Genetics, USA).

Functional analysis and data access

Functional annotation was performed according to EcoCyc Database or National Center for Biotechnology Information (16). In case where we found no corresponding gene in EcoCyc Database, the original identification B numbers were used (16).

Results

Whole genome expression profiles

To determine the whole genome expression profiles in wild type MG1655 throughout growth in LB medium, we measured gene expression changes in accordance with timecourses (0 to 12 h). As shown in Fig. 1A, growth of wild type MG1655 could be distinguished into four characteristic growth phases (A to D) during growth in LB. Clustering analysis showed all thirteen time conditions were successfully clustered according to known physiological growth status, in which four clustering like four growth phases (A to D of Fig. 1A) could be seen. As shown in Fig. 1B, the average correlations of the exponential phase (1 to 3 h), the entry of stationary phase (4 and 5 h), early stationary phase (6 and 7 h) and stationary phase (8 to 11 h) were 0.71, 0.86, 0.93 and 0.78, respectively. Interestingly, gene expression profile of seed culture (time 0 h) was closely clustered with end time point 12 h with correlation of 0.89.

These results propose that all genes of *E. coli* are coordinately regulated according to growth phase or growth rate at genome level.

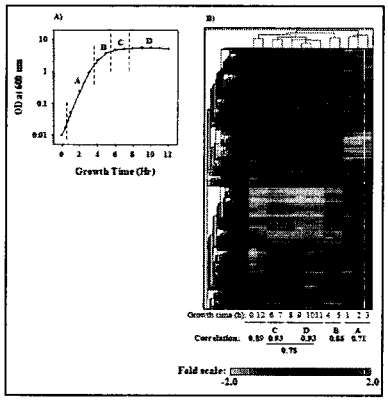

Fig. 1. Growth and gene expression profiles of *E. coli*. Cells were grown in LB medium at 37℃ with shaking for 12 h.

Analysis of growth phase-dependent genes

To identify the growth phase-dependent genes, we performed statistical analysis (ANOVA test, $p < 0.01$) with group comparison on the basis of three major groups of growth phases as determined by gene expression profiles. We identified 213 genes showing growth phase-dependent expression in the wild type. The 213 genes were hierarchically clustered into three characteristic subclasses according to their expression patterns (Fig. 2).

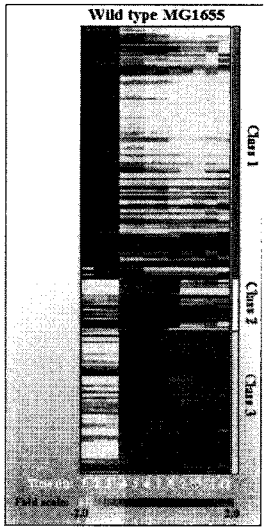

Fig. 2. Hierarchical clustering of growth phase-dependent genes in *E. coli*.

The genes of class 1 (121 genes) were highly elevated in the exponential phase and rapidly declined at the onset of the stationary phase. The majority of class 1 were involved in translation, post-translational modification. The genes of class 2 (24 genes) were significantly elevated at the onset of stationary growth phase and maintained during the early stationary phase. We found that 7 RpoS-dependent genes containing *rpoS* itself were clustered in class 2. The class 3 comprised stationary phase-induced 68 genes and the great portion of genes is unknown function. 11 of 68 genes were known RpoS-dependent genes. We also determined growth phase-dependent expression of known 256 operons and 208 regulatory genes during growth in LB medium.

Idenficiation of genes coregulated with *rpoS* by correlation analysis

The RpoS is one of the most well characterized top-level master regulator in *E. coli*. Microarray data revealed that the *rpoS* mRNA was rapidly induced at the onset of stationary phase, while gradually declined during the stationary phase (Fig. 3A). To identify coregulated genes with *rpoS* under the stationary phase, Pearson's correlation analysis was performed. Using a Peason correlation coefficient of 0.8 as the lower limit, we identified 193 coregulated genes (Fig. 3B) and found that 28 of 193 genes are within an operon.

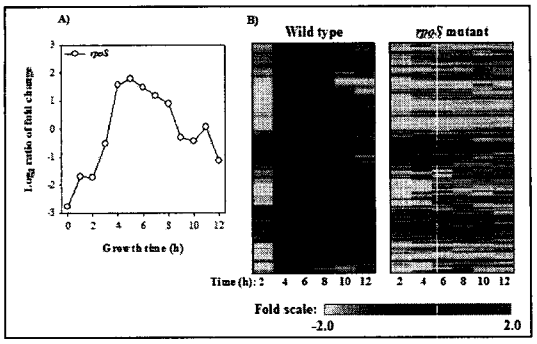

Fig. 3. Expression of *rpoS* in the wild type and comparison of RpoS-coregulated gene expression between wild type and *rpoS* mutant. Cells were grown in LB medium at 37°C.

We also found that 27 of 193 genes were known RpoS-dependent genes. To determine the effect of *rpoS*, the gene expression levels of 193 genes in the wild type was compared with *rpoS* mutant. The results revealed that the majority of genes were RpoS-dependent expression. 99 of 193 genes were novel RpoS-dependent stationary phase-induced genes, while 48 genes showed RpoS-independent expression. 19 genes were up-regulated in the *rpoS* mutant.

Discussion

In recent years, the huge data for *E. coli* gene expression profile has provided extensive biological information on a genome scale under various environmental and physiological conditions (3-6). We report here that new experimental design and whole genome expression profile of *E. coli*, especially in growth phase-dependent expression, throughout growth in LB medium.

For conventional microarray experiments, the expression levels detected for test RNA samples are directly compared to control RNA sample to calculate relative expression level. In addition, there is also prevailed simple pair-wise comparison between given conditions in *E. coli* genomic expression data. However, according to accumulation of expression data under different conditions, in order to compare more objectively gene expression levels of many different types of samples, it is necessary to common reference with ability to cross compare data. In this work, we prepared common reference RNA by combining the equal amount of RNAs pooled from thirteen time points (0 – 12 h) during growth in LB medium. To determine gene expression profile in accordance with growth transition, we indirectly compared common reference RNA pool with each RNA sample taken from every hour throughout growth (0 to 12 h), respectively. The true signals above background of common reference RNA pool labeled with Cy3 were detected more than 90% of 4,289 ORFs in more than seven of all thirteen hybridizations (data not shown). The results propose that common reference RNA used in this study was optimized to cover the broad range of genes in *E. coli*. The most important criterion for reference RNA pool is that cDNA synthesized from the reference RNA hybridizes to as many microarray probes as possible to make non-zero denominator. Therefore, we concluded that our common reference RNA pool is optimized to this criterion.

To ensure our experimental design (indirect comparison), we qualitatively compared data obtained by

this method with data taken from simple pair-wise comparison (direct comparison). For direct experiment, we compared RNAs extracted from cultures of 3 h (exponential phase) with 6 h (stationary phase) grown in LB medium. Meanwhile, to calculate fold difference of gene expression between 3 h and 6 h in indirect experiment, we estimated the ratios of fold changes measured by comparing 6 h and 3 h samples with common reference RNA. The correlation coefficient of 2,454 reliable genes in both comparisons was determined as 0.913, indicating that our experimental design is compatible with conventional direct comparison. In addition, to validate our microarray data, we performed semi-quantitative reverse transcription/PCR with four stationary phase-induced RpoS-dependent genes using same total RNA used in microarray experiments. The results showed that RT-PCR results were excellently matched with microarray results. Therefore, these results demonstrated that our experimental design is greatly optimized to genomic expression analysis of *E. coli* for comparison of multiple data sets from different conditions.

To determine whole gene expression profiles in accordance with growth in LB medium, we have extracted RNA samples from cultures every hour parallel with growth curve (Fig. 1A). Approximately 4000 reliable genes were hierarchically clustered without providing any preconception (Fig. 1B) and the condition trees revealed that classification of growth phase based on gene expression pattern was excellently matched with physiological growth phase and a high correlation between conditions in same growth phase. The results propose that gene expression patterns reflect on practical physiological status. Interestingly, the gene expression patterns of seed culture (0 h) were showed a high correlation ($r = 0.89$) with end point culture (12 h). Based upon gene expression profile, we categorized growth phases into three major groups and we performed statistical analysis to identify growth phase-dependent genes. We identified 213 growth phase-dependent genes showing significant differences across growth phases by statistical analysis ($p < 0.01$). The hierarchical clustering of statistically significant genes revealed three distinct classes, as shown in Fig. 2. When *E. coli* cells enter stationary phase, growth rate is rapidly decreased and occurred growth arrest (14). In this condition, the vast majority of genes involved in transcription and translation apparatus are markedly repressed, while in faster growing cells, these genes are highly induced (4). Our data are excellently matched with this known paradigm of gene regulation. It has been shown that the genes tightly coregulated under a variety of conditions have possibility that genes are functionally related or within an operon (5). To demonstrate this possibility, we performed correlation analysis using a *rpoS* as a model system. The correlation analysis (correlation coefficient ≥ 0.8) revealed that the number of 193 genes was coregulated parallel with *rpoS* expression. The genes showing coregulation with *rpoS* contained various genes with different biological function including metabolism, cellular processes and transcriptional regulation. 126 of 193 genes are known RpoS-dependent or novel RpoS-dependent genes. Our data provide that microarray analysis is useful tool for screening coregulated genes in *E. coli*.

In conclusion, we analyzed the whole genome expression profiles of *E. coli* throughout growth in LB medium in the wild type MG1655 using a high-density oligonucleotide microarray. We provide a new experimental design for studying genomic expression analysis to compare different data sets. The numerous genes are regulated depending on growth phase or growth rate parallel against cellular needs. This observation revealed that gene expression profiles essentially reflected on current physiological status of *E.*

coli under a given environment. Even though, we will not provide any details of regulatory mechanism in gene expression, our experimental design will provide a useful tool for understanding of gene regulation throughout growth in *E. coli* or other organisms.

Acknowledgments

This work was supported by R&D program of GenomicTree, Inc. We would like to thank Chiwang Yoon and Daekyung Yoon (GenomicTree, Inc.) for fabrication of HPV DNA microarrays.

References

1. Yan, W., J. M. Lee, C. Richmond, F. R. Blattner, J. A. Rafalski, and R. A. Larossa. 2001. Global impact of *sdiA* amplification revealed by comprehensive gene expression profiling of *Escherichia coli*. *J. Bacteriol.* **183:** 545-556.
2. Chang, D. E., D. J. Smalley, and T. Conway. 2002. Gene expression profiling of *Escherichia coli* growth transitions: an expanded stringent response model. *Mol. Microbiol.* **45:** 289-306.
3. Richmond, C. S., J. D. Glasner, R. Mau, H. Jin, and F. R. Blattner. 1999. Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acid Res.* **27**: 3821-3835.
4. Tao, H., C. Bausch, C. Richmond, F. R., Blattner, and T. Conway. 1999. Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. *J. Bacteriol.* **181**, 6425-6440.
5. Tani, T. H., A. Khodursky, R. M. Blumenthal, P. O. Brown, and R. G. Mathews. 2002. Adaptation to famine: a family of stationary-phase genes revealed by microarray analysis. *Proc. Natl. Acad. Sci. USA* **99:** 13471-13476.
6. Polen, T., D. Rittmann, V. F. Wendisch, and H. Sahm. 2003. DNA microarray analyses of the long-term adaptive response of *Escherichia coli* to acetate and propionate. *Appl. Environ. Microbiol.* **69:** 1759-1774.
7. Saier, M. H., T. M., Ramseier, and J. Reizer. 1996. Regulation of carbon utilization. In *Escherichia coli and Salmonella: cellular and molecular biology*, 2nd edition. (ed. Neidhardt, F. C., Curtiss, R., III., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E.), ASM press, Washington, D. C., pp. 1325-1343.
8. Magasanik, B. 1996. Regulation of nitrogen utilization. In *Escherichia coli and Salmonella: cellular and molecular biology*, 2nd edition. (eds. Neidhardt, F. C., Curtiss, R., III., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E.), ASM press, Washington, D. C., pp. 1344-1356.
9. Tong, S., A. Porco, T. Isturiz, and T. Conway. 1996. Cloning and molecular genetic characterization of the *Escherichia coli gntR*, *gntK*, and *gntU* genes of GntI, the main system for gluconate metabolism. *J. Bacteriol.* **178:** 3260-3269.

10. Farewell, A., K. Kvint, and T. Nyström. 1998. Negative regulation by RpoS: a case of sigma factor competition. *Mol. Microbiol.* **29:** 1039-1051.
11. Oh, T. J., I. L. Jung, and I. G. Kim. 2001. The *Escherichia coli* SOS gene *sbmC* is regulated by H-NS and RpoS during the SOS induction and stationary growth phase. *Biochem. Biophys. Res. Commun.* **288:** 1052-1058.
12. Hengge-Aronis, R. 1993. Survival of hunger and stress: the role of *rpoS* in early stationary phase gene regulation in *E. coli*. *Cell* **72**, 165-168.
13. Hengge-Aronis, R. 1996. Regulation of gene expression during entry into stationary phase. In *Escherichia coli and Salmonella*: *cellular and molecular biology*, 2nd edition. (eds. Neidhardt, F. C., Curtiss, R., III., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E.), ASM press, Washington, D. C., pp. 1497-1512.
14. Hengge-Aronis, R. 2002. Signal transduction and regulatory mechanisms involved in control of the sigma (S) (RpoS) subunit of RNA polymerase. *Microbiol. Mol. Biol. Rev.* **66:** 373-395.
15. Mukhopadhyay, S., and H. E. Schellhorn. 1997. Identification and characterization of hydrogen peroxide-sensitive mutants of *Escherichia coli*: genes that require OxyR for expression. *J. Bacteriol.* **179:** 330-338.
16. Blattner, F. R., G. Plunkett, III., C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, and G. F. Mayhew *et al.* 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**:1453-1474.