Differential Gene Expression by Two Strains of *Escherichia coli* (K12 and an Environmental Isolate) in Response to Temperature and Nutrient Stress Using Microarrays

by

Charmaine Wetherell
BSc(Hons), Okanagan University College, 2004

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTERS IN SCIENCE

in the Department of Biology

Supervisory Committee

Differential Gene Expression by two strains of *Escherichia coli* (K12 and an Environmental Isolate) in Response to Temperature and Nutrient Stress Using Microarrays

by

Charmaine Wetherell
BSc (Hons), Okanagan University College, 2004

Supervisory Committee

Dr. Real Roy (Department of Biology)
Supervisor

Dr. Ben Koop (Department of Biology)
Departmental Member

Dr. Edward Ishiguro (Department of Biochemistry & Microbiology)
Outside Member

Dr. Christian Chauret (Natural, Info & Math Sciences, Indiana University, USA)
Additional Member

Abstract

Supervisory Committee

Dr. Real Roy (Department of Biology)
Supervisor

Dr. Ben Koop (Department of Biology)
Departmental Member

Dr. Edward Ishiguro (Department of Biochemistry & Microbiology)
Outside Member

Dr. Christian Chauret (Natural, Info & Math Sciences, Indiana University, USA)
Additional Member

In this study we evaluated the use of microarray technology in Bacterial Source Tracking (BST), with the intent of identifying candidate genes to be used to differentiate between closely related strains of *Escherichia coli*. We anticipate that genes differentially expressed in response to stress by both a laboratory strain and environmental isolate could be used as marker genes on a microarray in BST. Using microarrays we characterized the transcriptional response of *E. coli* K12 MG1655 (K12), maintained for about 80 years in an artificial environment versus *E. coli* 43(C)-4A or E43, a strain recently isolated from the natural environment. The responses were to a temperature decrease from 37°C to 21°C, and to growth in a diluted LB broth (dLB). Overall we found that there were more genes differentially expressed between the strains than either strain's response to the stresses. At the 4-fold threshold, at reduced temperature there were only 26 genes differentially regulated by K12 and 9 by E43, respectively. In K12 the functions of some differentially expressed genes were linked to the general stress response and biofilm formation. A few genes differentially expressed by E43 were involved in the stress response. Similarly, in response to dLB there were 46

and 11 genes differentially expressed by K12 and E43 respectively. While it appeared that genes differentially expressed by K12 were involved in dealing with nutrient deficiencies, the genes differentially expressed by E43 did not show a similar pattern. Of these genes, none were obvious candidate genes for a microarray to be used in BST. However, we did find that 169, 286 and 254 genes were differentially expressed between K12 and E43 at 37°C, 21°C, and in dLB, respectively. Many of these genes were differentially expressed under all 3 growth conditions. Several of the genes differentially expressed between the strains were in the O-antigen-lipopolysaccharide gene family and are genes that could potentially be used on a microarray in BST. We found that E43, isolated from the natural environment, did not respond to the growth conditions in the same way as the model strain, *E. coli* K12, indicating that strains of *E. coli* isolated from the natural environment may not be identical to the model strain K12. It is suggested that other strains isolated from the natural environment be investigated. Such studies could also reveal genes differentially expressed between the strains that could be used on a microarray for use in BST.

Table of Contents

List of Abbreviations

AFLP	Amplified Fragment Length Polymorphism.
BSA	Bovine Serum Albumin
BST	Bacterial Source Tracking
dLB	LB diluted by a factor of 10 in water.
d-value	Deviation from the perfect correlation.
DTT	Dithiothreitol.
E43	*Escherichia coli* 43(C)-4A, an isolate of bovine origin recently taken from the natural environment.
IG	Intergenic regions of DNA.
K12	*Escherichia coli* K12 MG1655 (ATCC 700926).
LB	Luria-Bertani.
LPS	Lipopolysaccharide.
MM	Mismatch probe on an Affymetrix microarray, 25bp in length, designed with a single bp mismatch at position 13 from the target sequence, designed to detect the level on non-specific hybridization.
OD	Optical Density.
ORF	Open Reading Frame (bacterial gene set).
PCR-DGGE	Polymerase Chain Reaction-Denaturing Gradient Gel Electrophoresis.
PFGE	Pulsed-Field Gel Electrophoresis.
PM	Perfect Match probe on an Affymetrix microarray, 25bp in length, designed to hybridize to the target sequence.
RE	Restriction Enzyme.
REP-PCR	Repetitive element Polymerase Chain Reaction.
Rpos / sigma 38 / σ^S	The sigma factor known to regulate the shift into stationary phase, and is also associated with biofilm formation.
rpm	Revolutions per minute.
SAPE	Streptavidin Phycoerythrin (a staining reagent).
WRC	Wine Research Centre.

List of Tables

List of Figures

Acknowledgments

First and foremost I want to thank my supervisor, Réal Roy who challenged and guided my ideas, work, and writing at every stage. I am also grateful to Julian Davies, Rick Nordin, Melissa Hills, and Zongli Liu for their considerable assistance. In addition, I would like to gratefully acknowledge the financial contributions of the Canadian Institute for Health Research through a grant to Réal Roy and Asit Mazumder, the Natural Sciences and Engineering Research Council of Canada through a Discovery grant to Réal Roy, and the University of Victoria for a graduate student fellowship.

Dedication

To Herb who supported me throughout this work. And to my Parents who always encourage me.

Chapter 1 General Introduction

Using microarrays to find new genetic markers in *Escherichia coli* for the purpose of advancing Bacterial Source Tracking

1.1 Bacterial Source Tracking

Little is more necessary to human health than a clean water supply, not only for human consumption, but also for watering crops and livestock. Most communities obtain their water from the lakes, streams or reservoirs. Since these resources often span large areas of land, contamination cannot always be prevented. North American communities, and others, have developed methods for detecting contamination in water sources (American Public Health Association 1998, Ram *et al.* 2004). Once the contamination is detected, appropriate action, such as a boil-water advisory, can be taken. Besides pollution of water resources with synthetic chemicals and other substances, faecal contamination of drinking water supplies is of particular concern. Faecal pollution poses a threat to human health because faeces can be a vector for human pathogens such as *Cryptosporidium* spp., *Salmonella* spp., or even *Escherichia coli* O157:H7 (Rompre *et al.* 2002, Scott *et al.* 2002).

It is important to detect faecal contamination but direct screening for all potential pathogens is too difficult and costly (American Public Health Association 1998). Therefore, faecal indicators are used because they are more easily detected and are present in greater numbers than the pathogen. Detection of faecal indicators in water demonstrates the presence of faeces, indirectly indicating that pathogens may also be present (Rompre *et al.* 2002, Scott *et al.* 2002). While there are a few pathogenic strains of *E. coli*, most strains are benign if not beneficial to the mammalian gut (Rompre *et al.* 2002, Scott *et al.* 2002). *E. coli* species are considered reasonable indicators of faecal

contamination because they are present in greater numbers than most faecal pathogens, most species are non-pathogenic, and they are relatively easy to detect (Stender *et al.* 2001, Rompre *et al.* 2002, Scott *et al.* 2002). Although there are other microbial indicators of faecal contamination such as total coliforms, and *Enterococcus*, this introduction will mostly focus on *E. coli*, the most abundant of the coliforms (American Public Health Association 1998, Hagedorn *et al.* 1999, Scott *et al.* 2002).

It is now a standard practice for communities across North America to routinely test natural, drinking and waste-water for the presence of coliforms including *E. coli*. There are standardized methods for detection of faecal (American Public Health Association 1998, Kinzelman *et al.* 2003, Ram *et al.* 2004). However, detection of faecal contamination provides no information on the point source of the faecal pollution. Knowing the source of bacteria would provide valuable information for management of watersheds, reducing both economic loss and the risk to human health (Scott *et al.* 2002, Meays *et al.* 2004). Bacterial Source Tracking (BST) develops scientific methods for the identification of the point source of faecal contamination in the environment.

1.2 General methodology of BST

While detection methods have been developed to the point of becoming standardized, BST is still in its beginning stages as a science (American Public Health Association 1998). Techniques under study in BST can be divided into phenotypic and genotypic methods. The goal of both types of methods is to relate faecal bacteria to host categories. For example, one phenotypic method is antibiotic resistance patterning. The assumption of this method is that different host groups like cattle and wildlife have exposure to different antibiotics (Hagedorn *et al.* 1999, Harwood *et al.* 2000, Karasinski *et al.* 2005,

Vantarakis *et al.* 2006). The differential level of exposure of the hosts to various antibiotics should result in the bacterial populations of different host groups possessing a unique suite of resistance genes resulting in a detectable antibiotic resistance pattern. Bacteria from the environment can be cultured, analyzed, and their antibiotic resistance patterns compared to patterns of bacteria from a known origin (a library). This allows scientists to classify bacteria isolated from the environment into host categories and the source of the contamination to be identified. The method is relatively simple but requires isolation and cultivation of the bacteria being tested. A disadvantage is that antibiotic resistance genes are often carried on plasmids which are not bound to the bacterial genome and can be lost during cultivation (Scott *et al.* 2002). Another issue with this method of classification is that research indicates a larger than expected reservoir of antibiotic resistance genes in soil bacteria (Alonso *et al.* 2001, D'Costa *et al.* 2006, Martinez 2008). The presence of these genes may cause confusion in samples isolated from the environment.

Following the generally increased use of molecular techniques, genetic methods in BST are more numerous than phenotypic ones. Despite their large number, the underlying assumptions of the genetic methods are similar. Because bacteria generally reproduce asexually each daughter cell is a genetic copy of the parent. However, populations of *E. coli* existing in different host groups (such as cattle or wildlife) are subjected to different selection pressures such as diet and internal gut conditions. Combined with the high mutation rate and short generation time of bacteria, this leads to strains that likely have genetic variations between different host groups. The molecular markers that differentiate these strains can be detected by a variety of protocols. The

molecular-genetic branch of BST works on adapting the variety of molecular methods for grouping strains of bacteria into host categories and distinguishing between strains originating from different host groups.

1.3 Molecular techniques in BST

There are many molecular techniques used in BST. We will describe a few of the most popular of these techniques that use *E. coli* as a target. In *ribotyping*, genomic DNA is digested with one or two restriction enzymes (RE), electrophoresed, and bound to a membrane. The membrane is probed with 16S and/or 23S DNA fragments by Southern Blotting to produce a DNA-fingerprint for each bacterial strain (Carson *et al.* 2001, Scott *et al.* 2003). In *repetitive extragenic palindromic-PCR* (REP-PCR), DNA is amplified by PCR using primers based on short repetitive palindrome-elements present in bacterial DNA. Gel electrophoresis is then used to create the DNA fingerprints for each strain (Dombek *et al.* 2000, Johnson and O'Bryan 2000, Baldy-Chudzik *et al.* 2003, Seurinck *et al.* 2003). *Amplified fragment length polymorphism* (AFLP) uses restriction digestion of genomic DNA followed by ligation of adaptors to the sticky-ends of the RE fragments. PCR primers for the adaptors and part of the RE site are used to selectively amplify the fragments by PCR (Arnold *et al.* 1999, Savelkoul *et al.* 1999). The PCR products are electrophoresed and a DNA fingerprint for each strain is produced.

In *PCR-denaturing gradient gel electrophoresis* (PCR-DGGE), 16S DNA is amplified by PCR, followed by an acrylamide gel electrophoresis that has an increasing gradient of DNA denaturants (Boon *et al.* 2002, Seurinck *et al.* 2003). The denaturants melt double-stranded DNA into single-stranded DNA. Once melted, DNA migration on the gel will stop. PCR products of the same length but with different sequences will melt

at different points in the DNA-denaturant gradient, allowing for differentiation of fragments of the same size (Boon *et al.* 2002, Seurinck *et al.* 2003). In *pulsed-field gel electrophoresis* (PFGE), rare cutting RE's are used to digest the whole genome, followed by electrophoreses in alternately pulsed, perpendicular electrical fields (Hahm *et al.* 2003, Stoeckel *et al.* 2004). The pulsed electrical fields help separate large fragments of DNA on a gel to create a DNA fingerprint for each strain of bacteria. There are other molecular techniques used in BST, but all these techniques are used for the same purpose. The goal is to produce a DNA fingerprint unique enough to discriminate between strains originating from different host categories, but similar enough to group together strains from the same host category. Cost, reproducibility, and time to perform the procedure are factors in choosing any procedure that will be used in BST.

While all these methods have advantages each have disadvantages that would make them less than ideal for a standardized BST test. Some *E. coli* methods, such as ribotyping, and REP-PCR have a high rate of correct classification, but are limited to a relatively small geographical area in their application (Gordon 2001, Hartel *et al.* 2002, Scott *et al.* 2003, Meays *et al.* 2004, Soule *et al.* 2006). Also, the DNA fingerprints produced by REP-PCR suffer from issues concerning reproducibility between laboratories. The time to perform electrophoresis in PFGE is extremely long on account of the need to separate large DNA fragments. In general, the genetic fingerprinting methods depend on a large reference or training library of DNA-fingerprints for each potential host group so that DNA-fingerprints of strains from an unknown source may be compared to the library and classified into the correct host-group (Johnson *et al.* 2004, Meays *et al.* 2004). It is mainly the large library size required for the genetic

fingerprinting techniques that prevents their widespread application, since the creation of such libraries is a costly and time-consuming enterprise.

It should be mentioned that not all of the genetic techniques target the familiar faecal indicator organism *E. coli*; some have developed methods for faecal bacteria in the genera *Bacteroides* and *Prevotella* (Bernhard and Field 2000a, Bernhard and Field 2000b, Dick and Field 2004, Dick *et al.* 2005). They argue that these genera are better indicators of a potential human health hazard because they are present in greater numbers than faecal coliforms and are even less likely to survive in the environment than faecal coliforms (Bernhard and Field 2000a, Topp *et al.* 2003). *Bacteroides* and *Prevotella* are fastidious anaerobes which were extremely difficult to detect in faeces prior to the advent of DNA sequencing from environmental samples. For these bacteria, research has focused on isolation of the DNA coding for the small subunit RNA (Wood *et al.* 1998, Bernhard and Field 2000a, Bernhard and Field 2000b).

Since it is highly conserved, sequencing of the small subunit RNA or 16S rRNA has become the standard used for phylogenetic classification of bacteria (Woese 1987, Woese 1996). Species that are closely related will have more similar sequences than species that are more distantly related. In *E. coli* the 16S gene is about 1500 bp in length (Dorsch and Stackebrandt 1992). However, sequencing the 16S rDNA gene of *E. coli* does not reliably distinguish between strains or even between *E. coli* and species of *Salmonella* or *Shigella* (Fukushima *et al.* 2003, Kakinuma *et al.* 2003).

Citing this difficulty with this very popular gene some researchers have found that the *gyrB* gene is a better target for distinguishing between strains of *E. coli* (Fukushima *et al.* 2003, Kakinuma *et al.* 2003). It has the advantage of being a gene specific to *E. coli* and

its close relatives, eliminating the confusion from the presence of the small subunit rRNA gene from non-related bacteria (Fukushima *et al.* 2003, Kakinuma *et al.* 2003, Polissi *et al.* 2003). More studies on the *gyrB* gene need to be done. Furthermore, studies of the genetic distance between species using genes other than the 16S rRNA reveal clades different from the ones found using 16S rRNA (Blankenship 1992, Blankenship and Hartman 1998, Doolittle 1999, Blankenship 2001). These studies suggest that horizontal gene transfer may be of far more importance than previously thought. More research into the bacterial genome and the relationship between and within species will help in the development of better methods for BST.

As the study of BST progresses, more of the assumptions come to light and must be addressed to further the discipline. A recent paper pointed out that there is considerable variability in *E. coli* concentration and sources within a single 24 hour period which means that sampling for BST must be carefully planned so that it addresses the desired question (Meays *et al.* 2006). Another report by Barnes and colleagues (2004) questions whether or not there is a correlation between the coliform bacteria in host organisms deposited by faeces and coliform populations isolated in the environment. They used mathematical models to conclude that relative abundance of bacteria in the environment is not likely linked to the relative abundance of the bacteria from faecal deposits, but their research has not been directly tested. These studies indicate the need for more research and even for development of new methods that can be applied to BST.

Despite the diversity of research in the field of BST, all methods take advantage of differences in the DNA between species to differentiate between strains of faecal bacteria. However, each method only uses only one gene or relies on randomly

distributed patterns in the DNA sequence. It would be advantageous to increase the number of functional genes which could be used to distinguish between bacteria. Microarray technology provides a means of discovering many more strain-specific markers.

1.4 Microarray technology

Microarrays allow scientists to look at the global messenger-RNA (mRNA) profile of bacteria at a particular moment in time or to test for the presence of a large number of genes. Essentially, microarrays are Northern Blots except that the probes, not the target DNA, are fixed on nylon membranes or glass slides (Lucchini *et al.* 2001, Brown 2002). Improvements in printing technology allow for the simultaneous probing of thousands of genes. The probes of microarrays are affixed to a single platform in microscopic droplets and then total mRNA which has been converted to cDNA is labelled and hybridized to the fixed-probes. Once the labelled target-RNA has been hybridized to the chip, the fluorescent intensity of each spot is measured. The relative intensity of all the probes is proportional to the relative abundance of the mRNA in the target-sample (Lucchini *et al.* 2001, Brown 2002). Arrays that use probe for the presence of DNA sequences are also in use.

Many studies on *E. coli* using microarrays have been conducted using strains maintained for many years in an artificial laboratory environment. Most of these studies have focussed on the genetic regulation of a specific process within *E. coli*, often making comparisons between wild-type and mutant strains in order to determine the genetic response to a single factor. Some studies primarily looked at the effect of a particular gene on the global profile of mutant and wild-type strains (Arfin *et al.* 2000, Hung *et al.*

2002, Cooper *et al.* 2003, Li *et al.* 2003, Phadtare and Inouye 2004), while others focussed on how bacteria dynamically respond to nutritional or other environmental changes (Chang *et al.* 2002, Gonzalez *et al.* 2002, Polissi *et al.* 2003, Kao *et al.* 2005, Karlsson *et al.* 2005). Yet others compared the bacterial mRNA profiles under different conditions (Richmond *et al.* 1999, Tao *et al.* 1999, Oh and Liao 2000, Phadtare *et al.* 2002, Cheung *et al.* 2003, Shaw *et al.* 2003, Brokx *et al.* 2004, Ren *et al.* 2004, Gadgil *et al.* 2005, Kershaw *et al.* 2005, Kershaw *et al.* 2007). Finally, some applied microarray technology to differentiation between strains and genetically similar species (Kakinuma *et al.* 2003, Wu *et al.* 2003, Fukiya *et al.* 2004, Kostic *et al.* 2007).

In recent years some studies have focussed on the development of protocols for the use of microarrays in the environment (Cho and Tiedje 2002, Lemarchand *et al.* 2005, Soule *et al.* 2006). And a few have begun to directly apply microarray technology to environmental samples, recognizing the power of microarrays to probe thousands of genes (Sessitsch *et al.* 2006).

Some examples of the use of environmental microarrays follow. One study successfully used microarray technology to distinguish between pathogenic and non-pathogenic strains of *E. coli* (Wu *et al.* 2003). The authors used the sequenced genomes of both the non-pathogenic *E. coli* K12 MG1655 and the pathogenic *E. coli* O157:H7 EDL933 *E. coli* strains to find the strain-specific sequences that would be used on the microarray. It should be noted that DNA was sampled and used on the microarray. Another study evaluated the use of DNA gene arrays as a way to detect different functional groups of bacteria like methanotrophs and nitrogen fixers in the environment without culturing the bacteria (Wu *et al.* 2001). They suggested that microarrays will be

a potential tool for environmental studies. However, the authors also mentioned that PCR amplification is still a more sensitive means of detecting functional genes and that further work to improve the sensitivity and quantitation is needed (Wu *et al.* 2001). Yet another study evaluated the use of microarrays to detect a greater number of waterborne pathogens than can currently be detected by PCR amplification (Maynard *et al.* 2005). The authors of this study found that using microarrays introduced little or no PCR bias to the experiment when testing a sample with multiple strains. However they did find that detection sensitivity varied depending on the strains used. They also suggested that using genes specific to a limited type of bacteria rather than universal genes may be a way to increase sensitivity since they found the detection sensitivity was higher for *wecE*, a gene specific to the Enterobacteriaceae (Maynard *et al.* 2005). It is likely that microarray technology will be widely used in the future.

1.5 Microarrays and the *E. coli* stress response.

When *E. coli* enter the environment they are subjected to a number of environmental changes such as an increase in light, a decrease in temperature and likely a decrease in nutrient availability (Savageau 1983, Bennett *et al.* 1992, Flores *et al.* 2005, Hughes 2008, Vital *et al.* 2008). This means *E. coli*'s ability to respond to stress is essential for its survival in the natural environment (secondary habitat) so that in can re-colonize the guts of mammals (primary habitat). It is generally thought that viable *E. coli* neither live for longer than a few days, nor reproduce in the natural environment (Hughes 2008). For example, despite their regular deposition onto soil *E. coli* are not known to be a normal part of the soil bacterial community. While this theory probably remains generally true for *E. coli* in temperate climates, it is becoming evident that *E. coli* are able to survive

longer and possibly even reproduce in eutrophic tropical freshwater (Savageau 1983, Jimenez *et al.* 1989, Muniz *et al.* 1989, Desmarais *et al.* 2002, Ksoll *et al.* 2007, Hughes 2008). One paper even suggested that the strains isolated from tropical freshwater had become adapted to that environment and were, therefore, a population existing in the natural environment without the need for its primary habitat (Jimenez *et al.* 1989). Furthermore, at least one recent study has revealed that viable *E. coli* can survive in faecal matter under dry hot conditions for at least 45 days (Meays *et al.* 2005).

Since BST deals with *E. coli* in the environment the study of *E. coli*'s response to environmental stresses is especially relevant to BST. More information on how *E. coli* respond to environmental stress will be valuable to researchers so that they can better design methods in both sampling and identification of the point source of *E. coli*.

Despite microarrays being a relatively new technology, the use of microarrays has advanced the scientific knowledge of the *E. coli* genome regulation (Tjaden *et al.* 2002). Of particular interest to BST is microarray research on the *E. coli* stress response. It is well established that sigma factor (σ^S) is a general regulator of *E. coli*'s response to several stresses including starvation, extreme temperature, entry into stationary phase, and hyperosmolarity (Kabir *et al.* 2004, Patten *et al.* 2004, Weber *et al.* 2005). Even though it is a well-studied gene, microarray research is revealing new conditions under which RpoS-regulated genes are expressed as well as new genes previously not known to be under RpoS regulation (Olsen *et al.* 1993, Sugiura *et al.* 2003, Lacour and Landini 2004, Patten *et al.* 2004, Weber *et al.* 2005, Gualdi *et al.* 2007, White-Ziegler *et al.* 2008). Clearly microarray technology has expanded the information available on *E. coli* gene regulation.

Low temperature is a significant stress on *E. coli*, causing them to change their gene expression, along with other physiological changes (Sturgeon *et al.* 1976, Herendeen *et al.* 1979, Jones *et al.* 1987, Etchegaray and Inouye 1999, Mizunoe *et al.* 1999, Li *et al.* 2003, Polissi *et al.* 2003, Gadgil *et al.* 2005, White-Ziegler *et al.* 2008). *E. coli* cells are even grown at decreased temperature to improve the production of heterologous proteins (Gadgil *et al.* 2005). The response to cold-shock (at least as low as 15°C) has been well-studied (Graumann and Marahiel 1996, Etchegaray and Inouye 1999, Phadtare *et al.* 1999, Polissi *et al.* 2003, Gadgil *et al.* 2005), but a couple of other studies indicate that even small temperature changes (21-37°C) may affect gene expression in *E. coli* (Gadgil *et al.* 2005, White-Ziegler *et al.* 2008).

In the studies on the response of *E. coli* strains to low temperature, researchers have found that strains of *E. coli* maintained in a laboratory environment differentially expressed 9 % to 41 % of genes in different experiments (Li *et al.* 2003, Gadgil *et al.* 2005, White-Ziegler *et al.* 2008). In a different set of experiments on the adaptation of *E. coli*, Cooper and colleagues (2003) found 59 genes differentially expressed between strains grown for twenty thousand generations in glucose limiting environment and the ancestral strain, clearly showing adaptation of *E. coli* strains. In related experiments, it was demonstrated that *E. coli* also adapt by becoming temperature generalists and temperature specialists (Bennett *et al.* 1992, Leroi *et al.* 1994, Cooper *et al.* 2001). Clearly temperature is an important factor for *E. coli* and these experiments demonstrate not only physiological adaptation to its environment, but also genetic drift in *E. coli* over time in a selective environment. Since temperature adaptation is so important, genes

involved in adaptation to temperature stress would make good marker genes for a DNA microarray to be used in BST.

The other stress condition under examination in our research is the dilution of a complex culture medium. While various individual nutrient deficiencies have been studied in *E. coli*, there are few studies which attempt to approximate the natural soil and water environments (Savageau 1983, Fani *et al.* 1998, Mizunoe *et al.* 1999, Hughes *et al.* 2000, Flores *et al.* 2005, Maharjan *et al.* 2005). Because natural soil and water in North America are typically oligotrophic environments with a diverse set of nutrients available, *E. coli* typically encounters a complex set of nutrients at low concentrations when in the natural environment (secondary habitat). Therefore we chose to use a diluted complex culture medium (Luria-Bertani) rather than a defined culture medium to emulate the conditions *E. coli* must survive in its secondary habitat. In this study we used LB medium diluted 10 times in water (dLB) to approximate natural water habitat.

1.6 Research hypothesis and objectives

To our knowledge a direct comparison between a well-studied strain maintained in an artificial environment and an isolate recently retrieved from the natural environment has not yet been conducted using microarrays. While our ultimate goal is to identify candidate genes that could be used in the future to differentiate between strains of *E. coli* on a microarray used in BST, the direct comparison between these two strains will also provide valuable information about differences between strains maintained in artificial conditions and an isolate recently retrieved from a natural environment.

We chose to study both a lab strain of *E. coli* K12 MG1655 (hereafter referred to as K12), and a recent bovine environmental isolate called 43(C)-4A (hereafter referred to as

E43). We selected two phenotypically different strains so we could evaluate which genes were most essential for *E. coli*'s response to the environment. To identify candidate marker genes for a BST-microarray, we wanted to find genes that were highly differentially regulated by both K12 and E43 in response to both stresses (decreased temperature and dilute culture medium). We hypothesize that these genes will be the most essential for *E. coli* to respond to stress. In our experiment we will also compare gene expression between the two strains under all three conditions: 37°C in Luiria-Bertani (LB) broth, 21°C in LB broth and in dilute LB medium (dLB) at 37°C. We expect that there will be relatively few genes differentially expressed between the closely related strains of *E. coli* used in this experiment. However, any genes that are differentially expressed between the strains under all conditions may also be considered candidate genes for a DNA microarray.

Our specific objectives were as follows:

1. Is gene expression of *E. coli* K12 different when grown at 21°C rather than 37°C?
    - **a)** How many genes are differentially expressed?
    - **b)** Which functions are most affected?
2. Is the differential gene expression of 43(C)-4A (E43) between 21°C and 37°C similar to the expression profile of K12?
    - **a)** Are the number of genes differentially expressed similar to 1.a?
    - **b)** Are the functions affected the same as 1.b?
3. Is the gene expression of *E. coli* K12 different when grown in diluted LB broth (dLB) as compared to growth in L B broth?

a) Are the number of genes differentially expressed similar to 1.a?

b) Are the functions affected the same as 1.b?

4. Is the differential gene expression profile of E43 between dLB and LB similar to the expression profile of K12?

a) Are the number of genes differentially expressed similar to other comparable growth conditions (1 and 3)?

b) Are the functions affected the same as other comparable conditions (1 and 3)?

5. Are there genes differentially expressed between the K12 and E43?

a) How many genes are differentially expressed?

b) Are these genes differentially expressed under all growth conditions?

c) What functions are most affected?

This thesis has three main chapters, each looking at a specific set of questions. The first chapter is a general introduction to BST and microarrays and presents our research goals. The second chapter presents the general results of questions 1a, 2a 3a, 4a, and 5a. The third chapter presents the results of questions 1 and 2 in greater detail as well as addressing part of question 5. In the fourth chapter, the results of questions 3, 4, and 5 are discussed. The last chapter is a general conclusion in which we discuss the general significance of the differences in gene expression of the two *E. coli* strains and how these results could be useful in the context of BST studies.

Chapter 2

Overview of the global gene expression of *Escherichia coli* (K12 and 43(C)-4A) to low temperature and to dilute culture medium.

2.1 Abstract

In the field of Bacterial Source Tracking (BST) improved and new molecular techniques are needed to determine the point source of faecal contamination. In this report we evaluated the use of microarray technology in BST by determining how many genes are differentially regulated by two strains of *E. coli* in response to two different stresses. We compared the transcriptional response of *Escherichia coli*, 43(C)-4A also designated E43, recently isolated from the natural environment to both decreased temperature (21°C) and a decrease in nutrients with the responses of *E. coli* K12 MG1655 (K12), a strain maintained in an artificial environment. We found that both K12 and E43 differentially expressed few genes in response to 21°C (d-values 0.038 and 0.025 respectively) as well as dLB (d-values 0.052 and 0.041 respectively). By contrast, the d-values increased to 0.193, 0.232, and 0.200 at 37°C, 21°C and in dLB, respectively, when comparing gene expression between the strains. Therefore, we found more differences in gene expression between the strains than in either K12's or E43's response to the environmental stresses. We propose that the genes differentially expressed by both strains to both environmental stresses may be potential marker genes to be used on a microarray in BST. The genes differentially expressed between the strains may also be used for this purpose.

2.2 Introduction

In its history, North America has enjoyed an abundant supply of safe drinking water but there are ever increasing and complex demands on water sources all over North America (Bernhard and Field 2000a, Bernhard and Field 2000b, Dombek *et al.* 2000, Sauer *et al.* 2000, Albert *et al.* 2003, Jenkins *et al.* 2003). Faecal contamination of drinking water sources is of special concern because of pathogens that can be transmitted by faecal waste (Jenkins *et al.* 2003, Meays *et al.* 2005). Coliforms, including *Escherichia coli* are commonly used to detect faecal contamination in water because they are present in higher numbers than the pathogens they indicate, 2) relatively easy to detect by standard microbiological methods, 3) and not typically pathogenic (American Public Health Association 1998, Scott *et al.* 2002, Meays *et al.* 2004).

When contamination with a pathogen does occur, it is difficult to determine the point source so that the contamination can be eliminated. Although there are current standard methods for analyzing and detecting faecal indicators in water, there are no standardized laboratory methods that can determine the point source of the faecal contamination (American Public Health Association 1998, Rompre *et al.* 2002, Albert *et al.* 2003). Failure to identify sources of contamination can lead to continued contamination and subsequent risks to human health (Bernhard and Field 2000b). Such concern provides the incentive for studies into Bacterial Source Tracking (BST) and to develop new methods for determining the of point source faecal contamination in the environment (Dombek *et al.* 2000, Carson *et al.* 2001, Boon *et al.* 2002, Hartel *et al.* 2002, Dick and Field 2004, Johnson *et al.* 2004, Ram *et al.* 2004, Stoeckel *et al.* 2004). Generally, these methods group *E. coli* or other indicator bacteria into host categories by genetic or phenotypic markers. An environmental isolate of unknown origin is categorized by comparing its

markers to the library of markers from known host sources. Many methods, genetic and phenotypic, have been developed but no one method has been demonstrated to be superior to all the others (Scott *et al.* 2002, Meays *et al.* 2004).

While many molecular methods have been applied to BST with varying levels of success, they often rely on a single gene or short repeated sequences for differentiation and classification of *E. coli* genotypes into host-organism groups (Bernhard and Field 2000a, Carson *et al.* 2001, Seurinck *et al.* 2003). One method that uses non-coding regions of the genomic DNA is called *repetitive extragenic palindromic* -PCR (REP-PCR) (Dombek *et al.* 2000, Johnson and O'Bryan 2000, Baldy-Chudzik *et al.* 2003, Seurinck *et al.* 2003). REP-PCR relies on the differential dispersion between *E. coli* strains of short DNA palindromes. The regions between these palindromes are amplified by PCR and gel electrophoresis reveals the unique DNA fingerprints of different strains (Dombek *et al.* 2000, Johnson and O'Bryan 2000, Baldy-Chudzik *et al.* 2003, Seurinck *et al.* 2003). While REP-PCR does cover the whole genome, it provides no sequence information and has low reproducibility between laboratories (Albert *et al.* 2003, Baldy-Chudzik *et al.* 2003, Meays *et al.* 2004). In fact, each method currently under study in BST has advantages and disadvantages (Scott *et al.* 2002, Meays *et al.* 2004). BST would benefit from an increased number of genetic markers that could be used to classify bacteria into host groups.

Microarrays provide a means to determine the response of a particular organism to a specific stimulus on a whole-genome scale. Microarray technology has already been applied to artificially maintained strains of *E. coli* (Arfin *et al.* 2000, Gonzalez *et al.* 2002, Cooper *et al.* 2003, Kao *et al.* 2005, Karlsson *et al.* 2005) but less for recently

isolated strains. Since microarrays are a relatively new technology many studies have focused on a single lab strain of *E. coli*, often strain K12, or done wild type-mutant experiments (Chuang *et al.* 1993, Richmond, *et al.* 1999, Arfin *et al.* 2000, Cheung *et al.* 2003, Cooper *et al.* 2003, Brokx *et al.* 2004, Friedman *et al.* 2006). Other reports have taken another approach of applying microarray technology directly to DNA or RNA isolated from environmental samples (Wu *et al.* 2001, 2006, Zhou and Thompson 2002, Letowski *et al.* 2003, Zhou 2003, Ikeda *et al.* 2005, Sessitsch *et al.* 2006). To the best of our knowledge, no one has published a report comparing microarray data from an artificially maintained strain of *E. coli* and a strain recently isolated from the natural environment. Because microarrays can contain thousands of probes, they can be used to identify new genes previously not associated with a cellular process like temperature stress adaptation (Arfin *et al.* 2000, Polissi *et al.* 2003, Gadgil *et al.* 2005). Thus, using microarrays may not only identify new genetic markers of *E. coli* but also may provide a platform for BST in the future since it allows probing of so many different genetic regions at once.

In this report we evaluated an application of microarray technology by comparing *E. coli* K12 MG1655 to a recent isolate taken from the natural environment. We present data on how many genes were differentially expresses by both strains in response to the environmental stresses of low temperature and low nutrient concentration. We also compared the strains by determining the number of genes differentially transcribed between the strains in the same environment. It was expected that both strains would respond in a similar manner to both stresses. We also expected that few genes will be differentially expressed between strains (Porwollik *et al.* 2003).

2.3 Materials and methods

2.3.1 Isolation of strains and culture techniques

E. coli K12 MG1655 (ATCC 700926) was obtained from Dr. C. Chauret (University of Indiana, Kokomo). The environmental isolate originally from bovine faeces, 43(C)-4A, as well as several other strains were obtained from our collaborators (Drs. A. Mazumder and C. Meays, University of Victoria). The samples were originally taken from Duteau Creek Headwater, near Vernon, B.C., Canada, on June 30, 2004 by Agriculture Canada (Meays *et al.* 2005, Meays *et al.* 2006). The faecal samples of bovine origin were plated directly on McConkey agar within 24 hours of collection (Becton Dickson, Oakville, ON, Canada). Yellow, individual colonies were then isolated from MacConkey agar. Individual colonies were streaked on Luria-Bertani (LB) agar (Becton-Dickson, Oakville, ON, Canada).

To ensure purity, cultures were isolated twice by taking a single colony, streak-plating it on LB agar and growing it at 37°C. After streak-plating, the colonies were grown in LB broth overnight at 37°C and then brought to a final concentration of 20 % glycerol for storage at -80°C (Gerhardt 1994). All successive culture growth was done in regular LB broth or in dilute LB broth (dLB) where LB medium was diluted 10 times in sterile water. For experimental growth, a fresh aliquot of bacteria was always taken from the frozen stock culture. Cultures were grown with agitation (250 rpm) at 37°C or 21°C.

2.3.2 Selection of strains

Growth curves were measured by optical density (OD) at 600 nm on a UV/Vis spectrophotometer (Beckman/Coulter, Fullerton, CA, USA). Samples were taken every 30 or 60 minutes. The environmental isolate, *E. coli* Is. 43(C)-4A, hereafter E43, which

grew faster than K12 at 21°C and may grow slightly faster in dLB was selected for further investigation. To compare maximum growth rates, a standard curve of viable cells versus OD_{600} was created (data not shown). The equation from linear regression of viable cells against OD_{600} was used to calculate cell concentrations.

2.3.3 DNA extraction

DNA was extracted by using a modified method originally intended to isolate bacterial DNA from soil (Berthelet *et al.* 1996). Briefly, 500 µL of *E. coli* culture was combined in a total volume of 1550 µL containing 45.2 mM NaH_2PO_4 (pH8), 22.6 mM NaCl, 113 mM Tris-HCl (pH8), and 2.25 % SDS (w/v). The bacterial solution was mixed with 2.4 g of 0.1 mm Zirconia beads and agitated for 5 min. at the maximum speed (30 s^{-1}) on a Mini-Beadbeater (Retsch Biotech, Rheinische, Germany). The cell debris was removed by centrifugation at 11,500 rpm for 3 min on a microcentrifuge. The supernatant was collected and 0.4 volume of 7.5 M Ammonium acetate was added. The precipitate was removed by centrifugation. Only DNA with an OD_{260}/OD_{280} ratio greater than 1.6 was used and DNA was stored at -20°C for future use.

2.3.4 Confirmation of culture identity

Cultures were confirmed to be *E. coli* by amplification and dideoxy sequencing of the 16S rDNA sequence. F1 and R13 primers designed by Dorsch and Stackebrandt (1992) were used to amplify the *E. coli* DNA using a polymerase chain reaction (PCR) with the following protocol. A PCR mix from Sigma (St. Louise, MO, USA) was used to amplify the DNA. In a total volume of 50 µL, the PCR reaction contained 1X PCR mix (1.5 U Taq Polymerase, 10 mM Tris-HCl, 50 mM KCl, and 0.2 nM dNTP's); 2.5 mM $MgCl_2$, 0.2 µM of both F1 and R13 primers (Alpha DNA, Montreal, QC, Canada), and 100 ng of

template DNA. Reactions were amplified in a GeneAmp ® PCR System 2700 thermal cycler (Applied Biosystems, Streetsville, ON, Canada) using and initial denaturation step of 94°C for 1 min, followed by 30 cycles of denaturation at 94°C for 1min, annealing at 57.5°C for 1 min and extension at 72°C for 1min; the reaction was finished with a final extension step of 72°C for 3 min and reactions were held at 4°C. Quality and quantity of the 16S rDNA was confirmed by measuring the OD_{260}/OD_{280} ratio on a UV/Vis Spectrophotometer (Beckman-Coulter, Fullerton, CA, USA) and gel electrophoresis on a 1% (w/v) agarose (EMD Chemicals, Gibbstown, NJ, USA) gel run in 1X TAE buffer for 1 hour at 100V. Amplified DNA was stored at -20°C for future use.

The 1500bp sequences were purified using the Qia-quick PCR clean-up kit (Qiagen, Mississauga, ON, Canada) following the manufacture's instructions. The purified 16S rDNA was cloned into the pGEM-T plasmid using the pGEM kit (Promega, Madison, WI, USA). Briefly, the template DNA was ligated into the cut plasmid using a 1:1 insert DNA to plasmid ratio and a final concentration of 0.3 Weiss units/µL of T4 DNA ligase. Ligation reactions were incubated overnight at 4°C. *E. coli* JM109 cells were heat-shocked for 45-50 s at 42°C in the presence of the ligated plasmid. Cells were recovered in 950 µL of SOC medium (20 mg mL^{-1} Bacto-Tryptone, 5 mg mL^{-1} Bacto-Yeast extract, 10 mM NaCl, 2.5 mM KCl, 20 mM Mg^{2+}, and 20 mM glucose) at 37°C for 1.5 hrs. Transformed cells were plated on LB/ampicillin/IPTG/X-Gal plates and incubated overnight at 37°C. Transformed (white) *E. coli* JM109 colonies with the insert were used for DNA sequencing reactions.

Plasmids were isolated using the Qiagen plasmid purification kit (Qiagen, Mississauga, ON, Canada) according to manufacturer's instructions. Presence of the

desired insert was confirmed with a RE double digest using *PstI* and *NcoII*. Isolated plasmids were sent to the Centre for Biomedical Research (University of Victoria, Victoria, BC, Canada) for dideoxy sequencing. Sequences were analyzed by the NCBI BLAST program (http://www.ncbi.nlm.nih.gov/).

2.3.5 Small subunit RNA comparison

Since 16S rRNA sequencing is currently the standard method by which scientists determine the relatedness of bacteria, we sequenced the small subunit RNA firstly to confirm that the strains we were working with were *E. coli* and secondly to determine the similarity of our strains (Woese 1987, Woese 1996). Sequences were compared to the sequence submitted by Blattner and colleagues (Blattner *et al.* 1997) (NCBI Ascension #: U00096). Our lab strain of *E. coli* K12 MG1655 was 99.4% similar to the Blattner sequence. The environmental lab strain was 99.2% similar to our strain of *E. coli* K12 (Appendix A).

2.3.6 RNA extraction

RNA was extracted using a RiboPure^{TM}-Bacteria RNA extraction kit (Ambion, Austin, TX, USA) according to the manufacturer's instructions. All labware and surfaces were cleaned with RNaseZap (Ambion, Austin, TX, USA) to prevent degradation of RNA by RNAses. Briefly, cultures were harvested when they reached an OD_{600} of 0.85-1.0 or 0.4-0.45 in dLB (early stationary phase). Cell pellets were combined with 350 µL of RNA-wiz (containing phenol). The cells were lysed by mixing resulting solutions with the appropriate amount of ice-cold Zirconia beads (see *Ambion`s RiboPure^{TM}-Bacteria* Instruction Manual) in a Mini-Beadbeater (Retsch Biotech, Rheinische, Germany). RNA was separated from the other cell debris by the addition of chloroform and centrifugation

to separate the organic (mostly chloroform) phase and aqueous phase, containing RNA. RNA was further purified by the addition of 0.5 volumes of pure ethanol to the aqueous phase and binding of the RNA to an affinity column provided by the kit. The column was washed and RNA was eluted in 50 µL of a 10 mM Tris-Cl solution pre-heated to 91°C. RNA extract was stored at -80°C for future use.

RNA quantity and quality were checked by measuring the OD_{260}/OD_{280} ratio on a UV/Vis spectrophotometer and agarose (EMD Chemicals, Gibbstown, NJ, USA) gel electrophoresis. Only RNA with a ratio of 1.8-2.1 was used in subsequent reactions. RNA was loaded on 1 % (w/v) agarose gels, containing 1 X SYBR®Safe DNA stain (Invitrogen, Burlington, ON, Canada) and run for 2 hours at 50 V in 1 x TAE buffer. Images were taken on a gel visualization system (Alpha Imager, Alpha Innotech Corporation, San Leandro, CA, USA). Only high quality RNA, clearly showing the 16S, 23S ribosomal bands, and the messenger RNA smear was used in subsequent reactions.

2.3.7 Preparation of cDNA from total RNA

Preparation of biotin-labelled cDNA for hybridization to the microarray was as in Brokx and coworkers (2004) and followed the recommended procedure of the microarray manufacturer (Revision 5, Affymetrix, Santa Clara, CA, USA). Amplification of mRNA into cDNA was performed on 10 µg of total RNA (Gadgil *et al.* 2005). All incubation periods were performed in a Px2 thermal cycler (Thermo-Hybaid, Ashford, UK).

Primers were bound to the RNA template before synthesis in a solution of 25 ng/µL of Random primers (Invitrogen, Burlington, ON, Canada) and 2 µL of GeneChip® Eukaryotic Poly-A RNA controls (0.256 pM *lys*, 0.511 pM *phe*, 1.022 pM *thr*, and 3.833 pM *dap*, Affymetrix, Santa Clara, CA, USA). The primer solution was incubated on

control-tube mode at 70°C for 10 min, 25°C for 10 min; the solutions were chilled to 4°C.

The cDNA synthesis solution contained 1 X 1st Strand Buffer (Invitrogen, Burlington, ON, Canada), 10 mM DTT (Invitrogen, Burlington, ON, Canada), 0.5 mM dNTPs (Invitrogen, Burlington, ON, Canada), 0.5 U/µL SUPERase-In (Ambion, Austin, TX, USA), 25 U/µL SuperScirpt II™ (Invitrogen, Burlington, ON, Canada) and all 30 µL of the primer solution. This reaction mix was incubated on the control tube mode at 25°C for 10 min, 37°C for 60 min, 42°C for 60 min, 70°C for 10 min and cooled to 4°C. Any remaining RNA was removed by adding 20 µL of 1 N NaOH to the reaction mix, which was incubated again for 30 min at 65°C. This was followed by addition of 20 µL of 1 N HCl.

The cDNA was cleaned up using the MinElute PCR Purification column following the procedures provided by the manufacturer, except that the product was eluted in 12 µL of EB buffer (Qiagen, Mississauga, ON, Canada). The amount of cDNA produced was quantified by spectrophotometry on the UV-Vis spectrophotometer and only reactions with yields greater than 2µg of cDNA were used in subsequent procedures.

The cDNA was fragmented in a 20 µL reaction volume containing the following: 6 U of DNase I (Amersham Biosciences, Piscataway, NJ, USA) per µg of cDNA and 1 X concentration of One-Phor-All Buffer (Amersham Biosciences, Pscataway, NJ, USA). The reaction was incubated at 37°C for 10 min followed by 98°C for 10 min.

The cDNA was labelled using biotin, the GeneChip® DNA labelling reagent (Affymetrix, Santa Clara, CA, USA). The fragmented cDNA was incubated for 60 min at 37°C in a 50 µL solution containing 1 X reaction buffer, 0.3 mM biotin, and 1.2 Units

terminal deoxynucleotidyl transferase (Promega, Madison, WI, USA). The reaction was terminated by adding 2 μL of 0.5 M EDTA (Invitrogen, Burlington, ON, Canada).

To asses the success and completeness of the labelling reaction a gel-shift-assay was performed as directed by the recommended procedures (Affymetrix, Santa Clara, CA, USA). NeutrAvidin (Pierce Biotechnology, Rockford, IL, USA) is a heavy protein that binds to biotin. Biotin labelled RNA that is incubated in NeutrAvidin will migrate more slowly through a gel matrix than RNA that is unsuccessfully labelled with biotin or labelled RNA not incubated in NeutrAvidin. Successful labelling reactions will show a significant up-shift (indicating slow migration) in a gel electrophoresis if first incubated in NeutrAvidin as compared to a biotin-labelled sample not incubated in NeutrAvidin. Thus, the success of the labelling reaction was assessed by running a gel of two 200 ng aliquots of cDNA from the labelling reaction on a 2% agarose gel (EMD Chemicals, Gibbstown, NJ, USA). One of the aliquots was first incubated for 5 min with 5 μL of 2 mg/mL NeutrAvidin. Gels were electrophoresed for 10 min at 70 V followed by 50 min at 95 V. Only cDNA showing that the majority of cDNA was up-shifted in the aliquot with NeutrAvidin was used in subsequent experiments. The gels were stained with SYBR®Gold (Molecular Probes, Eugene, OR, USA).

2.3.8 Hybridization to the array

Biotinylated cDNA was sent to the WRC (UBC, Vancouver, Canada), where, the cDNA was hybridized to a GeneChip® *E. coli* Antisense genome array (Affymetrix, Santa Clara, CA, USA) by the Affymetrix recommended protocol for Prokaryotic arrays (Rosenow *et al.* 2001, Li *et al.* 2003, Karlsson *et al.* 2005). The array was hybridized for 16 hrs at 45°C and 60 rpm with a mixture containing 1X hybridization buffer, 50 pM B2

Control Oligo, 0.1 mg/mL Herring Sperm DNA, 0.5 mg/mL BSA, 7.8% DMSO, 0.5-7.0 μg labelled cDNA. The chips were washed with 200 μL of non-stringent wash buffer A (6 X SSPE (0.9 M NaCl, 60 mM NaH_2PO_4, 6 mM EDTA) , 0.01 % Tween-20) for 10 cycles of 2 mixes/cycle at 25°C, followed by 4 cycles of 15 mixes/cycle at 45°C with wash buffer B (100 mM MES, 0.1 M [NA^+], and 0.01 % Tween-20). The chips were stained with Streptavidin solution mix (1 X stain buffer, 2 mg/mL BSA, 10 μg/mL Streptavidin) for 600 s at 25°C and washed with 10 cycles of 4 mixes/cycle at 30°C with wash buffer A. The chip was stained with antibody solution mix (1 X MES stain buffer, 2 mg/mL BSA, 0.1 mg/mL Normal Goat IgG, and 5 μg/mL Anti-streptavidin Antibody, biotinylated) for 600 s at 25°C and SAPE solution mix (1 X MES stain buffer, 2 mg/mL BSA, 10 μg/mL Streptavidin Phycoerythrin) for 600 s at 25°C, successively. Finally, the chips were washed with 15 cycles of 4 mixes/cycle with wash buffer A at 30°C. Arrays were scanned using Affymetrix Microarray Suite 5.0 on an Agilent GeneArray® Scanner (Affymetrix, Santa Clara, CA, USA).

2.3.9 Data analysis

Gene expression, and comparison analysis of array data was performed on the Affymetrix Microarray Suite 5.0 (Affymetrix, Santa Clara, CA, USA) according to the "GeneChip® Expression Analysis Technical Manual" and as described in the GeneChip® "Expression Analysis Data Analysis Fundamentals" (Rosenow *et al.* 2001, Affymetrix 2002, Li *et al.* 2003, Karlsson *et al.* 2005). This portion of the data analysis was done at the WRC (UBC, Vancouver, BC, Canada). The analysis is briefly described below.

The gene chip contains probes 7312 ORF's and IG regions (Affymetrix, Santa Clara, CA, USA). The annotation file for the GeneChip® array used in this report can be found on the Affymetrix website on the Technical Documentation page under "Ecoli_ASv2 Annotations" (http://www.affymetrix.com/support/technical/annotationfilesmain.affx). Each ORF and IG is represented by 15 perfect match (PM) probes, about 25 bp in length. Each PM probe is accompanied by a mismatch probe (MM) that has a single base pair mismatch at position 13. The MM probes function as a measure of the non-specific binding and are used to calculate absolute intensity values as well as the detection and change calls (Li *et al.* 2003, Karlsson *et al.* 2005).

To determine if a gene was expressed or not both a *detection algorithm* and *signal algorithm* were used. The former generated a *detection call* for each ORF and the latter produced a value for the magnitude of each ORF's signal. Each probe of an ORF contributed to the final output of both the detection and signal algorithms.

The *detection algorithm* used the non-parametirc One-Sided Wlicoxon-signed rank test (Sokal and Rohlf 1981) to compare the discrimination score {(PM-MM)/(PM+MM)} of each probe to a small positive value and generate p-value for the ORF which determined whether the ORF was "present", "absent", or "marginal". This calculation used the stray signal value (MM) to determine the statistical significance of the ORF expression.

The magnitude of expression was calculated by a separate calculation using a one-step Tukey's biweight estimate. Once again the MM signal was used to eliminate the stray signal and was subtracted from each probe's signal before it was converted to the log value. The closer a signal was to the median, the more weight it was given before a

mean signal value was calculated for the ORF. The means were converted back to the linear scale for the final signal value.

Our experiment mostly involved comparisons between growth conditions or between strains. For comparisons between arrays, experimental arrays were always compared to control condition which was at 37°C in LB broth. When comparing gene expression between strains K12 was the baseline condition. All comparison analyses was done according to the

GeneChip® "Expression Analysis, Data Analysis Fundamentals" manual, (Affymetrix 2002). Before comparisons were done, experimental arrays were normalized to the control array and arrays were scaled to each other. Also, probes were normalized for probe-specific factors. Similarly to the expression analysis a *change call* as well as a *signal algorithm* were calculated for each ORF. The magnitude of differential expression values were obtained by averaging (one-step Tukey's biweight estimate (Sokal and Rohlf 1981)) comparisons between probes on the arrays and output in the natural log ($\log_2$) format. The natural logarithm format allows for simple interpretations since a $\log_2$ of 1 is a 2-fold change. The comparisons were done on a probe-to-probe basis. Instead of averaging the replicates and then comparing the experimental to the control, each experimental array was compared to each control array for a total of 16 comparisons. The *change calls* of "Increase", "Decrease", "Marginal" and "No Changes" were calculated by using a one-sided Wilcoxon-signed rank test in a similar manner to the *detection call* calculation.

Only genes that were given an "Increase" or "Decrease" call in all 16 comparisons were labelled differentially expressed and used in downstream analysis. Since detection

of genes and calculating the differential gene expression are done by independent calculations, any ORF with obvious contradictions between these calculations were eliminated in subsequent analysis. For example, it does not make sense for an ORF to be differentially expressed if the gene was not detected on either of the arrays being compared.

We used four biological replicates for each experimental array as well as for the controls. Data were exported into Microsoft Excel spreadsheets and all future analysis was done in Microsoft Excel 2003 (Microsoft, Mississauga, ON, Canada). For this study the 16 comparisons were averaged and all reported standard deviations are calculated on the basis of the 16 comparisons. Analysis of general trends was done on the complete data set. However IG regions were eliminated from the analysis of gene functions because the function of the majority of IG regions is unknown at this time. Also we chose a cut-off of $\log_2 2$ (4-fold) for the sake of simplifying the analysis to be sure that all genes considered were highly differentially regulated. This is higher than the two-fold differential expression that has been the cut-off in other papers (Arfin *et al.* 2000, Chang *et al.* 2002, Brokx *et al.* 2004). We also considered that highly differentially expressed genes are more likely to be important for the stress responses genes that we were looking for. Gene functions were found on the *E. coli* genome project BLAST and Annotation tools ASAP website (Glasner *et al.* 2003) (http://www.genome.wisc.edu/tools/asap.htm). The searchable database contains detailed annotation data on the *E. coli* genome and is updated as more research on the gene functions is conducted.

The scatter plot analysis was done according to the method used by Cooper and colleagues (2003) using the values for gene expression (is the gene expressed?) but not

the comparison values (is the gene differentially expressed?). The gene expression values were log transformed and averaged. These values were plotted to compare different arrays both within and between strains. D-values represent the deviation from the perfect correlation giving a measure of the divergence in expression between conditions or strains (Cooper *et al.* 2003). The d-values were calculated by subtracting the correlation from 1. Correlations were calculated in Microsoft Excel 2003.

2.4 Results

2.4.1 Culture selection and growth

E. coli K12 was selected because its genome has been sequenced and it had been well-studied on microarrays (Blattner *et al.* 1997). Several different bovine isolates sampled from the environment were grown at low temperature as well as strain K12. A strain which appeared to grow better under temperature stress than K12, which had a marked decrease in growth rate at 21°C, was selected (Fig. 2.1a). *E. coli* K12's maximum growth at 21°C was $1.7x10^8$ cells/mL/hr, while E43's maximum growth rate was $2.4x10^8$ cells/mL/hr at 21°C (data not shown). Interestingly *E. coli* Is. E43 also had a higher maximum growth rate at 37°C than K12, since K12's rate was $2.4x10^8$ cells/mL/hr while E43's was $7.0x10^8$ cells/mL/hr. Not only did E43 appear to recover from temperature stress more rapidly than K12, but it also grew at a faster rate during log phase in dilute medium (Fig. 2.1b and c**).** In LB, K12 grew to a maximum OD_{600} of 1.059, while E43 grew to 1.314 (Fig. 2.1a). By comparison, when both strains were grown in dLB, K12's maximum OD_{600} was 0.441 and E43's was 0.409 (Fig. 2.1.c). Of note is the comparatively rapid growth of strain E43 in dLB (Fig. 2.1c). In LB broth

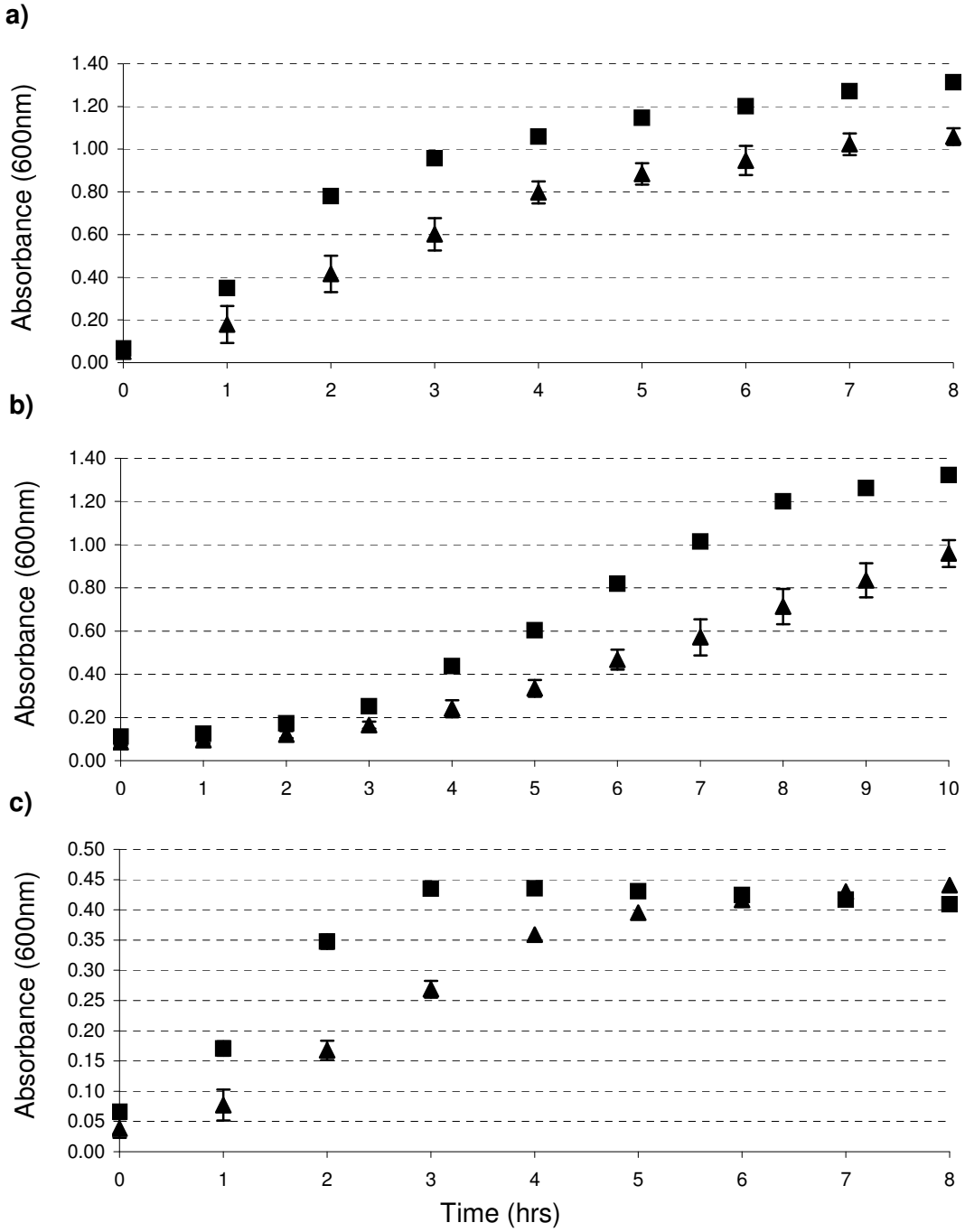

Figure 2.1 Growth of *Escherichia coli* K12 (triangles) and E43 (squares), an isolate from the natural environment, in batch culture at 250rpm at **a)** 37°C in LB broth, **b)** 21°C in LB broth, and **c**) 37°C in dLB broth. Error bars represent the standard deviation of 3 replicates**.**

E43's maximum rate was $1.4x10^8$ cells/mL/hr which was higher than K12's at only $1.0x10^8$ cells/mL/hr.

Because of the differing temperature and medium conditions, the strains were at different growth stage at any particular time point for sampling. Therefore, we chose to extract RNA when cells reached late-log to early-stationary phase (OD_{600} of 0.9-1.0) to ensure the same type of gene expression by both strains under all test conditions (Fig. 2.1b). Late-log phase was chosen so that growth conditions would more closely approximate the natural environment, since bacteria are rarely in logarithmic phase in the natural environment. However, when grown in dLB broth neither strain reached an OD_{600} of 0.90 (Fig. 2.1c). So when cells were grown in dLB harvesting was at an OD_{600} of about 0.45 (Fig. 2.1c). Twice as much culture was used in the RNA extraction to ensure enough total RNA was isolated. With this arrangement we were able to observe the differential gene expression of both strains in response to two different environmental growth conditions when compared to the control condition at 37°C in LB broth.

2.4.2 Global comparison

The overall pattern of differential gene expression is summarized in Fig. 2.2. The plots compare log transformed expression levels for each gene based on four replicate arrays. Note that each graph represents for each gene the log transformed expression levels of cells grown under optimal and stress conditions. The negative values represent the cases where the expression level was very low. As expected, most of the genes and IG regions are not differentially expressed by *E. coli* K12 between 21°C and 37°C as indicated by the linear relationship (Fig. 2.2a). However, the outliers from this linear pattern show that some genes are differentially expressed as indicated by the deviation

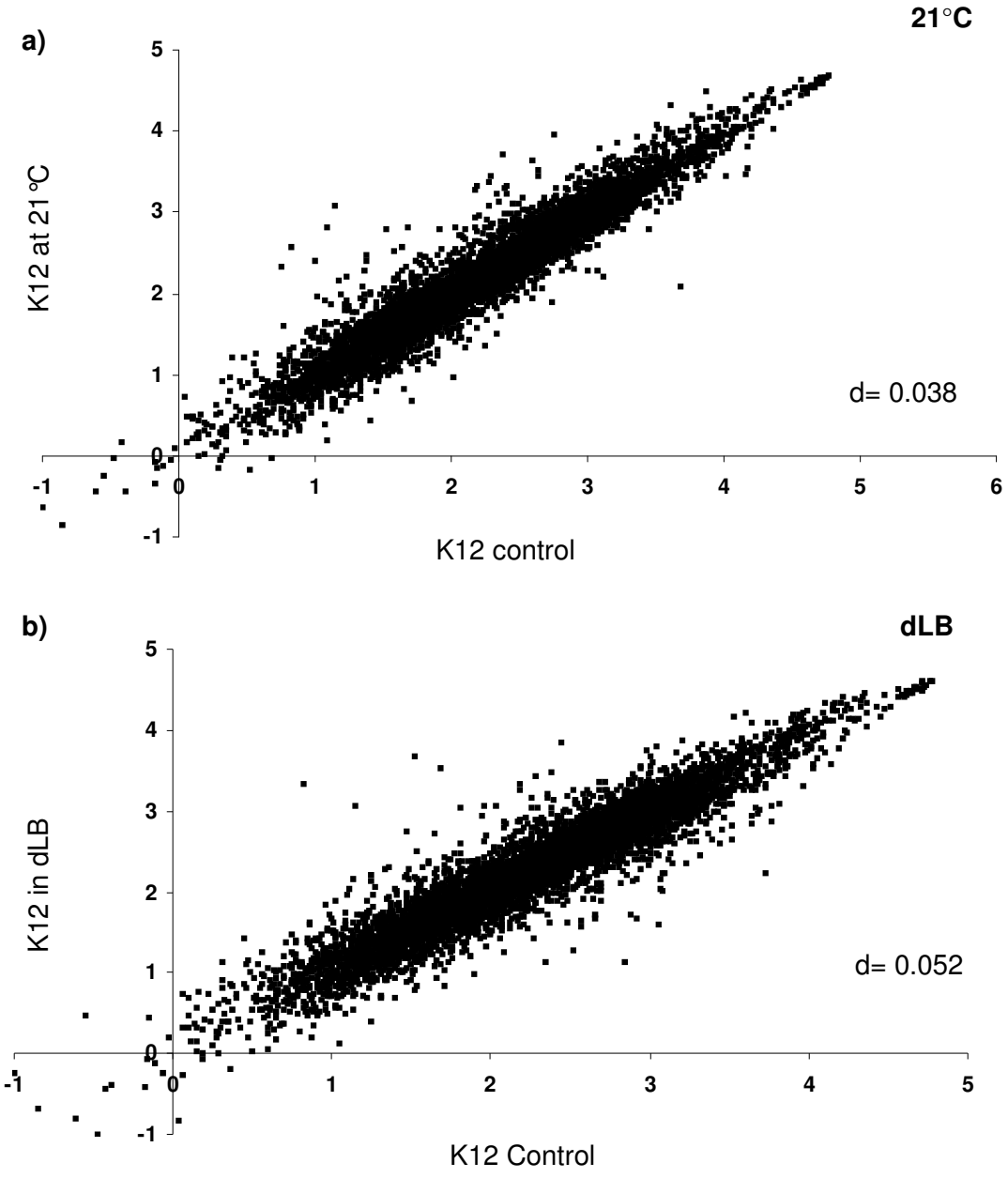

Figure 2.2 Comparison of gene expression in *Escherichia coli* K12 batch cultures in response to stress. The control condition was in LB broth at 37°C. Differential gene expression between **a)** 21°C and 37°C in LB broth and **b**) dLB broth and LB broth at 37°C. All data points were log transformed. Each data point represents a specific gene as measured from 4 replicate arrays.

from a perfect correlation (d=0.038) (Cooper *et al* 2003). Similarly, when expression levels were compared between regular and low nutrient concentrations it was found that the number of genes differentially expressed was greater (d=0.052) (Fig. 2.2 b). For strain E43 the d-values of the comparisons between 37°C and 21°C is 0.025 and between regular LB and dLB it is 0.041 (Fig. 2.3). These values show a strong linear relationship between the majority of spots on the microarrays of E43 (Fig. 2.3).

By contrast, when the log transformed expression values of the two strains are compared to each other in a single growth condition, the d-values increase to 0.193, 0.232, and 0.200 at 37°C, 21°C, and dLB, respectively (Fig. 2.4). The increase in the d-values shows that there are several genes not expressed at the same level between the strains under all tested growth conditions. Notably, there is the most variation from a linear trend at 21°C (Fig. 2.4). The shape of the scatter plots also indicates that the majority of genes that are differentially expressed between the strains are more highly expressed in *E. coli* K12.

2.4.3 Affymetrix analysis of the number of genes differentially expressed

The Affymetrix software was used to compare the arrays (see 2.3.9) and generate an "increase", "decrease", or "no-change" call. When the number of genes showing a change in expression between the control and experimental arrays or between replicates was tallied, we found that K12 had 10.2 to 16.7 % variability between replicates and a 24.7 to 27.1 % between test conditions (Table 2.1). The environmental strain had a 5.8 to 25.2 % variability between replicates and a 25.3 to 29.3 % difference between test conditions (Table 2.1). Between the different strains there was 39.5 % to 42.2 %

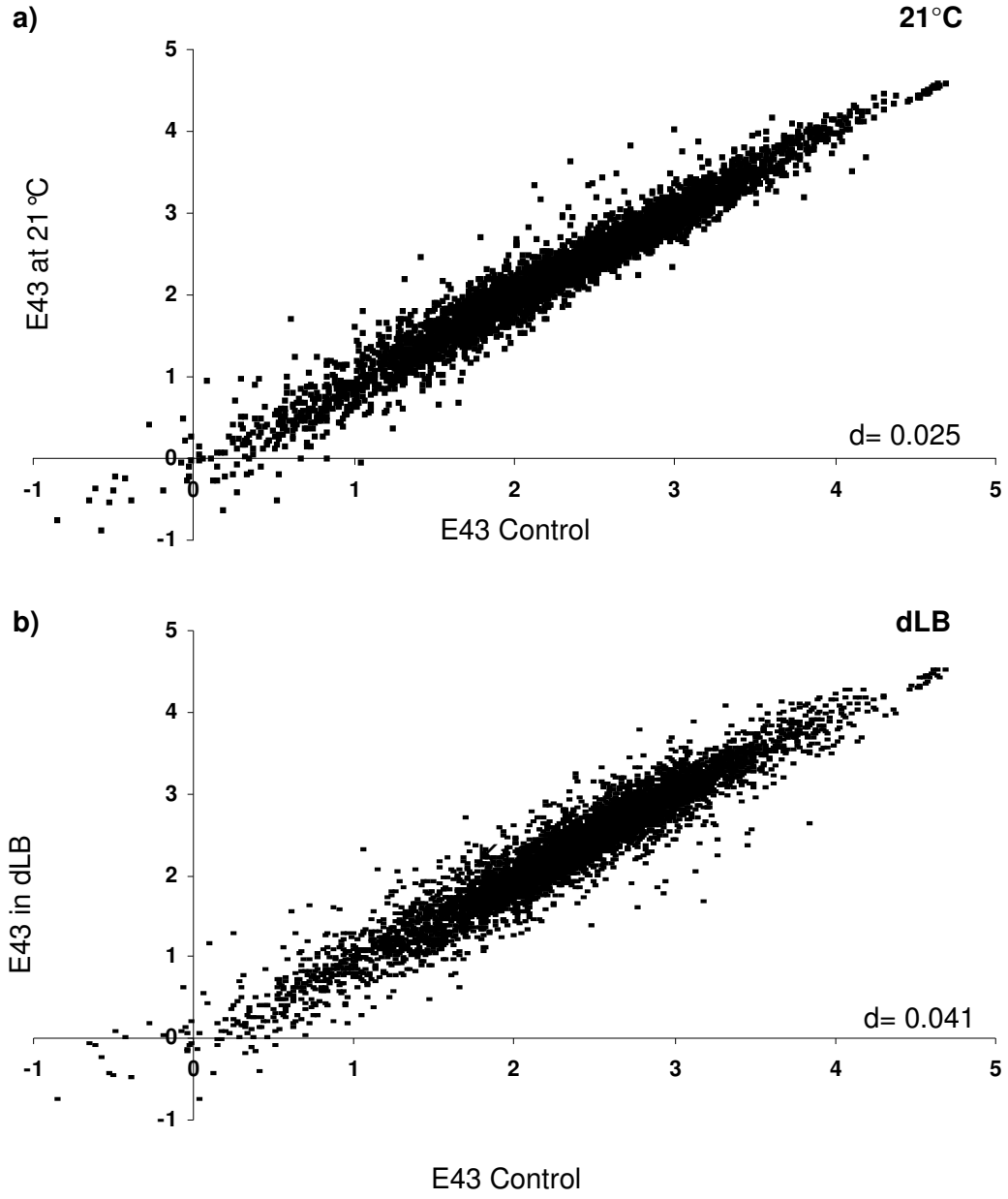

Figure 2.3 Comparison of gene expression in the environmental isolate *Escherichia coli* E43 grown in batch cultures. The control condition was in LB broth at 37°C. Differential gene expression between **a**) 21°C and 37°C in LB broth, and **b)** dLB broth and LB broth at 37°C. All data points were log transformed. Each data point represents a specific gene as measured from 4 replicate arrays.

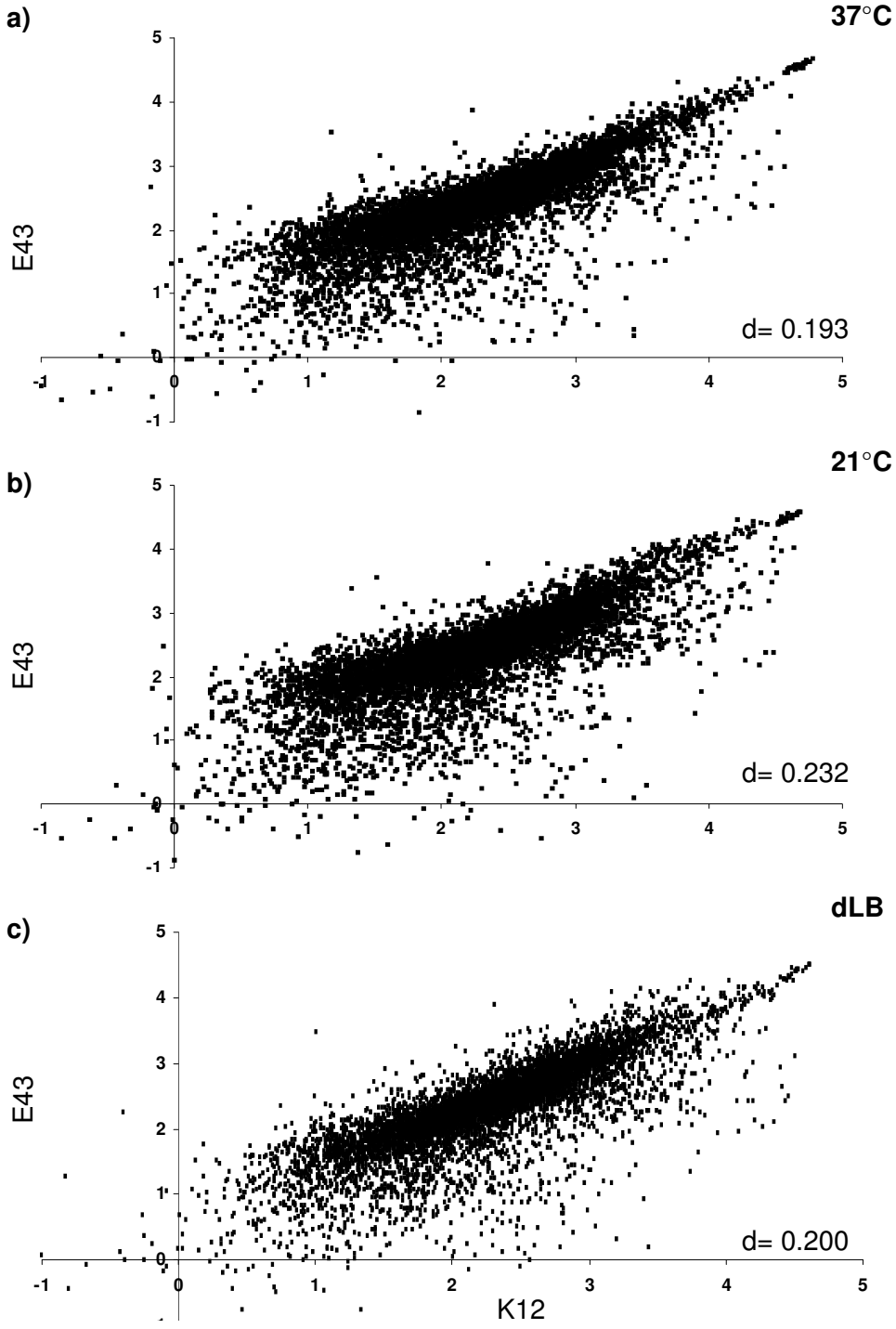

Figure 2.4 Comparisons of gene expression between *Escherichia coli* K12 and E43 grown in batch cultures. The base condition was *E. coli* K12 gene expression. Differential expression at **a)** 37°C, **b)** 21°C, and **c)** in dLB broth. All data points were log transformed. Each data point represents a specific gene as measured from 4 replicate arrays.

variability, depending on the growth conditions (Table 2.1). It is important to note that only genes that had an "increase" or "decrease" call for all 16 independent comparisons between replicates were used in future analysis. This step significantly decreased the percentage of genes that were called differentially expressed.

2.5 Discussion

2.5.1 Growth of the different strains

We deliberately chose two strains of *E. coli* which exhibited distinct phenotypes in their reaction to low temperature (Fig. 2.1). We can assert that the K12 strain should be well adapted to growth under controlled laboratory conditions, since it has been cultured in laboratories since 1922 (Riley *et al.* 2006). By contrast, we know little about the generational history of E43 and we can only suppose that its ancestors were subjected to a more variable environment, cycling between its host habitat (primary) and the environmental habitat (secondary). We also found that while both strains were confirmed to be *E. coli* by sequencing the small subunit RNA they were only 99.22% similar, indicating that these strains differ genetically to some extent.

2.5.2 Global trends

Figs. 2.2-2.4 show an overview of the genomic response to low temperature and dLB. It is important to stress that because we sampled the mRNA during late-log to stationary phase we did not test for the bacteria's immediate response to the experimental growth conditions, but rather what genes are regularly expressed at 21°C and in dLB broth as compared to gene expression in LB broth at 37°C. We found that when either *E. coli* K12 or E43 is subjected to 21°C or dLB broth there are few genes (24.7 to 29.3%)

Table 2.1 Variation in gene expression between replicates and experiments. Data are the mean percent of all genes on the array that increased or decreased in expression as determined by the Affymetrix software.

Array	K12-37°C	K12-21°C	K12-dLB	E43-37°C	E43-21°C	E43-dLB
K12-37°C	16.7					
K12-21°C	24.7	10.2				
K12-dLB	27.1	NC*	16.7			
E43-37°C	39.5	NC	NC	25.2		
E43-21°C	NC	42.2	NC	29.3	25.1	
E43-dLB	NC	NC	40.5	25.3	NC	15.8

* arrays Not Compared to each other

that are highly differentially expressed (Table 2.1). This finding was also reflected in the low d-values of 0.025 to 0.052 for the regression plots (Figs. 2.2 and 2.3). In response to both stresses, E43 had a lower d-value than K12, suggesting that E43 differentially regulates fewer genes in response to environmental stress than K12. Similarly, White-Ziegler and colleagues (2008) found that in *E. coli* K12, a similar number of genes, 297, that increased in expression at 23°C and Gadgil and colleagues (2005) found that about 9% of the genome was shifted between 37°C, 33°C and 28°C. Contrary to what one might expect, Polissi and colleagues (2003) found 20 genes had a 2-fold or greater change in mRNA transcription levels in response to cold shock (16°C); however, their experiment was done on a nylon-membrane format that lacked all non-coding sequences and operon leader and trailer sequences.

Furthermore, when contrasted to Figs. 2.2 and 2.3, it can be seen in Fig. 2.4 that there are significantly more genes differentially expressed between K12 and E43 than there are between the control condition and the environmental stresses. These differential mRNA levels between the two strains are present not only in the stress conditions, but also in the control condition at 37°C in LB broth (Fig. 2.4). The decrease in linear relationship of gene expression between the two strains, as expressed by larger d-values shows that these two strains differentially express some genes in all three growth conditions.

Many of these differentially expressed genes in Fig. 2.4 are expressed at a lower level in E43. It may be the case that these genes are not present in E43 since many of them were given the "absent" call by the Affymetrix calculations. However, it could also be the case that the genes are suppressed by E43 under most conditions or that the genes possess small sequence differences from the K12 sequence and therefore hybridize

poorly to the array. If the first case is true, it is a reasonable estimation that there are as several sequences unique to the environmental strain E43 that are not on the Affymetrix chip we used in this experiment because the chip's design was taken from the K12 MG1655 sequence.

Closely related strains sharing common genes but also having genes that are non redundant is well known. For example, Porwollik and colleagues (2003) have developed an array that consists of the core genes for *Salmonella* species and in addition has probes for species-specific genes of *S. typhi* and *S. typhimurium*. The authors postulated that these types of arrays could be developed and applied to any group of closely related species to yield valuable sequence and transcriptional information without the cost of sequencing each genome. Such an array or an array containing only strain-specific genes would be a valuable tool for distinguishing between strains found in the environment. The challenge would be in designing these arrays to be comprehensive enough to be practically applicable. It would be informative to run genomic DNA from strain E43 and other environmental isolates on the Affymetrix microarray designed from the genome sequence of *E. coli* K12 to determine what K12 genes are not in the genome of these strains.

2.5.3 Affymetrix analysis

Using the comparison analysis in the Affymetrix software package, we similarly found more variation between the strains themselves than either strain's response to the environmental stresses. The "change calls" were the criteria for whether a gene was differentially expressed or not. Without considering the magnitude of the change in gene expression, it is clear that the number of genes differentially expressed between strains is

higher than the number of genes differentially expressed between test conditions (Table 2.1). It is also interesting to note that there is more variability between replicates in E43 (15.8 to 25.2%) than in K12 (10.2 to 16.7). The environmental strain also had more variability between replicates under all three conditions. This may be reflective of the fact that the chip was designed for K12 strains and that E43 may have small sequence differences in some probes resulting in greater non-specific hybridization. Because the result for a particular gene is a composite of >15 short probes, Affymetrix arrays are sensitive to even single basepair differences. A definitive statement on this interpretation would require further studies with other environmental strains.

2.5.4 Genetic divergence

Thus far, these data show that despite the greater than 99% similarity between the test strains in this experiment at the small subunit RNA locus, there are differences between these two strains of *E. coli* (Fig. 2.1). Furthermore, the microarray data indicate that there may be more genetic differences between these strains (Fig. 2.4, Table 2.1). The link between genotypic and phenotypic adaptations within and between bacterial species has been notoriously difficult to demonstrate considering the differences between the species (Lenski and Travisano 1994). The strains we studied have grown in vastly different conditions, making a more controlled experiment comparing their genetics impossible. However, the study done by Lenski, Cooper, and colleagues (Lenski and Travisano 1994, Cooper *et al.* 2001, Cooper *et al.* 2003) demonstrated parallel mutations in the *spoT* gene of 8 of 12 originally identical *E. coli* populations, after 20, 000 generations (approximately 8.2 years) of growth in glucose medium at 37°C. They

believe that this *spoT* mutation was the genetic factor responsible for the temperature dependence these strains developed during the course of the experiment.

Furthermore, Cooper and collaborators (2003) discovered evidence of mutation leading to this adaptation in a controlled environment. Notably they found that the point mutations of the *spoT* gene were all in different locations between the 9 different populations that had acquired a mutation over 20,000 generations (Cooper *et al.* 2003). The differences in the *E. coli* strains used in our experiment may be due to similar mutations.

2.5.5 Conclusion

We found that there were more differences between strains regardless of the particular environment than between test conditions within a single strain. It is not surprising that only a select few genes are differentially expressed by a single strain, in response to an environmental stress. However it is surprising that these two strains of *E. coli* are not more similar in their response to environmental stress. Furthermore, it is interesting that two strains of the same species have such a high number of genes differentially regulated between them. These results do raise the question of how much we can extrapolate what is found in model organisms such as *E. coli* K12 to naturally-occurring organisms. As far as bacteria are concerned this study suggests a significant gap between the genetic expression of a lab strain maintained under artificial conditions for a long time and strains growing under natural conditions. This is not to say that we should abandon using model species, but that there should be an understanding that bacterial species, even very closely related strains are not genetically identical and may

not always behave in the same manner. These results indicate that it would be better to use an array designed from many different strains of *E. coli* in any BST application.

It is interesting to note that not only are there different numbers of genes differentially expressed by the strains in response to environmental stress, but also there are even fewer genes differentially regulated by both strains. This and other things including the high number of genes differentially expressed between the two strains regardless of environmental conditions will be the subject of the subsequent chapters.

Chapter 3

Global gene expression in *Escherichia coli* K12 and an environmental isolate (*E. coli* Is. E43) in response to temperature stress.

3.1 Abstract

We used microarrays to detect differential gene expression in an artificially maintained strain (K12) and a strain recently taken from the natural environment (E43) of *E. coli* in response to a decrease in temperature from 37°C to 21°C. Overall, 181 genes were differentially expressed when cells of *E. coli* K12 were grown at 21°C rather than 37°C, while only 65 genes were differentially expressed in *E. coli* E43 when grown at 21°C. By contrast, 582 and 1043 genes and IG regions were differentially expressed between the two strains at 37°C and 21°C, respectively. We only detected 9 genes differentially expressed by both strains in response to low temperature. There were 152 genes expressed between the two strains at both temperatures. We discovered that there were very few genes differentially expressed by both strains of *E. coli*, indicating that these two strains may have had different responses to the temperature stress. Further research into more environmental strains is needed.

3.2 Introduction

Escherichia coli, especially strain K12 MG1655, has been extensively studied (Riley *et al.* 2006). Not only are these gram-negative residents of mammalian intestines well-studied model of prokaryotic organisms, but also they are used in a variety of other applications due to their short doubling time and ease of cultivation. Despite the fact that this species may be the most studied at the molecular level, a standardized means to genetically differentiate between strains of *E. coli* is yet to be discovered (Fukushima *et al.* 2002, Kakinuma *et al.* 2003). In fact, more and more is being discovered about the

survival and reproduction of strains in the environment (Desmarais *et al.* 2002, Meays *et al.* 2005, Meays *et al.* 2006). Yet lack of understanding in these areas is an impediment especially for bacterial source tracking (BST). The goal of BST is to develop standardized methods for grouping enteric bacteria, especially *E. coli*, with their host groups, so that the host source of bacteria found in the environment may be identified. Therefore more information on the ecology of *E. coli* in the environment as well as greater knowledge about how to genetically differentiate between strains would be of benefit to BST. Because microarray technology is able to provide valuable sequence and gene expression information on thousands of genes in a single test, it may be used to gain more knowledge about the ecology of *E. coli* in the environment as well as a greater knowledge on how do differentiate and classify strains of *E. coli* (Porwollik *et al.* 2003, Porwollik *et al.* 2004, Soule *et al.* 2006, Wu *et al.* 2006).

So far, only a limited number of genes are used for any one molecular technique in BST research. Many applications take advantage of the dual nature of the small subunit ribosomal DNA which contains both hyper-variable and well conserved regions (Seurinck *et al.* 2003, Stoeckel *et al.* 2004). Methods that have had success in the field such as ribotyping and repetitive-element PCR depend on a large library of sequences and are limited to a particular geographical area (Hartel *et al.* 2002, Meays *et al.* 2004, Soule *et al.* 2006). While not all genetic BST methods depend on the 16S region of the genome, no one method has been championed by the majority of researchers (Scott *et al.* 2002, Meays *et al.* 2004, Stoeckel *et al.* 2004). So, BST research could benefit from a greater knowledge of *E. coli* genetics and a greater pool of genes that could be used to differentiate between strains.

Microarray technology allows probing of the messenger-RNA profile of entire genomes due to new printing technologies to provide valuable sequence information without the cost of actually sequencing every strain (Porwollik *et al.* 2003). These platforms are able to expand the genetic information available to researchers in a single experiment. It is no surprise that researchers have increasingly applied microarrays to the study of *E. coli*. A large portion of the published literature consists of research looking for specific biochemical processes in lab and mutant strains of K12 (Arfin *et al.* 2000, Chang *et al.* 2002, Gonzalez *et al.* 2002, Li *et al.* 2003, Polissi *et al.* 2003). Naturally, scientists have begun to apply microarray technology to environmental samples, recognizing its sensitivity and ability to probe hundreds and even thousands of genes (Wu *et al.* 2001, Wu *et al.* 2003, Lemarchand *et al.* 2005, Soule *et al.* 2006, Kostic *et al.* 2007).

Regardless of the preliminary nature of the application of microarray technology to environmental samples, the use of microarrays has advanced the scientific knowledge of how *E. coli* responds to its changing environment. It is well known that sigma factor RpoS (σ^S) is a general regulator of *E. coli*'s response to several stresses including starvation, extreme temperature, entry into stationary phase, and hyperosmolarity (Patten *et al.* 2004, Weber *et al.* 2005). The use of microarrays has helped researchers characterize how many genes are under RpoS control and even revealed new stress situations, like osmotic shock and biofilm formation, where RpoS genes are differentially regulated (Sugiura *et al.* 2003, Patten *et al.* 2004, Weber *et al.* 2005, White-Ziegler *et al.* 2008). Even more interestingly, White-Ziegler and colleagues (2008) have found that RpoS dependent genes were differentially regulated even in a small temperature shift

from 37°C to 23°C, leading them to postulate that temperature is a key factor in triggering the RpoS response.

To our knowledge a direct comparison between a strain maintained under artificial conditions for a long time and a strain recently isolated from a natural environment has not yet been conducted. Such a study may also be beneficial in the development of a new microarray that would allow library-independent BST. We believe that stress regulators or stress-related genes may provide good candidate genes for the differentiation of *E. coli* strains and chose temperature shift to 21°C as the first environmental stress to test. We used the lab strain of *E. coli* K12 MG1655, and a recent bovine environmental isolate called 43(C)-4A or E43. Both strains were exposed to the temperature change and their change in messenger RNA profile was detected using the commercially available Affymetrix microarrays.

3.3 Materials and methods

3.3.1 Strains, sequencing, and culture growth

The strains used in this study were pure cultures (See 2.3.1) of *E. coli* K12 MG1655 (ATCC 700926) and bovine isolate of *E. coli*, 43(C)-4A (E43), isolated from the natural environment on June 30, 2004. The previously PCR-amplified small subunit ribosomal DNA (ssu rDNA) was sequenced by cloning purified DNA into the pGEM-T plasmid using the pGEM kit (Promega, Madison, WI, USA) followed by sequencing based on the Sanger Method at the Centre for Biomedical Technology (University of Victoria). Sequences were analyzed by the NCBI BLAST program (http://www.ncbi.nlm.nih.gov/) and confirmed to be greater than 99% similar to other annotated *E. coli* sequences (See

2.3.4). All successive culture growth was done in LB broth. Cultures were grown with agitation (250 rpm) at the specified temperatures (37°C or 21°C).

For selection of an environmental strain that grows differently than *E. coli* K12 at low temperature (21°C), growth curve analysis was done on multiple environmental strains. The rates of growth were calculated in Microsoft Excel 2003 (Microsoft, Mississauga, ON, Canada) by the use of a standard curve of viable cell concentration versus cell Optical Density (OD) (See 2.3.9). Strain 43(C)-4A was selected because it recovered more rapidly at low temperature than K12 and grew at a faster rate.

3.3.2 RNA extraction

For microarray analysis, cultures were harvested when they reached an OD (600 nm) of 0.85-1.0. RNA was extracted from cell pellets using a RiboPure RNA extraction kit (Ambion, Austin, TX, USA) according to the manufacturer's instructions (See 2.3.6). All labware and surfaces were cleaned with RNaseZap (Ambion, Austin, TX, USA) to prevent degradation of RNA by RNAses . RNA quantity and quality were checked by measuring the OD_{260}/OD_{280} ratio on the UV/Vis spectrophotometer and gel electrophoresis. RNA extract was stored at -80°C for subsequent experimentation.

3.3.3 Preparation of cDNA from total RNA

Preparation of cDNA followed the recommended procedure by the microarray manufacturer (Revision 5, Affymetrix, Santa Clara, CA, USA) as described previously (See 2.3.7). Amplification of mRNA into cDNA was performed on 10 µg of total RNA (Gadgil *et al.* 2005). All incubation periods were performed in a Px2 thermal cycler (Thermo-Hybaid, Ashford, UK). The amount of RNA produced was quantified by UV spectrophotometry. Only reactions with yields greater than 2µg of cDNA were used in

subsequent procedures. The cDNA was fragmented and labelled using GeneChip DNA labelling reagent biotin (Affymetrix, Santa Clara, CA, USA) (See 3.2.7).

3.3.4 Hybridization to the microarray

Biotinylated cDNA was sent to the WRC (UBC, Vancouver, Canada), where, the cDNA was hybridized to a GeneChip® *E. coli* Antisense genome array containing probes for 7312 ORF's and IG regions (Affymetrix, Santa Clara, CA, USA). Hybridization was performed according to the recommended protocol for Prokaryotic arrays (Revision 5, Affymetrix, Santa Clara, CA, USA) (See 2.3.8). On the array, each ORF and IG is represented by 15 perfect match (PM) probes, about 25 bp in length. Each PM probe is accompanied by a mismatch probe (MM) that has a single base pair mismatch at position 13. This MM is meant to detect the level of non-specific hybridization. Arrays were scanned using Affymetrix Microarray Suite 5.0 (Suite 5.0) on an Agilent GeneArray® Scanner (Affymetrix, Santa Clara, CA, USA).

3.3.5 Microarray data analysis

The data for all genes were analyzed on Suite 5.0, provided by Affymetrix at the WRC (UBC, Vancouver, Canada). The array data were analyzed and compared according to the Affymetrix "Expression Analysis, Data Analysis Fundamentals" manual (Affymetrix 2002). In addition to calculating the magnitude of expression, a "detection call" was independently calculated using a one-sided Wilcoxon-signed rank test for whether the genes were "present" or "absent" (See 2.3.9). Four slides for each condition were used, representing 4 independent biological replicates.

Comparisons between arrays were done on a probe-to-probe basis using a one-step Tukey's biweight estimate and output in the natural log ($\log_2$) format. The natural

logarithm format allows for simple interpretations since a $\log_2$ of 1 is a 2-fold change. In addition, a "change call" was generated by an independent calculation in a similar manner to the "detection call" calculation for whether there was an "increase", "decrease" or "no change" in gene expression between arrays. In our experiment the baseline array was always the array representing the bacteria grown at 37°C in LB medium or in the case of comparisons between strains, K12 was the control.

We used four biological replicates for each experimental array as well as for the controls. Instead of averaging the replicates and comparing the experimental arrays to the control array, each experimental array was compared to each control array for a total of 16 comparisons as described in data analysis (Data Analysis Fundamentals, Affymetrix, 2002). Data were exported into Microsoft Excel spreadsheets and all future analysis was done in Microsoft Excel 2003 (Microsoft, Mississauga, ON, Canada). The actual number of ORFs and IG regions included in each spreadsheet was 7312/array. The annotation file for the GeneChip® array used in this report can be found on the Affymetrix website on the Technical Documentation page under “Ecoli_ASv2 Annotations" (http://www.affymetrix.com/support/technical/annotationfilesmain.affx). In this report the 16 comparisons were averaged and all reported standard deviations were computed between the 16 comparisons. Not only were both strains’ response to the environmental stress analyzed, but also the transcriptional profiles of the strains were compared to each other in both environmental conditions. An outline of the comparisons made can be seen in Fig. 3.1.

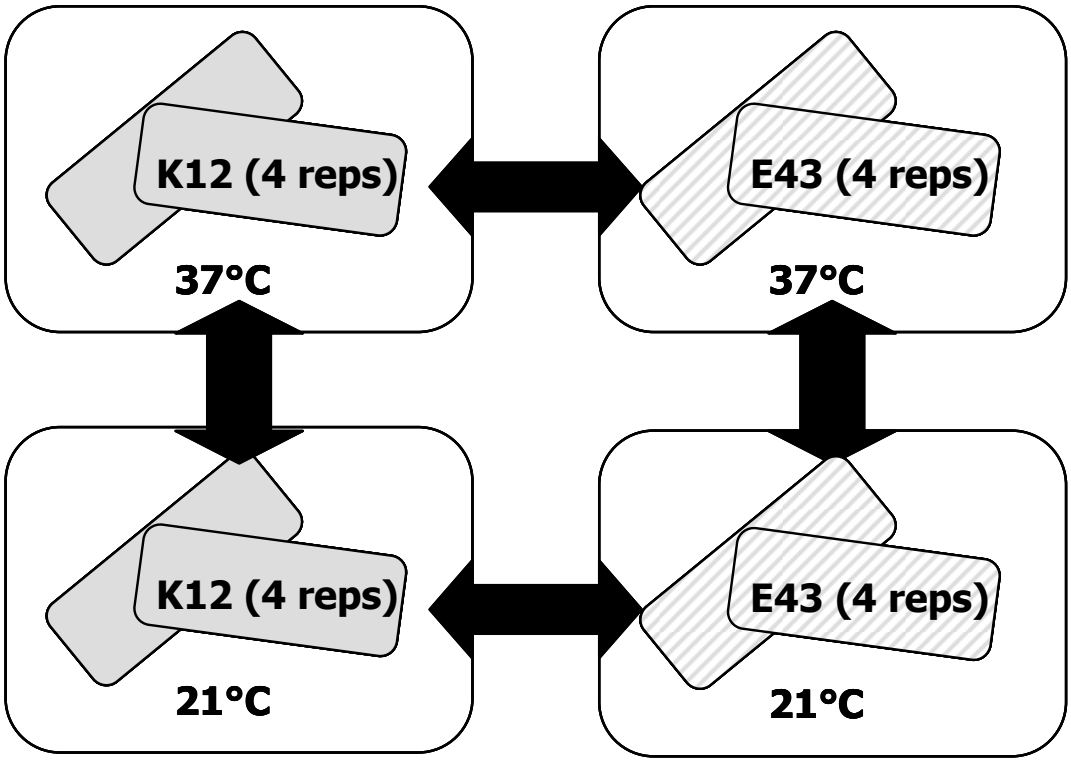

Figure 3.1 Schematic of the experimental design used to compare the transcriptional profiles of *Escherichia coli* K12 and E43.

3.4 Results

3.4.1 Growth curves and the environmental isolate

E. coli K12 was selected because it has been sequenced and it is already well-studied on microarrays (Blattner *et al.* 1997). The environmental isolate was selected for the following reasons. E43 grew to a maximum OD of about 1.3 at both temperatures while K12 only grew to a maximum OD of about 1.0 (Fig. 2.1) and because it grew at a faster rate than K12 (See 2.3.2). Also, E43 took only 8 hrs to reach late log phase at 21°C, while K12 took about 10 hours (Fig. 2.1).

3.4.2 General trends

Out of the whole array (7312 genes and IG regions), there were 181 genes (2.5 %) and IG regions differentially expressed by *E. coli* K12 in response to the temperature decrease, by contrast only 65 (0.9 %) genes and IG regions were differentially expressed by strain E43 (Fig. 3.2a). In both strains the most genes were differentially expressed at the $\log_2$-ratio of 1 (2-fold) and there were fewer genes in the other categories with the fewest being at the largest $\log_2$-ratio of 6. There was only one gene differentially expressed by K12 as high as $\log_2$-ratio of 6 (64-fold). The largest $\log_2$-ratio difference seen in E43 was 4 (Fig. 3.2a).

By contrast, there were a total of 582 (8.0 %) genes and IG regions differentially expressed between K12 and E43 at 37°C and 1043 (14.3 %) at 21°C. However, the pattern of the distribution of differentially expressed genes was similar with the most genes differentially expressed at both temperatures being at the $\log_2$-1 level (Fig. 3.2b). Between the two strains, there were 2 genes differentially expressed to a $\log_2$-ratio of 10 at 21°C and only one at 37°C.

3.4.3 Functional classifications

Since the majority of IG regions do not have known functions, they were excluded from the grouping into functional classifications. We also limited our classification to genes differentially expressed greater than the $\log_2$-2 level, since we considered highly differentially expressed genes to be of the greatest interest to our question about which genes could be potentially used in the development of BST microarrays. Under these criteria, there were 26 genes differentially expressed by K12 and 9 genes by E43. In K12, the function-categories with the most differentially expressed genes in response to decreased temperature were “metabolism” and “unknown” (both at 35 % of differentially expressed genes) (Fig. 3.3a). The next largest category was “transport” with 12 percent of the genes (Fig. 3.3a). There were no genes in the "cell structure" or "regulation" categories. In the case of E43, there was 1 more gene in the “unknown” category than in “cell structure” (33 %) and “regulation” (44%), E43 lacked genes in “information transfer”, “transport”, “cell process”, and “cell structure”. Overall there were 17 fewer genes differentially expressed by E43 than by K12 (Fig. 3.3a).

Similar to when all array probes were included in the analysis, there were more genes differentially expressed between K12 and E43 than in either strain’s response to temperature stress as shown by the scale in Fig. 3.3b. However, the relative percentages of differentially expressed genes remained similar in each category. At 37°C 107 of the 169 differentially expressed genes fell into the “unknown” and “metabolism” categories at 35 % and 28 %, respectively. Similarly, at 21°C 191 of the 286 genes were in the same categories, with "unknown" comprising 34 % of genes and "metabolism" having 33 % of the genes. Interestingly, the next largest category at both temperatures was the

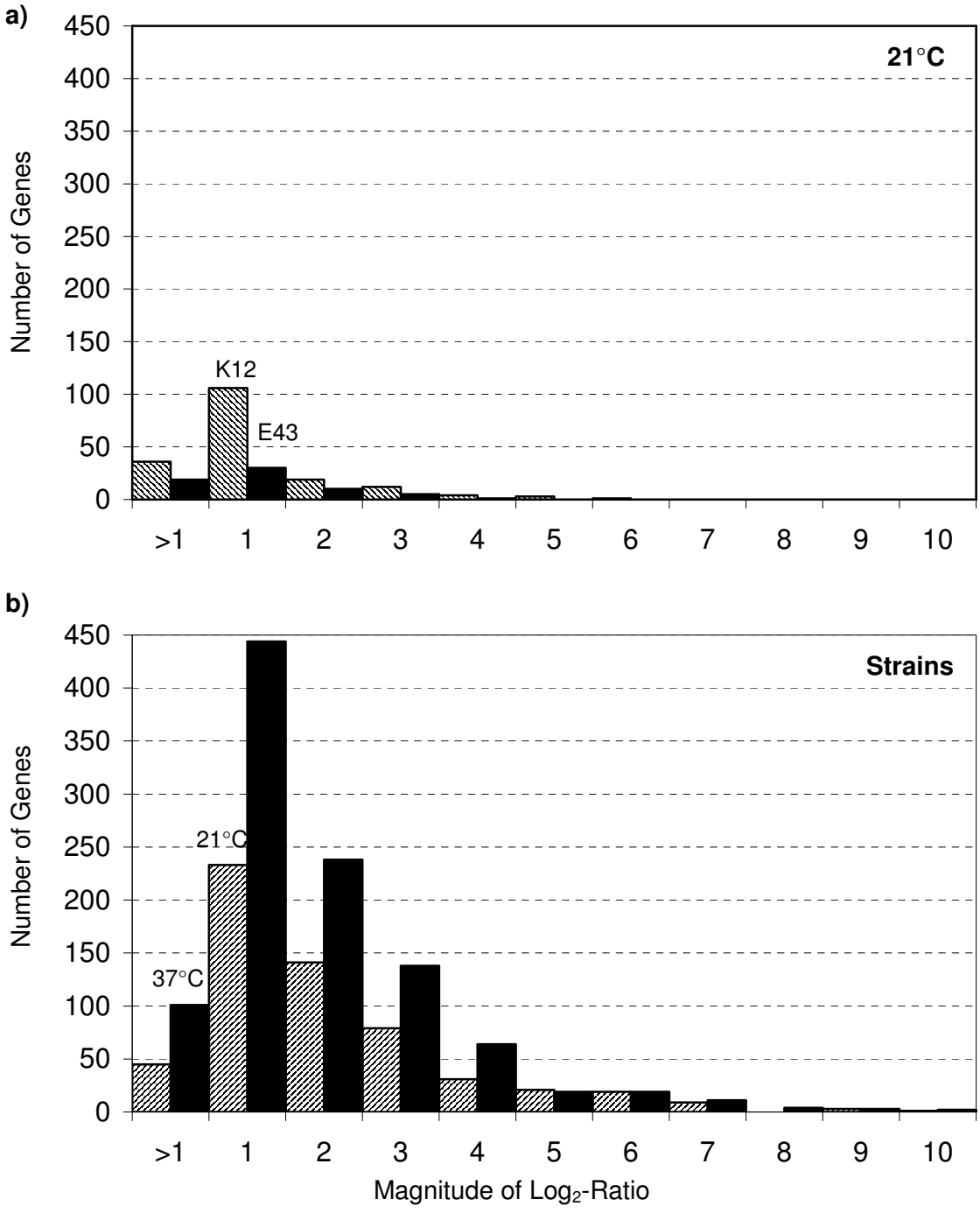

Figure 3.2 Distribution of genes and IG regions differentially transcribed by **a)** *Escherichia coli* K12 (hatched) and E43 (solid) in response to a decrease in temperature from 37°C to 21°C and between **b)** K12 and E43 at 37°C (hatched) and 21°C (solid). Cultures were grown in LB broth to an optical density of 0.9-1.0 (600 nm) with agitation at 250 rpm. RNA was hybridized to an GeneChip® *E. coli* Antisense microarray.

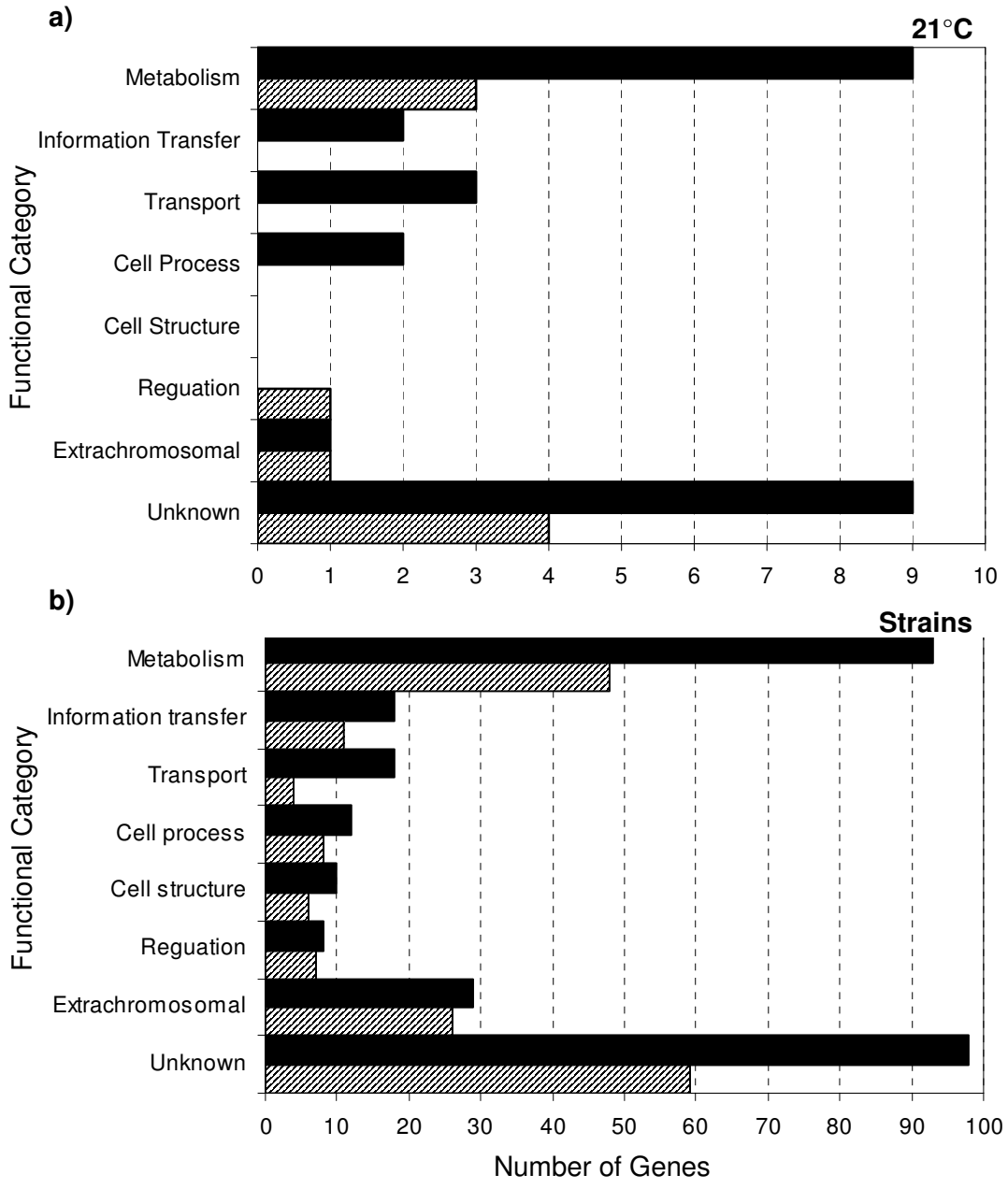

Figure 3.3 Functional categories of genes differentially expressed by **a)** *E. coli* K12 (solid) and E43 (hatched) in response to a temperature decrease from 37°C to 21°C and **b)** between *E. coli* K12 and E43 at 37°C (solid) and 21°C (hatched). Categories were defined according to the ASAP website (http://www.genome.wisc.edu/tools/asap.htm).

"extrachromosomal" category with 15 % at 37°C and 10 % at 21°C. The rest of the categories held fewer than 10 % of genes. Since there were more genes differentially expressed at 21°C than at 37°C, all functional categories have more genes at 21°C.

3.4.4 Genes highly differentially expressed in response to temperature stress

At a ratio of $\log_2$ 2, or 4-fold, there were 26 genes that were differentially expressed by K12 in response to temperature shock (Fig. 3.4, Table 3.1). Only 4 of these 26 genes were down-regulated in response to temperature shock. The gene *ymgC* had the greatest differential expression at $\log_2$ 5.42. Genes of note that were up-regulated by K12 are *csg* family with the notable absence of *csgG*. There were also 14 IG regions differentially expressed (data not shown). For comparison sake, the differential expression level of these genes by E43 is also shown on Fig. 3.4. It is apparent that most genes that were differentially expressed by K12 were not significantly differentially expressed by E43. In fact, for most of these genes the average differential expression level by E43 was smaller than the standard deviation.

As can be seen in Fig. 3.5 there were only 9 genes that were up-regulated by E43 at the $\log_2$-2 threshold. There were 6 IG regions differentially expressed (data not shown). With a $\log_2$ of 4.19, *gadB* was the gene most highly differentially expressed by E43. In contrast to the situation in K12, it appears in Fig. 3.6 that most of the genes differentially expressed by E43 are also differentially expressed by K12.

At the ratio threshold of $\log_2$ 2 only 3 genes were differentially regulated by both K12 and E43 in response to low temperature (Table 3.1). These genes are *ymgB, ycgZ, gadA*. Because of the stringent requirements for a gene to be selected as differentially expressed (if even one of the 16 comparisons had a "no change" or "marginal change"

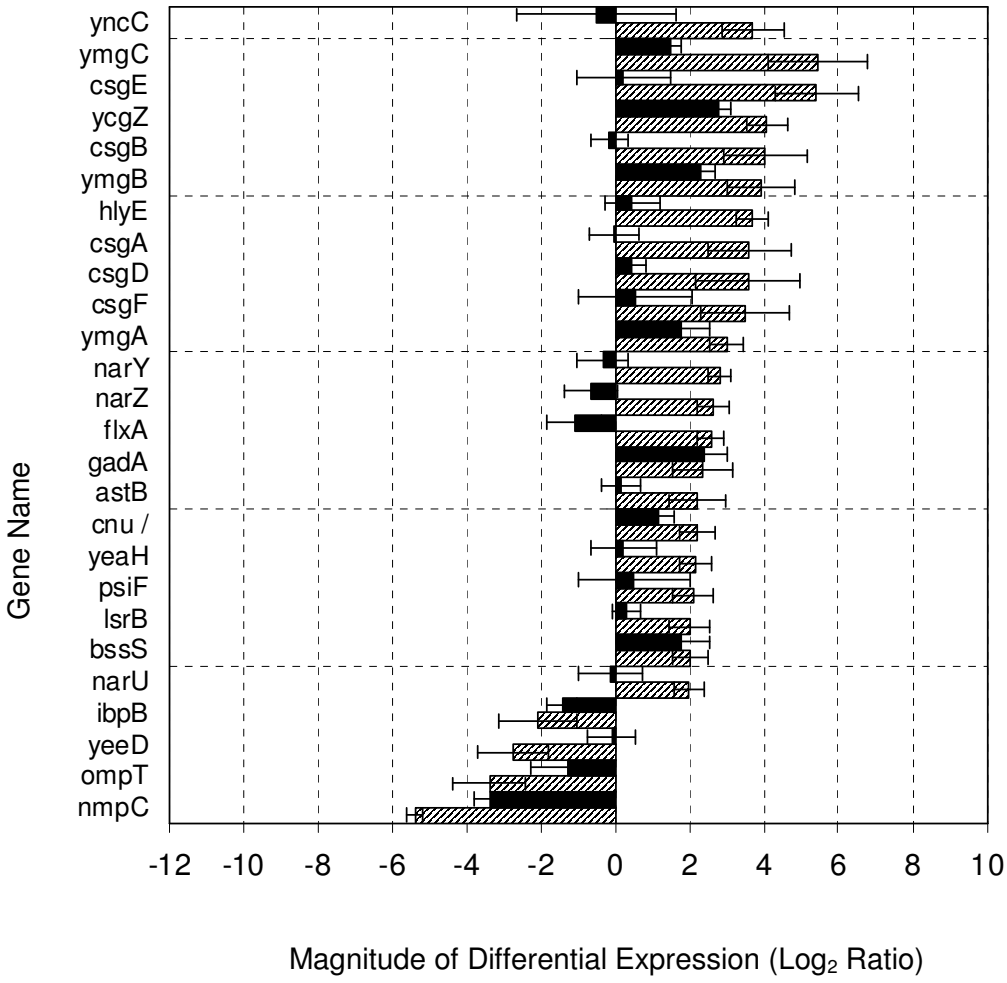

Figure 3.4 Genes of *Escherichia coli* K12 (hatched) that were differentially expressed at 21°C as compared to expression at 37°C, at the $\log_2 2$ threshold. Both strains were grown in batch culture in LB broth with agitation. The differential expression values for *E. coli* E43 (solid) are shown for comparison. Error bars represent the standard deviation.

Table 3.1 Functions of the genes differentially expressed by *Escherichia coli* K12 and E43 in response to a decrease in temperature from 37°C to 21°C. The magnitude of gene expression are the mean of comparisons between 4 replicates.

Gene Name	Blattner ID	Log_2Ratio K12	Stdev* K12	Log_2Ratio E43	Stdev E43	Gene Product
Cell Process						
hlyE	b1182	3.70	0.43	---**	---	hemolysin E
ibpB	b3686	-2.07	1.03	-1.43	0.29	heat shock chaperone
Information Transfer						
csgD	b1040	3.58	1.41	---	---	DNA-binding transcriptional regulator of adhesion determinants
yncC	b1450	3.71	0.83	---	---	predicted DNA-binding transcriptional regulator
Metabolism						
astB	b1745	2.21	0.75	---	---	succinylarginine dihydrolase
csgA	b1042	3.61	1.13	---	---	cryptic curlin major subunit
csgB	b1041	4.04	1.12	---	---	curlin nucleator protein, minor subunit in curli complex
gadA	b3517	2.36	0.82	2.38	0.69	glutamate decarboxylase A, PLP-dependent
gadB	b1493	---	---	4.19	0.68	glutamate decarboxylase B, PLP-dependent
nanA	b3225	---	---	2.30	0.66	N-acetylneuraminate lyase
narU	b1469	2.00	0.41	---	---	nitrate/nitrite transporter
narY	b1467	2.82	0.32	---	---	nitrate reductase 2 (NRZ), beta subunit
narZ	b1468	2.63	0.43	---	---	nitrate reductase 2 (NRZ), alpha subunit
ompT	b0565	-3.38	0.97	-1.27	0.43	DLP12 prophage; outer membrane protease VII (outer membrane protein 3b)
psiF	b0384	2.10	0.55	---	---	conserved protein
Regulation						
yhiE/gadE	b3512	---	---	2.99	0.63	acid-induced positive regulator of glutamate-dependent acid resistance
Transport						
csgE	b1039	5.42	1.14	---	---	predicted transport protein
csgF	b1038	3.50	1.18	---	---	predicted transport protein
lsrB	b1516	2.02	0.55	---	---	AI2 transporter -!- periplasmic-binding component of ABC superfamily

* stdev = standard deviation calculated by Microsoft excel

** not differentially expressed

Table 3.1 Continued.

Gene Name	Blattner ID	Log_2Ratio K12	Stdev* K12	Log_2Ratio E43	Stdev E43	Gene Product
Extrachromosomal						
cspI	b1552	---	---	2.01	0.88	Qin prophage; cold shock protein
nmpC	b0553	-5.38	0.20	---	---	DLP12 prophage; truncated outer membrane porin (pseudogene)
Unknown General Classification						
bssS / yceP	b1060	2.02	0.48	1.81	0.40	regulator of biofilm formation
cnu / ydgT	b1625	2.21	0.48	1.19	0.23	predicted regulator
flxA	no ID	2.57	0.36	---	---	unknown
hdeA	b3510	---	---	3.24	0.46	stress response protein acid-resistance protein
hdeB	b3509	---	---	3.44	0.81	acid-resistance protein
ycgZ	b1164	4.08	0.56	2.80	0.64	predicted protein
yeaH	b1784	2.16	0.42	---	---	conserved protein
yeeD	b2012	-2.75	0.96	---	---	conserved protein
ymgA	b1165	3.00	0.47	1.79	0.40	predicted protein

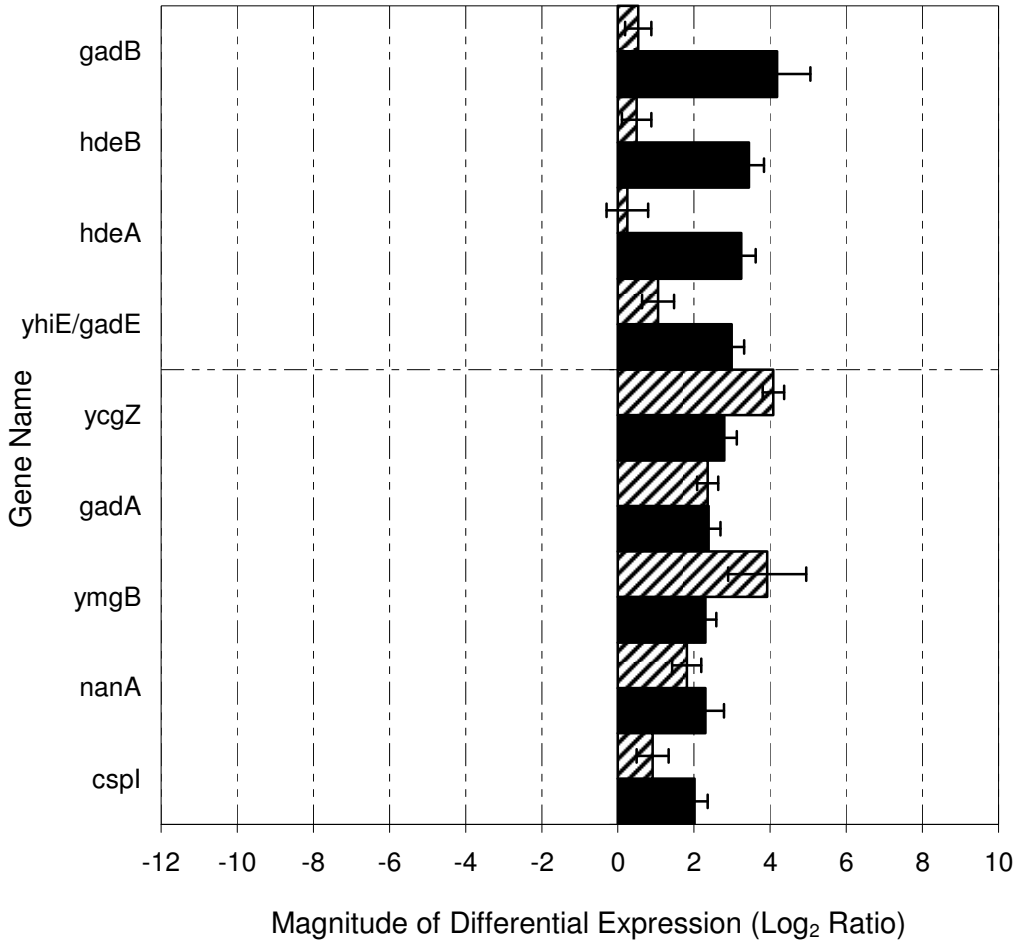

Figure 3.5 Genes of *Escherichia coli* E43 (solid) that were differentially expressed at 21°C as compared to expression at 37°C, at the $\log_2 2$ threshold. Both strains were grown in batch culture in LB broth with agitation. The differential expression values for *E. coli* K12 (hatched) are shown for comparison. Error bars represent the standard deviation.

call that gene was excluded) and for comparison purposes, if a gene was differentially expressed by K12 then we looked for that gene's data in E43. Even if that gene was independently selected as "differentially expressed" in E43, it could be below the $\log_2$-2 threshold. These results can be seen in Figs. 3.4 and 3.5 and Table 3.1. Most of the genes that were up-regulated by K12 did not correspond to any gene that was up-regulated by E43 except for *bssS/yceP*, *ymgA, ymgC*, *cnu/ydgT*, and the three genes mentioned above. For genes down-regulated *ompT,* and *ibpB* were all down-regulated in E43 as well, but to a lesser level than in K12.

3.4.5 Genes highly differentially expressed between the two strains

As already stated there are more genes differentially expressed between the two strains and this remains true at the 2-fold threshold. At 37°C there are 168 genes (137 IG regions) differentially expressed between the strains while at 21°C there are 288 (213 IG regions). However, what is different in this case is that there are a larger portion of genes in common between these two lists. There are 152 genes that appear on both lists, which is 90.5% of the genes at 37°C and 53% of the genes at 21°C. The additional genes differentially expressed at 21°C are thus demonstrated to be in addition to the genes normally differentially expressed between the two strains. At 37°C, 81.5% of the genes are expressed at a lower level in E43 and 73.6% are expressed at a lower level by E43 at 21°C (Appendix B, Tables B1-B3). As shown in Figs. 3.6 and 3.7 all of genes were decreased in expression by strain E43. These graphs show only genes differentially expressed above the threshold $\log_2 5$ for brevity's sake. It can be seen from these two graphs that even when a single gene was not detected as differentially expressed at both

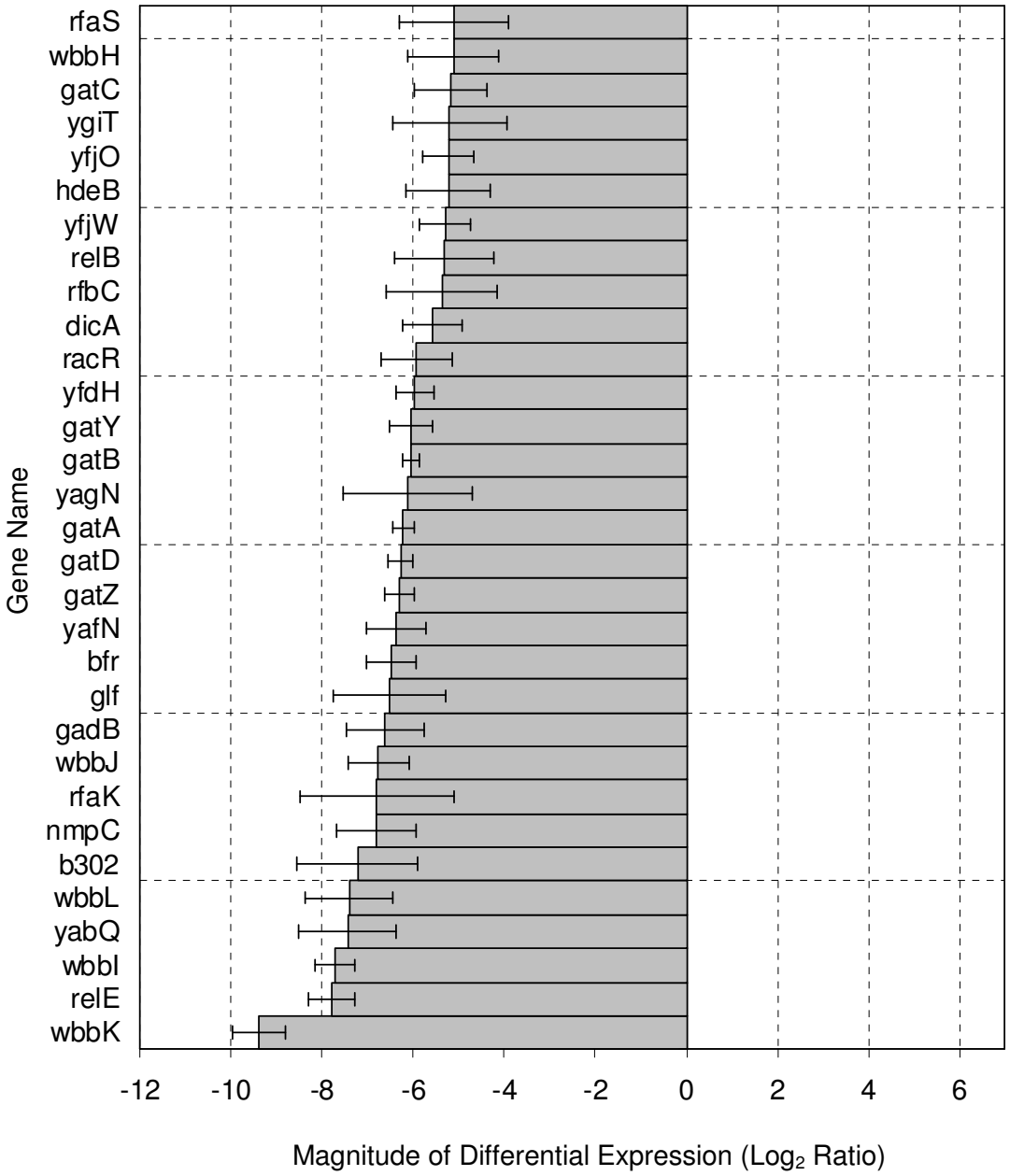

Figure 3.6 Genes differentially expressed between *Escherichia coli* K12 and E43 at 37°C at the $\log_2 5$ threshold. Expression in K12 was used as the baseline. Negative values indicate the gene was repressed by E43 as compared to K12. Both strains were grown in batch culture in LB broth with aeration. Error bars represent the standard deviation.

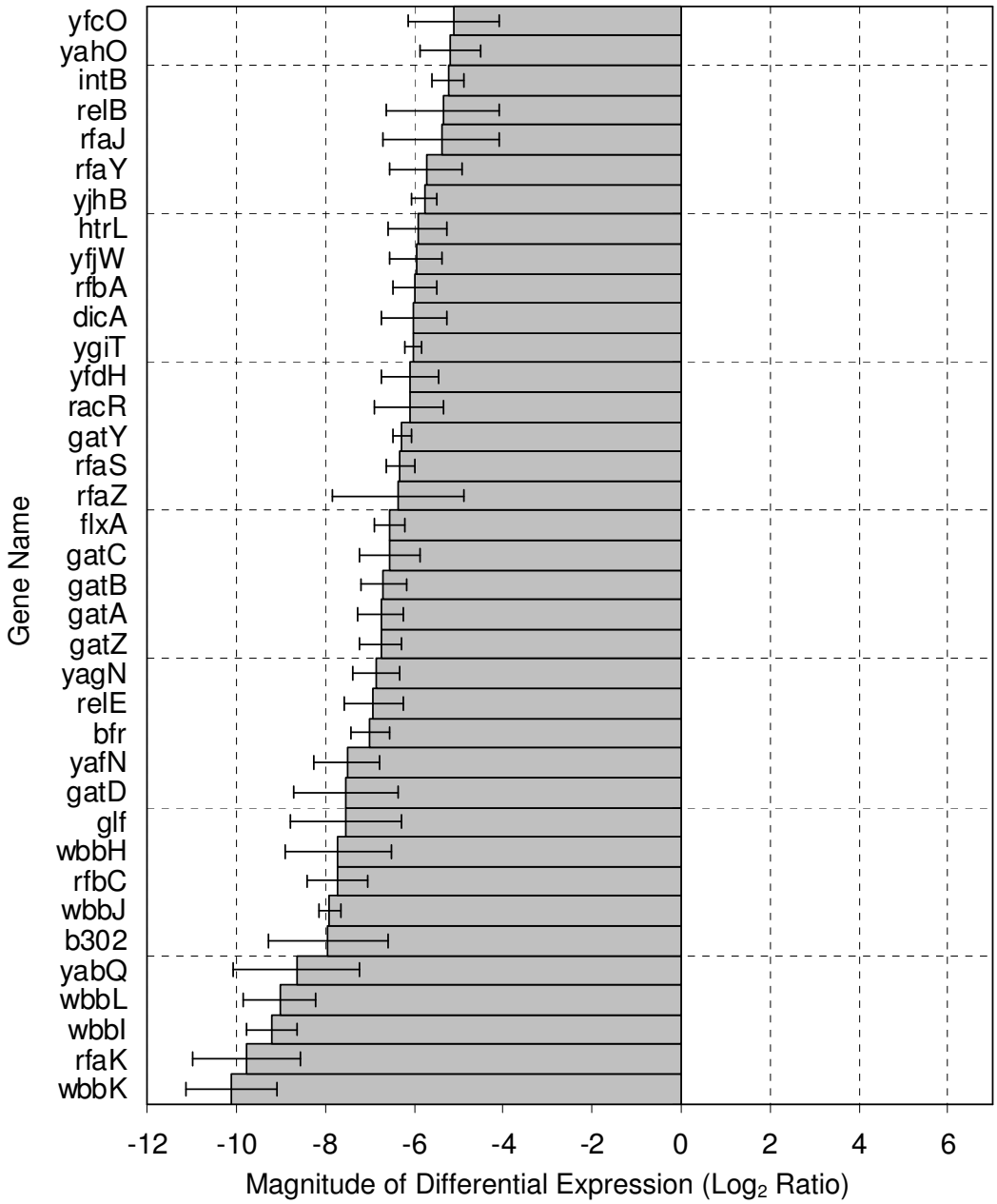

Figure 3.7 Genes differentially expressed between *Escherichia coli* K12 and E43 at 21°C at the $\log_2 5$ threshold. Expression in K12 was used as the baseline. Negative values indicate the gene was repressed by E43 as compared to K12. Both strains were grown in batch culture in LB broth with aeration. Error bars represent the standard deviation.

temperatures (due to stringent requirements), more often than not the differential expression at both temperatures followed the same pattern.

3.5 Discussion

3.5.1 Phenotypic differences

The phenotypic differences between the strains show that the strains generally react differently to their environment. It can be generally stated that the *E. coli* 43(C)-4A grows faster than K12 at 37°C and 21°C and that both strains have a longer lag phase at 21°C (Fig. 3.2). Although we have no direct knowledge of where or when these two strains of *E. coli* diverged from each other, we have demonstrated that there is at least one phenotypic difference between them which is that E43 recovers from growth in 21°C faster than K12. On the other hand, being different strains, both elicit a phenotypic response to a decrease in temperature. By using microarray technology we were able to compare and contrast E43's gene expression response to temperature to the well-studied lab strain K12. It was our hope that by using two distant *E. coli* strains we would be able to find some common genes that are elicited in response to environmental stress and that those genes could be used in the future to differentiate between strains of *E. coli*.

3.5.2 Comparing genetic results to other studies for K12

We found that a low percentage of the *E. coli* K12 genome (7312 genes and IG regions), 2.48% (181 genes), was differentially expressed in response to the temperature decrease. The environmental strain differentially expressed even less of its genome: 0.89% (65 genes). Furthermore, a closer analysis reveals that even fewer genes were differentially expressed by both strains. These results are in contrast to a study done by Gadgil, and colleagues (2005). In their study on the response of *E. coli* K12 to a

temperature shift from 37°C to 33°C and 28°C, they found that about 9% of the genome was altered in its expression in response to both temperature shifts (Gadgil *et al.* 2005). Of these they found that the largest percentage of genes was involved in energy metabolism which we similarly found as well (Fig. 3.4a). It is interesting that Gadgil and colleagues (2005) found a greater percentage of genes changed in response to a smaller temperature change than the one in our study, yet we found fewer genes differentially expressed. In fact, Gadgil, and colleagues (2005) found that despite a reduced glucose uptake, several TCA-cycle genes were up-regulated at lower temperatures while there was no significant change in any stress-related genes. These discrepancies may be due to different statistical analysis used by our respective reports. Our study also used different culture media than the Gadgil experiment. Comparisons are also more difficult because in addition to different selection criterion for genes to be scored as differentially expressed, different array platforms were used.

In contrast to our study and similarly to Gadgil, and colleagues (2005), White-Ziegler and colleagues (2008) found 297 genes that responded to a decreased temperature of 23°C. However, unlike Gadgil, Kapur and Hu (2005), White-Zielgler and colleagues (2008) found several genes related to stress responses like cold-shock that where differentially regulated in response to low temperature (23°C). The discrepancy in number of genes could be due to our stringent requirements for a gene to be differentially expressed and the high threshold. However, it may also be due to the difference in culture medium used. We used a rich LB broth while both of the other experiments used a minimal medium, in which it is expected that *E. coli* will need to express a greater number of genes than in a complex medium. We used LB broth because it is a complex

medium that approximates the heterogeneity of nutrients found in the environment but also allows comparisons between the strains.

3.5.3 Genes differentially regulated by both strains

Of the three genes differentially expressed by both strains relatively little is known (Tables 3.1 and 3.2). Little to no information has been compiled in either NCBI or the ASAP databases. However, *ycgZ* was shown to be regulated by the RpoS network, well known to regulate *E. coli*'s stress response including cold shock, starvation, hyperosmolarity, and stationary phase (Weber *et al.* 2005, White-Ziegler *et al.* 2008). While another project found that *ycgZ* was involved in *E. coli*'s response to cold shock (Polissi *et al.* 2003). Similarly, *ymgB* has been linked to *E. coli* K12's response to low temperature but was found to be independent of the RpoS regulon (White-Ziegler *et al.* 2008). *YmgB* is also thought to play a critical role in the suppression and regulation of biofilm formation (Lee *et al.* 2007). Another study on the *E. coli* response to low temperature found that some biofilm genes were up-regulated as well as *ymgB*, but in mutant studies they determined that unlike the biofilm *csg* genes, *ymgB* was expressed independently of the RpoS and DsrA regulators (White-Ziegler *et al.* 2008). However, though the exact function of *ymgB* is unknown, it may code for a regulatory protein of acid resistance in *E. coli*. This gene is also known to suppress biofilm formation, and has been linked to acid resistance (Lee *et al.* 2007). Lee and colleagues (2007) found that in *ymgB* mutants, biofilm formation was significantly increased and hypothesize that its mode of action is probably to suppress indole production thereby increasing the cell motility. It is worth noting that *ymgB* is also linked to acid-resistance in *E. coli* since it

behaves similarly to other acid resistance genes and may even be a regulator of acid-resistance (Lee *et al.* 2007).

Finally, *gadA* is also known to be part of *E. coli*'s acid resistance genes. It is also thought to be RpoS dependent and is even situated in an "acid resistance island" in *E. coli*'s genome which may have secondary regulators besides σ^S (Weber *et al.* 2005). So, all three genes differentially expressed by both strains are involved in the organism's response to stress. However, they are all involved in the response to different aspects of stress and have not previously been linked to each other. It is interesting that only these three, relatively unstudied genes are the ones most highly differentially expressed by both the lab and environmental strains of *E. coli* used in this study.

As to the genes differentially expressed by K12 at the cut off $\log_2 2$ and differentially expressed to a lower level by E43, *bssS/yceP* is also implicated in the regulation of biofilm formation. It is very likely that *bssS/yceP* is a global regulator of the cell through several mechanisms involving indole, cAMP and Autoinducer-2 (AI-2) (Domka *et al.* 2006). Researchers also suspect that *bssS/yceP* affects the *rpoS* gene since it was repressed by 3.7 fold in the *yceP* mutant but the details of the relationship between these regulators needs more research (Domka *et al.* 2006). Little is known about the role of *ymgA* but it is probably also involved in one part of the *E. coli* stress response and biofilm formation because of its link to *ymgB* and since several of the genes already in this study have been linked to the stress response and biofilm formation. In the same way, the function of *ymgC* is unknown; however, it is repressed by the quorum sensing gene AI-2, indicating that it is also linked to biofilm formation (DeLisa *et al.* 2001). The exact function of *Cnu/ydgT* has yet to be determined except that it is known to interact

with H-NS (Paytubi *et al.* 2004) and it is almost certainly involved in gene regulation of the *oriC* locus where replication of the bacterial DNA chromosome begins (Kim *et al.* 2005). H-NS are the histone-like proteins that are involved in condensation of the bacterial chromosome as well as transcriptional control of several genes (Ussery *et al.* 2001). Except for *Cnu/ydgT* all the genes up-regulated by both strains are involved in either biofilm formation or in a stress response.

Of the genes repressed by both strains at the lower temperature, *ompT* is likely an outer membrane protease (Table 3.1). That these would be down-regulated makes some sense because we have already determined that the bacteria should be responding to some stress. Also since they are entering stationary phase, it is likely that the cells would not need as many porins and would be dying or shifting gene expression towards a more sessile state. Finally, *ibpB*, is known to be a heat-shock chaperone so it is no surprise that this gene is repressed at lower temperatures. At this point it is not known why this particular gene is so highly repressed (Table 3.1) and why other genes of this type are not also down-regulated. Conversely, White-Ziegler colleagues (2008) found that *ibpA* was increased in expression at lower temperature (23°C).

3.5.4 Genes differentially regulated at low temperature by each strain

Only 26 genes changed expression levels at 21°C in K12, 10 of which were similarly changed in E43. Of these genes of K12, 9 are involved in the organism's metabolism. Several have to do with the main metabolic functions like energy and amino acid metabolism, but 3 of these genes are also involved in the biosynthesis of outer membrane proteins: *ompT, csgB,* and *csgA*. In fact, the entire *csg* family except for *csgG* was increased in expression at 21°C (Table 3.1). *CsgA* and *csgB* code for curli amyloid fibers

which are part of the extracellular matrix required for biofilm formation and the rest of the *csg* genes are involved in the regulation of *csgA* and *B* (Barnhart and Chapman 2006, Wang *et al.* 2007). *CsgD* is likely the main transcriptional activator of *csgA* and *B* but has been shown to have other transcriptional functions possibly unrelated to the *csg* family such as regulation of the *yhiU* operon which has to do with the O-antigen of lipopolysaccharides (Gualdi *et al.* 2007). It is unknown why these genes should activate at 21°C but not at 37°C when both cultures were at the same cell density. It is possible that the longer time required for the cells to reach the required cell density allowed these cells grown at 21°C to express these biofilm related genes. Similarly, White-Ziegler and co-workers (2008) found that several biofilm genes were up-regulated including the *csg* family. In light of all the similar data reported on studies of *E. coli* K12 and its closely related lab strains, it is surprising to find there are no *csg* genes differentially expressed by the environmental strain.

Other genes of interest are the *narZ,Y,*and *U* which are all increased in expression by strain K12 and are very likely also dependent on RpoS (Weber and Jung 2002b, White-Ziegler *et al.* 2008). Also, *ymgA*, in addition to the aforementioned *ymgB* was also increased in expression and has been implicated in the suppression of biofilm formation but has not been as well-studied as *ymgB* (Lee *et al.* 2007). It can also be assumed that *ymgA* is important for acid resistance in *E. coli* (Lee *et al.* 2007).

It is obvious that very few gene expression levels were changed at the lower temperature by the environmental strain (Table 3.1). While only *gadA* was increased by K12, *gadAB*, *yhiE/gadE,* and *hdeAB* which are also a part of acid-resistance were increased by the environmental isolate. As already stated, *gadAB* are σ^S dependent, but

they were also found to be dependent on the regulators *gadX* and *gadW* as conditions moved from stationary phase to more acid-stressed conditions (Weber *et al.* 2005). *GadE* is also a regulator of the "acid fitness island" (Bergholz *et al.* 2007). Since neither *gadX*, nor *gadW* were detected, and sampling time was at early stationary phase it can be assumed that the *gad* genes expressed by E43 were still under the influence of RpoS in our study. It is also worth mentioning that the *gad* genes are not found in other *Enterobacteriaceae* or in *Salmonella enteric* (Bergholz *et al.* 2007). *HdeAB* are known to be acid stress chaperones since mutant strains of each are less fit under acidic conditions (Kern *et al.* 2007).

3.5.5 Genes differentially expressed between K12 and E43

Perhaps what is most interesting about the genes differentially expressed between the two strains is the sheer number of them. A look at the total number of the genes expressed differently by these two strains shows that 7.96 % and 14.26 % of the genes are expressed differently at 37°C and 21°C, respectively (Fig. 3.3). Furthermore, our analysis showed that a significant portion (140) of the genes are differentially expressed at both temperatures (Appendix B, Table B1). These genes are probably differentially expressed no matter the environmental conditions. A reasonable deduction from this evidence is that a certain percentage of the genes of these two strains are either different in sequence or regulation. This implies that these two strains while being the same species, as supported by the sequencing of the small subunit rRNA, are at the very least reacting differently to their environments.

Of the types of genes differentially expressed between these two bacteria, there is no one pattern that stands out. At 37°C there are 14 genes of the O-antigen of LPS family

that are expressed at a lower level in E43 versus K12 and 18 genes of the same family at 21°C. It makes sense that these genes are differentially expressed between the two strains since the O-antigen is known to differ between strains. It is also obvious that the genes would be suppressed or not expressed in E43 since the microarray chip we used was designed from *E. coli* K12's genome. This may make an interesting target for BST since these genes are only present in gram-negative bacteria. Given that there are several LPS O-Antigen genes in each strain (*htrL*, *rfaBFJSYZ*, *rfbABX*, and *wbbH*, to list a few), there is potential to create a presence-absence microarray based on these genes. The O-Antigen LPS region is involved in bacteria-host recognition as the O-antigen is highly immunogenic. It is therefore likely that these variable genes have adapted to their host environment, meaning that these genes may be specific to host groups. The greater number of genes would hopefully allow for greater discrimination, having unique patterns for each strain or host type.

Another gene is the ABC region of which the main function is transporting small molecules across the cell membrane. It is interesting because all of the differentially expressed genes, 6 at 37°C and 12 at 21°C, are down-regulated by the strain E43, implying that this region is also variable between strains. This probably means that the genes are not present in the environmental strain or it may mean that they are used in different situations or that the environmental strain does not respond to the temperature stress in the same manner as K12 does.

Otherwise there are 13 differentially expressed genes linked with adaptation to stress at 37°C and 14 at 21°C suggesting that these two strains regulate the reaction to stress differently. Twelve of these genes are differentially expressed between the strains at both

temperatures which suggests that these genes are different in sequence. It is also noteworthy that since we used short 25bp probes rather than the longer oligos, our results are sensitive to single base pair mutations.

3.5.6 General conclusions

Overall it can be seen that the environmental strain is not identical to *E. coli* K12 in its reaction to low temperature. However it differs not so much in the genes it expresses but in the genes it does not express. This can be seen in both the fact that it differentially expressed very few genes in response to the temperature stress and in the fact that most of its genes were down-regulated when compared to K12. This result suggests that extrapolation of results from model organisms to natural isolates should be made cautiously. However, to confirm this more research on a greater number of environmental isolates would need to be done. Microarrays provide a good platform for future investigations. Our research also revealed that it might be best to find out what genes are always present or absent in *E. coli* strains originating from different hosts. This approach would provide a suite of genes available for a simple presences absence test to identify the host organism of an unknown bacterium.

Chapter 4

Whole genome response of *Escherichia coli* K12 and an environmental isolate in rich and poor media.

4.1 Abstract

We investigated the response of *Escherichia coli* K12 and an environmental isolate, E43, to growth in a low nutrient concentration medium. We grew the bacteria in LB broth and in LB diluted by 1:10 in sterile water (dLB). RNA, isolated during early stationary phase, was hybridized to a commercially available microarray for a comparative analysis of both strains' response to dLB. Only a few genes were differentially expressed (at the $\log_2 2$ threshold 46 by strain K12 and 11 by strain E43). Analysis of genes differentially expressed, such as *prpBCD*, *yjaI/zraP*, and *bfd* revealed that strain K12 was likely dealing with glucose and metal-cofactor deficiencies while the few genes differentially expressed by strain E43 were different from strain K12 and did not reveal any specific metabolic pattern. Between the two strains at the $\log_2 2$ threshold, there were 169 genes differentially expressed in LB broth and 254 genes in dLB broth, of these 119 were differentially expressed under both conditions. We were looking for stress-related genes differentially expressed by both strains, but found that there were more genes differentially expressed between the strains. These genes highly differentially expressed between the strains may be good candidates for a future BST microarray used to differentiate between strains.

4.2 Introduction

While *Escherichia coli* is already a well-studied organism, especially K12 and its mutants, there is yet much to learn about the lifecycle of *E. coli* strains and how they survive in the natural environment. Considering the different diets and eating patterns of

their hosts, different strains of *E. coli* are assumed to have adapted to that host environment and accumulated genetic differences, via mutation, between strains (Scott *et al.* 2002). However, all strains must cycle between a primary habitat (host mammal) and a secondary habitat (natural environment) (Savageau 1983, Jimenez *et al.* 1989, Muniz *et al.* 1989, Hughes 2008) and may respond similarly to stress. It is also possible different strains may vary in their genetic responses to a particular stress or multiple stresses. A variety of studies have been done on the response of *E. coli* strains maintained in artificial laboratory environments to different stresses like heat shock, cold shock, osmotic stress, nitrogen and carbon starvation (Matin *et al.* 1989, Fani *et al.* 1998, Fraley *et al.* 1998, Conter *et al.* 2001, Weber and Jung 2002a, Kabir *et al.* 2004, Nachin *et al.* 2005, Weber *et al.* 2005). There is evidence that bacterial strains have common genetic responses to many stresses (Conter *et al.* 2001, Patten *et al.* 2004, Yamada 2004). The RpoS sigma factor (σ^s) is responsible for the global regulation of many genes involved in the stress response as well as in the transition to stationary phase (Kabir *et al.* 2004, Patten *et al.* 2004, Weber *et al.* 2005). Recently σ^s regulated genes have been associated with stresses like low temperature which was not previously known to induce a sigma-S response (Olsen *et al.* 1993, Conter *et al.* 1997, Sugiura *et al.* 2003, Kabir *et al.* 2004, Weber *et al.* 2005, White-Ziegler *et al.* 2008).

Since *E. coli* are regularly deposited in the natural environment with faeces, they are not typically pathogenic, and because they are easy to detect, *E. coli* are commonly used as indicators of faecal contamination (Rompre *et al.* 2002, Scott *et al.* 2002). Detection of faecal contamination in water is of special concern because faeces may contain several human pathogens (Topp *et al.* 2003). There are already standard methods for detecting

faecal contamination of water (American Public Health Association 1998). However, the standardized detection methods of *E. coli* do not reveal any information on the host source of the bacteria. The goal of Bacterial Source Tracking (BST) is to identify the source of faecal contamination. *E. coli* has, therefore, been the subject of many studies in the field of BST (Scott *et al.* 2002, Simpson *et al.* 2002, Meays *et al.* 2004).

BST studies are focused on developing methods that will help determine the host source of faecal contamination in environmental samples. Determining the host organism aids in eliminating the source of the contamination. For example, on a stream with both cattle ranges and septic fields it would be useful to be able to distinguish between human and bovine contamination of that stream. There are many techniques being applied to BST but no single method stands out as the standard method (Wood *et al.* 1998, Dombek *et al.* 2000, Sabat *et al.* 2000, Carson *et al.* 2001, Scott *et al.* 2002, Hahm *et al.* 2003, Soule *et al.* 2006, Vantarakis *et al.* 2006). One potential method that could be used, but has not been widely studied in BST is microarray technology. An advantage of microarrays is the possibility of screening several thousand genes in a single test. However, relatively few genes are currently used to differentiate between strains of *E. coli* so it may be advantageous to identify more genes that can be used to differentiate between strains of *E. coli*.

When *E. coli* transitions from its primary mammalian host habitat to its secondary habitat, it experiences several stresses such as changes in temperature and light and nutrient availability (Meays *et al.* 2005). Especially when transferred to water, one of the major changes *E. coli* experience is a dilution of nutrient concentration such that its growth is limited (Vital *et al.* 2008). *E. coli* must be adapted to survival in such

conditions in order to re-colonize its primary habitat. Since this is a common occurrence for many *E. coli* strains, they may respond with the same genetic response to this stress. An analysis of different strains responses to this stress may reveal genes that can be used to differentiate between the strains on a BST microarray.

To our knowledge there is has been no previous research exploring the genetic response of *E. coli* to a low concentration of a rich culture medium. In this paper we used Luria Bertani broth diluted by 1:10 in water (dLB) to simulate a potential environmental condition that *E. coli* may encounter in the outdoors where there are a dilute concentration of a complex mixture of nutrients. Using *E. coli* K12 as a control, we looked at the response of a strain of bovine origin recently taken from the natural environmental isolate, designated 43(C)-4A or E43. Using a recently isolated environmental strain allowed us to compare its physiological response to an *E. coli* strain maintained for 86 years in a controlled laboratory environment (Riley *et al.* 2006). We used two different strains to determine which genes are differentially expressed by both bacteria and could be used in a potential microarray designed to be used in BST. We used commercially available microarrays from Affymetrix to determine if *E. coli* E43 would differentially express the same genes as *E. coli* K12 MG1655 in response to dLB.

4.3 Methods and materials

4.3.1 Strains, sequencing, and culture growth

The strains used in this study were pure cultures (See 2.3.1) of *E. coli* K12 MG1655 (ATCC 700926) and a bovine isolate of *E. coli*, 43(C)-4A or hereafter E43, cloned and isolated from the environment on June 30, 2004. The small subunit ribosomal DNA was PCR-amplified and sequenced as previously described (See 2.3.4). Cultures were grown

in batches with agitation (250 rpm) at 37°C. All maintenance culture growth was done in LB broth (2 parts BD-Triptone, 1 part BD-Yeast, 2 parts NaCl) (Becton-Dickson, Oakville, ON, Canada) broth. Experimental culturing was done in LB broth or in LB diluted by 1:10 in sterile distilled water (dLB).

Growth curve analysis was done on both strains of *E. coli* to determine when to sample for RNA extraction (during stationary phase). Growth rates were calculated in Microsoft Excel 2003 (Microsoft, Mississauga, ON, Canada) by using a standard curve of viable cell concentration and Optical Density (OD) (See 2.3.2).

4.3.2 RNA extraction

For microarray analysis, cultures were harvested when they reached an OD_{600} of 0.4 to 0.45 (stationary phase, Fig. 2.1c). RNA was extracted as described in 2.3.6. RNA quantity and quality were checked by measuring on the UV/Vis spectrophotometer and gel electrophoresis. RNA extract was stored at -80°C for future experimentation.

4.3.3 Preparation of cDNA from total RNA

Preparation of cDNA followed the recommended procedure by the microarray manufacturer (Revision 5, Affymetrix, Santa Clara, CA, USA) as described previously (See 2.3.7). Amplification of mRNA into cDNA was performed on 10 µg of total RNA (Gadgil *et al.* 2005). All incubation periods were performed in a Px2 thermal cycler (Thermo-Hybaid, Ashford, UK). The amount of RNA produced was quantified by UV spectrophotometry. Only reactions with yields greater than 2µg of cDNA were used in subsequent procedures. The cDNA was labelled using GeneChip® DNA labelling reagent biotin (Affymetrix, Santa Clara, CA, USA) (See 3.2.7).

4.3.4 Hybridization to the microarray

Biotinylated cDNA was sent to the WRC (UBC, Vancouver, Canada), where the cDNA was hybridized to a GeneChip® *E. coli* Antisense genome array as described in 2.3.8. On the array, each ORF and IG is represented by 15 perfect match (PM) probes, about 25 bp in length. Each PM probe is accompanied by a mismatch probe (MM) that has a single base pair mismatch at position 13 meant to detect the level of non-specific hybridization. Arrays were scanned using Affymetrix Microarray Suite 5.0 (Suite 5.0) on an Agilent GeneArray® Scanner (Affymetrix, Santa Clara, CA, USA) (See 2.3.8).

4.3.5 Microarray data analysis

The data for all genes were analyzed on Suite 5.0, provided by Affymetrix at the WRC (UBC, Vancouver, Canada). The array data were analyzed and compared according to the "Expression Analysis, Data Analysis Fundamentals" manual (Affymetrix 2002) (See 3.3.5 and 2.3.9). Four slides for each condition were used, representing 4 independent biological replicates.

Comparisons between arrays were done on a probe-to-probe basis as described in 3.3.5. The baseline array was always the array representing the bacteria grown at 37°C in LB medium or in the case of comparisons between strains, K12 was the control.

Data were exported into Microsoft Excel spreadsheets and all future analysis was done in Microsoft Excel 2003 (Microsoft, Mississauga, ON, Canada) as in 2.3.9. Each experimental array was compared to each control array for a total of 16 comparisons and then the comparison values were averaged. Standard deviations were calculated on the basis of these comparison values. An outline of the comparisons made can be seen in Fig. 4.1.

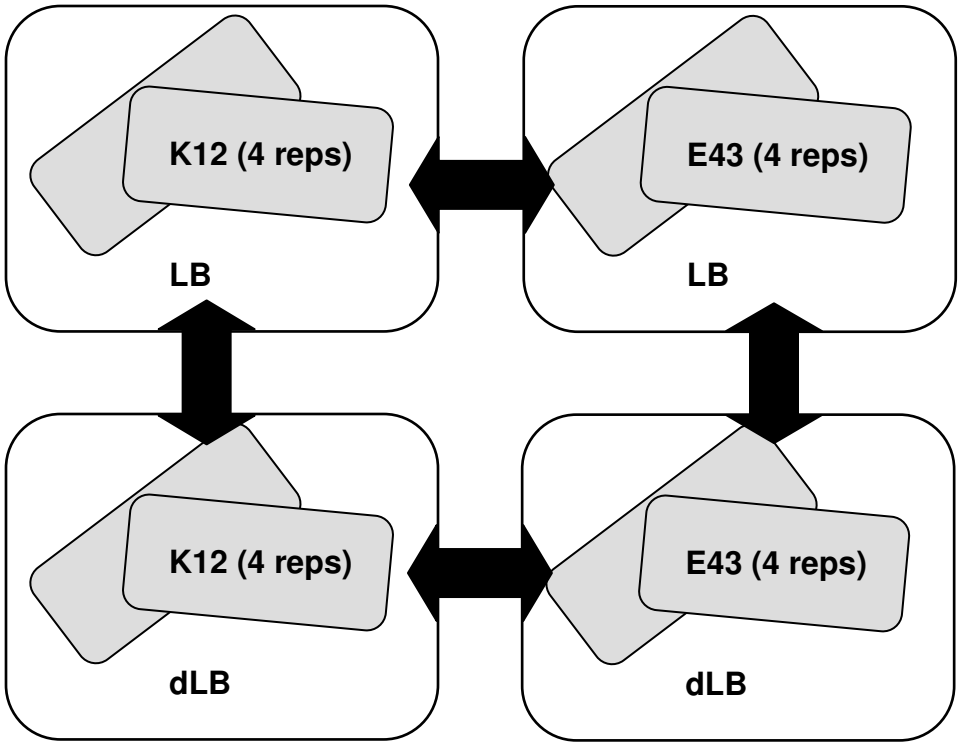

Figure 4.1 Schematic of the experimental design used to compare the transcriptional profiles of *Escherichia coli* K12 and E43.

4.4 Results

4.4.1 General trends

For comparisons between LB and dLB, growth in LB broth was the baseline or control condition. When the two strains genetic profiles were compared to each other, *E. coli* K12 was the baseline array. As can be seen in a comparison of Fig. 4.2 there were more genes differentially expressed between the two strains of *E. coli* than there were differentially expressed by either strain in response to dLB. In the comparison between growth in dLB versus LB broth, at a $\log_2$-ratio of 2 to 3.9 K12 changed the expression of 58 genes and IG regions (Fig. 4.2a). At the same level E43 changed 16 genes and IG regions (Fig. 4.2a). Predictably as the magnitude of the differential expression increased, the number of genes/IG regions altered in expression decreased in an almost asymptotic pattern. Except for a few outliers (eg. *csgE* with a $\log_2$-ratio of 8 in K12), the highest magnitude of differential expression for both strains was $\log_2 4$. Overall K12 differentially transcribed 310 (4.24% of 7312) genes and E43 altered the expression of 84 (1.15% of 7312) genes (Fig. 4.2a).

By contrast, when gene expression was compared between the strains there were a total of 582 (7.96%) differentially expressed genes in LB and 586 (8.01%) differentially expressed in dLB out of the total 7312 probes on the array (Fig. 4.2b). In both LB and dLB greatest number of genes differentially expressed was in the 2-fold ($\log_2$-ratio 1) category. As the magnitude of the $\log_2$-ratio increased the number of differentially expressed genes decreased. In each medium condition there was one gene differentially expressed at $\log_2$-ratio 10 (1024-fold), the highest magnitude of differential expression

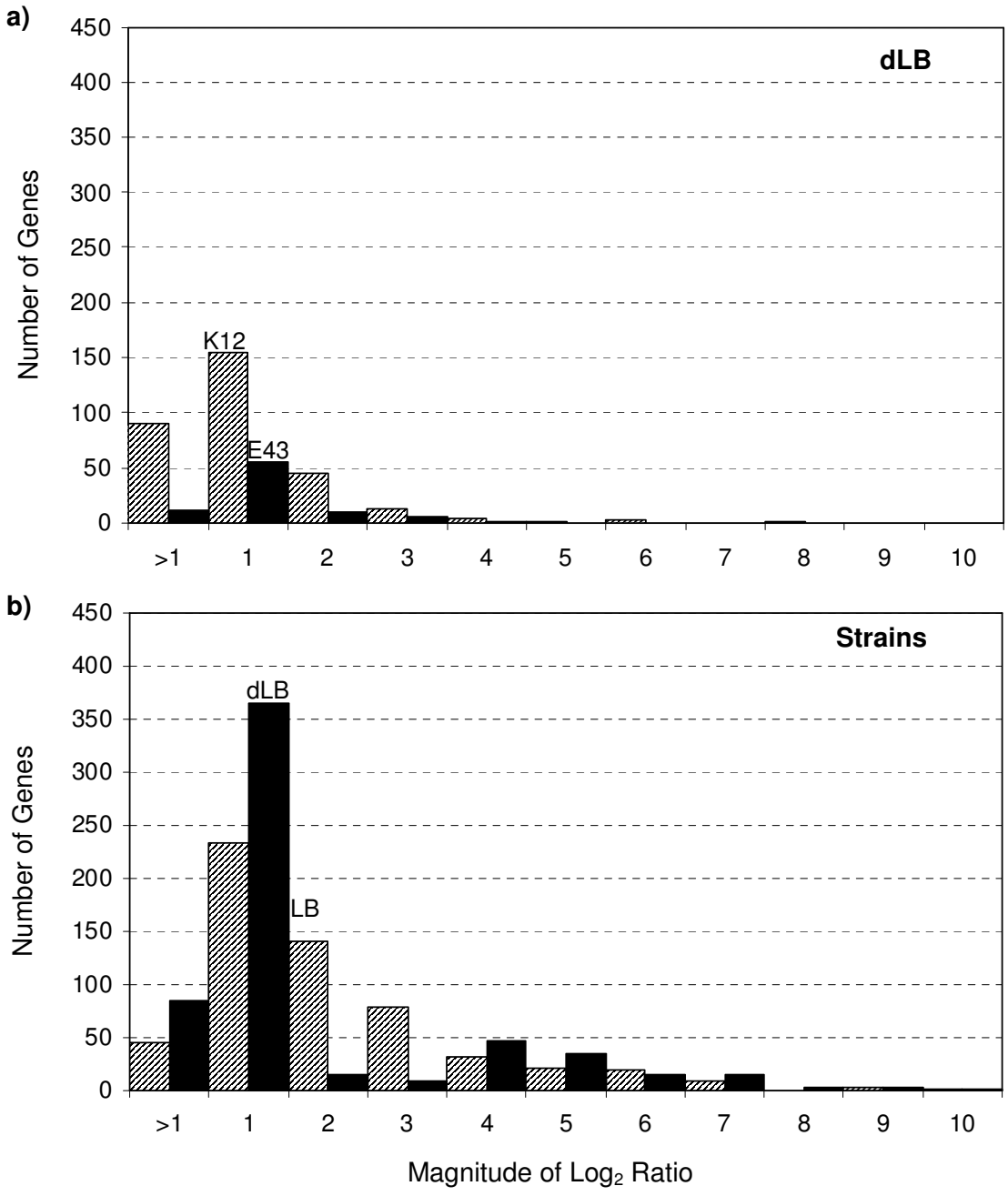

Figure 4.2 Distribution of genes and IG regions differentially transcribed **a)** by *Escherichia coli* K12 (hatched) and E43 (solid) in response to growth in dLB broth, **b)** between K12 and E43 in LB broth (hatched), and in dLB broth (solid). Cultures were grown at 37°C to early stationary phase with aeration at 250rpm. RNA extract was hybridized to a GeneChip® *E. coli* Antisense microarray.

(Fig. 4.2b). However, in dLB there was an irregular pattern from the expected number of genes in each category because there were fewer genes differentially expressed in the groups of $\log_2$-ratio 2 and 3 than there were in $\log_2$-ratio 4, 5, 6, and 7.

4.4.2 Functional classifications

Since the majority of IG regions do not have functional classifications, they were excluded from the subsequent analysis. Also, because we considered those genes that are highly differentially expressed to be of greatest interest for our study, we limited ourselves to a change in expression of 4-fold ($\log_2$-ratio of 2) and higher making the $\log_2$2-ratio our threshold. These choices significantly reduce the number of genes used in the analysis. Under these criteria, there were 46 and 11 genes differentially expressed in dLB by K12 and E43, respectively (Fig. 4.3a; Table 4.1). For both strains the majority of these genes fall into the "Metabolism" or "Unknown" categories. K12 had 28 % of differentially expressed genes in "metabolism" and 22 % of genes in "unknown", while E43 had 55 % of genes in "metabolism" and 18 % in "unknown". While K12 had at least one gene in every functional category, E43 did not have any genes in 3 of the 8 categories due to the fact that so few genes were differentially expressed by strain E43. Overall there were 35 more genes differentially expressed by K12 than E43 (Fig. 4.3a).

When comparing genes between strains, there were 169 genes differentially expressed in LB and 254 in dLB at the 4-fold threshold (Fig. 4.3b). The category with the greatest number of genes was "Metabolism" in dLB at 46 genes (35 % of differentially expressed genes) with "unknown" having 31 % of genes. However in LB broth, the greatest number of genes, 59 (35 %), were in the "Unknown" category, with

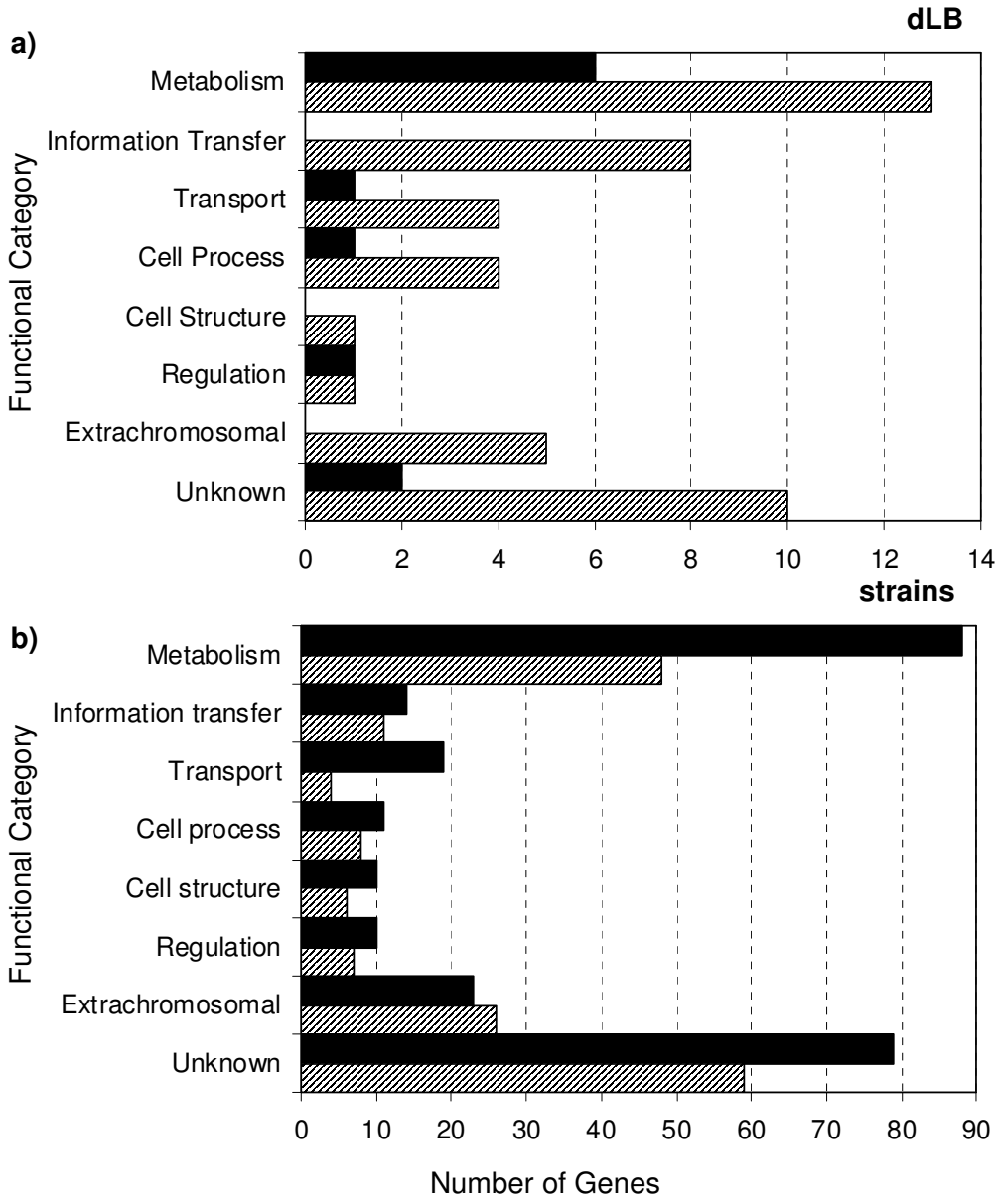

Figure 4.3 Functional categories of genes differentially expressed by **a)** *Escherichia coli* K12 (hatched) and E43 (solid) in dLB broth, **b)** between *E. coli* K12 and E43 in LB (hatched) and dLB (solid). Categories were defined according to the ASAP website (http://www.genome.wisc.edu/tools/asap.htm).

Table 4.1 Functions of the genes differentially expressed by *Escherichia coli* K12 and E43 grown in dLB broth as compared to gene expression in LB broth. The magnitude of differential gene expression is the mean of comparisons between 4 replicates.

Gene Name	Blattner ID	Log_2Ratio		Gene Product
		K12	E43	
Cell Process				
crcA	b0622	3.14	---*	palmitoyl transferase for Lipid A
ftnA	b1905	-2.41	---	ferritin iron storage protein (cytoplasmic)
ibpA	b3687	-2.75	-2.08	heat shock chaperone
ibpB	b3686	-2.50	-1.80	heat shock chaperone
Cell Structure				
ybhG	b0795	2.24	---	predicted membrane fusion protein (MFP) component of efflux pump, membrane anchor
Information Transfer				
csgD	b1040	6.61	---	DNA-binding transcriptional regulator of adhesion determinants
evgA	b2369	-2.10	-1.37	DNA-binding response regulator in two-component regulatory system with EvgS
fhuF	b4367	2.47	---	ferric iron reductase involved in ferric hydroximate transport
osmE	b1739	-2.45	---	DNA-binding transcriptional activator
pspA	b1304	-3.01	-1.47	regulatory protein for phage-shock-protein operon
pspB	b1305	-2.33	---	DNA-binding transcriptional regulator of psp operon
pspC	b1306	-2.37	---	DNA-binding transcriptional activator
soxS	b4062	-2.24	---	DNA-binding transcriptional dual regulator
Metabolism				
aceA	b4015	1.87	2.98	isocitrate lyase
aceB	b4014	1.09	2.21	malate synthase A
ansB	b2957	-2.99	---	periplasmic L-asparaginase II
csgA	b1042	4.00	---	cryptic curlin major subunit
csgB	b1041	4.10	---	curlin nucleator protein, minor subunit in curli complex
cysN	b2751	2.23	---	sulfate adenylyltransferase, subunit 1
dppA	b3544	2.53	1.89	dipeptide transporter -!- periplasmic-binding component of ABC superfamily
manZ	b1819	---	2.25	mannose-specific enzyme IID component of PTS
proV	b2677	-3.43	-3.32	glycine betaine transporter subunit -!- ATP-binding compoent of ABC superfamily
proX	b2679	-3.36	-2.98	glycine betaine transporter subunit; periplasmic-binding component of ABC superfamily

*gene not differentially expressed by this strain

Table 4.1 Continued.

Gene Name	Blattner ID	Log_2Ratio		Gene Product
		K12	E43	
Metabolism				
prpC	b0333	2.63	---	2-methylcitrate synthase
prpD	b0334	2.73	---	2-methylcitrate dehydratase
tdcA	b3118	-2.32	---	DNA-binding transcriptional activator
ygeW	b2870	-4.83	-3.47	conserved protein
ygeY	b2872	-5.02	-3.12	predicted peptidase
ygjG	b3073	2.55	---	putrescine:2-oxoglutaric acid aminotransferase, PLP-dependent
csgG	b1037	3.84	---	outer membrane channel lipoprotein
yjaI / zraP	b4002	-4.64	-4.16	Zn-binding periplasmic protein
Extrachromosomal				
cspI	b1552	2.05	---	Qin prophage; cold shock protein
nmpC	b0553	-2.22	---	DLP12 prophage; truncated outer membrane porin (pseudogene)
pspD	b1307	-2.17	---	peripheral inner membrane phage-shock protein
pspE	b1308	-2.63	---	thiosulfate:cyanide sulfurtransferase (rhodanese)
ynfN	b1551	2.98	---	Qin prophage; predicted protein
Unknown General Classification				
fecI	b4293	2.22	---	KpLE2 phage-like element; RNA polymerase, sigma 19 factor
rcnR	b2105	-2.23	---	transcriptional repressor of rcnA
uspF	b1376	-2.23	---	stress-induced protein, ATP-binding protein
xdhA	b2866	-2.10	---	xanthine dehydrogenase, molybdenum binding subunit
ycgF	b1163	2.24	---	predicted FAD-binding phosphodiesterase
ydcL	b1431	2.59	-0.64	predicted lipoprotein
yddV	b1490	2.46	---	predicted diguanylate cyclase
ygfK	b2878	-2.53	---	predicted oxidoreductase, Fe-S subunit
yjgF	b4243	-2.03	---	ketoacid-binding protein
yodA	b1973	---	-3.38	conserved metal-binding protein
yrbL	b3207	-1.72	-2.79	predicted protein

* gene not differentially expressed by this strain

the second highest category being “Metabolism” at 28 % of differentially expressed genes. Except for the “Extrachromosomal” category, the rest of the categories fewer than 10 % of genes in them (Fig. 3.4b). The "Extrachromosomal" category had 15 % and 9 % of genes in LB and dLB respectively.

4.4.3 Genes highly differentially expressed in diluted LB broth

After eliminating the IG regions and limiting the scope to genes differentially expressed above a $\log_2 2$ threshold, *E. coli* K12 up-regulated 22 genes and down-regulated 24 genes while E43 up-regulated 4 genes and down-regulated 7 genes while growing in dLB broth compared to gene expression in LB broth (Figs. 4.4 and 4.5; Table 4.1). The general categories of these genes have already been discussed and an outline of the more specific functions follows.

Genes of interest that are differentially expressed by K12 are the heat shock proteins *ibpAB*; the *pspABCDE* family which responds to membrane-potential stress, specifically phage shock,; the curlin protein gene family *csgABDEFG*; *prpBCD* of propionate metabolism; and genes involved in metal accumulation and incorporation including *bfd*, *cysN*, *yjal/zraP*, *fecI*, *rcnR*, *ftnA*, *fhuF*, and *ygfK* (Table 4.1; Figs. 4.4 and 4.5). As already noted, E43 differentially expressed far fewer genes above the $\log_2 2$ threshold but did differentially regulate the expression of several metal ion regulatory genes including, *ygeW*, *yjal/zraP*, and *yodA*. In addition *aceA* and *B* were up-regulated in dilute medium.

In dLB there were only 4 genes differentially expressed by both strains of *E. coli* which are *ibpA, proV*, *ygeW*, and *yjal/zraP* (Figs. 4.4 and 4.5; Table 4.1). However, the number of genes differentially expressed by both strains increases to 7 at a lower $\log_2$-ratio and includes *aceAB*, and *yrbL* (Fig. 4.5; Table 4.1). Furthermore, several genes like

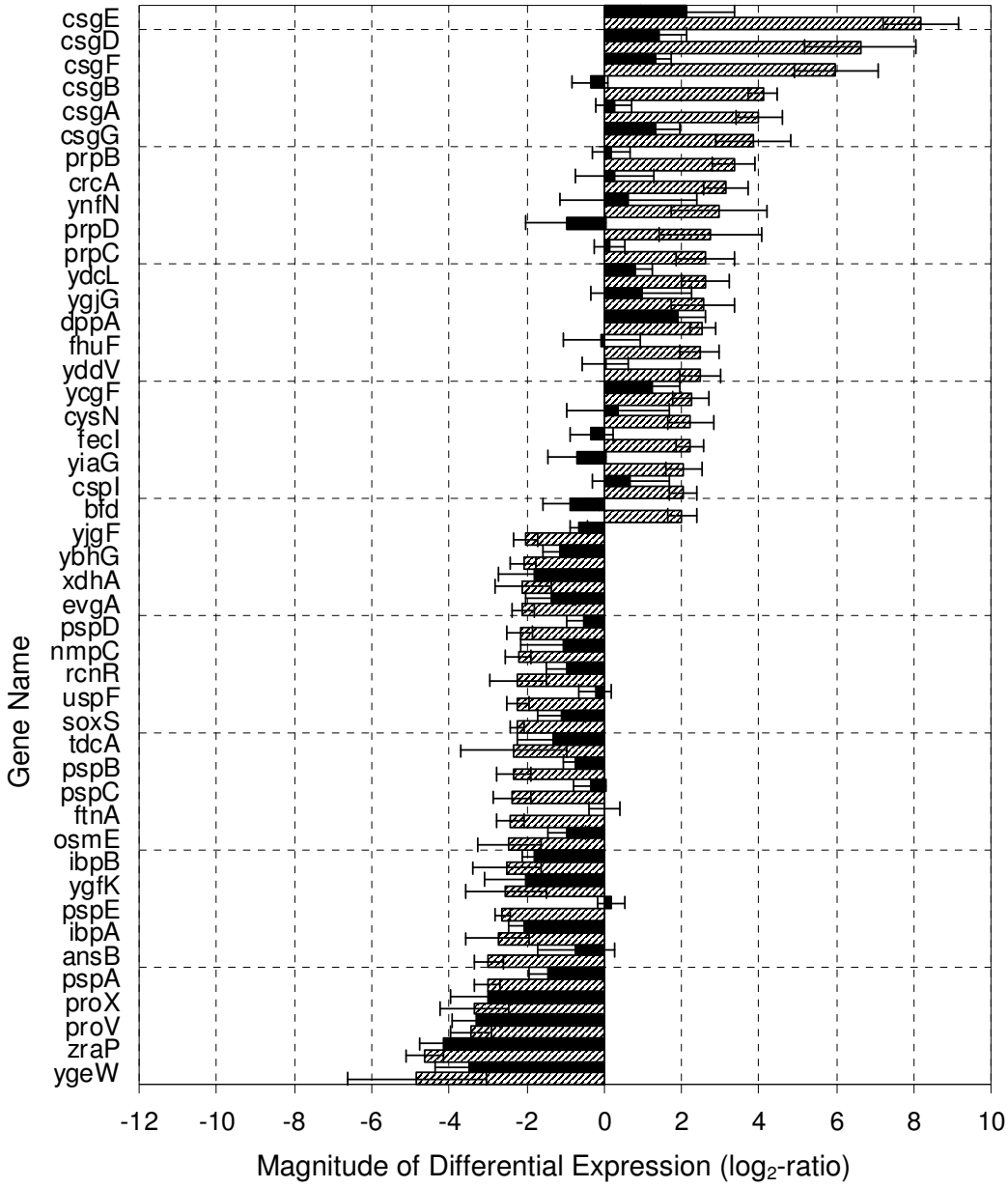

Figure 4.4 Genes differentially expressed by *Escherichia coli* K12 (hatched) above the $\log_2 2$ threshold, in dLB broth as compared to expression in LB broth. Differential expression results for *E. coli* E43 (solid) are shown for comparison. Both strains were grown in batch culture at 37°C with aeration. Error bars represent the standard deviation.

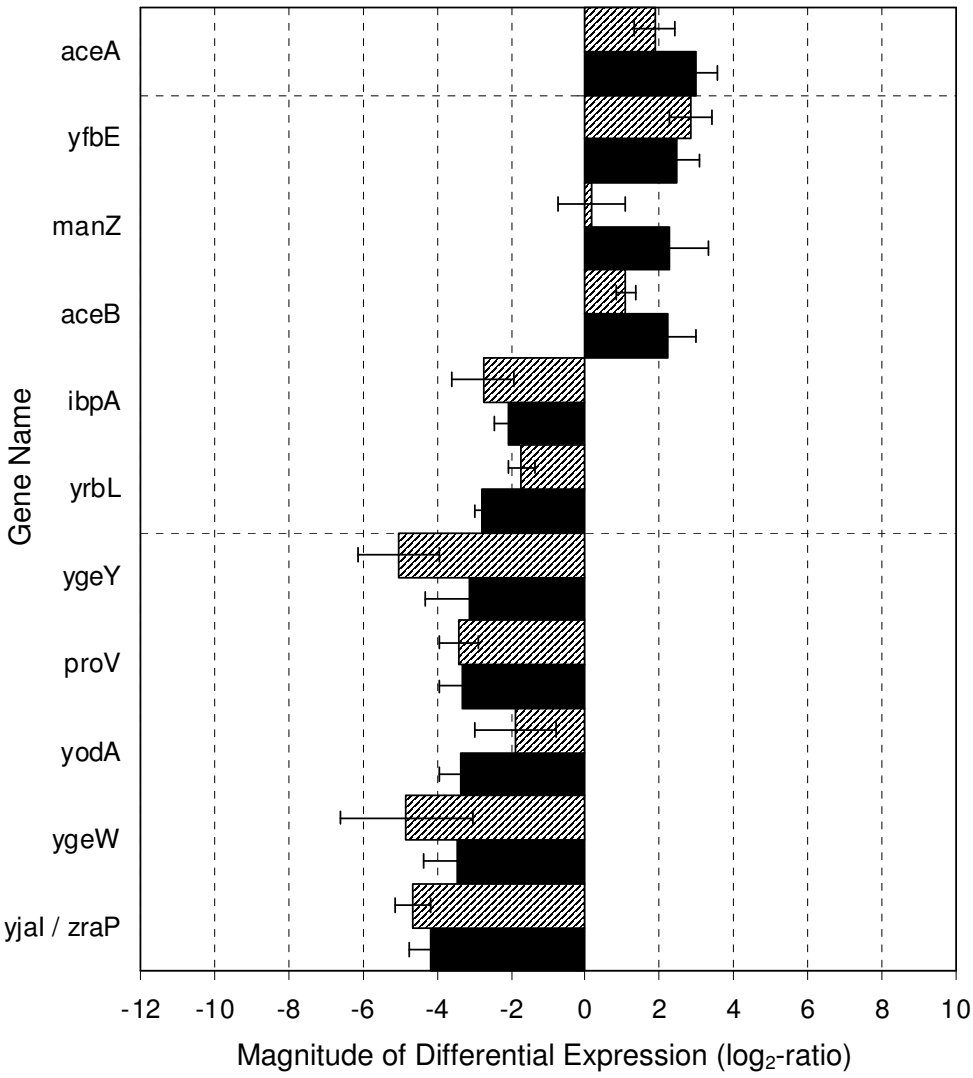

Figure 4.5 Genes differentially expressed by *Escherichia coli* E43 (solid) at the $\log_2 2$ threshold, in dLB broth as compared to expression in LB broth. Differential expression results for *E. coli* E43 (hatched) are shown for comparison. Both strains were grown in batch culture at 37°C with aeration. Error bars represent the standard deviation.

ygeY and *dppA* were originally eliminated from the differentially expressed gene list in one of the two strains because they had one or more “No Change” call. Since these genes were detected as differentially expressed in one strain and have only 1 or 2 calls of 16 comparisons that are "No Change" it is reasonable to infer that these genes are differentially expressed by both strains.

4.4.4 Genes highly differentially expressed between the two strains

At the $\log_2$2-ratio threshold, there were 169 genes differentially expressed by *E. coli* E43 in LB broth (compared to genes expressed by *E. coli* K12), and there were 254 genes differentially expressed by E43 in dLB broth (Fig. 2.3b, Appendix C, Tables C1, C2 and C3). Clearly this was a much higher number than the number of genes differentially expressed by either strain in response to dLB. Unlike each strain’s response to growth in dLB where there were few genes differentially transcribed above a magnitude of $\log_2$4, there were a large number of genes which were differentially expressed above $\log_2$5 between the strains.

Above $\log_2$5, there were 32 differentially expressed genes between K12 and E43 in LB broth and 47 in dLB broth (Figs. 4.6 and 4.7). Of these genes, 29 were differentially expressed under both conditions at a magnitude of $\log_2$5 (Table 4.2). However, of the 48 genes on Table 4.2 only 2 are not detected, at some level, as differentially expressed in LB broth. Similarly, 119 genes were differentially expressed in both conditions at the $\log_2$2 threshold (Table C3, Appendix C) and a considerable number of these genes overlap. The general trend can be seen in Figs. 4.6 and 4.7 where the genes differentially expressed under each nutrient condition (LB or dLB broth) are displayed with the differential expression level of the opposite nutrient condition displayed for comparison:

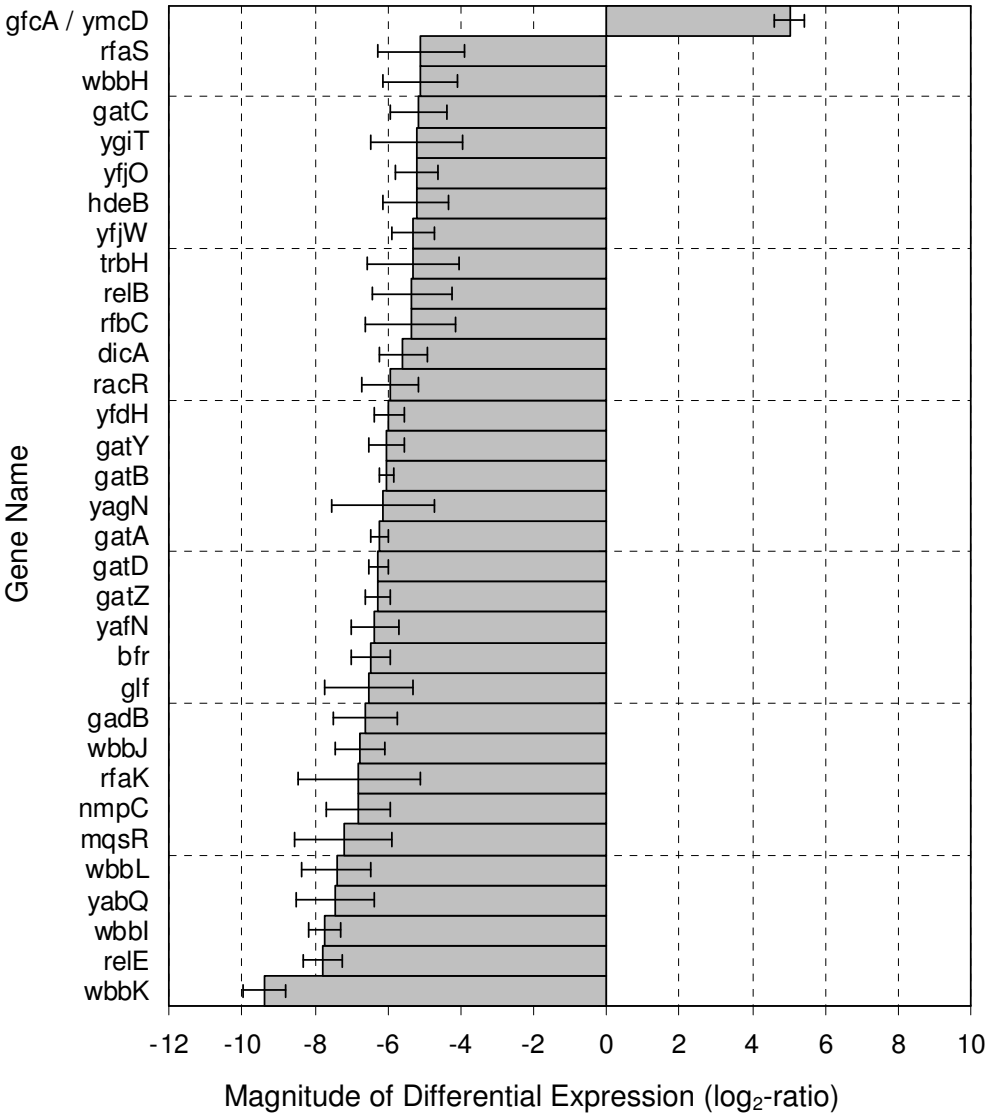

Figure 4.6 Genes differentially expressed between *Escherichia coli* K12 and E43 at a $\log_2 5$ threshold in LB broth. Both strains were grown in batch culture at 37°C with aeration. Error bars represent the standard deviation.

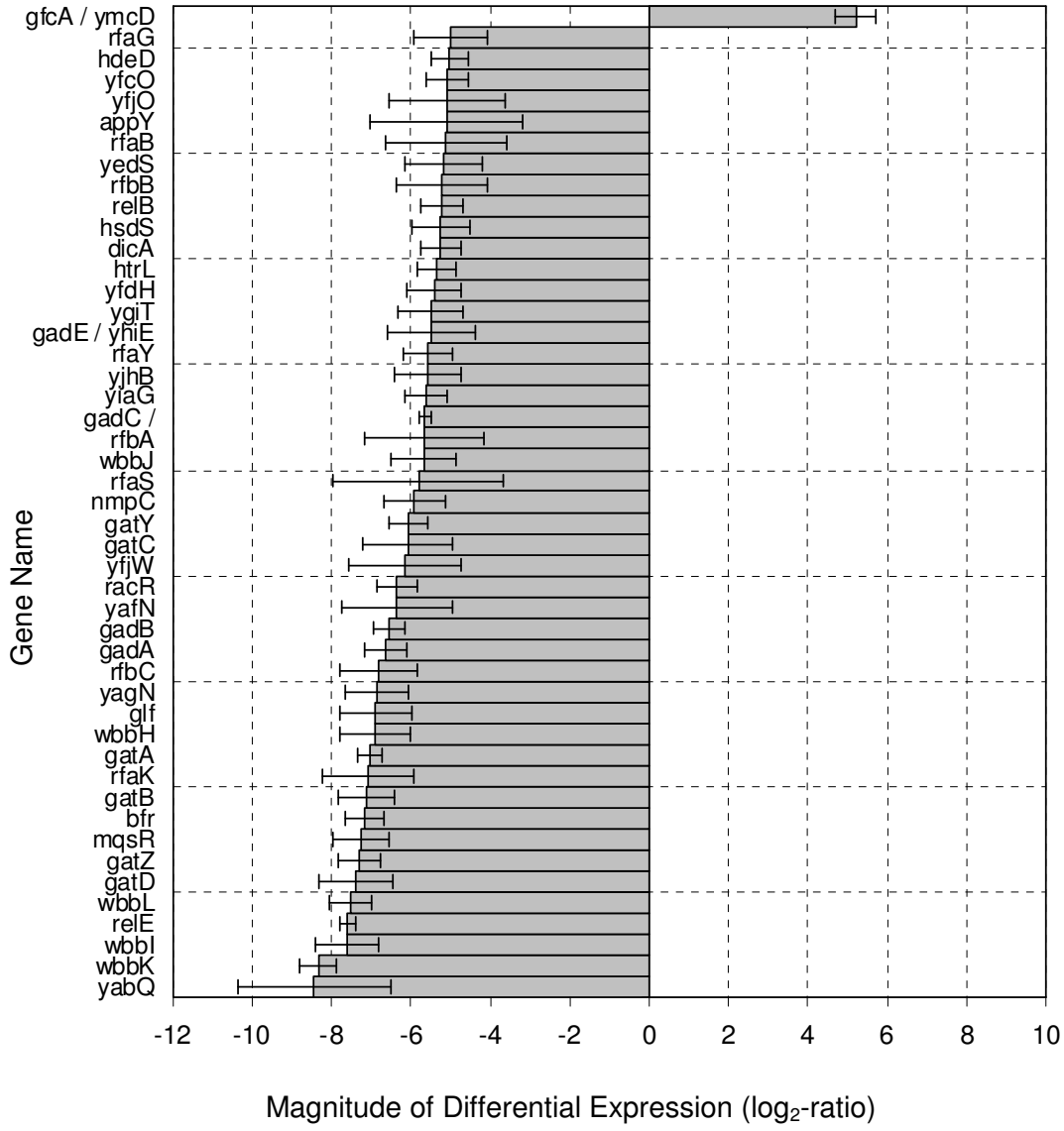

Figure 4.7 Genes differentially expressed between *Escherichia coli* K12 and E43 at a $\log_2 5$ threshold in dLB broth. Both strains were grown in batch culture at 37°C with aeration. Error bars represent the standard deviation.

Table 4.2 Functions of the genes differentially expressed between *Escherichia coli* K12 and E43 grown in LB broth and dLB broth at the $\log_2 5$ (32-fold) level. The magnitude of differential gene expression is the mean of comparisons between 4 replicates.

Gene Name	Blattner ID	$\log_2$ Ratio		Gene Product
		LB	dLB	
Cell Process				
bfr	b3336	-6.49	-7.17	bacterioferritin, iron storage and detoxification protein
Cell Structure				
gfcA / ymcD	b0987	5.03	5.21	predicted protein
yedS	b1964	---*	-5.17	predicted protein, N-ter fragment (pseudogene)
rfaB	b3628	-4.31	-5.11	UDP-D-galactose:(glucosyl)lipopolysaccharide-1,6-D-galactosyltransferase
Information Transfer				
hsdS	b4348	-3.89	-5.24	specificity determinant for hsdM and hsdR
appY	b0564	---	-5.10	DLP12 prophage; DNA-binding transcriptional activator
relE	b1563	-7.79	-7.59	Qin prophage; toxin of the RelE-RelB toxin-antitoxin system
relB	b1564	-5.33	-5.22	Qin prophage; bifunctional antitoxin of the RelE-RelB toxin-antitoxin system and transcriptional repressor
Metabolism				
gadA	b3517	-4.07	-6.63	glutamate decarboxylase A, PLP-dependent
gadB	b1493	-6.62	-6.55	glutamate decarboxylase B, PLP-dependent
gatA	b2094	-6.21	-7.02	galactitol-specific enzyme IIA component of PTS
gatB	b2093	-6.05	-7.12	galactitol-specific enzyme IIB component of PTS
gatC	b2092	-5.16	-6.07	galactitol-specific enzyme IIC component of PTS
gatD	b2091	-6.27	-7.39	galactitol-1-phosphate dehydrogenase, Zn-dependent and NAD(P)-binding
gatY	b2096	-6.04	-6.05	D-tagatose 1,6-bisphosphate aldolase 2, catalytic subunit
gatZ	b2095	-6.29	-7.31	D-tagatose 1,6-bisphosphate aldolase 2, subunit
glf	b2036	-6.52	-6.89	UDP-galactopyranose mutase, FAD/NAD(P)-binding
htrL	b3618	-4.28	-5.35	predicted protein
rfaG	b3631	-3.70	-5.00	glucosyltransferase I
rfaK / waaU	b3623	-6.80	-7.06	lipopolysaccharide core biosynthesis
rfaS	b3629	-5.10	-5.81	lipopolysaccharide core biosynthesis protein

*gene not differentially expressed under this conditions

Table 4.2 Continued.

Gene Name	Blattner ID	Log_2 Ratio LB	Log_2 Ratio dLB	Gene Product
Metabolism				
rfaY	b3625	-4.41	-5.56	lipopolysaccharide core biosynthesis protein
rfbA	b2039	-4.21	-5.66	glucose-1-phosphate thymidylyltransferase
rfbB	b2041	-4.31	-5.21	dTDP-glucose 4,6 dehydratase, NAD(P)-binding
rfbC	b2038	-5.38	-6.81	dTDP-4-deoxyrhamnose-3,5-epimerase
wbbH	b2035	-5.11	-6.90	O-antigen polymerase
wbbJ	b2033	-6.76	-5.68	predicted acyl transferase
Regulation				
yhiE / gadE	b3512	-4.62	-5.50	acid-induced positive regulator of glutamate-dependent acid resistance
ygiT	b3021	-5.20	-5.49	predicted DNA-binding transcriptional regulator
yiaG	b3555	-3.10	-5.62	predicted transcriptional regulator
xasA / gadC	b1492	-4.55	-5.64	predicted glutamate:gamma-aminobutyric acid antiporter
yjhB	b4279	-4.15	-5.59	KpLE2 phage-like element; predicted transporter
Extrachromosomal				
dicA	b1570	-5.58	-5.24	Qin prophage; predicted regulator for DicB
nmpC	b0553	-6.82	-5.91	DLP12 prophage; truncated outer membrane porin (pseudogene)
racR	b1356	-5.93	-6.35	Rac prophage; predicted DNA-binding transcriptional regulator
yagN	b0280	-6.13	-6.84	CP4-6 prophage; predicted protein
yfdH	b2351	-5.97	-5.41	CPS-53 (KpLE1) prophage; bactoprenol glucosyl transferase
yfjO	b2631	-5.22	-5.09	CP4-57 prophage; predicted protein
yfjW	b2642	-5.30	-6.14	CP4-57 prophage; predicted inner membrane protein
Unknown General Classification				
hdeB	b3509	-5.23	-4.68	acid-resistance protein
hdeD	b3511	-3.52	-5.02	acid-resistance membrane protein
msqR	b3022	-7.22	-7.26	quorum-sensing regulator
wbbI	b2034	-7.73	-7.61	conserved protein
wbbK	b2032	-9.38	-8.32	lipopolysaccharide biosynthesis protein
wbbL	b2031	-7.41	-7.52	lipopolysaccharide biosynthesis protein, N-ter fragment (pseudogene)
yabQ	b0057	-7.44	-8.43	C-terminal fragment of a predicted protein (pseudogene)
yafN	b0232	-6.37	-6.35	predicted antitoxin of the YafO-YafN toxin-antitoxin system
yfcO	b2332	-4.15	-5.08	predicted protein

*gene not differentially expressed under this conditions

if a gene is differentially expressed under one condition it is almost always differentially expressed to a greater or lesser extent in the other condition. This can especially be seen in Figs. 4.4 and 4.5 where the bars are almost never similar in length.

The gene products at the $\log_2 5$ threshold can be seen in Table 4.2. The largest group of genes differentially expressed between the bacteria at the $\log_2 5$ level is a cluster of 12 genes involved in synthesis of O-antigen lipopolysaccharide, most of which are in the "Metabolism" category (data not shown). The next largest group is the Phosphotransferase system involving 3 genes in the "Metabolism" category (Table 4.2). At the $\log_2 2$ level there are a few more genes that fall under the O-antigen lipopolysaccharide and Phosphotranferase system. Also some genes differentially expressed in both conditions or in dLB broth alone involved in iron storage and acquisition. So it can be seen that while there are a few groupings that can be made in the genes differentially expressed between the two strains there are a large number of genes involved in divergent processes that do not fall into any known gene expression pattern (Appendix C).

4.5 Discussion

4.5.1 General response to dLB

The two strains tested in this report have different histories. One has been in a lab since 1922 (Riley *et al.* 2006) and the other has an unknown history being isolated from the environment in June, 2004. However, we demonstrated that the two strains do not differentially express the same genes in response to growth in dLB broth. On a growth curve they do appear to grow in a similar manner in dLB, having a lower growth rate and maximum OD versus growth in LB (See 2.4.1). In terms of gene expression, the one

thing the strains seem to have in common is that they do not differentially express many genes while growing in dLB as compared to growth in LB. However *E. coli* K12 does differentially express more genes that E4. Even more interestingly they do not differentially express the same genes in dLB.

4.5.2 Genes differentially regulated by both strains

There were 46 genes differentially expressed by K12 in dLB at the $\log_2 2$ threshold while there were only 11 genes differentially expressed by E43 (Figs. 4.4 and 4.5). In Figs. 4.4 and 4.5 it is shown that the majority of genes differentially expressed by one strain are not differentially expressed by another strain unless the $\log_2 2$ threshold is lowered or the requirement for all "change calls" to be "increase" or "decrease" is relaxed. Originally, we were looking for genes that were highly differentially expressed by both strains in response to an environmental stress in the hopes that these genes would provide sensible marker genes to be used on a microarray designed to differentiate between strains, and therefore host sources, of *E. coli*.

In this experiment we found that only 4 genes were detected as differentially expressed by both bacteria (*ibpA, proV*, *ygeW*, and *yjal/zraP*). None of these genes are obviously linked to one another in terms of function since *ibpA* codes for a heat-shock protein (Allen *et al.* 1992); *proV* is a small ion transporter and is a part of the *proU* operon which is usually increased to deal with high osmolarity (Rajkumari and Gowrishankar 2002, Weber and Jung 2002a); and *yjal/zraP* is a zinc-binding protein which is up-regulated to increase zinc-tolerance (Noll *et al.* 1998); the function of *yegW* is unknown but we hypothesize that it is likely involved in a stress response (Table 4.1). These genes were all repressed by both strains in dLB. It can also be said that *proV*,

ibpA, and *yjal/zraP* are all genes involved in stress response and are down-regulated because they are not needed to respond to this particular stress. For example, the cells are not likely to need more zinc-sequestering proteins in a zinc limiting environment so *yjal/zraP* is repressed (Weber and Jung 2002a).

Two genes that can be said to be up-regulated by both strains are *aceAB*. These genes are up-regulated in response to glucose-limitation which is consistent with the low-nutrient concentration of our experiment and the late sampling time (Maharjan *et al.* 2005). However, Maharjan and colleagues (2005) found that the role of *aceA* was predominantly negative since *aceA* mutants had an increased fitness under glucose limiting conditions. But, the effect of *aceA* was changed from negative to positive if there were mutations on some other part of the genome, indicating that *aceA* might be part of a redundant glyoxylate pathway. There are no other genes as highly up-regulated and involved in the glyoxylate pathway as *aceAB* in our experiment.

4.5.3 Each strain's response to growth in dLB

The genes differentially expressed solely by K12 are more numerous than the genes differentially expressed by E43 (Table 4.1, Fig. 4.4). First of all, the *psp* family which is induced during membrane stress, possibly maintaining membrane potential, is actually down-regulated under this particular stress (Kobayashi *et al.* 2007). It is also known that the *pspF* regulon can be induced by various membrane stresses like extreme heat, or ethanol treatment, and is also found to be induced under biofilm formation. It is interesting to see that group down-regulated in response to low nutrient concentration (Beloin *et al.* 2004, Jovanovic *et al.* 2006). *PrpBCD* are all up-regulated but *prpP* is not differentially regulated, nor is the regulator *prpR* (Table 4.1). However, the operon's

main function is in propionate metabolism, namely for using propionate as the sole carbon source and energy source so the up-regulation of *prpBCD* is evidenced that glucose is lacking in the cells' environment (Lee *et al.* 2005). This is linked to the up-regulation of *aceAB*. It is notable that *prp* genes are not up-regulated in E43, perhaps indicating that it has alternate mechanisms for utilizing alternative carbon sources.

The *prp* group is likely under sigma-54 control and was up-regulated. Also under sigma-54 regulation, *ygjG,* which is known to be induced during nitrogen starvation(Samsonova *et al.* 2003), is up-regulated. A repressed gene that may also be under sigma-54 influence is *xdhA*. This gene is involved in a purine salvage pathway (Xi *et al.* 2000). Based on these data, it is clear that that *E. coli* K12 cells are attempting to deal with several nutrient deficiencies.

Correspondingly, there are several genes involved in iron usage by the cells that are also differentially regulated, including *bfd*, *fecI*, *rcnR*, *ftnA*, *fhuF*, and *ygfK* (Table 4.1). But, the iron metabolism regulon, *fur*, is not detected as differentially regulated which may be due to our sampling in stationary phase since *fur* is typically up-regulated in anticipation of stationary phase (Hantke 2002, Braun *et al.* 2003, Iwig *et al.* 2006). However, the iron storage proteins, *bfd* and *ftnA* are down-regulated, indicating the cells had no further use for iron storage (Table 4.1). Of the genes that are up-regulated *fhuF* has an unclear function but probably helps liberate iron from siderophores (Hantke 2002). Two genes, *rcnR* and *fecI*, are directly or indirectly controlled by the Fur repressor, responsible for iron uptake. Only these two are involved in iron acquisition (Braun *et al.* 2003, Iwig *et al.* 2006).

It is interesting that although cells of both strains when growing in dLB, were sampled when the cells were in early stationary phase (Fig.2.1c), they exhibited few genetic clues that they were going into stationary phase. We did find that the cells, especially K12, exhibit signs of glucose starvation as well as a need for iron and zinc. However, they only seem to be differentially regulating a small portion of the potential genes that could be induced or repressed in response to the stress, especially any RpoS regulated genes. This lack of reaction indicates that these cells may already be in some sort of low-level stasis or that they have been starved for essential nutrients for several hours prior to sampling. But if this is the case, why are there not more genes differentially regulated between LB and dLB broth, since the cells in LB have yet to completely enter the stationary phase and have grown in a rich medium? This question merits further investigation.

A possible avenue of inquiry is the following. A major difference between these two culture conditions is that there are probably more metabolic wastes present in the LB broth than in dLB broth due to a higher cell density in the LB broth (Fig. 2.1c). Therefore, though the cells grew with different concentrations of nutrients, they may be experiencing similar states. In LB broth, there is probably high metabolic waste at sampling time perhaps putting many of the bacteria into a stasis like state. In dLB broth there are low metabolic wastes but fewer nutrients available overall, perhaps also putting the bacteria into a stasis like state at the time of sampling. Sampling at different times during the growth of the bacteria may show more genes differentially expressed between the strains.

It is clear that the environmental isolate, E43, does not highly differentially regulate very many genes in response to low nutrient concentration (Table 4.1, Fig. 4.5). The only gene worth noting here is *yodA* which has been re-named *zinT* by Kershaw (2007) and is also involved in zinc homeostasis (Kershaw *et al.* 2007). This indicates that both strains are dealing with limited metal availability.

4.5.4 Comparison to other stress responses

It is interesting to note that this particular stress does not seem to have many similarities with other stresses. It seems clear that the bacteria, especially K12 is dealing with limited availability of glucose, iron and zinc but there is very little commonality between these differentially regulated genes to *E. coli*'s response to glucose or nitrogen starvation (Matin *et al.* 1989, Fraley *et al.* 1998, Conter *et al.* 2001, Kabir *et al.* 2004). We found that a few genes were differentially expressed in response to both low temperature (21°C) and to dLB. The curli amyloid proteins *csgABDEFGI*, were up-regulated and the porin *nmpC*, and heat-shock protein *ibpB* were repressed (Tables 3.1 and 4.1). The curli amyloid protein has been associated with other stresses like biofilm-formation and low temperature (Barnhart and Chapman 2006, Wang *et al.* 2007). Since all these genes are involved in some sort of stress response, this hints that they are important for K12 but there is not strong enough evidence supporting that any global regulator is responsible for *E. coli*'s growth and survival under low nutrient conditions.

4.5.5 Differences between the strains

Whatever the differences between these two strains, they still expressed the majority of their genome in a similar manner with only a small percentage of the total genes that were differentially expressed (Fig. 4.2). However, it must be noted that there are a subset

of genes that are differentially expressed between the two strains under each environmental condition (Figs. 4.6 and 4.7, Table 4.2).

As can be seen in Table 4.1 there are a number of genes differentially regulated by *E. coli* K12 that are not differentially regulated by strain E43. Furthermore, we noted that there were a significantly higher number of genes differentially regulated between the strains as compared to the number of genes differentially regulated by either strain in response to growth in dLB broth (Table 4.2 and Appendix C). This is similar to the case of *E. coli* K12 and E43's response to 21°C (Chapter 3) where there also were more genes differentially expressed between the strains than there were in response to growth in 21°C. However, among the genes highly differentially regulated in response to dLB ($\log_2 5$ and higher) it is difficult to see many specific patterns or gene function groupings.

In genes that were differentially expressed by between the strains in both LB and dLB, there were 5 genes related to lipopolysaccharide synthesis which is not unexpected since these genes are known to be variable between strains (Table 4.2). In addition there are 3 genes involved in the sugar-transport phosphotransferase system (PTS), dependent on phosphoenolpyruvate (PEP) (Table 4.2). There are a few genes involved in stress that are differentially expressed between the strains, like *bfr*, *hdeB* and *hdeD*. But there are no other discernable systems. Despite the apparent odd grouping of genes, these may be important candidates for a microarray used to distinguish between strains. Because these genes are so highly differentially regulated it would be easier to detect differences in the environment and it may also indicates sequence differences between the strains, allowing for the use of DNA from an environmental sample rather than the more difficult isolation of envirohmental RNA.

When considering the gene expression profile from a temperature shift (from 37°C to 21°C), we found that there were a number of genes that were differentially regulated between the strains in LB and dLB and also differentially expressed at 21°C (Chapter 3). Table 4.3 shows the genes highly differentially regulated, above $\log_2 5$, between the strains under each condition. While some genes were more highly differentially regulated under one environmental state than the others, there were only 3 genes, *yedS*, *hsdS*, and *appY,* not differentially regulated under all three conditions (Table 4.3). This result suggests that there are a significant number of genes which are differentially regulated between the strains under several growth conditions. To see if these genes are always differentially expressed between strains, it would be necessary to run a microarray experiment using DNA from several of strains to see if the genes are absent in any of them. Also, if any genes were present in all strains it would be important to determine if the genes are differentially expressed between strains under a greater number of growth conditions.

The function of the genes differentially expressed by the strains appear to have no overall metabolic pattern. There are large proportion of genes which can be expected to show differential regulation, such as the aforementioned lipopolysaccharide genes (LPS) which are involved in cell-cell recognition. Of interest to this paper, we also found that some genes involved in stress responses were differentially expressed (see especially the "Unknown General Classification" category of Table 4.3). There are few if any other discernible patterns in the gene function/gene product data. Due to the lack of a pattern such as a metabolic pathway in the genes differentially expressed between the strains, it

Table 4.3 Genes differentially expressed between *Escherichia coli* K12 and E43 grown at 37°C in LB broth (control), at 21°C in LB broth, and in dLB broth at 37°C. The magnitude of differential gene expression is the mean of comparisons between 4 replicates.

Gene Name	Blattner ID	Log_2Ratio 37°C	21°C	dLB	Gene Product
Cell Process					
bfr	b3336	-6.49	-7.00	-7.17	bacterioferritin, iron storage and detoxification protein
Cell Structure					
gfcA / ymcD	b0987	5.03	4.52	5.21	predicted protein
yedS	b1964	---*	-4.84	-5.17	predicted protein, N-terminal fragment (pseudogene)
rfaB	b3628	-4.31	-6.93	-5.11	UDP-D-galactose:(glucosyl)lipopolysaccharide-1,6-D-galactosyltransferase
Information Transfer					
hsdS	b4348	-3.89	---	-5.24	specificity determinant for hsdM and hsdR
appY	b0564	---	---	-5.10	DLP12 prophage; DNA-binding transcriptional activator
relE	b1563	-7.79	-5.35	-7.59	Qin prophage; toxin of the RelE-RelB toxin-antitoxin system
relB	b1564	-5.33	2.43	-5.22	Qin prophage; bifunctional antitoxin of the RelE-RelB toxin-antitoxin system and transcriptional repressor
Metabolism					
gadA	b3517	-4.07	-4.00	-6.63	glutamate decarboxylase A, PLP-dependent
gadB	b1493	-6.62	-2.96	-6.55	glutamate decarboxylase B, PLP-dependent
gatA	b2094	-6.21	-6.75	-7.02	galactitol-specific enzyme IIA component of PTS
gatB	b2093	-6.05	-6.69	-7.12	galactitol-specific enzyme IIB component of PTS
gatC	b2092	-5.16	-6.56	-6.07	galactitol-specific enzyme IIC component of PTS
gatD	b2091	-6.27	-7.53	-7.39	galactitol-1-phosphate dehydrogenase, Zn-dependent and NAD(P)-binding
gatY	b2096	-6.04	-6.27	-6.05	D-tagatose 1,6-bisphosphate aldolase 2, catalytic subunit
gatZ	b2095	-6.29	-6.75	-7.31	D-tagatose 1,6-bisphosphate aldolase 2, subunit
glf	b2036	-6.52	-7.54	-6.89	UDP-galactopyranose mutase, FAD/NAD(P)-binding
htrL	b3618	-4.28	-2.49	-5.35	predicted protein
hyaA	b0972	-3.08	-5.92	-4.29	hydrogenase 1, small subunit
rfaG	b3631	-3.70	-4.57	-5.00	glucosyltransferase I
rfaK / gapC / waaU	b3623	-6.80	-9.76	-7.06	lipopolysaccharide core biosynthesis

* gene not differentially expressed under this condition

Table 4.3 Continued.

Gene Name	Blattner ID	Log_2Ratio 37°C	21°C	dLB	Gene Product
Metabolism					
rfaS	b3629	-5.10	-4.63	-5.81	lipopolysaccharide core biosynthesis protein
rfaY	b3625	-4.41	-6.31	-5.56	lipopolysaccharide core biosynthesis protein
rfaZ	b3624	-4.06	-5.73	-4.86	lipopolysaccharide core biosynthesis protein
rfbA	b2039	-4.21	-6.37	-5.66	glucose-1-phosphate thymidylyltransferase
rfbB	b2041	-4.31	-6.00	-5.21	dTDP-glucose 4,6 dehydratase, NAD(P)-binding
rfbC	b2038	-5.38	-4.66	-6.81	dTDP-4-deoxyrhamnose-3,5-epimerase
wbbH	b2035	-5.11	-4.43	-6.90	O-antigen polymerase
wbbJ	b2033	-6.76	-9.20	-5.68	predicted acyl transferase
wrbA	b1004	-2.37	-10.11	-2.21	predicted flavoprotein in Trp regulation
ybaS	b0485	-2.85	-5.19	-3.43	predicted glutaminase
Regulation					
gatR / rfaL	b3622	-2.93	-9.76	-4.83	O-antigen ligase
yhiE / gadE	b3512	-4.62	-2.82	-5.50	acid-induced positive regulator of glutamate-dependent acid resistance
ygiT	b3021	-5.20	-6.03	-5.49	predicted DNA-binding transcriptional regulator
yiaG	b3555	-3.10	-4.47	-5.62	predicted transcriptional regulator
xasA / gadC	b1492	-4.55	-3.38	-5.64	predicted glutamate:gamma-aminobutyric acid antiporter
Transport					
yjhB	b4279	-4.15	-5.77	-5.59	KpLE2 phage-like element; predicted transporter
Extrachromosomal					
dicA	b1570	-5.58	-6.01	-5.24	Qin prophage; predicted regulator for DicB
intF	b0281	-4.84	-5.24	-3.50	CP4-6 prophage; predicted phage integrase
nmpC	b0553	-6.82	-2.44	-5.91	DLP12 prophage; truncated outer membrane porin (pseudogene)
racR	b1356	-5.93	-2.91	-6.35	Rac prophage; predicted DNA-binding transcriptional regulator
yagN	b0280	-6.13	-7.51	-6.84	CP4-6 prophage; predicted protein
yfdH	b2351	-5.97	-6.09	-5.41	CPS-53 (KpLE1) prophage; bactoprenol glucosyl transferase
yfjN	b2630	-3.49	-9.02	-3.76	CP4-57 prophage; RNase LS
yfjO	b2631	-5.22	-3.97	-5.09	CP4-57 prophage; predicted protein
yfjW	b2642	-5.30	-4.88	-6.14	CP4-57 prophage; predicted inner membrane protein

*gene not differentially expressed under this conditions

Table 4.3 Continued.

Gene Name	Blattner ID	Log_2Ratio 37°C*	21°C	dLB	Gene Product
Unknown General Classification					
hdeB	b3509	-5.23	-2.06	-4.68	acid-resistance protein
hdeD	b3511	-3.52	-2.24	-5.02	acid-resistance membrane protein
msqR / pgaB	b3022	-7.22	-7.94	-7.26	quorum-sensing regulator
wbbI	b2034	-7.73	-7.71	-7.61	conserved protein
wbbK	b2032	-9.38	-7.90	-8.32	lipopolysaccharide biosynthesis protein
wbbL	b2031	-7.41	-2.91	-7.52	lipopolysaccharide biosynthesis protein, N-ter fragment (pseudogene)
yabQ / yabP_2	b0057	-7.44	-3.88	-8.43	C-terminal fragment of a predicted protein (pseudogene)
yafN	b0232	-6.37	-8.65	-6.35	predicted antitoxin of the YafO-YafN toxin-antitoxin system
yahO	b0329	-4.34	-6.85	-4.37	predicted protein
yeaG	b1783	-3.04	-6.11	-3.83	conserved protein with nucleoside triphosphate hydrolase domain
yfcO	b2332	-4.15	-5.12	-5.08	predicted protein
ygaF	b2660	-2.67	-5.96	-2.74	predicted enzyme

is difficult to infer reasons for the differential expression. However, these genes may still be used on a microarray to differentiate between strains since they are so highly differentially expressed. Testing of more strains and conditions would be necessary before any of these genes were used on a DNA microarray in BST.

4.5.6 Conclusions

To our knowledge, there are no published papers using a low concentration rich medium as a stress to test the global gene expression response of *E. coli*. The results of our previous study on *E. coli* K12 and E43's response to low temperature seemed to confirm other research implying that low temperature is an inducer of the general stress response (White-Ziegler *et al.* 2008). This is especially evidenced in the genes linked to biofilm formation and to Rpos sigma factor (Table 3.1). It is important to note that *E. coli* K12 up-regulated *csgABDEFG* in dLB just as it did in LB at 21°C (Tables 3.1 and 4.1) which has been linked to general stress responses like biofilm formation. In general, this particular stress has not been studied and while there have been some studies on survival of *E. coli* cells under starvation conditions they usually focus on starvation of a single nutrient or are quite different in design making direct comparison difficult (Matin *et al.* 1989, Fraley *et al.* 1998, Kim *et al.* 2000, Conter *et al.* 2001, Dubey *et al.* 2003, Kabir *et al.* 2004). It is also evident from our results that the response of K12 to dLB was quite different from its response to other stresses because it differentially expressed some genes responsible for metal regulation.

We showed that the response to several hours of exposure to low nutrient concentration did not affect the growth pattern but did reduce the overall density of the bacteria. We also demonstrated that the two strains do not respond in the same way to the stress and that K12 differentially regulates more genes than the environmental isolate, E43.

It would be interesting to do more studies under low nutrient concentrations, especially during the early middle and late parts log-phase conditions to see what genes

the bacteria are expressing under “rapid” growth in comparison to other bacteria in log phase. The results may reveal that the cells are already coping with depletion of nutrient and perhaps we would see some up-regulation of genes normally associated with stationary-phase in log-phase.

Chapter 5 Conclusion

5.1 Differences between K12 and E43

Overall, we found that more genes were differentially expressed between the two strains in all three conditions: 37°C, 21°C and dLB, than were differentially expressed by the strains in response to either environmental stress (Figs. 3.2 and 4.2). It is expected that different strains will not be perfectly identical (Jimenez *et al.* 1989, Bennett *et al.* 1992, Lenski and Travisano 1994, Leroi *et al.* 1994, Cooper *et al.* 2001, Cooper *et al.* 2003, Porwollik *et al.* 2004). However we did not anticipate that so few genes would be differentially expressed by the environmental strain, E43, in response to either 21°C or dLB. Nor could we predict that the K12 and E43 would differentially express almost entirely different sets of genes in response to the growth conditions used in this experiment. Since our study also demonstrates that the natural isolate E43 does not respond in the same way to stress as the artificially maintained K12 (Fig 3.5, 3.6, 4.5 and 4.6), we suggest that the gene expression of artificially maintained model strains of bacteria such as *E. coli* K12 cannot simply be applied to strains existing in the environment.

Lenski and colleagues demonstrated adaptation of an *E. coli* strain to glucose limitation over 10,000 to 20,000 generations of growth in a glucose-limiting environment (Lenski and Travisano 1994, Cooper *et al.* 2001, Cooper *et al.* 2003). Furthermore, they were able to show that the adaptation was due to genetic mutation (Lenski and Travisano 1994, Cooper *et al.* 2001, Cooper *et al.* 2003). Our results suggest that the two strains of *E. coli*, K12 and E43, are adapted to growth in their respective environments because they differentially express different sets of genes in response to growth at 21°C and in

dLB. As evidence, we determined that there was a subset of genes that were differentially expressed between K12 and E43 under all culture conditions (Table 4.3). In our opinion this further supports differences in gene expression between model strains and strains existing in the natural environment.

A popular theory is that strains exposed to constant conditions will become specialists for those conditions with a corollary decrease in ability to adapt to fluctuating growth conditions (Huey and Hertz 1984, Cooper *et al.* 2001). On the other hand, strains of bacteria that are exposed to constantly changing conditions will become generalists, growing equally well under many conditions. While there is little empirical support for this theory it has not yet been disproved and since strains in the environment do seem to be adapted to different conditions, the theory persists (Lenski and Travisano 1994). It is our suggestion that K12 is specialized for the generally consistent laboratory conditions, while E43 remains a generalist. This helps to explain why E43 does not respond to stress in the same way as K12.

5.2 Stress response genes to be put on microarray for BST

In this study we had stringent requirements for a gene to be differentially expressed as well as a high threshold for a gene to be called differentially expressed ($\log_2 2$ or 4-fold). We found that only one gene, *ibpB*, was differentially expressed by both strains in response to the stress conditions (Tables 3.1 and 4.1) but less stringent analyses may reveal more genes that are differentially expressed by both strains in response to 21°C and dLB. While this gene was differentially expressed at a lower magnitude by E43, in both cases it was repressed in response to the stresses. This means that it would likely be expressed at a low level in the environment and is not a likely candidate gene for use on a

BST microarray unless DNA was the target instead of mRNA. However, the gene *ibpB* would have to be sequenced in more strains of *E. coli* before it could be used to differentiate between them.

The major genes that K12 differentially expressed in response to each stress conditions were the *csgABDEF*, with *csgG* being differentially expressed in dilute medium conditions but not under low temperature. The *csg* or curli amyloid fiber genes have been associated with the stress response before and are linked to the regulation of *rpoS* regulon (Olsen *et al.* 1993, Gualdi *et al.* 2007, White-Ziegler *et al.* 2008). As research into the genetics of *E. coli* continues, a link between biofilms and stress responses is being established and our study supports this link in K12 (Domka *et al.* 2007, Gualdi *et al.* 2007, White-Ziegler *et al.* 2008). This would seem to make the *csg* genes potential marker genes that are up-regulated in response to several stress conditions. However, E43 did not up-regulate these genes in response to 21°C or dLB. Interestingly, the *csg* genes were down-regulated in E43 as compared to expression in K12 under all three conditions, albeit the magnitude of differential expression was lower than the cut-off for genes to be scored as differentially expressed. So, while the genes were not differentially expressed by both strains in response to the stress conditions they were differentially regulated between the strains meaning that they may still be potential marker genes on a BST microarray. Additional research on other environmental strains would be necessary before definitively establishing the *csg* genes as biomarkers.

The other gene differentially expressed by K12 under each stress is *nmpC*, which codes for an extrachromosomal porin protein (Tables 3.1 and 4.1). Unlike the *csg* genes, the *nmpC* gene is repressed in response to both 21° C and dLB. It is probable that this

gene is repressed under stress conditions because the bacteria are preparing for long-term survival and do not need as many porin proteins which facilitate the influx of water (Coll *et al.* 1994, Prilipov *et al.* 1998).

Even if strain E43 is not considered in the analysis, we have found very few genes differentially expressed under two stresses by *E. coli* K12. Due to the dynamic nature of stress responses the case may be that these genes should not be used on a microarray designed to classify strains of *E. coli* into host categories. Since environmental conditions are in constant flux it is probable that gene expression of *E. coli* in the environment is also in flux, making it difficult to consistently isolate the same mRNA. Furthermore, if we have difficulty in finding commonalities in the stress response under well controlled laboratory conditions it will be even more difficult to differentiate between strains of bacteria using microarrays in the environment. This combined with the difficulty of isolating mRNA from environmental samples like soil and water may make mRNA microarrays unfeasible for this particular application. This does not, however, eliminate the possibility of using a DNA microarray (Soule *et al.* 2006).

These results are all the more surprising when it is considered that the concept of global stress regulators in *E. coli* is popular and well supported in the literature (Fani *et al.* 1998, Conter *et al.* 2001, Rajkumari and Gowrishankar 2002, Weber and Jung 2002a, Cheung *et al.* 2003, Sugiura *et al.* 2003, Kabir *et al.* 2004, Nachin *et al.* 2005, Weber *et al.* 2005, White-Ziegler *et al.* 2008). However, we found that neither K12 nor E43 differentially expressed any global stress regulators in response to either stress. One possible explanation of this result is that the many reports that found global stress regulators, like RpoS, studied the transition into stress conditions while we chose a later

sampling time because we wanted to see what genes were differentially expressed during the stress condition, not in the transition. However, there may also be more to be discovered about gene regulation by σ^S. For example, the genes regulated by σ^S in response to extreme temperature a relatively well known but White-Ziegler and colleagues (2008) postulated that even a small decrease in temperature such as a shift to 23°C could be an important factor in triggering the σ^S response. They made this postulation because they found that a number of RpoS-dependent genes were differentially regulated at 23°C and even found several genes under σ^S control that were previously not known to be RpoS-dependent genes (White-Ziegler *et al.* 2008). Therefore, several of the genes found as differentially regulated in our study may be RpoS-dependent, but not yet connected with the σ^S-stress response.

5.3 Genes differentially regulated between the strains

Another finding from our work is the considerable number of genes differentially regulated between the strains (Figs. 3.3 and 4.2). This was true not only under the control condition (in LB at 37°C), but also under both low temperature and low nutrient concentration. In fact, we found that the same genes were differentially regulated under all three growth conditions (Table 4.3).

As to the function of the genes differentially expressed by the strains, there appears to be no clear pattern. There are large proportion of genes which can be expected to show differential regulation, such as the lipopolysaccharide genes (LPS) which are involved in cell-cell recognition. Of interest to this thesis, we also found that some genes involved in stress responses were differentially expressed (see especially the "Unknown General Classification" category of Table 4.3). While 66-67 % of genes differentially

expressed between the strains fall into the function categories of “metabolism” and “unknown general classification”, overall the genes appeared to be a random selection from various pathways showing no clear cluster of genes responding to or regulating a specific process. More environmental isolates would have to be investigated to reveal if there are any genes that are consistently differentially expressed between different strains. Any genes revealed by this process are potential genes to be used on a BST microarray.

5.4 General conclusions

It is estimated that there is a certain proportion of the genome that is different even between closely related strains of bacteria (Porwollik *et al.* 2003). Our report seems to confirm this hypothesis and reveal that not only hypervariable genes like the LPS genes but also more central genes like *gadAB* are differentially expressed between the strains. These differences between strains may be able to provide library independent probes that could be used on a DNA microarray that would be used in the environment to universally classify strains of *E. coli* into point-source categories. Ideally these markers would be present in all strains of one host category but absent in another. Since that scenario is improbable, it would be necessary for any such microarray to have several probes for each host category so that the absence of any one probe does not lead to a false result. To reach this goal, a large number of strains of *E. coli* from a variety of known sources would need to be evaluated by microarrays, using a common stress such as temperature or the shift to stationary phases. This research would also give researchers a better understanding of gene expression in strains which are not well-studied model strains.

It is already established that *E. coli* differentially regulate genes in response to their changing environment and that *E. coli* can survive a considerable number of extreme

stresses for long periods of time. Our report shows that the genes differentially expressed in response to different stress can vary between strain. It would be interesting to determine whether each strain has its own set of adaptations, or if there are a finite number and combination of genetic adaptations. We also do know the extent of the difference between strains. There may be relatively little difference between the genetic sequence of some organisms, but the regulation of the encoded genes may add a further layer of variation. We found that while the two strains under study here were similar in the 16S rDNA sequence they did not differentially regulate the same genes in response to the same stress. As already stated this merits further study. As microarray technology develops and decreases in cost, such studies on multiple environmental strains will be possible, giving microbiologists a better understanding of microbial gene regulation that is not dependent on a few model strains. As the differences and similarities of *E. coli* genes and gene regulation between strains becomes known, increasingly better microarray applications for BST will be developed.

Literature cited

-

Affymetrix. 2002. GeneChip® Expression Analysis, Data Analysis Fundamentals. http://www.affymetrix.com/support/downloads/manuals/data_analysis_fundamentals_manual.pdf.

Albert, J.M., Munakata-Marr, J., Tenorio, L., and Siegrist, R.L. 2003. Statistical evaluation of bacterial source tracking data obtained by rep-PCR DNA fingerprinting of *Escherichia coli*. Environ. Sci. Technol. **37:** 4554-4560.

Allen, S.P., Polazzi, J.O., Gierse, J.K., and Easton, A.M. 1992. 2 novel heat-shock genes encoding proteins produced in response to heterologous protein expression in *Escherichia coli*. J. Bacteriol. **174:** 6938-6947.

Alonso, A., Sanchez, P., and Martinez, J.L. 2001. Environmental selection of antibiotic resistance genes. Environ. Microbiol. **3:** 1-9.

American Public Health Association. 1998. Standard methods for the examination of water and wastewater. 20th ed. American Public Health Association, Washington, DC.

Arfin, S.M., Long, A.D., Ito, E.T., Tolleri, L., Riehle, M.M., Paegle, E.S., and Hatfield, G.W. 2000. Global gene expression profiling in *Escherichia coli* K12 - The effects of integration host factor. J. Biol. Chem. **275:** 29672-29684.

Arnold, C., Metherell, L., Willshaw, G., Maggs, A., and Stanley, J. 1999. Predictive fluorescent amplified-fragment length polymorphism analysis of *Escherichia coli*: High-resolution typing method with phylogenetic significance. J. Clin. Microbiol. **37:** 1274-1279.

Baldy-Chudzik, K., Niedbach, J., and Stosik, M. 2003. Rep-PCR fingerprinting as a tool for the analysis of genomic diversity in *Escherichia coli* strains isolated from an aqueous/freshwater environment. Cell. Mol. Biol. Lett. **8:** 793-798.

Barnes, B., and Gordon, D.M. 2004. Coliform dynamics and the implications for source tracking. Environ. Microbiol. **6:** 501-509.

Barnhart, M.M., and Chapman, M.R. 2006. Curli biogenesis and function. Annu. Rev. Microbiol. **60:** 131-147.

Beloin, C., Valle, J., Latour-Lambert, P., Faure, P., Kzreminski, M., Balestrino, D., Haagensen, J.A.J., Molin, S., Prensier, G., Arbeille, B., and Ghigo, J.M. 2004. Global impact of mature biofilm lifestyle on *Escherichia coli* K12 gene expression. Mol. Microbiol. **51:** 659-674.

Bennett, A.F., Lenski, R.E., and Mittler, J.E. 1992. Evolutionary adaptation to temperature. 1. Fitness responses of *Escherichia coli* to changes in its thermal environment. Evolution **46:** 16-30.

Bergholz, T.M., Tarr, C.L., Christensen, L.M., Betting, D.J., and Whittam, T.S. 2007. Recent gene conversions between duplicated glutamate decarboxylase genes (*gadA* and *gadB*) in pathogenic *Escherichia coli*. Mol. Biol. Evol. **24:** 2323-2333.

Bernhard, A.E., and Field, K.G. 2000a. A PCR assay to discriminate human and ruminant feces on the basis of host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA. Appl. Environ. Microbiol. **66:** 4571-4574.

Bernhard, A.E., and Field, K.G. 2000b. Identification of nonpoint sources of faecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from faecal anaerobes. Appl. Environ. Microbiol. **66:** 1587-1594.

Berthelet, M., Whyte, L.G., and Greer, C.W. 1996. Rapid, direct extraction of DNA from soils for PCR analysis using polyvinylpolypyrrolidone spin columns. FEMS Microbiol. Lett. **138:** 17-22.

Blankenship, R.E. 2001. Molecular evidence for the evolution of photosynthesis. Trends Plant Sci. **6:** 4-6.

Blankenship, R.E. 1992. Origin and early evolution of photosynthesis. Photosynthesis Res. **33:** 91-111.

Blankenship, R.E., and Hartman, H. 1998. The origin and evolution of oxygenic photosynthesis. Trends Biochem. Sci. **23:** 94-97.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., ColladoVides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., and Shao, Y. 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277:** 1453-&.

Boon, N., De Windt, W., Verstraete, W., and Top, E.M. 2002. Evaluation of nested PCR-DGGE (denaturing gradient gel electrophoresis) with group-specific 16S rRNA primers for the analysis of bacterial communities from different wastewater treatment plants. FEMS Microbiol. Ecol. **39:** 101-112.

Braun, V., Mahren, S., and Ogierman, M. 2003. Regulation of the Fecl-type ECF sigma factor by transmembrane signalling. Curr. Opin. Microbiol. **6:** 173-180.

Brokx, S.J., Ellison, M., Locke, T., Bottorff, D., Frost, L., and Weiner, J.H. 2004. Genome-wide analysis of lipoprotein expression in *Escherichia coli* MG1655. J. Bacteriol. **186:** 3254-3258.

Brown, T.A. 2002. Genomes. 2nd ed. BIOS Scientific Publishers Ltd., Oxford, UK.

Carson, C.A., Shear, B.L., Ellersieck, M.R., and Asfaw, A. 2001. Identification of faecal *Escherichia coli* from humans and animals by ribotyping. Appl. Environ. Microbiol. **67:** 1503-1507.

Chang, D.E., Smalley, D.J., and Conway, T. 2002. Gene expression profiling of *Escherichia coli* growth transitions: an expanded stringent response model. Mol. Microbiol. **45:** 289-306.

Cheung, K.J., Badarinarayana, V., Selinger, D.W., Janse, D., and Church, G.M. 2003. A microarray-based antibiotic screen identifies a regulatory role for supercoiling in the osmotic stress response of *Escherichia coli*. Genome Res. **13:** 206-215.

Cho, J.C., and Tiedje, J.M. 2002. Quantitative detection of microbial genes by using DNA microarrays. Appl. Environ. Microbiol. **68:** 1425-1430.

Chuang, S.E., Daniels, D.L., and Blattner, F.R.1993. Global regulation of gene-expression in *Escherichia coli*. J. Bacteriol. **175:** 2026-2036.

Coll, J.L., Heyde, M., and Portalier, R. 1994. Expression of the *nmpC* Gene of *Escherichia coli* K12 is modulated by external Ph-identification of cis-acting regulatory sequences involved in this regulation. Mol. Microbiol. **12:** 83-93.

Conter, A., Menchon, C., and Gutierrez, C. 1997. Role of DNA supercoiling and RpoS sigma factor in the osmotic and growth phase-dependent induction of the gene *osmE* of *Escherichia coli* K12. J. Mol. Biol. **273:** 75-83.

Conter, A., Gangneux, C., Suzanne, M., and Gutierrez, C. 2001. Survival of *Escherichia coli* during long-term starvation: effects of aeration, NaCl, and the *rpoS* and *osmC* gene products. Res. Microbiol. **152:** 17-26.

Cooper, T.F., Rozen, D.E., and Lenski, R.E. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. **100:** 1072-1077.

Cooper, V.S., Bennett, A.F., and Lenski, R.E. 2001. Evolution of thermal dependence of growth rate of *Escherichia coli* populations during 20,000 generations in a constant environment. Evolution **55:** 889-896.

D'Costa, V.M., McGrann, K.M., Hughes, D.W., and Wright, G.D. 2006. Sampling the antibiotic resistome. Science **311:** 374-377.

DeLisa, M.P., Wu, C.F., Wang, L., Valdes, J.J., and Bentley, W.E. 2001. DNA microarray-based identification of genes controlled by autoinducer 2-stimulated quorum sensing in *Escherichia coli*. J. Bacteriol. **183:** 5239-5247.

Desmarais, T.R., Solo-Gabriele, H.M., and Palmer, C.J. 2002. Influence of soil on faecal indicator organisms in a tidally influenced subtropical environment. Appl. Environ. Microbiol. **68:** 1165-1172.

Dick, L.K., and Field, K.G. 2004. Rapid estimation of numbers of faecal Bacteroidetes by use of a quantitative PCR assay for 16S RNA genes. Appl. Environ. Microbiol. **70:** 5695-5697.

Dick, L.K., Bernhard, A.E., Brodeur, T.J., Domingo, J.W.S., Simpson, J.M., Walters, S.P., and Field, K.G. 2005. Host distributions of uncultivated faecal Bacteroidales bacteria reveal genetic markers for faecal source identification. Appl. Environ. Microbiol. **71:** 3184-3191.

Dombek, P.E., Johnson, L.K., Zimmerley, S.T., and Sadowsky, M.J. 2000. Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. Appl. Environ. Microbiol. **66:** 2572-2577.

Domka, J., Lee, J., and Wood, T.K. 2006. YliH (BssR) and YceP (BssS) regulate *Escherichia coli* K12 biofilm formation by influencing cell signaling. Appl. Environ. Microbiol. **72:** 2449-2459.

Domka, J., Lee, J.T., Bansal, T., and Wood, T.K. 2007. Temporal gene-expression in *Escherichia coli* K12 biofilms. Environ. Microbiol. **9:** 332-346.

Doolittle, W.F. 1999. Lateral genomics. Trends Biochem. Sci. **24:** M5-M8.

Dorsch, M., and Stackebrandt, E. 1992. Some modifications in the procedure of direct sequencing of PCR amplified 16S-rDNA. J. Microbiol. Methods **16:** 271-279.

Dubey, A.K., Baker, C.S., Suzuki, K., Jones, A.D., Pandit, P., Romeo, T., and Babitzke, P. 2003. *CsrA* regulates translation of the *Escherichia coli* carbon starvation gene, *cstA*, by blocking ribosome access to the *cstA* transcript. J. Bacteriol. **185:** 4450-4460.

Etchegaray, J.P., and Inouye, M. 1999. CspA, CspB, and CspG, major cold shock proteins of *Escherichia coli*, are induced at low temperature under conditions that completely block protein synthesis. J. Bacteriol. **181:** 1827-1830.

Fani, R., Gallo, R., Fancelli, S., Mori, E., Tamburini, E., and Lazcanoz, A. 1998. Heterologous gene expression in an *Escherichia coli* population under starvation stress conditions. J. Mol. Evol. **47:** 363-368.

Flores, S., Flores, N., de Anda, R., Gonzalez, A., Escalante, A., Sigala, J.C., Gosset, G., and Bolivar, F. 2005. Nutrient-scavenging stress response in an *Escherichia coli* strain lacking the phosphoenolpyruvate: Carbohydrate phosphotransferase system, as explored by gene expression profile analysis. J. Mol. Microbiol. Biotechnol. **10:** 51-63.

Fraley, C.D., Kim, J.H., McCann, M.P., and Matin, A. 1998. The *Escherichia coli* starvation gene *cstC* is involved in amino acid catabolism. J. Bacteriol. **180:** 4287-4290.

Friedman, S.M., Hossain, M., Hasson, T.H., and Kawamura, A. 2006. Gene expression profiling of intrinsic thermotolerance in *Escherichia coli*. Curr. Microbiol. **52:** 50-54.

Fukiya, S., Mizoguchi, H., Tobe, T., and Mori, H. 2004. Extensive genomic diversity in pathogenic *Escherichia coli* and *Shigella* strains revealed by comparative genomic hybridization microarray. J. Bacteriol. **186:** 3911-3921.

Fukushima, M., Kakinuma, K., and Kawaguchi, R. 2002. Phylogenetic analysis of *Salmonella*, *Shigella*, and *Escherichia coli* strains on the basis of the *gyrB* gene sequence. J. Clin. Microbiol. **40:** 2779-2785.

Fukushima, M., Kakinuma, K., Hayashi, H., Nagai, H., Ito, K., and Kawaguchi, R. 2003. Detection and identification of *Mycobacterium* species isolates by DNA microarray. J. Clin. Microbiol. **41:** 2605-2615.

Gadgil, M., Kapur, V., and Hu, W.S. 2005. Transcriptional response of *Escherichia coli* to temperature shift. Biotechnol. Prog. **21:** 689-699.

Gerhardt, P. 1994. Methods for general and molecular bacteriology. American Society for Microbiology, Washington, DC.

Glasner, J.D., Liss, P., Plunkett III, G., Darling, A., Prasad, T., Rusch, M., Byrnes, A., Gilson, M., Biehl, B., Blattner, F.R., and Perna, N.T. 2003. ASAP, a systematic annotation package for community analysis of genomes. Nucleic Acids. Res. **31:** 147-151.

Gonzalez, R., Tao, H., Shanmugam, K.T., York, S.W., and Ingram, L.O. 2002. Global gene expression differences associated with changes in glycolytic flux and growth rate in *Escherichia coli* during the fermentation of glucose and xylose. Biotechnol. Prog. **18:** 6-20.

Gordon, D.M. 2001. Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. Microbiology-Sgm **147:** 1079-1085.

Graumann, P. and Marahiel, M.A. 1996. Some like it cold: Response of microorganisms to cold shock. Arch. Microbiol. **166:** 293-300.

Gualdi, L., Tagliabue, L., and Landini, P. 2007. Biofilm formation-gene expression relay system in *Escherichia coli*: Modulation of sigma(s)-dependent gene expression by the CsgD regulatory protein via sigma(s) protein stabilization. J. Bacteriol. **189:** 8034-8043.

Hagedorn, C., Robinson, S.L., Filtz, J.R., Grubbs, S.M., Angier, T.A., and Reneau, R.B. 1999. Determining sources of faecal pollution in a rural Virginia watershed with antibiotic resistance patterns in faecal streptococci. Appl. Environ. Microbiol. **65:** 5522-5531.

Hahm, B.K., Maldonado, Y., Schreiber, E., Bhunia, A.K., and Nakatsu, C.H. 2003. Subtyping of foodborne and environmental isolates of *Escherichia coli* by multiplex-PCR, rep-PCR, PFGE, ribotyping and AFLP. J. Microbiol. Methods **53:** 387-399.

Hantke, K. 2002. Members of the Fur protein family regulate iron and zinc transport in *Escherichia coli* and characteristics of the Fur-regulated FhuF protein. J. Mol. Microbiol. Biotechnol. **4:** 217-222.

Hartel, P.G., Summer, J.D., Hill, J.L., Collins, J.V., Entry, J.A., and Segars, W.I. 2002. Geographic variability of *Escherichia coli* ribotypes from animals in Idaho and Georgia. J. Environ. Qual. **31:** 1273-1278.

Harwood, V.J., Whitlock, J., and Withington, V. 2000. Classification of antibiotic resistance patterns of indicator bacteria by discriminant analysis: Use in predicting the source of faecal contamination in subtropical waters. Appl. Environ. Microbiol. **66:** 3698-3704.

Herendeen, S.L., Vanbogelen, R.A., and Neidhardt, F.C. 1979. Levels of Major Proteins of Escherichia coli during growth at different temperatures. J. Bacteriol. **139:** 185-194.

Huey, R.B. and Hertz, P.E. 1984. Is a jack-of-all-trades a master of none? Evolution **38:** 441-444.

Hughes, B.S. 2008. Evolutionary patterns of *Escherichia coli* growth in seawater determined with a host to coast environmental laboratory analog. Aquat. Microb. Ecol. **53:** 243-255.

Hughes, T.R., Marton, M.J., Jones, A.R., Roberts, C.J., Stoughton, R., Armour, C.D., Bennett, H.A., Coffey, E., Dai, H.Y., He, Y.D.D., Kidd, M.J., King, A.M., Meyer, M.R., Slade, D., Lum, P.Y., Stepaniants, S.B., Shoemaker, D.D., Gachotte, D., Chakraburtty, K., Simon, J., Bard, M., and Friend, S.H. 2000. Functional discovery via a compendium of expression profiles. Cell **102:** 109-126.

Hung, S.P., Baldi, P., and Hatfield, G.W. 2002. Global gene expression profiling in *Escherichia coli* K12 - The effects of leucine-responsive regulatory protein. J. Biol. Chem. **277:** 40309-40323.

Ikeda, M., Yamaguchi, N., Tani, K., and Nasu, M. 2005. Development of phylogenetic oligonucleotide probes for screening foodborne bacteria. J. Health Sci. **51:** 469-476.

Iwig, J.S., Rowe, J.L., and Chivers, P.T. 2006. Nickel homeostasis in *Escherichia coli* - the *rcnR-rcnA* efflux pathway and its linkage to *nikR* function. Mol. Microbiol. **62:** 252-262.

Jenkins, M.B., Hartel, P.G., Olexa, T.J., and Stuedemann, J.A. 2003. Putative temporal variability of *Escherichia coli* ribotypes from yearling steers. J. Environ. Qual. **32:** 305-309.

Jimenez, L., Muniz, I., Toranzos, G.A., and Hazen, T.C. 1989. Survival and activity of *Salmonella typhimurium* and *Escherichia coli* in tropical fresh-water. J. Appl. Bacteriol. **67:** 61-69.

Johnson, J.R., and O'Bryan, T.T. 2000. Improved repetitive-element PCR fingerprinting for resolving pathogenic and nonpathogenic phylogenetic groups within *Escherichia coli*. Clin. Diagn. Lab. Immunol. **7:** 265-273.

Johnson, L.K., Brown, M.B., Carruthers, E.A., Ferguson, J.A., Dombek, P.E., and Sadowsky, M.J. 2004. Sample size, library composition, and genotypic diversity among natural populations of *Escherichia coli* from different animals influence accuracy of determining sources of faecal pollution. Appl. Environ. Microbiol. **70:** 4478-4485.

Jones, P.G., Vanbogelen, R.A., and Neidhardt, F.C. 1987. Induction of proteins in response to low temperature in *Escherichia coli*. J. Bacteriol. **169:** 2092-2095.

Jovanovic, G., Lloyd, L.J., Stumpf, M.P.H., Mayhew, A.J., and Buck, M. 2006. Induction and function of the phage shock protein extracytoplasmic stress response in *Escherichia coli*. J. Biol. Chem. **281:** 21147-21161.

Kabir, M.S., Sagara, T., Oshima, T., Kawagoe, Y., Mori, H., Tsunedomi, R., and Yamada, M. 2004. Effects of mutations in the *rpoS* gene on cell viability and global gene expression under nitrogen starvation in *Escherichia coli*. Microbiology-Sgm **150:** 2543-2553.

Kakinuma, K., Fukushima, M., and Kawaguchi, R. 2003. Detection and identification of *Escherichia coli*, *Shigella*, and *Salmonella* by microarrays using the *gyrB* gene. Biotechnol. Bioeng. **83:** 721-728.

Kao, K.C., Tran, L.M., and Liao, J.C. 2005. A global regulatory role of gluconeogenic genes in *Escherichia coli* revealed by transcriptome network analysis. J. Biol. Chem. **280:** 36079-36087.

Karasinski, J., Andreescu, S., Sadik, O.A., Lavine, B., and Vora, M.N. 2005. Multiarray sensors with pattern recognition for the detection, classification, and differentiation of bacteria at subspecies and strain levels. Anal. Chem. **77:** 7941-7949.

Karlsson, F., Malmborg-Hager, A.C., Albrekt, A.S., and Borrebaeck, C. 2005. Genome-wide comparison of phage M13-infected vs. uninfected *Escherichia coli*. Can. J. Microbiol. **51:** 29-35.

Kern, R., Malki, A., Abdallah, J., Tagourti, J., and Richarme, G. 2007. *Escherichia coli hdeB* is an acid stress chaperone. J. Bacteriol. **189:** 603-610.

Kershaw, C.J., Brown, N.L., Constantinidou, C., Patel, M.D., and Hobman, J.L. 2005. The expression profile of *Escherichia coli* K-12 in response to minimal, optimal and excess copper concentrations. Microbiology-Sgm **151:** 1187-1198.

Kershaw, C.J., Brown, N.L., and Hobman, J.L. 2007. Zinc dependence of *zinT* (*yodA*) mutants and binding of zinc, cadmium and mercury by ZinT. Biochem. Biophys. Res. Commun. **364:** 66-71.

Kim, M.S., Bae, S.H., Yun, S.H., Lee, H.J., Ji, S.C., Lee, J.H., Srivastava, P., Lee, S.H., Chae, H., Lee, Y., Choi, B.S., Chattoraj, D.K., and Lim, H.M. 2005. Cnu, a novel oriC-binding protein of *Escherichia coli*. J. Bacteriol. **187:** 6998-7008.

Kim, S.K., Kimura, S., Shinagawa, H., Nakata, A., Lee, K.S., Wanner, B.L., and Makino, K. 2000. Dual transcriptional regulation of the *Escherichia coli* phosphate-starvation-inducible *psiE* gene of the phosphate regulon by *phoB* and the cyclic AMP (cAMP)-cAMP receptor protein complex. J. Bacteriol. **182:** 5596-5599.

Kinzelman, J., Ng, C., Jackson, E., Gradus, S., and Bagley, R. 2003. Enterococci as indicators of Lake Michigan recreational water quality: Comparison of two methodologies and their impacts on public health regulatory events. Appl. Environ. Microbiol. **69:** 92-96.

Kobayashi, R., Suzuki, T., and Yoshida, M. 2007. *Escherichia coli* phage-shock protein A (*pspA*) binds to membrane phospholipids and repairs proton leakage of the damaged membranes. Mol. Microbiol. **66:** 100-109.

Kostic, T., Weilharter, A., Rubino, S., Delogu, G., Uzzau, S., Rudi, K., Sessitsch, A., and Bodrossy, L. 2007. A microbial diagnostic microarray technique for the sensitive detection and identification of pathogenic bacteria in a background of nonpathogens. Anal. Biochem. **360:** 244-254.

Ksoll, W.B., Ishii, S., Sadowsky, M.J., and Hicks, R.E. 2007. Presence and sources of faecal coliform bacteria in epilithic periphyton communities of lake superior. Appl. Environ. Microbiol. **73:** 3771-3778.

Lacour, S., and Landini, P. 2004. Sigma(S)-dependent gene expression at the onset of stationary phase in *Escherichia coli*: Function of sigma(S)-dependent genes and identification of their promoter sequences. J. Bacteriol. **186:** 7186-7195.

Lee, J., Page, R., Garcia-Contreras, R., Palermino, J.M., Zhang, X.S., Doshi, O., Wood, T.K., and Peti, W. 2007. Structure and function of the *Escherichia coli* protein *ymgB*: a protein critical for biofilm formation and acid-resistance. J. Mol. Biol. **373:** 11-26.

Lee, L.J., Barrett, J.A., and Poole, R.K. 2005. Genome-wide transcriptional response of chemostat-cultured *Escherichia coli* to zinc. J. Bacteriol. **187:** 1124-1134.

Lemarchand, K., Berthiaume, F., Maynard, C., Harel, J., Payment, P., Bayardelle, P., Masson, L., and Brousseau, R. 2005. Optimization of microbial DNA extraction and purification from raw wastewater samples for downstream pathogen detection by microarrays. J. Microbiol. Methods **63:** 115-126.

Lenski, R.E., and Travisano, M. 1994. Dynamics of adaptation and diversification - a 10,000-generation experiment with bacterial populations. Proc. Natl. Acad. Sci. U.S.A. **91:** 6808-6814.

Leroi, A.M., Lenski, R.E., and Bennett, A.F. 1994. Evolutionary Adaptation to Temperature. 3. Adaptation of *Escherichia coli* to a temporally varying environment. Evolution **48:** 1222-1229.

Letowski, J., Brousseau, R., and Masson, L. 2003. DNA microarray applications in environmental microbiology. Anal. Lett. **36:** 3165-3184.

Li, Y., Cole, K., and Altman, S. 2003. The effect of a single, temperature-sensitive mutation on global gene expression in *Escherichia coli*. RNA **9:** 518-532.

Lucchini, S., Thompson, A., and Hinton, J.C.D. 2001. Microarrays for microbiologists. Microbiology-Sgm **147:** 1403-1414.

Maharjan, R.P., Yu, P.L., Seeto, S., and Ferenci, T. 2005. The role of isocitrate lyase and the glyoxylate cycle in *Escherichia coli* growing under glucose limitation. Res. Microbiol. **156:** 178-183.

Martinez, J.L. 2008. Antibiotics and antibiotic resistance genes in natural environments. Science **321:** 365-367.

Matin, A., Auger, E.A., Blum, P.H., and Schultz, J.E. 1989. Genetic basis of starvation survival in nondifferentiating bacteria. Annu. Rev. Microbiol. **43:** 293-316.

Maynard, C., Berthiaume, F., Lemarchand, K., Harel, J., Payment, P., Bayardelle, P., Masson, L., and Brousseau, R. 2005. Waterborne pathogen detection by use of oligonucleotide-based microarrays. Appl. Environ. Microbiol. **71:** 8548-8557.

Meays, C.L., Broersma, K., Nordin, R., and Mazumder, A. 2005. Survival of *Escherichia coli* in beef cattle faecal pats under different levels of solar exposure. Rangeland Ecology & Management **58:** 279-283.

Meays, C.L., Broersma, K., Nordin, R., and Mazumder, A. 2004. Source tracking faecal bacteria in water: a critical review of current methods. J. Environ. Manage. **73:** 71-79.

Meays, C.L., Broersma, K., Nordin, R., Mazumder, A., and Samadpour, M. 2006. Spatial and annual variability in concentrations and sources of *Escherichia coli* in multiple watersheds. Environ. Sci. Technol. **40:** 5289-5296.

Mizunoe, Y., Wai, S.N., Takade, A., and Yoshida, S. 1999. Restoration of culturability of starvation-stressed and low-temperature-stressed *Escherichia coli* O157 cells by using H_2O_2-degrading compounds. Arch. Microbiol. **172:** 63-67.

Muniz, I., Jimenez, L., Toranzos, G.A., and Hazen, T.C. 1989. Survival and activity of *Streptococcus faecalis* and *Escherichia coli* in tropical fresh-water. Microb. Ecol. **18:** 125-134.

Nachin, L., Nannmark, U., and Nystrom, T. 2005. Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility. J. Bacteriol. **187:** 6265-6272.

Noll, M., Petrukhin, K., and Lutsenko, S. 1998. Identification of a novel transcription regulator from *Proteus mirabilis*, PMTR, revealed a possible role of *yjaI* protein in balancing zinc in *Escherichia coli*. J. Biol. Chem. **273:** 21393-21401.

Oh, M.K., and Liao, J.C. 2000. Gene expression profiling by DNA microarrays and metabolic fluxes in *Escherichia coli*. Biotechnol. Prog. **16:** 278-286.

Olsen, A., Arnqvist, A., Hammar, M., Sukupolvi, S., and Normark, S. 1993. The Rpos sigma factor relieves H-Ns-mediated transcriptional repression of *csgA*, the subunit gene of fibronectin-binding curli in *Escherichia coli*. Mol. Microbiol. **7:** 523-536.

Patten, C.L., Kirchhof, M.G., Schertzberg, M.R., Morton, R.A., and Schellhorn, H.E. 2004. Microarray analysis of RpoS-mediated gene expression in *Escherichia coli* K12. Molecular Genetics and Genomics **272:** 580-591.

Paytubi, S., Madrid, C., Forns, N., Nieto, J.M., Balsalobre, C., Uhlin, B.E., and Juarez, A. 2004. YdgT, the Hha paralogue in *Escherichia coli*, forms heteromeric complexes with H-NS and StpA. Mol. Microbiol. **54:** 251-263.

Phadtare, S., and Inouye, M. 2004. Genome-wide transcriptional analysis of the cold shock response in wild-type and cold-sensitive, quadruple-csp-deletion strains of *Escherichia coli*. J. Bacteriol. **186:** 7007-7014.

Phadtare, S., Kato, I., and Inouye, M. 2002. DNA microarray analysis of the expression profile of *Escherichia coli* in response to treatment with 4,5-dihydroxy-2-cyclopenten-1-one. J. Bacteriol. **184:** 6725-6729.

Polissi, A., De Laurentis, W., Zangrossi, S., Briani, F., Longhi, V., Pesole, G., and Deho, G. 2003. Changes in *Escherichia coli* transcriptome during acclimatization at low temperature. Res. Microbiol. **154:** 573-580.

Porwollik, S., Frye, J., Florea, L.D., Blackmer, F., and McClelland, M. 2003. A non-redundant microarray of genes for two related bacteria. Nucleic Acids Res. **31:** 1869-1876.

Porwollik, S., Boyd, E.F., Choy, C., Cheng, P., Florea, L., Proctor, E., and McClelland, M. 2004. Characterization of *Salmonella enterica* subspecies I genovars by use of microarrays. J. Bacteriol. **186:** 5883-5898.

Prilipov, A., Phale, P.S., Koebnik, R., Widmer, C., and Rosenbusch, J.P. 1998. Identification and characterization of two quiescent porin genes, *nmpC* and *ompN*, in *Escherichia coli* B-E. J. Bacteriol. **180:** 3388-3392.

Rajkumari, K., and Gowrishankar, J. 2002. An N-terminally truncated RpoS (sigma(S)) protein in *Escherichia coli* is active in vivo and exhibits normal environmental regulation even in the absence of rpoS transcriptional and translational control signals. J. Bacteriol. **184:** 3167-3175.

Ram, J.L., Ritchie, R.P., Fang, J., Gonzales, F.S., and Selegean, J.P. 2004. Sequence-based source tracking of *Escherichia coli* based on genetic diversity of beta-glucuronidase. J. Environ. Qual. **33:** 1024-1032.

Ren, D., Bedzyk, L.A., Thomas, S.M., Ye, R.W., and Wood, T.K. 2004. Gene expression in *Escherichia coli* biofilms. Appl. Microbiol. Biotechnol. **64:** 515-524.

Richmond, C.S., Glasner, J.D., Mau, R., Jin, H.F., and Blattner, F.R. 1999. Genome-wide expression profiling in *Escherichia coli* K-12. Nucleic Acids Res. **27:** 3821-3835.

Riley, M., Abe, T., Arnaud, M.B., Berlyn, M.K.B., Blattner, F.R., Chaudhuri, R.R., Glasner, J.D., Horiuchi, T., Keseler, I.M., Kosuge, T., Mori, H., Perna, N.T., Plunkett, G., Rudd, K.E., Serres, M.H., Thomas, G.H., Thomson, N.R., Wishart, D., and Wanner, B.L. 2006. *Escherichia coli* K-12: a cooperatively developed annotation snapshot - 2005. Nucleic Acids Res. **34:** 1-9.

Rompre, A., Servais, P., Baudart, J., de-Roubin, M.R., and Laurent, P. 2002. Detection and enumeration of coliforms in drinking water: current methods and emerging approaches. J. Microbiol. Methods **49:** 31-54.

Rosenow, C., Saxena, R.M., Durst, M., and Gingeras, T.R. 2001. Prokaryotic RNA preparation methods useful for high-density array analysis: comparison of two approaches. Nucleic Acids Res. **29:** e112.

Sabat, G., Rose, P., Hickey, W.J., and Harkin, J.M. 2000. Selective and sensitive method for PCR amplification of *Escherichia coli* 16S rRNA genes in soil. Appl. Environ. Microbiol. **66:** 844-849.

Samsonova, N.N., Smirnov, S.V., Altman, I.B., and Ptitsyn, L.R. 2003. Molecular cloning and characterization of *Escherichia coli* K12 *ygjG* gene. Bmc Microbiology **3:**

Sauer, T.J., Daniel, T.C., Nichols, D.J., West, C.P., Moore, P.A., and Wheeler, G.L. 2000. Runoff water quality from poultry litter-treated pasture and forest sites. J. Environ. Qual. **29:** 515-521.

Savageau, M.A. 1983. *Escherichia coli* habitats, cell-types, and molecular mechanisms of gene-control. Am. Nat. **122:** 732-744.

Savelkoul, P.H.M., Aarts, H.J.M., de Haas, J., Dijkshoorn, L., Duim, B., Otsen, M., Rademaker, J.L.W., Schouls, L., and Lenstra, J.A. 1999. Amplified-fragment length polymorphism analysis: the state of an art. J. Clin. Microbiol. **37:** 3083-3091.

Scott, T.M., Rose, J.B., Jenkins, T.M., Farrah, S.R., and Lukasik, J. 2002. Microbial source tracking: Current methodology and future directions. Appl. Environ. Microbiol. **68:** 5796-5803.

Scott, T.M., Parveen, S., Portier, K.M., Rose, J.B., Tamplin, M.L., Farrah, S.R., Koo, A., and Lukasik, J. 2003. Geographical variation in ribotype profiles of *Escherichia coli* isolates from humans, swine, poultry, beef, and dairy cattle in Florida. Appl. Environ. Microbiol. **69:** 1089-1092.

Sessitsch, A., Hackl, E., Wenzl, P., Kilian, A., Kostic, T., Stralis-Pavese, N., Sandjong, B.T., and Bodrossy, L. 2006. Diagnostic microbial microarrays in soil ecology. New Phytol. **171:** 719-736.

Seurinck, S., Verstraete, W., and Siciliano, S.D. 2003. Use of 16S-23S rRNA intergenic spacer region PCR and repetitive extragenic palindromic PCR analyses of *Escherichia coli* isolates to identify nonpoint faecal sources. Appl. Environ. Microbiol. **69:** 4942-4950.

Shaw, K.J., Miller, N., Liu, X.J., Lerner, D., Wan, J., Bittner, A., and Morrow, B.J. 2003. Comparison of the changes in global gene expression of *Escherichia coli* induced by four bactericidal agents. J. Mol. Microbiol. Biotechnol. **5:** 105-122.

Simpson, J.M., Santo Domingo, J.W., and Reasoner, D.J. 2002. Microbial source tracking: State of the science. Environ. Sci. Technol. **36:** 5279-5288.

Sokal, R.R., and Rohlf, F.J. 1981. Biometry: the principles and practice of statistics in biological research. 2nd ed., W.H. Freeman and Company, San Fransisco, CA.

Soule, M., Kuhn, E., Loge, F., Gay, J., and Call, D.R. 2006. Using DNA microarrays to identify library-independent markers for bacterial source tracking. Appl. Environ. Microbiol. **72:** 1843-1851.

Stender, H., Broomer, A.J., Oliveira, K., Perry-O'Keefe, H., Hyldig-Nielsen, J.J., Sage, A., and Coull, J. 2001. Rapid detection, identification, and enumeration of *Escherichia coli* cells in municipal water by chemiluminescent in situ hybridization. Appl. Environ. Microbiol. **67:** 142-147.

Stoeckel, D.M., Mathes, M.V., Hyer, K.E., Hagedorn, C., Kator, H., Lukasik, J., O'Brien, T.L., Fenger, T.W., Samadpour, M., Strickler, K.M., and Wiggins, B.A. 2004. Comparison of seven protocols to identify faecal contamination sources using *Escherichia coli*. Environ. Sci. Technol. **38:** 6109-6117.

Sturgeon, J.A., Buttke, T.M., Duckworth, D.H., and Ingram, L.O. 1976. Characterization of the envelope of a low temperature conditional cell division mutant of *Escherichia coli* Strain K12. Abstracts of the Annual Meeting of the American Society for Microbiology **76:** I94.

Sugiura, M., Aiba, H., and Mizuno, T. 2003. Identification and classification of two-component systems that affect rpoS expression in *Escherichia coli*. Bioscience Biotechnology and Biochemistry **67:** 1612-1615.

Tao, H., Bausch, C., Richmond, C., Blattner, F.R., and Conway, T. 1999. Functional genomics: Expression analysis of *Escherichia coli* growing on minimal and rich media. J. Bacteriol. **181:** 6425-6440.

Tjaden, B., Saxena, R.M., Stolyar, S., Haynor, D.R., Kolker, E., and Rosenow, C. 2002. Transcriptome analysis of *Escherichia coli* using high-density oligonucleotide probe arrays. Nucleic Acids Res. **30:** 3732-3738.

Topp, E., Welsh, M., Tien, Y.C., Dang, A., Lazarovits, G., Conn, K., and Zhu, H. 2003. Strain-dependent variability in growth and survival of *Escherichia coli* in agricultural soil. FEMS Microbiol. Ecol. **44:** 303-308.

Ussery, D., Larsen, T.S., Wilkes, K.T., Friis, C., Worning, P., Krogh, A., and Brunak, S. 2001. Genome organisation and chromatin structure in *Escherichia coli*. Biochimie **83:** 201-212.

Vantarakis, A., Venieri, D., Komninou, G., and Papapetropoulou, M. 2006. Differentiation of faecal *Escherichia coli* from humans and animals by multiple antibiotic resistance analysis. Lett. Appl. Microbiol. **42:** 71-77.

Vital, M., Hammes, F., and Egli, T. 2008. *Escherichia coli* O157 can grow in natural freshwater at low carbon concentrations. Environ. Microbiol. **10:** 2387-2396.

Wang, X.A., Smith, D.R., Jones, J.W., and Chapman, M.R. 2007. In vitro polymerization of a functional *Escherichia coli* amyloid protein. J. Biol. Chem. **282:** 3713-3719.

Weber, A., and Jung, K. 2002a. Profiling early osmostress-dependent gene expression in *Escherichia coli* using DNA macroarrays. J Bacteriol. **184:** 5502-5507.

Weber, A., and Jung, K. 2002b. Profiling early osmostress-dependent gene expression in *Escherichia coli* using DNA microarrays. J. Bacteriol. **184:** 5502-5507.

Weber, H., Polen, T., Heuveling, J., Wendisch, V.F., and Hengge, R. 2005. Genome-wide analysis of the general stress response network in *Escherichia coli*: sigma(S)-dependent genes, promoters, and sigma factor selectivity. J. Bacteriol. **187:** 1591-1603.

White-Ziegler, C.A., Um, S., Prez, N.M., Berns, A.L., Malhowski, A.J., and Young, S. 2008. Low temperature (23 degrees C) increases expression of biofilm-, cold-shock- and RpoS-dependent genes in *Escherichia coli* K12. Microbiology-Sgm **154:** 148-166.

Woese, C.R. 1996. Phylogenetic trees: Whither microbiology? Current Biology **6:** 1060-1063.

Woese, C.R. 1987. Bacterial Evolution. Microbiol. Rev. **51:** 221-271.

Wood, J., Scott, K.P., Avgustin, G., Newbold, C.J., and Flint, H.J. 1998. Estimation of the relative abundance of different *Bacteroides* and *Prevotella* ribotypes in gut samples by restriction enzyme profiling of PCR-amplified 16S rRNA gene sequences. Appl. Environ. Microbiol. **64:** 3683-3689.

Wu, C.F., Valdes, J.J., Bentley, W.E., and Sekowski, J.W. 2003. DNA microarray for discrimination between pathogenic 0157 : H7 EDL933 and non-pathogenic *Escherichia coli* strains. Biosens. Bioelectron. **19:** 1-8.

Wu, L.Y., Liu, X., Schadt, C.W., and Zhou, J.Z. 2006. Microarray-based analysis of subnanogram quantities of microbial community DNAs by using whole-community genome amplification. Appl. Environ. Microbiol. **72:** 4931-4941.

Wu, L.Y., Thompson, D.K., Li, G.S., Hurt, R.A., Tiedje, J.M., and Zhou, J.Z. 2001. Development and evaluation of functional gene arrays for detection of selected genes in the environment. Appl. Environ. Microbiol. **67:** 5780-5790.

Xi, H.L., Schneider, B.L., and Reitzer, L. 2000. Purine catabolism in *Escherichia coli* and function of xanthine dehydrogenase in purine salvage. J. Bacteriol. **182:** 5332-5341.

Zhou, J.H. 2003. Microarrays for bacterial detection and microbial community analysis. Curr. Opin. Microbiol. **6:** 288-294.

Zhou, J.Z., and Thompson, D.K. 2002. Challenges in applying microarrays to environmental studies. Curr. Opin. Biotechnol. **13:** 204-207.

Appendix A Sequence Alignments

Sequence Alignments done on the Blast Program of the NCBI website:
http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=megaBlast&SHOW_DEFAULTS=on&BLAST_SPEC=blast2seq&LINK_LOC=align2seq

Alignments done in February 2008.

***E. coli* K12 *MG1655* (NCBI** U00096) **versus *E. coli* 43(C)-4A (Bovine origin)**

```
Score = 2834 bits (1474),  Expect = 0.0
 Identities = 1520/1533 (99%), Gaps = 4/1533 (0%)99.15%

 Strand=Plus/Plus

Query  1     GAGTTTGATCCTGGCTCAGATTGGACGCTGGCGGCAGGCCTAACACATGCAAGTCGAACG  60
             |||||||||| |||||||||||| ||||||||||||||||||||||||||||||||||||
Sbjct  9     GAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGAACG  68

Query  61    GTAACAGGAAACAGCTTGCTGTTTCGCTGACGAGTGGCGGACGGGTGAGTAATGTCTGGG  120
             ||||||||||  ||||||||  || ||||||||||||||||||||||||||||||||||||
Sbjct  69    GTAACAGGAAGAAGCTTGCTTCTTTGCTGACGAGTGGCGGACGGGTGAGTAATGTCTGGG  128

Query  121   AAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATACCGCATAACGTCGC  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  129   AAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATACCGCATAACGTCGC  188

Query  181   AAGACCAAAGAGGGGGACCTTCGGGCC-CTTGCCATCGGATGTGCCCAGATGGGATTAGC  239
             |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Sbjct  189   AAGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCGGATGTGCCCAGATGGGATTAGC  248

Query  240   TAGTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGATGACC  299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  249   TAGTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGATGACC  308

Query  300   AGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATT  359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  309   AGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATT  368

Query  360   GCACAATGGGCGCAAGC-TGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGTTG  418
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Sbjct  369   GCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGTTG  428

Query  419   TAAAGTACTTTCAGCGGGGAGGAAGGGAGTAAAGTTAATACCTTTGCTCATTGACGTTAC  478
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  429   TAAAGTACTTTCAGCGGGGAGGAAGGGAGTAAAGTTAATACCTTTGCTCATTGACGTTAC  488

Query  479   CCGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAG  538
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  489   CCGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAG  548

Query  539   CGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTAAGTCAGATGTGAA  598
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  549   CGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTAAGTCAGATGTGAA  608

Query  599   ATCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGG  658
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct  609   ATCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGG  668

Query  659   GGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGA  718
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  669   GGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGA  728

Query  719   AGGCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGAT  778
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  729   AGGCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGAT  788

Query  779   TAGATACCCTGGTAGTCCACGCCGTAAACGATGTCGACTTGGAGGTTGTGCCCTTGAGGC  838
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  789   TAGATACCCTGGTAGTCCACGCCGTAAACGATGTCGACTTGGAGGTTGTGCCCTTGAGGC  848

Query  839   GTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAA  898
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  849   GTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAA  908

Query  899   ACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCA  958
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  909   ACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCA  968

Query  959   ACGCGAAGAACCTTACCTGGTCTTGACATCCACGGAAGTTTTCAGAGATGAGAATGTG-C  1017
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct  969   ACGCGAAGAACCTTACCTGGTCTTGACATCCACGGAAGTTTTCAGAGATGAGAATGTGCC  1028

Query  1018  TTCGGGAACCGTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGG  1077
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1029  TTCGGGAACCGTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGG  1088

Query  1078  GTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGCCAGCGGTCCGGCCGGGAACT  1137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1089  GTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGCCAGCGGTCCGGCCGGGAACT  1148

Query  1138  CAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTCAAGTCATCATGGC  1197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1149  CAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTCAAGTCATCATGGC  1208

Query  1198  CCTTACGACCAGGGCTACACACGTGCTACAATGGCGCATACAAAGAGAAGCGACCTCGCG  1257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1209  CCTTACGACCAGGGCTACACACGTGCTACAATGGCGCATACAAAGAGAAGCGACCTCGCG  1268

Query  1258  AGAGCAAGCGGACCTCATAAAGTGCGTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCC  1317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1269  AGAGCAAGCGGACCTCATAAAGTGCGTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCC  1328

Query  1318  ATGAAGTCGGAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAATACGTTCCCGGGC  1377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1329  ATGAAGTCGGAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAATACGTTCCCGGGC  1388

Query  1378  CTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCAAAAGAAGTAGGTAGCTTAA-  1436
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Sbjct  1389  CTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCAAAAGAAGTAGGTAGCTTAAC  1448

Query  1437  CTTCGGGAGGGCGCTTACCACTTTGTGATTCATGACTGGGGTGAAGTCGCAACAAGGTAA  1496
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Sbjct  1449  CTTCGGGAGGGCGCTTACCACTTTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAA  1508

Query  1497  CCGTAGGGGAACCTGCGGCTGGATCACCTCCTT  1529
              |||||||||||||||| ||||||||||||||
Sbjct  1509  CCGTAGGGGAACCTGCGGTTGGATCACCTCCTT  1541
```

43(C)-4A (bovine origin) versus K12 used in this research (ATCC: 700926)

```
Score = 2859 bits (1487),  Expect = 0.0
 Identities = 1521/1533 (99%), Gaps = 2/1533 (0%) 99.22%
 Strand=Plus/Minus

Query  1     GAGTTTGATCCTGGCTCAGATTGGACGCTGGCGGCAGGCCTAACACATGCAAGTCGAACG  60
             |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Sbjct  1532  GAGTTTGATCCTGGCTCAGATTGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGAACG  1473

Query  61    GTAACAGGAAACAGCTTGCTGTTTCGCTGACGAGTGGCGGACGGGTGAGTAATGTCTGGG  120
             ||||||||||  ||||||||  || ||||||||||||||||||||||||||||||||||||
Sbjct  1472  GTAACAGGAAGAAGCTTGCTTCTTTGCTGACGAGTGGCGGACGGGTGAGTAATGTCTGGG  1413

Query  121   AAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATACCGCATAACGTCGC  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1412  AAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATACCGCATAACGTCGC  1353

Query  181   AAGACCAAAGAGGGGGACCTTCGGGCC-CTTGCCATCGGATGTGCCCAGATGGGATTAGC  239
             |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Sbjct  1352  AAGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCGGATGTGCCCAGATGGGATTAGC  1293

Query  240   TAGTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGATGACC  299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1292  TAGTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGATGACC  1233

Query  300   AGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATT  359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1232  AGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATT  1173

Query  360   GCACAATGGGCGCAAGCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGTTGT  419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1172  GCACAATGGGCGCAAGCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGTTGT  1113

Query  420   AAAGTACTTTCAGCGGGGAGGAAGGGAGTAAAGTTAATACCTTTGCTCATTGACGTTACC  479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1112  AAAGTACTTTCAGCGGGGAGGAAGGGAGTAAAGTTAATACCTTTGCTCATTGACGTTACC  1053

Query  480   CGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGC  539
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1052  CGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGC  993

Query  540   GTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTAAGTCAGATGTGAAA  599
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  992   GTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTAAGTCAGATGTGAAA  933

Query  600   TCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGGG  659
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  932   TCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGGG  873

Query  660   GGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAA  719
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Sbjct  872   GGTAGAATTCCAGGTGTAGTGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAA  813

Query  720   GGCGGCCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGATT  779
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  812   GGCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGATT  753

Query  780   AGATACCCTGGTAGTCCACGCCGTAAACGATGTCGACTTGGAGGTTGTGCCCTTGAGGCG  839
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  752   AGATACCCTGGTAGTCCACGCCGTAAACGATGTCGACTTGGAGGTTGTGCCCTTGAGGCG  693
```

```
Query  840   TGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAAA  899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  692   TGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAAA  633

Query  900   CTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAA  959
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Sbjct  632   CTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATCCGATGCAA  573

Query  960   CGCGAAGAACCTTACCTGGTCTTGACATCCACGGAAGTTTTCAGAGATGAGAATGTGCTT  1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  572   CGCGAAGAACCTTACCTGGTCTTGACATCCACGGAAGTTTTCAGAGATGAGAATGTGCTT  513

Query  1020  CGGGAACCGTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGT  1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  512   CGGGAACCGTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGT  453

Query  1080  TAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGCCAGCGGTCCGGCCGGGAACTCA  1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  452   TAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGCCAGCGGTCCGGCCGGGAACTCA  393

Query  1140  AAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTCAAGTCATCATGGCCC  1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  392   AAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTCAAGTCATCATGGCCC  333

Query  1200  TTACGACCAGGGCTACACACGTGCTACAATGGCGCATACAAAGAGAAGCGACCTCGCGAG  1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  332   TTACGACCAGGGCTACACACGTGCTACAATGGCGCATACAAAGAGAAGCGACCTCGCGAG  273

Query  1260  AGCAAGCGGACCTCATAAAGTGCGTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCAT  1319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  272   AGCAAGCGGACCTCATAAAGTGCGTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCAT  213

Query  1320  GAAGTCGGAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAATACGTTCCCGGGCCT  1379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  212   GAAGTCGGAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAATACGTTCCCGGGCCT  153

Query  1380  TGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCAAAAGAAGTAGGTAGCTTAACTT  1439
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  152   TGTACGCACCGCCCGTCACACCATGGGAGTGGGTTGCAAAAGAAGTAGGTAGCTTAACTT  93

Query  1440  CGGGAGGGCGCTTACCACTTTGTGATTCATGACTGGGGTGAAGTCGCAACAAGGTAACCG  1499
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Sbjct  92    CGGGAGGGCGCTTACCACTTTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAACCG  33

Query  1500  TAGGGGAACCTGCGGCTGGATCACCTCCTTTCT  1532
              |||||||| |||||||||||||||||||||||
Sbjct  32    TAGGGGAA-CTGCGGCTGGATCACCTCCTTTCT  1
```

Appendix B

Table B1. Genes differentially expressed between *E. coli* K12 and E43 at both 37°C and

Gene Name	Blattner ID	Log_2Ratio 37°C	Stdev*	Log_2Ratio 21°C	Stdev	Gene Product
Cell Process						
uspB / yhiO	b3494	-2.15	0.56	-2.82	0.41	predicted universal stress (ethanol tolerance) protein B
bfr	b3336	-6.49	0.539	-7.00	0.436	bacterioferritin, iron storage and detoxification protein
osmC	b1482	-2.43	0.27	-3.56	0.447	osmotically inducible, stress-inducible membrane protein
osmB	b1283	-2.12	0.35	-3.26	0.342	lipoprotein
osmY	b4376	-2.62	0.41	-2.52	0.38	periplasmic protein
slp	b3506	-3.87	0.664	-3.10	1.438	outer membrane lipoprotein
cspH	b0989	3.28	0.896	2.83	0.31	stress protein, member of the CspA-family
gidB	b3740	2.82	0.58	3.40	0.822	methyltransferase, SAM-dependent methyltransferase, glucose-inhibited cell-division protein
Cell Structure						
gfcA / ymcD	b0987	5.03	0.401	4.52	0.553	predicted protein
pgaB	b1023	3.48	1.346	-3.49	0.337	putatative polysaccharide N-deacetylase/carbohydrate esterase
gfcE	b0983	3.84	0.663	3.71	0.635	predicted exopolysaccharide export protein
rfaB	b3628	-3.52	0.968	-6.93	0.668	UDP-D-galactose:(glucosyl)lipopolysaccharide-1,6-D-galactosyltransferase
rfaJ	b3626	-3.93	1.045	-3.52	0.503	UDP-D-glucose:(galactosyl)lipopolysaccharide glucosyltransferase
yibJ	b3595	2.68	0.62	4.81	1.745	predicted Rhs-family protein
Information Transfer						
dps	b0812	-3.80	0.615	-2.46	0.27	Fe-binding and storage protein
etp	b0982	3.30	0.770	-2.20	0.31	phosphotyrosine-protein phosphatase
gadX	b3516	-2.12	0.74	-2.64	0.12	DNA-binding transcriptional dual regulator
osmE	b1739	-2.99	0.58	-3.03	0.283	DNA-binding transcriptional activator
relB	b1564	-5.33	1.087	2.43	0.25	Qin prophage; bifunctional antitoxin of the RelE-RelB toxin-antitoxin system and transcriptional repressor
relE	b1563	-7.79	0.522	-5.35	1.277	Qin prophage; toxin of the RelE-RelB toxin-antitoxin system
Metabolism						
aldB	b3588	-2.17	0.35	-3.34	0.288	aldehyde dehydrogenase B
chiA	b3338	-3.40	0.527	-2.23	0.33	periplasmic endochitinase
cmtB	b2934	2.24	0.56	2.25	0.41	PTS system putative mannitol-specific EIIA component
dcuB	b4123	3.08	0.611	2.29	0.98	C4-dicarboxylate antiporter
fbaB	b2097	-3.33	0.698	-3.38	0.266	fructose-bisphosphate aldolase class I
gadA	b3517	-4.07	1.081	-4.00	0.202	glutamate decarboxylase A, PLP-dependent
gadB	b1493	-6.62	0.852	-2.96	0.20	glutamate decarboxylase B, PLP-dependent

* Standard Deviation calculated in Microsoft Excel.

Table B1. continued.

Gene Name	Blattner ID	Log_2Ratio 37°C	Stdev*	Log_2Ratio 21°C	Stdev	Gene Product
Metabolism						
gatA	b2094	-6.21	0.238	-6.75	0.505	galactitol-specific enzyme IIA component of PTS
gatB	b2093	-6.05	0.190	-6.69	0.503	galactitol-specific enzyme IIB component of PTS
gatC	b2092	-5.16	0.794	-6.56	0.668	galactitol-specific enzyme IIC component of PTS
gatD	b2091	-6.27	0.278	-7.53	1.177	galactitol-1-phosphate dehydrogenase, Zn-dependent and NAD(P)-binding
gatY	b2096	-6.04	0.470	-6.27	0.206	D-tagatose 1,6-bisphosphate aldolase 2, catalytic subunit
gatZ	b2095	-6.29	0.335	-6.75	0.479	D-tagatose 1,6-bisphosphate aldolase 2, subunit
glf	b2036	-6.52	1.232	-7.54	1.252	UDP-galactopyranose mutase, FAD/NAD(P)-binding
htrL	b3618	-4.28	1.139	-2.49	0.28	predicted protein
hyaA	b0972	-3.08	0.497	-5.92	0.665	hydrogenase 1, small subunit
hyaB	b0973	-2.07	0.49	-3.63	0.348	hydrogenase 1, large subunit
hyaC	b0974	-2.62	0.64	-2.59	0.30	hydrogenase 1, b-type cytochrome subunit
hyaD	b0975	-2.41	0.55	-2.71	0.23	protein involved in processing of HyaA and HyaB proteins
hyaE	b0976	-2.09	0.64	-2.49	0.27	protein involved in processing of HyaA and HyaB proteins
kgtP	b2587	-2.61	0.39	-4.38	0.736	alpha-ketoglutarate transporter
mcrB	b4346	-4.59	1.152	-2.23	0.24	5-methylcytosine-specific restriction enzyme McrBC, subunit McrB
mglA	b2149	2.35	0.56	-4.50	0.773	fused methyl-galactoside transporter subunits of ABC superfamily: ATP-binding components
mglC	b2148	2.22	0.97	2.55	0.53	membrane component of an ABC superfamily methyl-galactoside transporter
mqsR	b3512	-4.62	1.078	-3.44	0.378	acid-induced positive regulator of glutamate-dependent acid resistance
oppB	b1244	-2.31	0.84	-4.65	0.476	oligopeptide transporter subunit; membrane component of ABC superfamily
otsB	b1897	-3.02	0.735	-3.47	0.264	trehalose-6-phosphate phosphatase, biosynthetic
psiF	b0384	-2.17	0.60	-2.40	0.28	lipopolysaccharide biosynthesis protein (pseudogene)
relA	b2784	2.32	0.17	-3.67	0.309	(p)ppGpp synthetase I/GTP pyrophosphokinase
rfaG	b3631	-3.70	0.649	-4.57	0.895	glucosyltransferase I
rfaI	b3627	-2.70	0.99	-4.68	0.311	UDP-D-galactose:(glucosyl)lipopolysaccharide-alpha-1,3-D-galactosyltransferase
rfaS	b3629	-5.10	1.198	-4.63	0.881	lipopolysaccharide core biosynthesis protein
rfaY	b3625	-4.41	1.908	-6.31	0.335	lipopolysaccharide core biosynthesis protein
rfaZ	b3624	-4.06	0.969	-5.73	0.809	lipopolysaccharide core biosynthesis protein

Table B1. Continued.

Gene Name	Blattner ID	Log_2Ratio 37°C	Stdev*	Log_2Ratio 21°C	Stdev	Gene Product
Metabolism						
rfbA	b2039	-4.21	1.073	-6.37	1.484	glucose-1-phosphate thymidylyltransferase
rfbB	b2041	-4.31	0.496	-6.00	0.496	dTDP-glucose 4,6 dehydratase, NAD(P)-binding
rfbC	b2038	-5.38	1.227	-4.66	0.751	dTDP-4-deoxyrhamnose-3,5-epimerase
rfbX	b2037	-3.27	0.322	-7.73	0.682	predicted polisoprenol-linked O-antigen transporter
rpiR	b4089	-3.93	1.211	-4.60	0.558	DNA-binding transcriptional repressor
talA	b2464	-2.46	0.66	-3.13	0.337	transaldolase A
wbbH	b2035	-5.11	0.998	-4.43	0.536	O-antigen polymerase
wbbJ	b2033	-6.76	0.677	-9.20	0.560	predicted acyl transferase
wrbA	b1004	-2.37	0.54	-10.11	1.016	predicted flavoprotein in Trp regulation
ybaS	b0485	-2.85	0.33	-5.19	0.677	predicted glutaminase
ybaT	b0486	-2.37	0.42	-2.55	0.31	predicted transporter
yeeR	b1416	-2.14	0.60	-3.07	0.306	C-terminal fragment of glyceraldehyde-3-phosphate dehydrogenase C (pseudogene)
Regulation						
gatR/ rfaL	b3622	-2.93	0.63	-9.76	1.218	O-antigen ligase
ygiT	b3021	-5.20	1.265	-6.03	0.200	predicted DNA-binding transcriptional regulator
bolA	b0435	-2.38	0.32	-2.38	0.32	predicted protein (pseudogene)
gapC/ rhaK	b3623	-6.80	1.685	-5.39	1.297	lipopolysaccharide core biosynthesis
yjhU	b4295	-4.25	0.832	-3.84	0.270	KpLE2 phage-like element; predicted DNA-binding transcriptional regulator
Transport						
gadC/ xasA	b1492	-4.55	0.765	-2.42	0.31	predicted glutamate:gamma-aminobutyric acid antiporter
ydcV	b1443	-2.39	0.59	-2.94	0.49	membrane component of an ABC superfamily predicted spermidine/putrescine transporter
ydcS	b1440	-2.85	0.84	-4.16	0.933	periplasmic-binding component of an ABC superfamily predicted spermidine/putrescine transporter
msyB	b1051	-2.01	0.42	2.76	0.74	predicted protein
Extrachromosomal						
yfjN	b2630	-3.49	0.391	-9.02	0.814	CP4-57 prophage; RNase LS
yagN	b0280	-6.13	1.420	-7.51	0.733	CP4-6 prophage; predicted protein
yfdH	b2351	-5.97	0.422	-6.09	0.657	CPS-53 (KpLE1) prophage; bactoprenol glucosyl transferase
dicA	b1570	-5.58	0.662	-6.01	0.730	Qin prophage; predicted regulator for DicB
intF	b0281	-4.84	0.941	-5.24	0.351	CP4-6 prophage; predicted phage integrase
yfjW	b2642	-5.30	0.570	-4.88	0.894	CP4-57 prophage; predicted inner membrane protein
intB	b4271	-2.63	1.14	-4.68	0.635	KpLE2 phage-like element; predicted integrase
yfjO	b2631	-5.22	0.566	-3.97	0.134	CP4-57 prophage; predicted protein
yffL	b2443	-3.76	0.789	-3.94	1.201	CPZ-55 prophage; predicted protein
ykfB	b0250	-4.37	0.502	-3.86	0.539	CP4-6 prophage; predicted protein

Table B1. Continued.

Gene Name	Blattner ID	Log_2Ratio 37°C	Stdev*	Log_2Ratio 21°C	Stdev	Gene Product
Extrachromosomal						
gadC	b0556	2.10	1.10	-3.81	1.523	DLP12 prophage; predicted murein endopeptidase
yffR	b2449	-2.66	0.88	-3.64	0.569	CPZ-55 prophage; predicted protein
gatR_2	b2090	-2.03	0.40	-3.06	0.330	DNA-binding transcriptional repressor of galactitol utilization, N-ter fragment (pseudogene)
racR	b1356	-5.93	0.787	-2.91	0.36	Rac prophage; predicted DNA-binding transcriptional regulator
intA	b2622	-4.91	0.844	-2.57	0.46	CP4-57 prophage; integrase
nmpC	b0553	-6.82	0.872	-2.44	0.15	DLP12 prophage; truncated outer membrane porin (pseudogene)
ybcM	b0546	-4.03	1.007	-2.07	0.36	DLP12 prophage; predicted DNA-binding transcriptional regulator
borD	b0557	2.31	0.66	2.04	0.85	DLP12 prophage; predicted lipoprotein
cspF	b1558	2.33	0.81	2.88	1.22	Qin prophage; cold shock protein
ybcW	b0559	3.71	1.610	3.09	0.858	DLP12 prophage; predicted protein
pinR/ ybcY	b0562	3.79	0.975	3.85	0.953	DLP12 prophage; predicted SAM-dependent methyltransferase
gatR_1	b1374	2.09	0.67	2.04	0.23	Rac prophage; predicted site-specific recombinase
Unknown General Classification						
yafN	b0232	-6.37	0.660	-8.65	1.425	predicted antitoxin of the YafO-YafN toxin-antitoxin system
msqR / pgaB	b3022	-7.22	1.313	-7.94	1.346	quorum-sensing regulator
wbbK	b2032	-9.38	0.572	-7.90	0.252	lipopolysaccharide biosynthesis protein
wbbI	b2034	-7.73	0.442	-7.71	1.209	conserved protein
yahO	b0329	-4.34	0.605	-6.85	0.526	predicted protein
yeaG	b1783	-3.04	0.393	-6.11	0.779	conserved protein with nucleoside triphosphate hydrolase domain
ygaF	b2660	-2.67	1.00	-5.96	0.589	predicted enzyme
yfcO	b2332	-4.15	0.743	-5.12	1.029	predicted protein
yeaQ	b1795	-2.15	0.73	-4.70	0.357	conserved inner membrane protein
yiaH	b3561	3.43	0.804	-4.47	0.314	conserved inner membrane protein
gfcD	b0984	3.59	0.547	-4.42	0.686	conserved protein
yccJ	b1003	-2.20	0.68	-4.41	0.964	predicted protein
yodD	b1953	-3.84	1.164	-4.17	0.771	predicted protein
yabP_2	b0057	-7.44	1.074	-3.88	1.074	C-terminal fragment of a predicted protein (pseudogene)
yjdN	b4107	-2.35	0.94	-3.67	0.444	conserved protein
ygaU	b2665	-2.61	0.62	-3.62	0.152	predicted protein
yhfG	b3362	-3.64	1.507	-3.60	0.537	predicted protein
yebV	b1836	-3.37	0.601	-3.58	0.425	predicted protein
yodC	b1957	-3.10	0.699	-3.47	0.244	predicted protein
elaB	b2266	-3.07	0.529	-3.45	0.262	conserved protein

Table B1. Continued.

Gene Name	Blattner ID	Log_2Ratio 37°C	Stdev*	Log_2Ratio 21°C	Stdev	Gene Product
Unknown General Classification						
yabP_1	b0056	-3.24	0.778	-3.38	0.519	N-terminal fragment of a predicted protein (pseudogene)
ybgS	b0753	-2.49	0.67	-3.33	0.435	conserved protein
ygaM	b2672	-2.52	0.52	-3.28	0.433	predicted protein
ydiZ	b1724	-2.04	0.53	-3.01	0.258	predicted protein
wbbL	b4571	-7.41	0.954	-2.91	0.42	
ygiW	b3024	-3.72	0.477	-2.83	0.24	conserved protein
yqjE	b3099	-2.00	0.39	-2.65	0.23	conserved inner membrane protein
yegP	b2080		0.438	-2.54	0.40	predicted protein
yqjC	b3097	-2.23	0.48	-2.47	0.39	conserved protein
hdeD	b3511	-3.52	0.665	-2.24	0.35	acid-resistance membrane protein
yiaG	b3555	-3.10	0.653	-4.47	0.31	predicted transcriptional regulator
hdeB	b3509	-5.23	0.913	-2.06	0.31	acid-resistance protein
yodA	b1973	3.55	0.930	2.24	0.47	conserved metal-binding protein
ybjP	b0865	-3.14	0.782	2.36	0.67	predicted lipoprotein
yiiG	b3896	2.09	0.57	2.53	0.44	conserved protein
yqgC	b2940	2.00	0.53	2.57	1.05	predicted protein
yjbJ	b4045	-3.41	0.542	2.77	1.25	predicted stress response protein
gfcB	b0986	3.09	0.330	2.86	0.18	predicted outer membrane lipoprotein
yigB	b3812	3.44	0.461	2.94	0.54	FMN phosphatase
yibG	b3596	3.52	1.373	2.99	0.87	conserved protein
ycgB	b1188	-2.09	0.32	3.80	1.434	conserved protein
gapC_2	b2001	2.53	0.83	3.81	0.949	CP4-44 prophage; predicted membrane protein
hdeA	b3510	-4.90	0.550	3.83	0.385	stress response protein acid-resistance protein

Table B2. Genes differentially expressed between *E. coli* K12 and E43 at 37°C only.

Gene Name	Blattner ID	Log_2 Ratio	Stdev*	Gene Product
Cell structure				
gfcA	b0987	5.03	0.401	predicted protein
gfcE	b0983	3.84	0.663	predicted exopolysaccharide export protein
pgaB	b1023	3.48	1.346	putatative polysaccharide N-deacetylase/carbohydrate esterase
yiaA	b3562	2.02	1.02	conserved inner membrane protein
Information Transfer				rhsB_b3482_at
alaT	b3853	-2.29	0.96	tRNA-Ala(UGC)
hsdS	b4348	-3.89	0.539	specificity determinant for hsdM and hsdR
pspA	b1304	-2.01	0.45	DNA-binding transcriptional activator
pspC	b1306	-2.20	0.70	Qin prophage; bifunctional antitoxin of the RelE-RelB toxin-antitoxin system and transcriptional repressor
sra	b1480	-3.25	0.401	
Metabolism				
csrA	b2696	-2.14	0.72	post transcriptional regulatory RNA-binding protein
gapC_2	b1416	-2.14	0.60	C-terminal fragment of glyceraldehyde-3-phosphate dehydrogenase C (pseudogene)
gatR_1	b2090	-2.03	0.40	
waaU	b3623	-6.80	1.685	lipopolysaccharide core biosynthesis
Regulation				
bfd	b3337	-2.27	0.47	bacterioferritin-associated ferredoxin
yeiL	b2163	2.00	1.07	DNA-binding transcriptional activator of stationary phase nitrogen survival
yiaG	b3555	-3.10	0.653	predicted transcriptional regulator
g. Transport				
ompC	b2215	-2.61	0.61	outer membrane porin protein C
Extrachromosomal				
insD-2	b1402	-2.29	1.20	IS2 insertion element transposase InsAB'
pinR	b1374	2.09	0.67	Rac prophage; predicted site-specific recombinase
racR	b1356	-5.93	0.787	Rac prophage; predicted DNA-binding transcriptional regulator
rzpD	b0556	2.10	1.10	DLP12 prophage; predicted murein endopeptidase
yagJ	b0276	-3.88	0.451	CP4-6 prophage; predicted protein
ybcY	b0562	3.79	0.975	DLP12 prophage; predicted SAM-dependent methyltransferase
yeeR	b2001	2.53	0.83	CP4-44 prophage; predicted membrane protein
yfdQ	b2360	2.82	1.02	CPS-53 (KpLE1) prophage; predicted protein
Unknown general classification				
elaB	b2266	-3.07	0.529	conserved protein
gfcC	b0985	3.74	0.901	conserved protein
mqsR	b3022	-7.22	1.313	quorum-sensing regulator
yabP_2	b0057	-7.44	1.074	C-terminal fragment of a predicted protein (pseudogene)
yiaH	b3561	3.43	0.804	conserved inner membrane protein
yibG	b3596	3.52	1.373	conserved protein

*** Standard Deviation calculated in Microsoft Excel.**

Table B3. Genes differentially expressed between *E. coli* K12 and E43 at 21°C only.

Gene Name	Blattner ID	Log_2 Ratio	Stdev*	Gene Product
Cell Process				
fic	b3361	-2.75	0.23	stationary-phase protein, cell division
hlyE	b1182	-4.10	0.533	hemolysin E
ygeH	b2852	3.97	1.761	predictedtranscriptional regulator
yjaZ	b3989	2.38	0.55	heat shock protein
Cell Structure				
blc	b4149	-2.90	0.28	outer membrane lipoprotein (lipocalin)
rfaI	b3627	-3.52	0.503	UDP-D-galactose:(glucosyl)lipopolysaccharide-alpha-1,3-D-galactosyltransferase
yaiV	b0375	3.42	1.060	predicted DNA-binding transcriptional regulator
yccZ	b0983	3.81	0.436	predicted exopolysaccharide export protein
ycdR	b1023	3.80	1.434	putatative polysaccharide N-deacetylase/carbohydrate esterase
yedS	b1964	-4.84	1.379	predicted protein, N-ter fragment (pseudogene)
yiaD	b3552	3.69	0.433	predicted outer membrane lipoprotien
yidC	b3705	2.21	0.92	cytoplasmic insertase into membrane protein, Sec system
ymcD	b0987	4.52	0.553	predicted protein
DNA sites				
rhsB	b3482	3.25	1.085	rhsB element core protein RshB
rhsE	b1456	2.62	0.83	rhsE element core protein RshE
Information Transfer				
argW	b2348	5.17	0.690	tRNA-Arg(CCU)
argX	b3796	2.84	0.70	tRNA-Arg(CCG)
cbpA	b1000	-2.65	0.28	curved DNA-binding protein, DnaJ homologue that functions as a co-chaperone of DnaK
csgD	b1040	-2.33	0.77	DNA-binding transcriptional regulator of adhesion determinants
glyU	b2864	3.61	0.674	tRNA-Gly(CCC)
mrr	b4351	-4.23	0.641	methylated adenine and cytosine restriction protein
proM	b3799	2.55	1.21	tRNA-Pro(UGG)
soxR	b4063	2.12	1.12	DNA-binding transcriptional dual regulator, Fe-S center for redox-sensing
sufB	b1683	-3.21	0.557	component of SufBCD complex
ycdI /// ydcI	b1422	3.44	0.220	predicted DNA-binding transcriptional regulator
ydjF	b1770	-3.77	0.502	predicted DNA-binding transcriptional regulator
yihW	b3884	-2.40	0.80	predicted DNA-binding transcriptional regulator
Metabolism				
aceA	b4015	-2.55	0.49	isocitrate lyase
aceB	b4014	-2.08	0.56	malate synthase A
adhP	b1478	-2.38	0.29	alcohol dehydrogenase, 1-propanol preferring
astB	b1745	-2.75	0.23	succinylarginine dihydrolase
astD	b1746	-2.19	0.31	succinylglutamic semialdehyde dehydrogenase
btuE	b1710	-2.33	0.22	predicted glutathione peroxidase
c1843 /// gapC	b1416	-3.07	0.306	C-terminal fragment of glyceraldehyde-3-phosphate dehydrogenase C (pseudogene)

*** Standard Deviation calculated in Microsoft Excel.**

Table B3. Continued.

Gene Name	Blattner ID	Log_2 Ratio	Stdev	Gene Product
Metabolism				
tdcD	b3115	3.07	0.790	propionate kinase/acetate kinase C, anaerobic
tktB	b2465	-2.64	0.45	transketolase 2, thiamin-binding
ybaY	b0453	-2.55	0.44	predicted outer membrane lipoprotein
ycaC	b0897	-2.70	0.50	predicted hydrolase
ydeN	b1498	2.09	0.30	conserved protein
yodB	b1974	3.49	0.741	predicted cytochrome
Regulation				
chpB	b4225	-3.52	0.732	toxin of the ChpB-ChpS toxin-antitoxin system
yahA	b0315	2.15	0.27	predicted DNA-binding transcriptional regulator
yjjM	b4357	3.58	1.375	predicted DNA-binding transcriptional regulator
Transport				
acrF	b3266	3.59	1.197	multidrug efflux system protein
c3427 /// ygdQ	b2832	3.00	0.945	predicted inner membrane protein
c3761 /// ygiS	b3020	-3.96	0.576	predicted transporter subunit: periplasmic-binding component of ABC superfamily
csgE	b1039	-4.53	1.232	predicted transport protein
csgF	b1038	-2.05	0.73	predicted transport protein
frlB	b3371	2.02	1.00	fructoselysine-6-P-deglycase
gspM	b3334	-2.50	0.64	general secretory pathway component, cryptic
lsrA	b1513	-2.43	0.58	fused AI2 transporter subunits of ABC superfamily: ATP-binding components
mntH	b2392	-2.00	0.25	proton-dependent divalent metal cation transporter
sufC	b1682	-2.87	0.31	transport protein associated with Fe-S cluster assembly
xasA /// gadC	b1492	-3.38	0.519	predicted glutamate:gamma-aminobutyric acid antiporter
yeiU	b2174	2.49	0.53	undecaprenyl pyrophosphate phosphatase
yhiD	b3508	-2.14	0.31	predicted Mg(2+) transport ATPase inner membrane protein
yjhB	b4279	-5.77	0.289	KpLE2 phage-like element; predicted transporter
yohC	b2135	-2.80	0.55	predicted inner membrane protein
Extrachromosomal				
c0651 /// c1474 /// ybcY	b0562	3.84	1.512	DLP12 prophage; predicted SAM-dependent methyltransferase
c1274 /// c3664 /// yeeR	b2001	3.81	0.949	CP4-44 prophage; predicted membrane protein
c1563 /// rzpD	b0556	2.04	0.85	DLP12 prophage; predicted murein endopeptidase
c3146 /// pinQ /// pinR	b1374	2.04	0.23	Rac prophage; predicted site-specific recombinase
flxA	b1566	-6.54	0.339	Qin prophage; predicted protein
racR	b1356	-6.11	0.779	Rac prophage; predicted DNA-binding transcriptional regulator

Table B3. Continued.

Gene Name	Blattner ID	Log_2 Ratio	Stdev	Gene Product
Extrachromosomal				
yfdG	b2350	-2.17	0.53	CPS-53 (KpLE1) prophage; bactoprenol-linked glucose translocase (flippase)
yfdI	b2352	-4.88	2.042	CPS-53 (KpLE1) prophage; predicted inner membrane protein
yffS	b2450	-2.06	0.28	CPZ-55 prophage; predicted protein
yfjH	b2623	-3.05	1.162	CP4-57 prophage; predicted protein
yfjU	b2638	-3.12	0.772	CP4-57 prophage; conserved protein
Unkonwn General Classification				
ais	b2252	2.93	0.91	conserved protein
b3022	b3022	-7.94	1.346	quorum-sensing regulator
*b3254**		3.79	0.988	
c1843 /// gapC		-3.53	0.511	
c4375 /// yiaG		-4.47	0.314	
csiD		-2.97	1.14	
csiE	b2535	-2.11	0.46	stationary phase inducible protein
elaB	b2266	-3.45	0.262	conserved protein
FPLM prot B		-4.08	0.843	
FPLT trb		-4.86	1.162	
hchA	b1967	-2.05	0.21	Hsp31 molecular chaperone
nudI /// yfaO	b2251	2.35	0.35	pyrimidine deoxynucleoside triphosphate hydrolase
pinH	b2648	3.92	1.579	predicted invertase fragment (pseudogene)
pptA	b1461	-2.06	0.50	4-oxalocrotonate tautomerase
ProtA		-6.45	0.392	
ProtE		-5.40	0.472	
sufA	b1684	-2.57	0.43	Fe-S cluster assembly protein
sufE	b1679	-2.44	0.41	sulfur acceptor protein
*trs5**		-4.43	0.536	
uspC /// yecG	b1895	2.34	0.32	universal stress protein
wecH		2.99	0.87	
yabQ	b0057	-8.65	1.425	C-terminal fragment of a predicted protein (pseudogene)
yafO	b0233	-3.68	1.318	predicted toxin of the YafO-YafN toxin-antitoxin system
yahO		-5.19	0.677	
yaiA	b0389	-2.17	0.49	predicted protein
ybgA	b0707	-2.61	0.50	conserved protein
*ycdF**		-3.49	0.337	
ycfH	b1100	-3.60	0.974	predicted metallodependent hydrolase
ydgA	b1614	-2.22	0.16	conserved protein

Table B3. Continued.

Gene Name	Blattner ID	Log_2 Ratio	Stdev	Gene Product
Unkonwn General Classification				
ydhQ	b1664	-2.43	0.20	conserved protein
ydiH	b1685	-2.03	0.29	predicted protein
yeaH	b1784	-3.86	0.281	conserved protein
yegS	b2086	-2.49	0.70	conserved protein
yehE	b2112	-2.08	0.52	predicted protein
yfbM	b2272	3.19	0.722	predicted protein
yfeS	b2420	-4.51	0.893	conserved protein
yghW	b2998	3.15	0.371	predicted protein
yhcE	b3217	-4.36	0.412	predicted protein, N-ter fragment (pseudogene)
yhcO	b3239	-2.48	0.61	predicted barnase inhibitor
yhdU	b3263	3.17	0.724	predicted membrane protein
yhhA	b3448	-2.29	0.22	conserved protein
yhhZ	b3442	3.66	0.949	conserved protein
yibA	b3594	2.30	0.88	lyase containing HEAT-repeat
yjdI	b4126	-2.37	0.28	conserved protein
yjdJ	b4127	-2.72	0.30	predicted acyltransferase with acyl-CoA N-acyltransferase domain
yjhC	b4280	-4.66	0.663	KpLE2 phage-like element; predicted oxidoreductase
ykgE	b0306	2.00	0.29	predicted oxidoreductase
ykgK	b0294	4.03	0.996	predicted regulator
ykgM	b0296	4.45	0.950	predicted ribosomal protein
ymgC	b1167	-2.33	0.57	predicted protein
ymgE	b1195	-2.89	0.91	predicted inner membrane protein
yneL	b1506	2.70	1.26	predicted transcriptional regulator
ynhG	b1678	-3.10	0.228	conserved protein
yqjD	b3098	-2.57	0.41	conserved protein

Appendix C

Table C1. The genes differentially expressed between *E. coli* K12 and E43 grown in LB broth and dLB broth.

Gene Name	Blattner ID	Log_2 Ratio-LB	Stdev* LB	Log_2 Ratio-dLB	Stdev dLB	Gene Product
Cell Process						
bfr	b3336	-6.49	0.539	-7.17	0.47	bacterioferritin, iron storage and detoxification protein
chpR	b2783	3.28	0.896	2.01	0.57	antitoxin of the ChpA-ChpR toxin-antitoxin system
osmC	b1482	-2.43	0.27	-2.60	0.40	osmotically inducible, stress-inducible membrane protein
osmY	b4376	-2.62	0.41	-3.47	0.33	periplasmic protein
slp	b3506	-3.87	0.664	-3.31	0.14	outer membrane lipoprotein
uspB / yhiO	b3494	-2.15	0.56	-3.71	0.21	predicted universal stress (ethanol tolerance) protein B
Cell Structure						
gfcA / ymcD	b0987	5.03	0.401	5.21	0.51	predicted protein
gfcE	b0983	3.84	0.663	2.88	0.71	predicted exopolysaccharide export protein
pgaB	b1023	3.48	1.346	3.01	0.55	putatative polysaccharide N-deacetylase/carbohydrate esterase
rfaB	b3628	-3.52	0.968	-5.11	1.51	UDP-D-galactose:(glucosyl)lipopolysaccharide-1,6-D-galactosyltransferase
rfaJ	b3626	-3.93	1.045	-4.02	1.08	UDP-D-glucose:(galactosyl)lipopolysaccharide glucosyltransferase
yaiV	b0375	2.02	1.02	2.63	0.86	predicted DNA-binding transcriptional regulator
Information Transfer						
dps	b0812	-3.80	0.615	-3.90	0.48	Fe-binding and storage protein
gadX	b3516	-2.12	0.74	-2.46	0.31	DNA-binding transcriptional dual regulator
hsdS	b4348	-3.89	0.539	-5.24	0.74	specificity determinant for hsdM and hsdR
relB	b1564	-5.33	1.087	-5.22	0.53	Qin prophage; bifunctional antitoxin of the RelE-RelB toxin-antitoxin system and transcriptional repressor
relE	b1563	-7.79	0.522	-7.59	0.20	Qin prophage; toxin of the RelE-RelB toxin-antitoxin system
sra	b1480	-3.25	0.401	-2.57	0.64	30S ribosomal subunit protein S22
Metabolism						
aldB	b3588	-2.17	0.35	-3.04	0.46	aldehyde dehydrogenase B
gapC_2	b1416	-2.14	0.60	-3.68	0.37	C-terminal fragment of glyceraldehyde-3-phosphate dehydrogenase C (pseudogene)
gatR_1	b2090	-2.03	0.40	-2.88	0.42	DNA-binding transcriptional repressor of galactitol utilization, N-ter fragment (pseudogene)
cmtB	b2934	2.24	0.56	2.72	0.21	PTS system putative mannitol-specific EIIA component
gadA	b3517	-4.07	1.081	-6.63	0.53	glutamate decarboxylase A, PLP-dependent
gadB	b1493	-6.62	0.852	-6.55	0.40	glutamate decarboxylase B, PLP-dependent
gatA	b2094	-6.21	0.238	-7.02	0.32	galactitol-specific enzyme IIA component of PTS

*** Standard Deviation calculated in Microsoft Excel.**

Table C1. Continued.

Gene Name	Blattner ID	Log_2 Ratio-LB	Stdev* LB	Log_2 Ratio-dLB	Stdev dLB	Gene Product
Metabolism						
gatB	b2093	-6.05	0.190	-7.12	0.71	galactitol-specific enzyme IIB component of PTS
gatC	b2092	-5.16	0.794	-6.07	1.13	galactitol-specific enzyme IIC component of PTS
gatD	b2091	-6.27	0.278	-7.39	0.93	galactitol-1-phosphate dehydrogenase, Zn-dependent and NAD(P)-binding
gatY	b2096	-6.04	0.470	-6.05	0.48	D-tagatose 1,6-bisphosphate aldolase 2, catalytic subunit
gatZ	b2095	-6.29	0.335	-7.31	0.53	D-tagatose 1,6-bisphosphate aldolase 2, subunit
glf	b2036	-6.52	1.232	-6.89	0.90	UDP-galactopyranose mutase, FAD/NAD(P)-binding
htrL	b3618	-4.28	1.139	-5.35	0.49	predicted protein
hyaA	b0972	-3.08	0.497	-4.29	0.45	hydrogenase 1, small subunit
hyaB	b0973	-2.07	0.49	-3.12	0.25	hydrogenase 1, large subunit
hyaC	b0974	-2.62	0.64	-3.73	0.11	hydrogenase 1, b-type cytochrome subunit
hyaD	b0975	-2.41	0.55	-3.80	0.35	protein involved in processing of HyaA and HyaB proteins
hyaE	b0976	-2.09	0.64	-3.61	0.24	protein involved in processing of HyaA and HyaB proteins
kgtP	b2587	-2.61	0.39	-2.35	0.83	alpha-ketoglutarate transporter
mcrB	b4346	-4.59	1.152	-3.13	0.59	5-methylcytosine-specific restriction enzyme McrBC, subunit McrB
mglA	b2149	2.35	0.56	2.21	0.52	fused methyl-galactoside transporter subunits of ABC superfamily: ATP-binding components
mglC	b2148	2.22	0.97	2.03	0.39	membrane component of an ABC superfamily methyl-galactoside transporter
otsB	b1897	-3.02	0.735	-4.35	0.43	trehalose-6-phosphate phosphatase, biosynthetic
psiF	b0384	-2.17	0.60	-2.27	0.44	conserved protein
relA	b2784	2.32	0.17	2.65	0.21	(p)ppGpp synthetase I/GTP pyrophosphokinase
rfaG	b3631	-3.70	0.649	-5.00	0.91	glucosyltransferase I
waaU	b3623	-6.80	1.685	-7.06	1.14	lipopolysaccharide core biosynthesis
rfaS	b3629	-5.10	1.198	-5.81	2.15	lipopolysaccharide core biosynthesis protein
rfaY	b3625	-4.41	1.908	-5.56	0.62	lipopolysaccharide core biosynthesis protein
rfaZ	b3624	-4.06	0.969	-4.87	0.58	lipopolysaccharide core biosynthesis protein
rfbA	b2039	-4.21	1.073	-5.66	1.52	glucose-1-phosphate thymidylyltransferase
rfbB	b2041	-4.31	0.496	-5.21	1.16	dTDP-glucose 4,6 dehydratase, NAD(P)-binding
rfbC	b2038	-5.38	1.227	-6.81	0.96	dTDP-4-deoxyrhamnose-3,5-epimerase
rfbX	b2037	-3.27	0.322	-2.99	0.89	predicted polisoprenol-linked O-antigen transporter
talA	b2464	-2.46	0.66	-3.37	0.38	transaldolase A

Table C1. Continued.

Gene Name	Blattner ID	Log_2 Ratio-LB	Stdev* LB	Log_2 Ratio-dLB	Stdev dLB	Gene Product
Metabolism						
wbbH	b2035	-5.11	0.998	-6.90	0.89	O-antigen polymerase
wbbJ	b2033	-6.76	0.677	-5.68	0.82	predicted acyl transferase
wrbA	b1004	-2.37	0.54	-2.21	0.18	predicted flavoprotein in Trp regulation
ybaS	b0485	-2.85	0.33	-3.43	0.46	predicted glutaminase
ybaT	b0486	-2.37	0.42	-2.91	0.23	predicted transporter
Regulation						
bfd	b3337	-2.27	0.47	-4.79	0.59	bacterioferritin-associated ferredoxin
bolA	b0435	-2.38	0.32	-3.14	0.26	regulator of penicillin binding proteins and beta lactamase transcription (morphogene)
yiaG	b3555	-3.10	0.653	-5.62	0.53	predicted transcriptional regulator
ygiT	b3021	-5.20	1.265	-5.49	0.81	predicted DNA-binding transcriptional regulator
gadE	b3512	-4.62	1.078	-5.50	1.11	acid-induced positive regulator of glutamate-dependent acid resistance
yjhU	b4295	-4.25	0.832	-3.55	0.46	KpLE2 phage-like element; predicted DNA-binding transcriptional regulator
Trasnport						
ompC	b2215	-2.61	0.61	-3.66	0.41	outer membrane porin protein C
ydcS	b1440	-2.85	0.84	-4.24	1.08	predicted spermidine/putrescine transporter subunit -!- periplasmic-binding component of ABC superfamily
ydcV	b1443	-2.39	0.59	-2.84	0.72	membrane component of an ABC superfamily predicted spermidine/putrescine transporter
Extrachromosomal						
borD	b0557	2.31	0.66	3.17	1.09	DLP12 prophage; predicted lipoprotein
ybcY	b0562	3.79	0.975	4.44	0.46	DLP12 prophage; predicted SAM-dependent methyltransferase
dicA	b1570	-5.58	0.662	-5.24	0.53	Qin prophage; predicted regulator for DicB
intA	b2622	-4.91	0.844	-4.61	1.70	CP4-57 prophage; integrase
intB	b4271	-2.63	1.14	-4.09	1.07	KpLE2 phage-like element; predicted integrase
intF	b0281	-4.84	0.941	-3.50	0.49	CP4-6 prophage; predicted phage integrase
nmpC	b0553	-6.82	0.872	-5.91	0.77	DLP12 prophage; truncated outer membrane porin (pseudogene)
racR	b1356	-5.93	0.787	-6.35	0.51	Rac prophage; predicted DNA-binding transcriptional regulator
yagN	b0280	-6.13	1.420	-6.84	0.80	CP4-6 prophage; predicted protein
ybcM	b0546	-4.03	1.007	-4.57	1.90	DLP12 prophage; predicted DNA-binding transcriptional regulator
ybcW	b0559	3.71	1.610	4.16	0.50	DLP12 prophage; predicted protein
yfdH	b2351	-5.97	0.422	-5.41	0.69	CPS-53 (KpLE1) prophage; bactoprenol glucosyl transferase
yfdQ	b2360	2.82	1.02	3.89	0.76	CPS-53 (KpLE1) prophage; predicted protein
yffL	b2443	-3.76	0.789	-3.57	1.28	CPZ-55 prophage; predicted protein

Table C1. Continued.

Gene Name	Blattner ID	Log_2 Ratio-LB	Stdev* LB	Log_2 Ratio-dLB	Stdev dLB	Gene Product
Extrachromosomal						
yffR	b2449	-2.66	0.88	-3.70	0.63	CPZ-55 prophage; predicted protein
yfjO	b2631	-5.22	0.566	-5.09	1.46	CP4-57 prophage; predicted protein
yfjW	b2642	-5.30	0.570	-6.14	1.43	CP4-57 prophage; predicted inner membrane protein
ykfB	b0250	-4.37	0.502	-4.67	0.54	CP4-6 prophage; predicted protein
Unknown General Classification						
yabQ / yabP_2	b0057	-7.44	1.074	-8.43	1.93	C-terminal fragment of a predicted protein (pseudogene)
yafN	b0232	-6.37	0.660	-6.35	1.38	predicted antitoxin of the YafO-YafN toxin-antitoxin system
ybdK	b0581	-3.14	0.782	-2.03	0.36	gamma-glutamyl:cysteine ligase
ybgS	b0753	-2.49	0.67	-2.19	0.36	conserved protein
gfcD	b0984	3.59	0.547			conserved protein
gfcB	b0986	3.09	0.330	3.62	1.49	predicted outer membrane lipoprotein
yccJ	b1003	-2.20	0.68	-2.00	0.27	predicted protein
ycgB	b1188	-2.09	0.32	-3.58	0.44	conserved protein
ydiZ	b1724	-2.04	0.53	-2.38	0.37	predicted protein
yeaG	b1783	-3.04	0.393	-3.83	0.49	conserved protein with nucleoside triphosphate hydrolase domain
yeaQ	b1795	-2.15	0.73	-2.12	0.30	conserved inner membrane protein
yebV	b1836	-3.37	0.601	-4.23	0.33	predicted protein
yodD	b1953	-3.84	1.164	-4.89	0.44	predicted protein
yodC	b1957	-3.10	0.699	-3.13	0.57	predicted protein
wbbK	b2032	-9.38	0.572	-8.32	0.47	lipopolysaccharide biosynthesis protein
wbbI	b2034	-7.73	0.442	-7.61	0.81	conserved protein
yegP	b2080	-3.19	0.438	-3.83	0.37	predicted protein
elaB	b2266	-3.07	0.529	-3.93	0.65	conserved protein
yfcO	b2332	-4.15	0.743	-5.08	0.53	predicted protein
ygaF	b2660	-2.67	1.00	-2.74	0.77	predicted enzyme
ygaU	b2665	-2.61	0.62	-3.17	0.39	predicted protein
ygaM	b2672	-2.52	0.52	-2.68	0.48	predicted protein
mqsR	b3022	-7.22	1.313	-7.26	0.72	quorum-sensing regulator
ygiW	b3024	-3.72	0.477	-2.60	0.28	conserved protein
yqjC	b3097	-2.23	0.48	-2.56	0.09	conserved protein
yqjE	b3099	-2.00	0.39	-2.50	0.31	conserved inner membrane protein
yhfG	b3362	-3.64	1.507	-4.09	0.57	predicted protein
hdeB	b3509	-5.23	0.913	-4.68	0.79	acid-resistance protein
hdeA	b3510	-4.90	0.550	-4.51	0.74	stress response protein acid-resistance protein
hdeD	b3511	-3.52	0.665	-5.02	0.47	acid-resistance membrane protein
yigB	b3812	3.44	0.461	3.09	0.47	FMN phosphatase
yjbJ	b4045	-3.41	0.542	-4.38	0.66	predicted stress response protein
yjdN	b4107	-2.35	0.94	-4.02	0.45	conserved protein

Table C2. The genes differentially expressed between *E. coli* K12 and E43 grown in LB broth onlv.

Gene Name	Blattner ID	Log_2 Ratio LB	Stdev* LB	Gene Product
Cell process				
gidB	b3740	2.82	0.58	methyltransferase, SAM-dependent methyltransferase, glucose-inhibited cell-division protein
osmB	b1283	-2.12	0.35	lipoprotein
Cell structure				
yiaA	b3562	2.02	1.02	conserved inner membrane protein
DNA Sites				
yibJ	b3595	2.68	0.62	predicted Rhs-family protein
Information Transfer				rhsB_b3482_at
alaT	b3853	-2.29	0.96	tRNA-Ala(UGC)
etp	b0982	3.30	0.770	phosphotyrosine-protein phosphatase
osmE	b1739	-2.99	0.58	DNA-binding transcriptional activator
pspA	b1304	-2.01	0.45	regulatory protein for phage-shock-protein operon
pspC	b1306	-2.20	0.70	DNA-binding transcriptional activator
Metabolism				
chiA	b3338	-3.40	0.527	periplasmic endochitinase
csrA	b2696	-2.14	0.72	post transcriptional regulatory RNA-binding protein
dcuB	b4123	3.08	0.611	C4-dicarboxylate antiporter
fbaB	b2097	-3.33	0.698	fructose-bisphosphate aldolase class I
oppB	b1244	-2.31	0.84	oligopeptide transporter subunit; membrane component of ABC superfamily
rbsD	b3748	-2.09	0.25	predicted cytoplasmic sugar-binding protein
rpiR	b4089	-3.93	1.211	DNA-binding transcriptional repressor
Regulation				
yeiL	b2163	2.00	1.07	DNA-binding transcriptional activator of stationary phase nitrogen survival
Transport				
msyB	b1051	-2.01	0.42	predicted protein
insD-2	b1402	-2.29	1.20	IS2 insertion element transposase InsAB'
Extrachromosomal				
yagJ	b0276	-3.88	0.451	CP4-6 prophage; predicted protein
rzpD	b0556	2.10	1.10	DLP12 prophage; predicted murein endopeptidase
pinR	b1374	2.09	0.67	Rac prophage; predicted site-specific recombinase
cspF	b1558	2.33	0.81	Qin prophage; cold shock protein
yeeR	b2001	2.53	0.83	CP4-44 prophage; predicted membrane protein
yfjN	b2630	-3.49	0.391	CP4-57 prophage; RNase LS

*** Standard Deviation calculated in Microsoft Excel.**

Table C2. contiued

Gene Name	Blattner ID	Log_2 Ratio LB	Stdev* LB	Gene Product
Unknown General Classification				
yabP_1	b0056	-3.24	0.778	N-terminal fragment of a predicted protein (pseudogene)
gfcC	b0985	3.74	0.901	conserved protein
ybjP	b0865	-3.14	0.782	predicted lipoprotein
yodA	b1973	3.55	0.930	conserved metal-binding protein
yiaH	b3561	3.43	0.804	conserved inner membrane protein
wbbL	b4571	-7.41	0.954	
ygdI	b2809	-2.31	0.26	predicted protein
yqgC	b2940	2.00	0.53	predicted protein
yibG	b3596	3.52	1.373	conserved protein
yiiG	b3896	2.09	0.57	conserved protein

Table C3. Genes differentially expressed between *E. coli* K12 and E43 grown only in dLB.

Gene Name	Blattner ID	Log_2 Ratio	Stdev*	Gene Product
Cell Process				
aidB	b4187	-3.44	0.26	isovaleryl CoA dehydrogenase
fic	b3361	-3.13	0.23	stationary-phase protein, cell division
ftnA	b1905	2.77	0.42	ferritin iron storage protein (cytoplasmic)
yecI	b1902	2.98	0.28	predicted ferritin-like protein
ygeH	b2852	4.05	0.61	predictedtranscriptional regulator
Cell Structure				
mdtE	b3513	-2.12	0.39	multidrug resistance efflux transporter
rfaI	b3627	-4.09	1.62	UDP-D-galactose:(glucosyl)lipopolysaccharide-alpha-1,3-D-galactosyltransferase
yaiV	b0375	2.63	0.86	predicted DNA-binding transcriptional regulator
yedS_1	b1964	-5.17	0.98	predicted protein, N-ter fragment (pseudogene)
yehD	b2111	2.02	0.31	predicted fimbrial-like adhesin protein
Information Transfer				
appY	b0564	-5.10	1.94	DLP12 prophage; DNA-binding transcriptional activator
csgD	b1040	-4.02	0.49	DNA-binding transcriptional regulator of adhesion determinants
fhuF	b4367	-2.20	0.86	ferric iron reductase involved in ferric hydroximate transport
gspB	b3322	-2.96	1.26	part of gsp divergon involved in type II protein secretion
sufB	b1683	-2.94	0.27	component of SufBCD complex
yehC	b2110	2.10	0.80	predicted periplasmic pilin chaperone
yehU	b2126	2.18	0.31	predicted sensory kinase in two-component system with YehT
yihW	b3884	-2.36	0.51	predicted DNA-binding transcriptional regulator
Metabolism				
adhP	b1478	-2.99	0.35	alcohol dehydrogenase, 1-propanol preferring
allD	b0517	2.60	0.82	ureidoglycolate dehydrogenase
appB	b0979	-2.00	0.52	cytochrome bd-II oxidase, subunit II
appC	b0978	-2.40	0.28	cytochrome bd-II oxidase, subunit I
cfa	b1661	-2.18	0.36	cyclopropane fatty acyl phospholipid synthase (unsaturated-phospholipid methyltransferase)
csgA	b1042	-3.34	0.44	cryptic curlin major subunit
csgB	b1041	-3.58	0.42	curlin nucleator protein, minor subunit in curli complex
cysH	b2762	-2.50	0.54	3'-phosphoadenosine 5'-phosphosulfate reductase
cysM	b2421	-2.01	0.31	cysteine synthase B (O-acetylserine sulfhydrolase B)
cysU	b2424	-2.30	0.37	sulfate/thiosulfate transporter subunit; membrane component of ABC superfamily
fliC	b1923	-3.55	0.56	flagellar filament structural protein (flagellin)
frc	b2374	2.58	0.45	formyl-CoA transferase, NAD(P)-binding
gabD	b2661	-2.26	0.69	succinate-semialdehyde dehydrogenase I, NADP-dependent
gabT	b2662	-2.67	0.79	4-aminobutyrate aminotransferase, PLP-dependent
glcC	b2980	2.15	0.56	DNA-binding transcriptional dual regulator, glycolate-binding
glnA	b3870	2.84	0.70	glutamine synthetase
glnK	b0450	2.85	1.09	nitrogen assimilation regulatory protein for GlnL, GlnE, and AmtB

*** Standard Deviation calculated in Microsoft Excel**

Table C3. Continued.

Gene Name	Blattner ID	Log_2 Ratio	Stdev*	Gene Product
Metabolism				
gltK	b0653	2.54	0.25	membrane component of an ABC superfamily glutamate/aspartate transporter
hyaF	b0977	-3.04	0.27	protein involved in nickel incorporation into hydrogenase-1 proteins
hybF	b2991	2.18	0.58	protein involved with the maturation of hydrogenases 1 and 2
sucA	b0726	2.39	0.64	2-oxoglutarate decarboxylase, thiamin-requiring
mglB	b2150	2.11	0.45	periplasmic-binding component of an ABC superfamily methyl-galactoside transporter
napA	b2206	3.87	0.28	nitrate reductase, periplasmic, large subunit
napD	b2207	3.07	0.70	assembly protein for periplasmic nitrate reductase
napF	b2208	2.40	0.45	ferredoxin-type protein, predicted role in electron transfer to periplasmic nitrate reductase (NapA)
ndk	b2518	2.00	0.31	multifunctional nucleoside diphosphate kinase and apyrimidinic endonuclease and 3'-phosphodiesterase
phoA	b0383	2.14	0.59	bacterial alkaline phosphatase
phr	b0708	-2.10	0.57	deoxyribodipyrimidine photolyase, FAD-binding
poxB	b0871	-2.31	0.50	pyruvate dehydrogenase (pyruvate oxidase), thiamin-dependent, FAD-binding
prpB	b0331	-3.16	0.39	2-methylisocitrate lyase
prpC	b0333	-2.06	0.32	2-methylcitrate synthase
prpD	b0334	-2.96	1.08	2-methylcitrate dehydratase
rfaL	b3622	-4.83	2.12	O-antigen ligase
rfaQ	b3632	-4.80	1.31	lipopolysaccharide core biosynthesis protein
sdhB	b0724	2.65	0.51	succinate dehydrogenase, FeS subunit
speA	b2938	-2.13	0.42	biosynthetic arginine decarboxylase, PLP-binding
sucB	b0727	2.57	0.74	dihydrolipoyltranssuccinase
sufS	b1680	-2.87	0.39	selenocysteine lyase, PLP-dependent
tktB	b2465	-3.13	0.40	transketolase 2, thiamin-binding
tnaA	b3708	2.96	0.48	tryptophanase/L-cysteine desulfhydrase, PLP-dependent
tnaC	b3707	3.32	0.60	tryptophanase leader peptide
yajG	b0434	2.02	0.48	predicted lipoprotein
ycaC	b0897	-4.00	0.35	predicted hydrolase
ydeN	b1498	2.33	0.43	conserved protein
ygjG	b3073	-2.62	0.49	putrescine:2-oxoglutaric acid aminotransferase, PLP-dependent
yodB	b1974	3.79	0.81	predicted cytochrome
Regulation				
arnB	b2253	2.60	0.67	UDP-4-amino-4-deoxy-L-arabinose alpha-ketoglutarate aminotransferase
yddM	b1477	-2.58	0.37	predicted DNA-binding transcriptional regulator
yehT	b2125	2.18	0.40	predicted response regulator in two-component system withYehU
yfeD	b2399	2.01	0.30	predicted DNA-binding transcriptional regulator

Table C3. Continued.

Gene Name	Blattner ID	Log2 Ratio	Stdev*	Gene Product
Transport				
csgE	b1039	-4.77	1.05	predicted transport protein
csgF	b1038	-3.78	0.42	predicted transport protein
csgG	b1037	-2.82	0.35	outer membrane channel lipoprotein
mdtL	b3710	2.06	0.88	multidrug efflux system protein
mntH	b2392	-2.56	0.35	proton-dependent divalent metal cation transporter
nhaA	b0019	-2.56	0.50	sodium-proton antiporter
sufC	b1682	-2.67	0.53	transport protein associated with Fe-S cluster assembly
trkG	b1363	-2.76	1.29	Rac prophage; potassium transporter subunit
gadC	b1492	-5.64	0.14	predicted glutamate:gamma-aminobutyric acid antiporter
yahN	b0328	2.37	0.73	neutral amino-acid efflux system
ydcU	b1442	-2.04	0.46	membrane component of an ABC superfamily predicted spermidine/putrescine transporter
yhiD	b3508	-3.53	0.80	predicted Mg(2+) transport ATPase inner membrane protein
yhiM	b3491	-2.83	0.40	conserved inner membrane protein
yicE	b3654	2.29	0.88	predicted transporter
yjhB	b4279	-5.59	0.83	KpLE2 phage-like element; predicted transporter
yohC	b2135	-3.54	0.23	predicted inner membrane protein
Extrachromosomal				
insL-1	b0016	-4.75	0.76	IS186/IS421 transposase
intZ	b2442	-3.75	1.18	CPZ-55 prophage; predicted integrase
pspE	b1308	2.08	0.30	thiosulfate:cyanide sulfurtransferase
yfdP	b2359	2.10	0.47	CPS-53 (KpLE1) prophage; predicted protein
yfjU	b2638	-4.51	1.22	CP4-57 prophage; conserved protein
Unknown General Classification				
yafU	b0218	2.60	1.02	predicted inner membrane protein (pseudogene)
iraP	b0382	2.05	0.63	anti-adaptor protein
ybaA	b0456	-2.41	0.42	conserved protein
ybdK	b0581	-2.03	0.36	gamma-glutamyl:cysteine ligase
ybgA	b0707	-2.49	0.60	conserved protein
ycfH	b1100	-3.99	1.30	predicted metallodependent hydrolase
yciI	b1251	2.02	0.42	predicted enzyme
yciG	b1259	-2.83	0.53	predicted protein
ydaM	b1341	-2.01	0.31	predicted diguanylate cyclase, GGDEF domain signalling protein
ydcH	b1426	2.16	0.62	predicted protein
ydhV	b1673	2.47	0.57	predicted oxidoreductase
ynhG	b1678	-2.02	0.38	conserved protein
sufA	b1684	-2.17	0.41	Fe-S cluster assembly protein
yeaH	b1784	-3.07	0.26	conserved protein
yeaK	b1787	3.41	0.11	conserved protein
yedF	b1930	2.09	0.55	conserved protein
hchA	b1967	-2.66	0.39	Hsp31 molecular chaperone

Table C3. Continued.

Gene Name	Blattner ID	Log2 Ratio	Stdev*	Gene Product
Unknown General Classification				
wbbL_1	b2031	-7.52	0.54	lipopolysaccharide biosynthesis protein, N-ter fragment (pseudogene)
yegS	b2086	-2.80	0.36	conserved protein
arnC	b2254	2.20	0.21	Undecaprenyl-phosphate 4-amino-4-deoxy-L-arabinose transferase
yfbM	b2272	2.26	0.61	predicted protein
yfeS	b2420	-4.56	0.85	conserved protein
yqeF	b2844	2.45	0.52	predicted acyltransferase
yqjD	b3098	-2.40	0.12	conserved protein
yhcE_1	b3217	-4.02	1.05	predicted protein, N-ter fragment (pseudogene)
yhcO	b3239	-3.77	0.55	predicted barnase inhibitor
yhdU	b3263	3.44	0.91	predicted membrane protein
yheN	b3345	-3.03	0.95	tRNA 2-thiouridine synthesizing protein, subunit of sulfur transfer complex, TusD
cbrB	b3716	-3.60	0.27	predicted inner membrane protein
yjeI	b4144	2.11	0.42	conserved protein
yjhC	b4280	-4.42	0.65	KpLE2 phage-like element; predicted oxidoreductase
		-4.96	0.38	
		-4.53	0.24	
yabP		-4.20	0.42	predicted protein (pseudogene)
		-2.64	0.52	
		-2.35	0.37	
		-2.06	0.30	
		-4.12	0.13	
		3.01	0.58	
		-2.00	0.16	
		2.34	0.82	
		2.00	0.49	
		-7.03	0.50	
		-6.14	0.30	
		-3.76	1.01	
		-4.37	0.86	
		-2.03	0.23	

Blank cells indicate no information available for that gene.